The development of a simple basal area increment model

Georg E. Kindermann[*]

August 30, 2011

Abstract

In most cases forest practice in Austria use yield tables to predict the growth of their forests. Common yield tables show the increment of pure even-aged stands which are treated in a way the table developer recommends. The usage of these tables in stands which are either uneven-aged, mixed or treated in another way, may lead to inaccurate predictions. To avoid these problems, forest growth models have been developed. Until now they are not widely used in Austria. One reason may be, that most of the models need some input parameters which are usually not gathered by companies. In this work a basal area increment per hectare model has been developed which is based on the input parameters: diameter at breast height, hight to diameter ratio, top height at age 100 years and a selection out of several simple competition indices (growing space, basal area of larger trees, competing basal area, crown cross sectional area, crown competition factor, d/dg, d–dg, basal area and stand density index) which are distance independent. The model parameterization was done with seven diferent statistical methods (linear regession, linear mixed effect model, resistant linear regression, local polynomial regression, lazy learning model, random forest model and neural network model). By using only few input-parameters it should be possible to parameterize this model for many local areas by using inventory data sets of the specific region. The model works in pure and mixed stands of spruce and beech at the Rosaliengebirge. The observed average diameter increment per 5 years is 18.1 mm for spruce and 21.1 mm for beech. The average difference of the predicted and observed diameter-increment on a validation dataset is 0.3 mm for spruce and -0.3 mm for beech within 5 years and the estimated additional spread caused by the model is ± 4.5 mm/5 years for spruce and ± 4.0 mm/5 years for beech.

Keywords: Tree – Forest growth, Basal area increment, Competition index, Decision support

1 Introduction

In forest management the decision whether or not a thinning should be made is usually based on the experience of the forest surveyor. To support this experience the profit or loss, which is caused by a planned thinning, should be predictable. The growth response to the thinning can be described with tables or growth models. The tables in Kramer and Akça (2002) show the loss of growth for a certain degree of stocking. They are simple to use but they do not distinguish between tree size, yield level or production class.

In most cases in Austria forest companies use common yield tables to predict the volume of their stands. Forest growth models are only used in few cases, and when they are used, the model was served by the model developer. Also it is common that the models need input parameters which are usually not collected by companies. So they are forced to collect this additional information or to make a good guess what the value would be in their woods, if they want to use these models. For spatial explicit models, up until now, there is no other way than to generate the positions of every tree because the measurement of the coordinates of large areas will take too much

[*]International Institute for Applied Systems Analysis (IIASA), Schlossplatz Ĩ, A-2361 Laxenburg, Austria

time. STRUGEN (Pretzsch, 1997) is a popular program which is capable generating tree positions. There are only few tree-species which have been published that the program can generate. STRUGEN generates the mixture of different species and the distance between the trees is random as long as a threshold distance is not fallen below. It is not possible to produce a uniformly or a clumped structure of the tree positions. Pretzsch (1995); Mailly et al. (2003) pointed out that spatial information can improve an increment model. Biging and Dobbertin (1995); Windhager (1998); Holmes and Reed (1991) on the other hand have come to the conclusion, that there is no difference in the correlation between distance independent and distance dependent competition measurements. In fact it is unimportant whether a spatial explicit model is better or not, as long as the spatial information is not available for large areas.

Monserud and Sterba (1996) developed a distance independent basal area increment model for Austria. This model is based on the Austrian national forest inventory plots to predict the increment on these plot for different scenarios. It was also used on inventory plots which did not belong to the national forest inventory (Sterba and Monserud, 1997). But for this purpose, additional observations were necessary. To avoid these additional observations, a model which uses:

- Tree diameter at breast height – 1.3 m above ground (*dbh*)

- Tree height (*h*)

- Yield level

- Stand density/competition

will be parameterized.

To decide which stand density index and competition index respectively should be used, some of these indices will be compared regarding their benefit in the increment model.

The relation between basal area increment per standing area and dbh , h, yield level and stand density need not to be linear. To take this into consideration the following model types have been parameterized:

- Linear regression

- Linear Mixed-Effects Models (Venables and Ripley, 2002)

- Robust Fitting of Linear Models (Venables and Ripley, 2002)

- Local polynomial regression fitting (Cleveland et al., 1992)

- Lazy learning for local regression (Bontempi et al., 1999)

- Breiman and Cutler's random forests for classification and regression (Breiman, 2001)

- Neural network (Zell, 1994)

2   Datasets

The dataset was collected in the experimental forest of the university of natural resources and applied life sciences. It is situated in the Rosaliengebirge, 60 km south of Vienna. The observed forest has an area of 930 ha. The elevation ranges form 320 m to 725 m. The main species are Norway spruce (*Picea abies (L.) H. Karst.*) (41 % of the area) and beech (*Fagus sylvatica L.*) (36 %). Scots pine (*Pinus sylvestris L.*) has an allotment of 10 % and European larch (*Larix decidua Mill.*) 5 % of the area.

In 1989 the inventory points have been established and marked permanently with a metal pipe. Every point represents an area of 2 ha and they are located on a regular raster. So the distance from one point to the next is approximately 141 m. In the first year every fifth point,

Table 1: *Overview of observed data*

	Parameterization	Validation	∑
Number of points	278	139	417
Points with spruce	177	85	262
Points with beech	141	71	212
5 year increments spruce	1726	883	2609
5 year increments beech	1233	482	1715
H_{100} spruce [m]	35.6 (24–43)	36.3 (24–43)	35.8 (24–43)
H_{100} beech [m]	30.9 (20–39)	31.5 (22–38)	31.0 (20–39)
dbh spruce [cm]	30.5 (6–66)	29.5 (6–66)	30.2 (6–66)
dbh beech [cm]	31.8 (6-63)	33.9 (6–66)	32.4 (6–66)
G/ha spruce [m^2/ha]	39.6 (4–76)	40.2 (4–64)	39.8 (4–76)
G/ha beech [m^2/ha]	34.7 (4–64)	35.3 (4–64)	35.0 (4–64)
h/d-ratio spruce	81 (38–147)	81 (42–125)	81 (38–147)
h/d-ratio beech	84 (40–173)	81 (46–174)	83 (40-174)
i_g/ha spruce [m^2/ha/5 year]	5.3	5.7	5.4
i_g/ha beech [m^2/ha/5 year]	4.6	5.0	4.7

For variables with a unit the arithmetic mean is shown. The minimum and maximum is given in brackets. H_{100}...top height at the age of 100 years (see section 3.1), *G/ha* ...basal area, *h/d-Ratio* ...tree height divided by the dbh and i_g...basal area increment.

with the point numbers 1, 6, 11, . . . was observed. The next year point 2, 7, 12, . . . and so on was observed. So every year a fifth of the points has been measured and the time from one observation to the next, for a certain point, took 5 years. The last observation of the dataset comes from 2004.

On these points an angle-count method (Bitterlich, 1948) with a multiplication factor of $4\,m^2$/ha was used to select the trees to measure. The distance and angle to the center of the plot of every selected tree with a dbh of 6 cm or larger was measured. So it was possible to assign the measurement data of different years to a single tree. Beside the position also the species, dbh and tree height was observed.

The use of the angel-count method has the advantage, compared to fixed area plots, that small trees are not overrepresented. This means, that the number of observed trees is more or less the same for small and for large trees. So there is no dbh-range overrepresented and no need for a weighted regression.

The dbh was measured with a calliper. The measure position of the calliper lies 1.3 m above a nail, which was placed at the base of the tree. The measure direction points to the center of the sample plot. The tree height was measured in the first years with the Relascope and until now with a Haglöf Vertex.

417 Points have been observed. Every third point , with all measured trees on this point, has been taken out of this dataset and put into a validation dataset. An overview of the dataset is shown in table 1. The distinction between spruce and beech does not mean, that on those plots only spruce or beech are represented. It means, that this basal area is the total of any species observed on plots where at least one beech and spruce respectively was represented.

The basal area increment per hectare was assessed the following way. The observed basal area of every tree was divided by the basal area of the stand. The result of this division can be described as the growing space of the tree (equation 1), which is proportional to its basal area proportion. For those trees which have been re-measured at the next sampling period their basal area increment can be calculated. By dividing the observed basal area increment of a tree by its growing space, the basal area increment per area for a single tree (equation 2) can be calculated.

Table 2: *Estimated measurement error and average increment of the dbh*

dbh	all	≤ 15	$>15 \leq 30$	$>30 \leq 45$	>45
Spruce					
Error [cm]	± 0.69	± 0.60	± 0.70	± 0.65	± 0.85
id_5 [cm/5 years]	1.81	1.26	1.98	1.72	2.03
N	1134	122	375	486	151
Beech					
Error [cm]	± 0.68	± 0.58	± 0.54	± 0.76	± 0.72
id_5 [cm/5 years]	2.11	0.88	1.88	2.36	2.71
N	811	112	218	332	149

Error…measurement error, id_5…average diameter increment in five years and N…number of trees.

$$\text{Growing space} = \frac{d^2 \cdot \frac{\pi}{4}}{G/ha} \tag{1}$$

$$i_g/ha = \frac{i_g}{\text{Growing space}} = \frac{d_5^2 \cdot \frac{\pi}{4} - d_0^2 \cdot \frac{\pi}{4}}{\frac{d_0^2 \cdot \frac{\pi}{4}}{G_0/ha}} = \left(\frac{d_5^2}{d_0^2} - 1 \right) \cdot G_0/ha \tag{2}$$

i_g is the difference of the basal area for a single tree between two measurements. d_0 is the dbh at the first observation and d_5 is the dbh at the re-measurement which was in this case 5 years later. G_0 is the basal area of the stand at the same time when d_0 is observed.

In this approach recruitment/ingrowth and mortality where not considered. So the basal area increment of the whole stand may be above or below the average basal area increment of the individual trees. This approach gives the possibility to share a mortality and ingrowth model around this basal area increment model. This will lead to a model which consists of independent modules which work together but each single module can also be replaced by another without the necessity to change the other modules.

In Table 2 the estimated dbh measurement error for spruce and beech is shown. The first column is over all dbh-ranges and the following columns are for a certain dbh-range.

The average measurement error of the observed diameter increment in 5 years for spruce is 54 % ($0.69 \times \sqrt{2}/1.81$) and for beech 46 % ($0.68 \times \sqrt{2}/2.11$). The measurement error was estimated by creating a linear regression for every tree which has at least been re-measured for 3 times. In this regression diameter was a function of the observation year. For the period of 10 years the assumption, that the diameter increment is linear, was made. The standard deviation of these regressions was taken as the measurement error. The estimated average measurement error of ± 0.7 cm is above the published error from Elzinga et al. (2005) which has a value of ± 0.471 cm. This lower error may be a result of using diameter tapes. Also the deviation from a linear diameter increment increases the standard deviation.

3   Methods

The basal area increment per hectare for a single tree will be described by its dbh, h/d-ratio, yield level and competition. So the model will be formulated as:

$$i_g/ha = f(\text{dbh, h/d} - \text{ratio, yield level, competition}) \tag{3}$$

Bearing in mind, that the model should be held simple to make the practical approach possible, the parameters of the model should be collected during a standard inventory by the company. As practice shows, even the dbh was not measured by the forest surveyor. Nevertheless this model will use this input parameters because if they were not collected they could roughly be estimated.

4

Also it will be possible to reduce the created model for a company if they are not willing to collect a certain parameter.

The parameterization of a model to the needs of a certain company and to the growth response of their woods should be no problem nowadays. Sterba and Monserud (1997) also re-calibrated their model by multiplying the forecast with a factor so that the prediction would not be biased. The prediction may be even better if the model is re-calibrated for a specific region. So the purpose of the developed model, with the coefficients of this specific region of the observed data, is not to be used outside this region without reparametrisation to the region where it should be used. The core of the model keep's the same only the estemated coefficients need an update. Because the model needs only usual observed data this should be possible without much aditional effort.

It should be emphasized, that this model does not describe the basal area increment of a single tree. It describes the basal area increment for a single tree divided by its growing space. So it will predict the basal area increment per hectare for a single tree. By predicting the basal area of one tree it can be assumed that the increment of a tree with a dbh of 100 cm will be higher than by a tree with a dbh of 1 cm. This assumption can be made because it is usual, that large trees occupy more area than small trees and the productivity of large trees should not decrease so dramatically, that the basal area increment of large trees was at the same level as that of small trees.

To see the effect of different parameters, first a model will be built which consists only of one input parameter. After the one parameter model, a model which consists only of tree measurements and a description of the site, will be built. This model should be simple and small and can consist of dbh, height, h/d-ratio and the top height at age 100. Afterwards different competition indices will be added to this model.

The model should not have an input parameter like the share compared to other trees. The reason for this is, by using this approach, the model has to know the share of each other species or speciesgroup. In this certain case only two species will be modeld in pure and mixed stands. But the aim is, that the model can be extended without affecting the rest of the model. By using the information of other trees, this is not possible. So it has to be proofed, if the model delivers unbiased results for pure and mixed stands. Because this approach splits up the stand for the species proportional to their basal area, an overestimation for on species and an underestimation for the other can be corrected by weighting this splitup process so that the overestimated species get a lower and the underestimated species get a larger share of the stand.

3.1 Yield level

The observation of the yield level, described as the average overall growth until age 100 years, will take a long time and may still not be accurate because thinning may have an influence on the growth. So it is common, to describe it with the dominant height at a certain age, the site-index. For most species in Austria this age is 100 years. On the other hand a certain top height at age 100 can still have a different volume growth. This is described by the production class. So, Assmann and Franz (1972) made yield tables for a high, middle and low production class. According to Assmann (1961) there is a high relation between the natural degree of stocking (the maximum possible basal area) and the production class. For a stand it is not possible to determine its production class. In most cases for a company it will be sufficient to be right in the average. So some stands will be above and some below the prediction. In this approach the site index is described with the top height at age 100 and the production class is considered by using the observed data from the region where the model will be used. In this way an average production class is realized.

In this study top height curves have been created with the observed dataset. By plotting the observed top height over the known age, a cloud of points, which have the shape of a cornucopia, is produced. In this case not only points but also short lines can be produced, because the plots have been observed more than one time. With this information, top height curves have been made for spruce and beech. For every plot the top height has been calculated according to Pollanschütz (1973) as the average height of the thickest trees which have together $12\,m^2/ha$ of

the basal area. This height has been calculated for every species separately as long as it has at least one tree which belongs to these thickest trees. On some plots it is possible that a species has no top height. In this case the species was not taken into consideration. In practice it is possible to derive a site index for a certain species by knowing the site index of other species.

The height curves have not been smoothed by a specific function. The curves where drawn directly to the plot and the height at every 10 year has been read out of the plot. These points have been joined together with a polynomial local regression and the top height for every age from 0 to 160 years has been calculated. This matrix of age and height was used to determine the top height at age 100 years for every plot by interpolating or extrapolating between the given top height curves.

3.2 Density and competition indices

Many competition and stand density indices have been developed in the past. An overview of distance independent indices are given by Biging and Dobbertin (1995); Holmes and Reed (1991); Mailly et al. (2003); Windhager (1998). In these articles also a comparison between different indices is made. Biging and Dobbertin (1995); Windhager (1998); Holmes and Reed (1991) come to the conclusion, that distance independent indices are as good as distance dependent indices. Mailly et al. (2003) on the other hand say that the spatial information has a benefit for a basal area increment model. Likewise Pretzsch (1995) emphasizes that the spatial stand structure has an influence on the growth of a tree. In this study the answer whether the spatial structure has an influence on the tree growth or not, will not be discussed because the stand structure usually is not noticed by a forest inventory.

Some models use, beside a stand density index, a measurement of the effect of release cutting. The difference between the competition index, calculated with all trees in the stand, and the other, calculated only with these trees which are also present at the next observation period, can describe these effect.

The competition indices in table 3 are used for the basal area increment model. The basal area of larger trees (BAL) (Wykoff, 1990) adds up the basal area of trees which have a larger dbh than the observed tree. So it gives a smaller tree a high and a bigger tree a low value. Because this work uses trees observed on sample plots, every observed tree represents many trees per hectare. So the simpe approach by sorting the trees by their dbh and starting with the largest bhd and summing up the basal area will not be satisfiable. Here the assumption was made, that one observed tree represent's the average tree and half of thees represented trees are slightly larger and the oter half is slightly smaler. This means, that the largest tree, which represents $4\,\mathrm{m}^2$ has a BAL of $2\,\mathrm{m}^2$. This will be roughly corect but coming to a correct BAL, these $2\,\mathrm{m}^2$ have to be reduced by half of the basal area of the observed tree, because one tree can not be competed by himself. The same method was used in the case, that two trees have the same dbh on one plot.

The competing basal area (CBA) works the same way, like the BAL, with the difference, that it counts only the basal area which is larger than its own dbh multiplied by 0.7. The shown calculation of the CBA is only one possibility to calculate the competition from above. It is also possible to change the condition for computing the CBA from $d_j > d_i \cdot c_1$ to $d_j > d_i \cdot c_1 + c_2$, where c_1 can be a positive value starting at 0 and c_2 can be a positive or negative number. So it will be possible to consider not only the larger trees. It is also possible to take smaller trees into account. On the other hand it is also possible to look only at trees which are larger than the observed tree.

The crown competition factor (CCF) (Krajicek et al., 1961) is a density index for a whole stand. It summarizes the crown area of open grown trees for all trees and divides it by the observed area. So if 50 % are covered by the potential crown area the CCF is 50. The potential crown area for a single tree was calculated after Hasenauer (1997). There the crown width of an open grown tree, with the same dbh as the observed tree, can be estimated.

The crown competition factor of larger trees (CCFl) (Ritchie and Hann, 1985, 1986) computes the CCF for a subject tree, where only trees which have a larger dbh than the subject tree are considered. So it will show how much a tree is suppressed by the others.

6

Table 3: Examined competition indices

Index	Condition	Equation
BAL_i	$d_j > d_i$	$\sum g_j / A_{obs}$
CBA_i	$d_j > d_i \cdot 0.7$	$\sum (g_j - g_i \cdot 0.49) / A_{obs}$
CCF	—	$100 \cdot \sum pca_j / A_{obs}$
$CCFl_i$	$d_j > d_i$	$100 \cdot \sum pca_j / A_{obs}$
$CC_{66,i}$	—	$\sum ca_{0.66h_i,j} / A_{obs}$
d/dg_i	—	d_i / d_g
$d\text{-}dg_i$	—	$d_i - d_g$
G/ha	—	$\sum g_j / A_{obs}$
GGm	—	$\frac{G/ha}{G_{max}/ha}$
SDI_R	—	$N/ha \cdot (25/d_g)^{-1.605}$
Growing space$_i$	—	$\frac{g_i}{G/ha}$

i. . . the observed tree, j. . . the competing tree or all trees of a stand if there is no condition to select competitors, A_{obs}. . . area of the observation, g. . . basal area of a tree, *pca*. . . potential crown area of an open grown tree, $ca_{0.66h_i}$. . . crown area at 66 % of the height of three i, d. . . dbh, d_g. . . diameter of the stem with the mean basal area, G_{max}. . . maximum basal area, *N/ha*. . . stem number per hectare, *BAL*. . . basal area of larger trees, *CBA*. . . competing basal area, *CCF*. . . crown competition factor, *CCFl*. . . crown competition factor of larger trees, CC_{66}. . . crown closure at 66 % of the subject tree's height and SDI_R. . . stand density index.

The crown closure at 66 % of the subject trees height (CC_{66}) (Biging and Dobbertin, 1995) is similar to the CCF with the difference, that the present crown size instead of the potential crown size is used and the crown closure is examined at a certain height above ground. To calculate this index a crown model is needed. In this study the crown shape from Pretzsch (2002) was used. There the crown was divided into two parts, the sunlit and the shade crown. To calculate the CC_{66} only the part of the sunlit crown was used. If the estimated crown width lies beyond the light crown, the maximum crown width was used. To compute the crown width with Pretzsch (2002) also the maximum crown width and the height to the crown base was needed. These were calculated concerning to Pretzsch et al. (2002).

The two indices d/dg and d-dg describe the social status of a tree. d/dg describes the relative and d-dg the absolute difference to the mean diameter. In the first case (d/dg) a value above 1 belongs rather to the dominant trees and a tree with a value below 1 belongs rather to the dominated trees. Using d-dg, a tree with a value above 0 will belong to the dominant trees and a tree with a value below 0 will belong to the dominated trees.

The basal area per hectare (G/ha) can be estimated very fast with a Relascope (Bitterlich, 1948). So it will be observed in many cases. If yield tables are used, the degree of stocking is calculated by dividing the observed basal area by the basal area of the table. Thereby it is considered, that fully stocked young stands have a lower basal area than fully stocked old stands. Unfortunately most yield tables are made for pure even aged stands and so they can only be used in mixed uneven aged stands, by assuming separate pure and even aged sub-stands. Kindermann (1998) showed, that there are many ways to do this and their results differ in a wide range. The main problem according to Sterba (1991) is, that the given basal area in the yield table is in most cases not at a certain level of the potential possible basal area.

By dividing the observed basal area with the maximal possible basal area (G_{max}), a stand density index, which ranges from 0 to 1, was created (GGm). The main problem is how to come to the G_{max}. A way is to use equation (5) from Goulding (1968); Sterba (1975, 1987, 1991) to get

Table 4: Coefficients for the CD–Rule

Species	a_0	a_1	b_0	b_1
Spruce (eq. 6)	0.000002148	0.5789	0.6025	—
Beech (eq. 6)	0.000002742	0.6734	0.8662	—
Spruce (eq. 4)	0.0000026078	0.557195	0.511004	-0.9210
Beech (eq. 4)	0.0000025307	0.663855	0.837084	-1.0973

the maximum basal area.

$$d_g = \frac{1}{A \cdot N + B} = \frac{1}{a_0 \cdot h_o^{a_1} \cdot N + b_0 \cdot h_o^{b_1}} \tag{4}$$

$$G_{max} = \frac{\pi}{16 \cdot A \cdot B} = \frac{\pi}{16 \cdot a_0 \cdot b_0 \cdot h_o^{a_1+b_1}} \tag{5}$$

$$d_g = \frac{1}{a_0 \cdot h_o^{a_1} \cdot N + b_0 \cdot h_o^{\frac{a_1}{-0.605}}} \tag{6}$$

$$G_{max} = \frac{\pi}{16 \cdot a_0 \cdot b_0 \cdot h_o^{\frac{0.395}{-0.605} \cdot a_1}} \tag{7}$$

h_o is the top height, N are the number of trees per hectare and a_0, a_1, b_0 and b_1 are coefficients which need to be estimated. The estimation of the coefficients was done by collecting d_g, h_o and N and parameterize equation (4). Considering the work of Reineke (1933) a reduction to 3 parameters as shown in equation (6) and (7) was possible (Sterba, 1981, 1987). Sterba and Monserud (1993) applied this concept to uneven aged mixed species stands. Unfortunately this approach needs a maximum basal area for a habitat type and the asymptotic dominant height, which are not available. So in this study the parameters of equation (6) were estimated for beech and spruce on plots where only beech and spruce respectively were present. The results are shown in table 4, where also the coefficients for equation (4) are taken from Kindermann (1998).

The stand density index (SDI_R) (Reineke, 1933) supposes to calculate the number of trees of a stand at the same density like the observed stand and a $d_g = 25$ cm. The equation of Reineke (1933) is based on trees per acre and an average dbh in inches. The conversion into the metric system was taken from Daniel and Sterba (1980). The SDI was developed for pure even aged stands. Instead of using the d_g Stage (1968) used diameter classes and used them in the equation of Reineke which leads to equation (8) where every measured dbh was used.

$$SDI_S = \sum N/ha_i \cdot \frac{25}{d_i}^{-1.605} \tag{8}$$

Woodal et al. (2002) came to the result that the SDI_S is biased in uneven-aged pure stands. For the purpose of comparison SDI is used here also in uneven aged and mixed stands. Both forms, SDI_R and SDI_S, have been used in the increment model.

If it is possible to find a maximum SDI, it is possible to create a density index which has values from 0 to 1 by dividing the observed SDI by the maximum SDI. This looks similar to the index where G/ha is divided by G_{max}/ha. And in fact Sterba (1981, 1985) showed that it is also possible to predict the SDI_{max} with the values from equation (4), which was used before to get G_{max}/ha. So the SDI_{max} can be calculated with equation (9).

$$SDI_{max} = \frac{b_0}{a_0} \cdot (2 \cdot b_0)^{-1.605} \cdot 25^{-1.605} \tag{9}$$

Using the values of the three parameter model (eq. 6) from table 4 the SDI_{max} for spruce is 1186 and for beech is 746. This will work in pure stands but not in mixed. To avoid the problem

of mixed stands, the SDI is calculated for a single tree. This is possible if the growing space of the tree is known. Equation (1) gives a very simple approach to calculate the growing space for a single tree. With this it is possible to calculate the trees per hectare (N/ha). By using the N/ha for a single tree and its dbh in equation (8) the SDI for a single tree can be calculated (SDI_A).

The growing space of a single tree can also be used as a competition index. Knowing only the growing space is in most cases not very informative, but by knowing the growing space and the dbh or tree height, a single tree can be treated like a pure, even aged stand. In this work the growing space was calculated as the proportion of the basal area of a tree to the basal area of the stand. This is a simple approach. It splits up the whole growing space with no overlap or unstocked area. The sum of the growing space of each tree in the stand will be the area of the whole stand. In some cases it will be helpful if there is beside the stocked area also the possibility to have unstocked open space. Furthermore it could be possible that the growing space, even in pure stands, is not linear proportional to the basal area and that different species in one stand have different relations between their growing space and their basal area. Also it is possible that the sum of growing space of a stand is larger than the area of a stand. For a first approach, by using the growing space, it should be good enough by calculating it by its basal area proportion.

3.3 Regression

All the statistics where done with R (R Development Core Team, 2005) version 2.1.1. Additional used packages were MASS (version 7.2-16), lazy (version 1.2-14), nlme (version 3.1-60), nnet (version 7.2-16) and randomForest (version 4.5-11).

To see the benefit of an observed variable the approach was made first to parameterize a model with 2/3 of the dataset and then try this model on the validation dataset and calculate the correlation between the observation and the prediction. With this method a bias or a deviation from the 1:1 line has no influence on the ranking of the models. That's important because both, a bias or a deviation from the 1:1 line, can be eliminated with a simple linear regression. So bias and deviation from the 1:1 line sould have no influence on the ranking. It is only important that the observations and the predictions from a model lie close to a straight line and exact this will be measured with the correlation.

As it can not be assumed, that the relation between any variable and the predicted basal area increment is linear or has the same relation like any other variable, every used model type is not set to a fixed function. It was tried to make every time a model which fits the relation in a good way. So if a linear regression model was built, the degree of a polynomial has been changed for every variable and also for every variable combination each time separately.

The linear model (*Lm*) was used as a traditional linear regression. But as mentioned before the independent variables need not to be used only in the linear form. It was tried to improve the regression by adding polynomial terms. Some variables would be used as x, x^2, x^3, ... as long as their regression-coefficient was significant. The same degree of polynomials was used in the robust fitting of linear models (*RLm*).

A distinction between fixed and random effects can be made, so linear mixed-effects models (*Lme*) were also used. The selection of a tree, which to observe, can be treated as a random effect and the observations on this tree (e.g. dbh, h/d-ratio, competition) as the fixed effects to the observed increment. In this study, 4/5 of the points are observed 3 times and 1/5 4 times within the observationperiod of 15 years. This makes a maximum of 3 observed increments. So a model with only 2 independent values has the same number of parameters as observations when an intercept value is used. Most trees have only 2 observations of the increment and so 1 independent value would lead to the situation that the number of parameters and observations is the same. So the decision was made that the sample point selection is the random part of the model.

The local polynomial regression fitting (*Loess*) fits many cases by default quite well but by changing the degree of smoothing with the parameter "span", an improvement can be seen. The "span" was changed until the deviation was as low as possible and the calculated curves produced smooth curves.

Lazy learning for local regression (*Lazy*) and Breiman and Cutler's random forests for classification and regression (*RF*) have been used with default values. By looking at the prediction it can be seen, that these do not produce smooth curves. So by plotting one variable against another the result jitters.

The here used neural network (*NN*) was a single-hidden-layer neural network with a linear output unit. The number of iterations was not limited. The iteration process was every time allowed to converge. The size of the net, described by the number of units in the hidden layer, was changed, so that the produced net fitted well but did not over-learn. The net was trained as many times necessary until the residuals dispersed homogeneously. If the residuals are not homogeneous the net tries to create some classes and does not show the relation in a continuous way. This can happen because the initialization of a net is made by random numbers and sometimes the fitting algorithm finds a local minimum which does not describe the relation in a sufficient way.

4 Results

4.1 One parameter model

By starting developing the basal area increment model, a very simple model, which consists only of one input parameter, was built. The difference between these models and a linear correlation between the observed increment and the input parameter is, that the models can also follow a non linear relation. The results of one parameter models for spruce are shown in table 5 and for beech in table 6.

It can be seen, that the tree height is the best predictor of the basal area increment for spruce and the growing space is the best for beech. The dbh and the growing space have also a great correlation for spruce. GGm, SDI_R and SDI_S have medium correlation for spruce and beech. For spruce also d-dg and for beech the CCF and SDI_A have a medium correlation. So it seems, that many competition indices did not provide too much information to a basal area increment model.

The linear regression, the mixed effect model, the resistant regression, the local polynomial regression and the neural net have in most cases similar correlations. Lazy learning and random forest have in most cases a not so good correlation.

4.2 Model with tree measurements

The effect of the combination of dbh, height, h/d-ratio and the top height at age 100 for spruce is shown in table 7 and for beech in table 8.

It can be seen, that the correlation does not improve so much, compared to the single parameter model, and that there exist some single parameter models which have a correlation level similar to the two or three parameter models. But on the other hand the variables used are significant and show an expected effect on the model. It can also be seen, that the lazy learning model and the random forest have improved and are in some cases not so far behind the other methods as in the one parameter model.

Because there is not so much difference between these models the decision which to take and afterwards to add different competition or density measurements, cannot be made only by looking at the correlation. So the first choice was to use the dbh because it is easy to observe and carries much information. The second choice goes to the h/d-ratio because this describes in an easy manner the competition history of a tree. If it is high, there has been much competition on the tree and the crown length is short. If it is low, there has not been so much competition on that tree and it will have a long crown. At last, the top height at age 100 was taken. This variable is used because it is expected, that the growth on a good site is higher than on a bad one. The tree height was not used, because it can be calculated with the dbh and the h/d-ratio. Furthermore there would be a high correlation between dbh and tree height which leads to the point, that using both, would give no or only little additional information.

Table 5: *Correlation between predicted and observed increment for spruce on the validation dataset – 1 parameter models*

Variable	Lm	Lme	RLm	Loess	Lazy	RF	NN
area	0.37	0.32	0.38	0.39	0.28	0.25	0.39
BAL	0.12	0.13	0.13	0.13	0.30	0.25	0.11
CBA	0.12	0.12	0.12	0.12	0.06	0.03	0.12
CC66	0.09	0.04	0.11	0.08	0.03	0.03	0.08
CCFl	0.03	0.03	0.03	0.02	0.03	-0.03	0.02
CCF	0.08	0.15	0.10	0.07	0.05	0.14	0.16
dbh	0.38	0.24	0.38	0.37	0.23	0.16	0.38
d/dg	0.08	0.10	0.08	0.08	0.07	0.06	0.12
d-dg	0.25	0.14	0.25	0.26	0.08	0.02	0.27
GGm3	0.28	0.29	0.27	0.29	0.12	0.17	0.28
GGm4	0.28	0.20	0.26	0.28	0.13	0.17	0.27
G/ha	0.09	0.07	0.09	0.10	-0.03	0.03	0.12
h/d	0.03	0.05	0.02	0.00	0.03	0.04	0.03
h	0.39	0.38	0.39	0.40	0.35	0.41	0.41
SDI_R	0.28	0.28	0.27	0.27	0.17	0.14	0.27
SDI_S	0.32	0.31	0.31	0.31	-0.08	0.06	0.06
SDI_A	0.22	0.19	0.22	0.23	0.17	0.11	0.22
H_{100}	0.18	—	0.17	0.15	-0.01	-0.10	0.18

area...growing space of a single tree, BAL...basal area of larger trees, CBA...competing basal area, CC66...crown cross sectional area at 66 % of the tree height, CCFl...crown competition factor of larger trees, CCF...crown competition factor, dbh...diameter at breast height, d/dg...dbh divided by d_g, d-dg...dbh subtracted by d_g, GGm3...basal area divided by the maximum basal area from the 3 parameter model and GGm4 from the 4 parameter model, G/ha...basal area, h/d...height to the diameter ration, h...tree height, SDI_R...stand density index according to Reineke, SDI_S...stand density index according to Stage, SDI_A...stand density index per growing space, H_{100}...top height at age 100 years, Lm...linear regression model, Lme...linear mixed effects model, RLm...resistant linear regression, Loess...local polynomial regression, Lazy...lazy learning model, RF...random forest model and NN...neural network model.

Table 6: *Correlation between predicted and observed increment for beech on the validation dataset – 1 parameter models*

Variable	Lm	Lme	RLm	Loess	Lazy	RF	NN
area	0.35	0.35	0.35	0.39	0.26	0.22	0.40
BAL	0.12	0.09	0.09	0.08	0.28	0.20	0.09
CBA	0.06	0.02	0.04	0.08	0.05	0.06	0.08
CC66	0.13	0.13	0.13	0.15	0.18	0.07	0.14
CCFl	0.09	0.04	0.07	0.10	-0.01	-0.03	0.10
CCF	0.35	0.34	0.35	0.35	0.20	0.13	0.35
dbh	0.25	0.20	0.25	0.23	0.10	0.08	0.26
d/dg	0.17	0.16	0.17	0.18	0.12	0.07	0.17
d-dg	0.13	0.14	0.12	0.13	0.09	0.06	0.17
GGm3	0.34	0.34	0.34	0.34	0.27	0.27	0.34
GGm4	0.34	0.34	0.34	0.34	0.19	0.16	0.34
G/ha	0.24	0.23	0.24	0.24	0.15	0.20	0.24
h/d	0.25	0.18	0.24	0.26	0.14	0.12	0.29
h	0.25	0.11	0.24	0.27	0.07	0.18	0.27
SDI_R	0.34	0.34	0.34	0.34	0.05	0.20	0.34
SDI_S	0.35	0.35	0.35	0.35	0.03	0.25	0.36
SDI_A	0.35	0.23	0.35	0.35	0.21	0.15	0.30
H_{100}	0.19	—	0.19	0.18	-0.07	0.02	0.17

Table 7: *Correlation between predicted and observed increment for spruce on the validation dataset – 2–3 parameter models*

Variable	Lm	Lme	RLm	Loess	Lazy	RF	NN
dbh + hd	0.40	0.38	0.40	0.37	0.34	0.34	0.41
dbh + h	0.39	0.38	0.38	0.38	0.34	0.37	0.41
dbh + H_{100}	0.38	0.35	0.38	0.37	0.21	0.19	0.39
dbh + hd + H_{100}	0.42	0.41	0.42	0.42	0.31	0.34	0.42

Table 8: *Correlation between predicted and observed increment for beech on the validation dataset – 2–3 parameter models*

Variable	Lm	Lme	RLm	Loess	Lazy	RF	NN
dbh + hd	0.35	0.32	0.33	0.34	0.30	0.26	0.38
dbh + h	0.30	0.25	0.26	0.33	0.23	0.25	0.26
dbh + H_{100}	0.32	0.24	0.27	0.32	0.11	0.13	0.28
dbh + hd + H_{100}	0.41	0.39	0.38	0.41	0.25	0.33	0.39

Table 9: Correlation between predicted and observed increment for spruce on the validation dataset –
4 parameter models (dbh + hd + H_{100} + Variable)

Variable	Lm	Lme	RLm	Loess	Lazy	RF	NN
area	0.47	0.41	0.47	0.46	0.37	0.43	0.50
BAL	0.43	0.43	0.43	0.43	0.32	0.35	0.43
CBA	0.43	0.42	0.43	0.43	0.31	0.37	0.42
CC66	0.43	0.42	0.43	0.43	0.32	0.39	0.43
CCFl	0.42	0.42	0.42	0.43	0.36	0.35	0.43
CCF	0.40	0.44	0.42	0.41	0.35	0.45	0.50
d/dg	0.49	0.44	0.49	0.47	0.35	0.46	0.51
d-dg	0.49	0.46	0.49	0.47	0.33	0.44	0.46
GGm3	0.51	0.49	0.51	0.52	0.44	0.45	0.51
GGm4	0.51	0.49	0.51	0.52	0.44	0.45	0.51
G/ha	0.47	0.46	0.47	0.48	0.33	0.41	0.47
SDI_R	0.51	0.49	0.51	0.50	0.22	0.43	0.51
SDI_S	0.52	0.50	0.52	0.51	0.23	0.44	0.51
SDI_A	0.48	0.46	0.48	0.47	0.41	0.40	0.50

4.3 Models with tree measurement and competition indices

In table 9 and table 10 the correlation of the 4 parameter models for spruce and beech respectively
are shown. Many indices improve the correlation from about 0.4 to 0.5. It can also be seen, that
the random forest model comes closer and closer to the group of the best models and that the
linear mixed effect model belongs no longer to the best models.

The growing space, GGm and SDI bring the correlation to 0.5 for spruce and beech. For spruce
also the CCF and d/dg and for beech the basal area bring the correlation up to 0.5.

It can also be seen, that some combinations have a worse correlation than the 3 parameter
model. The reason for this is, that the correlation was calculated on an independent validation
data set and so it is possible that the correlation will decrease even if there are more input param-
eters for a model.

It seems, that the model can further improve, if it uses more than one competition index. By a
comparison of the indices two groups of indices can be formed. One describes the stand density,
the other group describes the social ranking by giving the largest trees a low and the smallest
trees a high competition index. So it seems, by combining indices from the two different groups,
additional information can be provided to describe the increment in a better way.

In table 11 the correlations of the models which have two density indices are shown. Here
the correlation reaches up to 0.58 for spruce and 0.57 for beech. The given number in this table
is from the best model which was in the most cases either the linear regression or the neural net.
The highest correlation for spruce can be observed for the combination of G/ha and GGm and
the combination of SDI_R and d/dg which have a correlation of 0.58. For beech the best model
seems to be the combination of the growing space with the CBA or the CCFl. There are some
models which reach a correlation of 0.56.

4.4 Model with consideration of thinning effects

Because of the fact, that the shown correlation is only an estimation and many estimated correla-
tions are close together, no best model can be chosen out of the compared models however. One
model has to be chosen and brought to a mature state. The decission goes to the model which
uses G/ha and CBA as a measurement of competition. This model has good but not the best
correlations but it uses parameters which are easy to understand and can be handled, out in the
woods. This is important, because the results of the model should be used by the forester who

Table 10: *Correlation between predicted and observed increment for beech on the validation dataset – 4 parameter models (dbh + hd + H_{100} + Variable)*

Variable	Lm	Lme	RLm	Loess	Lazy	RF	NN
area	0.49	0.42	0.48	0.50	0.46	0.48	0.52
BAL	0.41	0.38	0.38	0.45	0.31	0.37	0.44
CBA	0.40	0.37	0.37	0.47	0.31	0.39	0.44
CC66	0.35	0.30	0.35	0.34	0.26	0.29	0.32
CCFl	0.39	0.39	0.38	0.41	0.28	0.35	0.38
CCF	0.47	0.45	0.46	0.48	0.34	0.41	0.47
d/dg	0.47	0.41	0.45	0.44	0.35	0.44	0.41
d-dg	0.46	0.42	0.44	0.45	0.41	0.42	0.40
GGm3	0.51	0.50	0.50	0.53	0.42	0.47	0.51
GGm4	0.51	0.50	0.50	0.53	0.43	0.47	0.50
G/ha	0.48	0.47	0.47	0.51	0.49	0.47	0.50
SDI_R	0.50	0.49	0.50	0.52	0.38	0.48	0.50
SDI_S	0.51	0.50	0.50	0.53	0.37	0.47	0.50
SDI_A	0.49	0.45	0.47	0.51	0.34	0.49	0.52

Table 11: *Correlation between predicted and observed increment for beech (left/down) and spruce (right/up) on the validation dataset – 5 parameter models (dbh + hd + H_{100} + Variable-A + Variable-B)*

|	area	BAL	CBA	CC66	CCFl	CCF	d/dg	d-dg	GGm3	GGm4	G/ha	SDI_R	SDI_S
SDI_A	0.48	0.54	0.56	0.49	0.54	0.50	0.56	0.54	0.48	0.48	0.48	0.53	0.56
area	—	0.53	0.55	0.52	0.53	0.48	0.55	0.54	0.50	0.48	0.47	0.51	0.52
BAL	0.56	—	0.44	0.45	0.43	0.51	0.51	0.50	0.55	0.55	0.55	0.55	0.56
CBA	0.57	0.45	—	0.44	0.43	0.51	0.52	0.49	0.56	0.56	0.56	0.56	0.56
CC66	0.55	0.38	0.37	—	0.43	0.50	0.50	0.48	0.48	0.48	0.49	0.52	0.53
CCFl	0.57	0.42	0.43	0.46	—	0.53	0.52	0.50	0.54	0.54	0.54	0.55	0.55
CCF	0.52	0.49	0.48	0.47	0.49	—	0.51	0.52	0.50	0.48	0.49	0.52	0.52
d/dg	0.56	0.50	0.52	0.48	0.47	0.49	—	0.52	0.57	0.57	0.54	0.58	0.57
d-dg	0.54	0.49	0.49	0.47	0.45	0.48	0.46	—	0.53	0.53	0.53	0.55	0.55
GGm3	0.52	0.52	0.53	0.51	0.53	0.50	0.56	0.55	—	0.48	0.58	0.53	0.55
GGm4	0.53	0.52	0.53	0.51	0.53	0.50	0.55	0.55	0.50	—	0.58	0.53	0.55
G/ha	0.51	0.52	0.54	0.55	0.53	0.52	0.54	0.54	0.56	0.56	—	0.53	0.55
SDI_R	0.52	0.55	0.53	0.51	0.56	0.49	0.55	0.54	0.54	0.54	0.52	—	0.52
SDI_S	0.51	0.53	0.54	0.51	0.56	0.50	0.56	0.54	0.56	0.56	0.53	0.52	—
SDI_A	0.51	0.54	0.55	0.52	0.55	0.52	0.55	0.55	0.52	0.52	0.51	0.54	0.56
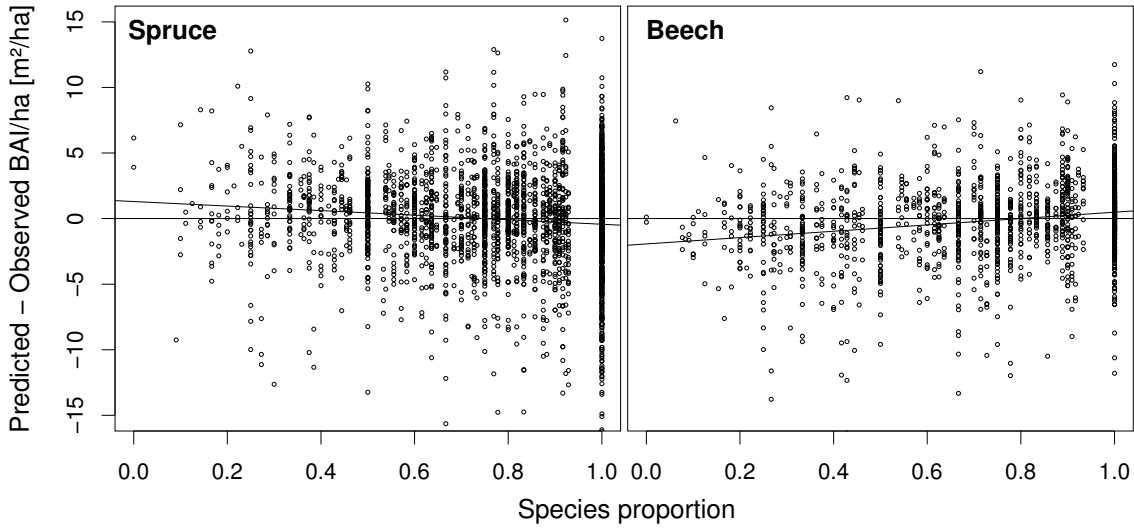
Figure 1: *Prediction-difference at different species proportions*

has to answer the question whether or not a thinning should be done. And when it will be done, how much of the wood should be harvested. The best model is the model which describes the observed relation in a good manner and is used.

Some models take also into consideration that a thinning has an effect on the growth. Usually this is described by a release in the competition. So in the model with G/ha and CBA also the release of G/ha ($\Delta G/ha$) and CBA (ΔCBA) was added.

In figure 1 the difference between the predicted and the observed basal area increment of a neural net model by different species proportion of spruce and beech is shown. Unfortunately a trend can be seen. The increment of spruce is overestimated and the increment of beech underestimated if they are in mixed stands. This unwanted trend can be corrected if the growing space of a tree is calculated with differently weighted basal area to each species. It turns out, that a multiplication of the basal area of beech by two removes this trend. So it seems that the model can now be used in mixed species stands of spruce and beech. By using this weighted calculation of the growing space, the basal area increment shown in table 1 also has to be recalculated. So the average increment per hectare and 5 years for spruce is for the parameterization dataset $5.9\,m^2$, the validation dataset $6.2\,m^2$ and for both $6.0\,m^2$. For beech it was $4.0\,m^2$ on the parameterization dataset, $4.1\,m^2$ on the validation and $4.1\,m^2$ for both together.

The coefficients of this model realized as a neural network are shown in table 12. How to use these coefficients for beech is described in equation (10, 11 and 12).

$$
\begin{aligned}
h_1 \quad = \quad & -240.0 - 3.024 \cdot d + 1.839 \cdot hd - \\
& -0.1172 \cdot H_{100} - 1.354 \cdot G/ha + 4.120 \cdot CBA - \\
& -0.3901 \cdot \Delta G/ha - 2.576 \cdot \Delta CBA
\end{aligned}
\tag{10}
$$

$$
\begin{aligned}
h_2 \quad = \quad & -0.9278 - 0.06094 \cdot d - 0.004598 \cdot hd + \\
& +0.01358 \cdot H_{100} + 0.05740 \cdot G/ha - 0.06140 \cdot CBA + \\
& +0.002037 \cdot \Delta G/ha - 0.04168 \cdot \Delta CBA
\end{aligned}
\tag{11}
$$

$$
i_g/ha \quad = \quad 1.139 - 3.214 \cdot \frac{1}{1 + e^{-h_1}} + 19.24 \cdot \frac{1}{1 + e^{-h_2}}
\tag{12}
$$

In table 13 the difference and deviation of the estimated diameter increment and the observed diameter increment is shown. A comparison with table 2 shows, that a dbh deviation of $\pm 0.69\,cm$ for spruce and $\pm 0.68\,cm$ for beech will be caused by the measurement error , which is equivalent

15

Table 12: Weights of the neural net with linear output

Spruce				
b→h1	d→h1	hd→h1	H_{100}→h1	G/ha→h1
-2.997	-0.04625	-0.008464	0.03122	0.02788
CBA→h1	ΔG/ha→h1	ΔCBA→h1	b→o	h1→o
-0.03126	-0.008228	0.008863	0.7345	144.4
Beech				
b→h1	d→h1	hd→h1	H_{100}→h1	G/ha→h1
-240.0	-3.024	1.839	-0.1172	-1.354
CBA→h1	ΔG/ha→h1	ΔCBA→h1	b→h2	d→h2
4.120	-0.3901	-2.576	-0.9278	-0.06094
hd→h2	H_{100}→h2	G/ha→h2	CBA→h2	ΔG/ha→h2
-0.004598	0.01358	0.05740	-0.06140	0.002037
ΔCBA→h2	b→o	h1→o	h2→o	
-0.04168	1.139	-3.214	19.24	

to a diameterincrement error of ±0.98 cm and ±0.96 cm respectively. If the estimated deviation of the measurement error is subtracted from the estimated prediction error of the model, an error of ±0.13 cm for spruce and ±0.10 cm for beech over all dbh–classes can be observed. If a dbh measurement error of ±0.47 cm, acording to Elzinga et al. (2005), is assumed, the deviation caused from the model will be ±0.45 cm for spruce and ±0.40 cm for beech. The lowest deviation can be calculated for the dbh-range of 15 cm to 30 cm. It is also shown that the average bhd increment in 5 years is overestimated by 0.03 cm for spruce and underestimated by 0.03 cm for beech. It is also shown that there is no trend of under– or overestimation by different dbh classes. The differences of -0.34 cm for the dbh-class larger than 45 cm and 0.14 cm for the dbh-class 15 cm to 30 cm is significant for spruce.

Also the correlation of the predicted and the observed diameterincrement (r_{id}) and the observed and predicted basal area increment (r_{ig}) has been calculated. The correlation of the diameterincrement is lower than the correlation of the basal area increment per hectare. The correlation of the basal area increment is above the correlation of the basal area increment per hectare.

In figure 2 the response of the basal area increment for spruce and in figure 3 for beech, to a different basal area, is shown. The diagrams were made for a tree whith a dbh = 25 cm, h/d-ratio = 80, H_{100} = 30 m, CBA = 30 % of the basal area, ΔGha = 0 and ΔCba = 0. The response of this tree can be seen in every picture as the solid line. It goes from $4\,m^2/ha/5$ years at a basal area of $10\,m^2/ha$ to $10.5\,m^2/ha/5$ years for spruce and for beech from $3.5\,m^2/ha/5$ years to $12\,m^2/ha/5$ years.

In the picture "dbh" in figure 2 it is shown, that a tree with a smaller dbh has a higher basal area increment than a tree with a larger dbh. Spruce shows a stronger effect to dbh than beech. An increase of the h/d-ratio causes a lower increment. Spruce reacts stronger to the h/d-ratio than beech. So to keep the increment high the h/d-ratio should be low. Also the site index shows the expected influence. Better sites show a higher basal area increment than poorer sites. Here it can be observed that beech is not so sensitive to the yield level as spruce.

Trees with a high competing basal area (CBA) show a lower increment than trees with a low one. In the graph "CBA" 30 % means, that 30 % of the basal area belong to the CBA. A tree in a stand of $20\,m^2/ha$ has therefore a CBA of $6\,m^2/ha$ (20×0.3). Dominant trees have a higher basal area increment per hectare than suppressed trees. Here beech and spruce show a similar reaction.

In the last two graphs (ΔG/ha, ΔCBA) the influence of a thinning to the increment on the growing space (eq. 1) of the remaining trees is shown. The point has to be emphasized that this graphs show the increment on the growing space of a single tree. So if some trees are harvested

Table 13: Estimated prediction error on the validation dataset of the dbh

dbh	all	≤15	>15 ≤30	>30 ≤45	>45
Spruce					
Error [cm]	±1.11	±1.17	±1.05	±1.07	±1.33
Δid_5 [cm]	0.03	0.06	0.14*	0.01	-0.34*
N	883	112	341	347	83
r_{id}	0.47	0.61	0.50	0.44	0.18
r_{ig}	0.61	0.62	0.54	0.47	0.21
Beech					
Error [cm]	±1.06	±1.05	±0.96	±1.07	±1.13
Δid_5 [cm]	-0.03	-0.19	-0.08	0.11	-0.13
N	482	67	120	186	109
r_{id}	0.54	0.68	0.53	0.41	0.27
r_{ig}	0.70	0.70	0.63	0.46	0.25

*…difference at 5 % significance level, *Error*…standard deviation between observation and prediction, Δid_5…difference between observation and prediction, *N*…number of trees, r_{id}…correlation between observed and predicted diameter increment and r_{ig}…correlation between observed and predicted basal area increment.

the growing space of a tree keeps the same and the former growing space of the removed trees has no increment. Until the next simulation period the free area is of course partitioned to the remaining trees. The graph "ΔGha" shows the reaction if a percentage of the basal area is removed and the competing basal area (CBA) stays the same. This is similar to a thinning from below. It can be seen, that the increment of the remaining trees of spruce increases if a thinning from below was done. Beech shows no reaction. In the graph "ΔCBA" the same percentage, like in the graph of "ΔGha", of the basal area has been removed but in this case 95 % of the removed basal area belongs to the CBA. So this is similar to a thinning from above. Beech shows an increase of its increment if the CBA is reduced and spruce shows no reaction.

Figure 4 shows graphs for a pure stand of spruce (left) and beech (right) respectively with dbh = 25 cm, h/d-ratio = 80, H_{100} = 30 m, CBA = 30 % of G/ha and without any thinning. The thick solid line shows the basal area increment at different basal areas and the thick dotted line shows the diameter increment for these trees. In the upper pictures it is shown, with the thin lines, how the increment reacts if 40 % of the basal area is removed and the assumption was made that 50 % of this removed basal area belongs to the competing basal area. It can be seen, that the overall basal area increment decreases. In the dense part the loss of increment is below 40 %. This is because the remaining trees are increasing their increment as a reaction to the harvest, which can be seen in the higher diameter increment. It also can be seen, that beech reacts much more than spruce. In the graphs below, the effect of an increasing h/d-ratio is shown. This is of interest because dense stands have low diameter increments and lead to high h/d-ratios. It can be seen that the increment decreases if the h/d-ratio increases. For beech this effect is not as high as for spruce.

5   Discussion

The comparison of different competition indices with one fixed function or simply calculating the linear correlation, as it was done by some works in the past, would not show which index can provide most benefit to a basal area increment model. For example Holmes and Reed (1991) calculated the linear correlation between different competition indices and annual diameter growth. Biging and Dobbertin (1995) created one fixed equation where the potential growth is reduced by multiplying it with $\exp(-a\text{CI}^b)$, where CI is the competition index and *a* and *b* are coefficients
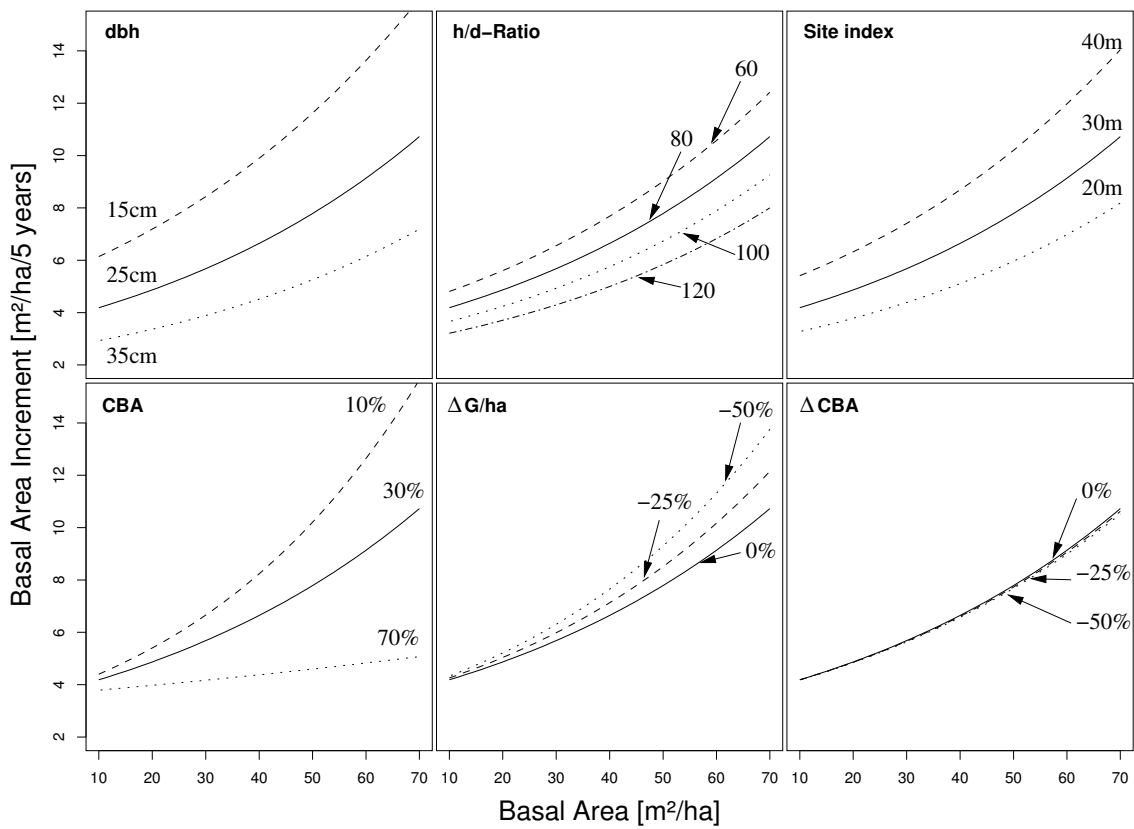
Figure 2: *Growth-response of spruce to different stand-densities*
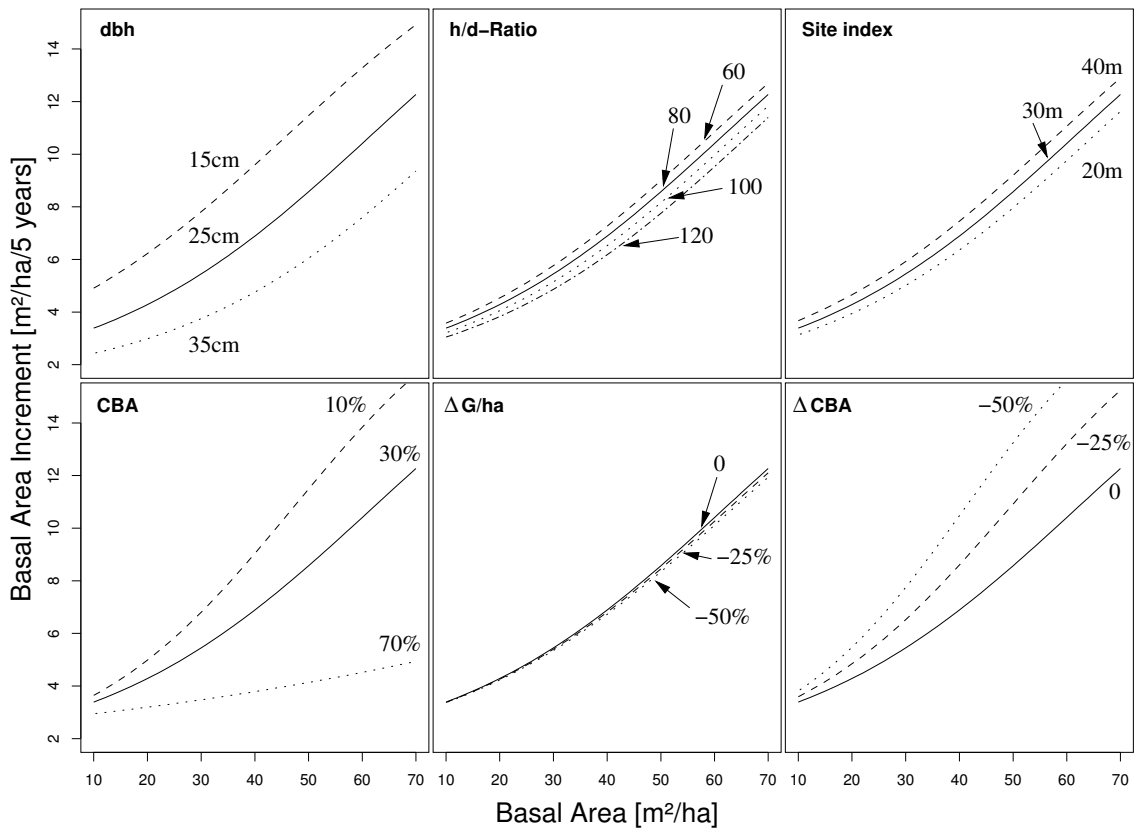
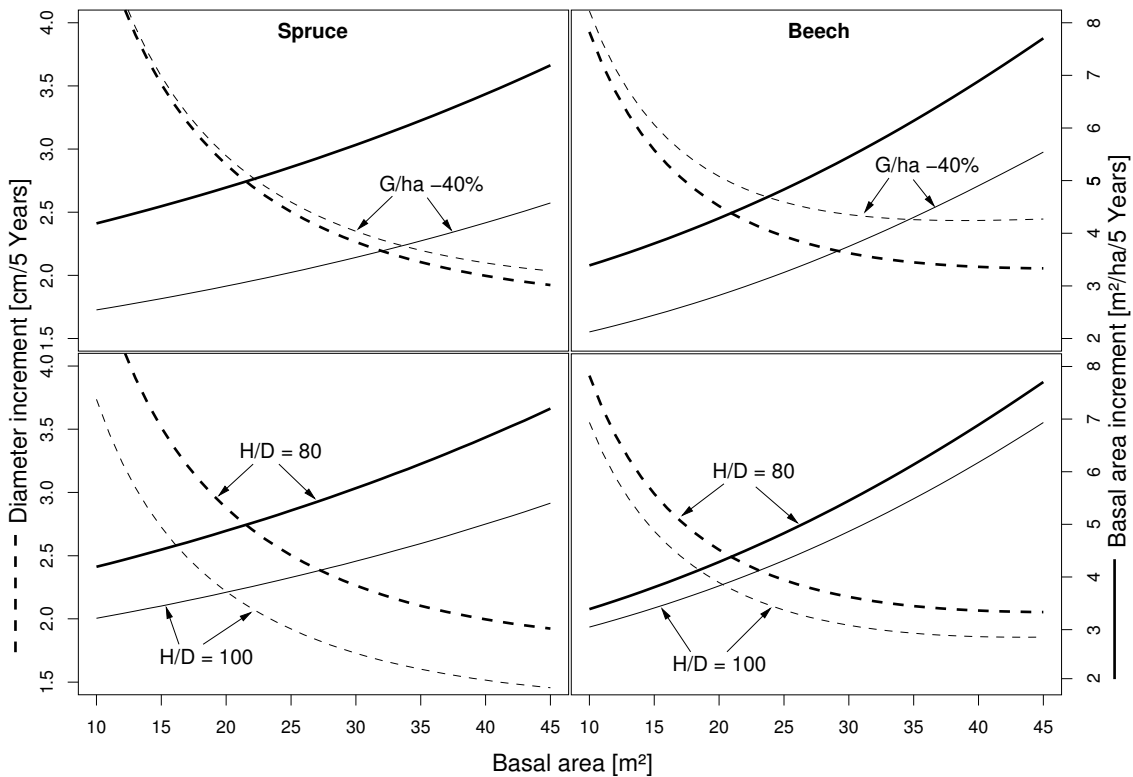*Figure 3:* *Growth-response of beech to different stand-densities*

Figure 4: Growth response to thinning and to different h/d-ratios

which have been parameterized for every competition index. This allows some transformation of the competition index but still can not follow any curve. Also Windhager (1998) and Mailly et al. (2003) used different competition indices in a linear equation to predict the logarithm of the basal area increment. This approach is able to find the competition index which has the best information for that specific model but cannot say which index has the most information on growth overall. If a variable is capable to describe a relation in a good manner, but the relation is not linear, this variable will not score well by calculating its linear correlation. Here were used seven different types of models and also, if it was possible, the model was not held to one fixed formula. It turned out that the results differ in a wide range between different model types. But it can also be seen that an index, which increases the correlation in one model, in most cases also increases it in the other models.

It seems also very important that the model improvement was not measured on the parameterization dataset. The improvement has to be measured on a validation dataset. This has especially to be mentioned when models are used which have a lot of internal parameters and are capable to reproduce any dataset, if every point can be distinguished by the input variables.

The used models are capable to predict also negative increments. This has the disadvantage that a negative prediction is possible but has also the advantage, that it can use also measurements which have a negative increment. By predicting the logarithm of the increment it is only possible to predict positive increments. This approach protects the model from a negative prediction. For parameterization all observations which have a negative increment, or an increment of zero, can not be used. If these observations are not used, or set to a low number, the average of the observed increment is increasing, which has to lead to a model which overestimates the increment. To correct this overestimation also the positive increment has to be reduced as long as the calculated average increment has the same value like the original observation before the negative observations have been modified.

In contrast to most of the other basal area increment models for individual trees, the approach

was made, to predict the basal area increment per area. Most other models predict the diameter increment or the basal area increment of single trees. To predict the increment per area for single trees the growing space for each tree has to be known. In the first try the growing space was calculated proportional to the basal area. This approach seems to work in pure stands but brings an underestimation for beech and an overestimation for spruce if they are mixed. By weighting the basal area this deficiency can be remedied. It turned out, that by multiplying the basal area of beech by two, the estimation of the increment for spruce and beech shows no longer a trend, if they are in mixed stands. The correction was necessary because the species proportion was not taken into the model like Andreassen and Tomter (2003) did. To leave the species proportion out of the model was made with intention, because if the proportion was taken into the model, it has not only an influence at which proportion the examined species is in the stand. It is also necessary to know the proportion of the other species in the forest. So the number of input parameters increases by the number of species or species groups the model should handle.

The proportion of 1:2 for spruce:beech in this case need not to be the same in other regions. Röhrl (1942) came also to the proportion of 1:2 between spruce and beech and Prodan (1959) comes to a relation of 1:1.8 for ages up to 60 years and 1:1.4 in stands over 60 years. So the solution by weighting the basal area of beech by 2 is similar to other approaches which split up the stand between different species.

Also, as mentioned before, some models predict the basal area increment for a single tree. This has been done for example by Wimberly and Bare (1996), Monserud and Sterba (1996) or Andreassen and Tomter (2003). By predicting the basal area increment, it is most likely that a tree with a large dbh has also a large increment. In table 13 the correlation of the basal area increment (r_{ig}) is higher than the correlation of the diameter increment (r_{id}) or the basal area increment per hectare. The assumption, that larger trees can have a higher basal area increment than smaler trees, is supported by this observation. So it is not surprising, that Monserud and Sterba (1996) or Andreassen and Tomter (2003) came to the conclusion, that the tree size describes most of the variation. Monserud and Sterba (1996) calculate that for spruce 44.2 % of 58.2 % and for beech 47.2 % of 63.0 % are described by the tree size. They mean by tree size the dbh and also the crown ratio. So it cannot be distinguished between the influence of dbh and crown ration.

In table 14 the relative importance compared to the most important variable is shown for the linear regression, the neural net and the random forest. The column for the linear model and the neural net was calculated by leaving one variable out of the model and observe how much the variance will increase by doing this. The increment of spruce, described by the neural network, seems to be described at most by the dbh. G/ha has only 24 % of the importance of the dbh and CBA only 14 %. For beech it seems that the importance is spread over different variables. Dbh has the most importance than CBA follows with 87 %, G/ha with 52 % and the h/d-ratio with 24 % of the importance of G/ha. The linear regression comes to similar results like the neural net. Some larger differences can be seen for beech by the variables G/ha and dbh. The column for the random forest model was calculated with the function "importance". Here also the variable which has the highest value gets 100 % and the others their proportion to this variable. In this case for spruce the dbh is still the most important variable but the distance to the others is not as large as by the other estimations. Here the importance goes from H_{100}, CBA and h/d which have an importance in the range of 60 %, down to G/ha, ΔCBA and ΔG/ha which are below 40 %. For beech this ranking goes from dbh over h/d, CBA, H_{100}, G/ha, ΔCBA to ΔG/ha.

So it can be seen that dbh plays every time an important role and the variables which describe the amount of thinning describes least. This is not surprising, because most of the time the variables ΔCBA and ΔG/ha have a value of zero or near to zero and so they can not provide any information in this cases. But nevertheless when a harvest has been done their information will be helpful for the model. The importance of some variables can also be estimated by looking at figure 2 and figure 3 which show the response of the model to the change of some variables.

It can be seen that the calculated models have never such high correlations like other basal area increment models. This may have several reasons. One is that the correlation was always calculated on the validation data set. Another is that the basal area increment per hectare was calculated and so the correlation between dbh and basal area increment was reduced. Also the

Table 14: Information content of input-variables in % of the most important variable

	Spruce			Beech		
	Linear	NN	R.Forest	Linear	NN	R.Forest
CBA	18.0	13.9	60.7	73.3	87.4	78.3
ΔCBA	0.1	0.1	19.7	3.1	2.6	24.5
dbh	100.0	100.0	100.0	73.5	100.0	100.0
G/ha	29.6	23.7	35.8	100.0	52.1	52.5
ΔG/ha	0.7	0.7	17.4	0.1	0.0	19.9
h/d	5.4	4.0	57.6	29.5	22.9	79.9
H_{100}	9.1	5.8	68.4	1.3	0.6	58.1

measurement error of the dbh is high compared to the observed increment (see table 2). The model prediction can not be better than the deviation of the measurement error. By comparing the deviation of the predicted dbh and the deviation caused by the measurement error, the additional error caused by the model is in the range of ± 4.5 mm/5 years for spruce and ± 4.0 mm/5 years for beech. This additional deviation is 25 % of the average increment for spruce and 19 % for beech. The observed bias of 0.3 mm/5 years over all dbh-ranges seems to be small enough to correspond to the needs of the forest practice.

As described by the used dataset, the regional extent of the observed area is not huge. The model relays on observations in an area of approximately 1000 ha. Also the top height curves have not been taken from yield tables. They have been made out of the observed data from this region. So these results should not be used unscrupulously in other regions. The model will also only handle those thinning regimes with confidence that are shown in the dataset.

The aim of this work was to show a possible template for creating regional models to supply forest surveyors with additional information on how their forests react on thinning. The area, where this template can be used, should be more or less homogeneous and comparable.

This model belongs to inventory data which were usually observed by a forest company. Unfortunately it needs a re-measurement of the same trees some time later, to observe their increment, which has occurred in the meantime. The time between the re-measurement need not to be 5 years. So it seems that permanent points which are measured from time to time are indispensable. But it is also possible to measure the diameter increment with an increment borer. By doing this, the model can be parameterized without any remeasurement. The only thing which must be estimated by this way is, the amount of basal area which has been removed in the last years.

It seems that the increment will be at the highest level if the stands are as dense as possible. This may only be true if the mortality was not taken into account. This model describes the increment per hectare only for those trees which will survive the next 5 years on their growing space. In fact the mortality will bring down the increment in dense stands so that a maximum stand density can not be exceeded. Also the diameter increment is in dense stands not as high as in spare stands. So the h/d-ratio will increase in dense stands. A high h/d-ratio on the other hand brings the increment down (see figure 2, 3 and 4). How the h/d-ratio is changing can be estimated if a height increment model is used in combination with a basal area increment model. So this model needs the combination with a height increment and a mortality model.

were provided by Oskar Franklin and Otto Eckmüllner.

References

Andreassen, K., Tomter, S. M., 2003. Basal area growth for individual trees of Norway spruce, Scots pine, birch and other broadleaves in Norway. Forest Ecology and Management 180 (1), 11–24.

Assmann, E., 1961. Waldertragskunde (The principles of forest yield study, 1970). BLV.

Assmann, E., Franz, F., 1972. Vorläufige Fichten–Ertragstafel für Bayern 1963, 2nd Edition. Forstliche Forschungsanstalt, Institut für Ertragskunde, München.

Biging, G. S., Dobbertin, M., 1995. Evaluation of competition indices in individual-tree growth models. Forest Science 41 (2), 360–377.

Bitterlich, W., 1948. Die Winkelzählprobe. Allgemeine forst- und holzwirtschaftliche Zeitung 59 (1/2), 4–5.

Bontempi, G., Birattari, M., Bersini, H., 1999. Lazy learning for modeling and control design. International Journal of Control 72 (7/8), 643–658.

Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.

Cleveland, W., Grosse, E., Shyu, W., 1992. Local regression models. In: Chambers, J., Hastie, T. (Eds.), Statistical Models in S. Wadsworth & Brooks, Ch. 8.

Daniel, T. W., Sterba, H., 1980. Zur Ansprache der Bestandesdichte. Allgemeine Forstzeitung 91 (6), 155–157.

Elzinga, C., Shearer, R. C., Elzinga, G., 2005. Observer variation in tree diameter measurements. Western Journal of Applied Forestry 20 (2), 134–137.

Goulding, C. J., 1968. Simulation techniques for a stochastic model of the growth of douglas-fir. Ph.D. thesis, University of British Columbia Vancouver Canada.

Hasenauer, H., 1997. Dimensional relationships of open-grown trees in Austria. Forest Ecology and Management 96 (3), 197–206.

Holmes, M. J., Reed, D. D., 1991. Competition Indices for Mixed Species Northern Hardwoods. Forest Science 37 (5), 1338–1349.

Kindermann, G., 1998. Die Flächenanteile der Baumarten. Master's thesis, Universität für Bodenkultur Wien.

Krajicek, J. E., Brinkman, K. A., Gingrich, S. F., 1961. Crown competition – A measure of density. Forest Science 7 (1), 36–42.

Kramer, H., Akça, A., 2002. Leitfaden zur Waldmesslehre. Sauerländer.

Mailly, D., Turbis, S., Pothier, D., 2003. Predicting basal area increment in a spatially expict, individual tree model: a test of competition measures with black spruce. Canadian journal of forest research 33 (3), 435–443.

Monserud, R. A., Sterba, H., 1996. A basal area increment model for individual trees growing in even- and uneven-aged forest stands in Austria. Forest Ecology and Management 80, 57–80.

Pollanschütz, J., 1973. Auswertung von Waldinventuren. In: 100 Jahre Hochschule für Bodenkultur in Wien, 4. Fachveranstaltungen der Studienrichtung Forst- und Holzwirtschaft, 1. Grundlagen der Forst- und Holzwirtschaft. Studienrichtung Forst- und Holzwirtschaft der Hochschule für Bodenkultur, pp. 355–368.

Pretzsch, H., 1995. Zum Einfluß des Baumverteilungsmusters auf den Bestandeszuwachs. Allgemeine Forst– und Jagdzeitung 166 (9–10), 190–201.

Pretzsch, H., 1997. Analysis and modeling of spatial stand structures. Methodological considerations based on mixed beech-larch stands in lower saxony. Forest Ecology and Management 97 (3), 237–253.

Pretzsch, H., 2002. Grundlagen der Waldwachstumsforschung. Parey.

Pretzsch, H., Biber, P., Ďurský, J., 2002. The single tree–based stand simulator SILVA: construction, application and evaluation. Forest Ecology and Management 162 (1), 3–21.

Prodan, M., 1959. Umrechnung von Massen in Flächenanteile. Forstarchiv, 110–113.

R Development Core Team, 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
URL http://www.R-project.org

Reineke, L. H., 1933. Prefecting a stand-density index for evenaged forests. Journal of Agricultural Research 46 (7), 627–638.

Ritchie, M. W., Hann, D. W., 1985. Equations for Predicting Basal Area Increment in Douglas-fir and Grand Fir. Tech. rep., Forest Research Lab, College of Forestry, Oregon state University, Research Bulletin 51.

Ritchie, M. W., Hann, D. W., 1986. Development of a tree height growth model for Douglas-fir. Forest Ecology and Management 15 (2), 135–146.

Röhrl, A. M., 1942. Forsteinrichtung. In: Rubner, K. (Ed.), Neudammer Forstliches Lehrbuch. Neumann–Neudamm.

Stage, A. R., 1968. A tree-by-tree measure of site utilization for grand fir related to stand density index. Tech. rep., USDA Forest Service Research Note INT-77.

Sterba, H., 1975. Assmanns Theorie der Grundflächenhaltung und die „Competition – Density – Rule" der Japaner Kira, Ando und Tadaki. Centralblatt für das gesamte Forstwesen 92 (1), 46–62.

Sterba, H., 1981. Natürlicher Bestockungsgrad und Reinekes SDI. Centralblatt für das gesamte Forstwesen 98 (1), 101–116.

Sterba, H., 1985. Das Ertragsniveau und der maximale Stand-Density-Index nach Reineke. Centralblatt für das gesamte Forstwesen 102 (2), 78–86.

Sterba, H., 1987. Estimating Potential Density from Thinning Experiments and Inventory Data. Forest Science 33 (4), 1022–1034.

Sterba, H., 1991. Zur Schätzung der Flächenanteile der Baumarten in Wuchsreihen. Centralblatt für das gesamte Forstwesen 108 (3), 297–308.

Sterba, H., Monserud, R. A., 1993. The maximum density concept applied to uneven-aged mixed-species stands. Forest Science 39 (3), 432–452.

Sterba, H., Monserud, R. A., 1997. Applicability of the forest stand growth simulator PROGNAUS for the Austrian part of the Bohemian Massif. Forest Ecology and Management 98 (1), 23–34.

Venables, W., Ripley, B., 2002. Modern Applied Statistics with S (Statistics & Computing). Springer-Verlag New York Inc.

Wimberly, M. C., Bare, B. B., 1996. Distance–dependent and distance–independent models of Douglas–fir and western hemlock basal area growth following silvicultural treatment. Forest Ecology and Management 89 (1), 1–11.

Windhager, M., 1998. Vergleich von distanzabhängigen und unabhängigen konkurrenzbeschreibenden Faktoren und ihre Auswirkung auf den Grundflächenzuwachs. Allgemeine Forst– und Jagdzeitung 170 (7), 133–136.

Woodal, C. W., Fiedler, C. E., Milner, K. S., 2002. Stand density index in uneven-aged ponderosa pine stands. Canadian journal of forest research 33 (1), 96–100.

Wykoff, W. R., 1990. A basal area increment model for individual conifers in the northern Rocky Mountains. Forest Science 36 (4), 1077–1104.

Zell, A., 1994. Simulation neuronaler Netze. Oldenbourg.