Available online at www.sciencedirect.com
ScienceDirect
Forest Ecology and Management 255 (2008) 2709–2717
Forest Ecology and Management
www.elsevier.com/locate/foreco

Comparison of treewise and standwise forest simulators by means of quantile regression

Antti Mäkinen *, Annika Kangas, Jouni Kalliovirta, Jussi Rasinmäki, Esko Välimäki

Department of Forest Resource Management, University of Helsinki, Latokartanonkaari 11, 00014 University of Helsinki, Finland

Received 7 September 2007; received in revised form 3 January 2008; accepted 15 January 2008

Abstract

Predicting forest development under varying treatment schedules forms the basis of forest management planning. The actual growth predictions are made with a forest simulator which includes growth equations and additional models for predicting a number of varying tree, forest and site properties. Forest growth simulators typically include either tree-level or stand-level growth models, but these two approaches have not been thoroughly compared. We set out here to compare these two approaches with the SIMO simulator framework in a small data set from southern Finland based on 60 sample plots in 30 stands, the development of which was known for 20 years. The stands chosen were very dense, so that the simulators could be tested under extreme conditions. The results show that the stand-level model is more accurate in almost all cases and its computational burden is much lower. It could therefore be advisable to use tree-level models for short-term predictions, which would ensure detailed information on forest structure for planning the near-future operations. Stand-level models would be more advisable in longer term predictions, especially when accurate volume estimates are considered more important than the forest structure. The errors observed in these simulators were analysed further by quantile regression, which allows empirical estimates of confidence intervals to be obtained for the simulator.

Keywords: Growth and yield models; Forest management planning

1. Introduction

Predictions of forest development based on the use of forest growth models of different types form the foundations for decision-making in forest management planning. Inaccurate or erroneous predictions of the future state of a forest holding can lead to unfavourable decisions, which can reduce the possible benefits for the forest owner. In order to make good decisions, one needs to have reliable, accurate information on the future situation.

In practice forest growth models are used for: 1. updating previously measured data sets in accordance with the present state and 2. predicting future forest development, in order to evaluate silvicultural treatments, management planning and harvest scheduling (Burkhart, 1993). The actual growth predictions are made with a forest simulator which includes growth equations and additional models for predicting a number of varying tree, forest and site properties.

Various types of forest growth model have been proposed, and they can basically be either empirical models, which means that they are estimated statistically from a measured data set, or mechanical process models, which are based on ecological theories and describe the eco-physiological processes of individual trees in detail. Growth models can also be categorized by their level of organization, which is usually that of either a single tree or a stand (Munro, 1974). Tree-level models predict the growth of an individual tree, and stand-level model predict the increment in an aggregate variable such as mean diameter or basal area. Tree-level models can then be further categorized as spatial (*distance-dependent*) or aspatial (*distance-independent*). Distance-dependent models use spatial indices, i.e. information about neighbouring trees and their locations, when predicting the growth of a single tree, whereas distance-independent models do not (Tomé and Burkhart, 1989; Vettenranta, 1999).

Other types of growth models include diameter distribution-based models (e.g. Bailey et al., 1981) and transition matrix models (e.g. Buongiorno and Mitchie, 1980; Kolström, 1993).

The tree-level growth models that have been used in Finland are mostly empirical and distance-independent, as it has been too expensive to acquire spatial information on the forest

* Corresponding author. Tel.: +358 40 583 6162.
E-mail address: antti.makinen@helsinki.fi (A. Mäkinen).

0378-1127/$ – see front matter
doi:10.1016/j.foreco.2008.01.048

structure for practical forest management planning purposes. The most up-to-date tree-level growth models in Finland, those presented by Hynynen et al. (2002), are based on extensive data representing all the main tree species and forest sites in the country. The history of stand-level growth models in Finland goes back to the yield tables introduced in the 19th century. Models of this type have been presented by Vuokila and Väliaho (1980) for pine (*Pinus sylvestris* L.) and spruce (*Picea abies*) and by Oikarinen (1983) and Saramäki (1977) for birches (*Betula pendula, Betula pubescens*). Other stand-level growth models for Finnish conditions include the models presented by Gustavsen (1977), Nyyssönen and Mielikäinen (1978) and Mielikäinen (1985). The old stand-level models have typically been estimated with a group of independent models, but in newer applications, simultaneous equations based on 3SLS, for instance, are used (e.g. Eerikäinen, 2002).

Tree-level growth models have replaced the stand-level models already in 1980s in Finland in practical forest planning The trend is towards single-tree level models also in other countries (e.g. HEUREKA, T and SILVA simulators developed in Sweden, Norway and Germany in recent years). Generally the concept of a tree-level model is considered to be simpler than that of a stand-level model, as an individual tree is easy to comprehend as a functional object. Tree-level growth models are capable of taking account of tree-level competition better than stand-level models, and they are needed when operating in uneven-aged or mixed-species forests (Garcia, 2001; Porté and Bartelink, 2002).

As tree-level growth models operate with individual trees, the stand has to be represented as a set of trees, and if the individual trees have not been measured, a list of them can be predicted from stand-level mean variables or laser scanning data by means of a theoretical size distribution model (e.g. Maltamo, 1997; Gobakken and Næsset, 2005). The number of additional models that are needed to predict the set of trees and all the independent variables for tree-level growth models is quite substantial. Although tree-level models can capture inter-tree processes in detail, the interactions between individual trees in a stand can be so complicated that tree-level models cannot necessarily take them all into account (Zeide, 1993). The sample plot size affects the measures of competition in tree-level growth models (Hynynen and Ojansuu, 2003) which can result in unexpected behaviour of the models and lead to an accumulation of errors, especially under extreme conditions, e.g. in very old or dense forests.

Stand-level models are usually simple, the computational cost is substantially lower than with tree-level models and they can predict forest growth with sufficient detail for many applications (Vanclay, 1995; Atta-Boateng and Moser, 2000). As the forests in the stand-level approach are presented by a number of aggregate mean variables, the heterogeneity inside stands cannot be assessed. One way to assess the stand's inner structure is to utilize diameter distribution models for predicting some of the stand variables. An argument in favour of stand-level models is that they are generally seen as being more stable, especially under extreme conditions. Even though it has been shown that tree-level growth models produce very good projections of forest development, stand-level models could provide a good complement for them under certain conditions (Garcia, 2001). It has also been noted that tree-level models may not be as accurate as stand-level models (Burkhart, 2003).

According to Vanclay and Skovsgaard (1997), the evaluation of a growth model should include at least evaluations of its logical and biological foundation, evaluation of its statistical properties, error characteristics and residuals, together with a sensitivity analysis. Gustavsen (1988, p. 151), in his comparison of stand-level growth models with the previous version of the MELA simulator based on tree-level models, found that the stand-level models were in general more accurate (RMSE% of volume 31.8 vs. 39 for tree-level models), but that the tree-level model was more accurate in some sub-regions. Shortt and Burkhart (1996) reported that merchantable volume projections with tree-level models were more accurate at short projection periods (3–6 years), but the stand-level model projections were more accurate at longer periods (9 years). There has been no extensive comparison of the performance of the finnish tree-level and stand-level growth models to date, however.

The aims of this study were (I) to examine the differences between predictions made with tree-level and stand-level growth models for a 20-year time period in dense sample plots, and (II) to test the use of quantile regression analysis for examining growth model prediction errors.

2. Materials and methods

2.1. Simulations

The simulations for study were produced using both tree-level and stand-level growth models for a time period of 20 years with reference data from sample plots. The simulators used were the SIMO simulation framework (Tokola et al., 2006) and the MOTTI stand simulator (Hynynen et al., 2002). SIMO (SIMulation and Optimization for next-generation forest planning), which was used as the test bench for the growth predictions obtained with both tree-level and stand-level growth models, is a flexible, adaptable and extendable simulation framework that has been developed at the University of Helsinki (Tokola et al., 2006; Rasinmäki et al., 2007). The term simulation framework is used here rather than simulator as SIMO works as a platform for implementing different forest simulators.

The framework consists of programmatic components which form the application itself, an XML (eXtensible Markup Language, McGrath, 2003)-based syntax that is used to define the simulation logic and an extendable model base which includes all the models and equations used in the simulations. The main idea behind the framework is that different types of simulator can be implemented on the same platform without the need for tedious programming. The simulation logic can be defined without any programming, and the extendable model base makes it very easy to re-use existing models. So far the SIMO model base includes approximately 400 empirical models describing forest properties and growth and yields at both the individual tree and stand level. The model base also includes a number of forest operation models, including thinnings, final harvests and other operations.

Two different simulators had been implemented in the SIMO framework and were used for the present purpose: a tree-level simulator and a stand-level simulator. These shared certain generic models, but the growth models were aspatial individual tree models for the tree-level simulator and stand models for the stand-level simulator. The forest management scenarios including different forest operations were disabled for the present purpose in the case of both simulators.

2.1.1. Tree-level simulator implementation

The growth and yield models that were used in the tree-level simulator were mainly the same as in the MOTTI and MELA simulators (Hynynen et al., 2002) and the simulator included growth models for all Finland's main tree species and forest types. As a large portion of these forests grow on peatlands, there are separate growth models for trees growing on mineral soil and peatlands. This group included a total of 21 individual growth models.

The dependent variables in these models were growth in height and basal area, while the independent variables included a substantial number describing the characteristics of a single tree, the stand, the geographical area and the competition factors, e.g. diameter and height of the tree, dominant diameter of the stand, growth in dominant height, crown ratio, dominant growth ratio, relative density factor, latitude of the stand, site class and site index.

As the tree-level models operate on individual trees, the trees have to be measured in the field or predicted with a distribution model. The tree-level simulator can construct trees by using a diameter or height distribution, but in this study the trees measured in the field were used to construct tree lists for the simulator. As the tree-level growth model predicts growth and new values for the individual tree variables, new stand level variables were aggregated from the tree list.

2.1.2. Stand-level simulator implementation

The growth models for pine and spruce used in the stand-level simulator were those of Vuokila and Väliaho (1980) and the growth models for birches those of Oikarinen (1983) and Saramäki (1977). These stand-level models included a number of individual regional models, as growth conditions vary across Finland. The models predicted the growth in a number of variables, including the basal area of the stand, basal area under bark, volume or dominant height, depending on the model and tree-species. The independent variables in these stand-level models included basal area, basal area under bark, stand age, site class, dominant height, length of growth period, stand volume, number of trees, mean diameter at breast height, temperature sum, etc. Other variables, some of which were used in the analyses of this study, were predicted from the growth models' results. For example, the mean height was predicted from dominant height values and mean diameter was predicted from mean height, mean age, temperature sum and site class in the predictions. In the beginning, mean height and diameter were assumed to be known. Although the growth of the stand in SIMO stand simulator was predicted with pure stand level models, the number of stems was calculated with a model that uses diameter distribution. However, the SIMO stand simulator can be considered a pure stand-level simulator as the growth is predicted at stand level. On average, a single stand-level model has 2 or 3 independent variables. This is quite a small number, and the variables have traditionally been fairly easy to measure or estimate. The total number of individual growth models in this group was 43.

2.1.3. MOTTI stand simulator

As a reference for the simulator implementations on the SIMO platform, the growth of the sample plots was also simulated with the MOTTI simulator (version 1.1), which is a well-established and validated forest simulator for Finnish conditions (Hynynen et al., 2005; Salminen et al., 2005). MOTTI is a stand level decision support tool which can be used to evaluate different forest management scenarios. Forest growth is evaluated with distance-independent tree-level models that predict the growth in tree diameter and height for a 5-year period. The models used in the core of the MOTTI simulator, described in detail by Hynynen et al. (2002), are based on an extensive, representative data set from across Finland, so that it is able to produce a reliable reference simulation for the present purposes. The MOTTI simulator was used mainly to validate the SIMO tree-level simulator implementation, and the main emphasis in the analysis was on comparing the tree-level and stand-level growth models when implemented in the SIMO framework.

2.2. Reference data

The reference data included 60 sample plots from 30 stands in central Finland measured in 2005. The stands were selected so that all of them were relatively dense compared with a normal commercially managed forest. Such a data set was selected in order to analyse the models under extreme conditions, where their accuracy has been questioned by users. Thus, the results do not provide unbiased estimate of RMSE in average Finnish forests, but they do give further insight to the behaviour of the models with respect to initial basal area. Another requisite for the selected stands was that there should not have been any human impact such as thinnings or other forestry operations during the past 20 years. Both spruce and pine-dominated stands were well represented, as also were mixed stands. Different age classes and site types were also well represented in the data.

The sample plots were divided into two types: sample tree plots and tally tree plots. All of the sample plots were circular, with a radius depending on the density of the stand (stems/ha), so that the maximum radius was 15.45 m and the minimum 4.89 m. Both the sample and tally tree plots had the same centre point, but the radius of a sample tree plot was half of that of a tally tree plot. The trees in the plots were labelled as sample trees and tally trees, respectively.

The total number of trees in the sample and tally plots was 2070, of which 490 were sample trees. The tally trees were measured for dbh (diameter at breast height) and their species was recorded. In the case of the sample trees the height, crown height, thickness of bark and health of the tree were also recorded. Every sample tree that was either pine or spruce was also cored at breast height to determine its exact age and to reconstruct its actual past growth. By analysing the annual rings in the core, the development of the tree's dbh during the past 20 years could be measured accurately. Plot level mean variables were calculated for the sample trees (Table 1). Dead trees were also measured and their time of death estimated, in order to form a reliable description of the mortality, i.e. decrease in the number of trees, in the dense stands.

The preprocessing of the reference data included predicting the heights, diameters and ages of the tally trees for the present time and for the past 20 years. This was done so that all of the trees, both tally and sample trees, could be used as input data in the tree-level simulations. For this purpose a multitude of mixed linear models were constructed and applied to produce the missing values. These models have been presented in more detail by Välimäki (2006). After this pre-processing, the reference data included 60 sample plots with 2070 trees for which the diameter at breast height (dbh), height (*h*), age (*a*) and thickness of the bark (t_b) were measured or predicted for the present time and for points in time 5, 10, 15 and 20 years ago. Stand-level aggregate values were computed from these tree-level variables for the same points in time.

2.3. Analysis of growth model performance

The performance of the tree-level and stand-level growth models was assessed by examining the differences between the simulated growth predictions and the reference data at different points in time during the 20-year simulation period. The analysis took place at sample plot level and the aggregate variables that were used in the comparisons were basal area per hectare (BA), number of trees per hectare (*N*), mean diameter at breast height (D_{MEAN}) and mean height (H_{MEAN}). As the stand level values in the reference data were considered correct, the prediction error for the simulated variables was ΔY, where *Y* is one of the variables of interest (Eq. (1)).

$$\Delta Y = Y_{\mathrm{REFERENCE}} - Y_{\mathrm{SIMULATED}}. \quad (1)$$

The prediction error ΔY was analysed statistically with basic statistical measures including the mean and standard deviation (S.D.). The prediction error ΔY was also modelled with quantile regression analysis. Regression analysis is usually employed for investigating possible relationships between variables, but Koenker and Bassett (1978) introduced an extension, known as quantile regression, which is a method for estimating the conditional quantiles of the distribution of a dependent variable in a linear regression model. Normal regression models can be problematic in cases where the variance is heterogeneous, as they may imply that there is no predictive relationship between the dependent and independent variables. There may actually be a close relationship in some parts of the distribution of the dependent variable, however, but normal regression can only provide models for the conditional mean function derived from the total distribution of the dependent variable. With quantile regression one can obtain models for the whole range of conditional quantile functions (Koenker and Hallock, 2001).

Table 1
Mean values for plot-level variables aggregated from the sample trees in the reference data

	Mean	Standard deviation	Min	Max
Basal area (m^2/ha)	36.9	7.6	23.1	54.5
Mean diameter (cm)	21.0	4.2	13.1	30.2
Mean height (m)	18.9	4.2	10.3	26.6
Mean age (years)	63.0	17.0	31.0	90.0

The normal linear regression function is of the form $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon$, where the dependent variable *Y* is estimated with a constant value β_0, *p* independent variables and a random parameter ε. The linear quantile regression method estimates a number of linear functions, where the τ refers to the parameters of the τ quantile and $\tau \in [0,1]$ (Eq. (2)). Thus $\tau = 0.95$, $Q_y(0.95|X)$ would denote the 95th percentile of the distribution of *y* conditional upon the values of *X* (Cade and Noon, 2003).

$$Q_y(\tau|X) = \beta_0(\tau)\chi_0 + \beta_1(\tau)c_1 + \beta_2(\tau)c_2 + \cdots \beta_p(\tau)\chi_p. \quad (2)$$

We used quantile regression here to examine the prediction error ΔY of the variables BA, D_{MEAN}, H_{MEAN} and *N* as a function of the field measurements of *Y* and as a function of the length of the simulation. Quantile regression was employed instead of normal regression analysis in order to find out the possible relationships that cannot be determined with normal regression analysis. In this case, as the dependent variable was the prediction error, the quantiles could be used to model the confidence limits for the model: i.e. 90% of the observed errors lie between the 95th and 5th quantiles and 50% between the 75th and 25th percentiles. This gives more insight into the error distribution and behaviour than does the RMSE of the errors. The aforementioned variables were chosen for the analysis because reference field measurements were available for them and because they have highly significant roles in both the tree-level and stand-level simulators. We used polynomial functions in which the model takes the form $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 X^3 + \cdots + \beta_p X^p$. All the models were fitted as second-order polynomial functions. The analysis was carried out with R software (http://www.r-project.org).

3. Results

3.1. Prediction errors

The mean prediction error ΔY for all the variables of interest showed notable variation when studied against the length of the simulation period. There was also a considerable difference in the mean prediction error between the SIMO tree-level simulator (SIMO tree), SIMO stand-level simulator (SIMO

Table 2
Relative mean prediction errors (mean) and standard deviations (S.D.) for simulations with tree-level (tree) and stand-level (stand) growth models implemented on the SIMO platform and with the MOTTI simulator

Simulation length (years)	H_{MEAN} (%)			D_{MEAN} (%)			BA (%)			N (%)		
	Motti	Simo tree	Simo stand	Motti	Simo tree	Simo stand	Motti	Simo tree	Simo stand	Motti	Simo tree	Simo stand
5												
Mean	−3.2	−4.5	−4.0	0.1	−2.3	−0.2	−4.8	−4.6	0.4	−9.0	−6.1	−3.0
S.D.	5.9	6.0	8.9	4.6	4.4	3.4	14.1	11.1	19.4	11.8	7.5	7.1
10												
Mean	−5.3	−6.2	−4.9	−0.8	−2.3	−0.3	−9.5	−6.3	−3.8	−12.6	−9.6	−3.9
S.D.	8.5	8.6	11.2	5.3	5.2	4.4	15.5	14.0	13.3	13.6	10.4	9.6
15												
Mean	−6.1	−6.7	−5.2	−0.1	−1.6	0.0	−8.7	−4.3	−3.4	−12.2	−8.7	−0.3
S.D.	10.1	10.3	12.9	6.5	5.4	4.9	17.4	16.4	12.0	15.9	13.7	12.8
20												
Mean	−6.4	−6.6	−5.2	1.2	−0.6	0.3	−5.3	−0.3	−0.2	−10.3	−6.1	−6.3
S.D.	11.4	11.6	14.3	9.6	5.5	5.2	19.4	17.9	12.8	16.7	14.3	17.8

stand) and MOTTI simulator (MOTTI). Despite the differences, comparable trends could be seen in the results of all three simulators. They all systematically underestimated the values of most of the variables examined, this effect being more pronounced with the tree-level simulator than with the stand-level simulator. The absolute value of the relative mean prediction error of all variables increased when the length of the simulation period was extended from 5 to 15 years, but the underestimation in the mean prediction error generally decreased, or in some cases changed to an overestimation, when the simulation period was 20 years. The overall underestimation trend can be seen in Table 2, which shows the prediction errors for all the variables of interest.

The standard deviation of the estimation error also had definite trends that were clearly present in the results of both the tree-level and stand-level simulators, becoming systematically larger as the length of the simulation period increased. The standard deviation of the prediction error was greater with the tree-level simulator than with the stand-level simulator in respect to BA and D_{MEAN}, and larger with the stand-level predictions for H_{MEAN} and N. The standard deviations of the prediction errors can also be seen in Table 2.

The prediction error for variable N was notably different from the general trend of underestimation which is visible in almost all of the results. As seen in Fig. 1, the predictions made with the SIMO stand simulator behave in a quite different manner from the MOTTI and SIMO tree-level predictions. Whereas the latter invariably underestimated the variable N, the stand-level simulator yielded underestimates for the first 15 years of simulation but overestimates thereafter. Also the standard deviations of SIMO stand simulator's estimates of N were notably high. It should also be noted that the SIMO tree and MOTTI tree simulations behaved slightly differently. Thus, although the models are in principle similar, the results are not. This is due to differences in the implementation of the simulators.

3.2. Quantile regression of prediction error

The prediction error ΔY was modelled as a function of the field measurements of the variable Y by the quantile regression method, estimating the conditional quantile functions as second-order polynomial functions of the field values or simulation times. The behaviour of the prediction error as a function of the field measurements varied greatly between the variables and between the three simulators. The prediction error ΔD_{MEAN} showed quite different behaviour when estimated with the tree-level and stand-level simulators. The relative mean prediction error is very small for both simulators, as can be seen in Table 2, but the error estimated by linear regression varies as a function of the field measurements specifically in the case of the tree-level simulator, implying a bias with respect to D_{MEAN} (Fig. 2). The normal fitted regression line for the prediction error with the stand-level simulator does not change noticeably as a function of the field measurements, implying non-biased predictions.

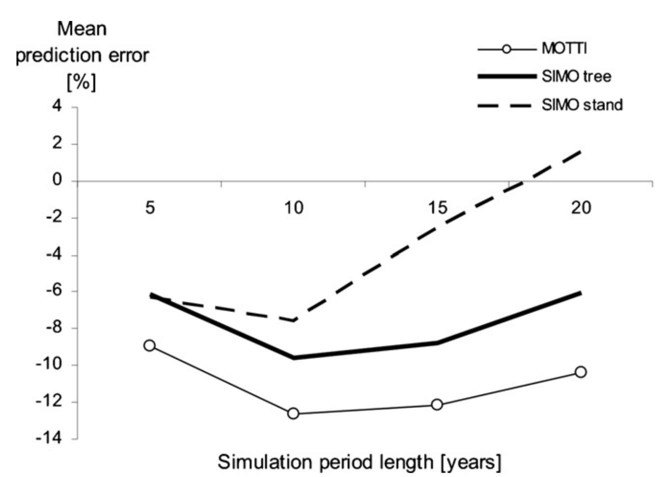

Fig. 1. Relative mean error in predicting the number of trees per hectare (N) plotted against the length of the simulation period with the SIMO tree, SIMO stand and MOTTI simulators.

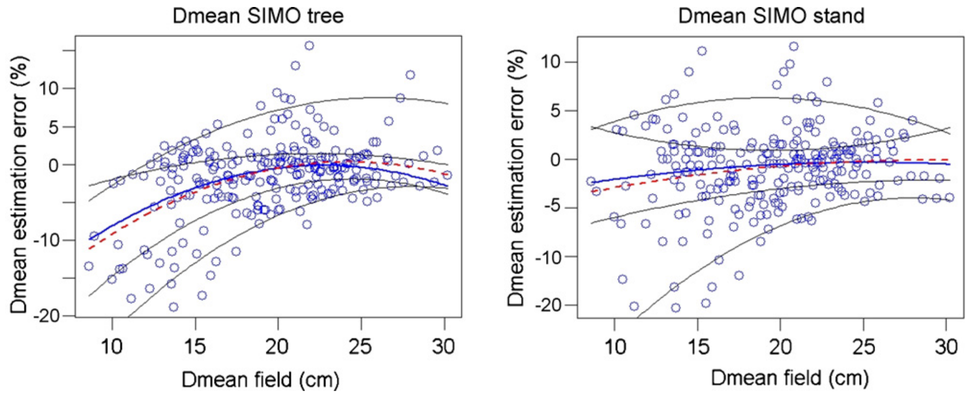

Fig. 2. Relative error in predicting the variable D_{MEAN} as a function of field measurements of D_{MEAN} for the SIMO tree (left) and SIMO stand (right) simulators. The red dotted line is a polynomial function estimated with normal regression, the blue line is the function for the quantile $Q_y(0.5|X)$ and the thin black lines are the quantiles $Q_y(0.05|X)$, $Q_y(0.25|X)$, $Q_y(0.75|X)$ and $Q_y(0.95|X)$ (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.).

Another clearly visible difference concerns the conditional quantile functions, which are quite different for the tree-level and stand-level simulator prediction errors. The conditional quantile functions for the ΔD_{MEAN} predicted with the SIMO tree simulator are all ascending and have relatively similar slopes, which points to fairly equal variance in the distribution of the prediction error. The SIMO stand simulator prediction error, however, does not show much change in the mean, but entails considerable changes in the slopes of the conditional quantile functions, which suggests unequal variance in the distribution of ΔD_{MEAN}. The variance of ΔD_{MEAN} clearly decreases as the field measurements of D_{MEAN} increases, which indicates that stand-level growth models yield more precise results in older stands or stands with a higher mean diameter.

Another variable with a prediction error that behaved differently as a function of the field measurements was the number of stems N. Both SIMO tree and stand simulators underestimated N markedly when a larger number of stems were measured in the field. SIMO stand simulator yielded considerable overestimates when the field measurement of N was small. This can be seen in Fig. 3, where the mean of the prediction error steadily decreases and the slopes of all of the conditional quantile functions are negative. The mean of the prediction error for N decreases strongly also in the SIMO stand simulator results, but the slopes of the conditional quantile functions vary from positive to negative, which indicates greater variation in ΔN.

When the prediction error ΔY for all the variables was modelled as a function of simulation length, all three simulators showed comparable behaviour. The general trend was that the variance in the prediction error ΔY increased as a function of simulation length. It can be seen in Fig. 4 that the ΔD_{MEAN} for the SIMO tree simulator is an underestimate at the beginning of the simulation but tends towards zero as the simulation length increases. The prediction error associated with D_{MEAN} in the SIMO stand simulator results shows an even mean value but increasing variance, which is visible in the slopes of the quantile regression functions. One slight departure from the general trend is that the tail of the distribution is longer in the upper quantiles, or overestimates, of the ΔD_{MEAN} obtained with the SIMO stand simulator. This is clearly visible in Fig. 4, where the interval between the upper quantiles is notably larger than between the lower quantiles. Most of the other results show the opposite behaviour, in which the distribution of ΔY is skewed so that the lower quantiles, representing underestimates, have a longer tail.

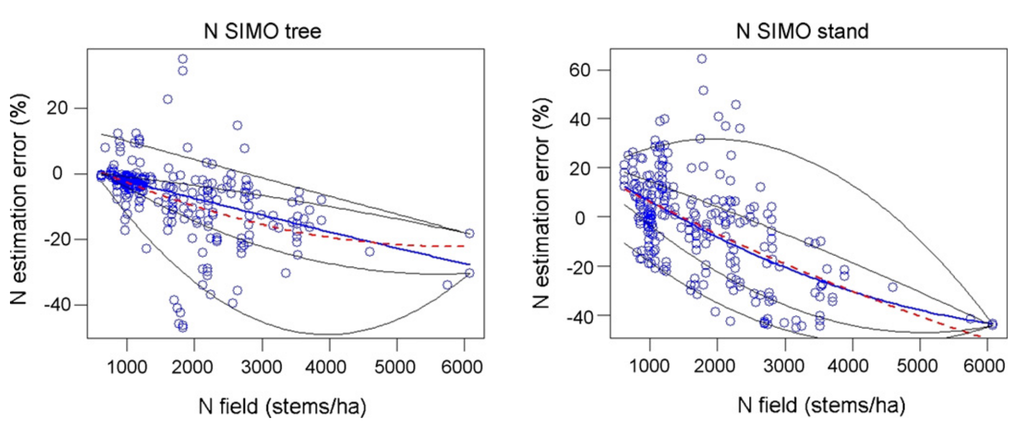

Fig. 3. Relative error in predicting variable N as a function of field measurements of N with the SIMO tree (left) and SIMO stand (right) simulators. The red dotted line is a polynomial function estimated by normal regression, the blue line is the function for the quantile $Q_y(0.5|X)$ and the thin black lines are the quantiles $Q_y(0.05|X)$, $Q_y(0.25|X)$, $Q_y(0.75|X)$ and $Q_y(0.95|X)$ (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.).

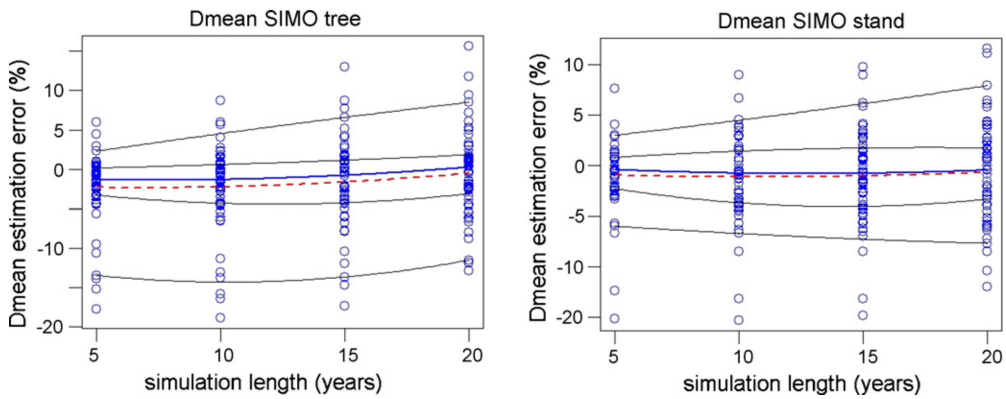

Fig. 4. Relative error in predicting the variable D_{MEAN} as a function of simulation length with the SIMO tree (left) and SIMO stand (right) simulators. The red dotted line is a polynomial function estimated by normal regression, the blue line is the function for the quantile $Q_y(0.5|X)$ and the thin black lines are the quantiles $Q_y(0.05|X)$, $Q_y(0.25|X)$, $Q_y(0.75|X)$ and $Q_y(0.95|X)$ (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.).

4. Discussion

Our aim here was to find possible differences between the predictions made with tree-level and stand-level growth models, especially when simulating the development of dense and unmanaged forest stands. We also wanted to test the quantile regression method for analysing prediction errors.

The errors in the prediction of all the variables of interest, i.e. basal area, mean diameter, mean height and number of stems, showed fairly similar behaviour, and with only a couple of exceptions, all three simulators produced systematic underestimates for all the variables of interest. The mean prediction error varied from −3.2 to −6.8% for H_{MEAN}, from 1.2 to −2.3% for D_{MEAN}, from 0.4 to −9.5% for BA and from 6.3 to −12.6% for N. Another marked trend was for the variance in ΔY to increase as a function of simulation length. The predictions made with the SIMO stand simulator were generally more accurate and precise than with the SIMO tree or MOTTI simulators for BA and D_{MEAN}, but less accurate and precise for H_{MEAN} and N, which had a higher standard deviation in its prediction error in the SIMO stand simulations. The most striking difference between the three simulators was in the error in predicting N, which was a definite underestimate in the SIMO tree and MOTTI simulations, but an overestimate with the SIMO stand simulator. The MOTTI and SIMO tree simulators, which both use tree-level growth models, yielded comparable predictions for all the variables of interest.

The mean error in predicting N showed somewhat different behaviour from the other variables, especially in the SIMO stand simulations, which yielded overestimates rather than underestimates. The reason for this is that the stand-level simulator does not include any explicit self-thinning model, which would ‘‘kill’’ trees in the stand under certain competition conditions. Tree deaths due to competition are only implicitly present in stand-level simulator, more specifically through the diameter distribution model used for predicting stem numbers with given BA and D_{MEAN}. With an underestimated D_{MEAN}, these models would yield a larger N for a given BA. The tree-level simulator does have a self-thinning model, however, which is driven by D_{MEAN} and N, so that underestimation of D_{MEAN} should produce less self-thinning and thus also a larger N. The situation is nevertheless complicated by the interactions between all models. In tree-level models, over-estimated self-thinning in the first period, based on true value of D_{MEAN}, is probably the main reason for constant underestimates for the other studied variables.

Quantile regression proved to be a useful tool for analysing the prediction errors in the simulations, providing more insight into the nature of these errors, as the variance in the distributions of ΔY was not equal but changed as a function of X in most cases. It would have been difficult to take this into account in a normal regression model, as instead of basic regression models, it would have been necessary to employ location scale models, which are regression models with unequal variance (Cade and Noon, 2003).

The present data included stands which had been selected because of their high densities or high basal areas, in order to test how the growth models behave under extreme conditions. Both the tree-level and stand-level growth models were empirical and we presumed that they would show unexpected behaviour when the data came close to the limits of the application area of the models. All of the models in both tree- and stand-level simulators are based on a representative data set that covers the whole of Finland for all main tree species. The modeling data probably did not have a lot of extremely dense forests, which is apparent in the results when simulating the development of such dense stands. The stand-level growth models seemed to be slightly more robust in the face of extreme data values than the tree-level growth models, for a number of possible reasons. The simulators, especially the tree-level simulator, are fairly complex systems with multiple data aggregation levels and a large number of equations for predicting growth and forest parameters. Even a single equation or sub-model in the simulator which reacts unexpectedly to extreme values of the variable can lead to an accumulation of errors (Hynynen and Ojansuu, 2003). The stand-level simulator is much simpler than the tree-level simulators, which may mean that it can handle extreme values in a more robust manner.

The data set included tree-level data measured in the field, which is not usually the case in practical forestry due to the high costs of field measurements. The presence of such input data should be an advantage for the tree-level simulator, as it implies accurate information about the stand structure and actual trees rather than trees generated with a distribution model to form a tree-list. In practise, however, the stand-level simulator still proved to be more robust than the tree-level simulator.

The length of the simulation period affected the prediction errors for all variables in all the simulators, with the variance in the prediction error increasing as a function of time, so that the predictions become less reliable and uncertainty attached to them grew as the length of the simulation increased. This was to be expected, as it can be regarded as normal behaviour for growth models (e.g. Kangas, 1997). The mean of the prediction error showed rather more inconsistent behaviour with respect to the different variables. The mean prediction error ΔY for all the variables increased during the first 15 years of simulation, but then trended slightly towards zero, which may be due to the fact that the stands had exceptionally high basal areas and the growth models do not converge well towards extreme values at the limits of their domains. This kind of underestimating behaviour is not surprising as the statistical growth models were used at the upper limits of their application domain, which is dictated by the modeling data. After 15 years the actual growth in the stand variables decreased more than the growth predicted by the simulators, which led to a decline in the mean prediction error.

As the stand-level models proved to be more robust and accurate, whereas tree-level models are more useful, e.g. for predicting the effects of silvicultural treatments, it would be most useful if these two approaches could be combined. One possibility would be to use stand-level models for predicting the level of growth and to distribute this growth among the individual trees (e.g. Ritchie and Hann, 1997; Qin and Cao, 2006), or another approach would be to use the different simulators consecutively, employing tree-level models for the first 10 years to obtain better information for planning immediate silvicultural measures and the more robust stand-level models for longer-term predictions. Yet another possibility might be to use the simulators in parallel, so that the development of stands that have just been thinned would be predicted with stand-level models, for instance, as accurate information in the stand structure in the immediate future may not be needed, and tree-level models could be used for stands that are approaching the next treatment. In this way it would be possible to extract maximum benefit from both simulator types. However, this topic requires more studying and testing of the different approaches for combining the two different simulator types. The availability of the SIMO simulation framework will make it possible to test the usefulness of these approaches in the future.

References

Atta-Boateng, J., Moser, J.W., 2000. A compatible growth and yield model for the management of mixed tropical rain forest. Can. J. Forest Res. 30, 311–323.

Bailey, R.L., Abernethy, N.C. Jones, E.P. 1981. Diameter distribution models for repeatedly thinned slash pine plantations. USDA Forest Service, general Technical Report SO-34, pp. 115–122.

Buongiorno, J., Mitchie, B.R., 1980. A matrix model of uneven-aged forest management. Forest Sci. 35, 548–556.

Burkhart, H.E., 1993. Tree and stand models in forest inventory. In: Nyyssönen, A., Poso, S., Rautala, J. (Eds.), Proceeding of the Ilvessalo Symposium on National Forest Inventories. IUFRO S4.02, Metsäntutkimuksen tiedonantoja 444, pp. 164–170.

Burkhart, H., 2003. Suggestions for choosing an appropriate level for modeling forest stands. In: Amaro, A., Reed, D., Soares, P. (Eds.), Modelling Forest Systems. CABI Publishing, p. 398.

Cade, B., Noon, B., 2003. A gentle introduction to quantile regression for ecologists. Front. Ecol. Environ. 1, 412–420.

Eerikäinen, K., 2002. A site dependent simultaneous growth projection model for Pinus kesiya plantations in Zambia and Zimbabwe. Forest Sci. 48, 58–529.

Garcia, O., 2001. On bridging the gap between tree-level and stand-level models. Forest biometry, modelling and information science. In: Proceedings of the IUFRO 4.11 Conference held at the University of Greenwich, June.

Gobakken, T., Næsset, E., 2005. Weibull and percentile models for lidar-based estimation of basal area distribution. Scand. J. Forest Res. 20, 490–502.

Gustavsen, N.G., 1977. Valtakunnalliset kuutiokasvuyhtälöt. Folia Forestalia 331 (in Finnish).

Gustavsen, H. 1998. Volymtillväxten och övre höjdens utveckling i talldominerade bestånd i Finland—en utvärdering av några modellers validitet i nuvarande skogar. Finnish Forest Research Institute, Research Notes 707. 190 p + app (in Swedish).

Hynynen, J., Ojansuu, R., Hökkä, H., Siipilehto, J., Salminen, H., Haapala, P., 2002. Models for Predicting Stand Development in MELA System. Finnish Forest Research Institute Research Papers, p. 835.

Hynynen, J., Ahtikoski, A., Siitonen, J., Sievänen, R., Liski, J., 2005. Applying the MOTTI simulator to analyse the effects of alternative management schedules on timber and non-timber production. Forest Ecol. Manage. 207, 5–18.

Hynynen, J., Ojansuu, R., 2003. Impact of plot size on individual-tree competition measures for growth and yield simulators. Can. J. Forest Res. 33, 455–465.

Kangas, A., 1997. On the prediction bias and variance of long-term growth predictions. Forest Ecol. Manage. 96, 207–216.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46 (1), 33.

Koenker, R., Hallock, K., 2001. Quantile regression. J. Econ. Perspect. 15, 143–156.

Kolström, T., 1993. Modelling the development of an uneven-aged stand of *Picea abies*. Scand. J. Forest Res. 8, 373–383.

Maltamo, M., 1997. Comparing basal area diameter distributions estimated by tree species and for the entire growing stock in a mixed stand. Silva Fenn. 31, 53–65.

McGrath, R.E., 2003. XML and Scientific File Formats National Center for Supercomputing Applications. University of Illinois, Urbana-Champaign, August.

Mielikäinen, K., 1985. Koivusekoituksen vaikutus kuusikon rakenteeseen ja kehitykseen. Commun. Inst. Forest. Fenn. 133 (in Finnish).

Munro, D.D., 1974. Forest growth models—a prognosis. Growth models for tree and stand simulation. In: Fries, J. (Ed.), Department of Forest Yield Research, Research Notes Nr 30. Royal College of Forestry, Stockholm, pp. 7–21.

Nyyssönen, A., Mielikäinen, K., 1978. Metsikön kasvun arviointi. Acta For. Fenn. 163 (in Finnish).

Oikarinen, M., 1983. Etelä-Suomen viljeltyjen rauduskoivikoiden kasvumallit. Commun. Inst. For. Fenn. 113. (in Finnish).

Porté, A., Bartelink, H.H., 2002. Modelling mixed forest growth: a review of models for forest management. Ecol. Model. 150, 141–188.

Qin, J., Cao, Q.V., 2006. Using disaggregation to link individual-tree and whole-stand growth models. Can. J. Forest Res. 36, 953–960.

Rasinmäki, J., Kalliovirta, J., Mäkinen, A. 2007. An adaptable simulation framework for multiscale forest resource data. Manuscript.

Ritchie, M.W., Hann, D.W., 1997. Implications of disaggregation in forest growth and yield modelling. Forest Sci. 43, 223–233.

Salminen, H., Lehtonen, M., Hynynen, J., 2005. Reusing legacy FORTRAN in the MOTTI growth and yield simulator. Comput. Electron. Agric. 49, 103–113.

Saramäki, J., 1977. Ojitettujen turvemaiden hieskoivikoiden kehitys Kainuussa ja Pohjanmaalla. Metsäntutkimuslaitoksen julkaisuja 91. 2 (in Finnish).

Shortt, J.S., Burkhart, H.E., 1996. A comparison of loblolly pine plantation growth and yield models for inventory updating. South. J. Appl. For. 20 (1), 15–22.

Tokola, T., Kangas, A., Kalliovirta, J., Mäkinen, A., Rasinmäki, J., 2006. SIMO—SIMulointi ja Optimointi uuteen metsäsuunnitteluun. Metsätieteen aikakauskirja 1, 60–65 (in Finnish).

Tomé, M., Burkhart, H.E., 1989. Distance-dependent competition measures for predicting growth of individual trees. Forest Sci. 35, 816–831.

Vanclay, J.K., 1995. Growth models for tropical forests: a synthesis of models and methods. Forest Sci. 41, 7–42.

Vanclay, J.K., Skovsgaard, J.P., 1997. Evaluating forest growth models. Ecol. Model. 98, 1–12.

Vettenranta, J., 1999. Distance-dependent models for predicting the development of mixed coniferous forests in Finland. Silva Fenn. 33, 51–72.

Vuokila, Y., Väliaho, H., 1980. Viljeltyjen havumetsiköiden kasvumallit. Metsäntutkimuslaitoksen julkaisuja 99 (2) (in Finnish).

Välimäki, E. 2006. Kasvumallien validointi ylitiheissä metsiköissä. Master's Thesis. Department of Forest Resource Management, University of Helsinki, 62 p., (in Finnish).

Zeide, B., 1993. Analysis of growth equations. Forest Sci. 39, 594–616.