Recruitment Models for Norway Spruce, Scots Pine, Birch and Other Broadleaves in Young Growth Forests in Norway

Nils Lexerød and Tron Eid

The objective of the present study was to develop recruitment models for Norway spruce, Scots pine, birch and other broadleaves in young growth forests in Norway. The models were developed from permanent sample plots established by the National Forest Inventory, and they will be included in a growth simulator that is part of a large-scale forestry scenario model. The modelling was therefore restricted to independent variables directly or indirectly available from inventories for practical forest management planning. A two-stage modelling approach that suited the stochastic nature of recruitment in boreal forests was used. Models predicting the probability of recruitment were estimated in a first stage, and conditional models for the number of recruits were developed in a second. The probability models as well as the conditional models were biologically realistic and logical. The goodness of fit tests revealed that the probability models fitted the data well, while the coefficients of determination for the conditional models were relatively low. No independent test data were available, but comparisons of predicted and observed number of recruits in different sub-groups of the data revealed few large deviations. The high level of large random errors was probably due to the great variability observed in number of recruits rather than inappropriate specifications of the models. Provided the generally high level of uncertainty connected to analysis performed with large-scale forestry scenario models and the stochastic nature of recruitment, the presented models seem to give satisfactory levels of accuracy.

1 Introduction

Recruitment is usually quantified by means of number of trees or seedlings that reach or exceed a specific threshold size over a certain period, as a result of different regeneration processes like establishment, growth and mortality of seedlings. A recruitment model predicts the number of trees or seedlings that pass a specific threshold size during a specified period of time. Usually, such models do not directly consider factors influencing seedling establishment (e.g. production, dispersion, predation and germination of seeds and competition from herbaceous vegetation), growth and mortality. Instead, recruitment models predict number of recruits directly from independent variables describing site and forest conditions.

Growth simulators and large-scale forestry scenario models are widely used to predict growth and yield of forest stands and to estimate the profitability of different harvesting regimes, and their influence on long-term timber production, stand structure and biodiversity. In Norway, several growth simulators and large-scale forestry scenario models covering even-aged forests have been developed from area-based growth models (e.g. GAYA (Hoen and Eid 1990) and AVVIRK (Eid and Hobbelstad 2000)). These models can, however, not predict, in a reliable way, stand development in uneven-aged forests, mainly because prediction of growth and yield is done independently of the tree size distribution. For this reason, a growth simulator able to predict stand development in any-aged forests has recently been developed in Norway (Gobakken et al. 2005). Another objective with the growth simulator is to consider stochastic regeneration, recruitment and mortality in projections. Hence, the mortality as well as the recruitment models needs to be stochastic.

The growth simulator consists of different sub-models, Fig. 1, including a regeneration model

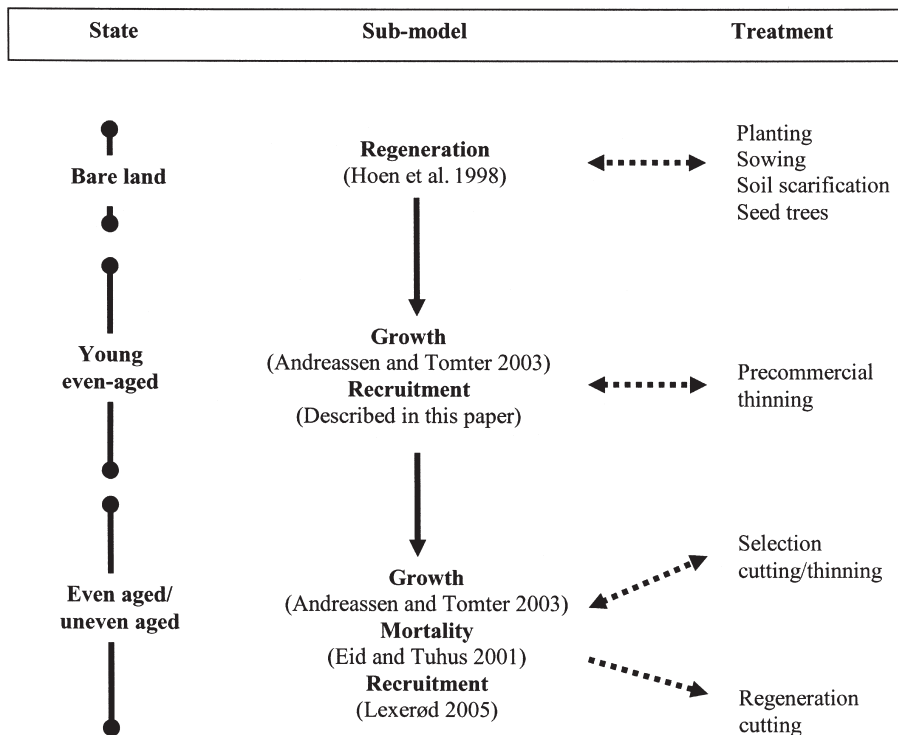| State	Sub-model	Treatment
Bare land	**Regeneration** (Hoen et al. 1998)	Planting Sowing Soil scarification Seed trees
Young even-aged	**Growth** (Andreassen and Tomter 2003) **Recruitment** (Described in this paper)	Precommercial thinning
Even aged/ uneven aged	**Growth** (Andreassen and Tomter 2003) **Mortality** (Eid and Tuhus 2001) **Recruitment** (Lexerød 2005)	Selection cutting/thinning Regeneration cutting

Fig. 1. Overview of the different sub-models in the growth simulator. The figure shows at which states the sub-models are used and how they interact with silvicultural treatments.

predicting number of seedlings and species composition after regeneration cutting. Independent variables in the regeneration model is site index, vegetation type and regeneration method (Hoen et al. 1998). In addition to the regeneration model, the growth simulator consists of growth models predicting basal area growth of individual trees (Andreassen and Tomter 2003), stochastic mortality models for individual trees (Eid and Tuhus 2001) and stochastic recruitment models covering even-aged as well as uneven-aged forests (Lexerød 2005). Plots with young growth were, however, not included in the data material used by Lexerød (2005), and consequently the models do not cover such forests. Hence, stochastic recruitment models covering young growth forests needed to be developed in order to improve projections with the simulator.

Several different approaches to recruitment modelling have been applied. Models predicting recruitment with one single equation is the most common approach (e.g. Moser (1972), Adams and Ek (1974), Buongiorno and Michie (1980), Hyink and Moser (1983) and Vanclay (1989)). Recruitment is the result of a stochastic process and, for several species and forest conditions in Norway, recruitment is either present during a particular growth period, or it is not. Hence, if both observations with and without recruitment are retained for estimation of one single equation, normal distribution of residuals will rarely appear. Thus, it will be difficult to evaluate the overall performance of the models. Alternatively, if only observations with recruitment are retained for model development, recruitment will be overestimated.

The stochastic nature of recruitment can be modelled by defining two separate states with the following characteristics; (i) observations can be placed with certainty in either of the two states and (ii) the dependent variable (i.e. number of recruits) is zero for all observations in one state. Such a system could be modelled by using a two-stage modelling approach that is previously applied to recruitment data by Vanclay (1992) and Lexerød (2005). Probability models predicting the probability that recruitment is present during the next simulation cycle are estimated in the first stage. In the second stage, conditional models are estimated in order to predict number of recruits given that recruitment is present.

Separate models for Norway spruce, Scots pine, birch and other broadleaves were developed based on data from permanent sample plots established by the National Forest Inventory (NFI) in Norway. The data covers all geographic regions, forest conditions and silvicultural prescriptions representative for the conditions under which the growth simulator will be used. The modelling was restricted to include independent variables directly or indirectly available from forest inventories for practical forest management planning.

2 Material and Methods

2.1 Data Collection and Preparation

According to the Norwegian classification system young growth is defined as development class II. Maximum total stand ages for development class II are 20, 20, 25, 30, 35, 45 and 55 years for site index 23, 20, 17, 14, 11, 8 and 6 respectively if the forest is dominated by conifers or other broadleaves. The maximum ages for birch are 15, 15, 20, 25, 25, 25 and 30 years, respectively. Site index (SI) is estimated according to the H_{40} system, which is based on the dominant height (i.e. mean height of the 100 largest trees ha^{-1}) at breast height age 40 years (Strand 1967, Tveite 1976, 1977).

The NFI in Norway established permanent sample plots between 1986 and 1993. All plots were remeasured from 1994 to 1998, and approximately 60% of the plots, with a representative distribution, were remeasured for the second time in the period 1999–2001. The permanent sample plots cover the forest area of Norway (Finnmark county excluded) in a 3 km × 3 km grid (NIJOS 2000). Diameters were unfortunately not recorded in development class II at the time of establishment. Hence, recruitment in young forest could only be calculated for the second growth period consisting of approximately 60% of plots. The time period between the measurements was five years, and the size of the circular sample plots was 250 m^2. The total number of available sample plots in development class II, after exclusion of non-productive areas (i.e. annual production < 1 m^3 ha^{-1}), was 897 (Table 1). Dominant tree

Table 1. Data summary [a].

	Spruce dominated			Pine dominated			Birch dominated			Other broadleaves dominated		
No. of observations	464			190			94			149		
	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.	Mean	Min.	Max.	Mean
ALT (m)	5.0	860.0	338.0	15.0	930.0	395.8	0.0	940.0	343.6	5.0	860.0	312.7
LAT (°N)	58.1	69.6	61.7	58.2	69.9	60.8	58.1	69.8	63.3	58.1	69.5	61.5
SI (m)	6.0	26.0	13.2	6.0	17.0	9.9	6.0	23.0	10.2	6.0	23.0	11.5
A (years)	1.0	45.0	16.6	2.0	44.0	20.0	2.0	30.0	13.6	2.0	48.0	14.2
$N_{Conifers}$ (ha^{-1})	240	7000	1772	80	9000	1854	0	3500	527	0	8000	963
$N_{Broadleaves}$ (ha^{-1})	0	8000	1053	0	7000	947	200	9000	3446	400	9500	3073
$H_{Conifers}$ (m)	0.1	8.5	2.5	0.0	>10.0	2.7	0.0	6.0	1.1	0.0	6.0	1.8
$H_{Broadleaves}$ (m)	0.0	9.0	1.9	0.0	8.5	2.0	0.1	7.5	2.5	0.3	8.5	2.5

[a] ALT: altitude; LAT: latitude; SI: Site index, dominant height at breast height age 40 years; A: age; N: number of trees; H: arithmetic mean height.

Table 2. Total number of observations, observations with recruitment and annual recruitment rate for observations with recruitment and for all observations by tree species.

Tree species	Total no. of observations	Observations with recruitment		Annual recruitment rate (iN ha^{-1} yr^{-1})	
		No.	%	Observations with recruitment	All observations
Norway spruce	897	510	56.9	46.8	26.6
Scots pine	897	228	25.4	34.4	8.7
Birch	897	436	48.6	44.7	21.7
Other broadleaves	897	166	18.5	47.2	8.7
All	897	754	84.1	78.3	65.9

All observations have a remeasurement interval of five years. Threshold size, dbh = 5 cm.

species (spruce, pine or broadleaves) were determined as the single tree species with the largest proportion of crown cover.

Tree species and diameter at breast height (dbh) were recorded for all trees with dbh ≥ 5 cm on each sample plot. In addition, numbers of trees (N) with height larger than 0.3 m was registered within an area of 1000 m^2 which included the 250 m^2 sample plot together with a buffer zone. Altitude (ALT), latitude (LAT), SI, stand age (A) and arithmetic mean height (H) were also based on the area of 1000 m^2. SI was estimated according to the H_{40} system described above for the dominant tree species on the plot, while number of trees and arithmetic mean height were measured separately for conifers and broadleaves. Vegetation type was registered according to the Norwegian classification of vegetation types (Fremstad 1997). Vegeta-

tion types having similar influence on recruitment of a particular species were clustered using a method described by Lexerød (2005). The cluster procedure resulted in three different clusters describing good (VT1), intermediate (VT2) and poor (VT3) conditions for recruitment.

Recruitment was calculated for each plot as number of seedlings that reached or exceeded dbh = 5 cm between the two measurements. The proportion of plots with recruitment and number of recruits per plot varied considerably between species (Table 2). Observations of recruits were converted into annual recruitment rates (iN ha^{-1} yr^{-1}) based on observations with recruitment only, and all observations, respectively. Norway spruce (*Picea abies* L.) and Scots pine (*Pinus sylvestris* L.) were treated separately while the broadleaved species were defined in

groups as either birch (*Betula* spp.) or as other broadleaves. Species defined as other broadleaves were grey alder (*Alnus incana* L.), black alder (*Alnus glutinosa* L.), rowan (*Sorbus aucuparia* L.), aspen (*Populus tremula* L.), beech (*Fagus sylvatica* L.), ash (*Fraxinus excelsior* L.), elm (*Ulmus glabra* L.), linden (*Tilia cordata* L.), maple (*Acer platanoides* L.) and oaks (*Quercus* ssp. (L.). A few foreign tree species were excluded because of small populations unsuitable for statistical analysis. These were Sitka spruce (*Picea sitchensis*), fir (*Abies* spp.), larch (*Larix* spp.) and Lodgepole pine (*Pinus contorta*).

Some trees were recruited because of measurement errors (e.g. a tree considered to be outside the plot at the beginning of the growth period was considered to be inside at the end, or a tree inside the plot, but not recorded at the initial measurement was recorded at the remeasurement). In order to exclude such measurement errors it was assumed that a recruited tree could not be larger at the end of the growth period than what could be expected from predicted diameter growth. Single tree basal area growth models (Andreassen and Tomter 2003) were used to predict the expected diameter of recruits if the recruits had a diameter just below the threshold size at the beginning of the growth period. The predicted diameters were used as maximum size of recruits. Recruits that were wind thrown or not alive at the time of remeasurement were also excluded to ensure compatibility with the mortality models.

2.2 Model Development

Since only trees with dbh ≥ 5 cm were recorded by the NFI, a threshold size of dbh = 5 cm was used (i.e. the models will predict number of seedlings that reach or exceed a dbh = 5 cm).

Probability of recruitment as well as number of recruits was assumed to depend on location, site conditions and stand characteristics. Independent variables were restricted to those directly or indirectly available from inventories for practical forest management planning. It was important to develop models that were biologically realistic in order to ensure a logical behaviour when applied outside the range of the modelling data. Thus, an appropriate biological interpretation was the

main criteria with respect to variable selection. Independent variables, with an appropriate biological interpretation, were selected among those describing location (ALT, LAT), site conditions (SI, VT1, VT2, VT3) and stand characteristics (A, N, PN_{sp}, H,) through visual analysis of data plots, residual analysis and forward and stepwise regression (SAS Institute Inc. 1999).

Given the binary variable 0 = absence and 1 = presence of recruitment, the probability of presence of recruitment over a five-year period $\pi(x)$ was modelled with the following logistic regression model (Agresti 1996):

$$\pi(x) = \frac{\exp^{\alpha + \beta x}}{1 + \exp^{\alpha + \beta x}} = \left[1 + \exp^{-(\alpha + \beta x)}\right]^{-1} \qquad (1)$$

where α and β are parameters defining the S-shaped function of x. The logistic model has a linear form for the logit of this probability;

$$\text{logit}[\pi(x)] = \log\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta x \qquad (2)$$

Model (2) was estimated for each species group with PROC LOGISTIC using maximum likelihood estimation with the logit link function and the Newton-Raphson optimization algorithm (SAS Institute Inc. 1999). The significance of the parameter estimates was tested by means of Wald statistics (i.e. $z^2 = \beta / \text{ASE}$ where β is the parameter estimate and ASE is the corresponding asymptotic standard error (Agresti 1996). The different models were evaluated with the Hosmer and Lemeshow goodness-of-fit test (Hosmer and Lemeshow 1989).

The relationship between the number of recruits (iN ha^{-1}5yr^{-1}) and the independent variables was expressed with the following conditional model:

$$iN = X'\beta + \varepsilon \qquad (3)$$

where X' is the transpose of the vector of independent variables, β is a vector of regression coefficients and ε is a vector of random errors. It was assumed that the error term ε had E(ε) = 0 and V(ε) = σ^2 and that the errors were uncorrelated. The assumption of constant variance is a basic requirement for regression analysis, and model

(3) was therefore log-transformed in order to stabilize variance;

$$\ln(iN) = X'\beta + \varepsilon \qquad (4)$$

Model (4) was estimated using PROC REG (SAS Institute Inc. 1999). In order to obtain unbiased estimates for the response variable when converted back to original scale, the intercept term was adjusted by adding half of the variance (Miller 1984). The relationship between the dependent variable and some of the independent variables were non-linear, and these variables were transformed in order to express a linear relationship. The final models were inspected with respect to multicollinearity by calculation of a condition index (Belsley et al. 1980). Since the intercept had no physical interpretation, data were centered before calculation of the condition index (Montgomery and Peck 1992).

2.3 Model Evaluation

Validation of a model should in general involve independent data preferably from controlled and replicated experiments measured over a long time (Vanclay and Skovsgaard 1997), or data representing another growth period, sample plot size or representation of forest conditions. Such data were, however, not available because dbh was not recorded in young forests at the time of plot establishment (i.e. only data from one growth period were available in young forests). In addition, no other suitable independent data were available at the present moment. The evaluation of the models was therefore based on the same data as the model development. The overall performance of the models was evaluated by examination of prediction graphs and residual plots, and by calculating the mean differences between predicted and observed number of recruits, together with the corresponding standard deviations, as recommended by Huang et al. (2003) and Kozak and Kozak (2003). In addition, the conditional models were evaluated by means of cross-validation and calculation of the predicted error sum of squares (PRESS). In all evaluations predicted number of recruits was calculated deterministically by multiplying the probability of recruitment with the expected number of recruits.

3 Results

Variables describing location (ALT, LAT), site (SI, VT1, VT2, VT3) and stand conditions (A, N, PN_{sp}) were tested as independent variables in the probability models. Neither ALT nor LAT had significant influence on the probability of recruitment for any species group (Table 3). The

Table 3. Parameter estimates for the probability models.

Variable [a]	Species/species group			
	Norway spruce	Scots pine	Birch	Other broadleaves
Intercept	−17.5779*** [b]	−14.6228***	−8.2528***	−6.8385***
ln (SI) (m)	1.4185***	1.1399***		1.0531***
ln (A) (yr)	2.0103***	0.9231***	1.1822***	0.2766*
ln (N) (ha^{-1})	0.8111***	0.7120***	0.2306*	0.2577*
ln (PN_{sp}) (%)	0.9393***	1.2627***	0.9530***	
Observations	897	897	897	897
Hosmer and Lemeshow statistics:				
Misclassified plots (%)	6.7	7.9	6.9	9.0
χ^2	13.9	5.7	9.2	2.3
$p > \chi^2$	0.08	0.68	0.33	0.97

Dependent variable $\pi(x)$ is the probability of recruitment over a period of 5 years. Threshold size, dbh = 5 cm.
[a] SI: Site index, dominant height at breast height age 40 years; A: age; N: number of trees; PN_{sp}: proportion of Norway spruce, Scots pine and broadleaves, respectively.
[b] * p < 0.10; ** p < 0.05; *** p < 0.01

Table 4. Parameter estimates, coefficient of determination (R^2), square root of mean square error (RMSE), coefficient of variation (C.V.), condition index (C.I.) and predicted error sum of squares (PRESS) for the conditional models for number of recruits.

Variable [a]	Species/species group			
	Norway spruce	Scots pine	Birch	Other broadleaves
Intercept (adj.)	–3.9717***	–0.66805 [ns] [b]	–0.8924 [ns]	15.5017**
ln (ALT) (m)				–0.2691***
ln (LAT) (°N)				–3.6831**
SI (m)	0.0631***	0.0547***		0.0444**
A (yr)	0.0391***	0.0175**	0.0294***	
ln (A) (yr)				0.2657**
N (ha^{-1})	–0.0001***		–0.0001 [ns]	–0.0002*
ln (N) (ha^{-1})	0.8335***	0.4145***	0.6204***	0.5514 [ns]
ln (PN_{sp}) (%)	0.4374***	0.4344***	0.2790***	0.3727***
Observations	510	228	436	166
R^2	0.402	0.310	0.155	0.194
Adj R^2	0.396	0.297	0.147	0.159
RMSE	0.697	0.755	0.898	0.907
C.V. (%)	13.8	16.0	18.2	18.4
C.I.	5.3	1.7	5.9	1.9
PRESS	252	147	406	143

Dependent variable is the natural logarithm of number of recruits during the next five years, ln (iN ha^{-1} 5yr^{-1}). Threshold size, dbh = 5 cm.
[a] ALT: altitude; LAT: latitude; SI: Site index, dominant height at breast height age 40 years; A: age; N: number of trees; PN_{sp}: proportion of Norway spruce, Scots pine and broadleaves, respectively.
[b] ns = not significant (p > 0.10); * p < 0.10; ** p < 0.05; *** p < 0.01

variables SI and VT indirectly describe site conditions as soil fertility, moisture and competition from herbs, all factors expected to influence recruitment. VT was, however, not significant for any species when N was included in the models, while the influence of SI was significant for all species except birch. The relationship between SI and the probability of recruitment was positive (i.e. the probability increased with increasing SI). Both A and N were significant for all species, and the probability of recruitment increased with increasing values of A and N. The probability of recruitment of a particular species significantly increased with increasing PN_{sp} for all species groups, except other broadleaves.

ALT and LAT were only included in the conditional models for other broadleaves (Table 4). The number of recruits decreased when ALT and LAT increased. SI had a significant effect on number of recruits for all species except birch, and number of recruits increased with increasing SI. The relationship with number of recruits where negative when SI was tested as independent variable in the birch model (i.e. number of recruits decreased

with increasing SI). Number of recruits were significantly influenced by A and PN_{sp} for all species groups, and number of recruits increased with increasing values of A and PN_{sp}. Number of recruits was expected to increase with increasing N up to an optimal level, with respect to number of recruits, and thereafter slowly decrease with increasing N due to increased competition and mortality. Such a relationship was included in all models except for Scots pine. The parameter estimates for the variable N in the birch model and for the transformation of N in the model for other broadleaves were not significant, but the variables were still included since the sign and magnitude of the parameter estimates were in agreement with the ones estimated for the other species groups.

The Hosmer and Lemeshow test revealed no significant differences between observed and expected number of plots with recruitment (Table 3). The proportion of misclassified plots was 6.7%, 7.9%, 6.9% and 9.0% for Norway spruce, Scots pine, birch and other broadleaves, respectively. The coefficient of determination (R^2)

for the conditional models was 0.40, 0.31, 0.16 and 0.19 for Norway spruce, Scots pine, birch and other broadleaves, respectively (Table 4). The coefficient of variation (C.V.) varied from 13.8 to 18.4%, and there were no sign of multicollinearity in any of the models.

The overall mean difference between observed and predicted number of recruits was 9.7, –1.3, –5.2 and –10.6 ha^{-1} 5 yr^{-1} for Norway spruce, Scots pine, birch and other broadleaves, respectively (Tables 5–8). This corresponds to an overestimation of 7.2% for Norway spruce and an underestimation of 3.0%, 4,8% and 24.3% for Scots pine, birch and other broadleaves, respectively. There were in general few large deviations between predicted and observed number of recruits in different parts of the material (Tables 5–8). In general, the spruce model seemed to slightly overestimate number of recruits when the observed recruitment rate was low, while the models for birch and other broadleaves underestimated number of recruits when forest conditions favoured recruitment of the respective species, as in broadleaved dominated forests.

4 Discussion

4.1 Data

The main advantage of the NFI data used for model development was that it covered all geographic regions, forest conditions and silvicultural practices in Norway. This was essential since the models will be included in a growth simulator supposed to cover all forest conditions and silvicultural treatments in Norway. Several recruitment models applied in growth simulators has been developed from NFI data to ensure a representative distribution of forest conditions and silvicultural practices (e.g. Sterba et al. (1997), Wikberg (2004) and Lexerød (2005))

The fact that all independent variables used in the presented models were registered on a 1000 m^2 large area including the sample plot (250 m^2) may have caused some discrepancies between the dependent variable (number of recruits) observed on the 250 m^2 plots and the independent variables. This is particularly relevant for N and PN$_{sp}$

since these variables may differ when observed on 250 m^2 compared to 1000 m^2. The problem is probably less important for SI and A since it is reasonable to assume only minor differences in these variables between a 250 m^2 sample plot and a 1000 m^2 large area surrounding it. These discrepancies have, however, probably not biased the parameter estimates, since the difference can be treated as random measurement errors. In addition, the relatively large number of plots has probably reduced the effect of extreme combinations of variables caused by the different sample plot sizes.

There were some uncertainties related to the use of the growth models by Andreassen and Tomter (2003), since the models do not cover development class II, and the accuracy of the models when extrapolated is unknown. Unfortunately, no models predicting diameter growth of individual trees in development class II were available, and the models by Andreassen and Tomter (2003) had to be used.

No independent data were available for model evaluation, and both data splitting and resampling techniques (e.g. cross-validation) were considered. It has, however, been questioned if such data splitting and resampling provides any additional information. Kozak and Kozak (2003), for instance, states that the goodness of fit statistics and lack of fit statistics calculated directly from the data used for model development provides a better and more reliable description of the prediction errors than statistics obtained from splitting the original data set. In addition, models developed from a subset of the entire data set will not be as reliable as models developed from the entire data set because some observations are disregarded. The quality of the model fit can also be assumed to reflect the quality of predictions when, as in this study, the data are collected through a well-designed and representative sampling-process, when the models will be used within the range of the data and when the models are biologically realistic (Huang et al. 2003). Thus, as long as the model fit was found satisfactory there are probably no reasons for a reduced confidence in the presented models despite the lack of independent test data.

Table 5. Mean differences ($\bar{D}$) and standard deviations (SD) for the differences between predicted and observed number of recruits of Norway spruce (ha^{-1} 5 yr^{-1}).

Part of the material	No. of observations	Observed mean	Predicted mean	$\bar{D}$	SD
Altitude (m)					
–150	186	149.5	146.0	–3.5	171.2
151–300	255	144.3	151.8	7.4	157.5
301–450	184	143.0	156.9	13.9	141.0
451–600	136	138.2	152.8	14.6	123.4
600–	136	71.5	92.6	21.1	99.7
Latitude (°N)					
–60	248	122.3	139.3	17.0	132.5
60–62	356	149.7	156.1	6.4	154.6
62–64	136	112.4	129.7	17.4	120.3
64–66	95	149.1	146.1	–2.9	150.6
66–	62	103.2	104.5	1.3	175.0
Site index (m)					
–9.5	236	39.2	59.2	20.1	72.2
9.6–12.5	285	122.0	119.8	–2.2	149.5
12.6–15.5	244	189.0	189.8	0.7	157.6
15.6–	132	222.1	255.1	33.0	196.3
Age class (yr)					
–5	97	12.4	36.1	23.7	53.5
6–10	176	52.3	81.7	29.4	90.5
11–15	163	128.8	137.4	8.6	122.1
16–20	190	210.1	185.3	–24.8	200.2
21–25	140	190.0	187.3	–2.7	153.9
25–	131	164.3	201.4	37.1	161.6
No. of trees (ha^{-1})					
–1000	78	47.7	57.6	9.9	55.1
1001–2000	262	116.0	124.8	8.8	111.3
2001–3000	210	157.9	172.5	14.6	153.2
3001–4000	128	159.7	172.4	12.7	189.6
4001–5000	83	174.9	150.3	–24.6	176.7
5001–	136	126.5	148.1	21.6	153.8
Forest type					
Spruce dom.	464	213.9	220.0	6.2	181.3
Pine dom.	190	36.8	44.2	7.4	89.8
Birch dom.	94	46.8	52.9	6.0	81.5
Broadleaved dom.	149	59.1	84.7	25.7	97.8
Region					
Eastern	573	135.3	142.2	6.9	139.7
Middle	160	157.3	149.6	–7.7	141.4
Western	81	110.1	163.3	53.2	162.6
Northern	83	94.5	113.9	19.4	161.6
All	897	133.2	142.8	9.7	144.9

Table 6. Mean differences ($\bar{D}$) and standard deviations (SD) for the differences between predicted and observed number of recruits of Scots pine (ha^{-1} 5 yr^{-1}).

Part of the material	No. of observations	Observed mean	Predicted mean	$\bar{D}$	SD
Altitude (m)					
–150	186	40.4	38.6	–1.8	85.2
151–300	255	64.5	50.9	–13.6	107.4
301–450	184	33.7	34.1	0.4	78.2
451–600	136	27.1	37.1	10.0	60.6
600–	136	39.7	48.5	8.8	88.7
Latitude (°N)					
–60	248	76.5	60.5	–16.0	121.6
60–62	356	39.9	47.5	7.6	79.8
62–64	136	36.5	35.7	–0.7	80.8
64–66	95	6.3	11.0	4.6	19.5
66–	62	8.4	4.4	–4.0	42.0
Site index (m)					
–9.5	236	47.1	51.3	4.1	76.2
9.6–12.5	285	55.3	51.5	–3.8	102.2
12.6–15.5	244	38.4	35.5	–2.9	96.6
15.6–	132	22.7	20.0	–2.7	56.1
Age class (yr)					
–5	97	3.3	10.7	7.4	29.8
6–10	176	25.7	28.2	2.5	85.2
11–15	163	53.0	46.5	–6.5	100.9
16–20	190	45.5	36.1	–9.4	85.6
21–25	140	56.3	56.7	0.5	79.5
21–	131	70.5	73.9	3.4	115.1
No. of trees (ha^{-1})					
–1000	78	20.0	29.4	9.4	42.1
1001–2000	262	39.8	37.2	–2.7	76.4
2001–3000	210	50.7	45.6	–5.0	99.0
3001–4000	128	36.6	41.6	5.1	71.8
4001–5000	83	38.6	38.5	–0.1	55.0
5001–	136	64.1	58.3	–5.8	133.1
Forest type					
Spruce dom.	464	12.1	15.8	3.8	37.7
Pine dom.	190	158.5	143.8	–14.7	177.2
Birch dom.	94	5.1	4.5	–0.6	18.8
Broadleaved dom.	149	20.4	19.9	–0.5	48.6
Region					
Eastern	573	58.9	56.0	–2.9	104.2
Middle	160	10.3	16.8	6.6	29.2
Western	81	41.0	35.6	–5.4	83.1
Northern	83	6.3	4.5	–1.8	36.8
All	897	43.7	42.4	–1.3	88.5

Table 7. Mean differences ($\bar{D}$) and standard deviations (SD) for the differences between predicted and observed number of recruits of birch (ha^{-1} 5 yr^{-1}).

Part of the material	No. of observations	Observed mean	Predicted mean	$\bar{D}$	SD
Altitude (m)					
–150	186	117.6	107.2	–10.4	196.4
151–300	255	117.0	101.3	–15.7	199.6
301–450	184	88.7	96.0	7.3	143.1
451–600	136	83.5	96.7	13.2	151.3
600–	136	132.1	118.6	–13.4	178.6
Latitude (°N)					
–60	248	91.0	97.0	6.1	168.6
60–62	356	103.4	95.8	–7.6	169.3
62–64	136	80.3	102.0	21.7	124.3
64–66	95	148.6	116.1	–32.6	216.9
66–	62	209.0	155.6	–53.4	270.6
Site index (m)					
–9.5	236	118.8	130.1	11.3	165.4
9.6–12.5	285	122.8	103.0	–19.8	192.9
12.6–15.5	244	94.4	91.4	–3.1	180.9
15.6–	132	85.5	78.5	–7.0	162.3
Age class (yr)					
–5	97	31.8	33.5	1.8	121.7
6–10	176	51.6	67.2	15.6	111.7
11–15	163	116.6	102.4	–14.2	184.9
16–20	190	149.9	112.8	–37.1	216.7
21–25	140	148.9	138.2	–10.7	201.1
25–	131	128.9	154.1	25.3	185.2
No. of trees (ha^{-1})					
–1000	78	30.8	48.4	17.6	73.0
1001–2000	262	81.7	74.6	–7.1	146.8
2001–3000	210	87.0	87.3	0.3	125.8
3001–4000	128	119.7	129.2	9.5	180.9
4001–5000	83	198.6	158.4	–40.2	273.1
5001–	136	172.6	157.2	–15.4	254.0
Forest type					
Spruce dom.	464	73.8	84.9	11.1	132.8
Pine dom.	190	74.1	88.6	14.5	137.3
Birch dom.	94	263.0	152.5	–110.5	293.2
Broadleaved dom.	149	163.2	148.8	–14.5	225.7
Region					
Eastern	573	98.4	98.0	–0.5	169.2
Middle	160	112.8	104.8	–7.9	177.3
Western	81	90.4	92.4	2.1	141.2
Northern	83	188.0	148.6	–39.4	257.4
All	897	108.5	103.4	–5.2	178.4

Table 8. Mean differences ($\bar{\mathrm{D}}$) and standard deviations (SD) for the differences between predicted and observed number of recruits of other broadleaves (ha^{-1} 5 yr^{-1}).

Part of the material	No. of observations	Observed mean	Predicted mean	$\bar{\mathrm{D}}$	SD
Altitude (m)					
–150	186	100.4	52.7	–47.8	269.0
151–300	255	38.4	34.9	–3.5	107.9
301–450	184	28.5	28.6	0.1	97.0
451–600	136	30.0	23.3	–6.7	162.8
600–	136	10.3	18.7	8.4	63.9
Latitude (°N)					
–60	248	64.4	40.0	–24.3	194.2
60–62	356	28.8	29.6	0.9	133.7
62–64	136	53.5	31.8	–21.7	168.3
64–66	95	24.8	31.4	6.5	84.7
66–	62	54.2	30.7	–23.5	191.6
Site index (m)					
–9.5	236	24.9	18.5	–6.4	114.7
9.6–12.5	285	34.4	27.1	–7.3	110.8
12.6–15.5	244	53.8	39.9	–13.9	199.0
15.6–	132	78.8	59.5	–19.2	219.7
Age class (yr)					
–5	97	19.8	20.5	0.7	57.4
6–10	176	27.5	27.5	0.0	96.9
11–15	163	68.5	40.2	–28.2	195.0
16–20	190	68.8	39.2	–29.7	230.9
21–25	140	46.9	37.1	–9.8	167.4
25–	131	12.5	27.9	15.4	39.8
No. of trees (ha^{-1})					
–1000	78	21.5	17.3	–4.2	68.9
1001–2000	262	31.0	25.7	–5.3	115.5
2001–3000	210	40.0	34.3	–5.7	162.9
3001–4000	128	62.2	42.0	–20.2	149.3
4001–5000	83	66.0	43.0	–23.0	204.5
5001–	136	55.6	40.0	–15.6	225.2
Forest type					
Spruce dom.	464	19.5	34.1	14.6	69.0
Pine dom.	190	11.8	23.2	11.4	53.0
Birch dom.	94	39.6	32.9	–6.7	107.6
Broadleaved dom.	149	162.4	42.7	–119.7	334.6
Region					
Eastern	573	36.7	31.8	–5.0	154.4
Middle	160	37.5	30.1	–7.4	108.1
Western	81	100.2	50.3	–50.0	235.1
Northern	83	48.7	31.2	–17.4	174.2
All	897	43.7	33.1	–10.6	158.6

4.2 Model Development and Evaluation

The relatively low R^2 for the conditional models (Table 4) were probably due to the large variability in observed number of recruits, rather than inappropriate specifications of the models. Residual plots indicated no serious violation of the assumption of constant variance, nor did they reveal any patterns that might indicate inappropriate specifications of the models. The relatively poor model fit for the species groups birch and other broadleaves were probably due to differences between species within the respective species groups. For instance, the species group other broadleaves consisted of 10 different species. The large deviations between predicted and observed number of recruits of birch and other broadleaves in forests dominated by these species were probably related to limitations in the description of dominant tree species. PN_{sp} in the models for birch and other broadleaves includes both species groups, and the effect of forest type on number of recruits of birches and other broadleaves had probably been estimated more precisely with a better description of species distribution.

A general problem related to recruitment models is that current recruitment depends on regeneration conditions of previous years (Leak and Graber 1976). The long time interval between seedling establishment and recruitment means that conditions (e.g. temperature, moisture) favouring recruitment of a particular species in the recruitment year are, most probably, different from the conditions under which the seedlings were established. Thus, if variables describing temperature and moisture in the year of recruitment are included in the models both number of recruits and species distribution may be biased (Yaussy et al. 1996). This also illustrates the weakness of using a relatively short growth period for model development.

The presented recruitment models are logical with respect to feedback from silvilcultural treatments. Planting and soil scarification increases N and A at a specific point in time because of a larger amount of seedlings and a shorter regeneration period. The positive relationship between N and A on one side and number of recruits on the other ensures that number of recruits increases with an increase in N and/or A. In addition, planting influences the species composition, and hence regulates number of recruits of a particular species through the variable $PN_{sp.}$ If, for example, Norway spruce is planted at a site with naturally regenerated birch and pine seedlings, then the proportion of Norway spruce increases, with increased recruitment of Norway spruce as a consequence. Precommercial thinning will also influence recruitment since stand density, as well as species composition, are represented in the models. If, for example, Scots pine is favoured at the expense of Norway spruce, the presented recruitment models will reflect this since the number of pines and spruces recruited will increase and decrease, respectively.

4.3 Model Application

The objective of the present study was to develop recruitment models for application in a growth simulator that will be a part of a large-scale forestry scenario model. Although stands or sample plots are used as basic units for calculations in such models, the target levels with respect to accuracy of the predictions are usually at national, regional or forest-estate level. Detailed studies of the forest dynamics at the stand level are seldom an important part of analysis performed with a large-scale forestry scenario models. It is also important to address the many sources of uncertainty related to the use of large-scale forestry scenario models in general, such as uncertainty related to input variables (e.g. Kangas and Kangas (1999) and Eid (2000)), to model errors of the numerous sub-models used for predictions (e.g. Kangas (1996)), to the stochastic processes of stand development (e.g. Pasanen (1998)) and to future price and cost variations (e.g. Ståhl (1994) and Leskinen and Kangas (1998)). The uncertainties related to the presented models should be seen in this perspective.

The differences between expected and observed number of recruits were relatively large in some strata, especially for birch and other broadleaves. The consequences of such large errors are, however, minor with respect to stand volume. An underestimation of, for example, 50 recruits ha^{-1} during a five year period will lead to an underestimation in total stem number after five years

of 2.5%, if the initial stem number is 1900 stems ha^{-1} and predicted recruitment is 50 recruits ha^{-1} instead of 100 ha^{-1}. In addition, most growth simulators include number of stems as an explanatory variable. The mortality sub-models will therefore compensate for an underestimation in number of recruits since total stand mortality will be underestimated. Diameter growth for each individual tree may also be overestimated if number of stems is underestimated. The overall underestimation considering volume at the stand level will therefore be lower than the corresponding underestimation of number of trees.

The developed models can be applied both deterministically and stochastically. The deterministic approach involves multiplying the probability of recruitment with the expected number of recruits. This approach will produce average values at the plot or stand level. If the models are used stochastically, the probability of recruitment within a plot is calculated. Whether or not recruitment is present is then determined by drawing a random number from a uniform distribution. If the stand is classified as recruited, the number of recruits is predicted directly with the conditional models. Separate models for each species group give a stochastic species distribution if the models are used stochastically. The stochastic approach should be restricted to plots with a plot size corresponding to the one used for model development (250 m^2). Mortality should be predicted before recruitment is added to the initial number of stems, if the models are used together with a mortality model, since the presented models predict net amount of recruits (i.e. all recruits are assumed to be alive at the end of the simulation period).

Although the data used for model development were based on 250 m^2 sample plots, the models can be applied deterministically to any plot size and also at the stand level. A small plot size means that the probability of recruitment is relatively small, while the conditional number of recruits per unit area will be relatively large. By calibrating conditional number of recruits per unit area with the probability of recruitment, an appropriate final level is determined. The opposite will be the case if the plot size is large or the models are applied at the stand level. A large plot size gives a large probability of recruitment, but a correspondingly low number of recruits per unit area, and

Table 9. Numbers of years the presented models are used for prediction of recruitment in young growth forests in the growth simulator described by Gobakken et al. (2005).

Site index	Species/species group			
	Norway spruce	Scots pine	Birch	Other broadleaves
6	47	53	25	50
8	40	47	25	45
11	37	40	25	40
14	33	33	25	35
17	30	30	25	30
20	27	27	25	25
23	23		25	20
26	20		25	15

accordingly an appropriate final level.

When the models are applied in the growth simulator described by Gobakken et al. (2005), the number of years that the models are used will vary with species groups and site index according to Table 9. The number of years that the models are used decreases with increasing site index for all species groups except birch, because of the positive effect of site index on the probability of recruitment as well as number of recruits. For birch there is no positive effect of site index, and hence the number of years to use the models does not vary with site index. Once the number of recruits is predicted with the probability and conditional models presented in this paper, each recruited tree needs to be given a diameter and height before its growth can be projected by the simulation system. Hence, the growth simulator includes a diameter distribution function and a diameter/height function (Gobakken et al. 2005)

Usually only larger stems are sampled in inventories for practical forest management planning, and smaller stems remain unsampled. This means that data "censorship" may occur when the threshold size used in the recruitment model is smaller than the minimum size of the trees sampled. Data "censorships" will bias prediction of stand development. One solution to avoid this is, of course, to sample all stems above the threshold size. Other possibilities are to develop models that either predicts the number and size of small

stems at the time of the inventory (Randall et al. 1988), or to make the recruitment models more flexible (see e.g. Shifley et al. (1993)).

5 Conclusion

Both the probability models and the conditional models were biologically realistic and logical. The large random error levels were probably due to lack of independent variables describing the stochastic process of seedling establishment rather than inappropriate model specification. Provided the general uncertainty connected to analysis performed with large-scale forestry scenario models, and the stochastic nature of recruitment, the presented models seem to give a satisfactory level of accuracy. However, since the models were developed from data collected over a relatively short time period they should be evaluated and, if necessary, revised or re-estimated when more data becomes available.

Acknowledgements

References

Adams, D.M. & Ek, A.R. 1974. Optimizing the management of uneven-aged forest stands. Canadian Journal of Forest Research 4(3): 274–287.

Agresti, A. 1996. An introduction to categorical data analysis. John Wiley & Sons, Inc., New York. 290 p.

Andreassen, K. & Tomter, S. 2003. Basal area growth models for individual trees of Norway spruce, Scots pine, birch and other broadleaves in Norway. Forest Ecology and Management 180: 11–24.

Belsley, D.A., Kuh, E. & Welsch, R.E. 1980. Regression diagnostics: identifying influential data and sources of collinearity. John Wiley & Sons, Inc., New York.

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

Eid, T. 2000. Use of uncertain inventory data in forestry scenario models and consequential incorrect harvest decisions. Silva Fennica 34(2): 89–100.

— & Hobbelstad, K. 2000. AVVIRK-2000 – a large scale scenario model for long-term investment, income and harvest analyses. Scandinavian Journal of Forest Research 15: 472–482.

— & Tuhus, E. 2001. Models for individual tree mortality in Norway. Forest Ecology and Management 154: 69–84.

— & Øyen, B.-H. 2003. Models for prediction of mortality in even-aged forests. Scandinavian Journal of Forest Research 18(1): 64–77.

Fremstad, E. 1997. Vegetasjonstyper i Norge. NINA temahefte 12: 1–297. (In Norwegian).

Gobakken, T., Lexerød, N. & Eid, T. 2005. A growth simulator based on models for individual trees. In: Forest Inventory and Planning in Nordic Countries. Proceedings of SNS Meeting at Sjusjøen, Norway, September 6–8, 2004. Ed. Hobbelstad, K. NIJOS rapport 9/2005: 243–248.

Hoen, H.F. & Eid, T. 1990. A model for analysis of treatment strategies for a forest applying standvice simulations and linear programming. Rep. Norw. For. Res. Inst. 9/90: 1–35. (In Norwegian with English summary).

— , Eid, T., Veisten, K. & Økseter, P. 1998. Økonomiske konsekvenser av tiltak for et bærekraftig skogbruk. Forutsetninger og metodebeskrivelse. Rapport fra skogforskningen – Supplement 6/98: 1–48. (In Norwegian).

Hosmer, D.W. & Lemeshow, S. 1989. Applied logistic regression. John Wiley and Sons, Inc., New York. 307 p.

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. CABI publishing, UK, London. p. 271–294.

Hyink, D.M. & Moser, J.W. 1983. A generalized framework for projecting forest yield and stand structure using diameter distributions. Forest Science 29(1): 85–95.

Kangas, A. 1996. On the bias and variance in tree volume predictions due to model and measurement errors. Scandinavian Journal of Forest Research 11(3): 281–290.

— & Kangas, J. 1999. Optimization bias in forest management planning solutions due to errors in forest variables. Silva Fennica 33(4): 303–315.

Kozak, A. & Kozak, R. 2003. Does cross validation provide additional information in the evaluation of regression models? Canadian Journal of Forest Research 33(6): 976–987.

Leak, W.B. & Graber, R.E. 1976. Seedling input, death, and growth in uneven-aged northern hardwoods. Canadian Journal of Forest Research 6: 368–374.

Leskinen, P. & Kangas, J. 1998. Modeling and simulation of timber prices for forest management planning calculations. Scandinavian Journal of Forest Research 13(4): 469–476.

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

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

Montgomery, D.C. & Peck, E.A. 1992. Introduction to linear regression analysis. John Wiley & Sons, Inc., New York. 527 p.

Moser, J.W. 1972. Dynamics of uneven-aged forest stands. Forest Science 18(3): 184–191.

NIJOS 2000. Skog 2000, Statistics of forest conditions and resources in Norway. Norwegian Institute of Land Inventory, Ås. 7/99. 84 p.

Pasanen, K. 1998. Integrating variation in tree growth into forest planning. Silva Fennica 32(1): 11–25.

Randall, B.L., Ek, A.R., Hahn, J.T. & Buchman, R.G. 1988. STEMS model projection capability with incomplete tree list input data. Northern Journal of Applied Forestry 5: 190–194.

SAS Institute Inc. 1999. SAS OnlineDOC (R), Version 8. SAS Institute Inc., Cary, NC. 956 p.

Shifley, S.R., Ek, A.R. & Burk, T.E. 1993. A generalized methodology for estimating forest ingrowth at multiple threshold diameters. Forest Science 39(4): 776–798.

Sterba, H., Golser, M., Schweiger, J. & Hasenauer, H. 1997. Modelling recruitment and growth of natural regeneration. Centralblatt für das gesamte Forstwesen 114(1): 11–33.

Strand, L. 1967. Høydekurver for bjørk. In: Braastad, H. (ed.). Produksjonstabeller for bjørk. Meddelelse fra Det norske Skogforsøksvesen 22. p. 291–296. (In Norwegian).

Ståhl, G. 1994. Optimal stand level forest inventories under deterministic and stochastic stumpage value assumptions. Scandinavian Journal of Forest Research 9: 405–412.

Tveite, B. 1976. Bonitetskurver for furu. Norsk institutt for skogforskning. Intern rapport. (In Norwegian).

— 1977. Bonitetskurver for gran. Meddelelser fra Norsk institutt for skogforskning 33: 1–84 (In Norwegian).

Vanclay, J.K. 1989. A growth model for north Queensland rainforests. Forest Ecology and Management 27: 245–271.

— 1992. Modelling regeneration and recruitment in tropical rainforest. Canadian Journal of Forest Research 22(9): 1235–1248.

— & Skovsgaard, J.P. 1997. Evaluating forest growth models. Ecological Modelling 98: 1–12.

Wikberg, P.-E. 2004. Occurence, morphology and growth of understory saplings in Swedish forests. Doctoral thesis. Department of Silviculture. Swedish University of Agricultural Sciences, Umeå.

Yaussy, D.A., Sutherland, E.K. & Hale, B.J. 1996. Rule-based, individual-tree regeneration model for forest simulators. In: Skovsgaard, J.P. & Johanssen, V.K. (eds.). Modelling regeneration success and early growth of forest stands. Proceedings from the IUFRO Conference, Copenhagen, 10–13 June. Danish Forest and Landscape Research Institute. p. 176–182.

Total of 39 references