Modeling maximum basal area of even-aged pure beech and spruce stands in northwestern Germany

German Title:
Modellierung maximaler Grundflächen gleichaltriger Buchen- und Fichtenreinbestände in Nordwestdeutschland

Author:
Renke Christian von Seggern

Master's Thesis at the
Faculty of Forest Sciences and Forest Ecology,
Georg-August-Universität Göttingen

Date of submission:
December 20, 2017

Supervisors:
Prof. Dr. Winfried Kurth
Prof. Dr. Jürgen Nagel

Contents

Table of abbreviations and symbols

Symbol	Description	Unit
$CV(\lambda)$	cross validation sum of squares for smoothing parameter λ	
D	average diameter (by basal area)	cm
diag($\mathbf{x}$)	diagonal matrix with main diagonal $\mathbf{x}$	
erf	error function	
exp	exponential function	
$\mathbb{E}(x)$	expectation of variable x	
G	basal area	m^2
GAM	generalized additive model	
GAMLSS	generalized additive model for location, scale, and shape	
GCV	generalized cross validation	
$GCV(\lambda)$	generalized cross validation sum of squares for smoothing parameter λ	
GLM	generalized linear model	
h_{100}	top height, height of dominant trees	m
$h_{100}(x)_{\mathrm{YC}\,1}$	stand age variable (top height at age x if the stand were yield class 1)	m
$\ell(x)$	log-likelihood of x	
LM	Linear Model	
ln	natural logarithm	
log	decadic logarithm	
N	stand density	ha^{-1}
PI	absolute productivity index of stand (h_{100} at age 100 a)	m
SCAM	shape constrained additive model	
sign	sign function	
tr($\mathbf{A}$)	trace of square matrix $\mathbf{A}$	

1 Zusammenfassung

Die maximale oder natürliche Bestandesgrundfläche ist eine vielseitig einsetzbare Größe und kann beispielsweise als Referenz zur Bestimmung von Durchforstungsstärken dienen. Sie lässt sich jedoch nur auf Nullgrad-Dauerversuchsflächen zweifelsfrei bestimmen, wobei derartig ermittelte Werte streng genommen nur auf Bestände vergleichbarer Produktivität übertragbar sind. Da derartige Versuchsflächen nur in begrenztem Umfang vorhanden sind, muss auf andere Methoden der Grundflächenermittlung zurückgegriffen werden. Hierfür bieten sich statistische Modelle an. Die Auswahl derartiger Modelle ist allerdings eng begrenzt. Insbesondere fehlt es bisher an einer Möglichkeit, die maximale Bestandesgrundfläche unter Berücksichtigung von Unterschieden in der Bestandesproduktivität modellieren. Diese Arbeit zielt darauf ab, diese Lücke zu schließen. Zur Anwendung kommen generalisierte additive Modelle (GAMs), formeingeschränkte additive Modelle (SCAMs) sowie generalisierte additive Modelle für Lage, Skalierung und Form (GAMLSSs). All diese Modelltypen stellen eine Erweiterung generalisierter linearer Modelle (GLMs) dar. Im Gegensatz zu letzteren bieten sie eine höhere Flexibilität. Insbesondere sind GAMLSSs nicht auf Wahrscheinlichkeitsverteilungen der Exponentialfamilie beschränkt und gestatten die unmittelbare Schätzung nicht nur des Lageparameters, sondern aller jeweiligen Verteilungsparameter. Diese erhöhte Flexiblität bringt jedoch eine weniger intuitive Interpretation der Modellergebnisse mit sich.

Um sicherzustellen, dass ein Modell *maximale* Bestandesgrundflächen vorhersagt, sollte dieses mithilfe von Daten angepasst werden, die ihrerseits in Beständen mit wenigstens näherungsweise maximaler Bestandesgrundfläche erhoben wurden. Eine Möglichkeit zur Gewinnung derartiger Daten besteht darin, sie aus einem größeren Datensatz auszuwählen. Als Auswahlkriterium wird in der vorliegenden Arbeit die Steigung in der von Reineke (1933) aufgestellten Gleichung verwendet, die mit artspezifischen Steigungen von Beständen verglichen wird, in denen Selbstdurchforstung stattfindet.

Es werden insgesamt 6 Modelle an je einen Buchen- (*Fagus sylvatica* L.) und einen Fichtendatensatz (*Picea abies* (L.) H.Karst) aus Nordwestdeutschland angepasst. Es kann gezeigt werden, dass eine Einschränkung der Modellflexibilität einerseits zu plausibleren Ergebnissen, andererseits aber zu einer verringerten Aussagekraft in Gestalt höherer AIC-Werte führt. Mögliche Schwächen der zugrundeliegenden Datensätze und ihre Auswirkung auf die Anwendbarkeit der Modelle werden diskutiert. Die erlangten Modelle werden mit bereits vorhandenden Ansätzen zur Modellierung maximaler Bestandesgrundflächen verglichen.

2 Introduction

The term "maximum basal area" refers to the site-specific highest possible basal area of living trees in a forst stand (Assmann 1970). It plays a key role in forest management and planning, e.g., as a reference for determining thinning intensity (Assmann 1961; Döbbeler & Spellmann 2002) or stand density (Spellmann et al. 1999) or as an estimator of yield level (Franz 1967). At the same time, permanent sample plots on which maximum basal area is recorded are few and far between and cover only a narrow selection of stand properties. Thus, the need for extrapolating maximum basal area for stands for which comparable experimental plots do not exist arises.

Basal area modeling in general has been approached from various directions, both via remote sensing imagery (e.g., Silva et al. (2017)) as well as using in situ measurements (e.g., Yue et al. (2012)). Models may either be sensitive to stand productivity (e.g., Castedo-Dorado et al. (2007)) or insensitive (e.g., Monserud & Sterba (1996)). However, only few models are available for predicting *maximum* basal area, with the method proposed by Sterba (1975) being a widely used example. Recently, Wördehoff (2016) and Wördehoff et al. (2014) extended the range of available methods by utilizing generalized additive models for location, scale, and shape (GAMLSSs, Rigby & Stasinopoulos (2001)), which are an extension of generalized additive models (GAMS, Hastie & Tibshirani (1991)). However, both methods only use top height as a predictor variable and do not take stand productivity into account. "Stand productivity" here means the total crop yield achieved for a given top height (Assmann 1970). The present study aims to fill this gap and provide an age- and productivity-sensitive method for predicting maximum basal area. Extending the method proposed by Wördehoff (2016), it uses GAMs, shape constrained additive models (SCAMs, which are an extension of GAMS (Pya 2010)), and GAMLSSs with varying constraints. Using smooth functions and splines to construct their function basis, these models provide greater flexibility than Generalized Linear Models (GLMs, Nelder & Wedderburn (1972)), albeit at the price of a less intuitive interpretation. In order to achieve age- and productivity-sensitivity of the models, suited stand properties need to be used as predictor variables. Past experience has shown that usage of top height as a predictor variable prohibits the inclusion of a distinct productivity-effect. Building on the work by Nagel (1999), this study attempts to split up the information contained in top height into an age variable and a productivity variable, which may then be used as predictor variables in model fitting.

In order to increase reliability of model predictions to be *maximum* basal area predictions, it is necessary to ensure that the training data used for model fitting belong to stands which have attained maximum basal area. The most reliable way to achieve this is through continued measurements on permanent sample plots which ideally undergo no thinning apart from the removal of dead trees. Unfortunately, such sample plots are not as numerous as might be required to obtain broadly applicable models. As an alternative, observations of stands which are approximately at maximum basal area may be selected from readily available data sets. Basal area itself is not a valid selection criterion in this regard and therefore has to be replaced by a closely related variable. The mechanism presented here uses the slope of the equation proposed by

Reineke (1933) as a criterion, by comparing observed values to a range of species-specific values of stands undergoing self-thinning.

6 models are fitted, 2 GAMs, one SCAM, and 3 GAMLSSs. The models differ in the degree of constraints imposed on them. Each model was fit to a European beech (*Fagus sylvatica* L.) (beech) and a Norway spruce (*Picea abies* (L.) H.Karst) (spruce) data set. The data sets consist of even-aged pure stands of the respective species. The data sets used were kindly provided by the Northwest German Forest Research Institute.

3 Material

3.1 Data selection

The present study aims at predicting maximum basal area of even-aged pure beech and spruce stands by fitting statistical models to real-world data sets. This approach necessitates that the data sets used for model fitting only contain observations meeting all of the following requirements:

1. The observations belong to an even-aged stand.
2. The observations belong to a pure stand of the respective species.
3. The observations belong to a stand with maximum basal area.

The provided data sets were assumed to meet the first requirement. The second and third requirement were, however, not met by all observations contained in the data sets. Therefore, a selection of observations had to be made. With respect to the second requirement, selection was based on crown projection area: an observation was considered a pure stand observation and consequently selected if either beech or spruce held 70 % or more of the stand's total crown projection area. The approach for selecting observations which meet the third requirement (termed "data selection mechanism") sets species-specific thresholds for the slope of the equation

$$\log(N) = s \cdot \log(D) + k, \tag{1}$$

where N is stand density per unit area, s is the slope, D is average diameter by basal area, and k is a species-specific constant (Reineke 1933). It examines observations belonging to the same sample plot in chronological order, while attempting to answer 3 questions:

1. Is the current observation preceded and/or followed by another observation?
2. If the current observation is preceded by another observation, is the slope of Equation (1) between the current and the previous observation greater than or equal to the species-specific lower threshold?

3. If the current observation is followed by another observation, is the slope of Equation (1) between the current and the following observation lower than or equal to the species-specific upper threshold?

An observation is considered to be a maximum basal area observation and consequently selected if and only if all questions are answered positively. While being based on literature values (cp. Table 10), the slope thresholds are essentially arbitrary and a compromise of 2 conflicting goals: do not select observations of stands which clearly are not subject to self-thinning; select enough observations to ensure reliable model results. The slope thresholds are reported in Table 1 and form a range of allowed slopes. For beech it lies between -2.91 and -0.9, whereas for spruce it lies between -2.82 and -0.65. Figure 1 may aid in understanding how the data selection mechanism works. It depicts the observed relationship between stand density and average diameter on a double-logarithmic scale before and after application of the data selection mechanism for each species. In all plots of the figure, the dashed and solid black lines exemplify the species-specific upper and lower slope thresholds, respectively. Thus, an observation from a plot in column A (observations before application of the data selection mechanims) is present in the corresponding plot of column B (observations after application of the data selection mechanism) if

- it is connected to a previous and/or following observation by a colored line
  and
- the slope of the line connecting it to the previous observation (if present) is higher than the slope of the solid black line
  and
- the slope of the line connecting it to the following observation (if present) is lower than the slope of the dashed black line.

The data sets of selected observations will be further described in the following sections.

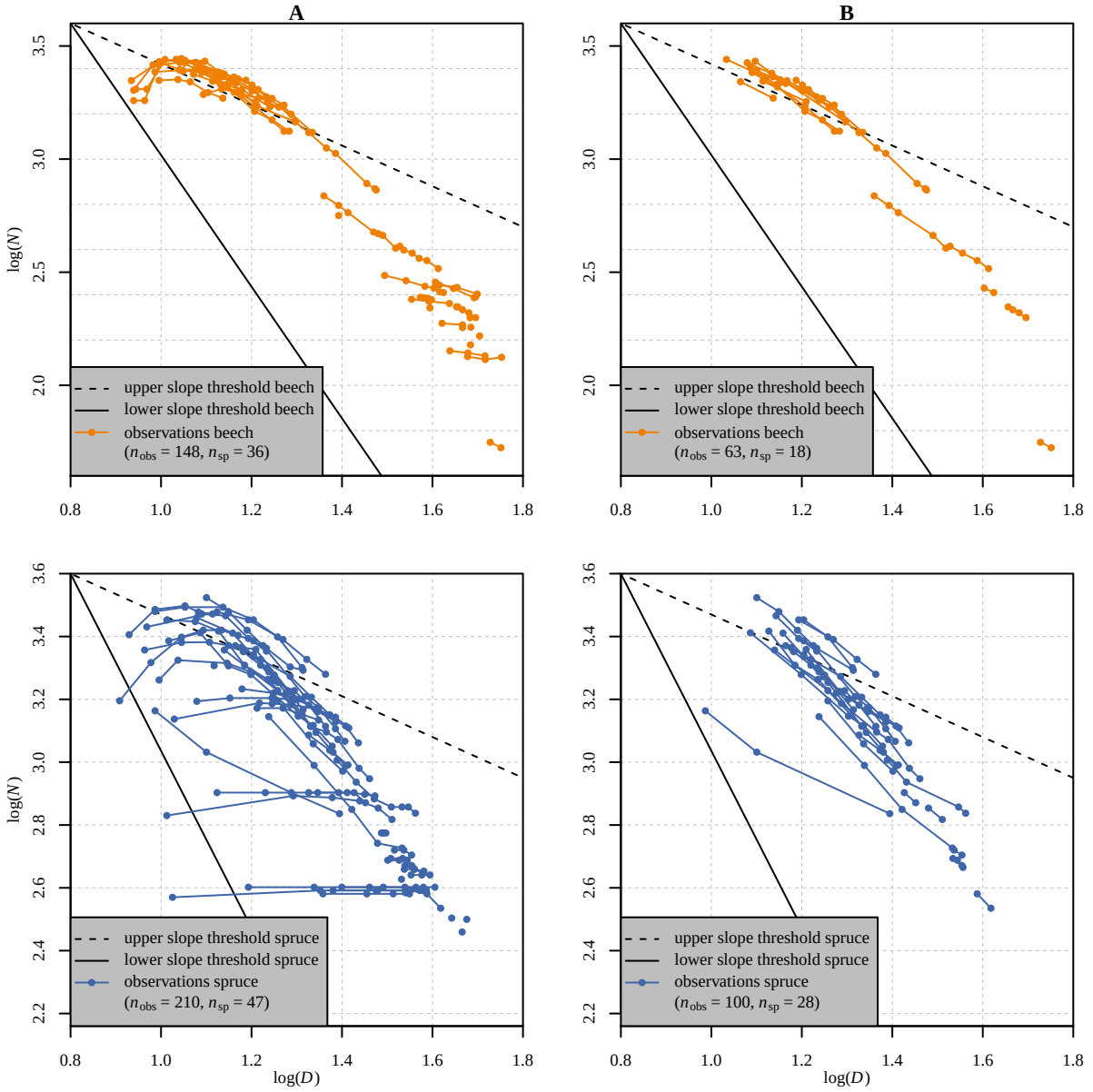

Figure 1: Observed relationship between stand density N and average diameter D in pure stands of beech (top row) and spruce (bottom row) on double-logarithmic scale before application of the data selection mechanism (column A) and after application of the data selection mechanism (column B). Each colored dot represents one observation. Colored lines connect observations belonging to the same sample plot. Dashed black lines exemplify the species-specific upper slope threshold used in the data selection mechanism. Solid black lines exemplify the species-specific lower slope threshold used in the data selection mechanism. See Table 1 for the threshold values.
n_{obs}: number of observations
n_{sp}: number of sample plots

Table 1: Species-specific lower and upper threshold for the slope s of Equation (1) used in the data selection mechanism.

	Beech	Spruce
Lower threshold	−2.91	−2.82
Upper threshold	−0.9	−0.65

3.2 Data sets

The data set size for beech is reported in Table 2. The data set comprises 18 sample plots, with a total of 63 observations and a mean of 3.5 observations per sample plot. The data set size for spruce is reported in Table 3. The data set comprises 28 sample plots, with a total of 100 observations and a mean of 3.6 observations per sample plot. In both data sets, the number of observations per sample plot ranges from 2 to 8.

Table 2: Number of observations per sample plot, total number of sample plots, and total number of observations in the beech data set.

	Sample plot ID	Number of observations
	00521004	4
	04221005	5
	08021003	2
	58321003	8
	8942102A	4
	8942102B	2
	89521002	5
	89621002	4
	89721006	2
	90421001	2
	99321000	4
	A1321300	2
	A8121011	2
	H1021001	2
	J5121001	4
	J5121005	5
	J5121007	4
	Z72NAT01	2
Total	18	63

Table 3: Number of observations per sample plot, total number of sample plots, and total number of observations in the spruce data set.

	Sample plot ID	Number of observations
	05451102	5
	06451102	5
	07151102	8
	07551103	7
	07551105	3
	4665111A	2
	4665112B	2
	4665113B	2
	4665114B	4
	4675112A	2
	4675113A	3
	4675113B	3
	47451104	5
	55751102	3
	87021515	2
	87021517	2
	87021522	2
	J6351121	2
	J6351141	3
	S1051103	3
	S1751101	3
	S1851101	4
	S1951101	3
	S2051102	3
	S2151101	2
	S2251101	5
	S2451102	4
	S2651104	8
Total	28	100

The geographical location and the altitude above sea level of sample plots are reported in Figures 2 and 3, respectively. The dots in the plots of both figures do not add up to the total number of sample plots of the respective species because some sample plots were part of the same trial

and therefore shared the same geographical location and altitude. The stands in both data sets are from northwest Germany. The beech sample plots are situated in Saarland, Rhineland-Palatinate, Hesse, North Rhine-Westphalia, and Lower Saxony. The spruce sample plots are situated in Rhineland-Palatinate, Hesse, North Rhine-Westphalia, Lower Saxony, and Schleswig-Holstein. In the beech data set, altitude above sea level spans 525 m, ranging from 40 m to 565 m, with a mean of 313.2 m. In the case of spruce it spans 730 m, ranging from 20 m to 750 m, with a mean of 410.5 m.

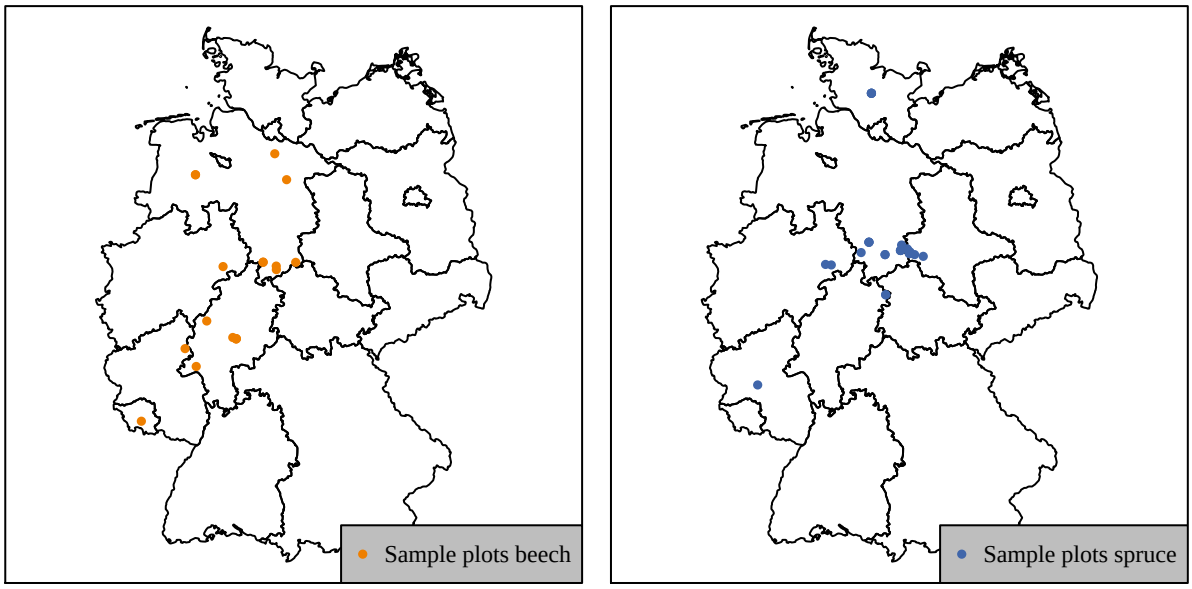

Figure 2: Geographical location of the sample plots of beech (left) and spruce (right) in Germany.

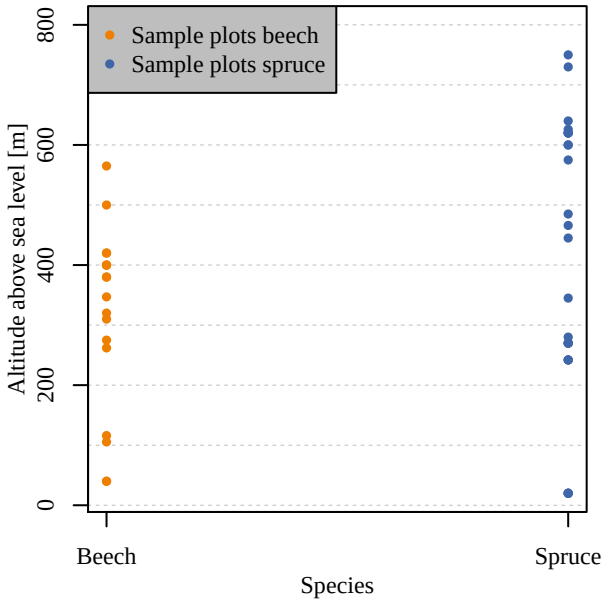

Figure 3: Altitude above sea level of sample plots.

In the following figures, dot color signifies the (fractional) yield class of the respective observation. Classification of observations into yield classes required multiple steps, which were undertaken separately for each species. First, for all observations the absolute productivity index

of stand (PI) was computed using the equation

$$PI = \frac{h_{100}(x) - \beta_0 - \beta_1 \cdot \ln(x) - \beta_2 \cdot \ln(x)^2}{\beta_3 + \beta_4 \cdot \ln(x)}, \tag{2}$$

where PI is approximately equal to top height (h_{100}) in meter at age 100 a, x is stand age in years, and $h_{100}(x)$ is top height in meter at age x and $\beta_0, \ldots, \beta_4$ are species-specific coefficients (Nagel 1999). The coefficients were taken from Nagel (1999) and are reported in Table 4. Next, a sequence of PI values was generated, ranging from yield class 4 to yield class −2, using an increment of 0.1 m, thus also covering fractional yield classes. Values of whole yield classes are reported in Table 5. Values for yield classes 4 to 1 were taken from Schober (1995) (moderate thinning), while values for yield classes 0 to −2 were linearly extrapolated from those for yield classes 2 and 1. The sequence was then restricted to the range between the best yield class needed to include the lowest observed PI value and the worst yield class needed to include the highest observed PI value. For beech, the restricted sequence ranged from yield class 3 to −2, whereas for spruce it ranged from yield class 4 to −2. A color palette matching this restricted sequence was then generated, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Each observed PI value was then mapped to the corresponding color. Figure 5 may aid in understanding how the mapping of colors was done.

Table 4: Species-specific coefficients of Equation (2) as reported in Nagel (1999).

Species	β_0	β_1	β_2	β_3	β_4
Beech	−75.659	23.192	−1.468	0	0.2152
Spruce	−49.872	7.3309	0.77338	0.52684	0.10542

Table 5: Absolute productivity index of stand (PI) of different yield classes (YC) for beech and spruce. Values for yield classes 4 to 1 were taken from (Schober 1995) (moderate thinning). Values for yield classes 0 to −2 were linearly interpolated from those of yield classes 2 and 1.

	PI [m]						
Species	YC 4	YC 3	YC 2	YC 1	YC 0	YC -1	YC -2
Beech	20.7	24.7	28.6	32.4	36.2	40	43.8
Spruce	23.5	27.2	31.2	35.1	39	42.9	46.8

Figure 4 shows the observed relationship between stand age and top height. As can be seen, several differences between the beech data set and the spruce data set exist. In the former, stands are notably older and cover a wider range of ages, with stand age ranging from 35 a to 153.6 a, spanning 118.6 a with a mean of all observations of 76.3 a and a mean of sample plot means of

77.6 a. In the latter, age ranges from 15 a to 113 a, spanning 98 a with a mean of all observations of 53.8 a and a mean of sample plot means of 53.9 a. Consequently, stands in the beech data set have a higher top height, with h_{100} ranging from 16.3 m to 39.2 m (difference: 22.9 m), but spruce stands cover a slightly wider range of top heights, with h_{100} ranging from 9.1 m to 33.3 m (difference: 24.2 m). In the case of beech, the maximal top height of all observations was reached by a sample plot of intermediate yield class (≈ 1) compared to the other sample plots in the data set. In the case of spruce, the maximal top height of all observations was reached by a sample plot of bad yield class (≈ 2) compared to the other sample plots in the data set. Apart from one observation of beech, top height increases with stand age.

Figure 5 shows the observed development of absolute productivity index of stand over stand age. For all sample plots in both data sets, the absolute productivity index of stand changed at least once during stand development. However, for several sample plots the direction of this change itself differs during stand development and there does not seem to be a general direction to which changes in absolute productivity index of stand adhere. The beech data set covers a narrower range of productivity indices and consequently yield classes than the spruce one, with *PI* ranging from 23.7 m to 38.3 m (difference: 14.6 m) and yield class ranging from 4 to -1 in the former, whereas for spruce *PI* ranges from 24.1 m to 45.2 m (difference: 21.1 m) and yield class from 4 to -2.

Figure 6 depicts the observed development of basal area over stand age for beech (top) and spruce (bottom). The beech data set has a higher minimal and a lower maximal basal area and covers a narrower range of basal areas compared to the spruce one, with basal area ranging from 12.6 $m^2\,ha^{-1}$ to 51.3 $m^2\,ha^{-1}$ (difference: 38.7 $m^2\,ha^{-1}$) in the former, and from 10.8 $m^2\,ha^{-1}$ to 79.8 $m^2\,ha^{-1}$ (difference: 69 $m^2\,ha^{-1}$) in the latter.

Figure 7 depicts the observed development of basal area over top height for beech (top) and spruce (bottom). In all sample plots of both species, basal area increases as top height increases.

One chief goal of this study was to ensure that the fitted models were capable of separating the effects of stand age and of absolute productivity index of stand on predicted basal area. However, this requires that corresponding predictor variables are available for model training. Top height (h_{100}) was considered an unsuited predictor variable for this, since past experience had shown that using this dimension as the predictor variable does not allow identification of a separate productivity index-effect in the model, at least when data set size is rather limited as is the case in the present study. Therefore, 2 new variables were calculated: a stand age variable and a productivity index variable, each of which was calculated in such a way as to exclude the effect of the other. The stand age variable ($h_{100}(x)_{\mathrm{YC}\,1}$) was calculated using the equation

$$h_{100}(x)_{\mathrm{YC}\,1} = \beta_0 + \beta_1 \cdot \ln(x) + \beta_2 \cdot \ln(x)^2 + PI_{\mathrm{YC}\,1} \cdot \big(\beta_3 + \beta_4 \cdot \ln(x)\big), \tag{3}$$

where $h_{100}(x)_{\mathrm{YC}\,1}$ is the top height in meter at age x if the stand were yield class 1, $PI_{\mathrm{YC}\,1}$ is the species-specific top height (h_{100}) at age 100 a of a stand of yield class 1 and all other terms have

the same meaning as in Equation (2), namely: x is stand age and $\beta_0, \ldots, \beta_4$ are species-specific coefficients which are reported in Table 4 (Nagel 1999). The value for $PI_{\mathrm{YC}\,1}$ was taken from Schober (1995) (moderate thinning) and was 32.4 m for beech and 35.1 m for spruce. By setting $PI_{\mathrm{YC}\,1}$ to a species-specific constant, it was possible to exclude any productivity index-effects from the stand age variable. The independence of the stand age variable from productivity index-effects is made apparent by Figure 8, which depicts the observed relationship between stand age and stand age variable: for each species, all observations follow the same curve, regardless of yield class. Notable differences with respect to the stand age variable exist between the data sets of both species: for beech it ranges from 13 m to 39 m, spanning 26 m with a mean of 25.6 m, while for spruce it ranges from 4.2 m to 38.1 m, spanning 33.9 m with a mean of 23.8 m. These range differences are not in correspondence with the differences in stand age between species: compared to spruce, the beech data set covers a wider range of stand age but a narrower range of stand age variable; the opposite is true for the spruce data set compared to beech.

The productivity index variable was calculated in 2 steps. First, the absolute productivity index of stand was calculated using Equation (2). Subsequently, the absolute productivity index of stand of yield class 1 was subtracted from this value:

$$PI_{\mathrm{diff}} = PI - PI_{\mathrm{YC}\,1}, \tag{4}$$

where PI_{diff} is the productivity index variable, PI has the same meaning as in Equation (2), and $PI_{\mathrm{YC}\,1}$ has the same meaning as in Equation (3), namely: PI is top height (h_{100}) in meter at age 100 a and $PI_{\mathrm{YC}\,1}$ is the species-specific top height (h_{100}) in meter at age 100 a of a stand of yield class 1 (32.4 m for beech and 35.1 m for spruce according to Schober (1995) (moderate thinning)). PI_{diff} is a measure of the performance of a stand relative to a reference stand (here: a stand of yield class 1 according to Schober (1995)): high values signify stands which outperform the reference stand, whereas low values indicate low performance stands. Since both PI and $PI_{\mathrm{YC}\,1}$ refer to a specific stand age (here: 100 a) rather than a variable one, PI_{diff} should only be influenced by stand productivity and largely be free of any stand age-effects. Figure 9 depicts the observed relationship between stand age and productivity index variable. Despite the attempt to exclude stand age-effects from it, the productivity index variable does exhibit an evident negative correlation with stand age, in that older stands tend to have lower PI_{diff} values. The correlation coefficient between stand age and productivity index variable was -0.665 for beech and -0.737 for spruce. However, no individual stand shows a clear worsening of productivity index variable over time. Therefore, it seems plausible that the correlation is due to the fact that older stands (which were established several decades ago) were not subject to the rather recent phenomenon of nitrogen fertilization through atmospheric deposition (Kenk & Fischer 1988), rather than being due to a true dependence of productivity index variable on stand age. The differences in observed yield classes between both species are also encountered in the productivity index variable: for beech it ranges from -6.7 m to 8.6 m, spanning 15.3 m with mean of 1.1 m, while for spruce it ranges from -11 m to 10.1 m, spanning 21.1 m with a mean of -0.2 m.

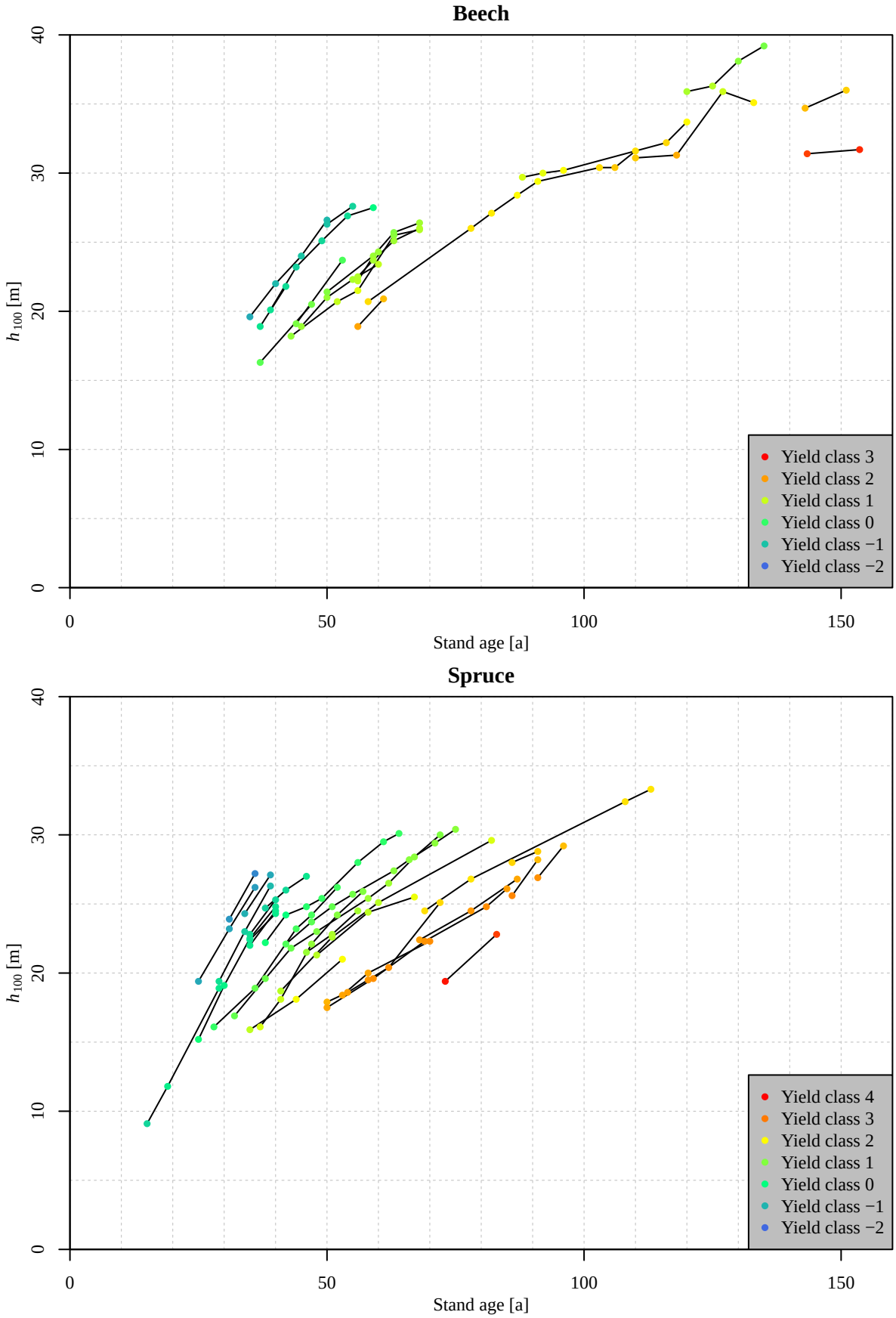

Figure 4: Observed relationship between stand age and top height (h_{100}) for beech (top) and spruce (bottom). Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Dot color signifies the (fractional) yield class of the respective observation, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

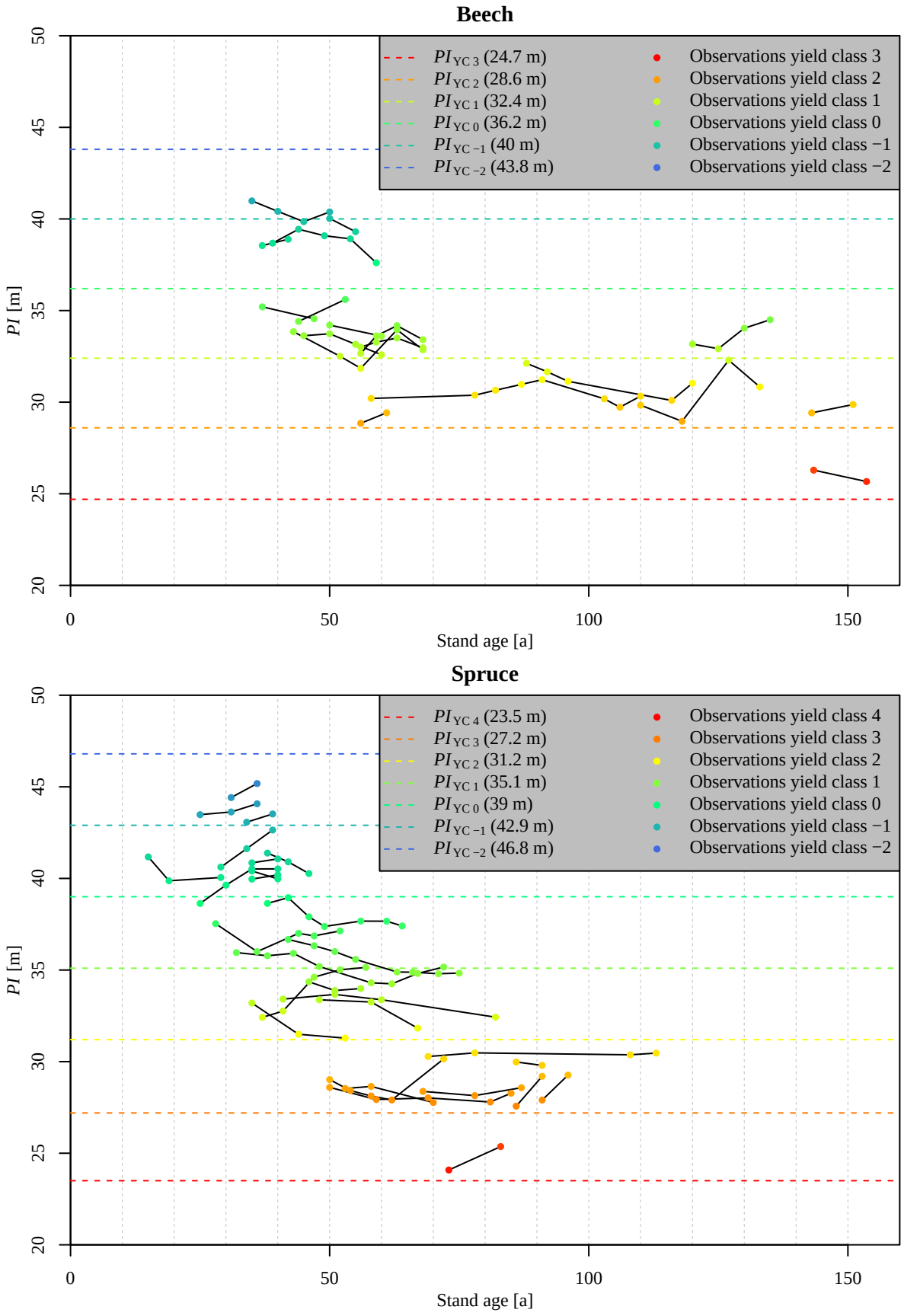

Figure 5: Observed relationship between stand age and absolute productivity index of stand (*PI*) for beech (top) and spruce (bottom). Dashed lines mark the *PI* of yield classes according to Table 5. Dots, black lines, and dot color have the same meaning as in Figure 4, namely: Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Dot color signifies the (fractional) yield class of the respective observation, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

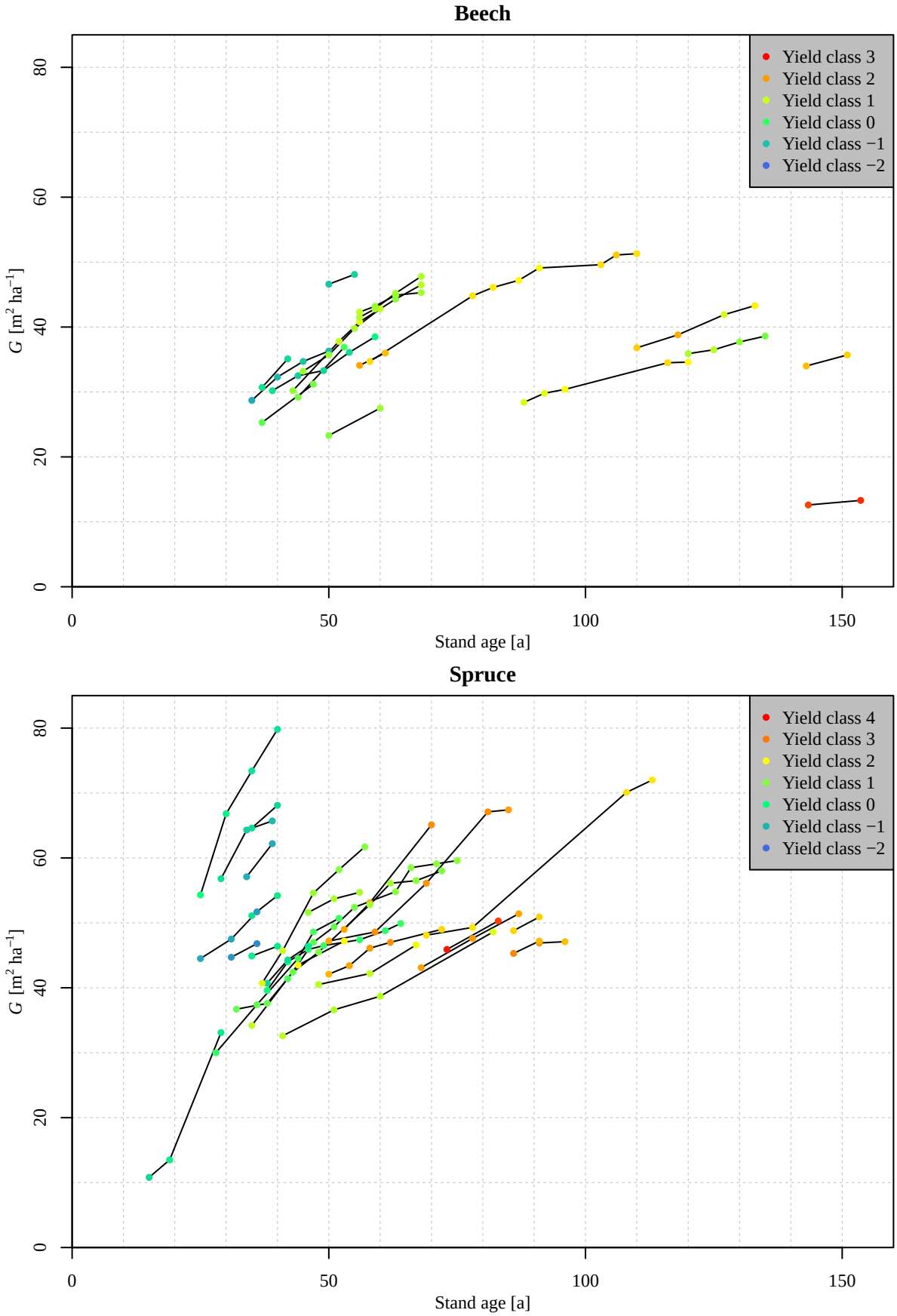

Figure 6: Observed relationship between stand age and basal area (G) for beech (top) and spruce (bottom). Dots, black lines, and dot color have the same meaning as in Figure 5, namely: Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Dot color signifies the (fractional) yield class of the respective observation, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

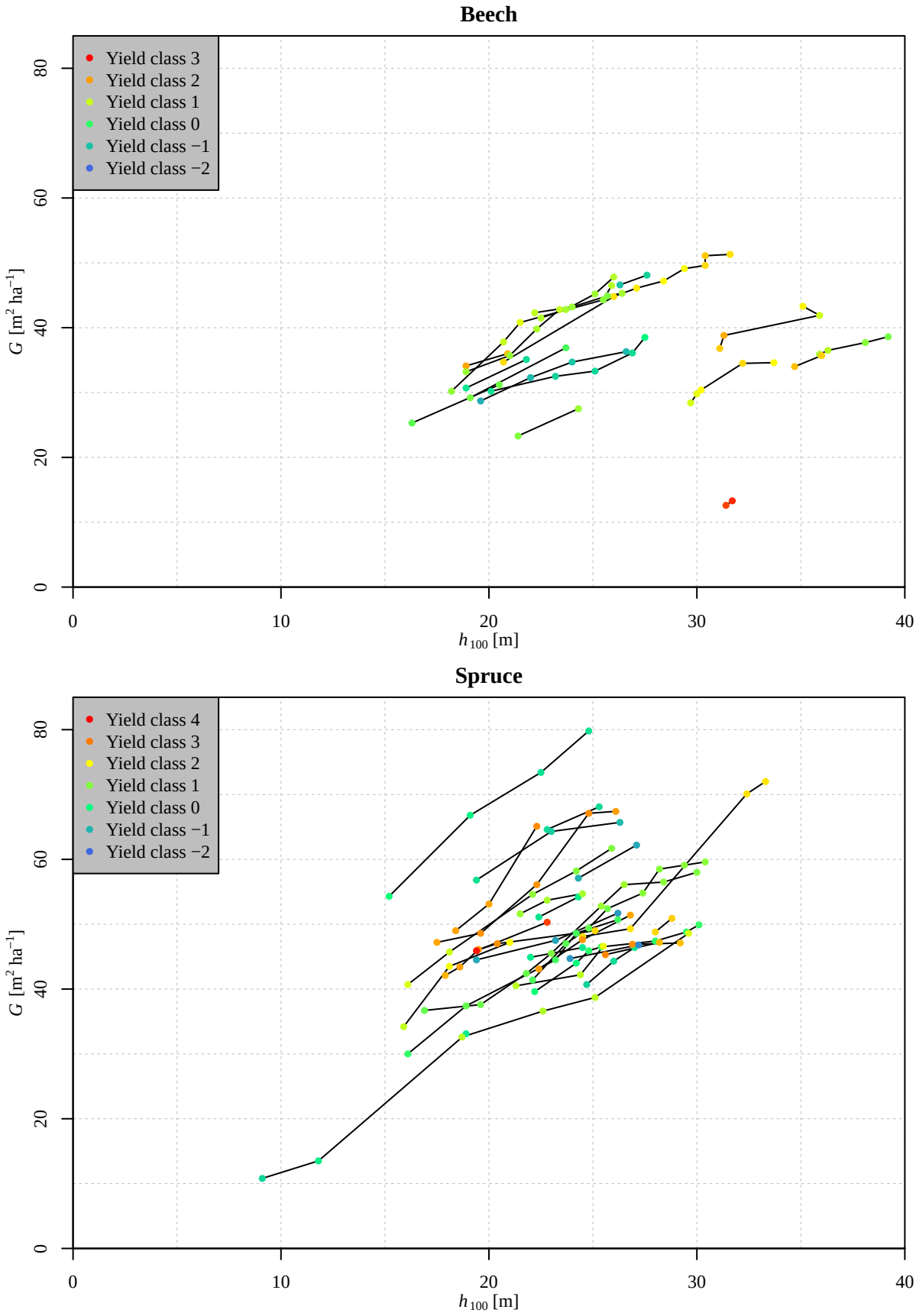

Figure 7: Observed relationship between top height (h_{100}) and basal area (G) for beech (top) and spruce (bottom). Dots, black lines, and dot color have the same meaning as in Figure 6, namely: Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Dot color signifies the (fractional) yield class of the respective observation, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

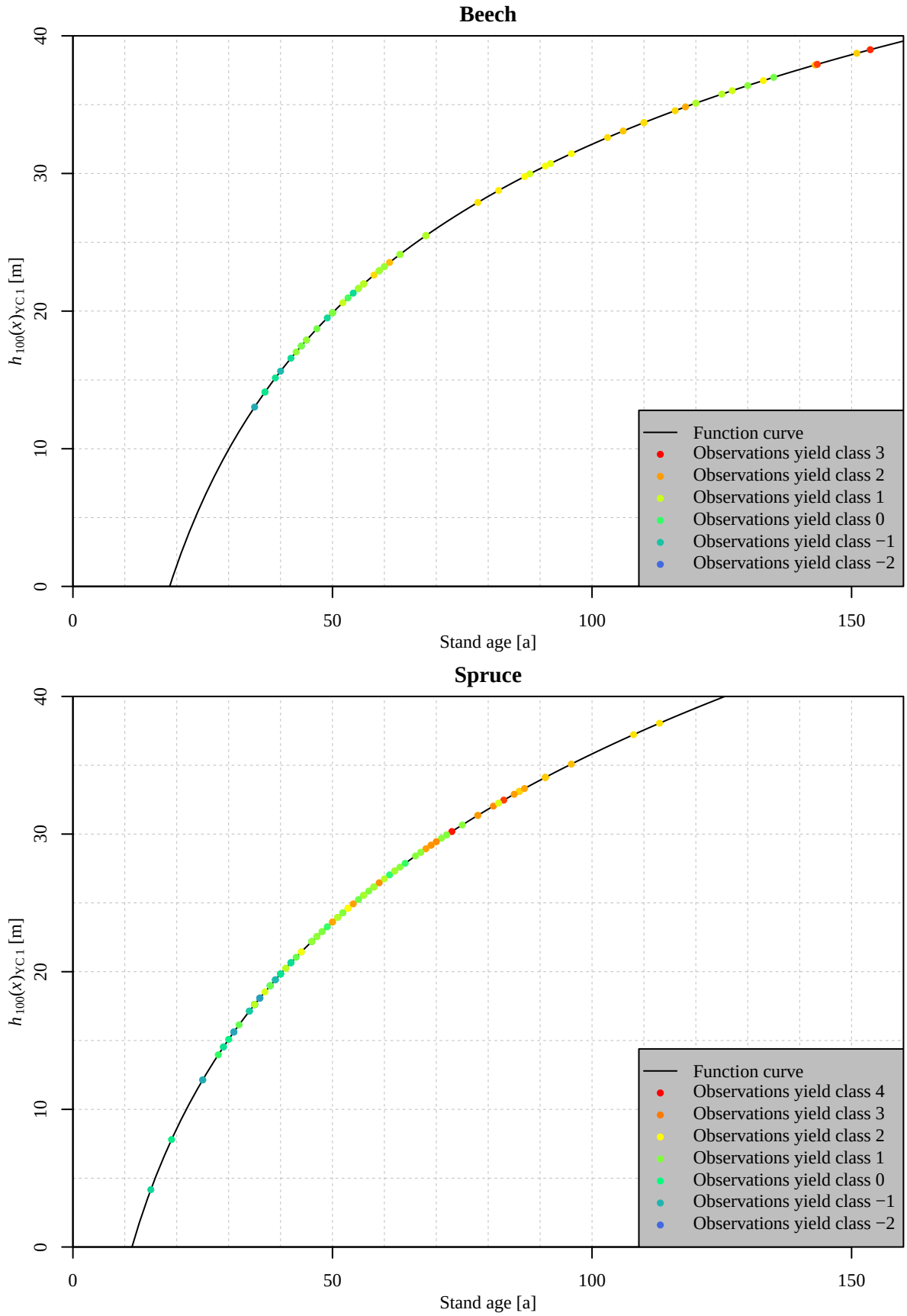

Figure 8: Observed relationship between stand age and stand age variable ($h_{100}(x)_{YC\,1}$) for beech (top) and spruce (bottom). Black lines depict the function given by Equation (3), which was used for calculating the stand age variable. Dots and dot color have the same meaning as in Figure 7, namely: Each dot represents one observation. Dot color signifies the (fractional) yield class of the respective observation, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

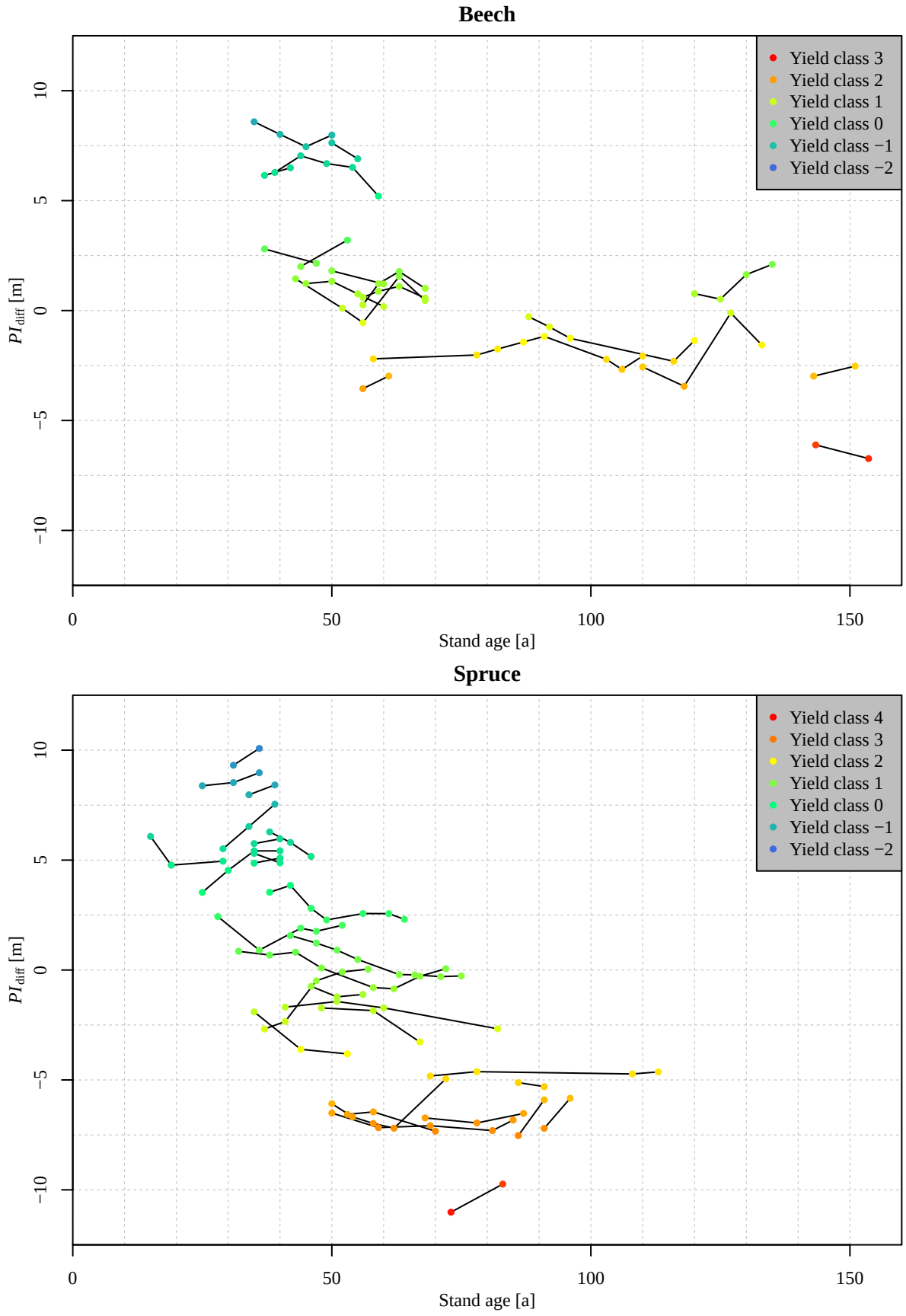

Figure 9: Observed relationship between stand age and productivity index variable (PI_{diff}) for beech (top) and spruce (bottom). Dots, black lines, and dot color have the same meaning as in Figure 7, namely: Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Dot color signifies the (fractional) yield class of the respective observation, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

4 Methods

Each model type used in this study (GAM, SCAM, GAMLSS) will be detailed separately in the following sections. Please note that, unless stated otherwise, mathematical symbols used for model explanation do not retain their meaning across model types. Subsequently, the settings used for calculating the models will be described.

4.1 GAMs

GAMs (Hastie & Tibshirani 1991; Wood 2006) are an extension of generalized linear models (GLMs) (Nelder & Wedderburn 1972), which in turn build upon linear models (LMs).

A simple LM assumes a metric random variable Y to linearly depend upon the metric variable X, such that

$$Y = X\beta + \varepsilon, \tag{5}$$

where β is the coefficient of X to be estimated and ε_i is an independent random variable such that $\mathbb{E}(\varepsilon_i) = 0$ and $\mathbb{E}(\varepsilon_i^2) = \sigma^2$. To allow testing of hypotheses related to the model described by Equation (5), additional assumptions about the distribution of Y and ε need to be made. Specifically, the residual term ε is assumed to follow normal distribution with a mean of zero and a variance of σ^2: $\varepsilon \sim N(0, \sigma^2)$, which is equivalent to assuming the response variable Y to follow normal distribution with a mean equal to the product of independent variable X and parameter β and a variance of σ^2: $Y \sim N(X\beta, \sigma^2)$ (Burkschat et al. 2012; Wood 2006).

A GLM has the structure

$$g(\mu_i) = \mathbf{x}_i\boldsymbol{\beta}, \tag{6}$$

where μ_i is the expectation of response variable Y_i, i.e., $\mu_i \equiv \mathbb{E}(Y_i)$, g is a smooth monotonic "link function", $\mathbf{x}_i$ is the ith row of model matrix $\mathbf{X}$, and $\boldsymbol{\beta}$ is a vector of unknown parameters (Nelder & Wedderburn 1972; Wood 2006). GLMs are an extension of LMs insofar as they also center around a "linear predictor", $\mathbf{X}\boldsymbol{\beta}$, but allow other link functions than the identity function and allow the distribution of the dependent variable to be any distribution from the exponential family, instead of only the normal distribution. The exponential family consists of distributions whose probability density function can be written as

$$f_\theta(y) = \exp\left(\frac{y\theta - b(\theta)}{a(\boldsymbol{\Phi})} + c(y, \boldsymbol{\Phi})\right), \tag{7}$$

where a, b, and c are arbitrary functions, $\boldsymbol{\Phi}$ is an arbitrary "scale" parameter, and θ is the so-called "canonical parameter" of the distribution (Wood 2006).

Building upon GLMs, a GAM has a structure similar to

$$g(\mu_i) = \mathbf{x}_i^* \boldsymbol{\theta} + f_1(x_{1,i}) + f_2(x_{2,i}) + f_3(x_{3,i}, x_{4,i}) + \dots, \tag{8}$$

where μ_i again is the expectation of response variable Y_i, i.e., $\mu_i \equiv \mathbb{E}(Y_i)$, the response variable Y_i follows any exponential family distribution, $\mathbf{x}_i^*$ is a row of the model matrix for any strictly parametric model components, $\boldsymbol{\theta}$ is the corresponding parameter vector, and the f_j are smooth functions of the covariates, x_k (Wood 2006). A smooth function may be considered an estimate of the actual functional relationship between the response variable and the predictor variable to which the smooth function is applied (Hastie & Tibshirani 1991). Unlike LMs or GLMs, a smooth function is nonparametric, i.e., it does not assume the response variable to follow any specific distribution. As an explanatory example, consider a model containing one smooth function of one predictor variable with the identity function as the link function,

$$y_i = f(x_i) + \varepsilon_i, \tag{9}$$

where y_i is a response variable, x_i is a predictor variable, f is a smooth function, and the ε_i follow the normal distribution with a mean of zero and a variance of σ^2 (Wood 2006). The x_i lie in the interval [0, 1]. In order to be able to estimate f, a space of known functions (called the "function basis"), of which f (or its estimate) is assumed to be an element, needs to be chosen, such that

$$f(x) = \sum_{i=1}^{q} b_i(x)\beta_i, \tag{10}$$

where q is the number of observations, $b_i(x)$ is the i^{th} element of the chosen function space, and β_i is an unknown parameter (Wood 2006). Substituting Equation (10) into Equation (9) yields a linear model which can then be fitted using established methods for LMs or GLMs.

4.1.1 Thin plate regression splines

The GAMs presented in this study employ thin plate regression splines (Wood 2003) for creating the basis of all their smooth functions. Thin plate regression splines are based on thin plate splines (Duchon 1977). In order to explain thin plate splines, consider the problem of estimating the smooth function $g(x)$, based on n observations $(y_i, \mathbf{x}_i)$ such that

$$y_i = g(\mathbf{x}_i) + \varepsilon_i, \tag{11}$$

where ε_i is a random error term and where $\mathbf{x}$ is a d-vector ($d \leq n$) (Wood 2006). Function g is then estimated by finding the function $\hat{f}$ which minimizes

$$\|\mathbf{y} - \mathbf{f}\|^2 + \lambda J_{m,d}(f), \tag{12}$$

where $\mathbf{y}$ is the vector of y_i data, $\mathbf{f} = \big(f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)\big)^{\mathrm{T}}$, λ is a smoothing parameter controlling the tradeoff between data fitting and smoothness of f, and $J_{m,d}(f)$ is a penalty

functional measuring the "roughness" of f (Wood 2006). The penalty is defined as

$$J_{m,d} = \int \cdots \int_{\Re^d} \sum_{\nu_1+\ldots+\nu_d=m} \frac{m!}{\nu_1! \ldots \nu_d!} \left(\frac{\partial^m f}{\partial x_1^{\nu_1} \ldots \partial x_d^{\nu_d}} \right)^2 \mathrm{d}x_1 \ldots \mathrm{d}x_d \tag{13}$$

(Wood 2006). If the restriction $2m > d$ is imposed, the function minimizing Equation (12) has the form

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^{n} \delta_i \eta_{m,d} \left(\left\| \mathbf{x} - \mathbf{x}_i \right\| \right) + \sum_{j=1}^{M} \alpha_j \boldsymbol{\Phi}_j(\mathbf{x}), \tag{14}$$

where $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$ are vectors of coefficients to be estimated, $\boldsymbol{\delta}$ being subject to the linear constraints that $\mathbf{T}^{\mathrm{T}}\boldsymbol{\delta} = 0$, where $T_{i,j} = \boldsymbol{\Phi}_j(\mathbf{x}_i)$ (Wood 2006). The $M = \binom{m+d-1}{d}$ functions, $\boldsymbol{\Phi}_i$, are linearly independent polynomials spanning the space of polynomials $\Re^d$ of degree less than m. The $\boldsymbol{\Phi}_i$ span the space of functions for which $J_{m,d}$ is zero. The $\eta_{m,d}$ functions are defined as

$$\eta_{m,d}(r) = \begin{cases} \frac{(-1)^{m+1+d/2}}{2^{2m-1}\pi^{d/2}(m-1)!(m-d/2)!} r^{2m-d} \log(r) & \text{if } d \text{ is even} \\ \frac{\Gamma(d/2-m)}{2^{2m}\pi^{d/2}(m-1)!} r^{2m-d} & \text{if } d \text{ is odd} \end{cases} \tag{15}$$

(Wood 2006). If matrix $\mathbf{E}$ is defined by $E_{i,j} \equiv \eta_{m,d}\left(\left\| \mathbf{x}_i - \mathbf{x}_j \right\| \right)$, then the thin plate spline fitting problem becomes

$$\text{minimize } \left\| \mathbf{y} - \mathbf{E}\boldsymbol{\delta} - \mathbf{T}\boldsymbol{\alpha} \right\|^2 + \lambda \boldsymbol{\delta}^{\mathrm{T}} \mathbf{E} \boldsymbol{\delta} \text{ subject to } \mathbf{T}^{\mathrm{T}}\boldsymbol{\delta} = 0, \tag{16}$$

with respect to $\boldsymbol{\delta}$ and $\boldsymbol{\alpha}$ (Wood 2006).

Thin plate regression splines build upon thin plate splines by truncating the space of the components with parameter $\boldsymbol{\delta}$, without changing the $\boldsymbol{\alpha}$ components. Let $\mathbf{E} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}$ be the eigendecomposition of $\mathbf{E}$, so that $\mathbf{D}$ is a diagonal matrix of eigenvalues of $\mathbf{E}$ arranged so that $\left| D_{i,i} \right| \geq \left| D_{i-1,i-1} \right|$ and the columns of $\mathbf{U}$ are the corresponding eigenvectors. Furthermore, let $\mathbf{U}_k$ denote the matrix consisting of the first k columns of $\mathbf{U}$ and let $\mathbf{D}_k$ denote the top right $k \times k$ submatrix of $\mathbf{D}$. By writing $\boldsymbol{\delta} = \mathbf{U}_k \boldsymbol{\delta}_k$, one can restrict $\boldsymbol{\delta}$ to the column space of $\mathbf{U}_k$. Thus, Equation (16) becomes

$$\text{minimize } \left\| \mathbf{y} - \mathbf{U}_k \mathbf{D}_k \boldsymbol{\delta}_k - \mathbf{T}\boldsymbol{\alpha} \right\|^2 + \lambda \boldsymbol{\delta}_k^{\mathrm{T}} \mathbf{D}_k \boldsymbol{\delta}_k \text{ subject to } \mathbf{T}^{\mathrm{T}} \mathbf{U}_k \boldsymbol{\delta}_k = 0 \tag{17}$$

with respect to $\boldsymbol{\delta}_k$ and $\boldsymbol{\alpha}$ (Wood 2006). In order to absorb the constraints, an orthogonal column basis, $\mathbf{Z}_k$, must be found, such that $\mathbf{T}^{\mathrm{T}}\mathbf{U}_k\mathbf{Z}_k = 0$. By writing $\boldsymbol{\delta}_k = \mathbf{Z}_k \tilde{\boldsymbol{\delta}}$, one can restrict $\boldsymbol{\delta}_k$ to this space. This yields the unconstrained problem that must be solved to fit the rank k approximation to the smoothing spline:

$$\text{minimize } \left\| \mathbf{y} - \mathbf{U}_k \mathbf{D}_k \mathbf{Z}_k \tilde{\boldsymbol{\delta}} - \mathbf{T}\boldsymbol{\alpha} \right\|^2 + \lambda \tilde{\boldsymbol{\delta}}^{\mathrm{T}} \mathbf{Z}_k^{\mathrm{T}} \mathbf{D}_k \mathbf{Z}_k \tilde{\boldsymbol{\delta}} \tag{18}$$

with respect to $\tilde{\boldsymbol{\delta}}$ and $\boldsymbol{\alpha}$ (Wood 2006).

4.1.2 Generalized cross validation

For the GAMs presented in this study, estimation of the smoothing parameter λ for each smooth function was achieved via generalized cross validation (GCV), which builds upon cross validation. In cross validation, each pair of observations $\left(x_i, y_i\right)$ is left out one at a time while estimating the smooth function at x_i using the remaining $n-1$ observations. This way, cross validation mimics the use of training and test data for model fitting. The quantity of interest is the cross validation sum of squares

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{f}_{\lambda}^{-i}\left(x_i\right)\right)^2, \tag{19}$$

where n is the total number of observations, y_i is the i^{th} observation of the response variable Y, x_i is the i^{th} observation of the response variable X, and $\hat{f}_{\lambda}^{-i}\left(x_i\right)$ is the smooth function fit at x_i, obtained with smoothing parameter λ and by leaving out $\left(x_i, y_i\right)$ (Hastie & Tibshirani 1991). The cross validation sum of squares is computed for several values of λ over a suitable range and the smoothing parameter value $\hat{\lambda}$ which minimizes $CV(\lambda)$ is selected. A formal definition of $\hat{f}_{\lambda}^{-i}\left(x_i\right)$ for linear smoothers is

$$\hat{f}_{\lambda}^{-i}\left(x_i\right) = \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{S_{i,j}(\lambda)}{1 - S_{i,i}(\lambda)} y_i, \tag{20}$$

where $S_{i,j}(\lambda)$ is the element of row i and column j of the smoother matrix $\mathbf{S}_{\lambda}$ for smoothing parameter λ. The smoother matrix $\mathbf{S}$ is the matrix which satisfies the equation

$$\hat{\mathbf{f}} = \mathbf{S}\mathbf{y}, \tag{21}$$

where $\hat{\mathbf{f}}$ is the vector of smooth function fits at observations $x_1, \ldots, x_n$ and $\mathbf{y}$ is the vector of observations of the response variable. Equation (20) implies that

$$\hat{f}_{\lambda}^{-i}\left(x_i\right) = \sum_{\substack{j=1 \\ j\neq i}}^{n} S_{i,j}(\lambda) y_i + S_{i,i}(\lambda)\hat{f}_{\lambda}^{-i}\left(x_i\right), \tag{22}$$

which in turn implies that

$$y_i - \hat{f}_{\lambda}^{-i}\left(x_i\right) = \frac{y_i - \hat{f}_{\lambda}\left(x_i\right)}{1 - S_{i,i}(\lambda)}. \tag{23}$$

Substituting Equation (23) into Equation (19) yields

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{f}_{\lambda}\left(x_i\right)}{1 - S_{i,i}(\lambda)}\right)^2. \tag{24}$$

GCV replaces $S_{i,i}(\lambda)$ in Equation (24) with its mean value $\mathrm{tr}\left(\mathbf{S}_\lambda\right)/n$, since it is easier to compute:

$$GCV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{f}_\lambda(x_i)}{1 - \mathrm{tr}\left(\mathbf{S}_\lambda\right)/n}\right)^2, \tag{25}$$

where tr(**A**) is the trace, i.e., the sum of the diagonal elements of square matrix **A** (Hastie & Tibshirani 1991).

4.2 Explanation of SCAMs

SCAMs (Pya 2010; Pya & Wood 2015) are an extenstion of GAMs. Instead of thin plate regression splines, they use penalized splines (P-splines) (Eilers & Marx 1996) to construct their smooth function basis, which in turn build upon basis splines (B-splines) (Boor 2001; Curry & Schoenberg 1947).

4.2.1 B-splines and P-splines

A B-spline is a pieceweise polyonmial. Adjacent polynomial pieces are joined at the so-called knots in a specific manner: both the ordinates as well as the first derivatives of the adjacent polynomials are equal at the knots (Eilers & Marx 1996). This ensures that the resulting B-spline is continuous in its value and first derivative. To define a B-spline, let the knot vector $\mathbf{T} = \left(t_0, t_1, \ldots, t_m\right)$ be a nondecreasing sequence with $t_i \in [0, 1]$ and define the control points $\mathbf{P}_0, \ldots, \mathbf{P}_n$. The degree is defined as $p \equiv m - n - 1$. Each control point is associated with a basis function, which is given by

$$N_{i,0}(x) = \begin{cases} 1 & \text{if } t_i \leq x < t_{i+1} \text{ and } t_i < t_{i+1} \\ 0 & \text{otherwise} \end{cases}, \tag{26}$$

and

$$N_{i,j}(x) = \frac{x - t_i}{t_{i+j} - t_i} N_{i,j-1}(x) + \frac{t_{i+j+1} - x}{t_{i+j+1} - t_{i+1}} N_{i+1,j-1}(x) \quad \text{for } j = 1, 2, \ldots, p. \tag{27}$$

Then the curve defined by

$$\mathbf{C}(x) = \sum_{i=0}^{n} \mathbf{P}_i N_{i,p}(x) \tag{28}$$

is a B-spline (Weisstein 2017a). This can then be used for regression by estimating the vector of control points. Commonly, this is done via the method of ordinary least squares (Bollaerts et al. 2006) by estimating the control points which, for q observations (x_j, y_j), minimize the equation

$$S = \sum_{j=1}^{q}\left(y_j - \sum_{i=0}^{n} \mathbf{P}_i N_{i,p}\left(x_j\right)\right)^2. \tag{29}$$

The concept of P-splines extends B-splines in 2 ways: on the one hand it sets a large number of equidistant knots, while on the other hand adding a penalty to Equation (29) in order to prevent overfitting. This penalty is based on high-order ($k \geq 2$) finite differences between the coefficients of adjacent B-splines. The least squares objective function to minimize thus becomes

$$S = \sum_{j=1}^{q} \left(y_j - \sum_{i=0}^{n} \mathbf{P}_i N_{i,p}(x_j) \right)^2 + \lambda \sum_{i=k+1}^{n} \left(\Delta^k \mathbf{P}_i \right)^2, \tag{30}$$

with $\Delta^k \mathbf{P}_i$ being the kth-order differences, i.e., $\Delta^k \mathbf{P}_i = \Delta^1 \left(\Delta^{k-1} \mathbf{P}_i \right)$ with $\Delta^1 \mathbf{P}_i = \mathbf{P}_i - \mathbf{P}_{i-1}$ and with λ being a smoothness parameter (Bollaerts et al. 2006; Eilers & Marx 1996).

4.2.2 SCAMs

A SCAM may have a structure like

$$g(\mu_i) = \mathbf{A}\boldsymbol{\theta} + \sum_j f_j(z_{j,i}) + \sum_k m_k(x_{k,i}), \tag{31}$$

where g is a known smooth monotonic link function, μ_i is the mean of the univariate response variable which follows an exponential family distribution, $\mathbf{A}$ is the model matrix, $\boldsymbol{\theta}$ is a vector of unknown parameters, f_j is an unknown smooth function of predictor variable z_j and m_k is an unknown shape constrained smooth function of predictor variable x_k (Pya & Wood 2015).

In order to explain shape constrained smooth functions, consider the case of a monotonically increasing smooth function, m, using a B-spline basis. Let

$$m(x) = \sum_{j=1}^{q} \gamma_j B_j(x), \tag{32}$$

where q is the number of basis functions, the B_j are B-spline basis functions of at least second order for representing smooth functions over interval $[a, b]$, based on equally spaced knots, and the γ_j are spline coefficients (Pya & Wood 2015). A sufficient condition for ensuring the smooth function m to be monotonically increasing (i.e., for $m'(x) \geq 0$) over $[a, b]$ is that $\gamma_j \geq \gamma_{j-1} \forall j$. This condition can be imposed by reparameterizing, so that

$$\boldsymbol{\gamma} = \boldsymbol{\Sigma}\tilde{\boldsymbol{\beta}}, \tag{33}$$

where $\boldsymbol{\beta} = \left[\beta_1, \beta_2, \ldots, \beta_q\right]^{\mathrm{T}}$ and $\tilde{\boldsymbol{\beta}} = \left[\beta_1, \exp(\beta_2), \ldots, \exp(\beta_q)\right]^{\mathrm{T}}$, while

$$\Sigma_{i,j} = \begin{cases} 0 & \text{if } i < j \\ 1 & \text{if } i \geq j \end{cases} \tag{34}$$

(Pya & Wood 2015). Thus, if $\mathbf{m} = [m(x_1), m(x_2), \ldots, m(x_n)]^{\mathrm{T}}$ is the vector of m values at the

observed points x_i and $\mathbf{X}$ is a matrix such that $X_{i,j} = B_j(x_i)$, then

$$\mathbf{m} = \mathbf{X}\boldsymbol{\Sigma}\tilde{\boldsymbol{\beta}} \tag{35}$$

(Pya & Wood 2015). Smoothness of m is ensured by penalizing the squared difference between adjacent β_j, starting from β_2, using $\left\|\mathbf{D}\boldsymbol{\beta}\right\|^2$ as the penalty, where $\mathbf{D}$ is the $(q-2)\times q$ matrix whose elements are given by

$$D_{i,j} = \begin{cases} -D_{i,j+1} = 1 & \text{if } i = 1, \ldots, q-2,\ j = i+1 \\ 0 & \text{otherwise} \end{cases}. \tag{36}$$

This penalty becomes zero when all β_j following β_1 are equal, thus ensuring the γ_j to form a uniformly increasing sequence and $m(x)$ to form an increasing straight line (Pya & Wood 2015).

Embedding the shape constrained smooth function $m(x)$ in a larger model requires an additional constraint on $m(x)$ in order to avoid it being confused with the intercept of the larger model. This can be achieved by imposing centering constraints on the model matrix columns, i.e., by setting the sum of the values of the smooth to zero: $\sum_{i=1}^{n} m(x_i) = 0$ (Pya & Wood 2015).

The SCAMs presented in this study constrain the smooth function to be increasing and concave. For this constraint, matrix $\boldsymbol{\Sigma}$ has to be adjusted so that

$$\Sigma_{i,j} = \begin{cases} 0 & \text{if } i = 1,\ j \geq 2 \\ 1 & \text{if } i \geq 1,\ j = 1 \\ i-1 & \text{if } i \geq 2,\ j = 2, \ldots, q-1+2 \\ q-j+1 & \text{if } i \geq 2,\ j = q-i+3, \ldots, q, \end{cases} \tag{37}$$

while matrix $\mathbf{D}$ has to be adjusted so that

$$D_{i,j} = \begin{cases} -D_{i,j+1} = 1 & \text{if } i = 1, \ldots, q-3,\ j = i+2 \\ 0 & \text{otherwise} \end{cases} \tag{38}$$

(Pya & Wood 2015).

In order to represent Equation (31) for computation, the following paragraphs use basis expansion, penalties, and identifiability constraints for all f_j as described in Wood (2006). Thus,

$$\sum_j f_j\left(z_{j,i}\right) = \mathbf{F}_i\boldsymbol{\gamma}, \tag{39}$$

where $\mathbf{F}$ is a model matrix determined by the basis functions and the constraints and $\boldsymbol{\gamma}$ is a vector of coefficients to be estimated. The penalties on the f_j are quadratic in $\boldsymbol{\gamma}$. Each m_k is represented by a model matrix of the form $\mathbf{X}\boldsymbol{\Sigma}$ and a corresponding coefficient vector (Pya & Wood 2015).

The model matrices for all m_k are then combined so that

$$\sum_k m_k\left(x_{k,i}\right) = \mathbf{M}_i\tilde{\boldsymbol{\beta}}, \tag{40}$$

where $\mathbf{M}$ is a model matrix and $\tilde{\boldsymbol{\beta}}$ is a vector containing both model coefficients (β_i) and exponentiated model coefficients ($\exp(\beta_i)$). The penalties are quadratic in the coefficients $\boldsymbol{\beta}$ (not in the $\tilde{\boldsymbol{\beta}}$). Thus, Equation (31) becomes

$$g\left(\mu_i\right) = \mathbf{A}_i\boldsymbol{\theta} + \mathbf{F}_i\boldsymbol{\gamma} + \mathbf{M}_i\tilde{\boldsymbol{\beta}}. \tag{41}$$

For fitting purposes, the model matrices may be combined column-wise into a single model matrix $\mathbf{X}$. Thus, Equation (41) becomes

$$g\left(\mu_i\right) = \mathbf{X}_i\tilde{\boldsymbol{\beta}}, \tag{42}$$

where $\tilde{\boldsymbol{\beta}}$ has been enlarged to now contain $\boldsymbol{\theta}$, $\boldsymbol{\gamma}$, and the original $\tilde{\boldsymbol{\beta}}$. Similarly there is a corresponding expanded model coefficient vector $\boldsymbol{\beta}$ containing $\boldsymbol{\theta}$, $\boldsymbol{\gamma}$, and the original $\boldsymbol{\beta}$. The penalties on the terms have the general form $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{S}_\lambda\boldsymbol{\beta}$, where $\mathbf{S}_\lambda = \sum_k \lambda_k\mathbf{S}_k$, and the $\mathbf{S}_k$ are the original penalty matrices expanded with zeros everywhere except for the elements which correspond to the coefficients of the kth smooth (Pya & Wood 2015).

The chosen probability distribution determines the form of the log-likelihood $\ell(\boldsymbol{\beta})$ of the model. In order to control model smoothness, the penalized version of the log-likelihood

$$\ell_p(\boldsymbol{\beta}) = \ell(\boldsymbol{\beta}) - \frac{\boldsymbol{\beta}^{\mathrm{T}}\mathbf{S}_\lambda\boldsymbol{\beta}}{2} \tag{43}$$

needs to be maximized. For this, let $V(\mu)$ be the variance of the chosen probability distribution, and define

$$\alpha\left(\mu_i\right) = 1 + \left(y_i - \mu_i\right)\left\{\frac{V'\left(\mu_i\right)}{V\left(\mu_i\right)} + \frac{g''\left(\mu_i\right)}{g'\left(\mu_i\right)}\right\}. \tag{44}$$

Maximization of the penalized log-likelihood is then achieved in the following way (Pya & Wood 2015):

1. Obtain an initial estimate of $\boldsymbol{\beta}$ by minimizing $\left\|g(\mathbf{y}) - \mathbf{X}\tilde{\boldsymbol{\beta}}\right\|^2 + \tilde{\boldsymbol{\beta}}^{\mathrm{T}}S_\lambda\tilde{\boldsymbol{\beta}}$ with respect to $\tilde{\boldsymbol{\beta}}$, subject to linear inequality constraints which ensure that $\tilde{\beta}_j > 0$ whenever $\tilde{\beta}_j = \exp(\tilde{\beta})$.
2. Set $k = 0$ and repeat steps 3–11 until convergence.
3. Evaluate $z_i = \left(y_i - \mu_i\right)g'\left(\mu_i\right)\Big/\alpha\left(\mu_i\right)$ and $w_i = \omega_i\alpha\left(\mu_i\right)\Big/\left\{V\left(\mu_i\right)g'^2\left(\mu_i\right)\right\}$, using the current estimate of μ_i.
4. Evaluate vectors $\tilde{\mathbf{w}} = |\mathbf{w}|$ and $\tilde{\mathbf{z}}$, where $\tilde{z}_i = \operatorname{sign}\left(w_i\right) z_i$.

5. Evaluate diagonal matrix $\mathbf{C}$ such that $C_{j,j} = \begin{cases} 1 & \text{if } \tilde{\beta}_j = \beta_j \\ \exp(\beta_j) & \text{otherwise.} \end{cases}$
6. Evaluate diagonal matrix $\mathbf{E}$ such that $E_{j,j} = \begin{cases} 0 & \text{if } \tilde{\beta}_j = \beta_j \\ \sum_i^n w_i g'(\mu_i)[\mathbf{XC}]_{i,j}(y_i - \mu_i) \Big/ \alpha(\mu_i) & \text{otherwise.} \end{cases}$
7. Let $\mathbf{I}^-$ be a diagonal matrix such that $I^-_{i,i} = \begin{cases} 1 & \text{if } w_i < 0 \\ 0 & \text{otherwise.} \end{cases}$
8. Letting $\tilde{\mathbf{W}}$ denote $\text{diag}(\tilde{\mathbf{w}})$, form the QR decomposition $\begin{bmatrix} \sqrt{\tilde{\mathbf{W}}}\mathbf{XC} \\ \mathbf{B} \end{bmatrix} = \mathbf{QR}$, where $\mathbf{B}$ is any matrix square root such that $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{S}_\lambda$.
9. Letting $\mathbf{Q}_1$ denote the first n rows of $\mathbf{Q}$, form the symmetric eigen-decomposition $\mathbf{Q}_1^{\mathrm{T}}\mathbf{I}^-\mathbf{Q}_1 + \mathbf{R}^{-\mathrm{T}}\mathbf{E}\mathbf{R}^{-1} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{T}}$.
10. Hence define $\mathbf{P} = \mathbf{R}^{-1}\mathbf{U}(\mathbf{I} - \boldsymbol{\Lambda})^{-1/2}$ and $\mathbf{K} = \mathbf{Q}_1\mathbf{U}(\mathbf{I} - \boldsymbol{\Lambda})^{-1/2}$.
11. Update the estimate of $\boldsymbol{\beta}$ as $\boldsymbol{\beta}^{[k+1]} = \boldsymbol{\beta}^{[k]} + \mathbf{P}\mathbf{K}^{\mathrm{T}}\sqrt{\tilde{\mathbf{W}}}\tilde{\mathbf{z}} - \mathbf{P}\mathbf{P}^{\mathrm{T}}\mathbf{S}_\lambda\boldsymbol{\beta}^{[k]}$ and increment k.

The SCAMs presented in this study optimize the GCV/un-biased risk estimator score (Craven & Wahba 1979; Wahba 1990) for selecting the estimate of the smoothing parameter vector $\boldsymbol{\lambda}$.

4.3 Explanation of GAMLSSs

GAMLSSs (Akanztiliotou et al. 2002; Rigby & Stasinopoulos 2001; Rigby & Stasinopoulos 2005) are an extension of GAMs and GLMs. In the case of GAMs and GLMs, only the mean of the response variable (i.e., the location parameter of the assumed probability distribution) is estimated directly from the predictor variables. Variance, skewness, and kurtosis (i.e., the scale and shape parameters of the assumed probability distribution), on the other hand, are in general functions of both the mean and a constant dispersion parameter and are thus estimated only indirectly through their dependence on the mean of the response variable (Rigby & Stasinopoulos 2001). GAMLSSs allow direct estimation of these distribution parameters as well.

In order to explain GAMLSSs, let $\mathbf{y}^{\mathrm{T}} = (y_1, y_2, \ldots, y_n)$ be the vector of independent observations y_i of the response variable, where n is the number of observations. Further, let the y_i follow the probability density function $f(y_i|\boldsymbol{\theta}^i)$ conditional on the vector of distribution parameters $\boldsymbol{\theta}^{i\mathrm{T}} = (\theta_{i,1}, \theta_{i,2}, \ldots, \theta_{i,p})$, where p is the number of distribution parameters. Let J_k be the number of explanatory variables related to the kth distribution parameter $\boldsymbol{\theta}_k$. A GAMLSS is then given

by a known monotonic link function $g_k(\cdot)$ relating $\boldsymbol{\theta}_k$ to explanatory variables and random effects through the additive model

$$\begin{aligned} g_k\left(\boldsymbol{\theta}_k\right) &= \boldsymbol{\eta}_k \\ &= \mathbf{X}_k\boldsymbol{\beta}_k \sum_{j=1}^{J_k} \mathbf{Z}_{j,k}\boldsymbol{\gamma}_{j,k}, \end{aligned} \tag{45}$$

where $\boldsymbol{\theta}_k$ and $\boldsymbol{\eta}_k$ are vectors of length n, $\mathbf{X}_k$ is a known design matrix of order $n \times J'_k$, $\boldsymbol{\beta}_k^{\mathrm{T}} = \left(\beta_{1,k}, \beta_{2,k}, \ldots, \beta_{J'_k,k}\right)$ is a parameter vector, $\mathbf{Z}_{j,k}$ is a fixed known $n \times q_{j,k}$ design matrix and $\boldsymbol{\gamma}_{j,k}$ is a $q_{j,k}$-dimensional random variable. The term $\mathbf{X}_k\boldsymbol{\beta}_k$ is the parametric component of $\boldsymbol{\eta}_k$, whereas the $\mathbf{Z}_{j,k}\boldsymbol{\gamma}_{j,k}$ terms are its additive components (Rigby & Stasinopoulos 2005).

For the explanation of model estimation, assume in Equation (45) that the $\boldsymbol{\gamma}_{j,k}$ have independent (prior) normal distributions with $\boldsymbol{\gamma}_{j,k} \sim N_{q_{j,k}}\left(\mathbf{0}, \mathbf{G}_{j,k}^{-}\right)$, where $\mathbf{G}_{j,k}^{-}$ is the (generalized) inverse of a $q_{j,k} \times q_{j,k}$ symmetric matrix $\mathbf{G}_{j,k} = \mathbf{G}_{j,k}\left(\lambda_{j,k}\right)$, which may depend on a vector of hyperparameters $\lambda_{j,k}$ and where if $\mathbf{G}_{j,k}$ is singular $\boldsymbol{\gamma}_{j,k}$ is understood to have an improper prior density function proportional to $\exp\left(-\frac{1}{2}\boldsymbol{\gamma}_{j,k}^{\mathrm{T}}\mathbf{G}_{j,k}\left(\lambda_{j,k}\right)\boldsymbol{\gamma}_{j,k}\right)$. For fixed $\lambda_{j,k}$s the $\boldsymbol{\beta}_k$s and the $\boldsymbol{\gamma}_k$s are estimated by maximizing a penalized log-likelihood function ℓ_p given by

$$\ell_p = \ell - \frac{1}{2}\sum_{k=1}^{p}\sum_{j=1}^{J_k} \boldsymbol{\gamma}_{j,k}^{\mathrm{T}}\mathbf{G}_{j,k}\left(\lambda_{j,k}\right)\boldsymbol{\gamma}_{j,k}, \tag{46}$$

where $\ell = \sum_{i=1}^{n} \log\left(f\left(y_i|\boldsymbol{\theta}^i\right)\right)$ is the log-likelihood function of the data given $\boldsymbol{\theta}^i$ (Rigby & Stasinopoulos 2005). For the GAMLSSs presented in this study, maximization of ℓ_p was achieved via the algorithm laid out by Rigby & Stasinopoulos (1996). Maximization of ℓ_p leads to the shrinking matrix $\mathbf{S}_{j,k}$, applied to partial residuals $\varepsilon_{j,k}$ to update the estimate of the additive predictor $\mathbf{Z}_{j,k}\boldsymbol{\gamma}_{j,k}$ within a backfitting algorithm, given by

$$\mathbf{S}_{j,k} = \mathbf{Z}_{j,k}\left(\mathbf{Z}_{j,k}^{\mathrm{T}}\mathbf{W}_{k,k}\mathbf{Z}_{j,k} + \mathbf{G}_{j,k}\left(\lambda_{j,k}\right)\right)^{-1}\mathbf{Z}_{j,k}^{\mathrm{T}}\mathbf{W}_{k,k} \tag{47}$$

for $j = 1, 2, \ldots, J_k$ and $k = 1, 2, \ldots, p$, where $\mathbf{W}_{k,k}$ is a diagonal matrix of iterative weights. Different types of additive terms in the linear predictor $\boldsymbol{\eta}_k$ lead to different forms of $\mathbf{Z}_{j,k}$ and $\mathbf{G}_{j,k}$.

The GAMLSSs presented in this study employ P-splines as the basis of their smooth functions, either unconstrained (Eilers & Marx 1996) or with a constraint of being monotone increasing (Bollaerts et al. 2006). Unconstrained P-splines have already been discussed (see section 4.2.1). To impose a monotone increasing constraint on a P-spline, an additional asymmetric penalty on the first-order differences has to be added to Equation (30). The least squares objective function to minimize thus becomes

$$S = \sum_{j=1}^{q}\left(y_j - \sum_{i=0}^{n}\mathbf{P}_i N_{i,p}\left(x_j\right)\right)^2 + \lambda\sum_{i=k+1}^{n}\left(\Delta^k\mathbf{P}_i\right)^2 + \kappa\sum_{i=2}^{n} w_{(\mathbf{P}_i)}\left(\Delta^1\mathbf{P}_i\right)^2 \tag{48}$$

with

$$w_{(\mathbf{P}_i)} = \begin{cases} 0 & \text{if } \Delta^1 \mathbf{P}_i \geq 0 \\ 1 & \text{otherwise} \end{cases}, \tag{49}$$

where κ is a user-defined constraint parameter for fine-tuning the constraint strength and where all other symbols retain the same meaning as in Equation (30), namely: q is the number of observations, (x_j, y_j) are the observations, n is the number of control points, $\mathbf{P}_i$ is the ith control point, $N_{i,p}(x_j)$ is the ith basis function of degree p evaluated at x_j, λ is a smoothness parameter, k is the order of differences, and $\Delta^k \mathbf{P}_i$ is the kth-order difference, i.e., $\Delta^k \mathbf{P}_i = \Delta^1\left(\Delta^{k-1}\mathbf{P}_i\right)$ with $\Delta^1 \mathbf{P}_i = \mathbf{P}_i - \mathbf{P}_{i-1}$ (Bollaerts et al. 2006).

4.4 Model overview

For each species, 6 different models were fitted: GAM1, GAM2, SCAM1, GAMLSS1, GAMLSS2, and GAMLSS3. Data analysis and model fitting was done using the `R` software suite (version 3.3.3) (R Core Team 2017) and several non-base `R` packages. Table 6 lists the non-base `R` packages used for data analysis and the corresponding references. Tables 7 and 8 provide an overview of the settings used for model fitting: Table 7 reports the `R` functions and formulas used for model fitting, while Table 8 reports the probability distributions assumed for each model. The settings for a given model were the same for beech and spruce. In the formulas reported in Table 7, the `~` character separates the dependent variable (on the left) from the predictor terms (on the right), while the `+` character separates individual predictor terms. Thus, in mathematical notation the formula for model GAM1 might be written as

$$g\left(\mu_{Gi}\right) = s_1\left(h_{100}(x)_{\text{YC }1i}\right) + s_2\left(PI_{\text{diff}i}\right), \tag{50}$$

where g is the link function, μ_{Gi} is the ith expected value of basal area, s_1 and s_2 are smooth functions, $h_{100}(x)_{\text{YC }1i}$ is the ith observation of the stand age variable, and $PI_{\text{diff}i}$ is the ith observation of the productivity index variable. In the case of GAM1 and GAM2, the `s(...)` formula terms are smooth functions using thin plate regression splines as their function basis. In the case of SCAM1, the `s(..., bs = "micv")` formula term is a smooth function using P-splines constrained to be increasing and concave as its function basis. In all smooth functions of the GAMs and SCAM, the setting of basis dimension and order of penalty is left to the smooth function. In the case of GAMLSS1 and GAMLSS2, the `ps(...)` formula terms are smooth functions using P-splines as their function basis. In the case of GAMLSS3, the `pbm(...)` term is a smooth function using P-splines constrained to be montonone increasing as its function basis. In all smooth functions of the 3 GAMLSSs, function basis degree was set to 3, function basis order was set to 2, the number of spline knots was set to 20, while selection of the smoothing parameter was left to the smooth function. For all GAMLSSs, the formula reported in Table 7 applies only to the location parameter of the assumed probability distribution. All other distribution parame-

ters were modeled as constants (i.e., with formula `G ~ 1`). Models GAM1 and GAMLSS1 use smooth functions for both the stand age variable as well as the productivity index variable. All other models use smooth functions only for the stand age variable while the productivity index variable is included as a linear predictor. None of the 6 models consider interactions between the predictor terms.

Models GAM1, GAM2, and SCAM1 assume a Gamma distribution. The probability density function and cumulative distribution function of a variable following Gamma distribution are given by

$$f_X(x|k,\theta) = x^{k-1}\frac{\exp\left(\frac{-x}{\theta}\right)}{\theta^k\Gamma(k)} \quad \text{for } x \geq 0,\ k,\theta > 0 \tag{51}$$

and

$$P(X \leq x|k,\theta) = 1 - \frac{\Gamma_{\mathrm{i}}\left(k, \frac{x}{\theta}\right)}{\Gamma(k)} \quad \text{for } x \geq 0,\ k,\theta > 0, \tag{52}$$

respectively, where k is the shape parameter, θ is the scale parameter, Γ is the complete Gamma function, and Γ_{i} is the incomplete Gamma function (Dormann 2013; Lindgren 1976; Weisstein 2017b). The models use the natural logarithm as the link function.
Models GAMLSS1, GAMLSS2, and GAMLSS3 assume a Box-Cox-Cole-Green distribution. This distribution is the Box-Cox transformation model presented by Cole & Green (1992) (Stasinopoulos & Rigby 2007). Based on the transformations suggested by Box & Cox (1964), the model first transforms the dependent variable y such that

$$x = \begin{cases} \frac{(y/\mu)^{\nu}-1}{\nu} & \text{if } \nu \neq 0 \\ \log\left(y/\mu\right) & \text{if } \nu = 0, \end{cases} \tag{53}$$

where x is the transformed variable, μ is the median of y, and ν is the Box-Cox power of y (Cole & Green 1992). The standard deviation score of x is given by

$$z = \frac{x}{\sigma}, \tag{54}$$

where z is the standard deviation score of x and σ is the coefficient of variation of y . It is assumed that z follows a standard normal distribution, i.e., a normal distribution with mean 0 and variance 1. Thus, the probability density function of z is given by

$$f_Z(z) = \frac{\exp\left(-\frac{z^2}{2}\right)}{\sqrt{2\pi}}, \tag{55}$$

while its cumulative distribution function is given by

$$P(Z \leq z) = \frac{\operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) + 1}{2} \tag{56}$$

(Henze 2013; Weisstein 2017c). Substituting z in Equations (55) and (56) using Equations (53) and (54) yields the probabilty density function and the cumulative distribution function of y

$$P(Y = y|\mu, \sigma, \nu) = \begin{cases} \frac{\exp\left(-\frac{\left((y/\mu)^{\nu}-1\right)^2}{2\nu^2\sigma^2}\right)}{\sqrt{2\pi}} & \text{if } \nu \neq 0 \\ \frac{\exp\left(-\frac{\log(y/\mu)^2}{2\sigma^2}\right)}{\sqrt{2\pi}} & \text{if } \nu = 0 \end{cases} \tag{57}$$

and

$$D(Y = y|\mu, \sigma, \nu) = \begin{cases} \frac{1}{2}\left(\operatorname{erf}\left(\frac{(y/\mu)^{\nu}-1}{\sqrt{2}\nu\sigma}\right) + 1\right) & \text{if } \nu \neq 0 \\ \frac{1}{2}\left(\operatorname{erf}\left(\frac{\log(y/\mu)}{\sqrt{2}\sigma}\right) + 1\right) & \text{if } \nu = 0, \end{cases} \tag{58}$$

respectively, where μ, σ, and ν have the same meaning as in Equations (53) and (54), namely: μ is the median of y, σ is the coefficient of variation of y, and ν is the Box-Cox power of y. The models use the natural logarithm as the link function for both μ and σ and the identity function as the link function for ν.

Table 6: Overview of the non-base R packages (including version number) used for data analysis and corresponding references.

Package (Version)	References
`gamlss` (5.0.4)	Rigby & Stasinopoulos (2005)
`gamlss.dist` (5.0.3)	Stasinopoulos (2017)
`mgcv` (1.8.22)	Wood (2003, 2004, 2011, 2017) and Wood et al. (2016)
`scam` (1.2.2)	Pya (2017)

Table 7: Overview of the R functions and formulas used for model fitting. The overview includes the model ID, the name of the R package (and its version number) which provided the model fitting function, the name of the R model fitting function, and the formula used in the model fitting function call. In the case of GAMLSS1, GAMLSS2, and GAMLSS3, the formula applies only to the location parameter of the assumed probability distribution. For all other distribution parameters, the formula was `G ~ 1`.
`G`: basal area vector
`SAV`: stand age variable vector (cp. Equation (3))
`PIV`: productivity index variable vector (cp. Equation (4))

Model ID	Package (Version)	Fitting function	Formula
GAM1	`mgcv` (1.8.22)	`gam`	`G ~ s(SAV) + s(PIV)`
GAM2	As above	As above	`G ~ s(SAV) + PIV`
SCAM1	`scam` (1.2.2)	`scam`	`G ~ s(SAV, bs = "micv") + PIV`
GAMLSS1	`gamlss` (5.0.4)	`gamlss`	`G ~ ps(SAV) + ps(PIV)`
GAMLSS2	As above	As above	`G ~ ps(SAV) + PIV`
GAMLSS3	As above	As above	`G ~ pbm(SAV) + PIV`

Table 8: Overview of the R distribution functions used for model fitting. The overview includes the model ID, the name of the R package (and its version number) which provided the distribution function, and the R call of the distribution function used in model fitting.

Model ID	Package (Version)	Distribution function call
GAM1	`stats` (3.3.3)	`Gamma(link = "log")`
GAM2	As above	As above
SCAM1	As above	As above
GAMLSS1	`gamlss.dist` (5.0.3)	`BCCGo()`
GAMLSS2	As above	As above
GAMLSS3	As above	As above

5 Results

Summary statistics for all models are reported in the Appendix in Tables 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 and 25. All linear effects are significant at the 5 % level. The AIC of all 6 models for each species is reported in Table 9. The lowest AIC for beech and spruce is achieved by models GAMLSS1 (380.189) and GAMLSS3 (705.311), respectively.

Table 9: Akaike Information Criterion (AIC) scores (Akaike 1998) of all models for beech and spruce. Numbers in parentheses give the model rank, from lowest to highest AIC.

	AIC	
Model ID	Beech	Spruce
GAM1	400.109 (3)	705.998 (2)
GAM2	430.491 (5)	718.058 (6)
SCAM1	442.729 (6)	717.144 (5)
GAMLSS1	380.189 (1)	707.125 (3)
GAMLSS2	399.319 (2)	710.280 (4)
GAMLSS3	412.600 (4)	705.311 (1)

In the following sections, the results of each model will be presented via the following plots:

- Estimated effect of the stand age variable term over stand age variable.
- Estimated effect of the productivity index variable term over productivity index variable.
- Quantile-quantile plot of model residuals.
- Model residuals over fitted values.
- Observed and predicted basal area over stand age.
- Observed and predicted basal area over top height.

Data on which model predictions are based were generated in 2 steps. First, using an increment of one year, a sequence of age values was generated, ranging from 2 a to 160 a. Then, corresponding sequences of top height values were calculated using the equation

$$h_{100}(x) = \beta_0 + \beta_1 \cdot \ln(x) + \beta_2 \cdot \ln(x)^2 + PI_i \cdot (\beta_3 + \beta_4 \cdot \ln(x)), \tag{59}$$

where PI_i is the absolute productivity index of stand for yield class i as reported in Table 5, with i ranging from worst to best observed yield class (3 to −2 for beech, 4 to −2 for spruce), and all other terms have the same meaning as in Equation (2), namely: $h_{100}(x)$ is top height in meter at age x, x is stand age, and $\beta_0, \ldots, \beta_4$ are species-specific coefficients as reported in Table 4 (Nagel

1999). Based on these top height sequences, corresponding sequences of stand age variable and productivity index variable were calcluated using Equations (3) and (4), respectively. These were then used as input for model predictions.

5.1 GAM1

Figure 10 depicts the estimated effect of the stand age variable smooth function over the stand age variable in model GAM1. The smooth function uses unconstrained thin plate regression splines as its basis. For beech, the estimated effect reaches its absolute maximum at roughly $h_{100}(x)_{\mathrm{YC}\,1} = 27.5\,\mathrm{m}$, which corresponds to an age of about 76 a. After that, the effect starts to decrease. In the case of spruce, the effect increases throughout all of the interval depicted, leading to the highest estimate at the end of the interval ($h_{100}(x)_{\mathrm{YC}\,1} = 40\,\mathrm{m}$), which corresponds to an age of about 125 a. Compared to spruce, the estimated effect of the stand age variable covers a narrower range in the beech model. This is in accordance with differences in the range of the stand age variable between species. The estimated effect does not show signs of asymptotic behavior in either species. The effective degrees of freedom of the smooth function are 2.66 and 6.39 for beech and spruce, respectively.

Figure 11 depicts the estimated effect of the productivity index variable smooth function over the productivity index variable in model GAM1. The smooth function also uses unconstrained thin plate regression splines as its basis. In the case of beech, the estimated effect shows a sharp increase for $-12.5\,\mathrm{m} \leq PI_{\mathrm{diff}} \leq -2.5\,\mathrm{m}$, starting at an estimate of approximately 3 before almost completely levelling off at an estimate of zero. In the case of spruce, the estimate starts at around zero and never strays very far from it in the interval depicted. The effective degrees of freedom of the smooth function are 4.87 and 6.48 for beech and spruce, respectively.

Figure 12 shows the quantile-quantile plots for model GAM1. In both species, response residuals follow the reference line fairly closely with some residuals lying outside the 90 % reference band.

Figure 13 shows response residuals over fitted values of model GAM1. In both species, variability of residuals appears to increase with fitted value.

Figure 14 shows basal area over stand age, both observed values as well as predictions of model GAM1. General behavior of the model is notably different for each species: in the case of beech, predicted basal area starts to decrease between stand ages of 70 a to 90 a, while in spruce it increases throughout the whole depicted stand age interval, especially between 80 a to 100 a. In both species, predicted basal area shows stratification depending on yield class. However, the order of curves does not coincide with yield class order: in the case of beech, yield class 1 lies above yield class 0, while in the case of spruce, yield class 1 lies above yield class -2 and yield class 3 lies above yield class 2. In beech, yield class 3 performs much worse than yield class 2. Other yield classes of beech and all yield classes of spruce do not exhibit as large distances

between adjacent yield classes.

Figure 15 shows basal area over top height, both observed values as well as predictions of model GAM1. General model behavior is the same as in Figure 14. Stratification of curves depending on yield class is visible. In the case of beech, the order of curves (excluding yield class 3) changes during the depicted age interval. For top heights up to 22 m, order of curves is almost the inverse of yield class order, with yield class 2 showing best performance and yield class -2 showing worst performance. Between top heights of 22 m to 35 m, order of curves shifts until eventually stratification follows yield class order. In contrast, spruce shows a highly erratic pattern of stratification, the only constant of which is that yield class -2 performs worst of all yield classes throughout.

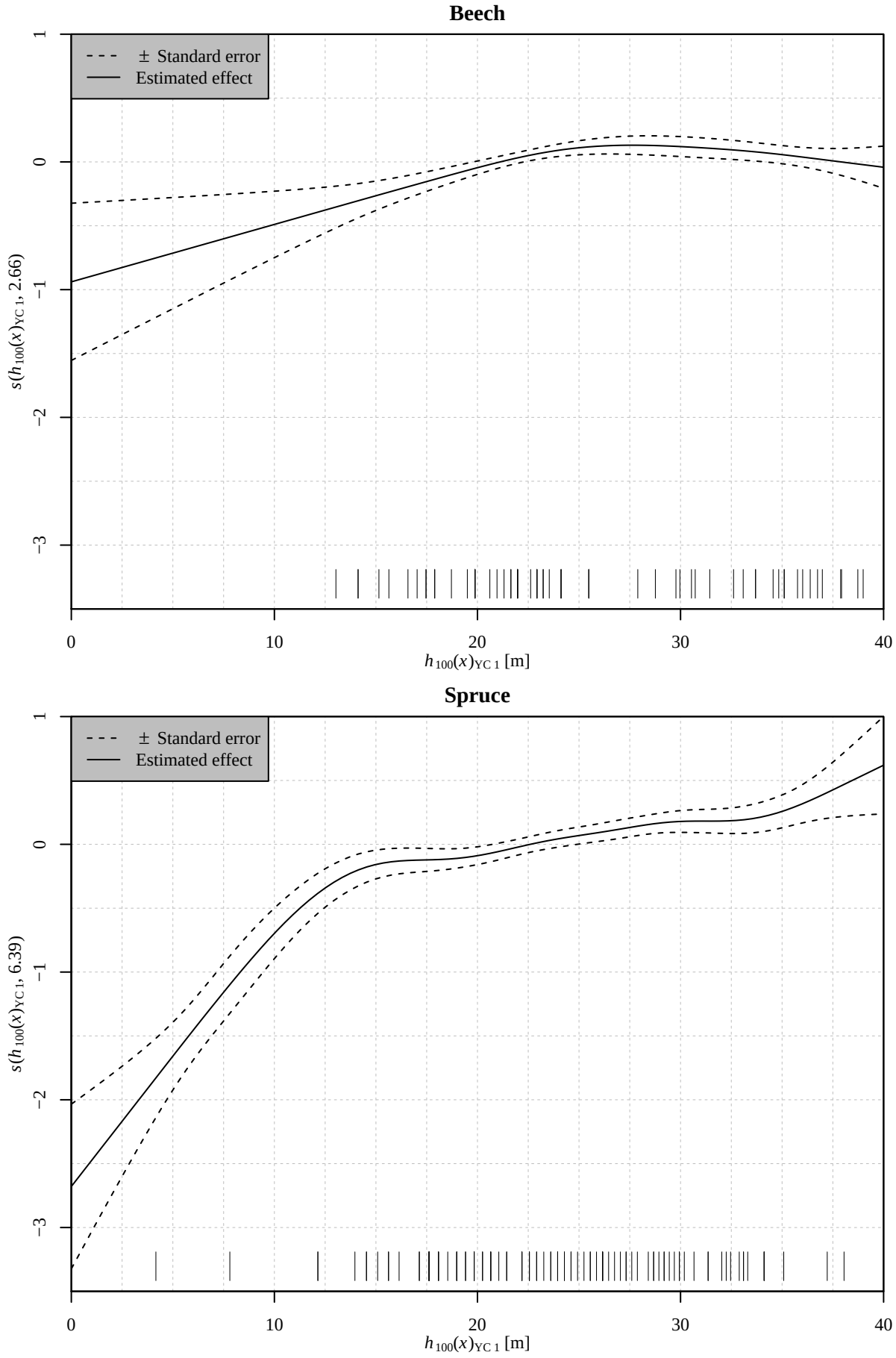

Figure 10: Estimated effect of the stand age variable smooth function ($s(h_{100}(x)_{\mathrm{YC}\,1}, \ldots)$) over stand age variable ($h_{100}(x)_{\mathrm{YC}\,1}$) in model GAM1 for beech (top) and spruce (bottom). Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Numbers in the y-axis titles are the effective degrees of freedom of the smooth function.

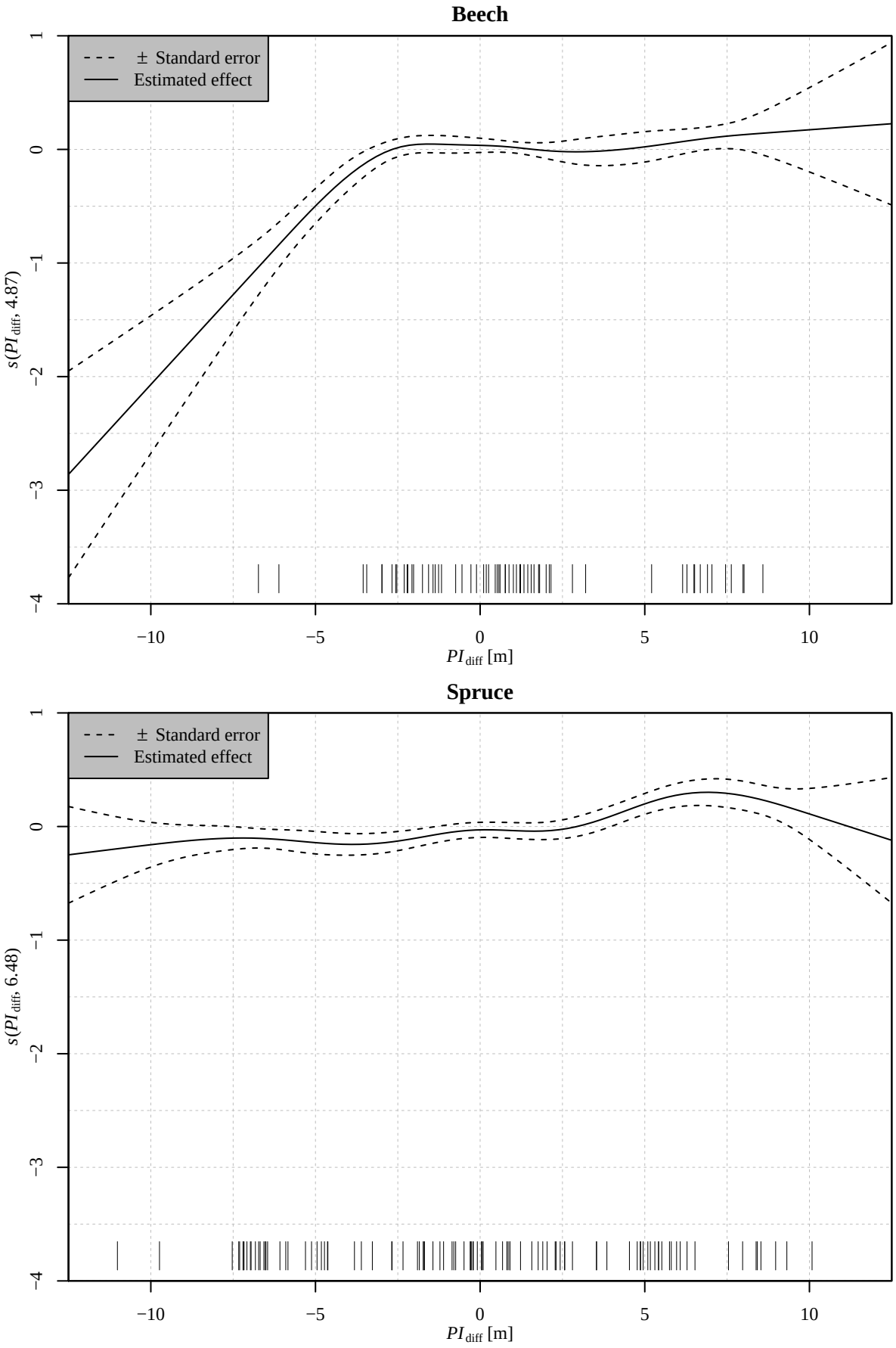

Figure 11: Estimated effect of the productivity index variable smooth function ($s(PI_{\text{diff}}, \ldots)$) over productivity index variable (PI_{diff}) in model GAM1 for beech (top) and spruce (bottom). Solid lines, dashed lines, vertical bars, and numbers in the y-axis titles have the same meaning as in Figure 10, namely: Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Numbers in the y-axis titles are the effective degrees of freedom of the smooth function.

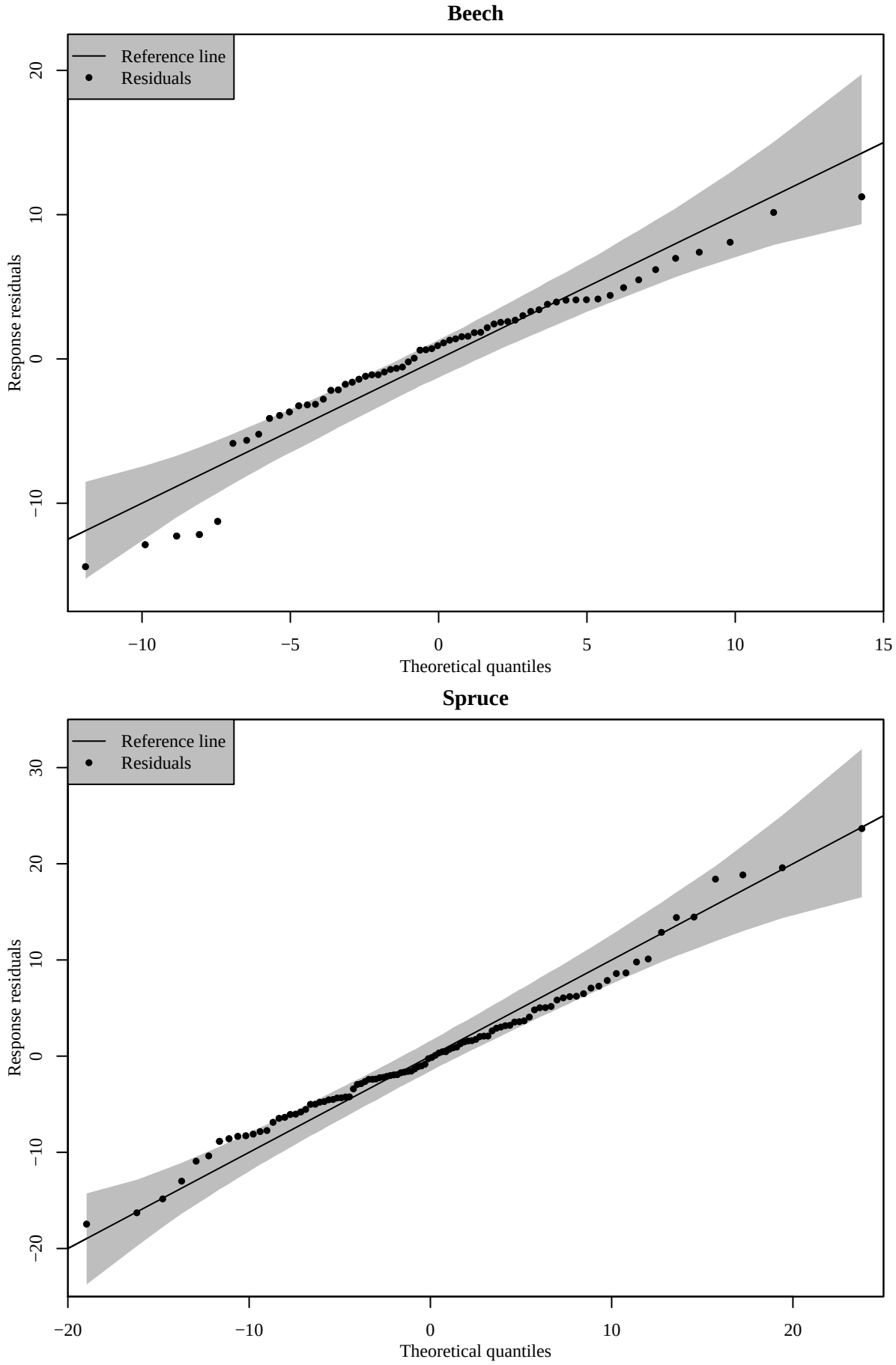

Figure 12: Quantile-quantile plot of residuals of model GAM1 for beech (top) and spruce (bottom). Black lines are reference lines. Black dots represent residuals. Gray shaded areas mark reference bands between the 0.05 and 0.95 quantiles of predictions (90 % level). Theoretical quantiles and reference bands are based on repeated predictions ($N = 10^4$) (Augustin et al. 2012). Note the different axis scaling in both plots.

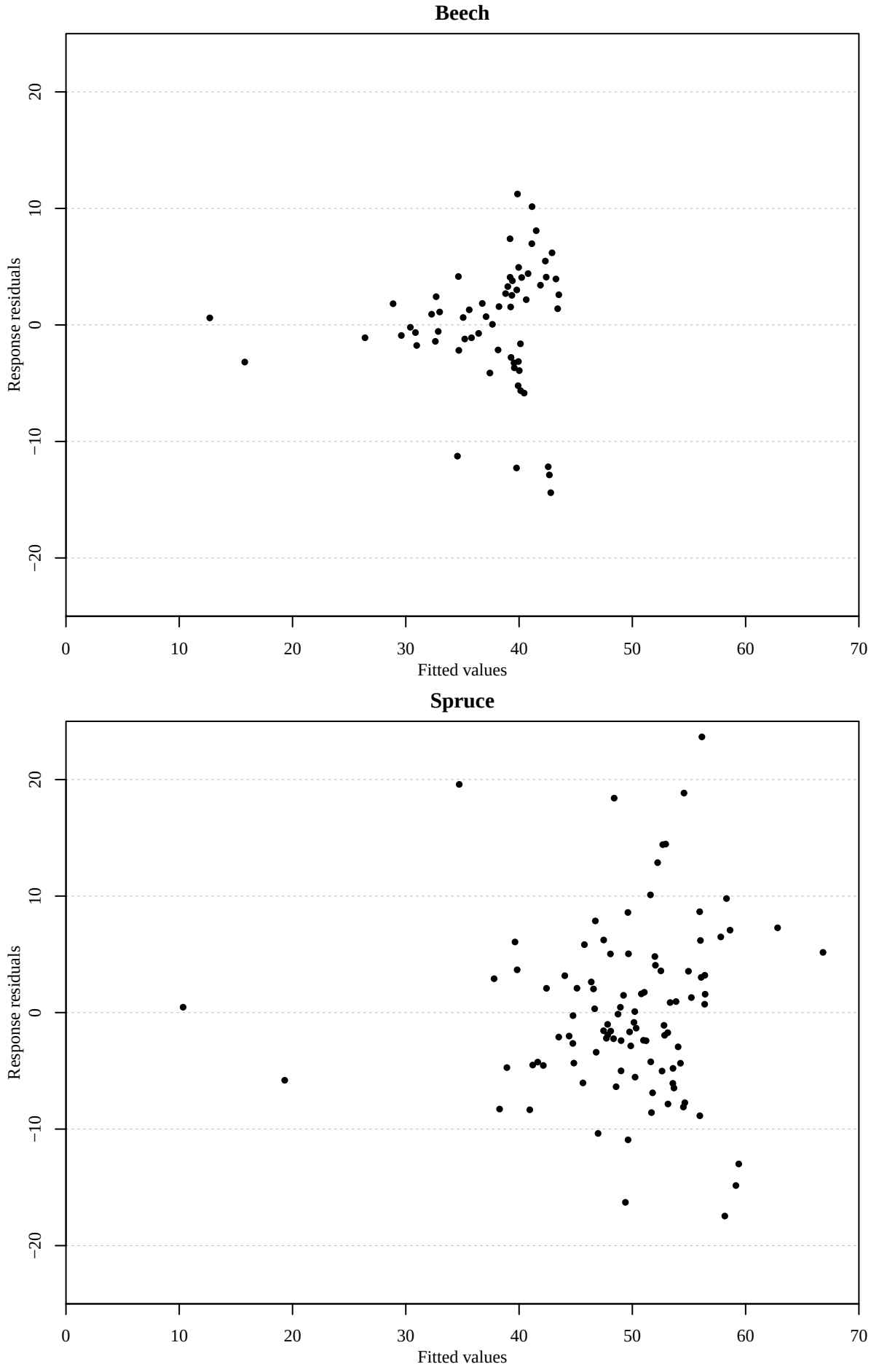

Figure 13: Observed values minus fitted values (Response residuals) over fitted values of model GAM1 for beech (top) and spruce (bottom).

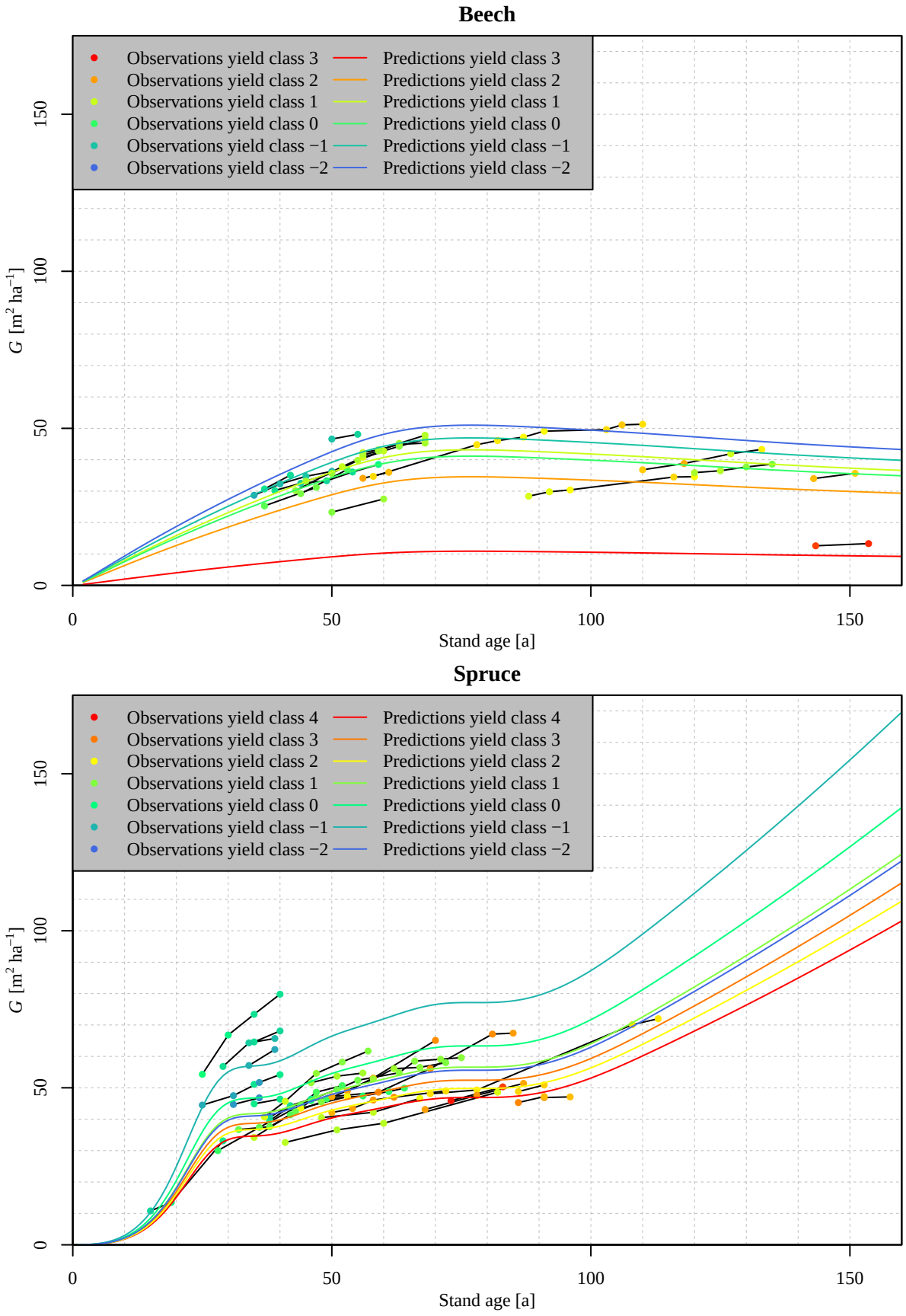

Figure 14: Basal area (*G*) over stand age of beech (top) and spruce (bottom). Colored lines represent predictions of model GAM1. Dots, black lines, and colors have the same meaning as in Figure 4, namely: Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

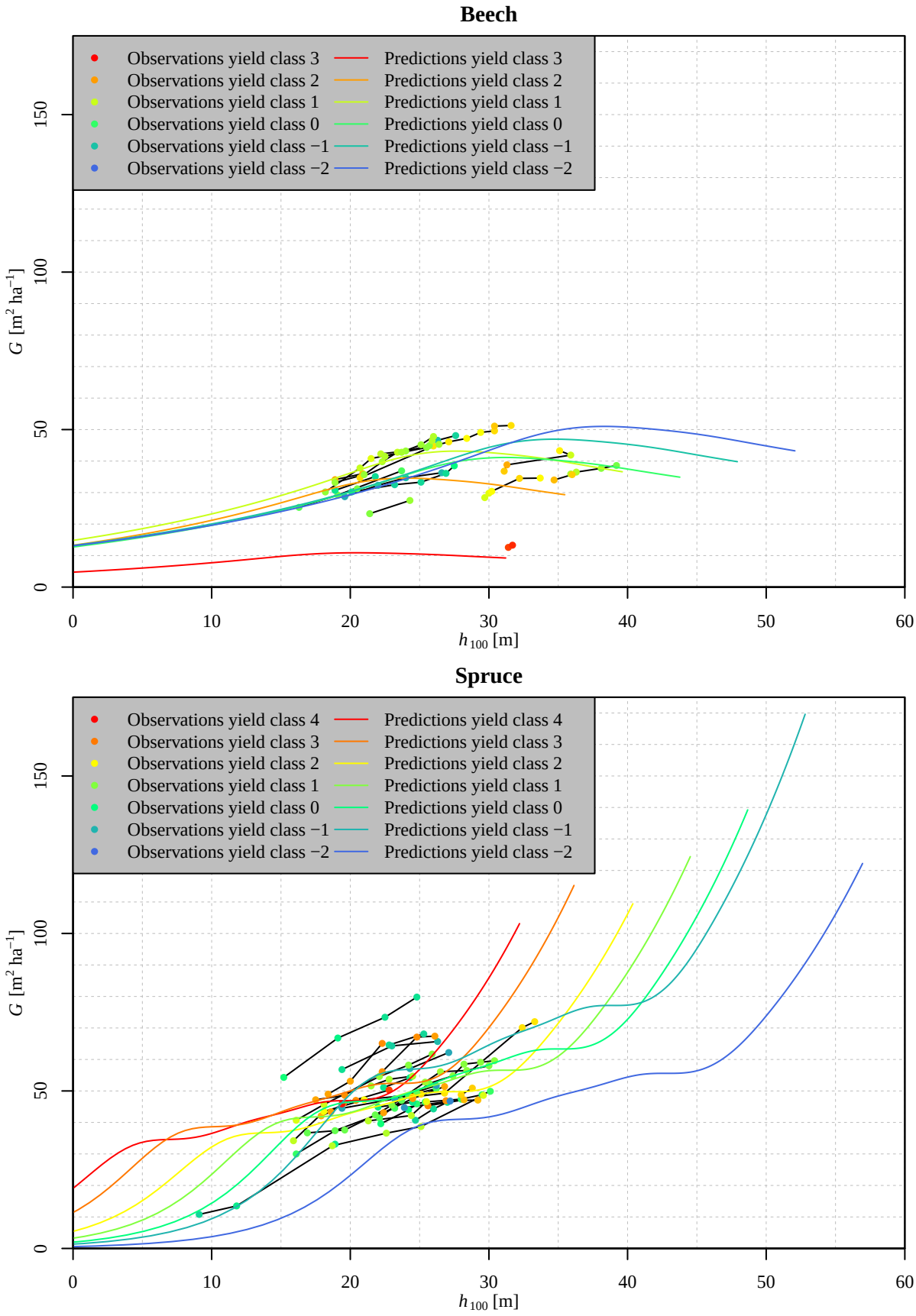

Figure 15: Basal area (G) over top height (h_{100}) of beech (top) and spruce (bottom). Colored lines, dots, black lines, and colors have the same meaning as in Figure 14, namely: Colored lines represent predictions of model GAM1. Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

5.2 GAM2

Figure 16 shows the estimated effect of the stand age variable smooth function over the stand age variable in model GAM2. The smooth function uses unconstrained thin plate regression splines as its basis. General characteristics of the curves are similar to those in the corresponding figure of model GAM1 (cp. Figure 10): for beech, the curve is markedly concave, reaching its maximum at approximately $h_{100}(x)_{\mathrm{YC}\,1} = 29\,\mathrm{m}$, which corresponds to an age of circa 85 a; for spruce, the curve exhibits a steep increase up to roughly $h_{100}(x)_{\mathrm{YC}\,1} = 15\,\mathrm{m}$, which is equivalent to an age of about 30 a, after which the curve's slope is reduced while remaining positive. As in model GAM1, the estimated effect covers a narrower range of values in beech than in spruce and does not show signs of asymptotic behavior in either species. The effective degrees of freedom of the smooth function are 3.7 and 5.32 for beech and spruce, respectively.

Figure 17 depicts the estimated linear effect of the productivity index variable over the productivity index variable in model GAM2. The effect is monotone increasing in both species, having its root at approximately 0 m.

Figure 18 shows the quantile-quantile plots for model GAM2. In beech, several residuals lie outside the 90 % reference band, scattered across the whole range of theoretical quantiles. In spruce, a few low quantile residuals lie outside the reference band.

Figure 19 shows response residuals over fitted values of model GAM2. In beech, residual variance appears to be constant over the range of fitted values depicted, whereas in spruce, residual variance appears to be higher for fitted values around 50 and lower for other values.

Figure 20 depicts basal area over stand age, both observed values as well as predictions of model GAM2. General model behavior is different between species: in beech, prediction curves are concave, reaching their maximum around a stand age of 80 a; in spruce, prediction curves are increasing throughout the depicted range of stand ages, with a steep increase from 0 a to 30 a and a slower increase thereafter. In both species, curves are clearly stratified depending on yield class, with curve order following yield class order at all times, i.e., a yield class shows higher basal area predictions than the next worse yield class.

Figure 21 shows basal area over top height, both observed values as well as predictions of model GAM2. As in Figure 20, the beech plot exhibits clear stratification of prediction curves depending on yield class. For ages up to 20 a, curve order is the inverse of yield class order. Between 20 a to 33 a, curve order switches, so that for ages above 30 a, curve order is the same as yield class order. In spruce, curve stratification is visible as well, but curve order is highly erratic, with yield class -2 performing worst and yield class 4 performing best for almost the entire range of stand ages depicted.

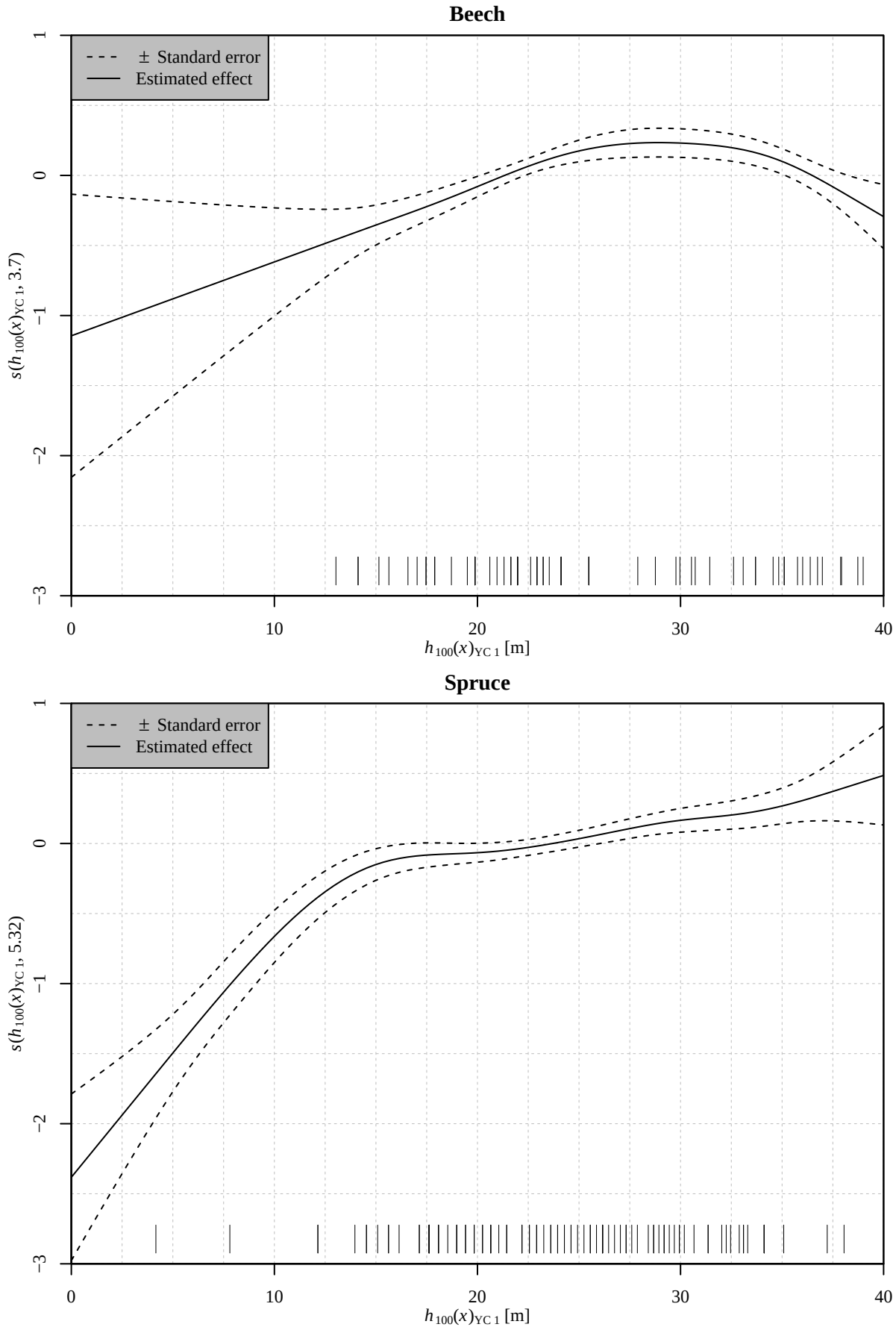

Figure 16: Estimated effect of the stand age variable smooth function ($s(h_{100}(x)_{\mathrm{YC}\,1}, \ldots)$) over stand age variable ($h_{100}(x)_{\mathrm{YC}\,1}$) in model GAM2 for beech (top) and spruce (bottom). Solid lines, dashed lines, vertical bars, and numbers in the y-axis titles have the same meaning as in Figure 11, namely: Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Numbers in the y-axis titles are the effective degrees of freedom of the smooth function.

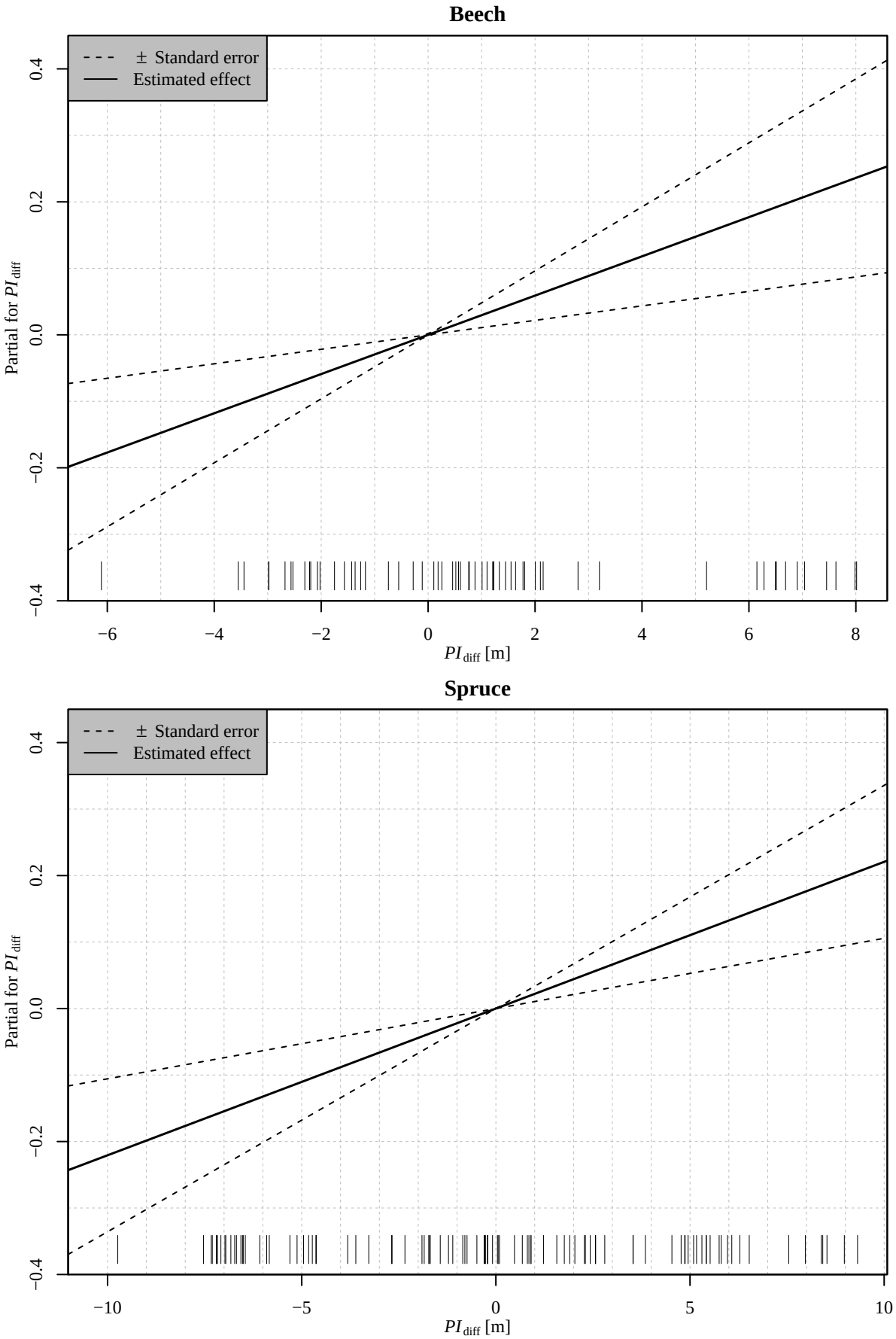

Figure 17: Estimated effect of the parametric productivity index variable term (Partial for PI_{diff}) over productivity index variable (PI_{diff}) in model GAM2 for beech (top) and spruce (bottom). Solid lines, dashed lines, and vertical bars have the same meaning as in Figure 16, namely: Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Note the different scaling of the x-axis in both plots.

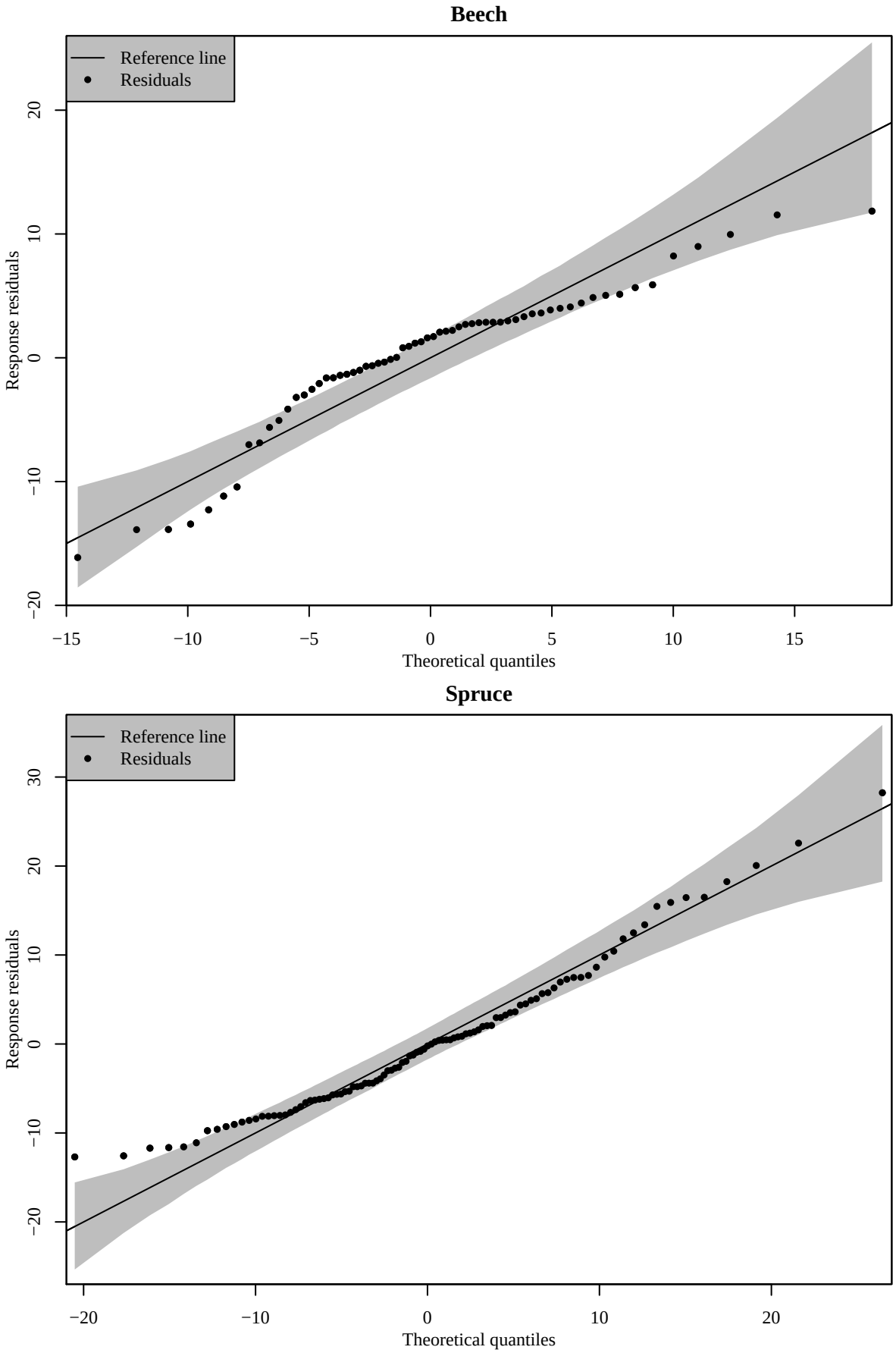

Figure 18: Quantile-quantile plot of residuals of model GAM2 for beech (top) and spruce (bottom). Black lines, black dots, and gray shaded areas have the same meaning as in Figure 12, namely: Black lines are reference lines. Black dots represent residuals. Gray shaded areas mark reference bands between the 0.05 and 0.95 quantiles of predictions (90 % level). Theoretical quantiles and reference bands are based on repeated predictions ($N = 10^4$) (Augustin et al. 2012). Note the different axis scaling in both plots.

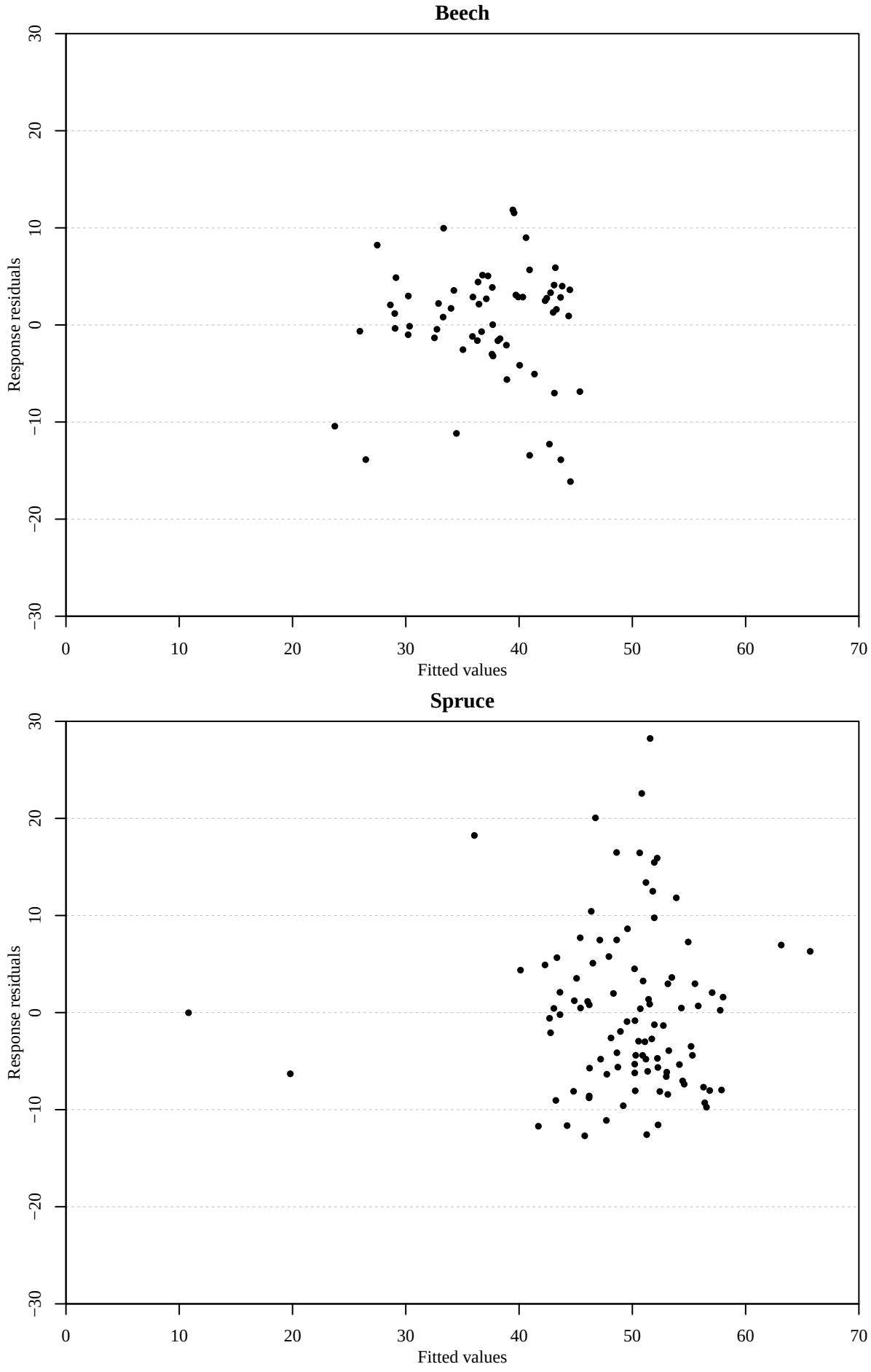

Figure 19: Observed values minus fitted values (Response residuals) over fitted values of model GAM2 for beech (top) and spruce (bottom).

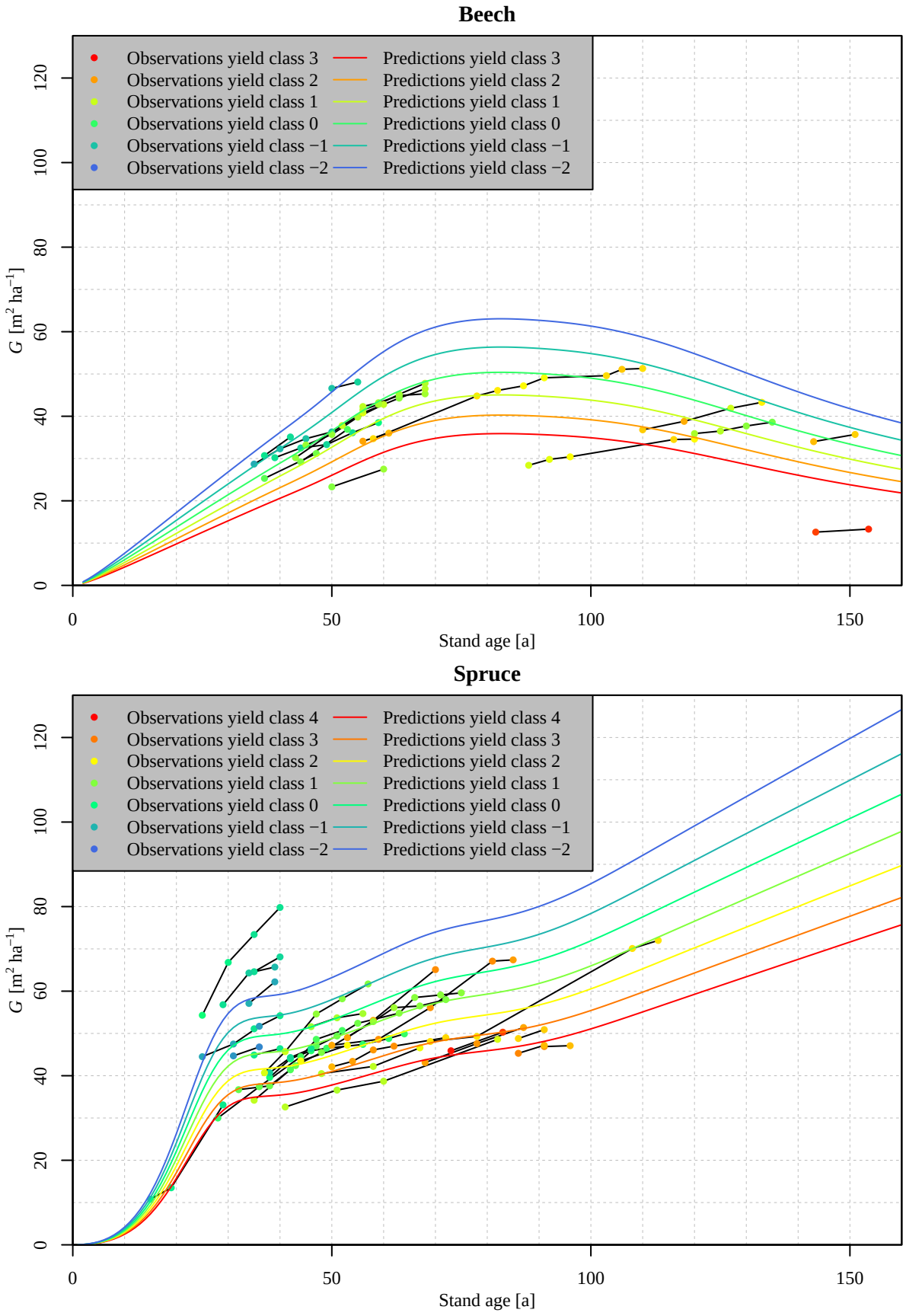

Figure 20: Basal area (G) over stand age of beech (top) and spruce (bottom). Colored lines represent predictions of model GAM2. Dots, black lines, and colors have the same meaning as in Figure 15, namely: Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

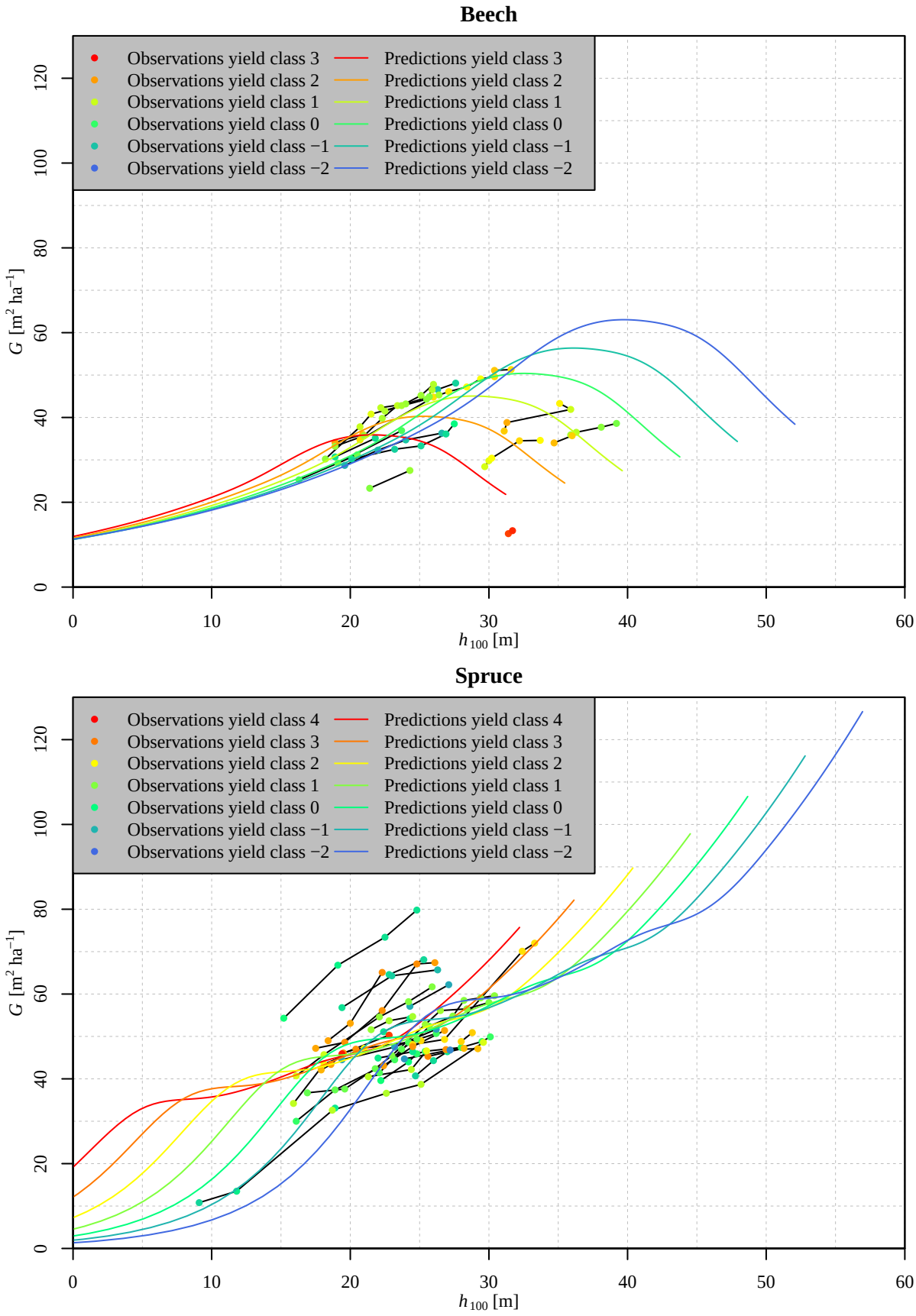

Figure 21: Basal area (G) over top height (h_{100}) of beech (top) and spruce (bottom). Colored lines, dots, black lines, and colors have the same meaning as in Figure 20, namely: Colored lines represent predictions of model GAM2. Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

5.3 SCAM1

Figure 22 shows the estimated effect of the stand age variable smooth function over the stand age variable in model SCAM1. The smooth function uses P-splines constrained to be increasing and concave as its basis. In the case of beech, the smooth function behaves asymptotic: for stand age variable values up to circa 25 m (which corresponds to an age of about 67 a), the smooth function value increases before levelling off. The effective degrees of freedom of the smooth function is 1.86. In the case of spruce, the overall curve shape is similar to the corresponding figure of model GAM2 (cp. Figure 16): the smooth function value increases steeply for stand age variable values up to 12.5 m (which corresponds to an age of about 25 a), after which the increase continues, but at a lower rate. The effective degrees of freedom of the smooth function is 3.01.

Figure 23 depicts the estimated linear effect of the productivity index variable over the productivity index variable in model SCAM1. As in model GAM2, the effect is monotone increasing in both species, having its root at approximately 0 m.

Figure 24 shows the quantile-quantile plots for model SCAM1. In beech, several residuals of low, intermediate, and high quantiles lie outside the 90 % reference band. In spruce, multiple residuals of low and intermediate quantiles lie outside the reference band.

Figure 25 shows response residuals over fitted values of model SCAM1. In beech, residual variance appears to be constant over the range of fitted values depicted, whereas in spruce, residual variance appears to be higher for fitted values around 50 and lower for other values.

Figure 26 depicts basal area over stand age, both observed values as well as predictions of model SCAM1. In both species, the overall shape of prediction curves follows closely that of the stand age variable smooth function as shown in Figure 22: in beech, predicted basal area increases up to an age of about 67 a at which it levels off; in spruce, predicted basal area increases steeply up to an age of about 25 a, after which the increase continues, albeit at a lower rate. In both species, curves for different yield classes are clearly separated from each other, with curve order being the inverse of yield class order in both species.

Figure 27 shows basal area over top height, both observed values as well as predictions of model SCAM1. In beech, stratification of prediction curves is rather weakly pronounced up to a top height of about 17 m, with curve order being the inverse of yield class order, i.e., worse yield classes outperforming better yield classes. Between top heights of 17 m to 31 m, curve order switches due to basal area of worse yield classes levelling off at lower top heights than that of better yield classes, so that for top heights above 31 m curve order follows yield class order. In spruce, a similar pattern is visible: for top heights up to 9 m, curve order is the inverse of yield class order; between a top height of 9 m to 26 m, curve order switches due to curve slopes of worse yield classes decreasing at lower top heights than that of better yield classes, so that for top heights above 26 m curve order follows yield class order.

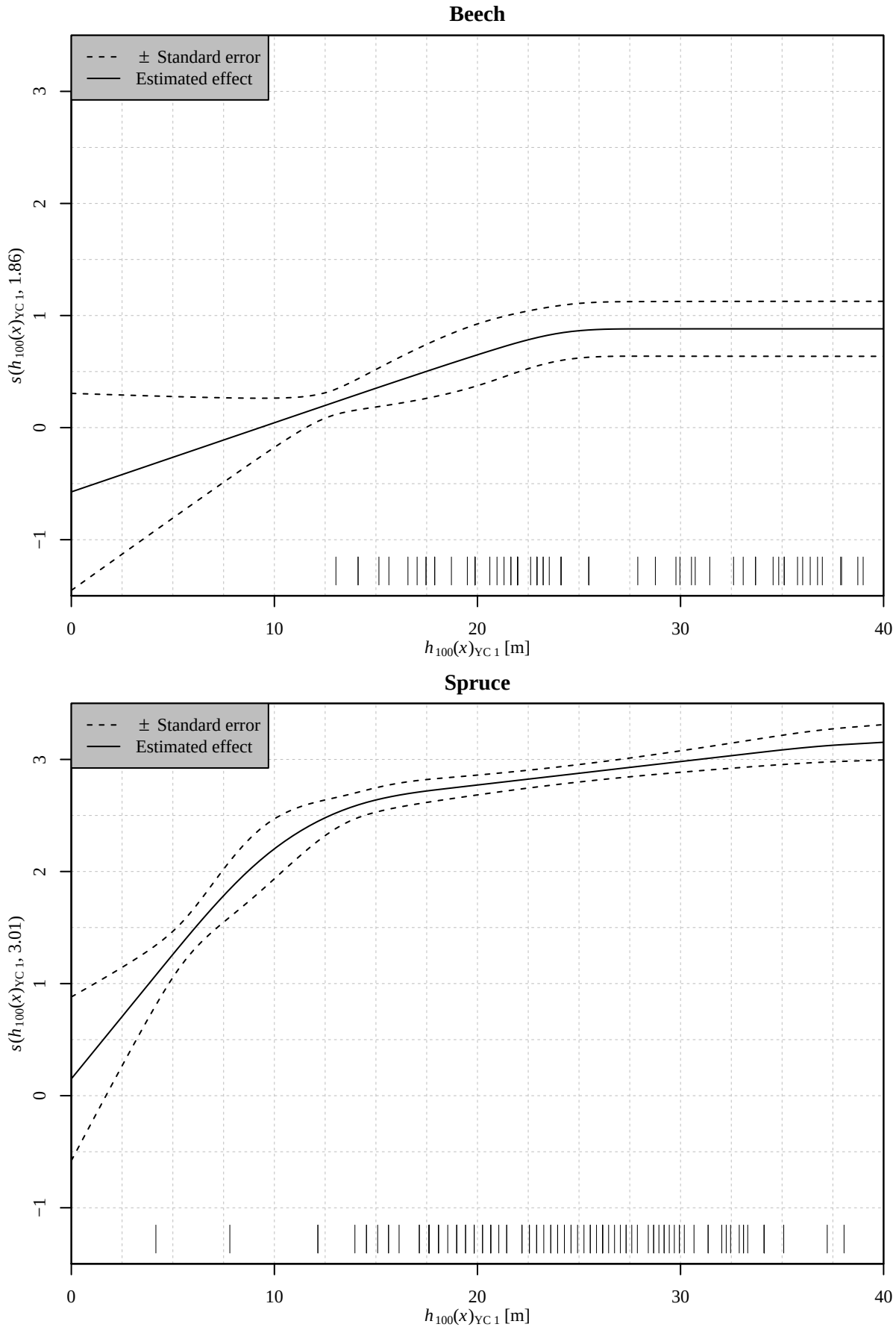

Figure 22: Estimated effect of the stand age variable smooth function ($s(h_{100}(x)_{\mathrm{YC}\,1}, \ldots)$) over stand age variable ($h_{100}(x)_{\mathrm{YC}\,1}$) in model SCAM1 for beech (top) and spruce (bottom). Solid lines, dashed lines, vertical bars, and numbers in the y-axis titles have the same meaning as in Figure 17, namely: Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Numbers in the y-axis titles are the effective degrees of freedom of the smooth function.

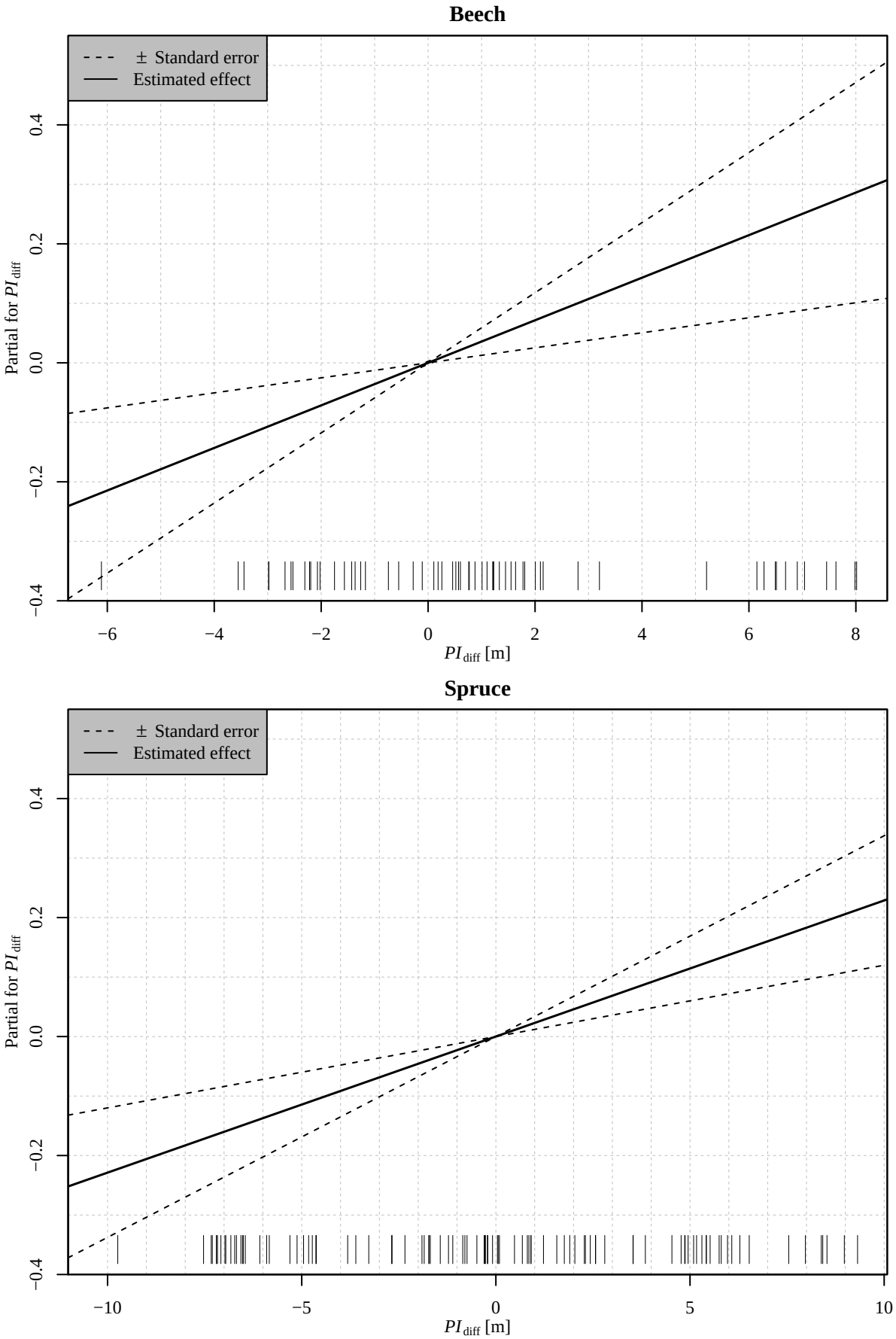

Figure 23: Estimated effect of the parametric productivity index variable term (Partial for PI_{diff}) over productivity index variable (PI_{diff}) in model SCAM1 for beech (top) and spruce (bottom). Solid lines, dashed lines, and vertical bars have the same meaning as in Figure 22, namely: Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Note the different scaling of the x-axis in both plots.

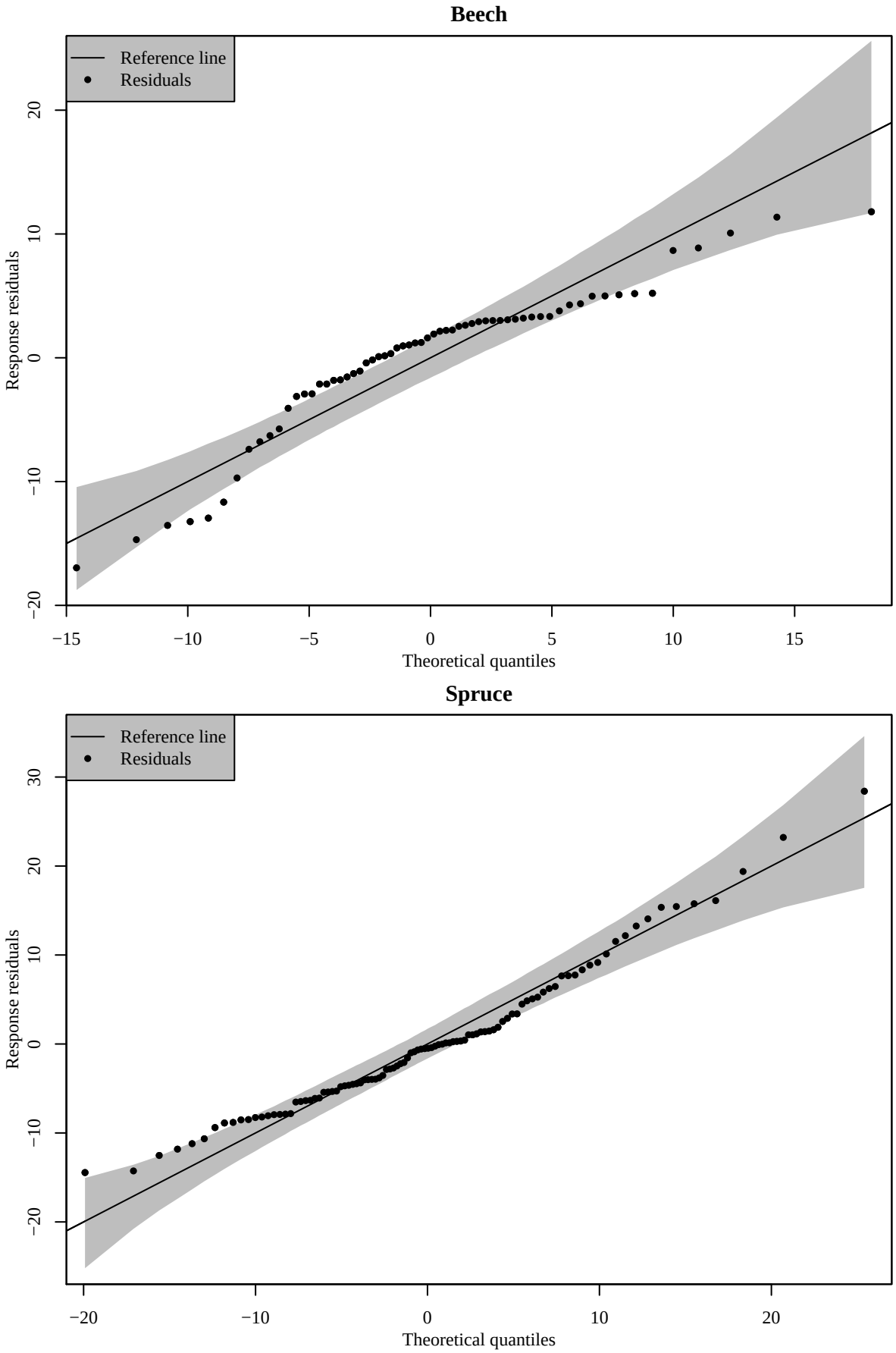

Figure 24: Quantile-quantile plot of residuals of model SCAM1 for beech (top) and spruce (bottom). Black lines, black dots, and gray shaded areas have the same meaning as in Figure 18, namely: Black lines are reference lines. Black dots represent residuals. Gray shaded areas mark reference bands between the 0.05 and 0.95 quantiles of predictions (90 % level). Theoretical quantiles and reference bands are based on repeated predictions ($N = 10^4$) (Augustin et al. 2012). Note the different axis scaling in both plots.

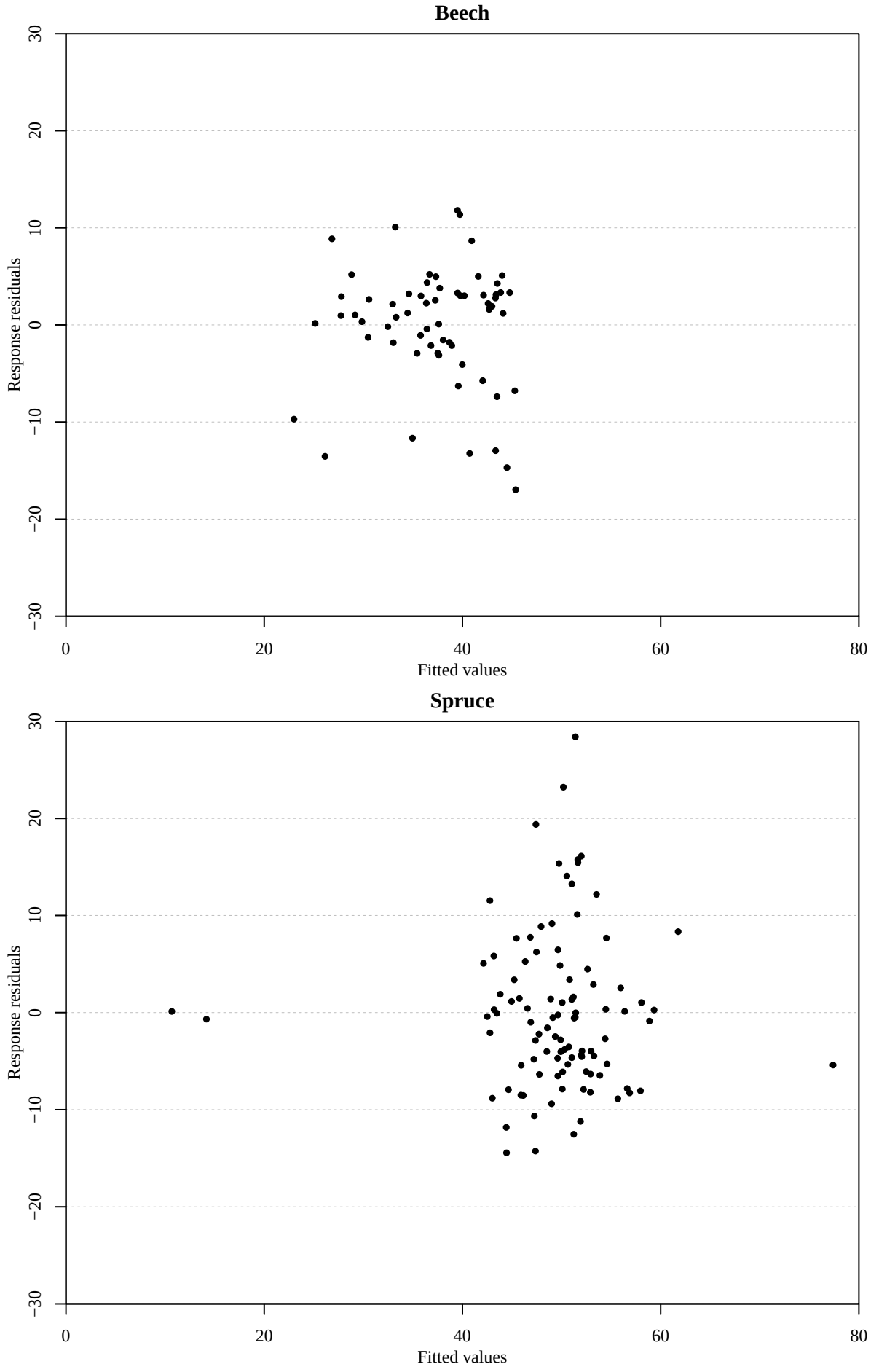

Figure 25: Observed values minus fitted values (Response residuals) over fitted values of model SCAM1 for beech (top) and spruce (bottom).

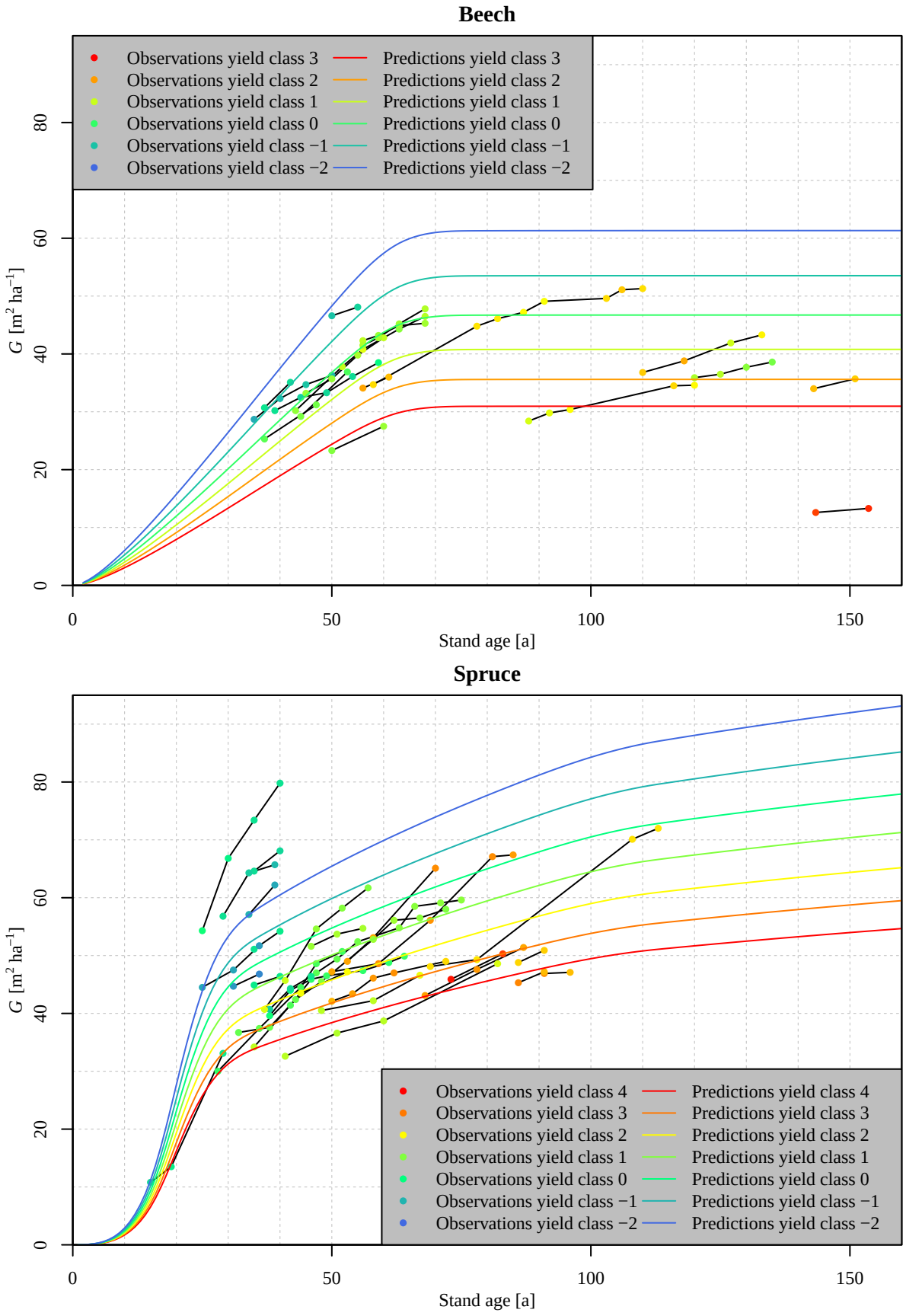

Figure 26: Basal area (G) over stand age of beech (top) and spruce (bottom). Colored lines represent predictions of model SCAM1. Dots, black lines, and colors have the same meaning as in Figure 21, namely: Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

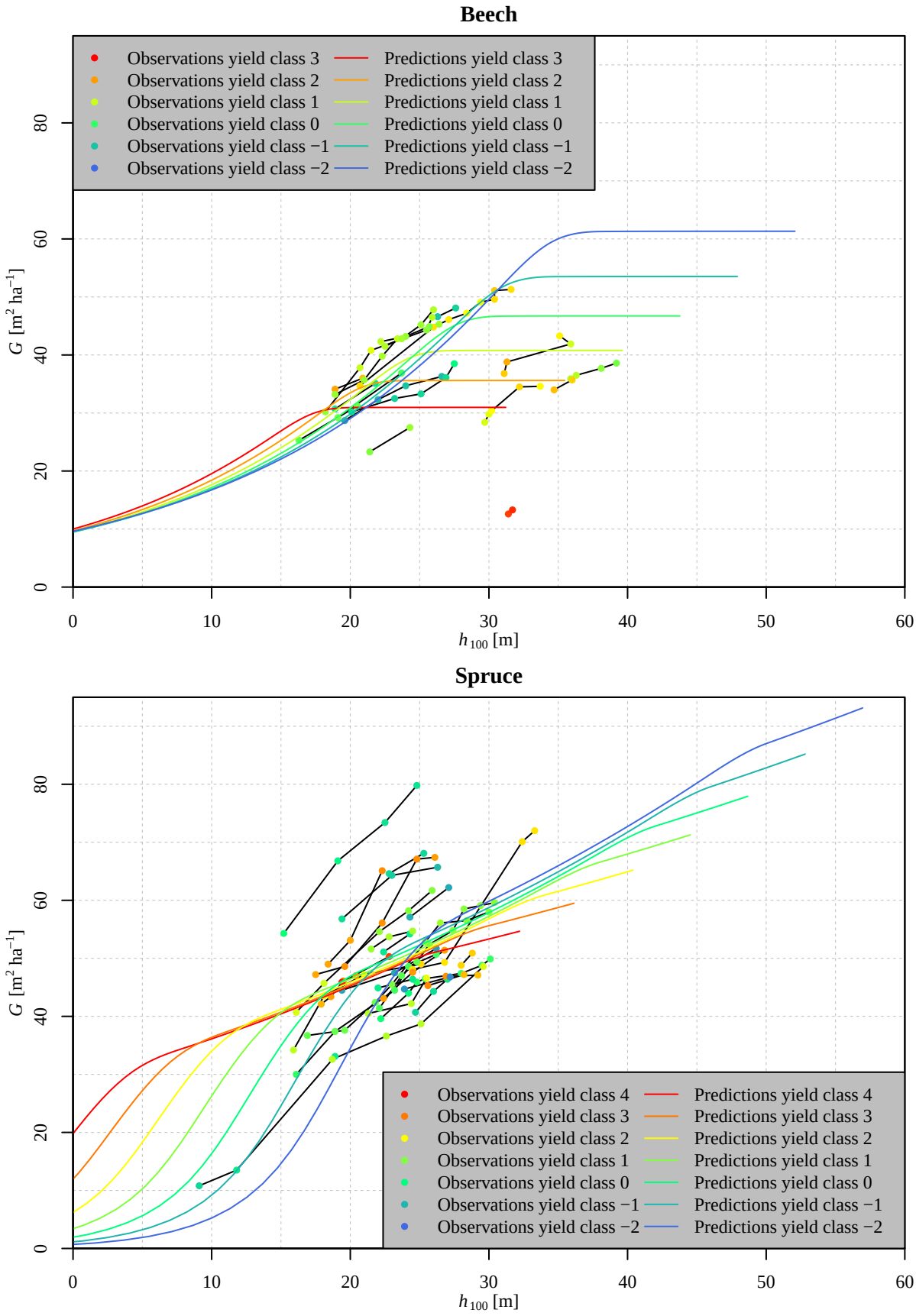

Figure 27: Basal area (G) over top height (h_{100}) of beech (top) and spruce (bottom). Colored lines, dots, black lines, and colors have the same meaning as in Figure 26, namely: Colored lines represent predictions of model SCAM1. Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

5.4 GAMLSS1

Figure 28 depicts the estimated effect of the stand age variable smooth function over the stand age variable in model GAMLSS1. The smooth function uses unconstrained P-splines as its basis. For beech, the effect is concave, reaching its maximum at a stand age variable value of circa 30 m, which corresponds to an age of about 90 a. In spruce, the effect increases steeply between stand age variable values of 5 m to 15 m (corresponding to an stand age range of 18 a to 30 a), after which the curve's slope decreases while remaining positive. At circa 33 m (which corresponds to an age of about 86 a) the slope again increases.

Figure 29 shows the estimated effect of the productivity index variable smooth function over the productivity index variable in model GAMLSS1. The smooth function uses unconstrained P-splines as its basis. In beech, the curve is rather irregularly formed: for productivity index variable values between -6 m to -3.5 m, the curve has a steep positive slope; afterwards, the slope decreases, eventually becoming negative, leading to a local minimum of the curve at a productivity index variable value of roughly 2.5 m; after this, the slope increases again. For spruce, the effect's curve has an almost convex shape, with the lowest and highest effect occurring at the left and right edge of the plot, i.e., at a productivity index variable value of circa -11 m and 10 m, respectively.

Figure 30 shows the quantile-quantile plots for model GAMLSS1. In both species, several residuals deviate form the reference line. However, only one to two of these deviations extend outside the 90 % reference band in either species.

Figure 31 depicts the normalized quantile residuals over fitted values of model GAMLSS1. In beech, residual variance appears to increase with fitted values, whereas in spruce, residual variance appears to be constant over the range of fitted values shown.

Figure 32 shows basal area over stand age, both observed values as well as predictions of model GAMLSS1. In beech, prediction curves of all yield classes are concave, reaching their maximum at a stand age of about 90 a. Prediction curves are also stratified, depending on yield class. However, curve order does not follow yield class order completeley, since the curve of yield class 1 lies above that of yield class 0. Additionally, distance between yield classes is irregular, with yield class 3 performing severely worse than yield class 2 and yield classes 1 and 0 performing almost the same. In spruce, shape of prediction curves largely resembles the shape of the estimated effect of the stand age variable smooth function shown in Figure 28: a steep slope for stand ages up to 30 a and a less steep but still positive slope for stand ages above 30 a. Curves also exhibit stratification based on yield class. However, distance between low performing yield classes is rather irregular compared to high performing ones, with yield classes 3 and 2 performing almost the same and yield class 4 performing only slightly worse than the former two.

Figure 33 shows basal area over top height, both observed values as well as predictions of

model GAMLSS1. For beech, yield class 3 performs worst throughout. For top heights up to 28 m, yield class 1 performs best, after which yield class −2 becomes the best performing yield class. For top heights above 32.5 m, curve order follows yield class order, whereas for lower top heights, curver order is rather erratic. In spruce, curve order for top heights up to 22 m is the inverse of yield class order, with yield class 4 performing best and yield class −2 performing worst. For top heights above 22 m, curve order is highly erratic, with yield class −2 never performing best.

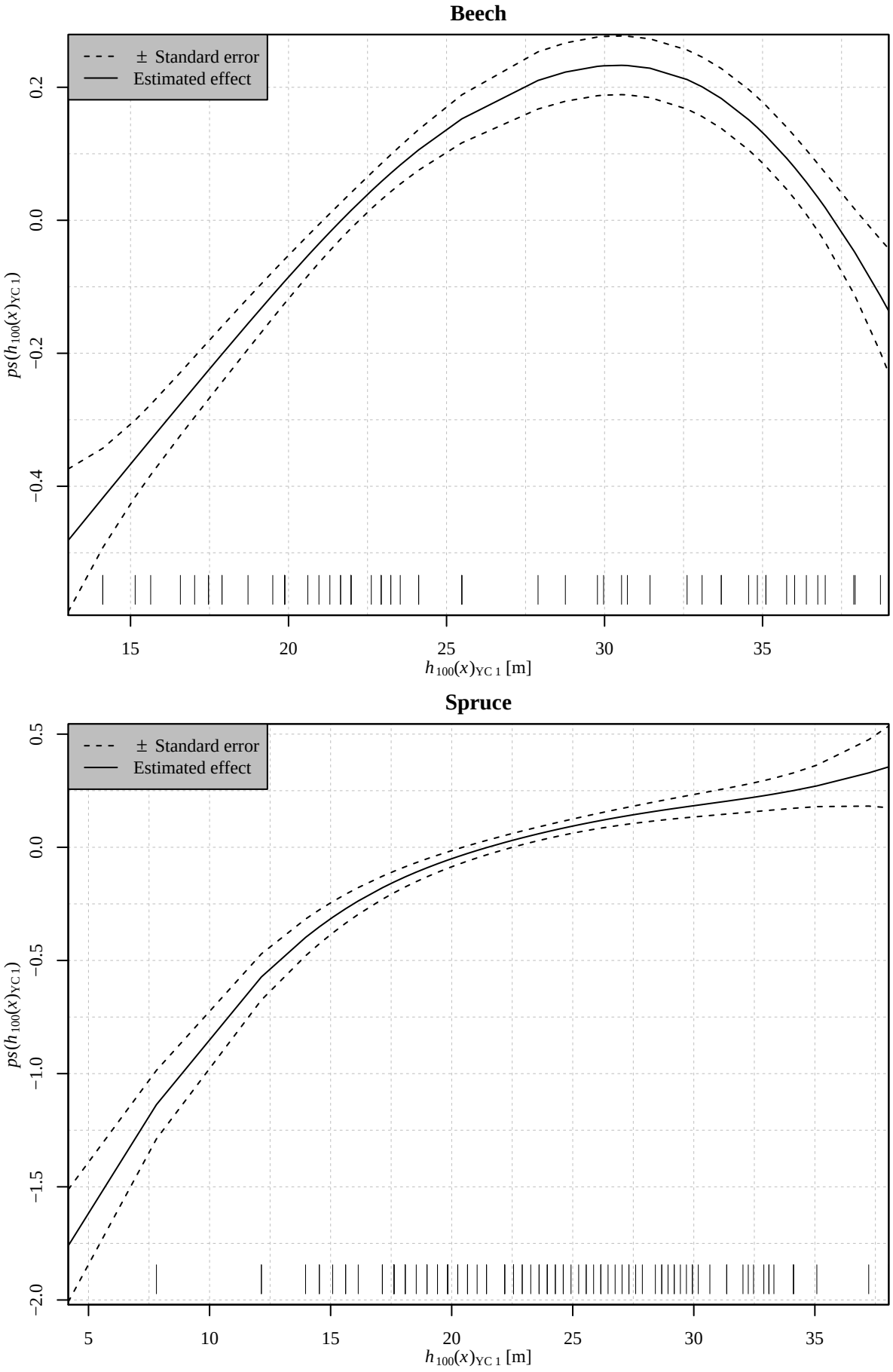

Figure 28: Estimated effect of the stand age variable smooth function ($ps(h_{100}(x)_{\mathrm{YC}\,1})$) over stand age variable ($h_{100}(x)_{\mathrm{YC}\,1}$) in model GAMLSS1 for beech (top) and spruce (bottom). Solid lines, dashed lines, and vertical bars have the same meaning as in Figure 23, namely: Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Note the different axis scaling in both plots.

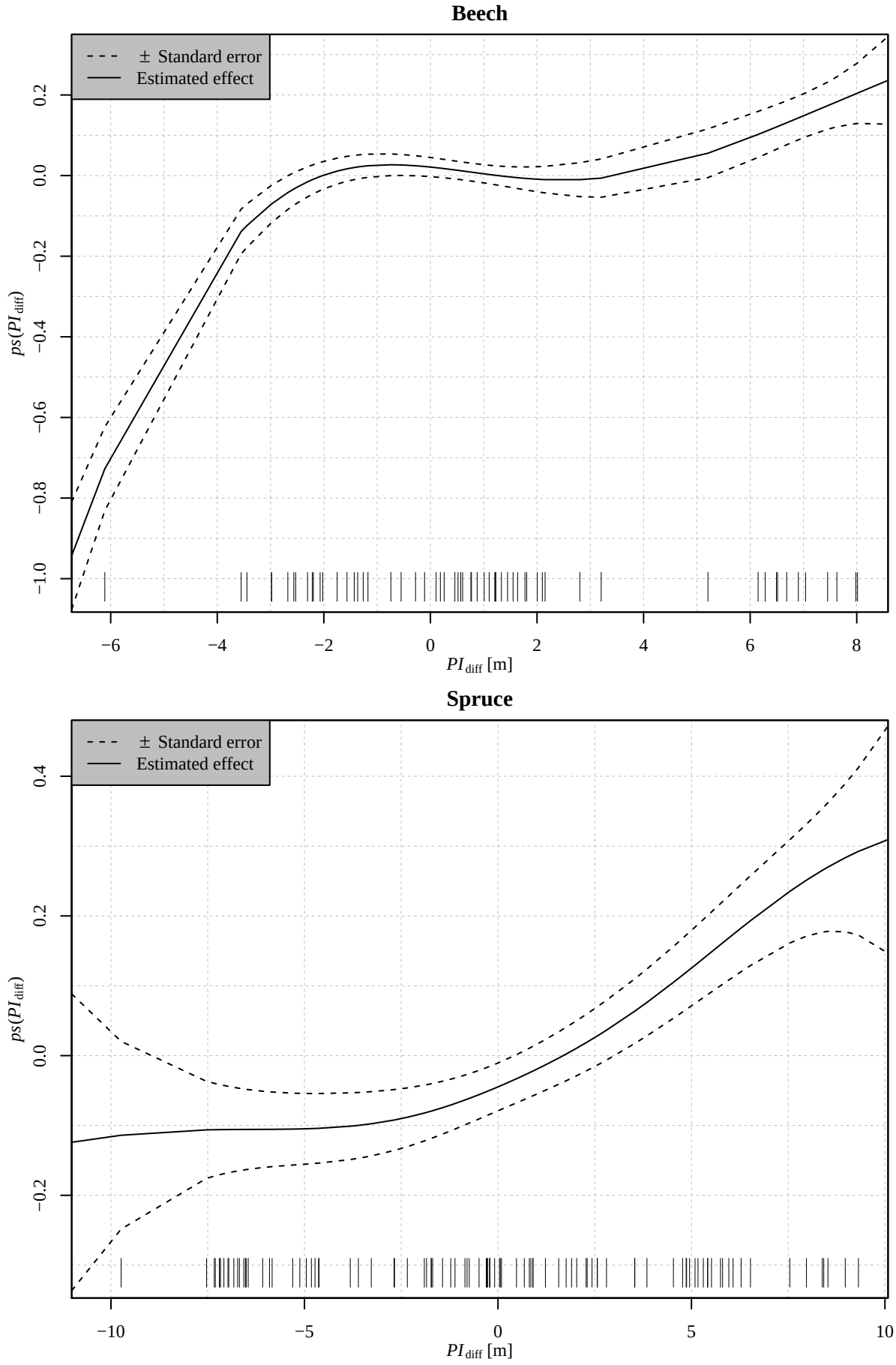

Figure 29: Estimated effect of the productivity index variable smooth function ($ps(PI_{diff})$) over productivity index variable (PI_{diff}) in model GAMLSS1 for beech (top) and spruce (bottom). Solid lines, dashed lines, and vertical bars have the same meaning as in Figure 28, namely: Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Note the different axis scaling in both plots.

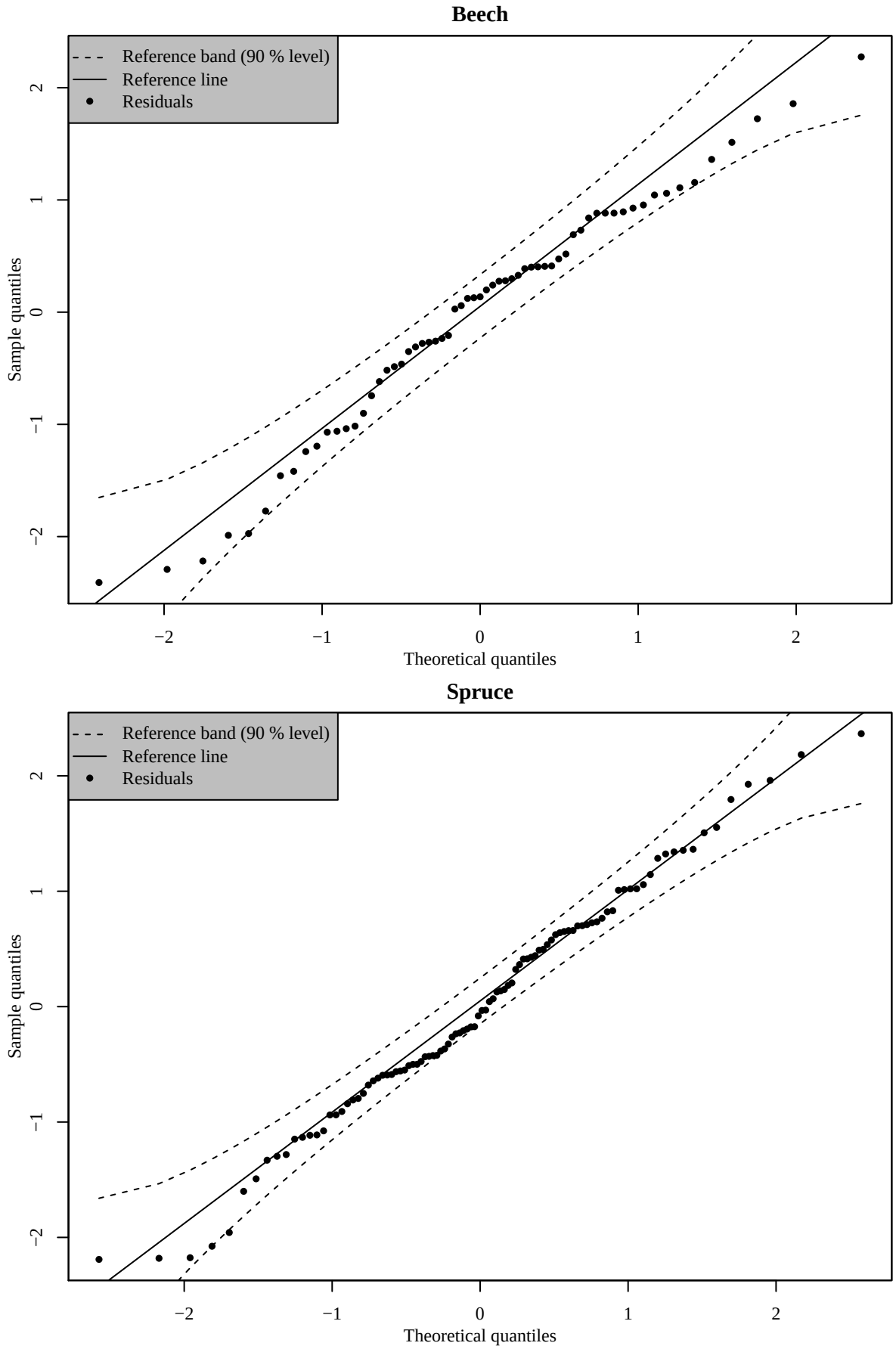

Figure 30: Quantile-quantile plot of the normalized quantile residuals of model GAMLSS1 for beech (top) and spruce (bottom). Solid lines are reference lines. Black dots represent residuals. Dashed lines mark reference bands between the 0.05 and 0.95 quantiles of predictions (90 % level). For a definition of normalized quantile residuals see Dunn & Smyth (1996). Reference bands are based on the standard errors of the order statistics of an independent random sample from the standard normal distribution (Fox 2016). Note the different axis scaling in both plots.

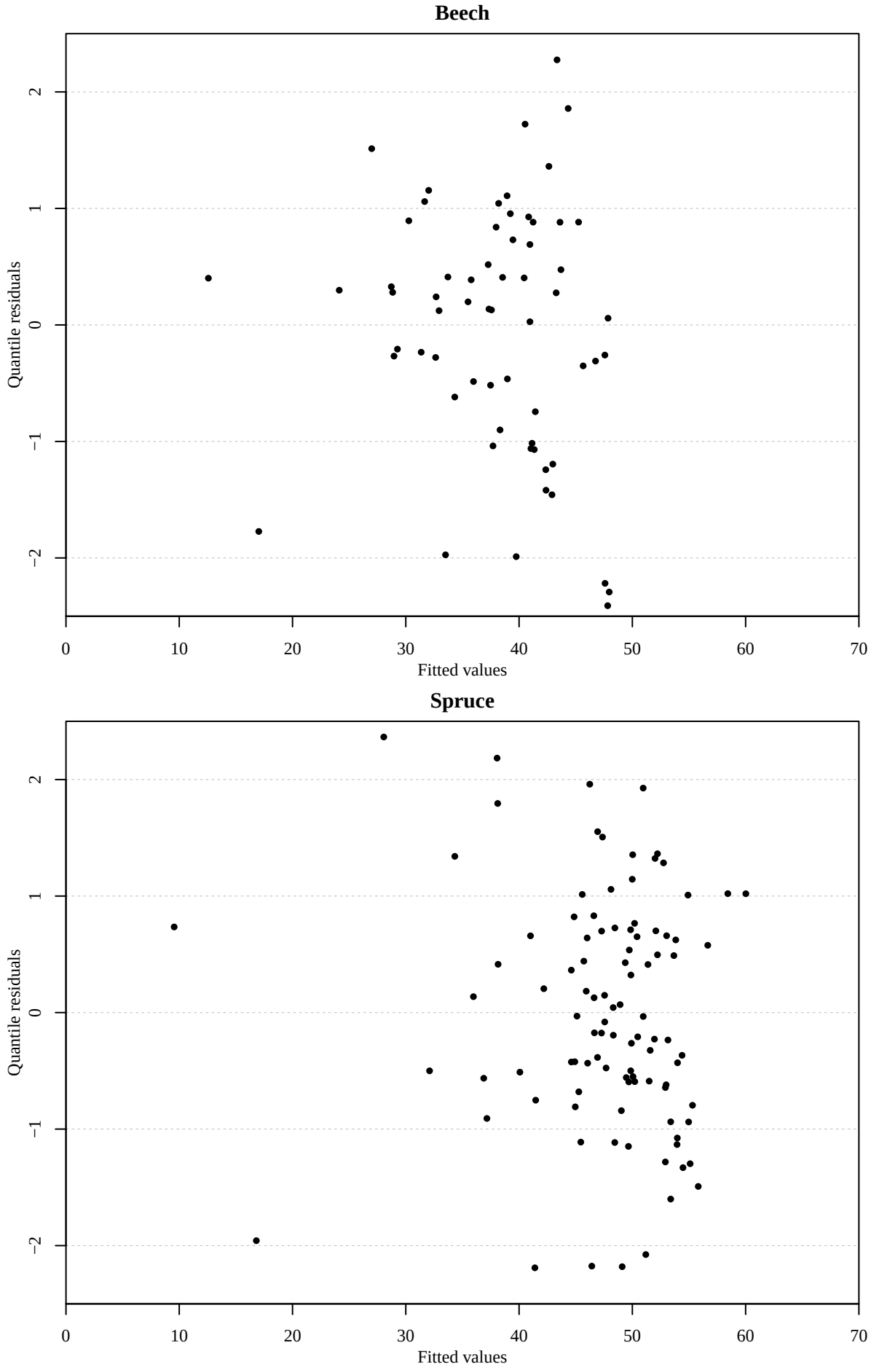

Figure 31: Normalized quantile residuals (Quantile residuals) over fitted values of model GAMLSS1 for beech (top) and spruce (bottom). For a definition of normalized quantile residuals see Dunn & Smyth (1996).

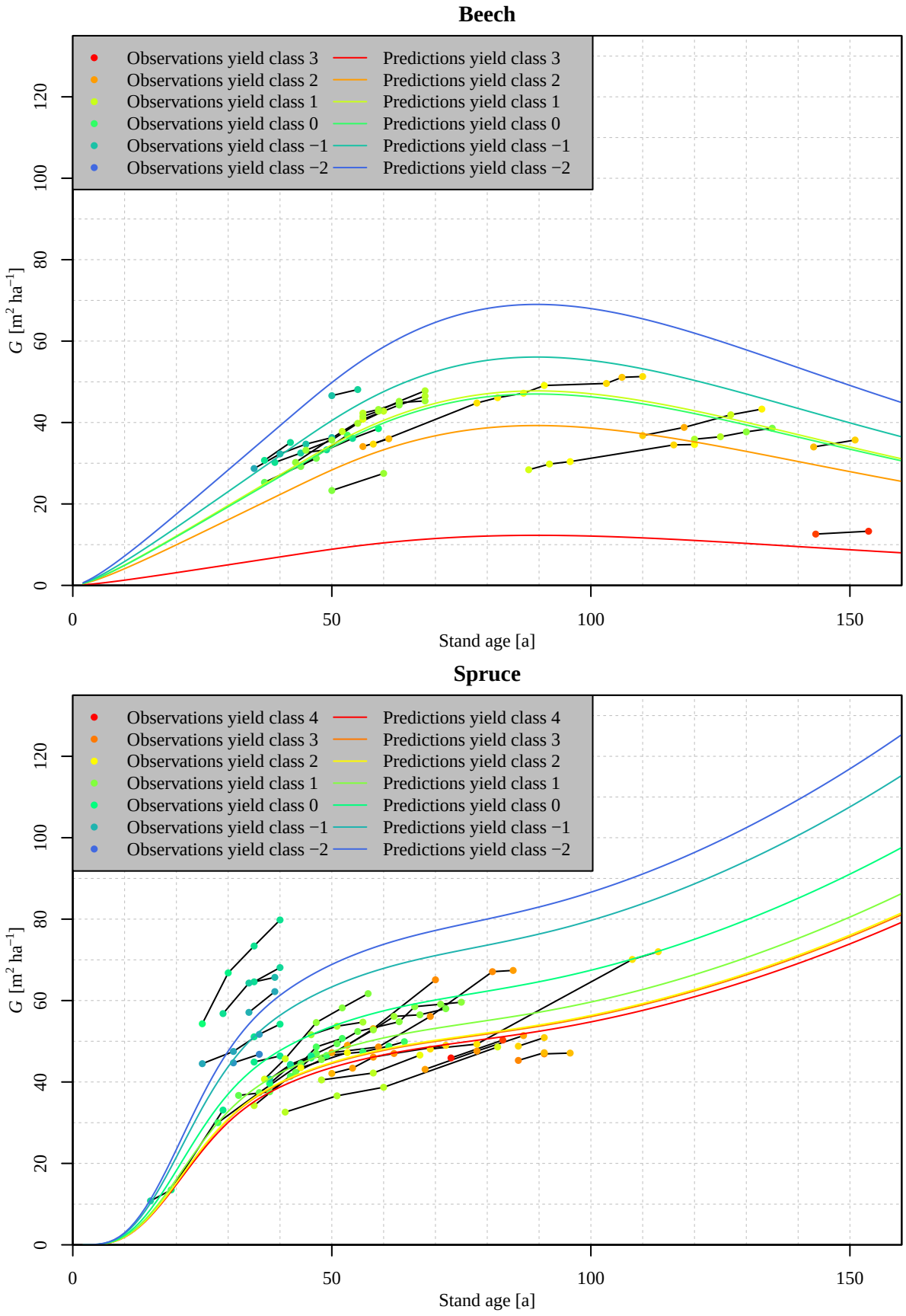

Figure 32: Basal area (G) over stand age of beech (top) and spruce (bottom). Colored lines represent predictions of model GAMLSS1. Dots, black lines, and colors have the same meaning as in Figure 27, namely: Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

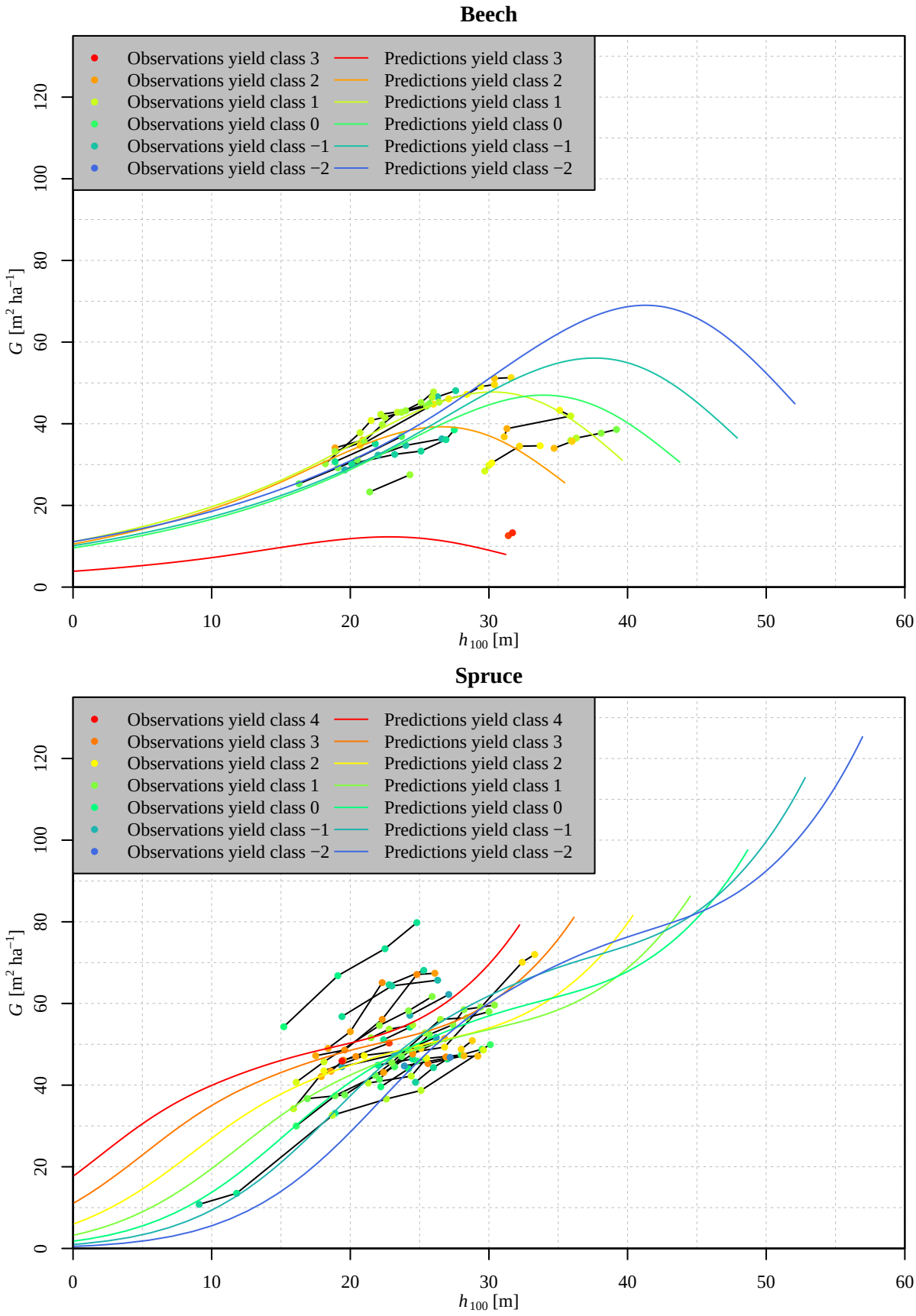

Figure 33: Basal area (G) over top height (h_{100}) of beech (top) and spruce (bottom). Colored lines, dots, black lines, and colors have the same meaning as in Figure 32, namely: Colored lines represent predictions of model GAMLSS1. Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

5.5 GAMLSS2

Figure 34 shows the estimated effect of the stand age variable smooth function over stand age variable in model GAMLSS2. The smooth function uses unconstrained P-splines as its basis. For both species, the curve's shape is similar to that in the corresponding plot of model GAMLSS1 (cp. Figure 28). For beech, the effect has a concave shape, reaching its maximum at a stand age variable value of approximately 31 m, which corresponds to an age of about 93 a. For spruce, initially the curve has a rather steep positive slope, between stand age variable values of 5 m to 12.5 m (which corresponds to an age range of 16 a to 25 a). Between 12.5 m to 20 m (25 a to 40 a), the slope gradually decreases, while remaining positive. At roughly 32.5 m (83 a) the slope begins to increase again for the remainder of the plot.

Figure 35 shows the estimated linear effect of the productivity index variable over productivity index variable in model GAMLSS2. In both species the effect is a monotone increasing line with its root at approximately 0 m.

Figure 36 shows the quantile-quantile plots of model GAMLSS2 for both species. In beech, 2 residuals deviate outside the 90 % reference band, whereas in spruce 7 residuals do. In either species, deviations from the reference line occur in the lower quantiles only, while the residuals of higher quantiles adhere fairly closely to it.

Figure 37 depicts the normalized quantile residuals over fitted values of model GAMLSS2. In both species, residual variance appears to be constant over the range of fitted values depicted.

Figure 38 shows basal area over stand age, both observations as well as predictions of model GAMLSS2. In both species, the general shape of the prediction curve is similar to the corresponding plot of model GAMLSS1 (cp. Figure 32). For beech, curves have a concave shape, each reaching their maximum at approximately 90 a. For spruce, curves have a rather steep slope between ages 15 a to 25 a. Between ages 25 a to 40 a the slope gradually decreases, while remaining positive, until at approximately 90 a slope gradually increases again. As in model GAMLSS1, prediction curves are stratified depending on yield class. A major difference to GAMLSS1, however, is that for both species, curve order at all times is the same as yield class order, meaning that any yield class shows higher predicted basal area than the next worse yield class.

Figure 39 depicts basal area over top height, both observations as well as predictions of model GAMLSS2. In beech, prediction curves have generally a concave shape, each reaching its maximum at a different top height. For top heights up to approximately 22.5 m, curve order is the inverse of yield class order, with yield class 3 performing best and yield class -2 performing worst. Between top heights of 22.5 m to 36 m, curve order switches, so that for top heights above 36 m, curve order is the same as yield class order, with yield class -2 performing best and yield class 3 performing worst. In spruce, prediction curves have a slanted wave-like form. Curve order generally is the inverse of yield class order, with predictions of adjacent yield classes

becoming almost identical between top heights of 13 m to 43 m and yield class 4 always outperforming yield class -2.

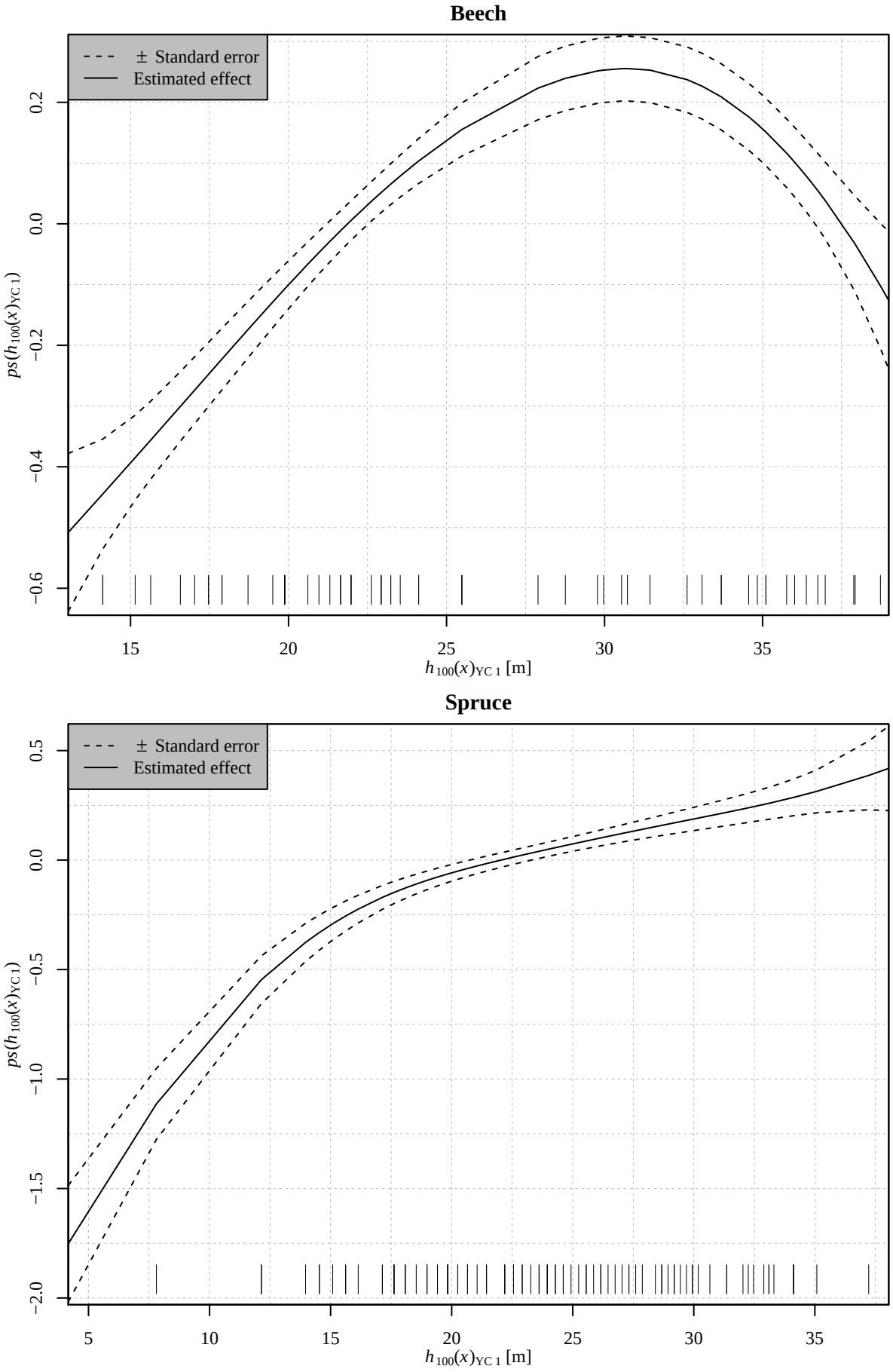

Figure 34: Estimated effect of the stand age variable smooth function ($ps(h_{100}(x)_{YC\,1})$) over stand age variable ($h_{100}(x)_{YC\,1}$) in model GAMLSS2 for beech (top) and spruce (bottom). Solid lines, dashed lines, and vertical bars have the same meaning as in Figure 29, namely: Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Note the different axis scaling in both plots.

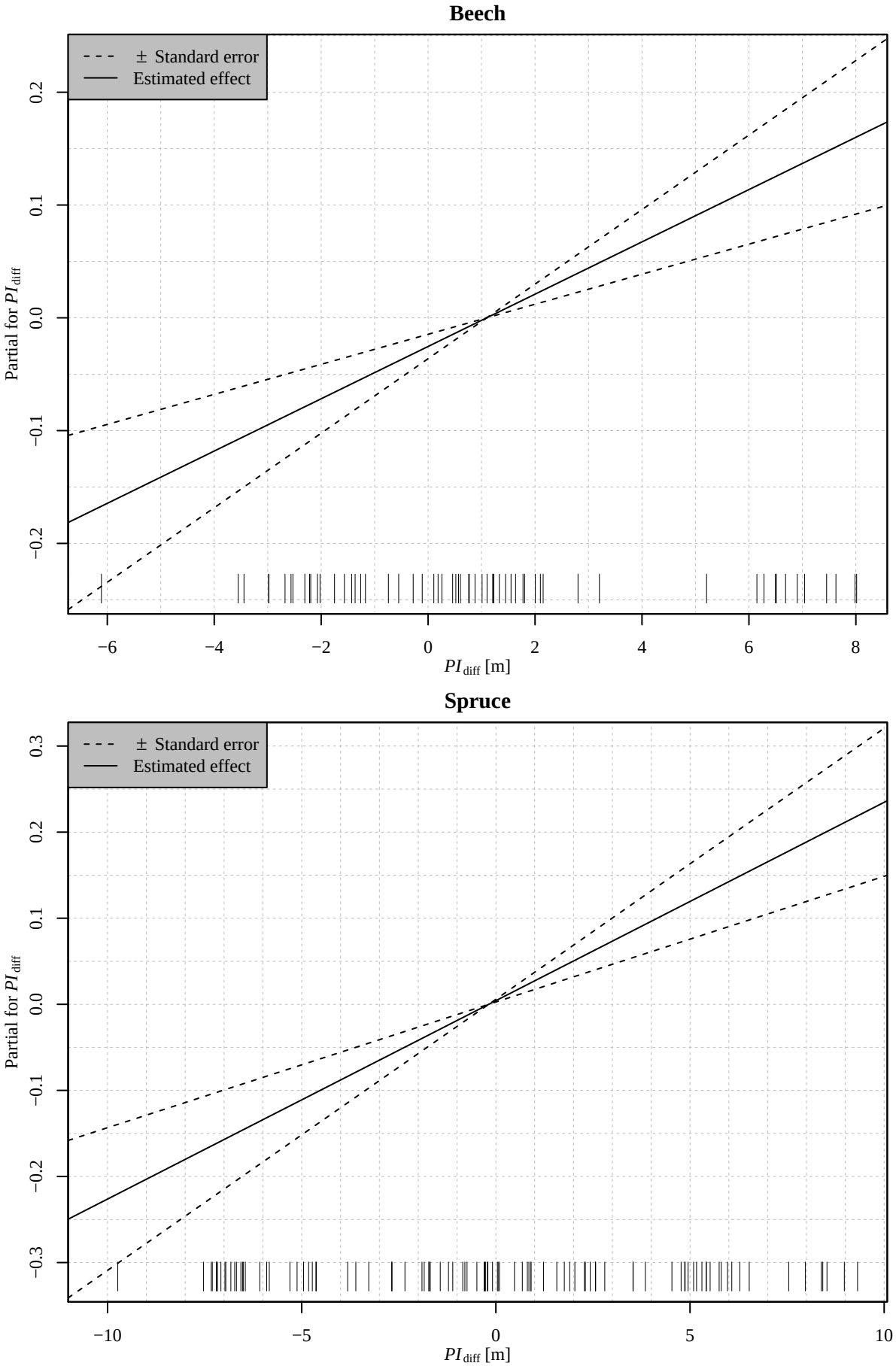

Figure 35: Estimated effect of the parametric productivity index variable term (Partial for PI_{diff}) over productivity index variable (PI_{diff}) in model GAMLSS2 for beech (top) and spruce (bottom). Solid lines, dashed lines, and vertical bars have the same meaning as in Figure 34, namely: Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Note the different axis scaling in both plots.

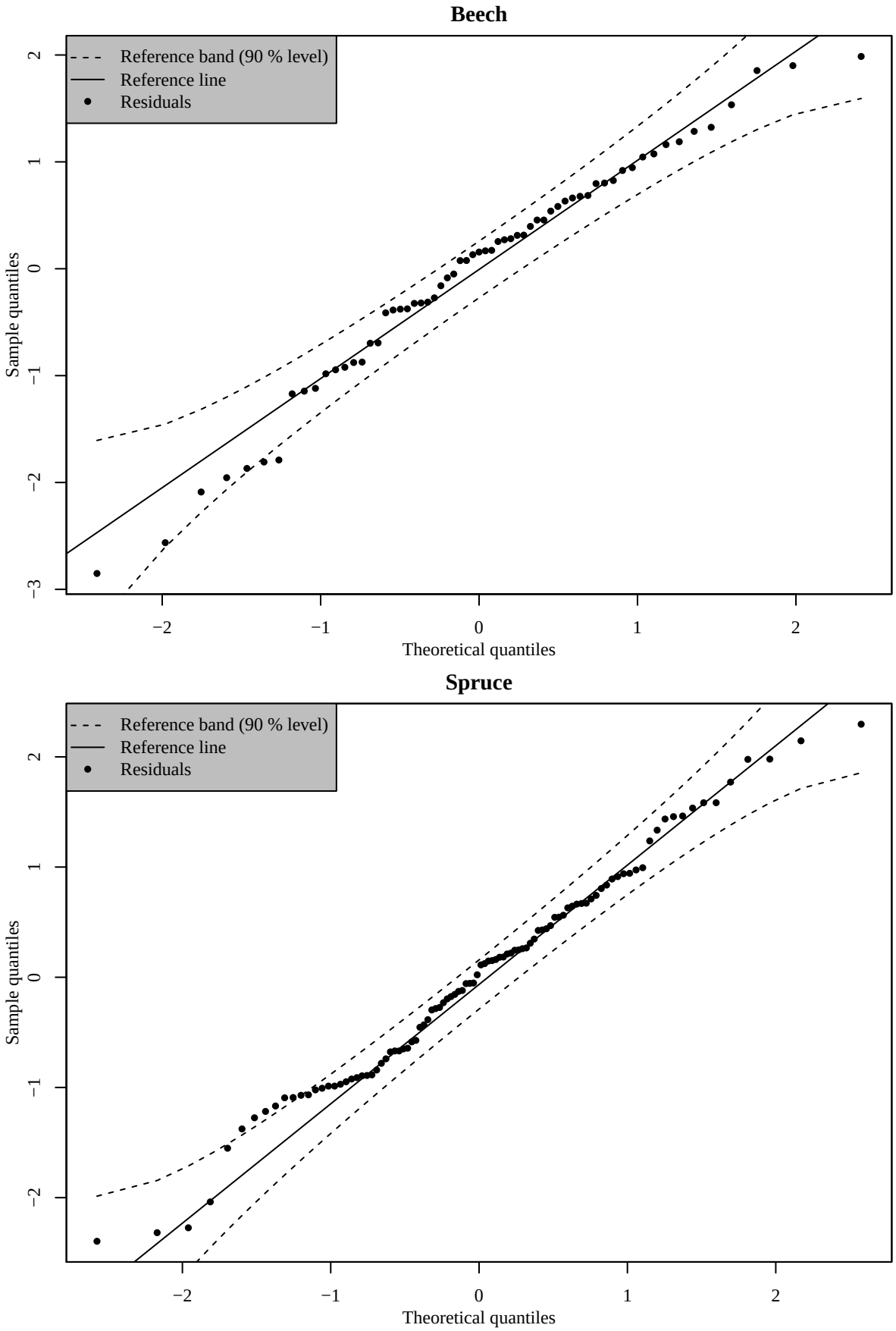

Figure 36: Quantile-quantile plot of the normalized quantile residuals of model GAMLSS2 for beech (top) and spruce (bottom). Solid lines, black dots, and dashed lines have the same meaning as in Figure 30, namely: Solid lines are reference lines. Black dots represent residuals. Dashed lines mark reference bands between the 0.05 and 0.95 quantiles of predictions (90 % level). For a definition of normalized quantile residuals see Dunn & Smyth (1996). Reference bands are based on the standard errors of the order statistics of an independent random sample from the standard normal distribution (Fox 2016). Note the different axis scaling in both plots.

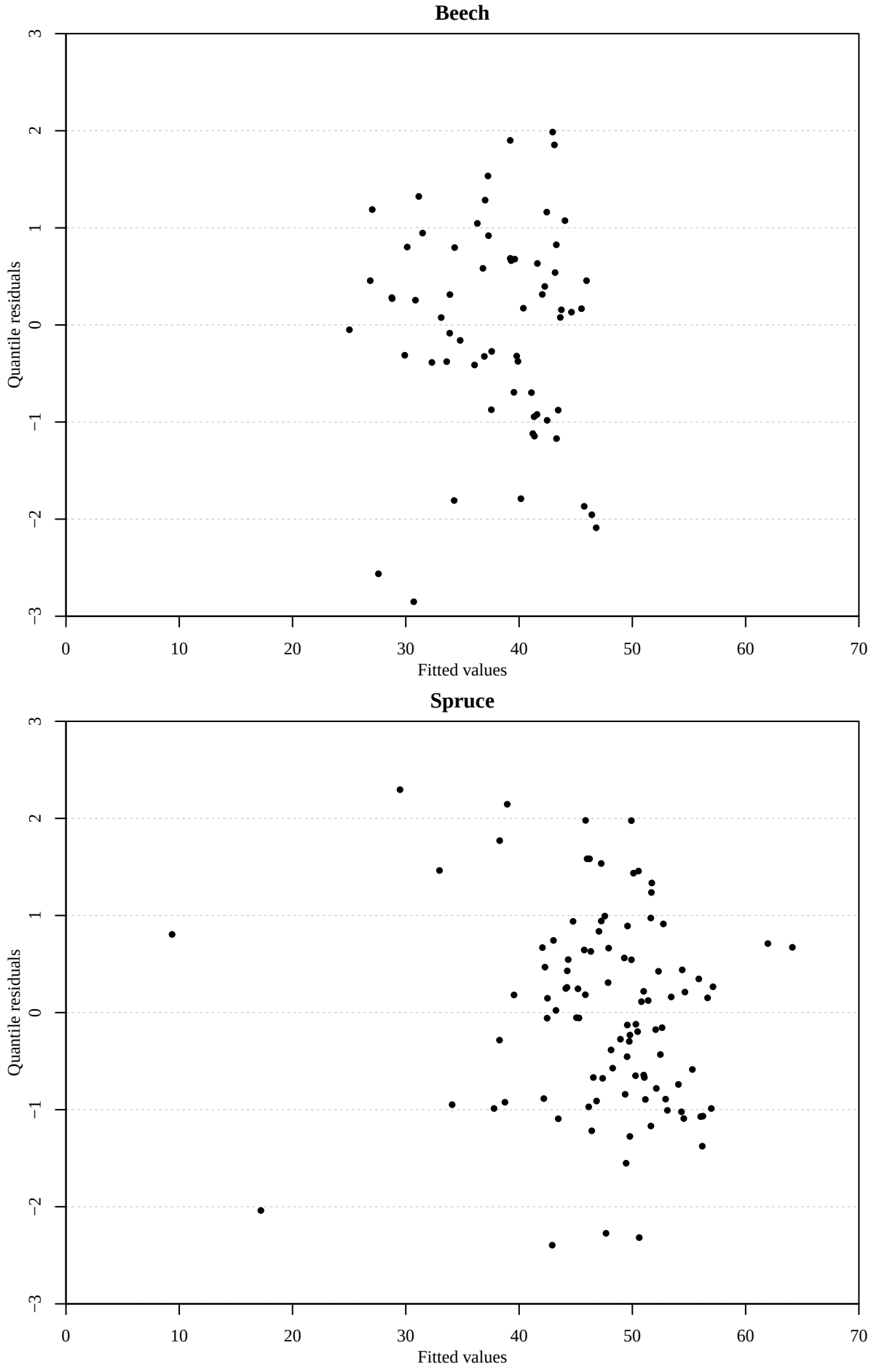

Figure 37: Normalized quantile residuals (Quantile residuals) over fitted values of model GAMLSS2 for beech (top) and spruce (bottom). For a definition of normalized quantile residuals see Dunn & Smyth (1996).

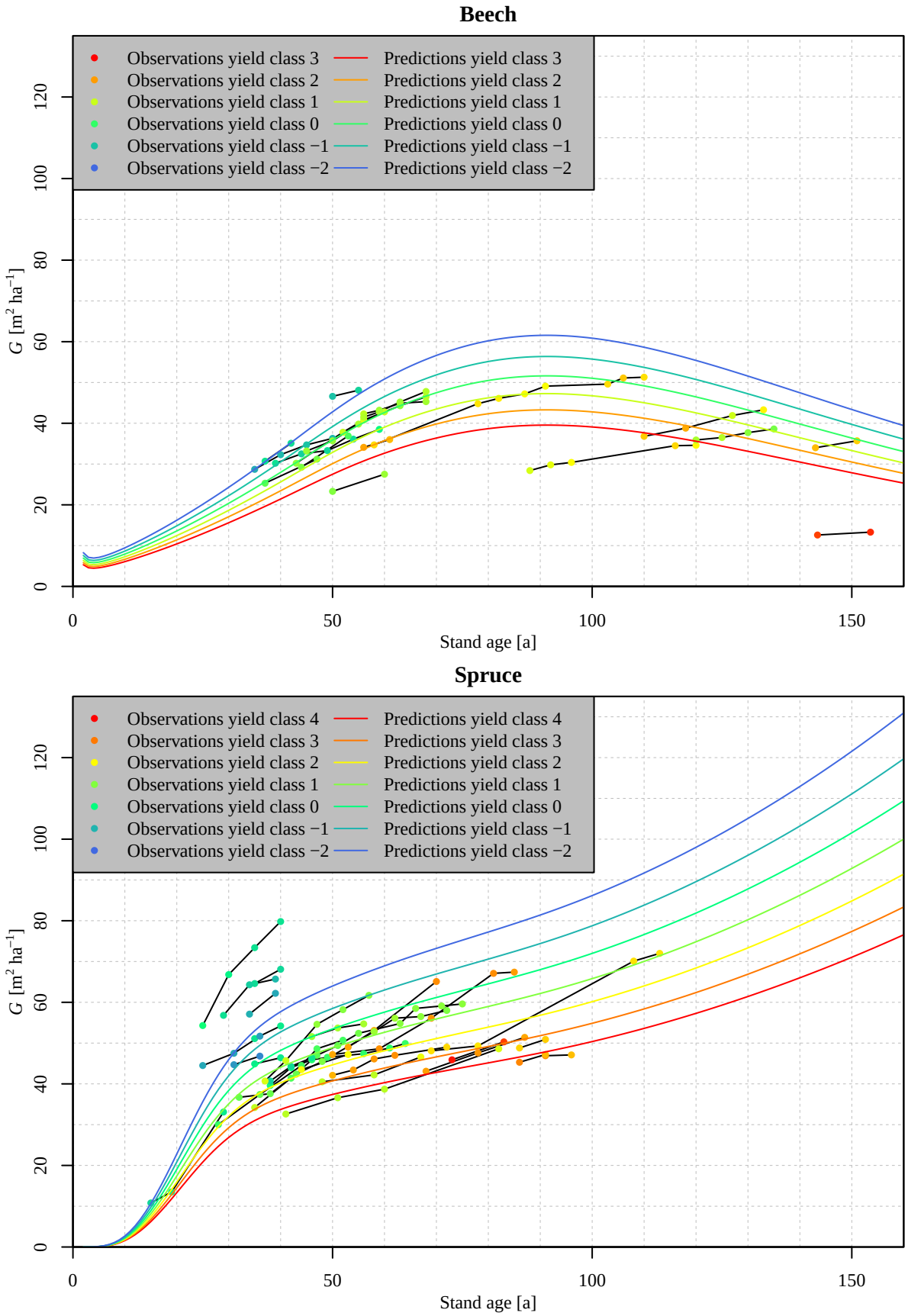

Figure 38: Basal area (G) over stand age of beech (top) and spruce (bottom). Colored lines represent predictions of model GAMLSS2. Dots, black lines, and colors have the same meaning as in Figure 33, namely: Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

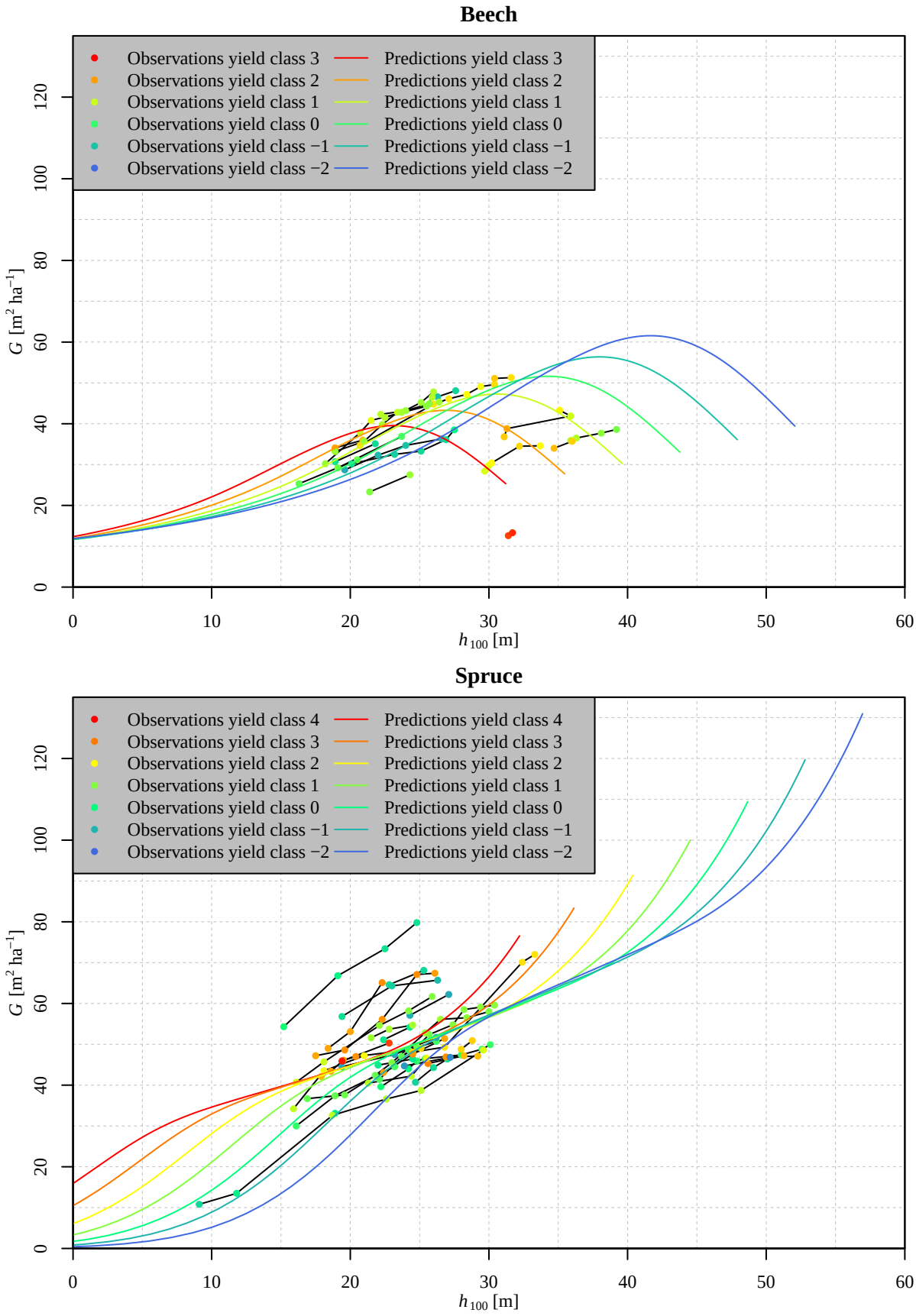

Figure 39: Basal area (G) over top height (h_{100}) of beech (top) and spruce (bottom). Colored lines, dots, black lines, and colors have the same meaning as in Figure 38, namely: Colored lines represent predictions of model GAMLSS2. Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

5.6 GAMLSS3

Figure 40 shows the estimated effect of the stand age variable smooth function over stand age variable in model GAMLSS3. The smooth function uses P-splines constrained to be monotone increasing as its basis. In beech, the effect is asymptotic: it initially increases before levelling off at about 27.5 m, which corresponds to a stand age of approximately 76 a. In spruce, the effect initially shows a steep slope, which gradually decreases between 12.5 m to 17.5 m (which corresponds to stand ages 25 a to 34 a) but remains positive afterwards.

Figure 41 shows the estimated linear effect of the productivity index variable over productivity index variable in model GAMLSS3. In both species, the effect is a monotone increasing line with its root at approximately 0 m.

Figure 42 shows the quantile-quantile plots of model GAMLSS3. In beech, one residual deviates outside the 90 % reference band, whereas in spruce, 5 residuals do. In both species, the outliers belong to lower quantiles.

Figure 43 depicts the normalized quantile residuals over fitted values of model GAMLSS3. In beech, residual variance appears to increase with fitted value, whereas in spruce, residual variance appears to be constant over the range of fitted values depicted.

Figure 44 shows basal area over stand age, both observations as well as predictions of model GAMLSS3. In beech, prediction curves are asymptotic: they initially increase before levelling off at about 80 a. In spruce, predicted basal area increases throughout the range of stand ages depicted, showing a very steep increase up to 30 a, a moderate increase between 30 a to 90 a, and a steep increase for ages above 90 a. In both species, prediction curves are stratified depending on yield class, with curve order being the same as yield class order, i.e., better yield classes outperforming worse ones.

Figure 45 shows basal area over top height, both observations as well as predictions of model GAMLSS3. In beech, prediction curves are asymptotic: they initially increase before levelling off at different top heights, depending on yield class. Prediction curves again exhibit stratification depending on yield class. For top heights up to 20 m, curve order is the inverse of yield class order, with yield class 3 performing best and yield class -2 performing worst. Between top heights of 20 m to 34 m, curve order switches due to worse yield classes levelling off at lower top heights than better yield classes. For top heights above 34 m, curve order is the same as yield class order, with yield class -2 performing best and yield class 3 performing worst. In spruce, prediction curves increase throughout the whole range of top heights depicted, while showing stratification based on yield class. Curve order is generally the inverse of yield class order, with yield class 4 always performing better than yield class -2. For top heights between 10 m to 40 m, curves of adjacent yield classes approach each other very closely, leading to almost identical basal area predictions for these yield classes.

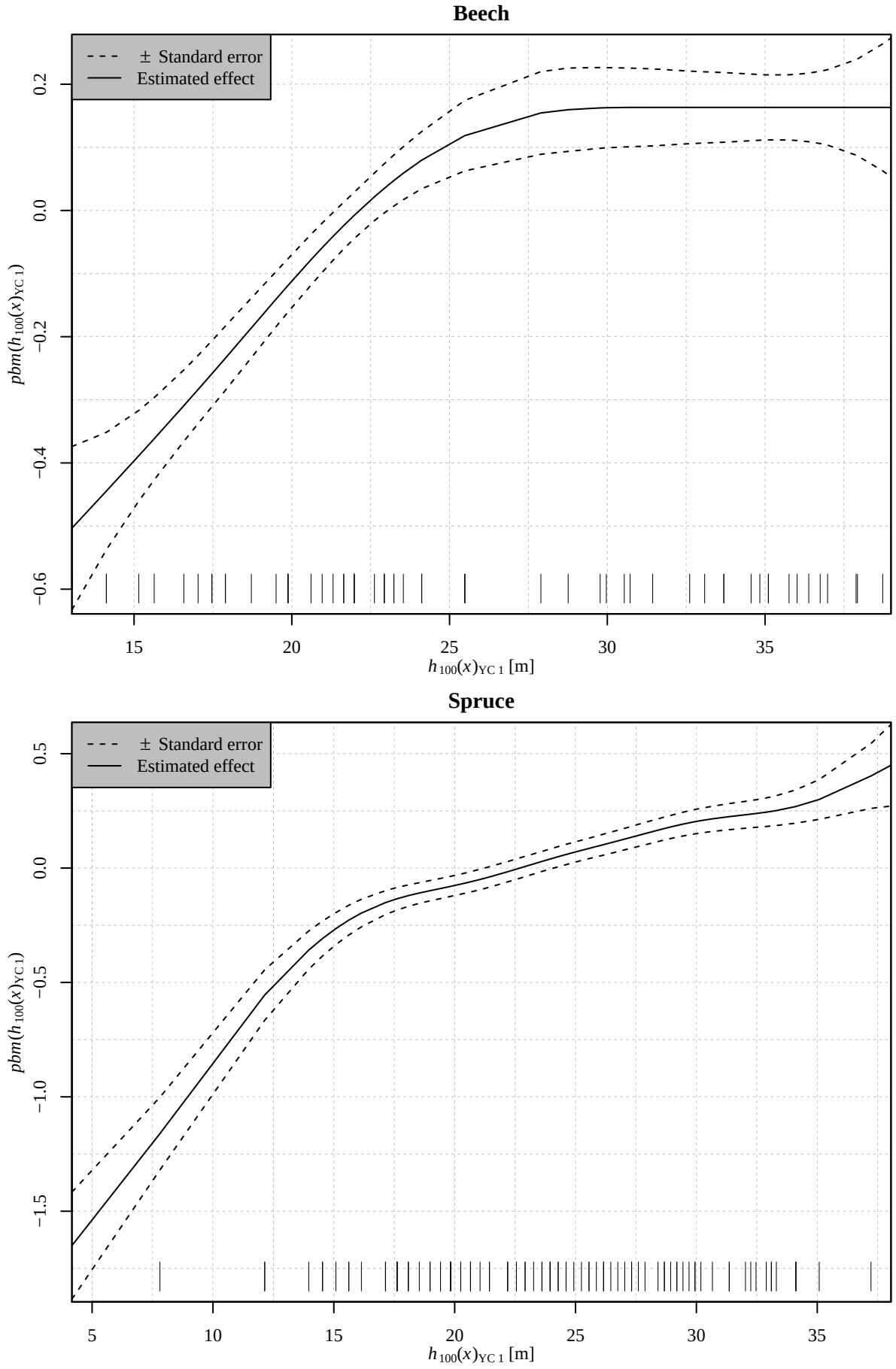

Figure 40: Estimated effect of the stand age variable smooth function ($pbm(h_{100}(x)_{\text{YC 1}})$) over stand age variable ($h_{100}(x)_{\text{YC 1}}$) in model GAMLSS3 for beech (top) and spruce (bottom). Solid lines, dashed lines, and vertical bars have the same meaning as in Figure 35, namely: Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Note the different axis scaling in both plots.

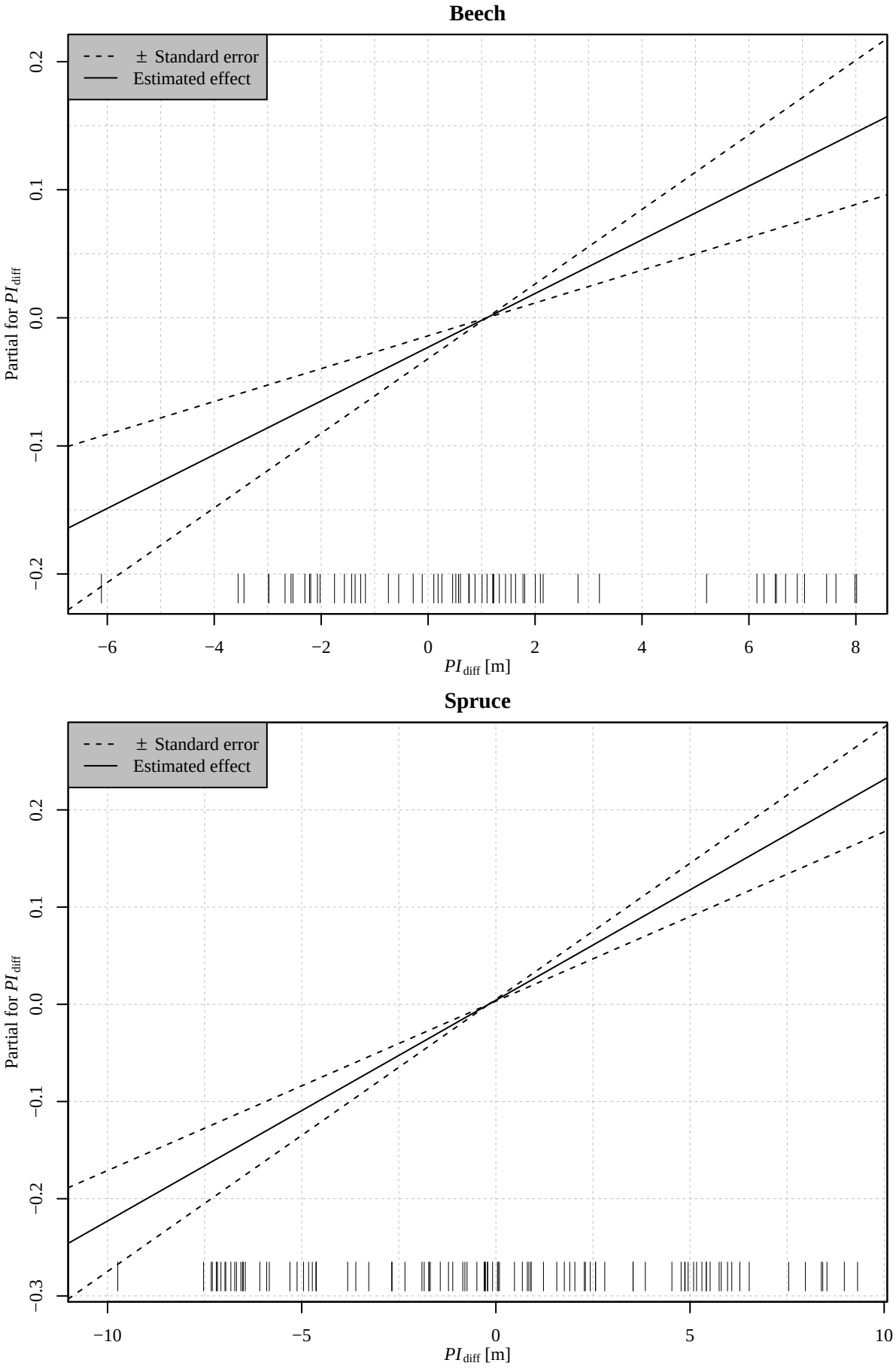

Figure 41: Estimated effect of the parametric productivity index variable term (Partial for PI_{diff}) over productivity index variable (PI_{diff}) in model GAMLSS3 for beech (top) and spruce (bottom). Solid lines, dashed lines, and vertical bars have the same meaning as in Figure 40, namely: Solid lines mark estimates. Dashed lines mark confidence bands of 2 standard errors width. Vertical bars mark observed values. Note the different axis scaling in both plots.

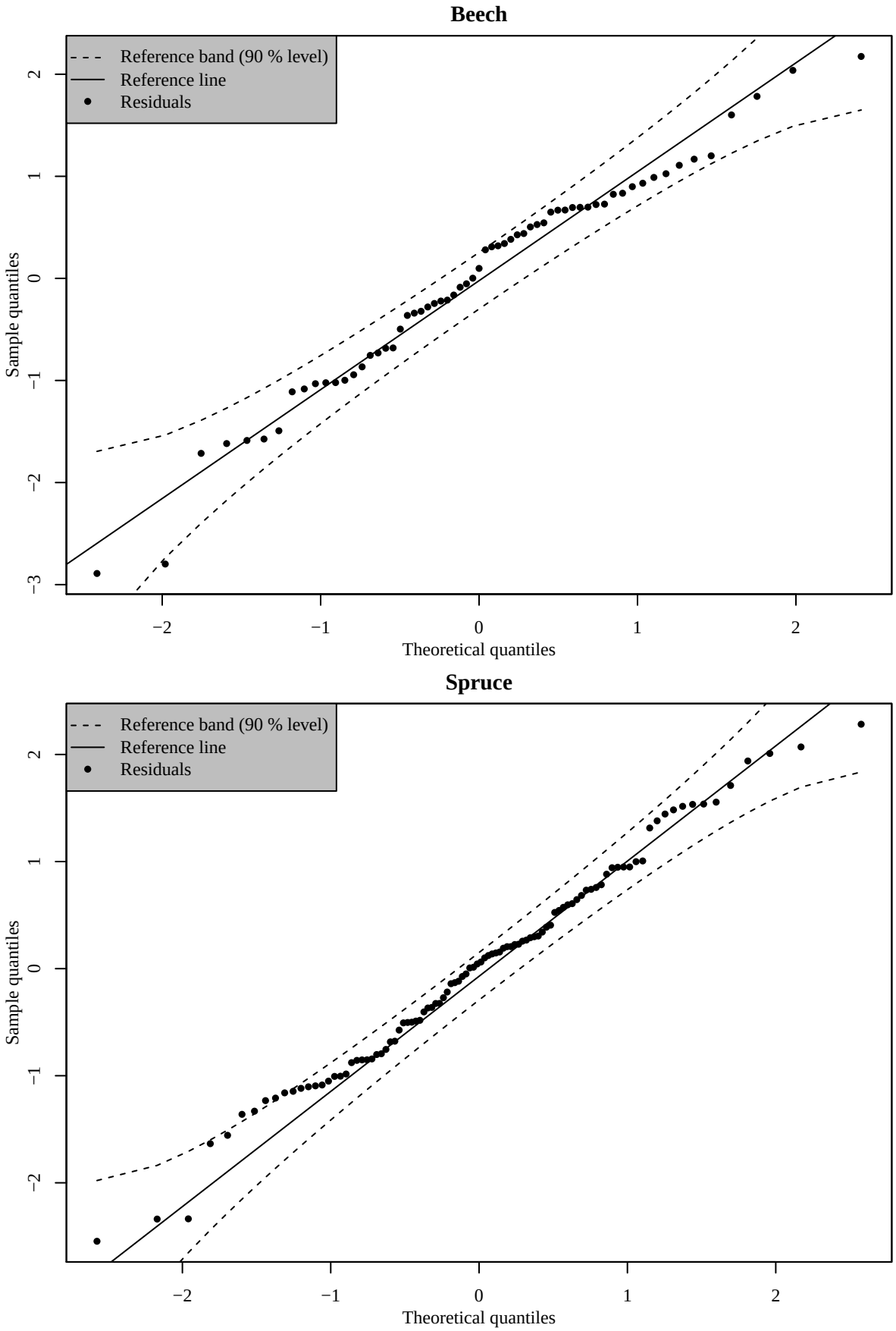

Figure 42: Quantile-quantile plot of the normalized quantile residuals of model GAMLSS3 for beech (top) and spruce (bottom). Solid lines, black dots, and dashed lines have the same meaning as in Figure 36, namely: Solid lines are reference lines. Black dots represent residuals. Dashed lines mark reference bands between the 0.05 and 0.95 quantiles of predictions (90 % level). For a definition of normalized quantile residuals see Dunn & Smyth (1996). Reference bands are based on the standard errors of the order statistics of an independent random sample from the standard normal distribution (Fox 2016). Note the different axis scaling in both plots.

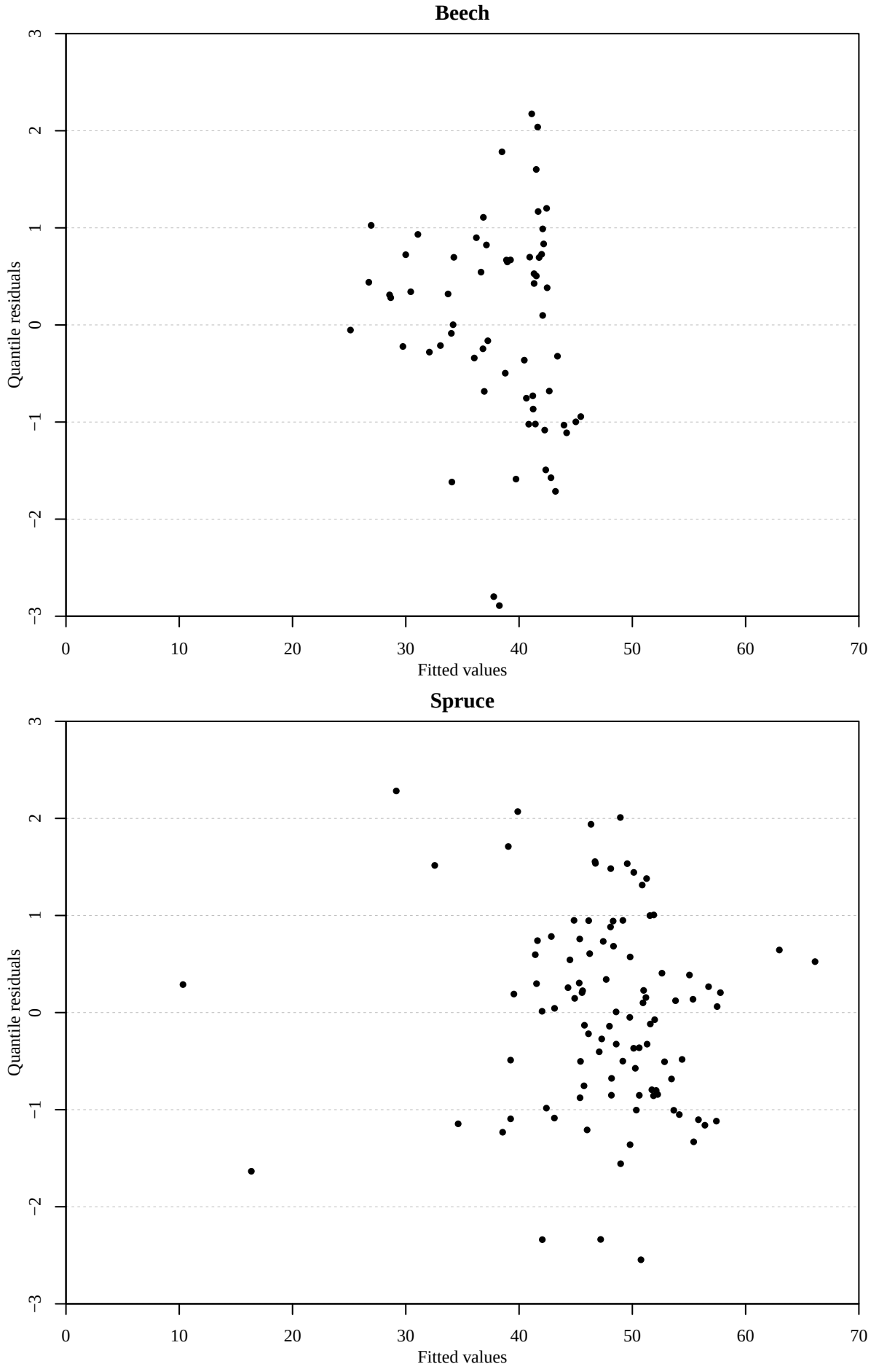

Figure 43: Normalized quantile residuals (Quantile residuals) over fitted values of model GAMLSS3 for beech (top) and spruce (bottom). For a definition of normalized quantile residuals see Dunn & Smyth (1996).

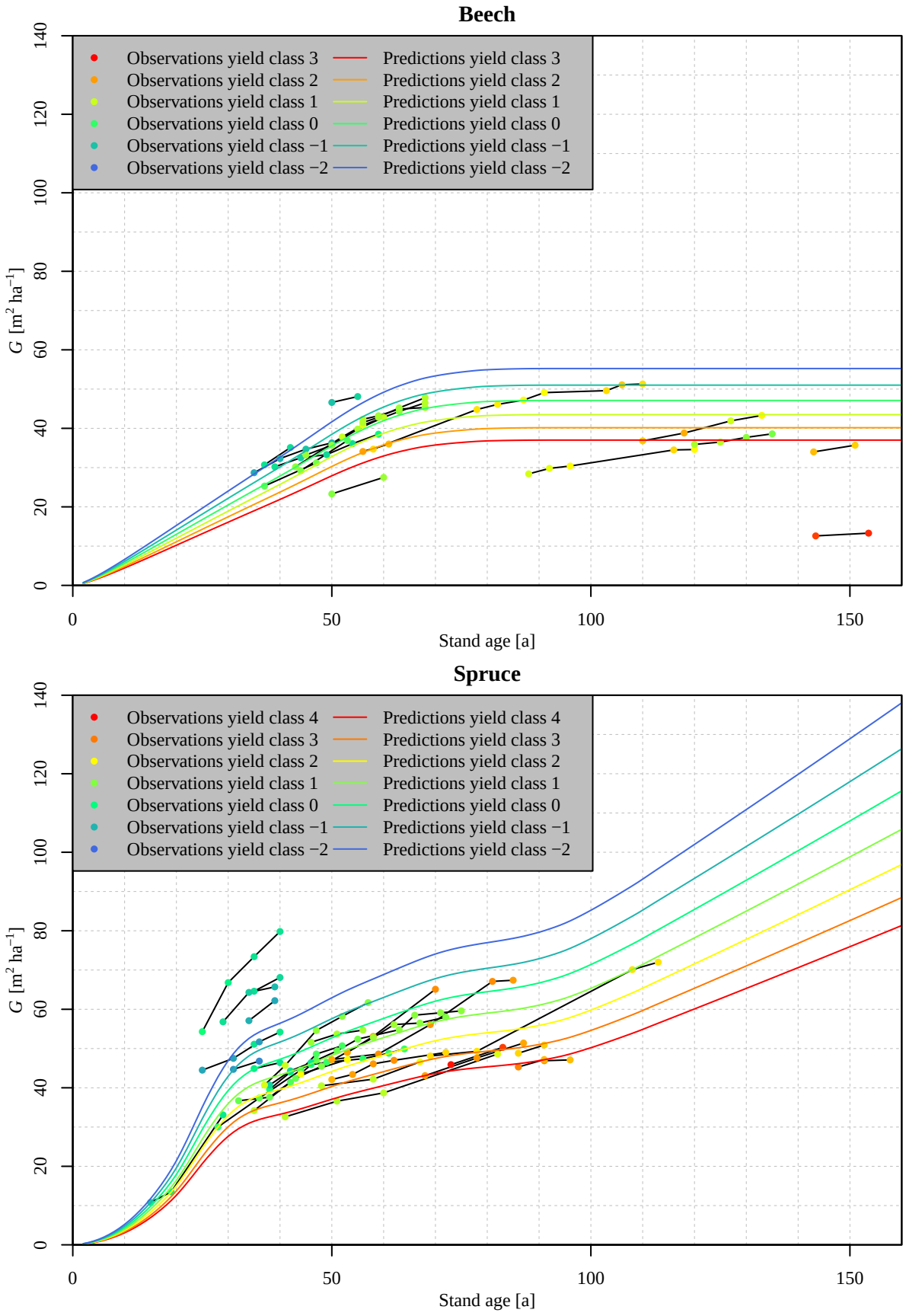

Figure 44: Basal area (G) over stand age of beech (top) and spruce (bottom). Colored lines represent predictions of model GAMLSS3. Dots, black lines, and colors have the same meaning as in Figure 39, namely: Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

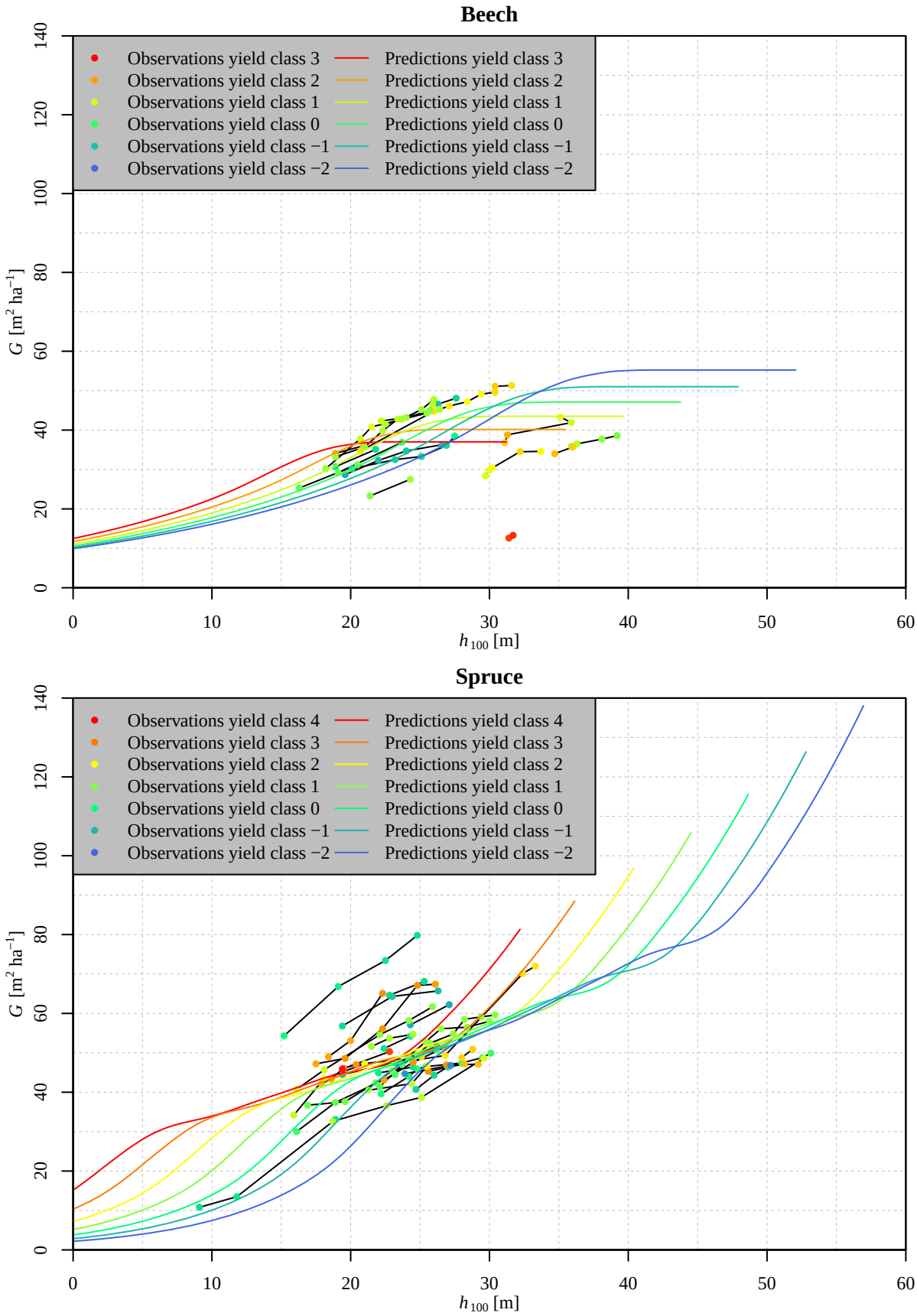

Figure 45: Basal area (G) over top height (h_{100}) of beech (top) and spruce (bottom). Colored lines, dots, black lines, and colors have the same meaning as in Figure 44, namely: Colored lines represent predictions of model GAMLSS3. Each dot represents one observation. Black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

6 Discussion

6.1 Data selection

In his definition of "maximum basal area", Assmann (1970) points out that maximum basal area is only achieved in stands which have not actively been thinned. In this context, "active thinning" means a reduction of stand density exceeding natural density-dependent mortality, also known as "natural thinning" (Society of American Foresters 1958) or "self-thinning" (Röhrig 1992). Self-thinning, by definition, is a naturally occurring phenomenon. However, the cause of thinning, be it natural inter-plant competition or human influence, is of no relevance to the present study since the end result is the same: a reduction of stand density. What is relevant is the actual extent or rate of thinning, since a thinning rate below or above the self-thinning rate will lead to the stand's basal area being above or below its natural maximum, respectively. Thus, data selection was based on the following rationale: a stand is considered to have maximum basal area, as long as its thinning rate is roughly equal to the self-thinning rate of the respective species. This, however, requires knowledge of the self-thinning modalities of forest stands in general and the species in question in particular.

According to Reineke (1933), the slope s of Equation (1) is a species-independent constant of -1.605. However, the universal applicability of Equation (1) has been called into question. As an alternative approach, Charru et al. (2012) proposed the use of a quadratic, rather than a linear logarithmic diameter term in the right hand side of Equation (1). A similar approach is taken by Schütz (2008), Schütz & Zingg (2010), and Zeide (1995), whose results suggested to add a quadratic logarithmic diameter term to the right hand side of Equation (1). Meyer (1938) found that for Ponderosa pine (*Pinus ponderosa* Douglas), the line reported by Reineke (1933) had to be changed to a slightly concave curve in order to fit observations.

Despite the apparent shortcomings of Equation (1), its general form has been upheld by other studies. Building on the findings of Drew & Flewelling (1979), VanderSchaaf (2010) and VanderSchaaf & Burkhart (2008) argued that Equation (1) is valid, but only during a specific phase of stand development. Similarly, Zeide (1985) suggested that Equation (1) is only applicable during the stage of full canopy closure. A different approach encountered in the literature is to use a species-specific slope rather than the constant of -1.605 reported by Reineke (1933) (Charru et al. 2012; MacKinney & Chaiken 1935; Pretzsch 2006; Pretzsch & Biber 2005; Río et al. 2001; Sterba 1987; Vacchiano et al. 2013; VanderSchaaf & Burkhart 2007; Vospernik & Sterba 2015; Zeide 1985, 1987). Table 10 lists several example slopes for beech and spruce stands undergoing self-thinning as reported in the literature. Pretzsch (2000, 2002) showed that the rule of Reineke (1933) may be considered a special case of the $-3/2$ power rule of Yoda et al. (1963), claiming that as long as the $-3/2$ power rule is valid, species-specific deviations from Reineke's constant of -1.605 are a result of species-specific diameter-biomass relationships.

Table 10: Species-specific values for the slope s of Equation (1) for beech and spruce as reported in the literature for stands undergoing self-thinning.

Source	Beech	Spruce
Charru et al. (2012)	−1.941	−1.878
Pretzsch (2006)	−1.873 to −1.723	−1.669 to −1.607
Pretzsch & Biber (2005)	−1.789	−1.664
Sterba (1987)		−1.737
Vacchiano et al. (2013)		−1.497
Vospernik & Sterba (2015)	−1.941	−1.753

Table 11 provides an overview of how the minimum, mean, and maximum values of the observed slopes are affected by the data selection mechanism. In both species, the maximum value of observed slopes is reduced noticeably due to the data selection mechanism: from 12.106 to −0.903 in beech and from 1.755 to −0.666 in spruce. In contrast, the minimum value of observed slopes remains unaffected by it in both species: for beech it remains at −2.027 while for spruce it remains at −1.958. Consequently, the mean value of observed slopes is reduced by the data selection mechanism in both species: from −0.568 to −1.327 in beech and from −0.649 to −1.322 in spruce. As can be seen from Figure 1, the data selection mechanism also led to a noticeable reduction in the number of observations and sample plots in both species: for beech the number of observations drops from 148 to 63, while the number of sample plots is halved from 36 to 18; for spruce the number of observations is reduced from 210 to 100, while the number of sample plots drops from 47 to 28.

Table 11: Observed minimum, mean, and maximum values for the slope s of Equation (1) for beech and spruce before application of the data selection mechanism (columns A) and after application of the data selection mechanism (columns B).

	Beech		Spruce	
	A	B	A	B
Minimum	−2.027	−2.027	−1.958	−1.958
Mean	−0.568	−1.327	−0.649	−1.322
Maximum	12.106	−0.903	1.755	−0.666

Both the lower and the upper slope thresholds used in the data selection mechanism (cp. Table 1) differ notably from the species-specific slopes of Equation (1) as reported in the literature (cp. Table 10). However, only the maximum of observed slopes is affected by the data selection

mechanism, whereas the minimum remains the same before and after its application in both data sets (cp. Table 11). The minimum is close to, albeit lower than, the slopes reported in the literature. In contrast, the maximum slope, both before and after application of the mechanism, notably exceeds even the highest slopes reported in the literature in both data sets. This suggests that the upper slope threshold may be too high in order to reliably exclude all observations of stands which were not (yet) subject to self-thinning, which in turn suggests that the predictions reported may underestimate the maximum basal area possible on sites similar to the sample plots. Lowering the upper thresholds to the highest values reported in the literature would have reduced the data set size for beech and spruce to 6 observations and 26 observations, respectively. Since using such small data sets would have severely hampered model fitting, lowering the upper slope thresholds was not an option.

6.2 Data sets

Although sample plots in both species cover a large range of geographical locations and altitudes (cp. Figures 2 and 3), neither of these properties could be harnessed as a predictor variable in either species, apparently due to the small data set sizes.

Considering Figure 4, the method of yield class classification yielded plausible results in both data sets: for a given top height, yield class gradually worsens as stand age increases. Both Equation (2), which was used for yield class classification of observed top heights, as well as Equation (59), which was used for test data generation, are based on the function developed by Nagel (1999). This ensured comparable yield class classification of training as well as test data.

The training data sets only partially cover the range of site characteristics encountered in forest stands in northwestern Germany. This fragmentariness of the training data sets reduces generalizability of model predictions insofar as predictions of stands that are poorly represented in the training data sets cannot be relied upon in the same manner as those for stands which are well mirrored in them can be. In the case of beech, stand ages below 30 a are completely absent, while in the case of spruce, only 2 observations of stand ages above 100 a are present (cp. Figure 4). Top heights below 15 m are completely absent from the beech data set and only represented by 2 observations in the spruce data set (cp. Figure 4). With respect to productivity, extreme sites are not as well represented in the training data sets as intermediate sites (cp. Figure 5), with sites below yield class 3 being completely absent from the beech data set. At the same time there is a strong negative correlation between stand age and absolute productivity index of stand in both data sets, meaning that young lowly productive stands as well as old highly productive stands are particularly underrepresented in the training data sets.

6.3 Models

The logarithm was chosen as the link function for the location parameter in all models in order to prevent unrealistic predictions of negative basal area.

The Gamma distribution was chosen for models GAM1, GAM2, and SCAM1 because past experience had shown it to be a suitable distribution for describing basal area. Similarly, the BCCG distribution was chosen for models GAMLSS1, GAMLSS2, and GAMLSS3 since it has successfully been used for describing unimodal distributions of site class (Albert & Schmidt 2009; Wördehoff 2016).

The confidence bands of the estimated smooth function effect in all models contained non-zero estimates (cp. Figures 10, 11, 16, 22, 28, 29, 34 and 40). This suggests that the stand age variable in all models and the productivity index variable in models GAM1 and GAMLSS1 is related to the response variable (Wood 2001).

Using the stand age variable as a predictor variable, rather than stand age directly, may reduce comparability of models across species, since its calculation (cp. Equation (3)) uses species-specific coefficients. Additionally, Equation (3) "reverses" the stand age ranges between species: in the beech data set stand ages span 118.6 a, while in the spruce data set they span 98 a; in contrast, the stand age variable spans 26 m in the former, but 33.9 m in the latter. While this result is scientifically sound (beech does not attain as high top heights as spruce does), it nevertheless obscures the underlying relations, namely that the beech data set coveres a wider range of stand ages than does the spruce one.

The sequences of models presented here, from GAM1 over GAM2 to SCAM1 and from GAMLSS1 over GAMLSS2 to GAMLSS3, may be seen as sequences of increasingly constrained models. GAM1 and GAMLSS1 contain the least constraints: the effects of stand age variable and productivity index variable on basal area are completely unrestricted and allowed to take on any shape. In models GAM2 and GAMLSS2, the productivity index variable is constrained to have a linear, rather than a smooth effect on basal area. Finally, in models SCAM1 and GAMLSS3 the stand age variable smooth function is also constrained, namely to be either increasing and concave or monontone increasing, respectively. The fact that model GAMLSS3 only imposes a monotone increasing constraint, rather than an increasing and concave constraint like model SCAM1, is due to the fact that no such constraint has been implemented in `R` package `gamlss` as of version 5.0.4. This imposition of increasingly strict constraints was motivated by the implausible results of the lesser restricted models. In model GAM1, yield class 1 in beech performs better than yield class 0 while yield class -2 in spruce performs worse than yield class 1 throughout all stand ages between 2 a to 160 a (cp. Figure 14). In model GAMLSS1, the same problem regarding predictions for beech persists (cp. Figure 32). Additionally, in both models the distance between yield classes 3 and 2 in beech and yield classes -1 and 0 in spruce is much larger than between other adjacent yield classes. Compared to the values reported in Schober

(1995) (moderate thinning), which are depicted in Figure 46, this irregularity in distance appears implausible. In models GAM2 and GAMLSS2, linearization of the productivity index variable eliminates these irregularities (cp. Figures 20 and 38). However, basal area predictions remain implausible, since in the case of beech, basal area is predicted to decrease after approximately 80 a in all yield classes, whereas in the case of spruce, basal area increases indefinitely in all yield classes. Both results are not in conjunction with values reported in Schober (1995) (moderate thinning), which suggest no decrease in basal area for beech at least up to 150 a for yield classes 1 to 4, and a decrease in basal area for yield classes 1 to 4 of spruce after 100 a the latest (cp. Figure 46). Similarly, Franz (1965) reports spruce stands to reach a maximum basal area at around 120 a. These remaining implausibilities were partly countered by the additional constraints imposed in models SCAM1 and GAMLSS3 (cp. Figures 26 and 44). Basal area predictions for beech show no decrease for stand ages up to 160 a, but rather an asymptotic behavior, reaching their maximum at around 67 a in the case of SCAM1 and around 80 a in the case of GAMLSS3. While neither of these values is in accordance with the basal area development reported by Schober (1995) (moderate thinning), the predictions for beech of models SCAM1 and GAMLSS3 nevertheless appear more plausible than those of models GAM2 and GAMLSS2, respectively. In the case of spruce, the difference between model SCAM1, which imposes both an increasing as well as a concavity constraint, and model GAMLSS3, which only imposes a monontone increasing constraint, becomes apparent. While in both models basal area predictions increase indefinitely, the increase is remarkably smaller in SCAM1, which predicts, at 160 a, a basal area of roughly $93\,m^2\,ha^{-1}$ for yield class -2 and of about $55\,m^2\,ha^{-1}$ for yield class 4, whereas GAMLSS3, for the same age and yield classes, predicts a basal area of approximately $138\,m^2\,ha^{-1}$ and $81\,m^2\,ha^{-1}$, respectively. In Figure 27, which depicts predicted basal area over top height for model SCAM1, in both species the order of prediction curves initially is the inverse of yield class order, but shifts during a certain range of top heights, so that finally it is the same as yield class order. The same is true for beech in model GAMLSS3 (cp. Figure 45). These shifts are due to the fact that prediction curves of low productivity yield classes level off or show a decrease in slope at lower top heights than do curves of higher productivity yield classes. The reason for this becomes apparent when considering Figure 47, which depicts the relationship between top height and stand age in the test data sets (on which model predictions were based): for a given top height, age increases as yield class worsens. Thus, the asymptotic behavior or the decrease in slope of the prediction curves occurs at lower top heights the worse a stand's yield class is, but still at approximately the same stand age in all yield classes for each species (cp. Figures 26 and 44). Similar shifts of curve order can be observed in the predictions of other models as well, owing to the same properties of the test data sets. All things considered, the chosen constraints improved plausibility of model predictions.

However, the more constrained models also generally achieve higher (i.e., worse) AIC scores than the less constrained models (cp. Table 9), with the notable exception of GAMLSS3 for spruce, which achieves the lowest AIC for that species. A reason for this discrepancy between plausibility and statistical power of the models may lie in the incompleteness of the training data

sets discussed above. At the same time, the GAMLSSs generally achieve a lower AIC than the correpsonding GAM or SCAM, with the exception of GAMLSS1 in spruce which achieves a slightly higher score than GAM1. This suggests that the increased uncertainty of GAMLSSs compared to GAMs or SCAMs is outweighed by their increased flexibility.

Since the training data sets only consist of pure even-aged stands, model predictions are also only valid for such stands. In order to acquire valid maximum basal area predictions for mixed and/or uneven-aged stands, the models would need to be refit using appropriate training data sets.

Since GAMLSSs directly estimate all parameters of the assumed probability distribution, they also offer the option of predicting specific distribution quantiles (Yee 2004). As shown by Wördehoff et al. (2014), this is particularly useful in the context of maximum basal area modeling insofar as models may be fit using non-maximum basal area observations, while predicting upper quantiles to obtain function envelopes, which may be considered a close approximation of true maximum basal area predictions. Since the training data sets in this study here were assumed to contain only maximum basal area observations after application of the data selection mechanism, this option was not examined. It nevertheless does warrant further research.

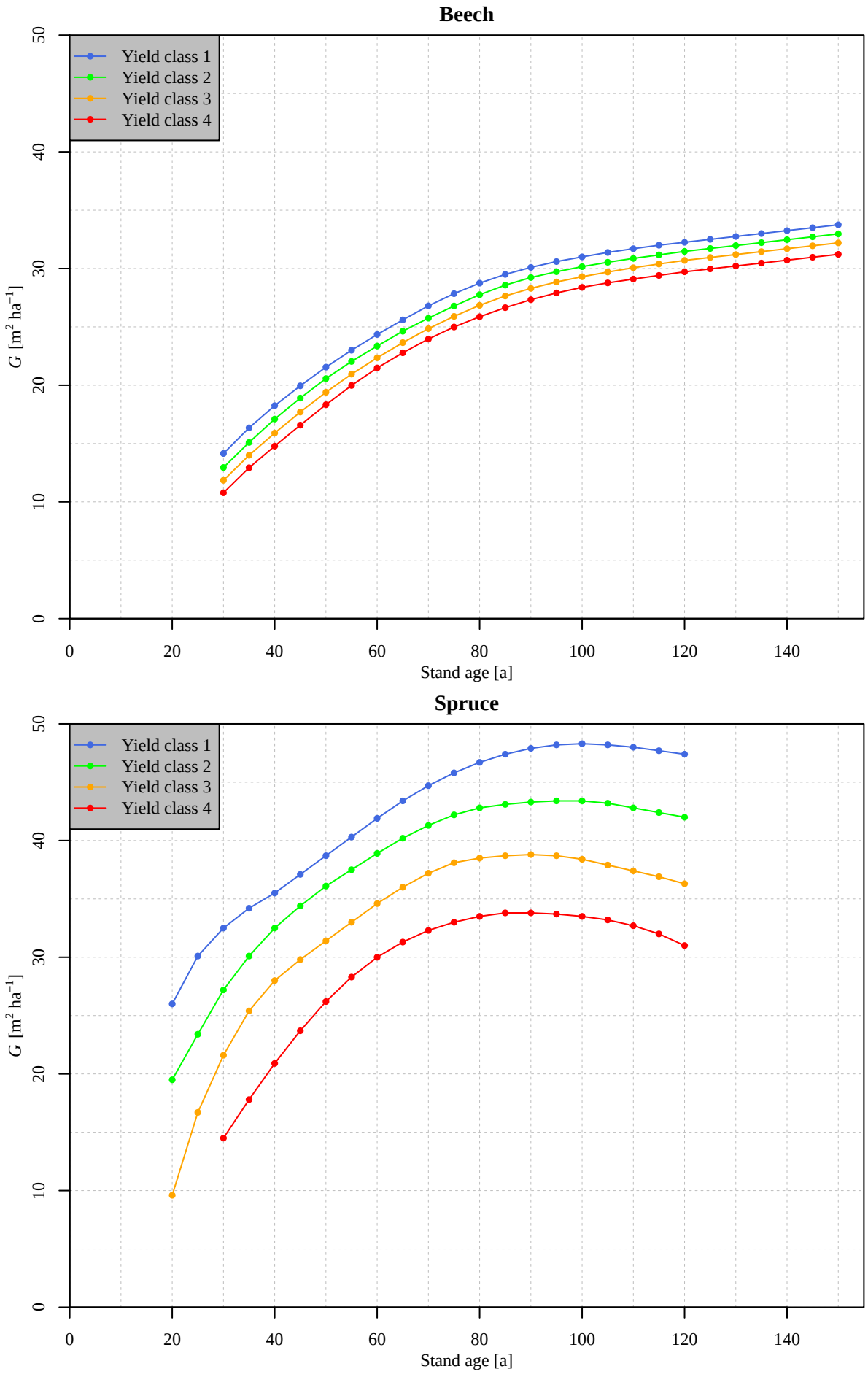

Figure 46: Basal area (G) over stand age of beech (top) and spruce (bottom) as reported in Schober (1995) (moderate thinning). Color signifies yield class.

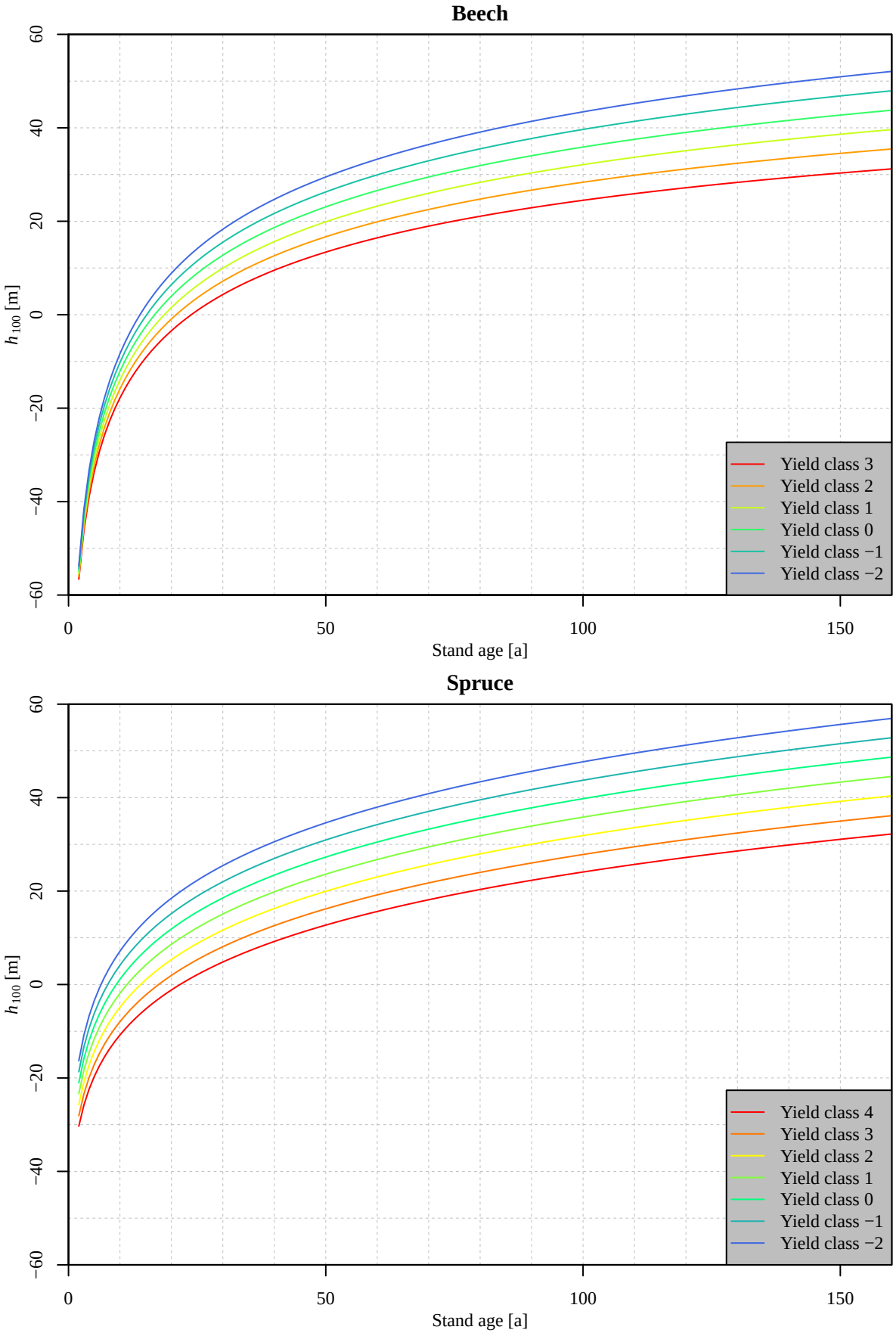

Figure 47: Top height (h_{100}) over stand age of beech (top) and spruce (bottom) in the test data. Test data was generated using Equation (59). Line color has the same meaning as in Figure 45, namely: Color signifies the (fractional) yield class of the test data, ranging from red (worst yield class observed in the training data) over yellow and green to blue (best yield class observed in the training data). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

6.4 Comparison to other approaches

While basal area modeling, both at individual tree level (e.g., Andreassen & Tomter (2003), Hein & Dhôte (2006), Jogiste (2000), Monserud & Sterba (1996), Nyström & Kexi (1997), Schröder et al. (2002), and Wimberly & Bare (1996)) as well as stand level (e.g., Castedo-Dorado et al. (2007), Chikumbo (2001), Chikumbo et al. (1999), and Eerikäinen & Maltamo (2003)), has been the subject of several studies, modeling of *maximum* basal area has not gathered as much attention. To the best knowledge of the author, sofar only two approaches have been made (Sterba (1975, 1981, 1987) and Wördehoff (2016)), none of which take differences in stand productivity into account. The method presented by Sterba results from applying the "competition-density rule" developed by Ando (1968), Ando et al. (1968), Kira et al. (1953), and Tadaki (1963) on mean diameter as suggested by Goulding (1972), rather than on mean volume. It allows prediction of maximum basal area using the equation

$$G_{max} = \frac{\pi}{16\ a_0\ b_0} {h_{100}}^{-(a_1+b_1)}, \tag{60}$$

where G_{max} is maximum basal area, a_0, a_1, b_0, b_1 are coefficients to be estimated, and h_{100} is top height. In order to use Equation (60) for predicting maximum basal area, the equation needs to be fit to maximum basal area observations so as to obtain valid estimates of $a_0, a_1, b_0,$ and b_1. Equation (60) uses only top height as an independent variable, thus offering no option to incorporate stand productivity as a predictor. As a consequence, predictions obtained from Equation (60) are strictly only valid for stands which are of approximately the same yield class as the observations used for fitting it. In order to obtain maximum basal area predictions for different yield classes, a new set of equation coefficients needs to be estimated, based on a data set representative of the respective yield class. Fitting of Equation (60) to the training data sets used in the present study was not successful. In order to nevertheless compare predictions of the most plausible models presented in this study (GAMLSS3 for beech and SCAM1 for spruce) with predictions of Equation (60), coefficient estimates reported by Döbbeler (2004) (Region Nordwest) and Wördehoff (2016) for beech and spruce, which are given in Tables 12 and 13, respectively, were used . The results are shown in Figure 48.

In the case of beech, Equation (60), particularly when using the Döbbeler (2004) estimates, predicts notably higher basal areas for lower top heights than does model SCAM1 for yield classes 3 to −2. This relation then switches with increasing top height, so that for top heights above 30 m and below 50 m, Equation (60) predicts lower basal areas than does model SCAM1 for yield class −2. Unlike those of model SCAM1, however, predictions of Equation (60) continue to increase at a top height of 60 m.

In the case of spruce, relations between predictions of Equation (60) and of model GAMLSS3 behave similarly to those described above. For lower top heights, the former predicts higher basal areas than does the latter for yield classes 2 to −2. For higher top heights, Equation (60) predicts lower basal areas than does GAMLSS3 for yield classes 3 to −2 (Wördehoff (2016) estimates) or for yield classes 0 to −2 (Döbbeler (2004) estimates).

Wördehoff (2016) used a GAMLSS (termed "GAMLSSW" in the following text) for modeling maximum basal area. The model assumes basal area to follow Box-Cox-Cole-Green distribution, with identity function, natural logarithm function, and identiy function as the link function for μ, σ, and ν, respectively. Similar to Equation (60), it only contains top height as a predictor variable. Like GAMLSS1 and GAMLSS2, it uses P-splines as the function basis for the smooth function. Figure 49 shows the predictions of GAMLSSW fit to the data sets used in the present study in comparison to predictions of models GAMLSS3 for beech and SCAM1 for spruce. GAMLSSW, due to its smooth function basis not being constrained, predicts increasing basal areas with increasing top height, showing no asymptotic behavior in either species. In the case of beech, the first third of the prediction curve somewhat resembles the curves of GAMLSS3 for the depicted yield classes, reaching a basal area minimax point of approximately $36\,\mathrm{m^2\,ha^{-1}}$ at a top height of roughly 27 m. For top heights above 27 m, the slope of the GAMLSSW curve gradually increases again, leading to a basal area prediction of $80\,\mathrm{m^2\,ha^{-1}}$ at a top height of 27 m. The overall shape of the prediction curve of GAMLSSW for spruce is similar to that for beech, albeit more compressed and having an inflection point, rather than a minimax point, at a basal area of around $49\,\mathrm{m^2\,ha^{-1}}$ and a top height of roughly 23 m. Up to a top height of 35 m (beech) and 28 m (spruce), GAMLSSW predictions are largely lower than those of GAMLSS3 for beech or SCAM1 for spruce. For higher top heights, GAMLSSW predictions gradually exceed all predictions for all depicted yield classes of the latter two models.

While models SCAM1 and GAMLSS3 do suffer from certain inconsistencies discussed above, they nevertheless offer greater flexibility in terms of incorporating stand productivity as a predictor of maximum basal area than do Equation (60) or GAMLSSW.

Table 12: Estimates of the coefficients of Equation (60) for beech as reported in Döbbeler (2004) (Region Nordwest) and Wördehoff (2016).

Source	a_0	a_1	b_0	b_1
Döbbeler (2004)	$1.0829 \cdot 10^{-7}$	1.5374	8.3652	−1.7365
Wördehoff (2016)	$2.616551 \cdot 10^{-7}$	1.368151	6.496417	−1.731867

Table 13: Estimates of the coefficients of Equation (60) for spruce as reported in Döbbeler (2004) (Region Nordwest) and Wördehoff (2016).

Source	a_0	a_1	b_0	b_1
Döbbeler (2004)	$1.28745 \cdot 10^{-6}$	0.7148	1.2842	−1.1914
Wördehoff (2016)	$4.913256 \cdot 10^{-6}$	0.4394706	0.3716977	−0.9097641

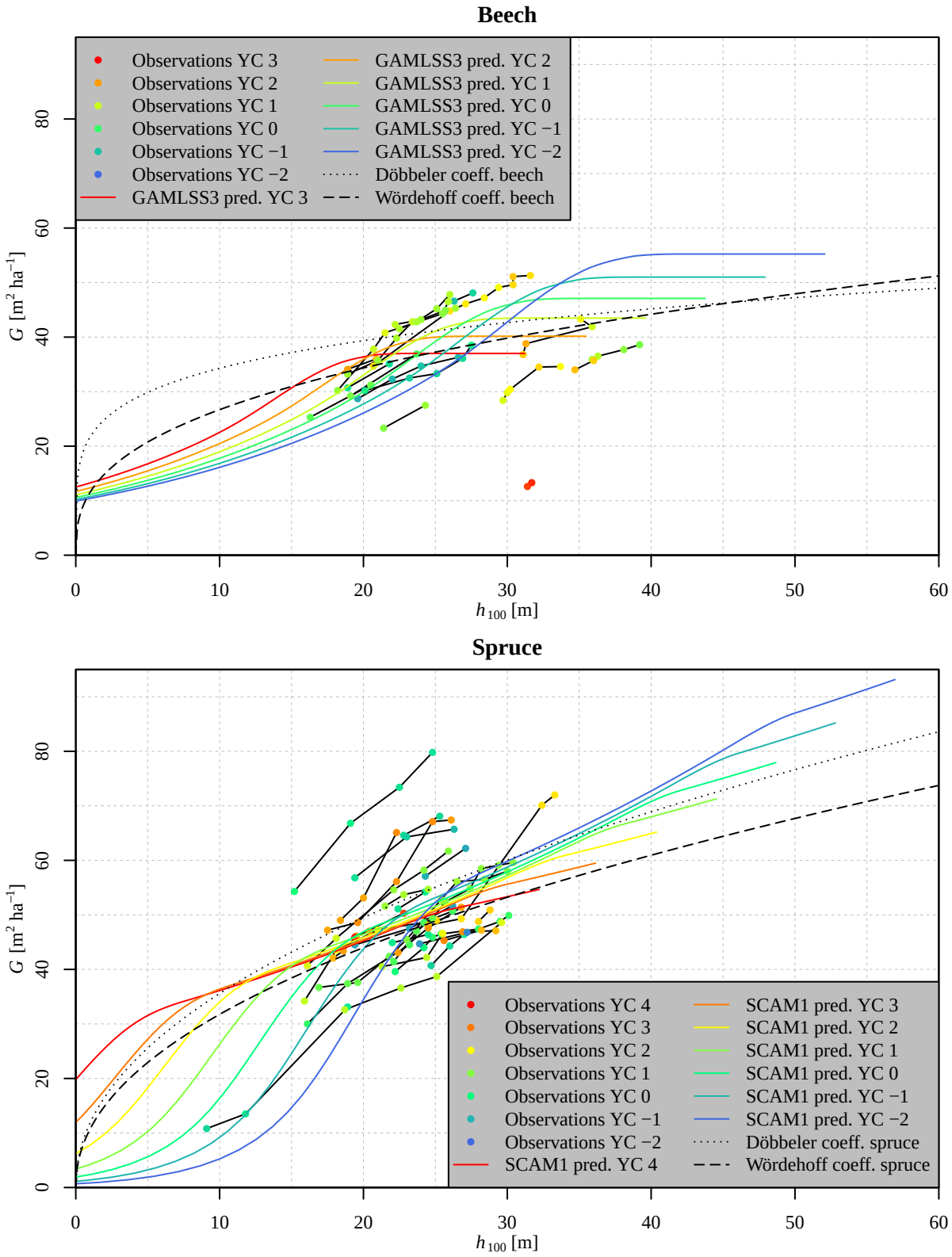

Figure 48: Basal area (G) over top height (h_{100}) of beech (top) and spruce (bottom). Depicted are observations as well as predictions of 3 different models per plot. Colored lines represent predictions of model GAMLSS3 (top) or of model SCAM1 (bottom). Dotted lines represent predictions of Equation (60), using the coefficient estimates reported by Döbbeler (2004) (Region Nordwest) for the respective species. Dashed lines represent predictions of Equation (60), using the coefficient estimates reported by Wördehoff (2016) for the respective species. Large dots, solid black lines, and colors have the same meaning as in Figure 45, namely: Each large dot represents one observation. Solid black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

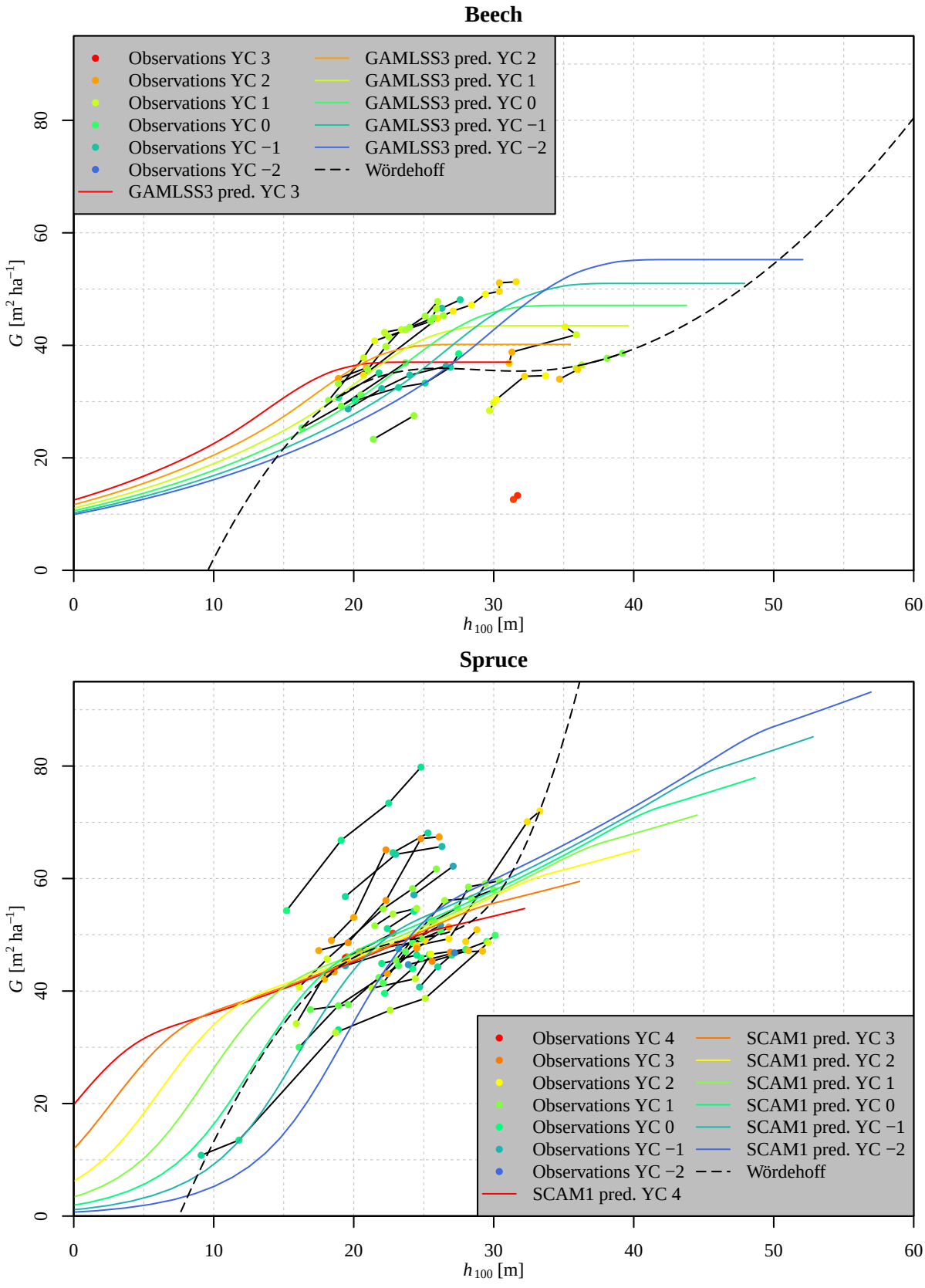

Figure 49: Basal area (G) over top height (h_{100}) of beech (top) and spruce (bottom). Depicted are observations as well as predictions of 2 different models per plot. Colored lines represent predictions of model GAMLSS3 (top) or of model SCAM1 (bottom). Dashed lines represent predictions of the GAMLSS reported by Wördehoff (2016) fitted to the respective species's training data set. Dots, solid black lines, and colors have the same meaning as in Figure 48, namely: Each dot represents one observation. Solid black lines connect observations belonging to the same sample plot. Color signifies the (fractional) yield class of the respective observation or prediction, ranging from red (worst yield class observed) over yellow and green to blue (best yield class observed). Yield class classification was based on absolute productivity index of stand as given by Equation (2) (rounded to one decimal digit), using Table 5 as reference. Note the different yield class ranges in both plots.

6.5 Top height and stand productivity

Following Baur (1881) (cited in Assmann (1970, p. 159)), this study rests on the so-called "site index hypothesis" (Skovsgaard & Vanclay 2008), which may be broken down into 2 assumptions:

1. Yield classes can be sufficiently distinguished based on top height at a given reference age (here: 100 a), i.e., their site index.
2. Stands of different yield classes attain different basal areas at a given age, while stands of the same yield class attain the same basal area at a given age.

These assumptions are based on the belief that top height in pure even-aged unthinned stands is largely independent of stem number (Skovsgaard & Vanclay 2008). They are, however, not undisputed. Assmann (1970) pointed out that stands sharing the same site index still show site-dependent differences in total crop yield as well as basal area. Nevertheless, site index continues to be a subject of research (Somarriba et al. 2001; Wang et al. 2005; Weiskittel et al. 2011) and was considered an indicator of site productivity by others (Karlsson et al. 1997; Monserud 1984; Rayner 1992). The goal of this study was not to determine whether or not site index is an appropriate measure for classfying site productivity, but rather to develop tools which allow modeling of maximum basal area while also taking differences in site productivity into account. These tools are applicable, regardless of whether the productivity is expressed in terms of site index or by other means. Adjusting the approach presented here to such other means would merely require replacing the productivity index variable with an appropriate alternative.

6.6 Test data

As can be seen in Figure 47, Equation (59) yields negative top heights for low stand ages (up to 25 a for yield class 3 and better of beech and up to 21 a for yield class 4 and better of spruce). This affects model results insofar, as in both species basal area predictions for a top height of 0 m exceed 0 $m^2\,ha^{-1}$, with the magnitude of excess depending on yield class (cp. Figures 15, 21, 27, 33, 39 and 45). Thus, model predictions for low stand ages are not as dependable as those for higher stand ages.

7 Conclusion

The presented data selection mechanism proved a valid way of removing non-maximum basal area observations from large data sets. It could be shown that GAMs, SCAMs, and GAMLSSs provide useful tools for modeling maximum basal area of pure even-aged beech and spruce stands from northwest Germany. The presented models use stand productivity and stand age

as predictor variables. This allows for more realistic predictions than previous modeling approaches, which only used top height as a predictor variable. Restriction of model flexibility led to improved plausibility of model predictions, but reduced statistical power of the models. Further model refinement would greatly benefit from increased data set sizes.

References

Akaike H. (1998). “Information Theory and an Extension of the Maximum Likelihood Principle”. In: *Selected Papers of Hirotugu Akaike*. Ed. by E. Parzen, K. Tanabe and G. Kitagawa. New York

Akanztiliotou K., Rigby R. A. and Stasinopoulos D. M. (2002). “The R implementation of Generalized Additive Models for Location, Scale and Shape”. In: *Statistical Modelling in Society: Proceedings of the 17th International Workshop on Statistical Modelling*. Ed. by M. Stasinopoulos, G. Touloumi, pp. 75–83. url: `http://www.gamlss.org/wp-content/uploads/2013/01/cretegamlss.pdf` (visited on June 6, 2017)

Albert M., Schmidt M. (2009). “Beurteilung der Anbauwürdigkeit von Baumarten unter Klimawandel mittels dreidimensionaler Ökogramme”. In: *Deutscher Verband Forstlicher Forschungsanstalten. Sektion Ertragskunde. Jahrestagung 25.- 27. Mai 2009 Ascona (Schweiz)*. Ed. by J. Nagel, pp. 83–94. url: `http://sektionertragskunde.fvabw.de/SektionErtragskundeBand2009.pdf` (visited on June 15, 2017)

Ando T. (1968). “Ecological Studies on the Stand Density Control in Even-Aged Pure Stand”. *Bulletin of the Government Forest Experiment Station* 210, pp. 1–153

Ando T., Hatiya K., Doi K., Kataoka H., Kato Y. and Sakaguchi K. (1968). “Studies on the System of Density Control of Sugi (*Cryptomerica japonica*) Stand”. *Bulletin of the Government Forest Experiment Station* 209, pp. 1–76

Andreassen K., Tomter S. M. (2003). “Basal area growth models for individual trees of Norway spruce, Scots pine, birch and other broadleaves in Norway”. *Forest Ecology and Management* 180 (1), pp. 11–24. doi: `10.1016/S0378-1127(02)00560-1`

Assmann E. (1961). *Waldertragskunde. Organische Produktion, Struktur, Zuwachs und Ertrag von Waldbeständen*. München et al.

– (1970). *The principles of forest yield study. Studies in the organic production, structure, increment and yield of forest stands*. Oxford et al.

Augustin N. H., Sauleau E.-A. and Wood S. N. (2012). “On quantile quantile plots for generalized linear models”. *Computational Statistics & Data Analysis* 56 (8), pp. 2404–2409. doi: `10.1016/j.csda.2012.01.026`

Baur F. v. (1881). *Das Forstliche Versuchswesen I*

Bollaerts K., Eilers P. H. C. and Aerts M. (2006). “Quantile regression with monotonicity restrictions using *P*-splines and the L_1-norm”. *Statistical Modelling* 6 (3), pp. 189–207. doi: `10.1191/1471082X06st118oa`

Boor C. de (2001). *A Practical Guide to Splines*. New York et al.

Box G. E. P., Cox D. R. (1964). “An Analysis of Transformations”. *Journal of the Royal Statistical Society. Series B (Methodological)* 26 (2), pp. 211–252. url: `http://www.jstor.org/stable/2984418`

Burkschat M., Cramer E. and Kamps U. (2012). *Beschreibende Statistik. Grundlegende Methoden der Datenanalyse*. Berlin and Heidelberg. doi: `10.1007/978-3-642-30013-4`

Castedo-Dorado F., Diéguez-Aranda U., Barrio-Anta M. and Álvarez-González J. G. (2007). "Modelling stand basal area growth for radiata pine plantations in Northwestern Spain using the GADA". *Annals of Forest Science* 64 (6), pp. 609–619. doi: 10.1051/forest:2007039

Charru M., Seynave I., Morneau F., Rivoire M. and Bontemps J.-D. (2012). "Significant differences and curvilinearity in the self-thinning relationships of 11 temperate tree species assessed from forest inventory data". *Annals of Forest Science* 69 (2), pp. 195–205. doi: 10.1007/s13595-011-0149-0

Chikumbo O. (2001). "A basal area model responsive to thinning for a plantation forest". *Environment International* 27 (2). Modelling & Sustainability, pp. 207–210. doi: 10.1016/S0160-4120(01)00084-8

Chikumbo O., Mareels I. M. and Turner B. J. (1999). "Predicting stand basal area in thined stands using a dynamical model". *Forest Ecology and Management* 116 (1), pp. 175–187. doi: 10.1016/S0378-1127(98)00449-6

Cole T. J., Green P. J. (1992). "Smoothing reference centile curves: The LMS method and penalized likelihood". *Statistics in Medicine* 11 (10), pp. 1305–1319. doi: 10.1002/sim.4780111005

Craven P., Wahba G. (1979). "Smoothing Noisy Data with Spline Functions". *Numerische Mathematik* 31 (4), pp. 377–403. doi: 10.1007/BF01404567

Curry H. B., Schoenberg I. J. (1947). "On spline distributions and their limits: the Pólya distribution functions". *Bulletin of the American Mathematical Society* 53, p. 1114. doi: 10.1090/S0002-9904-1947-08918-7

Döbbeler H. (2004). "Simulation und Bewertung von Nutzungsstrategien unter heutigen und veränderten Klimabedingungen mit dem Wuchsmodell SILVA 2.2". PhD thesis. Fakultät für Forstwissenschaften und Waldökologie, Georg-August-Universität Göttingen. url: http://hdl.handle.net/11858/00-1735-0000-0006-B114-3 (visited on July 11, 2017)

Döbbeler H., Spellmann H. (2002). "Methodological Approach to Simulate and Evaluate Silvicultural Treatments under Climate Change". *Forstwissenschaftliches Centralblatt* 121, Supplement 1, pp. 52–69

Dormann C. F. (2013). *Parametrische Statistik. Verteilungen, maximum likelihood und GLM in R*. Berlin and Heidelberg

Drew T. J., Flewelling J. W. (1979). "Stand Density Management: an Alternative Approach and Its Application to Douglas-fir Plantations". *Forest Science* 25 (3), pp. 518–532. url: http://www.ingentaconnect.com/content/saf/fs/1979/00000025/00000003/art00024

Duchon J. (1977). "Splines minimizing rotation-invariant semi-norms in Sobolev spaces". In: *Constructive Theory of Functions of Several Variables. Proceedings of a Conference Held at Oberwolfach April 25 – May 1, 1976*. Ed. by W. Schempp, K. Zeller. Vol. 571. Lecture Notes in Mathematics. Berlin et al.: Springer Berlin Heidelberg, pp. 85–100. doi: 10.1007/BFb0086566

Dunn P. K., Smyth G. K. (1996). "Randomized Quantile Residuals". *Journal of Computational and Graphical Statistics* 5 (3), p. 236. doi: 10.2307/1390802

Eerikäinen K., Maltamo M. (2003). "A percentile based basal area diameter distribution model for predicting the stand development of *Pinus kesiya* plantations in Zambia and Zimbabwe". *Forest Ecology and Management* 172 (1), pp. 109–124. doi: 10.1016/S0378-1127(02)00443-7

Eilers P. H. C., Marx B. D. (1996). "Flexible Smoothing with *B*-splines and Penalties". *Statistical Science* 11 (2), pp. 89–121. doi: 10.1214/ss/1038425655

Fox J. (2016). *Applied regression analysis and generalized linear models*. Los Angeles et al.

Franz F. (1965). "Ermittlung von Schätzwerten der natürlichen Grundfläche mit Hilfe ertragskundlicher Bestimmungsgrößen des verbleibenden Bestandes". *Forstwissenschaftliches Centralblatt* 84, pp. 357–386

– (1967). "Ertragsniveau-Schätzverfahren für die Fichte an Hand einmalig erhobener Bestandesgrößen". *Forstwissenschaftliches Centralblatt* 86 (2), pp. 98–125. doi: 10.1007/BF01822159

Goulding C. J. (1972). "Simulation techniques for a stochastic model of the growth of Douglas-fir". PhD thesis. University of British Columbia, Vancouver

Hastie T., Tibshirani R. (1991). *Generalized additive models*. Boca Raton, Florida

Hein S., Dhôte J.-F. (2006). "Effect of species composition, stand density and site index on the basal area increment of oak trees (*Quercus* sp.) in mixed stands with beech (*Fagus sylvatica* L.) in northern France". *Ann. For. Sci.* 63 (5), pp. 457–467. doi: 10.1051/forest:2006026

Henze N. (2013). *Stochastik für Einsteiger. Eine Einführung in die faszinierende Welt des Zufalls*. Wiesbaden

Jogiste K. (2000). "A Basal Area Increment Model for Norway Spruce in Mixed Stands in Estonia". *Scandinavian Journal of Forest Research* 15 (1), pp. 97–102. doi: 10.1080/02827580050160529

Karlsson A., Albrektson A. and Sonesson J. (1997). "Site Index and Productivity of Artificially Regenerated *Betula pendula* and *Betula pubescens* Stands on Former Farmland in Southern and Central Sweden". *Scandinavian Journal of Forest Research* 12 (3), pp. 256–263. doi: 10.1080/02827589709355408

Kenk G., Fischer H. (1988). "Evidence from Nitrogen Fertilisation in the Forests of Germany". *Environmental Pollution* 54 (3). Excess Nitrogen Deposition, pp. 199–218. doi: 10.1016/0269-7491(88)90112-1

Kira T., Ogawa H. and Sakazaki N. (1953). "Intraspecific competition among higher plants I. Competition-yield-density interrelationship in regularly dispersed populations". *Journal of the Institute of Polytechnics Osaka City University*. D, Biology 4, pp. 1–16

Lindgren B. W. (1976). *Statistical Theory*. New York

MacKinney A. L., Chaiken L. (1935). *A Method of Determining Density of Loblolly Pine Stands*. Tech. rep. 15. U.S. Deptartment of Agriculture, Forest Service, Appalachian Forest Experiment Station. 3 pp. url: https://archive.org/download/CAT31365986/CAT31365986.pdf (visited on July 10, 2017)

Meyer W. H. (1938). *Yield of even-aged stands of ponderosa pine*. Tech. rep. 630. U.S. Deptartment of Agriculture. url: https://naldc.nal.usda.gov/catalog/CAT86200625 (visited on July 12, 2017)

Monserud R. A. (1984). "Height Growth and Site Index Curves for Inland Douglas-fir Based on Stem Analysis Data and Forest Habitat Type". *Forest Science* 30 (4), pp. 943–965. url: http://www.ingentaconnect.com/content/saf/fs/1984/00000030/00000004/art00015

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 (1), pp. 57–80. doi: 10.1016/0378-1127(95)03638-5

Nagel J. (1999). *Konzeptionelle Überlegungen zum schrittweisen Aufbau eines waldwachstumskundlichen Simulationssystems für Nordwestdeutschland*. Vol. 128. Schriften aus der Forstlichen Fakultät der Universität Göttingen und der Niedersächsischen Forstlichen Versuchsanstalt. Frankfurt am Main

Nelder J. A., Wedderburn R. W. M. (1972). "Generalized Linear Models". *Journal of the Royal Statistical Society*. A (General) 135 (3), pp. 370–384. url: http://www.jstor.org/stable/2344614

Nyström K., Kexi M. (1997). "Individual tree basal area growth models for young stands of Norway spruce in Sweden". *Forest Ecology and Management* 97 (2). Regeneration Success and Early Growth of Forest Stands, pp. 173–185. doi: 10.1016/S0378-1127(97)00098-4

Pretzsch H. (2000). "Die Regeln von REINEKE, YODA und das Gesetz der räumlichen Allometrie". *Allgemeine Forst- und Jagdzeitung* 11, pp. 205–210. url: http://www.sauerlaender-verlag.com/CMS/uploads/media/AFJZ171__11__2000.pdf

– (2002). "A Unified Law of Spatial Allometry for Woody and Herbaceous Plants". *Plant Biology* 4 (2), pp. 159–166. doi: 10.1055/s-2002-25732

– (2006). "Species-specific allometric scaling under self-thinning: evidence from long-term plots in forest stands". *Oecologia* 146 (4), pp. 572–583. doi: 10.1007/s00442-005-0126-0

Pretzsch H., Biber P. (2005). "A Re-Evaluation of Reineke's Rule and Stand Density Index". *Forest Science* 51 (4), pp. 304–320. url: http://www.ingentaconnect.com/content/saf/fs/2005/00000051/00000004/art00004

Pya N. (2010). "Additive models with shape constraints". PhD thesis. University of Bath, Department of Mathematical Sciences

– (2017). *scam: Shape Constrained Additive Models*. R package version 1.2-2. url: https://CRAN.R-project.org/package=scam (visited on December 14, 2017)

Pya N., Wood S. N. (2015). "Shape constrained additive models". *Statistics and Computing* 25 (3), pp. 543–559. doi: 10.1007/s11222-013-9448-7

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. url: https://www.R-project.org/

Rayner M. (1992). "Evaluation of six site classifications for modelling timber yield of regrowth karri (*Eucalyptus diversicolor* F. Muell.)" *Forest Ecology and Management* 54 (1), pp. 315–336. doi: 10.1016/0378-1127(92)90020-A

Reineke L. H. (1933). “Perfecting a stand-density index for even-aged forests”. *Journal of Agricultural Research* 46, pp. 627–638

Rigby B., Stasinopoulos M. (2001). “The GAMLSS project: a flexible approach to statistical modelling”. In: *New Trends in Statistical Modelling: Proceedings of the 16th International Workshop on Statistical Modelling*. Ed. by B. Klein, L. Korsholm, pp. 249–256. url: http://www.gamlss.org/wp-content/uploads/2013/01/paper044.pdf

Rigby R. A., Stasinopoulos D. M. (1996). “A semi-parametric additive model for variance heterogeneity”. *Statistics and Computing* 6 (1), pp. 57–65. doi: 10.1007/BF00161574

– (2005). “Generalized additive models for location, scale and shape”. *Applied Statistics* 54, pp. 507–554

Río M. del, Montero G. and Bravo F. (2001). “Analysis of diameter–density relationships and self-thinning in non-thinned even-aged Scots pine stands”. *Forest Ecology and Management* 142 (1), pp. 79–87. doi: 10.1016/S0378-1127(00)00341-8

Röhrig E. (1992). “Strukturen von Waldbeständen”. In: *Waldbau auf ökologischer Grundlage*. Ed. by E. Röhrig, N. Bartsch and A. Dengler. Vol. 1: Der Wald als Vegetationsform und seine Bedeutung für den Menschen: mit 48 Tabellen. Hamburg, pp. 25–40

Schober R. (1995). *Ertragstafeln wichtiger Baumarten*. Frankfurt am Main

Schröder J., Soalleiro R. R. and Alonso G. V. (2002). “An age-independent basal area increment model for maritime pine trees in northwestern Spain”. *Forest Ecology and Management* 157 (1), pp. 55–64. doi: 10.1016/S0378-1127(00)00657-5

Schütz J.-P. (2008). “Ertragsniveau und maximale Bestockungsdichte als Grundlage für die Modellierung der natürlichen Mortalität”. In: *Deutscher Verband Forstlicher Forschungsanstalten. Sektion Ertragskunde. Jahrestagung 5.-8. Mai 2008 Trippstadt*. Ed. by J. Nagel, pp. 123–132. url: http://sektionertragskunde.fvabw.de/SektionErtragskundeBand2008.pdf

Schütz J.-P., Zingg A. (2010). “Improving estimations of maximal stand density by combining Reineke’s size-density rule and the yield level, using the example of spruce (*Picea abies* (L.) Karst.) and European Beech (*Fagus sylvatica* L.)” *Annals of Forest Science* 67 (5), p. 507. doi: 10.1051/forest/2010009

Silva C. A., Klauberg C., Hudak A. T., Vierling L. A., Fennema S. J. and Corte A. P. D. (2017). “Modeling and mapping basal area of *Pinus taeda* L. plantation using airborne LiDAR data”. en. *Anais da Academia Brasileira de Ciências* 89, pp. 1895–1905. url: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0001-37652017000401895&nrm=iso

Skovsgaard J. P., Vanclay J. K. (2008). “Forest site productivity: a review of the evolution of dendrometric concepts for even-aged stands”. *Forestry: An International Journal of Forest Research* 81 (1), pp. 13–31. doi: 10.1093/forestry/cpm041

Society of American Foresters (1958). *Forestry Terminology*. Washington, D. C.

Somarriba E., Valdivieso R., Vásquez W. and Galloway G. (2001). “Survival, growth, timber productivity and site index of *Cordia alliodora* in forestry and agroforestry systems”. *Agroforestry Systems* 51 (2), pp. 111–118. doi: 10.1023/A:1010699019745

Spellmann H., Nagel J. and Böckmann T. (1999). “Summarische Nutzungsplanung auf der Basis von Betriebsinventurdaten”. *Allgemeine Forst- und Jagdzeitung* 170, pp. 122–128

Stasinopoulos D. M., Rigby R. A. (2007). “Generalized Additive Models for Location Scale and Shape (GAMLSS) in R”. *Journal of Statistical Software* 23 (7), pp. 1–46. doi: 10.18637/jss.v023.i07

Stasinopoulos M. (2017). *gamlss.dist: Distributions for Generalized Additive Models for Location Scale and Shape*. R package version 5.0-3. url: https://CRAN.R-project.org/package=gamlss.dist (visited on December 14, 2017)

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

– (1981). “Natürlicher Bestockungsgrad und Reinekes SDI”. *Centralblatt für das gesamte Forstwesen* 98, pp. 101–116

– (1987). “Estimating Potential Density from Thinning Experiments and Inventory Data”. *Forest Science* 33, pp. 1022–1034

Tadaki Y. (1963). “The Pre-estimating of Stem Yield based on the Competition-Density Effect”. *Bulletin of the Government Forest Experiment Station* 154, pp. 1–19

Vacchiano G., Derose R. J., Shaw J. D., Svoboda M. and Motta R. (2013). “A density management diagram for Norway spruce in the temperate European montane region”. *European Journal of Forest Research* 132 (3), pp. 535–549. doi: 10.1007/s10342-013-0694-1

VanderSchaaf C. L. (2010). “Estimating Individual Stand Size–Density Trajectories and a Maximum Size–Density Relationship Species Boundary Line Slope”. *Forest Science* 56 (4), pp. 327–335. url: http://www.ingentaconnect.com/content/saf/fs/2010/00000056/00000004/art00001

VanderSchaaf C. L., Burkhart H. E. (2007). “Comparison of Methods to Estimate Reineke’s Maximum Size-Density Relationship Species Boundary Line Slope”. *Forest Science* 53 (3), pp. 435–442. url: http://www.ingentaconnect.com/content/saf/fs/2007/00000053/00000003/art00006

– (2008). “Using Segmented Regression to Estimate Stages and Phases of Stand Development”. *Forest Science* 54 (2), pp. 167–175. url: http://www.ingentaconnect.com/content/saf/fs/2008/00000054/00000002/art00006

Vospernik S., Sterba H. (2015). “Do competition-density rule and self-thinning rule agree?” *Annals of Forest Science* 72 (3), pp. 379–390. doi: 10.1007/s13595-014-0433-x

Wahba G. (1990). *Spline models for observational data*. Philadelphia, Pa.

Wang Y., Raulier F. and Ung C.-H. (2005). “Evaluation of spatial predictions of site index obtained by parametric and nonparametric methods—A case study of lodgepole pine productivity”. *Forest Ecology and Management* 214 (1), pp. 201–211. doi: 10.1016/j.foreco.2005.04.025

Weiskittel A. R., Crookston N. L. and Radtke P. J. (2011). “Linking climate, gross primary productivity, and site index across forests of the western United States”. *Canadian Journal of Forest Research* 41 (8), pp. 1710–1721. doi: 10.1139/x11-086

Weisstein E. W. (2017a). *B-Spline. From MathWorld–A Wolfram Web Resource.* url: http://mathworld.wolfram.com/B-Spline.html (visited on November 10, 2017)

– (2017b). *Gamma Distribution. From MathWorld–A Wolfram Web Resource.* url: http://mathworld.wolfram.com/GammaDistribution.html (visited on November 15, 2017)

– (2017c). *Standard Normal Distribution. From MathWorld–A Wolfram Web Resource.* url: http://mathworld.wolfram.com/StandardNormalDistribution.html (visited on November 15, 2017)

Wimberly M. C., Bare 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), pp. 1–11. doi: 10.1016/S0378-1127(96)03870-4

Wood S. N. (2001). “mgcv: GAM and Generalized Ridge Regression for R”. *R News* 1 (2), pp. 20–25

– (2003). “Thin Plate Regression Splines”. *Journal of the Royal Statistical Society*. B (Statistical Methodology) 65 (1), pp. 95–114. url: http://www.jstor.org/stable/3088828

– (2004). “Stable and Efficient Multiple Smoothing Parameter Estimation for Generalized Additive Models”. *Journal of the American Statistical Association* 99 (467), pp. 673–686. doi: 10.1198/016214504000000980

– (2006). *Generalized additive models: an introduction with R*. Texts in statistical science. Boca Raton, Florida

– (2011). “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models: Estimation of Semiparametric Generalized Linear Models”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (1), pp. 3–36. doi: 10.1111/j.1467-9868.2010.00749.x

– (2017). *Generalized additive models: An introduction with R*

Wood S. N., Pya N. and Säfken B. (2016). “Smoothing Parameter and Model Selection for General Smooth Models”. *Journal of the American Statistical Association* 111 (516), pp. 1548–1563. doi: 10.1080/01621459.2016.1180986

Wördehoff R. (2016). “Kohlenstoffspeicherung als Teilziel der strategischen Waldbauplanung erläutert an Reinbeständen verschiedener Baumarten in Niedersachsen”. PhD thesis. Fakultät für Forstwissenschaften und Waldökologie der Georg-August-Universität Göttingen

Wördehoff R., Schmidt M., Nagel R.-V. and Spellmann H. (2014). “Prognose der maximalen Bestandesgrundfläche mit Hilfe von Quantilsregressionen und Entwicklung eines grundflächengesteuerten Nutzungskonzeptes für die Baumarten Buche und Fichte in Nordwestdeutschland”. In: *Deutscher Verband Forstlicher Forschungsanstalten Sektion Ertragskunde Tagungsband 2014*, pp. 88–92

Yee T. W. (2004). “Quantile regression via vector generalized additive models”. *Statistics in Medicine* 23 (14), pp. 2295–2315. doi: 10.1002/sim.1822

Yoda K., Kira T., Ogawa H. and Hozumi K. (1963). “Self-thinning in Overcrowded Pure Stands under Cultivated and Natural Conditions. (Intraspecific Competition among Higher Plants XI)”. *Journal of Biology, Osaka City University* 14, pp. 107–129

Yue C., Kohnle U., Kahle H.-P. and Klädtke J. (2012). “Exploiting irregular measurement intervals for the analysis of growth trends of stand basal area increments: A composite model approach”. *Forest Ecology and Management* 263 (Supplement C), pp. 216–228. doi: 10.1016/j.foreco.2011.09.007

Zeide B. (1985). “Tolerance and self-tolerance of trees”. *Forest Ecology and Management* 13 (3), pp. 149–166. doi: 10.1016/0378-1127(85)90031-3

– (1987). “Analysis of the 3/2 Power Law of Self-Thinning”. *Forest Science* 33 (2), pp. 517–537. url: http://www.ingentaconnect.com/content/saf/fs/1987/00000033/00000002/art00024

– (1995). “A relationship between size of trees and their number”. *Forest Ecology and Management* 72 (2), pp. 265–272. doi: 10.1016/0378-1127(94)03453-4

A Appendix

A.1 Model statistics

Table 14: Coefficient estimates and statistics of parametric and smooth terms in model GAM1 for beech. Statistics of smooth terms are approximations.
edf: effective degrees of freedom
`PIV`: productivity index variable
Pr(x): probability of event x
`s(x)`: smooth function applied to `x` with thin plate regression splines as function basis
`SAV`: stand age variable
SE: standard error

Parametric model term	Estimate	SE	t value	Pr($>\|t\|$)
Intercept	3.59847	0.01775	202.7	$<2 \cdot 10^{-16}$
Smooth model term		**edf**	**F value**	**Pr($\geq F$)**
`s(SAV)`		2.663	7.68	$1.58 \cdot 10^{-4}$
`s(PIV)`		4.865	13.14	$5.45 \cdot 10^{-10}$

Table 15: Coefficient estimates and statistics of parametric and smooth terms in model GAM1 for spruce. Statistics of smooth terms are approximations.
edf: effective degrees of freedom
`PIV`: productivity index variable
Pr(x): probability of event x
`s(x)`: smooth function applied to `x` with thin plate regression splines as function basis
`SAV`: stand age variable
SE: standard error

Parametric model term	Estimate	SE	t value	Pr($>\|t\|$)
Intercept	3.88328	0.01618	240	$<2 \cdot 10^{-16}$
Smooth model term		**edf**	**F value**	**Pr($\geq F$)**
`s(SAV)`		6.391	22.829	$<2 \cdot 10^{-16}$
`s(PIV)`		6.480	4.675	$1.36 \cdot 10^{-4}$

Table 16: Coefficient estimates and statistics of parametric and smooth terms in model GAM2 for beech. Statistics of smooth terms are approximations.
edf: effective degrees of freedom
`PIV`: productivity index variable
Pr(x): probability of event x
`s(x)`: smooth function applied to `x` with thin plate regression splines as function basis
`SAV`: stand age variable
SE: standard error

Parametric model term	Estimate	SE	t value	Pr($>\|t\|$)
Intercept	3.573394	0.024359	146.70	$<2 \cdot 10^{-16}$
`PIV`	0.0295	0.009307	3.17	$2.45 \cdot 10^{-3}$
Smooth model term		**edf**	**F value**	**Pr($\geq F$)**
`s(SAV)`		3.702	11.12	$2.3 \cdot 10^{-7}$

Table 17: Coefficient estimates and statistics of parametric and smooth terms in model GAM2 for spruce. Statistics of smooth terms are approximations.
edf: effective degrees of freedom
`PIV`: productivity index variable
Pr(x): probability of event x
`s(x)`: smooth function applied to `x` with thin plate regression splines as function basis
`SAV`: stand age variable
SE: standard error

Parametric model term	Estimate	SE	t value	Pr($>\|t\|$)
Intercept	3.89044	0.01794	216.816	$<2 \cdot 10^{-16}$
`PIV`	0.02207	0.00575	3.838	$2.27 \cdot 10^{-4}$
Smooth model term		**edf**	**F value**	**Pr($\geq F$)**
`s(SAV)`		5.318	20.16	$<2 \cdot 10^{-16}$

Table 18: Coefficient estimates and statistics of parametric and smooth terms in model SCAM1 for beech. Statistics of smooth terms are approximations.
edf: effective degrees of freedom
`PIV`: productivity index variable
Pr(x): probability of event x
`s(x, bs = "micv")`: smooth function applied to `x` with P-splines constrained to be increasing concave as function basis
`SAV`: stand age variable
SE: standard error

Parametric model term	Estimate	SE	t value	Pr($>\|t\|$)
Intercept	2.8269	0.26641	10.611	$2.65 \cdot 10^{-15}$
`PIV`	0.03577	0.01157	3.091	$3.04 \cdot 10^{-3}$
Smooth model term		**edf**	**F value**	**Pr($\geq F$)**
`s(SAV, bs = "micv")`		1.857	8.657	$3.99 \cdot 10^{-4}$

Table 19: Coefficient estimates and statistics of parametric and smooth terms in model SCAM1 for spruce. Statistics of smooth terms are approximations.
edf: effective degrees of freedom
`PIV`: productivity index variable
Pr(x): probability of event x
`s(x, bs = "micv")`: smooth function applied to `x` with P-splines constrained to be increasing concave as function basis
`SAV`: stand age variable
SE: standard error

Parametric model term	Estimate	SE	t value	Pr($>\|t\|$)
Intercept	1.066116	0.429109	2.484	$1.47 \cdot 10^{-2}$
`PIV`	0.022874	0.005441	4.204	$5.92 \cdot 10^{-5}$
Smooth model term		**edf**	**F value**	**Pr($\geq F$)**
`s(SAV, bs = "micv")`		3.012	43.12	$<2 \cdot 10^{-16}$

Table 20: Coefficient estimates and statistics of model terms for all distribution parameters (μ, σ, ν) in model GAMLSS1 for beech. Standard errors for smooth function terms apply only to the linear effect. Standard errors for linear terms may not be accurate.
`PIV`: productivity index variable
Pr(x): probability of event x
`ps(x)`: smooth function applied to `x` with P-splines as function basis
`SAV`: stand age variable
SE: standard error

Model term	Estimate	SE	t value	$\Pr(>\|t\|)$
Distribution parameter μ				
Intercept	3.18399112	0.075858297	41.972879	$8.733187 \cdot 10^{-42}$
`ps(SAV)`	0.01553977	0.002412037	6.442593	$3.826335 \cdot 10^{-8}$
`ps(PIV)`	0.02583156	0.005074946	5.090018	$5.022636 \cdot 10^{-6}$
Distribution parameter σ				
Intercept	−2.03742179	0.400343606	−5.089183	$5.037397 \cdot 10^{-6}$
Distribution parameter ν				
Intercept	6.81971611	1.725801250	3.951623	$2.351630 \cdot 10^{-4}$

Table 21: Coefficient estimates and statistics of model terms for all distribution parameters (μ, σ, ν) in model GAMLSS1 for spruce. Standard errors for smooth function terms apply only to the linear effect. Standard errors for linear terms may not be accurate.
`PIV`: productivity index variable
Pr(x): probability of event x
`ps(x)`: smooth function applied to `x` with P-splines as function basis
`SAV`: stand age variable
SE: standard error

Model term	Estimate	SE	t value	Pr($>\|t\|$)
Distribution parameter μ				
Intercept	2.96336411	0.074718912	39.660161	$2.335117 \cdot 10^{-58}$
`ps(SAV)`	0.03720503	0.003084204	12.063092	$2.003461 \cdot 10^{-20}$
`ps(PIV)`	0.02587688	0.004038600	6.407388	$6.852684 \cdot 10^{-9}$
Distribution parameter σ				
Intercept	−1.92345892	0.083753138	−22.965813	$3.607049 \cdot 10^{-39}$
Distribution parameter ν				
Intercept	−2.34254041	0.794897386	−2.946972	$4.096469 \cdot 10^{-3}$

Table 22: Coefficient estimates and statistics of model terms for all distribution parameters (μ, σ, ν) in model GAMLSS2 for beech. Standard errors for smooth function terms apply only to the linear effect. Standard errors for linear terms may not be accurate.
`PIV`: productivity index variable
Pr(x): probability of event x
`ps(x)`: smooth function applied to `x` with P-splines as function basis
`SAV`: stand age variable
SE: standard error

Model term	Estimate	SE	t value	Pr($>\|t\|$)
Distribution parameter μ				
Intercept	3.14190153	0.086433909	36.350335	$3.830686 \cdot 10^{-40}$
`ps(SAV)`	0.01784155	0.002844072	6.273240	$5.803796 \cdot 10^{-8}$
`PIV`	0.02318045	0.005761870	4.023077	$1.769104 \cdot 10^{-4}$
Distribution parameter σ				
Intercept	−1.92886231	0.325428360	−5.927149	$2.108970 \cdot 10^{-7}$
Distribution parameter ν				
Intercept	5.32543126	1.348124323	3.950252	$2.243708 \cdot 10^{-4}$

Table 23: Coefficient estimates and statistics of model terms for all distribution parameters (μ, σ, ν) in model GAMLSS2 for spruce. Standard errors for smooth function terms apply only to the linear effect. Standard errors for linear terms may not be accurate.
`PIV`: productivity index variable
Pr(x): probability of event x
`ps(x)`: smooth function applied to `x` with P-splines as function basis
`SAV`: stand age variable
SE: standard error

Model term	Estimate	SE	t value	Pr($>\|t\|$)
Distribution parameter μ				
Intercept	2.94992436	0.079987669	36.879739	$6.530661 \cdot 10^{-57}$
`ps(SAV)`	0.03780157	0.003246622	11.643353	$8.618176 \cdot 10^{-20}$
`PIV`	0.02304275	0.004578450	5.032871	$2.386608 \cdot 10^{-6}$
Distribution parameter σ				
Intercept	−1.87150855	0.081816573	−22.874443	$9.804238 \cdot 10^{-40}$
Distribution parameter ν				
Intercept	−2.03254437	0.673142253	−3.019487	$3.276757 \cdot 10^{-3}$

Table 24: Coefficient estimates and statistics of model terms for all distribution parameters (μ, σ, ν) in model GAMLSS3 for beech. Standard errors for smooth function terms apply only to the linear effect. Standard errors for linear terms may not be accurate.
`PIV`: productivity index variable
Pr(x): probability of event x
SE: standard error

Model term	Estimate	SE	t value	Pr($>\|t\|$)
Distribution parameter μ				
Intercept	3.60925073	0.033309288	108.355685	$3.639668 \cdot 10^{-68}$
`PIV`	0.02097131	0.004052918	5.174374	$3.031625 \cdot 10^{-6}$
Distribution parameter σ				
Intercept	−1.81603770	0.293080207	−6.196385	$6.650035 \cdot 10^{-8}$
Distribution parameter ν				
Intercept	4.27957208	1.175984884	3.639139	$5.870832 \cdot 10^{-4}$

Table 25: Coefficient estimates and statistics of model terms for all distribution parameters (μ, σ, ν) in model GAMLSS3 for spruce. Standard errors for smooth function terms apply only to the linear effect. Standard errors for linear terms may not be accurate.
`PIV`: productivity index variable
Pr(x): probability of event x
SE: standard error

Model term	Estimate	SE	t value	Pr($>\|t\|$)
Distribution parameter μ				
Intercept	3.84737943	0.016426776	234.213908	$1.307827 \cdot 10^{-128}$
`PIV`	0.02270638	0.003076847	7.379756	$7.137219 \cdot 10^{-11}$
Distribution parameter σ				
Intercept	−1.90983216	0.083585185	−22.848931	$1.607009 \cdot 10^{-39}$
Distribution parameter ν				
Intercept	−2.30907594	0.766078634	−3.014150	$3.336363 \cdot 10^{-3}$

Hiermit versichere ich gemäß § 7 Abs. 5 der Master-Prüfungsordnung vom 23.09.2010, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

_____

Datum Unterschrift