

Produktivitätsmodell

”Motormanuelle Holzhauerei 2014”

Teil B: Analyse der Datensätze und Diskussion der Modellierung

Dario Pedolin, WSL Birmensdorf

10. Dezember 2014

Co-Autoren: Herbert Borchert
 Fritz Frutig
 Renato Lemm
 Oliver Thees
Leitung: Oliver Thees

Zusammenfassung

Anhand eines Datensatzes über die Leistungen von motormanuellen Holzschlägen soll ein Modell für die Berechnung der Zeitaufwände für die Holzhauerei erstellt werden. Das Modell erklärt 67% der im Datensatz gefundenen Varianz (N= 22'525 Holzschläge, totale Holzmenge = 10'360'727 m³ ohne Rinde).

Inhaltsverzeichnis

1 Einleitung	3
2 Analyse Produktivitäten	4
2.1 Beschreibung der Prozesse	4
2.2 Beschreibung der Einheiten und Masse	4
2.3 Erstellen und Testen des Modells	4
3 Diskussion	9
3.1 Diskussion der Datengrundlage	9
3.2 Diskussion des Modells	12
4 Modelle einsetzen	13
4.1 Formel, Koeffizienten und Systemgrenzen	13
4.2 Sensitivitätsanalysen	15
5 Anhang	22
5.1 Testen auf Verletzungen der Annahmen für multiple Regression .	22
5.2 Beschreibung der ungültigen Fälle wegen Zufällen für Harvester .	32
5.3 Beschreibung der 'Ausreisser'	34
5.4 Kreuzvalidierung	36
5.5 Beschreibung der im Modell benutzten Messgrößen	39
5.6 Regionen vs. Hangneigung des Reviers	52
5.7 Erweiterung des Modells für kleine und grosse BHD _{mit}	54

1 Einleitung

Ein Datensatz bestehend aus 23'458 Holzschlägen aus den Jahren 2007 bis 2013 aus deutschen Forstbetrieben wurde benutzt, um ein Modell für die Produktivität der motormanuellen Holzhauerei zu erstellen. In einer ersten Analyse wurden Ausreisser identifiziert. In einem zweiten Schritt wurden diese Ausreisser aus der Analyse ausgeschlossen und ein Modell für die gemessene Produktivität erstellt. Zusätzlich wurde aus einem unabhängigen Datenset mit Informationen zu den, den Daten zugrundeliegenden Regionen mittels Clusteranalyse Gruppen für die unabhängige Variable 'Region' gebildet. Die beiden Regionen sind 'Gebirge' und 'Flach- und Hügelland', nachfolgend 'Huegelland' genannt (siehe auch Anhang Tabelle 15). Das Modell wurde in der Residuenanalyse auf Verletzungen der Voraussetzungen für lineare Modellierung untersucht (Anhang Abbildung 11 bis 17). Abschliessend wurde eine Sensitivitätsanalyse für die Parameter durchgeführt. Die Qualität des Modells wurde ausserdem mit einer Kreuzvalidierung getestet (Anhang Abbildung 20).

2 Analyse Produktivitäten

2.1 Beschreibung der Prozesse

Der Prozess 'motormanuelle Holzhauerei' umfasst alle Arbeitsschritte des Fällens der zum Aushieb gekennzeichneten Bäume (Fällen, Entasten sowie in Rundholzsortimente aufbereiten incl. Anschreiben von Länge und Durchmesser). Der Output des Prozesses 'motormanuelle Holzhauerei' sind die Rundholzabschnitte am Ort der gefällten Bäume.

2.2 Beschreibung der Einheiten und Masse

Alle Zeiten (Stunden [h]) sind WPPH (Arbeitsplatz-Zeit, vergleiche Anhang Abbildung 26) und beinhalten alle Zeitaufwände der motormanuellen Holzhauerei. Die Mengenangaben [m^3] sind Kubikmeter verkaufsfähiges Holz ohne Rinde. ("Derbholt das nicht verkauffertig aufgearbeitet wird (X-Holz¹ nach Definition des EST und NH), geht nicht in die Holzmenge mit ein. Der Arbeitszeitaufwand für dieses Holz ist dennoch in den produktiven Stunden mit enthalten." (Quelle: internes Dokument des Datenlieferanten))

2.3 Erstellen und Testen des Modells

Die erfassten Messgrößen wurden mittels der R Methode 'lm' ('Fitting Linear Models') auf ihren Einfluss auf die abhängige Variable 'TAP' [m^3 o.R./WPPH] (= Technische Arbeitsproduktivität in m^3 ohne Rinde /WPPH) untersucht. Dabei wurden nur die 23'458 Fälle benutzt, deren gemessene Produktivität maximal 10% über dem EST² liegen³ (siehe Anhang Tabelle 10 für Beschreibung der Variable 'Produktivität' sowie Abschnitt 'Diskussion der Datengrundlage'). Nachfolgend werden die Ergebnisse der Modellierung vorgestellt.

Tabelle 1 zeigt die Koeffizienten und p-Werte der ersten Analyse des Modells. Das korrigierte R^2 für die Modellbildung beträgt 0.56.

Die Abbildung 1 zeigt den Vergleich der berechneten mit den gemessenen Werten. Für Punkte, die auf der grauen Linie liegen, ist der Schätzwert gleich gross wie der gemessene Wert. Liegen die Punkte links der Linie unterschätzt das Modell die Werte. Liegen sie rechts der Linie, werden die Produktivitäten vom Modell überschätzt. Zusätzlich sind jene Fälle mit aussergewöhnlich grossen Schätzfehlern (= Residue minus Mittelwert der Residuen grösser als zwei Standardabweichungen der Residuen) schwarz markiert.

Für die nachfolgenden Auswertungen wurden diese Fälle aus dem Datensatz

¹X-Holz sind Baumteile oder Bäume ab Brusthöhendurchmesser 7 cm m.R., die bearbeitet, jedoch nicht verwendet bzw. genutzt werden (z.B. wegen Fäule, Krümmungen oder Splitterhaltigkeit)

²Der EST (erweiterter Sorten Tarif) beschreibt die höchste, langfristig erreichbare Produktivität unter einfachen Bedingungen.

³Im Datensatz sind auch Arbeiten des Zufällens für Harvester erfasst, d.h. Fällen ohne Aufarbeiten. Datensätze mit gemessenen Produktivitäten die über dem EST + 10% (vergleiche Anhang Abbildung 30) liegen wurden deshalb von den nachfolgenden Analysen ausgeschlossen.

Tabelle 1: Resultat der Analyse des linearen Zusammenhangs und Schätzung der Modellkoeffizienten

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-8.5370	0.2165	-39.44	0.0000
RegionHuegelland	1.3723	0.2232	6.15	0.0000
log(BHD)	2.6559	0.0555	47.84	0.0000
pctLbh	3.5815	0.1499	23.89	0.0000
pctKiefer	-0.2026	0.3737	-0.54	0.5876
RegionHuegelland:log(BHD)	-0.1135	0.0573	-1.98	0.0476
log(BHD):pctLbh	-0.6309	0.0401	-15.75	0.0000
log(BHD):pctKiefer	0.1667	0.0974	1.71	0.0872

entfernt. Von den 933 Ausreisern⁴ sind 180 den drei Forstbetrieben 15, 45 und 34 zuordbar was jeweils 15.8%, 10.9% und 10.4% der von diesem Forstbetrieb erfassten Fälle entspricht (siehe Anhang Tabelle 5 für Beschreibung der Ausreisser).

Für die nachfolgenden Auswertungen wurden die in Abbildung 1 identifizierten Ausreisser entfernt. Es werden noch 22'525 Fälle benutzt.

⁴Die 933 Ausreisser entsprechen 4% der Fälle der Grundgesamtheit.

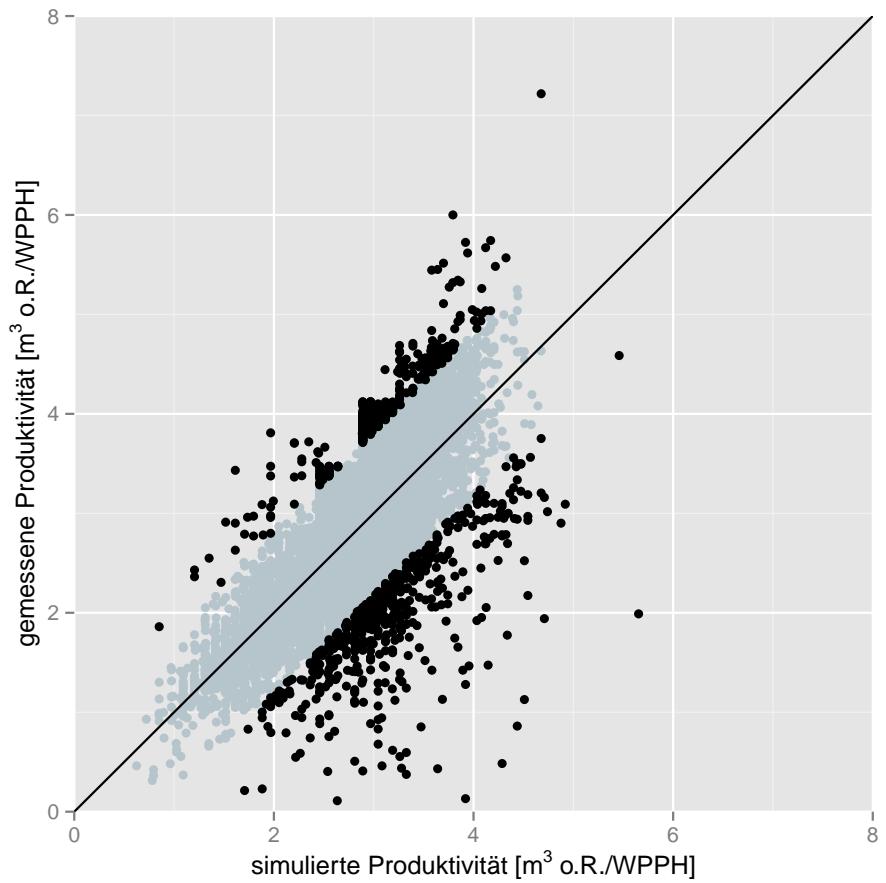


Abbildung 1: Vergleich der mittels des Modells berechneten mit der gemessenen Produktivität [$\text{m}^3 \text{ o.R.}/\text{WPPH}$]. Messwerte deren Residuen (= gemessener Wert - durch das Modell geschätzter Wert) mehr als zwei Standardabweichungen über dem Mittelwert der Residuen liegen, sind schwarz eingefärbt.

Tabelle 2: Resultat der Analyse des linearen Zusammenhangs und Schätzung der Modellkoeffizienten für den Datensatz ohne Ausreisser

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-8.8441	0.1797	-49.20	0.0000
RegionHuegelland	1.5564	0.1854	8.39	0.0000
log(BHD)	2.7332	0.0461	59.23	0.0000
pctLbh	2.5214	0.1281	19.69	0.0000
pctKiefer	-1.3273	0.3138	-4.23	0.0000
RegionHuegelland:log(BHD)	-0.1601	0.0476	-3.36	0.0008
log(BHD):pctLbh	-0.3171	0.0344	-9.21	0.0000
log(BHD):pctKiefer	0.4712	0.0819	5.75	0.0000

Tabelle 2 zeigt, dass vor allem die Variable 'BHD' (= mittlerer BHD nachfolgend 'BHD_{mit}') für die Produktivität des Fällens massgebend ist. Das korrigierte R² für die Modellbildung beträgt 0.67. Zusätzlich wurden die in Anhang Abbildungen 11 bis 17 gezeigten Tests durchgeführt, um auf Verletzungen der Voraussetzungen der Annahmen für einen linearen Zusammenhang zu testen. Zusätzlich wurde für das Modell eine Kreuzvalidierung durchgeführt (siehe Anhang Tabelle 20).

Die Abbildung 2 zeigt den Vergleich der berechneten mit den gemessenen Werten. Für Punkte, die auf der grauen Linie liegen, ist der Schätzwert gleich gross wie der gemessene Wert. Liegen die Punkte links der Linie, unterschätzt das Modell die Werte. Liegen sie rechts der Linie, werden die Produktivitäten vom Modell überschätzt.

Die mittlere Abweichung der berechneten Werte von den gemessenen beträgt 0.26 m³ o.R./WPPH (Median: 0.22 m³ o.R./WPPH). Um ein Mass für die relative Genauigkeit der durch das Modell berechneten Werte zu bekommen, wurde auch die relative Abweichung der berechneten Werte von den gemessenen Werten untersucht (d.h. (gemessener Wert - berechneter Wert)/ gemessener Wert; vergl. Park and Stefanski (1998)). Die mittlere relative Abweichung beträgt 9.4% (Median: 8%). Vergleiche auch Anhang Abbildung 17 und 18 für einen Überblick über die Verteilung der Residuen.

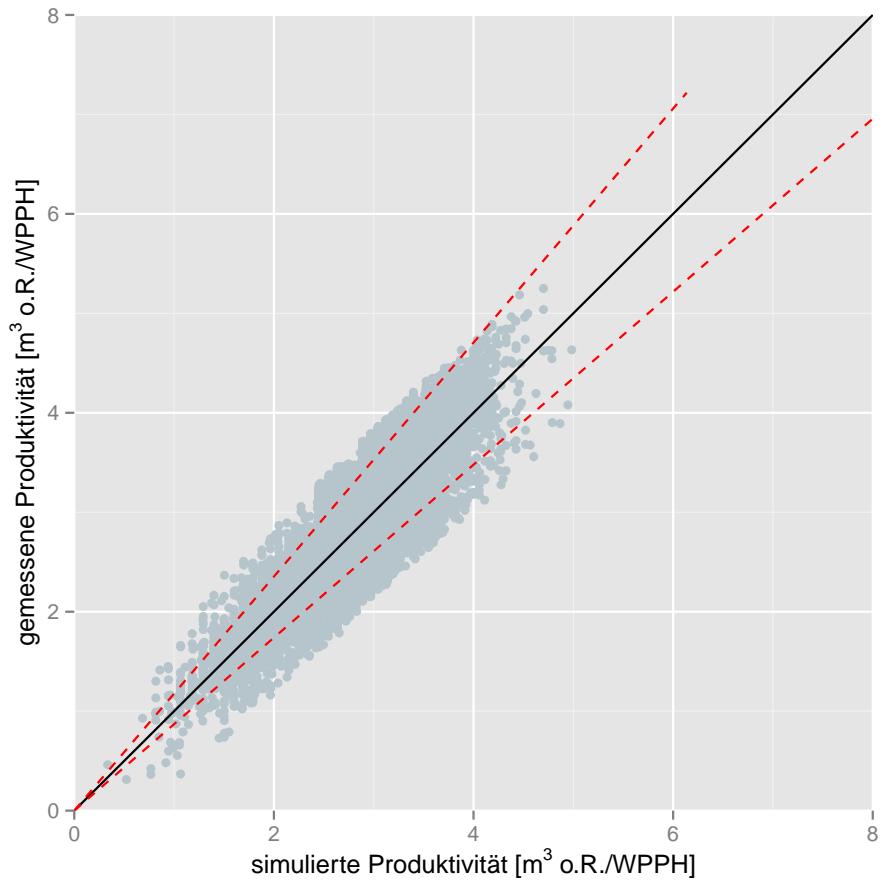


Abbildung 2: Vergleich der mittels des Modells berechneten Produktivität mit der gemessenen Produktivität. Die roten gestrichelten Linien grenzen den Bereich von 15% Abweichung ein.

3 Diskussion

3.1 Diskussion der Datengrundlage

Im Dataset wurde statt der unabhängigen Grösse 'BHD_{mit} [cm]' der Mittendurchmesser [cm] des liegenden Stammes erfasst. Mittels der im Anhang 27 gezeigten Formel (übernommen aus Krüke und Möhring (2003)) wurde der BHD_{mit} aus dem Mittendurchmesser geschätzt.

Die im Datensatz erfassten Fälle beinhalten nicht immer alle Arbeitsschritte des motormanuellen Fällens und Aufarbeitens. Es sind auch reines Zufällen für Harvester (wahrscheinlich Situationen, in denen die Rückegassen zu weit auseinander liegen, so dass Bäume nicht mehr vom Harvester erreicht werden können, bzw. Bäume mit eher kleinem Stückinhalt, deren manuelle Aufarbeitung sich nicht lohnen würde) erfasst worden. Diese Fälle können nur durch ihre übermäßig hohe Produktivität identifiziert werden (vergl. Abbildung 3). Für diese Auswertung wurde der Schwellenwert für die Analyse beim EST + 10% (Erweiterter Sorten Tarif, vergleiche Anhang Abbildung 30) gesetzt (vergleiche auch Anhang Abbildung 19 für die Verteilung der BHDs bei den ausgeschlossenen Fällen). Des Weiteren ist nicht bekannt auf wie viele Arbeiter sich die erfassten Zeitaufwände verteilen. Gearbeitet wurde vermutlich mehrheitlich in 2-oder 3-Mann Gruppen (Arbeitssicherheit).

Die entfernten Schläge sind über folgende Regionen und Baumarten verteilt (Tabelle 3):

Tabelle 3: Zuordnung der wegen zu hoher Produktivität entfernten Datensätze

	Gebirge	Huegelland	Sum
Fichte	29	1009	1038
Kiefer	0	5602	5602
Laubbaum	13	5007	5020
Sum	42	11618	11660

Die Zeitaufwände für gemischte Schläge (mehr als eine Baumart) wurden nicht für jede Baumart gesondert gemessen sondern proportional zu den Festmetern aufgeteilt. Damit diese Aufträge trotzdem für die Analysen benutzt werden können, wurde zuerst aus den "reinen" Fichten Schlägen (= nur Baumart Fichte) ein Modell für Fichten erstellt. Dieses Modell konnte anschliessend bei den "Fichte / andere Hauptbaumart" gemischten Schlägen benutzt werden um die Zeiten für Fichte zu berechnen und daraus die Zeit der anderen Baumart zu schätzen (Zeit andere Baumart = gemessene Zeit - geschätzte Zeit für Fichte).

Schliesslich wurden nach einer ersten Modellbildung jene Fälle als Ausreisser identifiziert, deren Residuen (= gemessener Wert - durch das Modell geschätzter Wert) mehr als zwei Standardabweichungen über dem Mittelwert der Residuen liegen. Die Ausreisser sind in allen Regionen vertreten, allerdings mit unterschiedlicher Häufigkeit (vergleiche Tabelle 4).

Tabelle 4 zeigt, dass die Ausreisser in den Regionen Gebirge übervertreten sind

Tabelle 4: Zuordnung der Ausreisser zu den Regionen

	Gebirge	Huegelland
Original	1486	21972
Ausreisser	75 (5%)	858 (4%)

(Mittelwert über alle 4%). 19% (180) dieser 933 Fälle konnte drei Forstbetrieben zugeordnet werden (vergleiche Tabelle 5 und Anhang Tabelle 7).

Tabelle 5: Die drei Forstbetriebe mit den meisten Ausreissern

Forstbetrieb	Fälle	Ausreisser	% Ausreisser
15	406	64	15.8 %
45	551	60	10.9 %
34	536	56	10.4 %

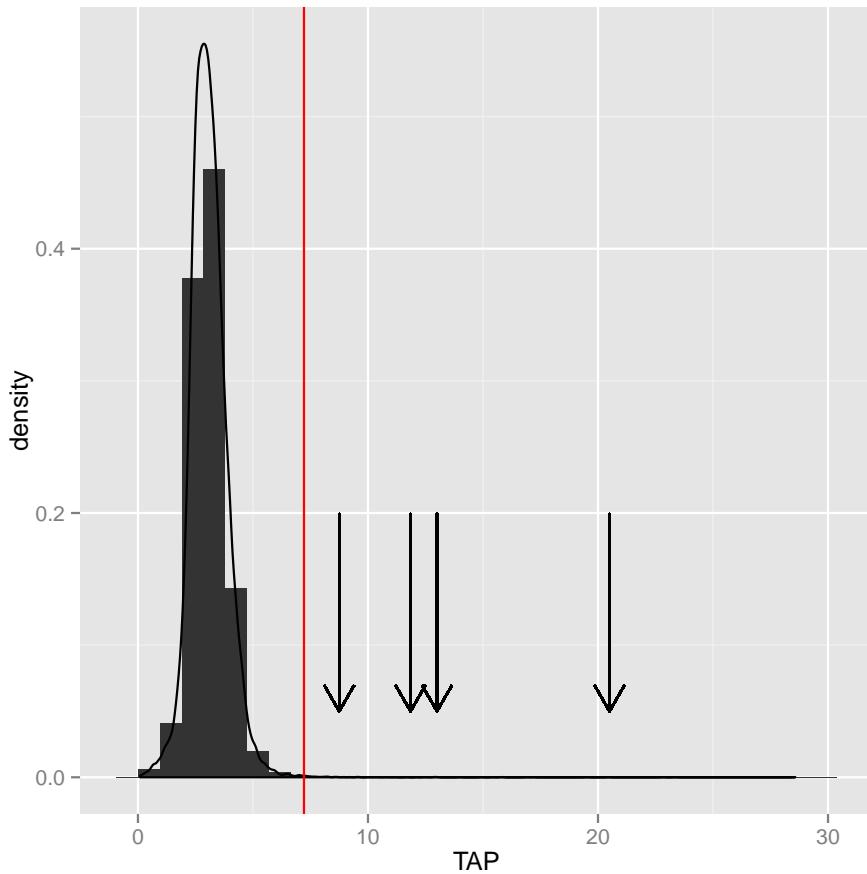


Abbildung 3: DichteVerteilung der Produktivitäten aller erfassten Schläge. Es sind sowohl Schläge mit Aufarbeiten als auch Holzschläge ohne Aufarbeiten (reines Zufällen für Harvester) erfasst , die Pfeile markieren im Datenset erfasste, extrem hohe Produktivitäten.

3.2 Diskussion des Modells

Das in Tabelle 2 gezeigte Modell für die Produktivität der motormanuellen Holzhauerei beinhaltet den durchschnittlichen BHD (BHD_{mit}) sowie den Anteil Laubholz, den Anteil Kiefer (Anteil Laub/ Kiefer- Hölzer am gesamten Volumen liegend) und die Wuchsregion. Abbildung 4 zeigt die gemessenen Produktivitäten und die simulierten Produktivitäten für die Regionen und unterschiedliche Laubholzanteile. Die Grafik zeigt, dass die simulierten Werte versus gemessene Werte auf der Diagonale liegen. Zweitens ist erkennbar, dass die gemessenen sowie die berechneten Produktivitäten in der Region Huegelland höher als in der Region Gebirge sind. Drittens ist der schwache Einfluss des Laubholzanteils erkennbar.

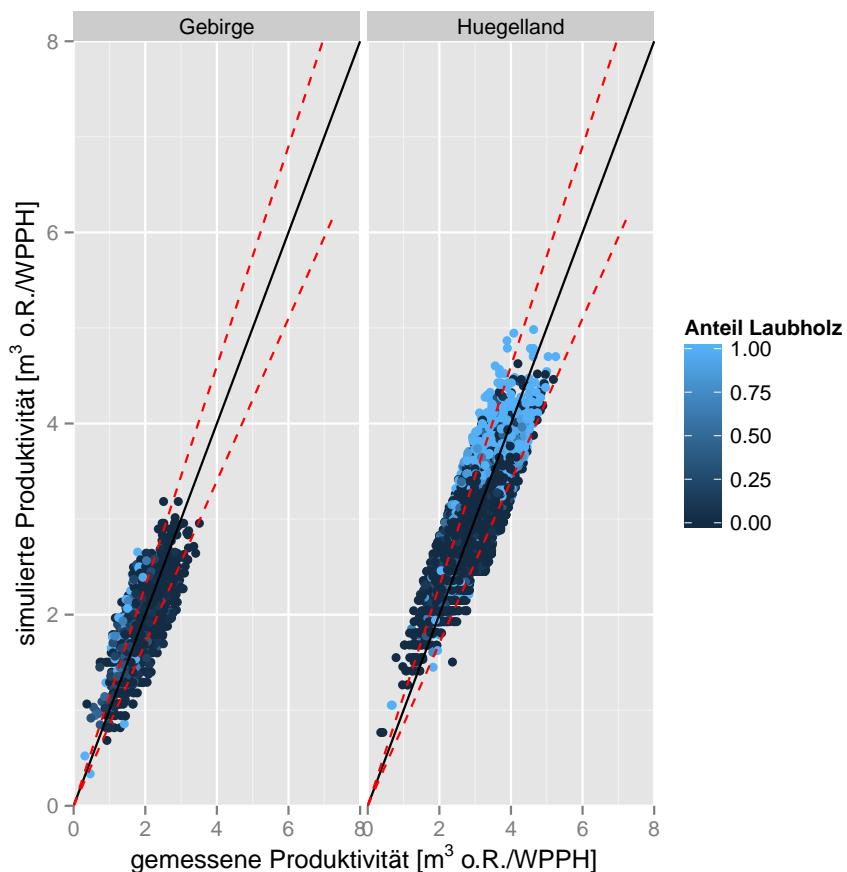


Abbildung 4: Simulierte Produktivität [m³ o.R./WPPH] vs. gemessene Produktivität [m³ o.R./WPPH] für die Regionen Gebirge und Huegelland.

4 Modelle einsetzen

4.1 Formel, Koeffizienten und Systemgrenzen

Die Modelle liefern sinnvolle Resultate für Wertebereiche, in denen sie gut definiert sind (das heisst Bereiche, die von den Trainingsdaten abgedeckt werden, siehe 'Beschreibung der Messgrößen im Anhang').

Der Einsatz der Modelle kann mittels Excel oder anderen Programmen erfolgen. Dabei werden für die kontinuierlichen Variablen (metrisches Skalenniveau) die Koeffizienten mit den Werten der Variable verrechnet, für die kategorialen Variablen (nominales Skalenniveau) die zu der Kategorie gehörenden Werte eingesetzt. Die Formel für 'motormanuelle Holzerei' lautet:

$$\text{Produktivität } \left[\frac{m^3 o.R.}{WPPH} \right] = a + b_{\text{Region}} \\ + c * \ln(\text{BHD [cm]}) \\ + d * \text{Anteil Laubholz [-]} \\ + e * \text{Anteil Kiefer [-]} \\ + f_{\text{Region}} * \ln(\text{BHD}_{\text{mit}} [cm]) \\ + g * \ln(\text{BHD}_{\text{mit}} [cm]) * \text{Anteil Laubholz [-]} \\ + h * \ln(\text{BHD}_{\text{mit}} [cm]) * \text{Anteil Kiefer [-]} \quad (1)$$

Formel 1 zeigt das Modell aus Tabelle 2 in mathematischer Form (die Zeiten in Stunden sind als 'Gesamtarbeitsplatzzeit' (WPSH in Abbildung 26) angegeben). Das Modell umfasst alle Teilprozesse der motormanuellen Holzhauerei. Die Koeffizienten dazu sind:

Tabelle 6: Koeffizienten für Modell Produktivität 'motormanuelles Fällen und Aufarbeiten'. Die Min. und Max. Werte geben die unteren und oberen Grenzen für die einzusetzenden Variablen an, für die das System definiert ist.

Koeff.	Kategorie	Koeffizienten		Parameter	
		Wert	Einheiten	Min.	Max.
a	Konstante	-8.84	[$\frac{m^3 o.R.}{h}$]		
b _{Region}	falls Gebirge	0	[$\frac{m^3 o.R.}{h}$]		
b _{Region}	falls Huegelland	1.56	[$\frac{m^3 o.R.}{h}$]		
c	BHD _{mit}	2.73	[$\frac{m^3 o.R.}{h*ln(cm)}$]	25 cm	100 cm
d	Anteil Laubholz	2.52	[$\frac{m^3 o.R.}{h}$]	0	1
e	Anteil Kiefer	-1.33	[$\frac{m^3 o.R.}{h}$]	0	1
f _{Region}	falls Region Gebirge : BHD _{mit}	0	[$\frac{m^3 o.R.}{h}$]	25 cm	82 cm
f _{Region}	falls Region Huegelland : BHD _{mit}	-0.16	[$\frac{m^3 o.R.}{h}$]	25 cm	100 cm
g	Anteil Laubholz * BHD _{mit}	-0.317	[$\frac{m^3 o.R.}{h*ln(cm)}$]	0	1
h	Anteil Kiefer * BHD _{mit}	0.471	[$\frac{m^3 o.R.}{h*ln(cm)}$]	0	1

4.2 Sensitivitätsanalysen

Die nachfolgenden Abbildungen 5 bis 10 der Sensitivitätsanalyse zeigen die Abhängigkeit der durch das Modell berechneten Produktivität vom BHD_{mit} (schwarze Punkte) für die Regionen und ausgewählte reine Baumartenzusammensetzungen. Die roten Punkte markieren die für die entsprechende Region/Baumart gemessenen Produktivitäten. Die Sensitivitätsanalyse zeigt wie sich die berechneten Werte in Abhängigkeit der Modellparameter verändern.

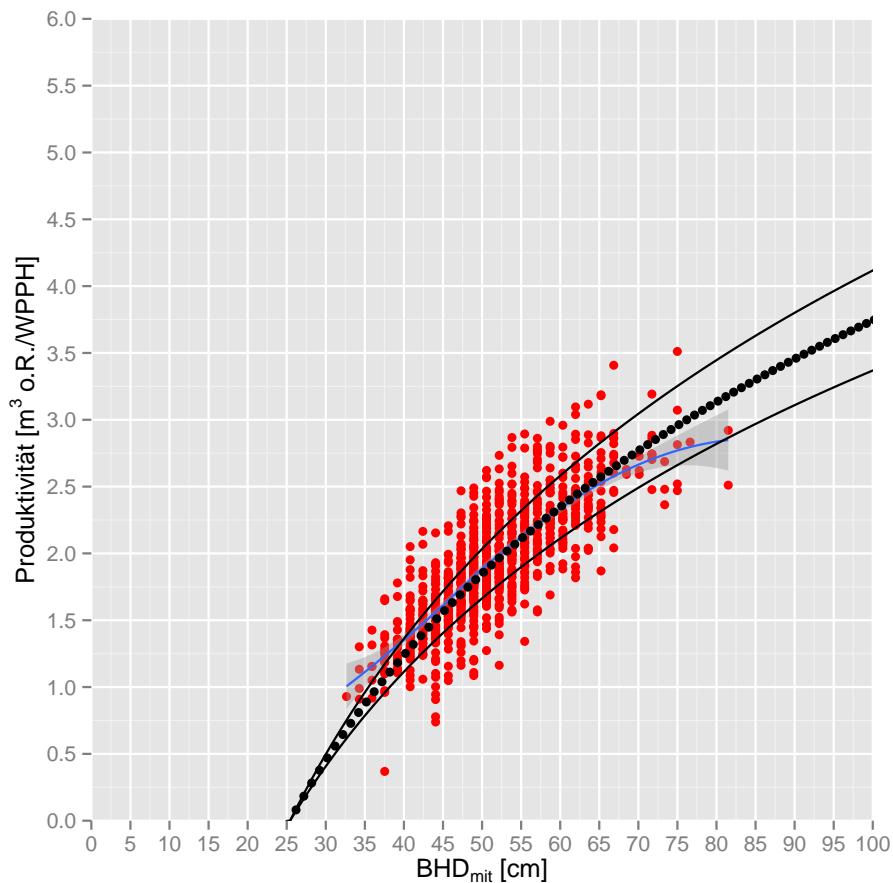


Abbildung 5: Sensitivitätsanalyse für Variable 'BHD' für Nadelbäume in der Region 'Gebirge'.

Abbildung 5 zeigt die Abhängigkeit der durch das Modell berechneten Produktivität vom mittleren BHD für Nadelholz in der Region 'Gebirge'. Die Messpunkte sind rot, die Werte aus dem Modell schwarz eingezeichnet (die blaue Linie steht für die Glättung der Messpunkte, die schwarzen Linien grenzen den Bereich +/- 10% für die Modellwerte ein).

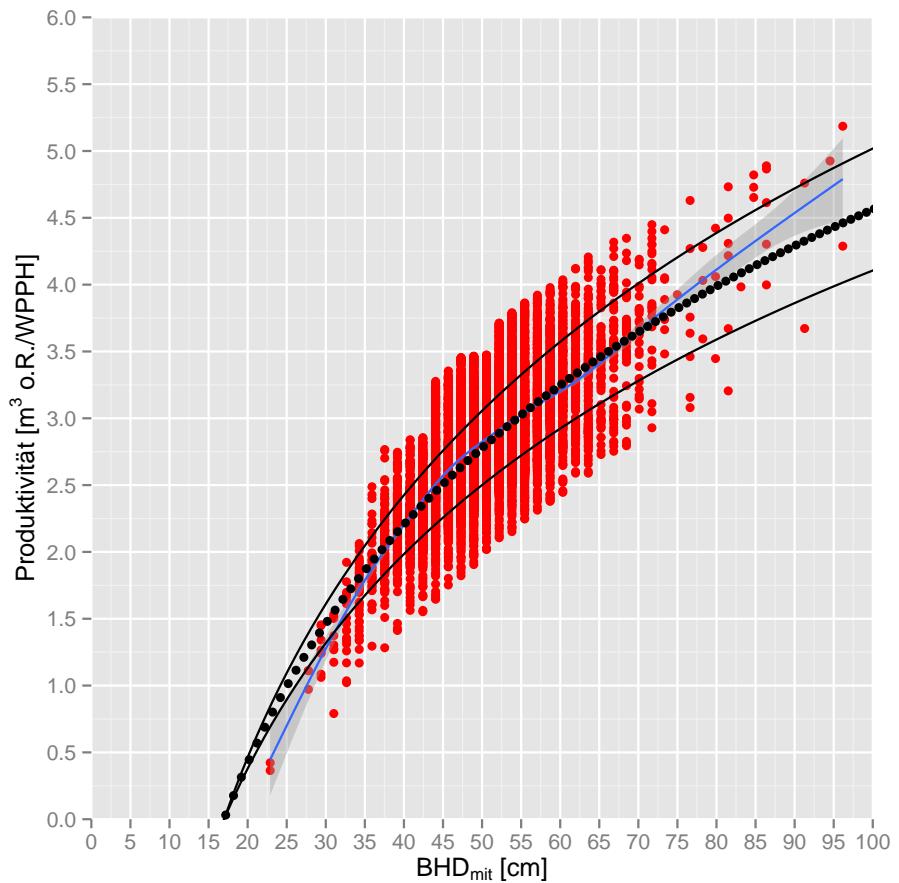


Abbildung 6: Sensitivitätsanalyse für Variable 'BHD' für Nadelbäume in der Region 'Huegelland'.

Abbildung 6 zeigt die Abhängigkeit der durch das Modell berechneten Produktivitäten vom mittleren BHD für Nadelholz in der Region 'Huegelland'. Die Messpunkte sind rot, die Werte aus dem Modell schwarz eingezeichnet (die blaue Linie steht für die Glättung der Messpunkte, die schwarzen Linien grenzen den Bereich +/- 10% für die Modellwerte ein).

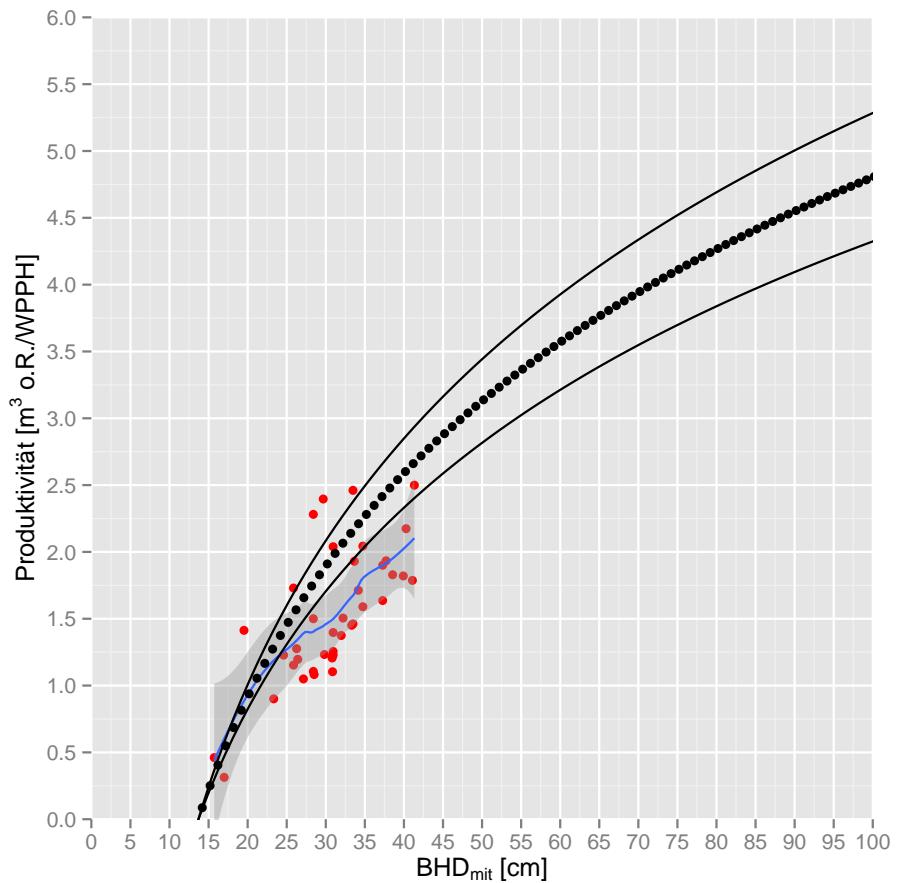


Abbildung 7: Sensitivitätsanalyse für Variable 'BHD_{mit}' für Laubbaum in der Region 'Gebirge'.

Abbildung 7 zeigt die Abhängigkeit der durch das Modell berechneten Produktivitäten vom mittleren BHD für Laubholz in der Region Gebirge. Die Messpunkte sind farbig eingezeichnet (und die blaue Linie steht für die Glättung der Messpunkte), die Werte aus dem Modell sind schwarz eingezeichnet (die schwarzen Linien grenzen den Bereich +/- 10% ein).

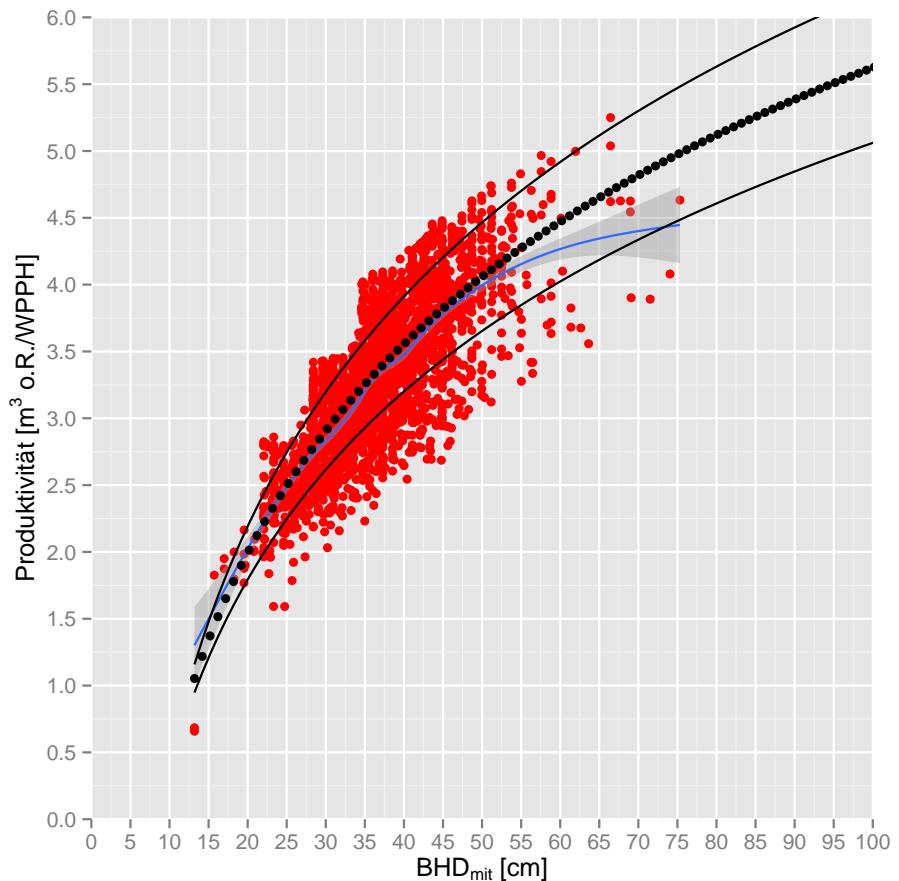


Abbildung 8: Sensitivitätsanalyse für Variable 'BHD_{mit}' für Laubbaum in der Region 'Huegelland'.

Abbildung 8 zeigt die Abhängigkeit der durch das Modell berechneten Produktivitäten von den mittleren BHD für Laubholz in der Region 'Huegelland'. Die Messpunkte sind farbig eingezeichnet (die blaue Linie steht für die Glättung der Messpunkte), die Werte aus dem Modell sind schwarz eingezeichnet (die schwarzen Linien grenzen den Bereich +/- 10% ein).

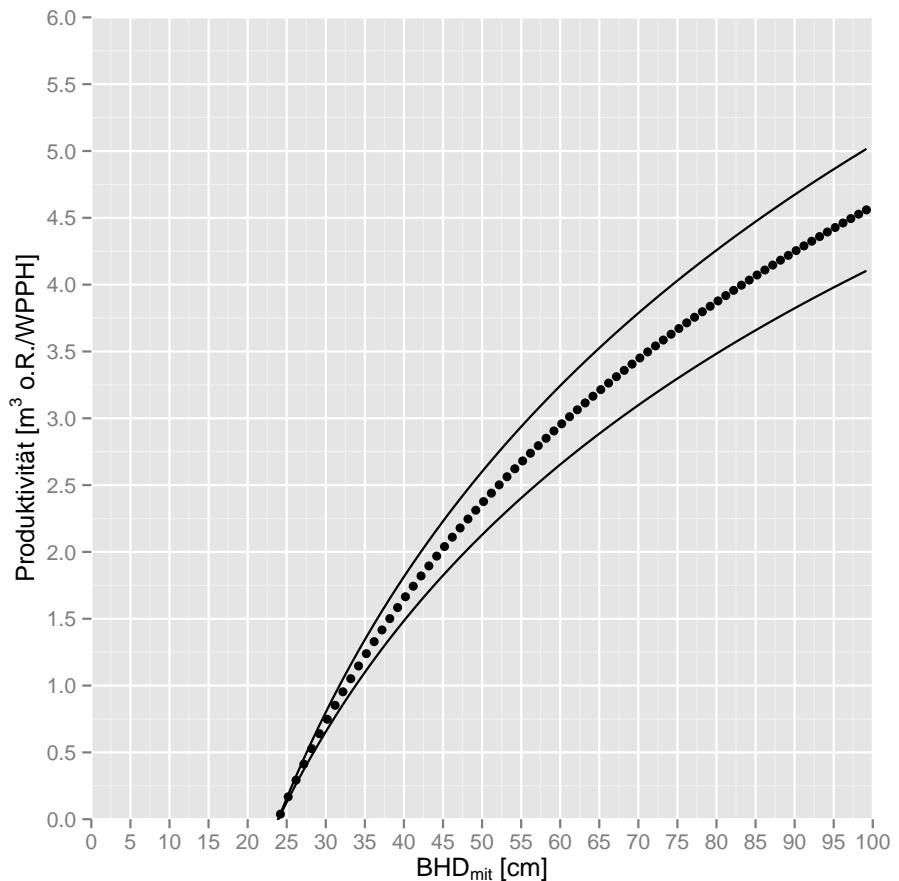


Abbildung 9: Sensitivitätsanalyse für Variable 'BHD_{mit}' für Kiefer in der Region 'Gebirge'.

Abbildung 9 zeigt die Abhängigkeit der durch das Modell berechneten Produktivitäten vom mittleren BHD für Kiefer in der Region 'Gebirge'. Für diese Region hat es zu wenige Daten mit Hauptbaumart Kiefer. Die Werte aus dem Modell sind schwarz eingezeichnet (die schwarzen Linien grenzen den Bereich +/- 10% ein).

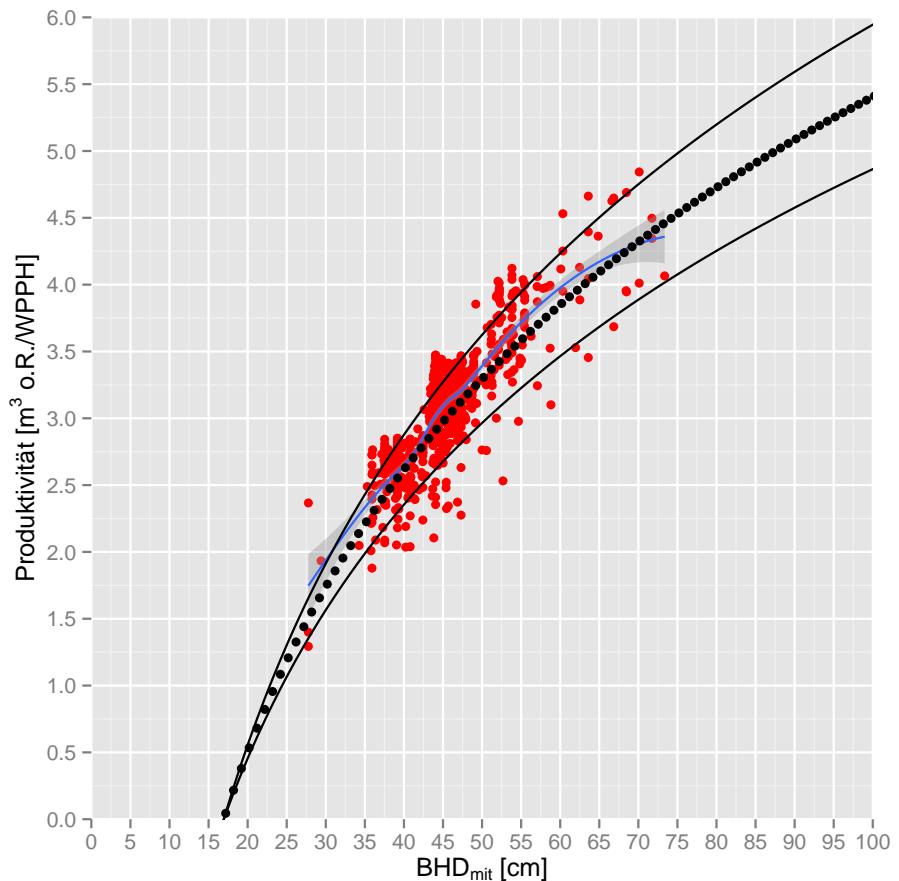


Abbildung 10: Sensitivitätsanalyse für Variable 'BHD_{mit}' für Kiefer in der Region 'Huegelland'.

Abbildung 10 zeigt die Abhängigkeit der durch das Modell berechneten Produktivitäten vom mittleren BHD für Kiefer in der Region 'Huegelland'. Die Messpunkte sind farbig eingezeichnet (die blaue Linie steht für die Glättung der Messpunkte), die Werte aus dem Modell sind schwarz eingezeichnet (die schwarzen Linien grenzen den Bereich +/- 10% ein).

Literatur

- Krüke, B. and Möhring, B. (2003). Rückrechnung des BHD des ausscheidenden Bestandes als Basisinformation des forstwirtschaftlichen Controllings. *Forst und Holz*, 19(58):567 – 570.
- Lemm, R., Erni, V., Frutig, F., Oswald, K., Riechsteiner, D., and Thees, O. (2003). *Produktivitätsmodelle für Verjüngung, Jungwaldpflege und Holzernte mit Hilfe komponentenbasierter Softwaretechnologie. Zusammenstellung der Grundlagen für die Programmierung*. Birmensdorf, Eidg. Forschungsanstalt für Wald, Schnee und Landschaft WSL. Internes Dokument.
- Park, H. and Stefanski, L. (1998). Relative-error prediction. *Statistics & Probability Letters*, 40(3):227 – 236.
- Schweizer, S., (Redaktion) (2011). *Schweizerischer Forstkalender 2012*. Verlag Huber Frauenfeld.
- Thees, O. and Lemm, R. (Hrsg.) (2009). *Management zukunftsähige Waldnutzung: Grundlagen, Methoden und Instrumente*. Birmensdorf, Eidg. Forschungsanstalt für Wald, Schnee und Landschaft WSL, vdf Hochschulverlag. 816 Seiten

5 Anhang

5.1 Testen auf Verletzungen der Annahmen für multiple Regression

Das Modell wurde auf folgende Anforderungen getestet:

1. Modellannahme 1: Es gibt einen Linearen Zusammenhang zwischen gemessenen Werten und mit linearem Modell simulierten Werten (Linearität).
2. Modellannahme 2: Die Residuen sind unabhängig von den abhängigen und den unabhängigen Variablen (sog. Heteroskedastie). Der Erwartungswert der Residuen ist für jeden Wert der Variable = 0 (in den Grafiken durch die blaue Linie (gleitendes Mittel) dargestellt).
3. Modellannahme 3: Falls die Residuen geordnet sind (zBsp. bei Zeitreihen), dürfen sie nicht voneinander abhängig sein (keine Autokorrelation, zBsp. auf grosse Residuen folgen kleine, etc). Da unsere Daten keine natürliche Reihenfolge haben, spielt diese Annahme keine Rolle, (bzw. umfasst das Datenset Daten aus 7 Jahren, aber es ist nur ein geringer Anstieg der Produktivität über die Jahre erkennbar. Außerdem ist der Unterschied der Residuen für die Jahre sehr gering (Vergleiche auch Anhang Abbildung 24 und 25)).
4. Modellannahme 4: Die Residuen sind normalverteilt um den Mittelwert 0.

Diese Anforderungen werden in den nachfolgenden Tests und Grafiken diskutiert. Auf die Anforderung '3: keine Autokorrelation' wird nicht weiter eingegangen.

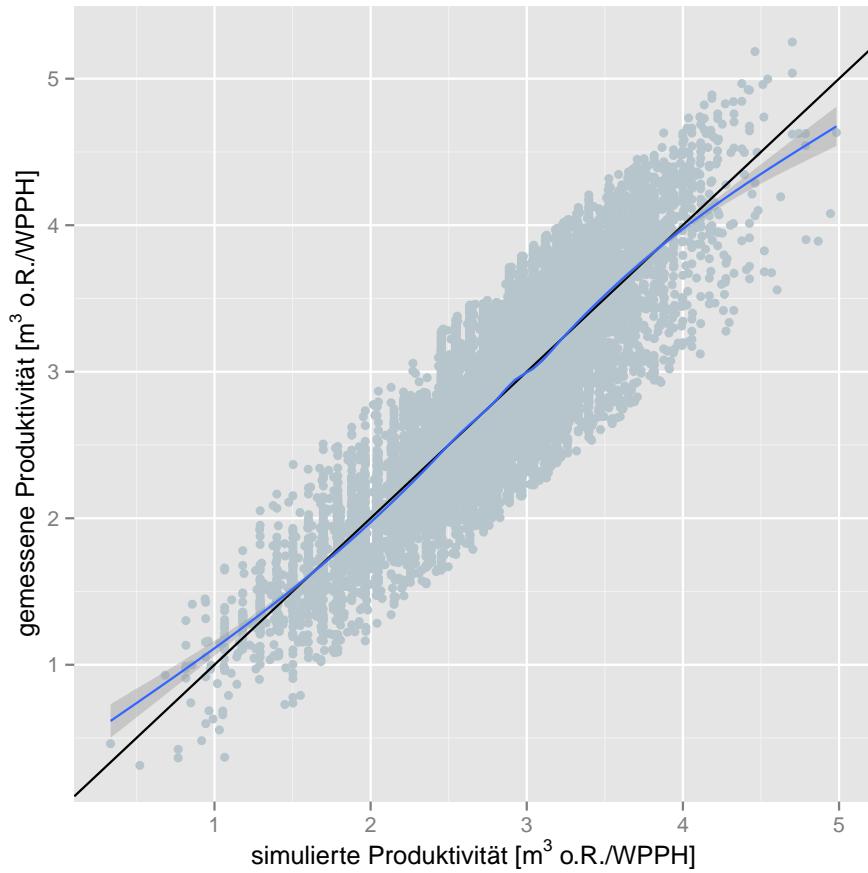


Abbildung 11: Test Modellannahmen 1: Linearität. Die Punkte streuen um die Diagonale, es gibt keine Anzeichen für einen systematischen Fehler.

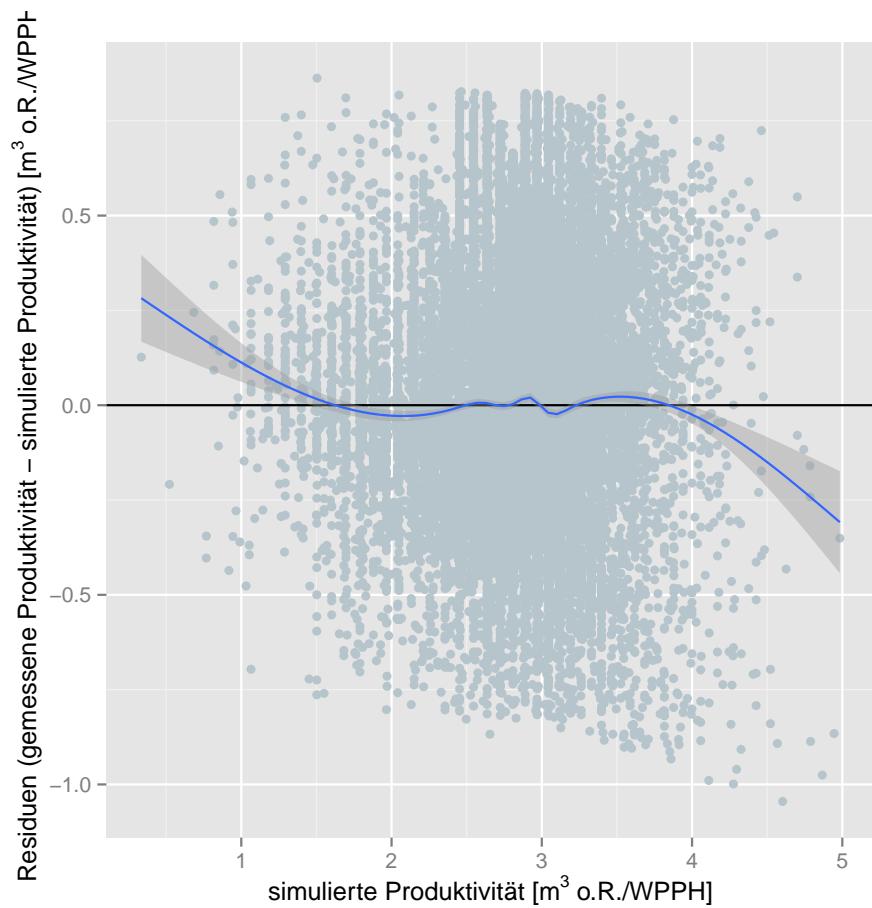


Abbildung 12: Modellannahme 1: Normalverteilung der Residuen. Tukey-Anscombe Plot für das Modell aus Tabelle 1. Der p-Wert der ANOVA der Mittelwertunterschiede der Residuen über die vier Quartile beträgt 0.328

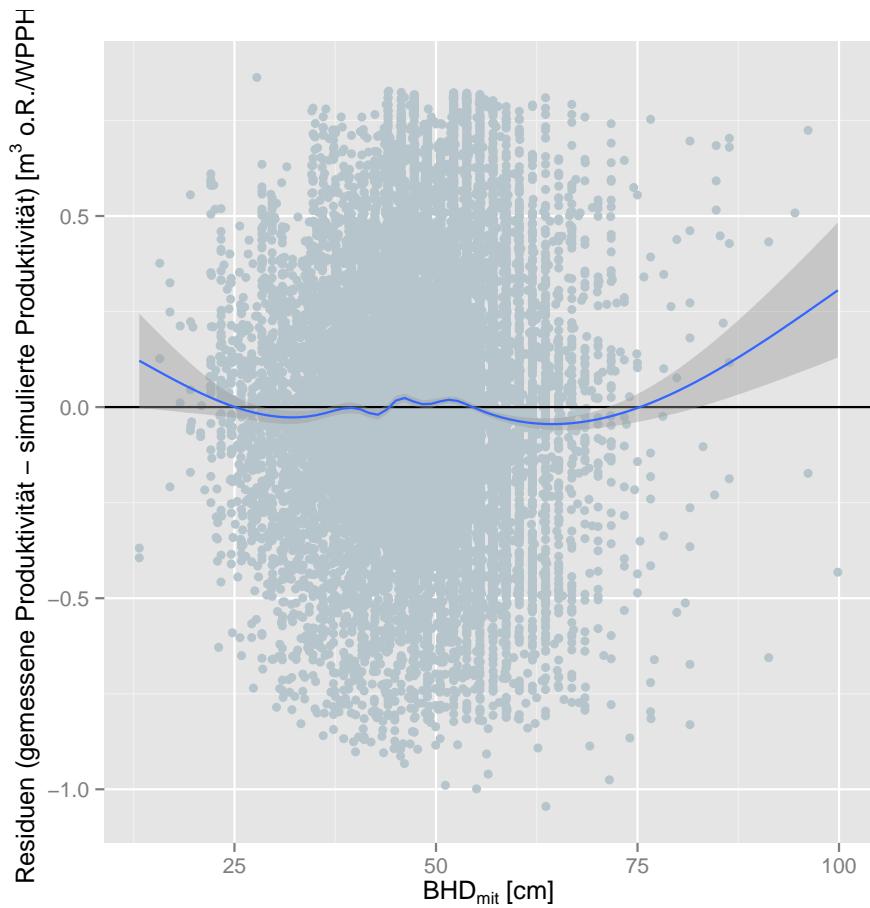


Abbildung 13: Test Modellannahmen 2: Korrelation der Residuen mit unabhängiger Variable 'BHD_{mit}'. Die blaue Linie steht für das gleitende Mittel. Der p-Wert der ANOVA der Mittelwertunterschiede der Residuen über die vier Quartile beträgt 0.734

Abbildung 15 und 17 zeigen, dass für die Variablen "Anteil Laubholz" und "Anteil Kiefer" die Residuen nicht unabhängig sind. Dies kann mit dem im Abschnitt "Diskussion der Datengrundlage" aufgeführten Umstand erklärt werden, dass die Produktivitäten für gemischte Schläge nicht exakt gemessen worden sind. Für die Variable "Anteil Laubholz" ist die Abhängigkeit der Residuen nur durch einen kleinen Peak bei geringe Laubholzanteilen (< 10%) erklärbar.)

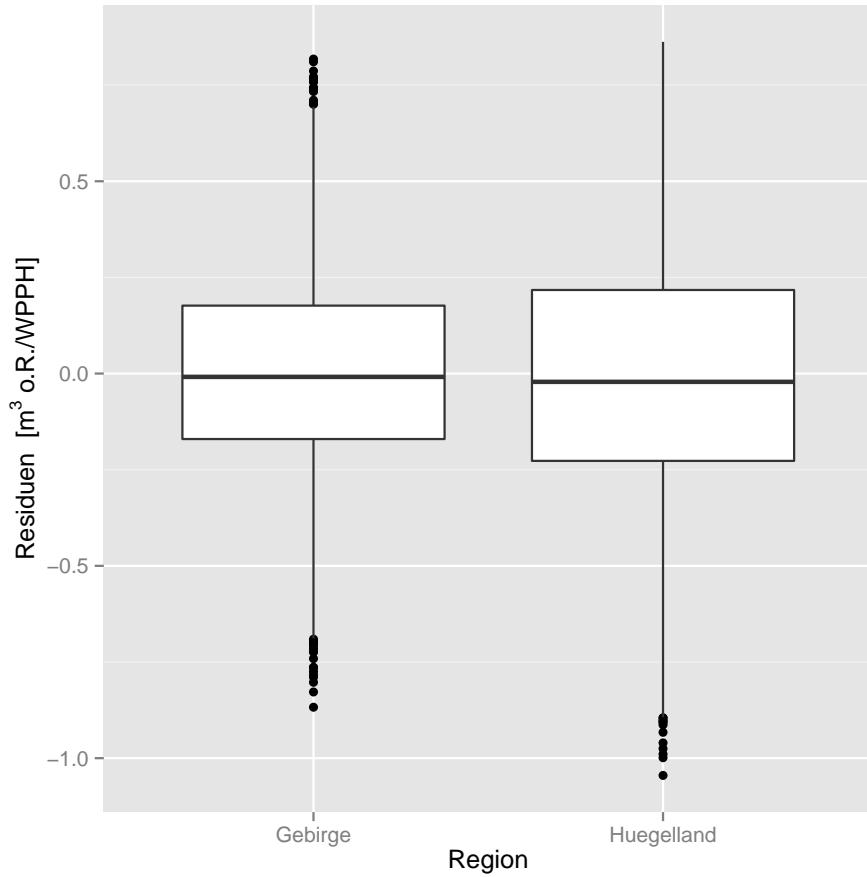


Abbildung 14: Test Modellannahmen 2: Korrelation der Residuen mit unabhängiger Variable 'Region'. Die weisse Box des Boxplots umfasst mindestens 50% der Fälle (die fette Linie markiert den Median), die vertikalen Linien sind maximal 2.5 x so lang wie die weisse Box und reichen vom tiefsten zum grössten gemessenen Wert. Messpunkte die ausserhalb dieser Grenze liegen werden als Ausreisser-Punkte eingezeichnet.

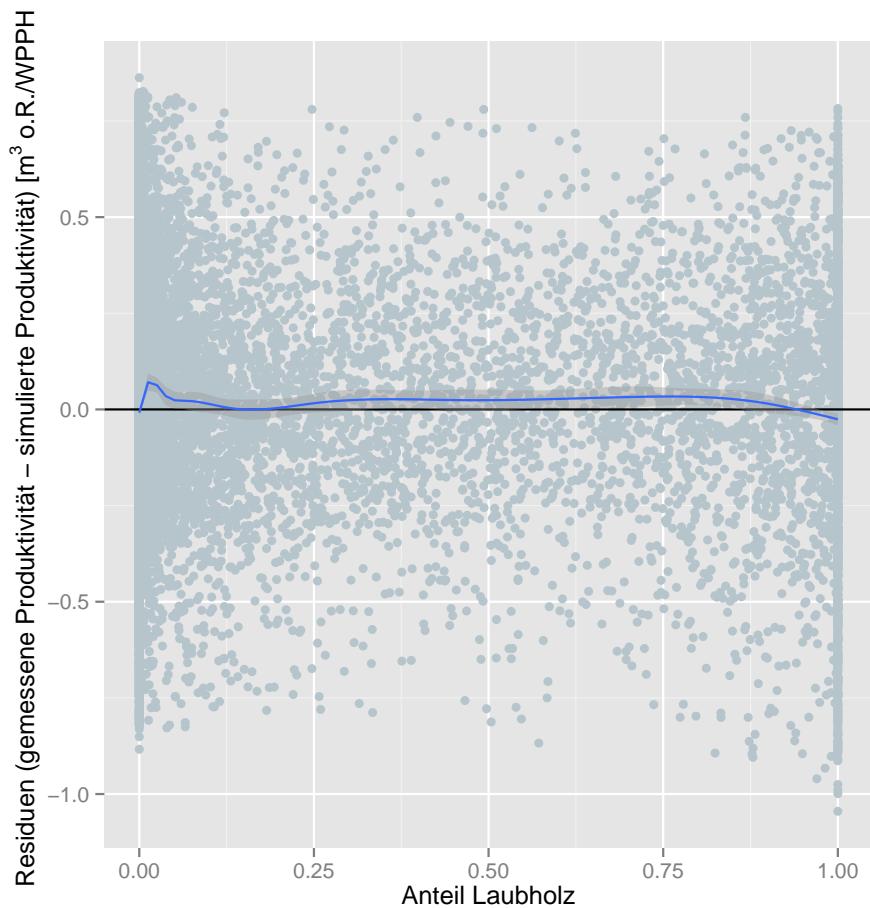


Abbildung 15: Test Modellannahmen 2: Korrelation der Residuen mit unabhängiger Variable 'Anteil Laubholz'. Die blaue Linie steht für das gleitende Mittel. Der p-Wert der ANOVA der Mittelwertunterschiede der Residuen über die vier Quartile beträgt < 0.001.

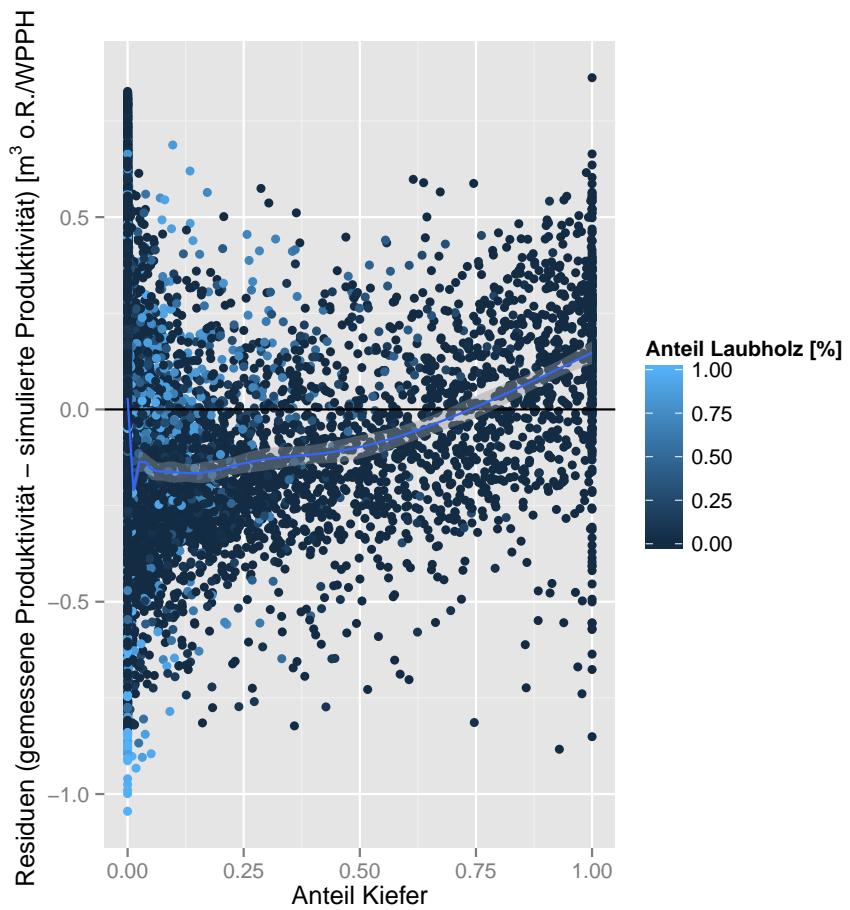


Abbildung 16: Test Modellannahmen 2: Korrelation der Residuen mit unabhängiger Variable 'Anteil Kiefer'. Die blaue Linie steht für das gleitende Mittel. Der p-Wert der ANOVA der Mittelwertunterschiede der Residuen über die vier Quartile beträgt < 0.001.

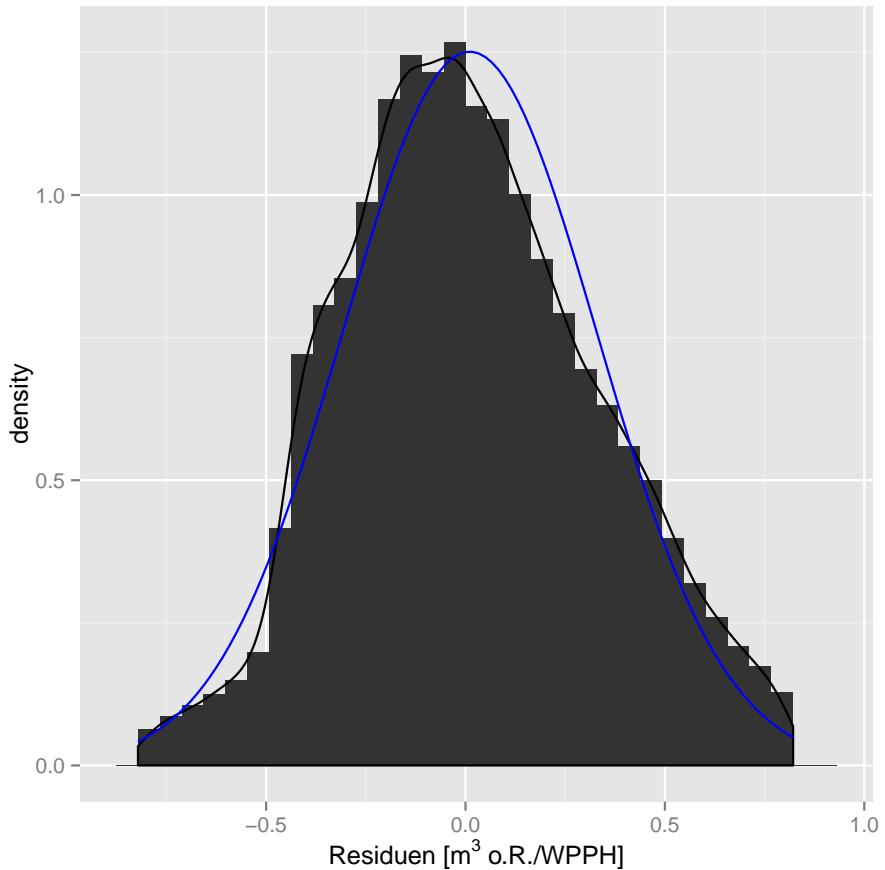


Abbildung 17: Test Modellannahmen 4: Normalverteilung der Residuen: Dichtefunktion der Residuen. Als Referenz ist die Dichtefunktion einer normalverteilten Variable blau eingezeichnet. Die Residuen sind annähernd normalverteilt.

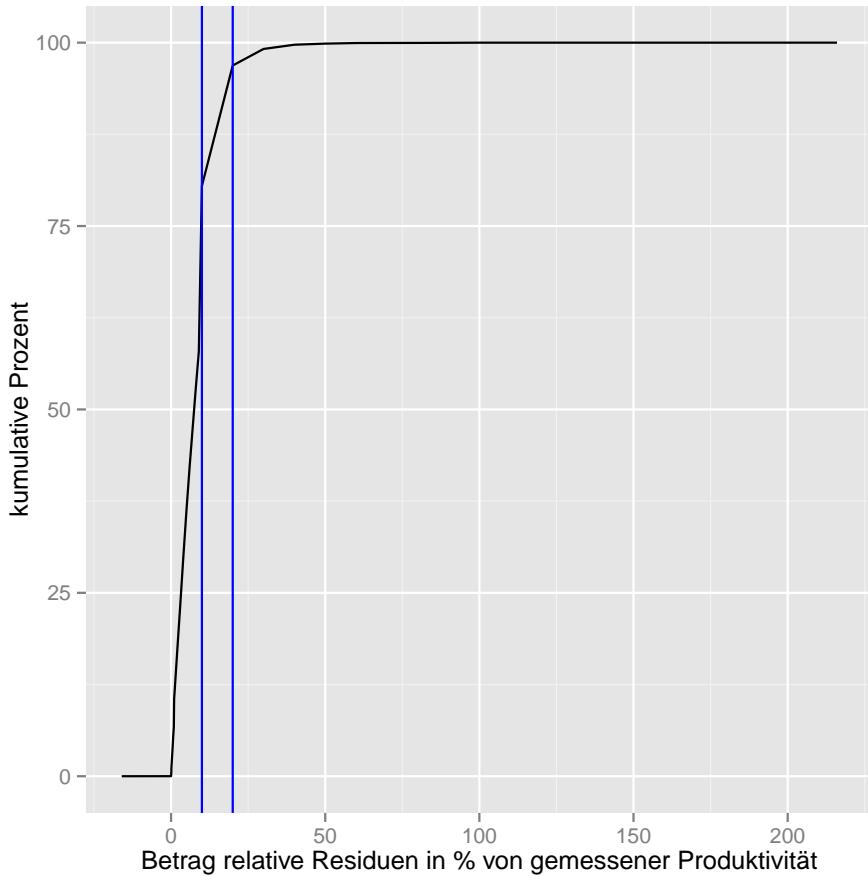


Abbildung 18: Kumulative Verteilung der Beträge der relativen Residuen. Die vertikalen blauen Linien markieren 10% bzw. 20% Abweichung. Bei 80% der Fälle ist der relative Schätzfehler kleiner gleich 10%.

5.2 Beschreibung der ungültigen Fälle wegen Zufällen für Harvester

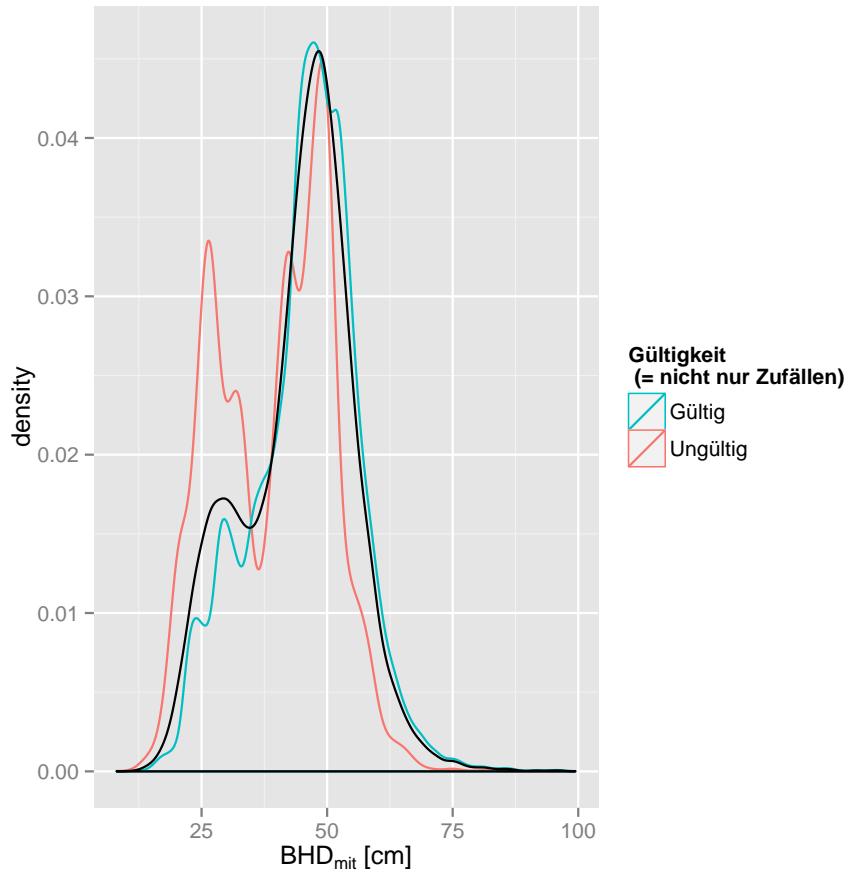


Abbildung 19: Vergleich der Verteilung der BHDs für die gültigen und ungültigen Fälle (die schwarze Linie zeigt die Verteilung in der Grundgesamtheit). Bäume mit kleineren BHDs sind überproportional oft nur zugefallen worden.

5.3 Beschreibung der 'Ausreisser'

Tabelle 7: Zuordnung der Anzahl Ausreisser zu den Organisationseinheiten

Forstbetrieb	Fälle	Ausreisser	Prozent Ausreisser
15	406	64	16
45	551	60	11
34	536	56	10
30	390	39	10
29	491	46	9
35	514	47	9
26	811	67	8
42	547	44	8
50	431	27	6
14	354	22	6
48	540	31	6
49	556	31	6
28	615	34	6
41	530	28	5
24	822	41	5
38	692	32	5
13	607	27	4
10	529	23	4
11	503	21	4
12	589	22	4
23	391	13	3
25	503	16	3
27	636	18	3
39	527	15	3
33	489	12	2
22	561	10	2
37	582	9	2
31	1000	13	1
16	730	9	1
19	432	5	1
20	894	11	1
32	841	9	1
44	941	10	1
18	203	2	1
40	702	6	1
17	446	3	1
21	631	3	0
36	580	3	0
43	591	2	0
47	633	2	0
51	131	0	0

5.4 Kreuzvalidierung

Die Variablen sowie die benutzten Fälle für das Trainings- und das Validierungsset werden im Abschnitt 'Beschreibung der im Modell benutzten Messgrößen' beschrieben (Zeilen 'Trainings- und Validierungsdaten').

Tabelle 8: Resultat der Analyse des linearen Zusammenhangs und Schätzung der Modellkoeffizienten für den Trainingsdatensatz

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-8.894	0.215	-41.336	0.000
RegionHuegelland	1.577	0.223	7.075	0.000
log(BHD)	2.746	0.055	49.677	0.000
pctLbh	2.528	0.153	16.566	0.000
pctKiefer	-1.577	0.364	-4.335	0.000
RegionHuegelland:log(BHD)	-0.165	0.057	-2.881	0.004
log(BHD):pctLbh	-0.318	0.041	-7.762	0.000
log(BHD):pctKiefer	0.538	0.095	5.668	0.000

Tabelle 8 gibt die Koeffizienten-Schätzungen für das Trainingsdatenset wieder. Das korrigierte R² für die Modellbildung beträgt 0.68.

Tabelle 9: Resultat der Analyse des linearen Zusammenhangs und Schätzung der Modellkoeffizienten für den Trainingsdatensatz - Unterschiede der Koeffizientenschätzungen

	p-Wert	alle Daten	trainings Daten	% Diff
(Intercept)	0.000	-8.844	-8.894	-0.6
RegionHuegelland	0.000	1.556	1.577	-1.3
log(BHD)	0.000	2.733	2.746	-0.4
pctLbh	0.000	2.521	2.528	-0.3
pctKiefer	0.000	-1.327	-1.577	-18.8
RegionHuegelland:log(BHD)	0.001	-0.160	-0.165	-3.1
log(BHD):pctLbh	0.000	-0.317	-0.318	-0.4
log(BHD):pctKiefer	0.000	0.471	0.538	-14.1

Tabelle 9 zeigt die Unterschiede für die Koeffizientenschätzung aus allen Daten (N= 22525) vs. Koeffizientenschätzung aus Trainingsdaten (N= 15756). Der p-Wert bezieht sich auf die Schätzung aus allen Daten.

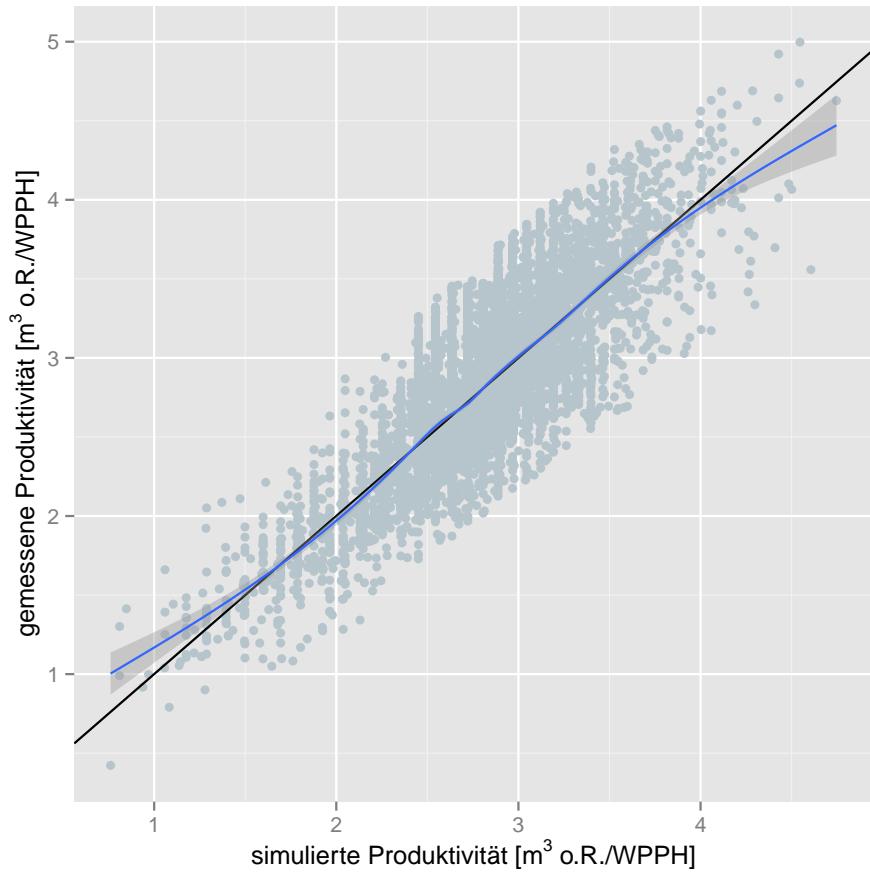


Abbildung 20: Der Korrelationskoeffizient zwischen simulierter Produktivität [$\text{m}^3 \text{ o.R./WPPH}$] und gemessener Produktivität [$\text{m}^3 \text{ o.R./WPPH}$] für das Validierungsset beträgt 0.81.

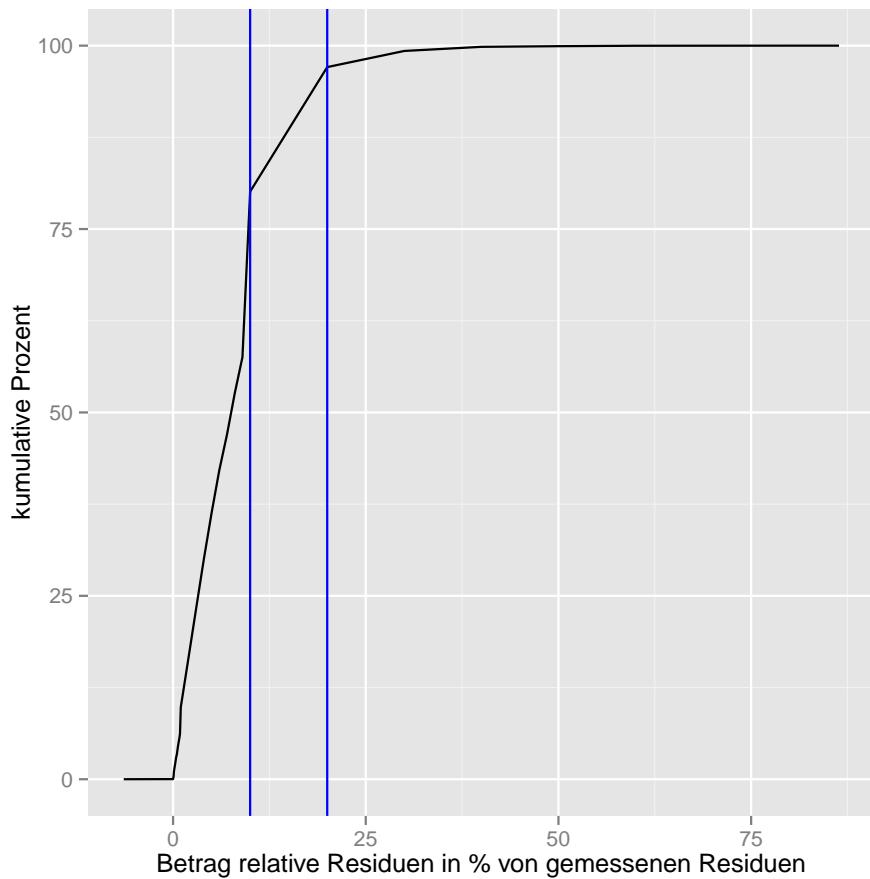


Abbildung 21: Kumulative Verteilung der Beträge der relativen Residuen aus der Kreuzvalidierung. Die vertikalen blauen Linien markieren 10% bzw. 20% Abweichung.

5.5 Beschreibung der im Modell benutzten Messgrössen

Für die nachfolgenden Tabellen gelten folgende Abkürzungen: n = Anzahl gültiger Datensätze für diese Variable, mean = Mittelwert, sd = Standardabweichung, median = Median, min = kleinster Wert, max = grösster Wert, skew = Schiefe (wie stark neigt die Verteilung nach links (+) oder rechts (-)), se = Standardfehler des Mittelwerts.

Tabelle 10: Beschreibung der Produktivität [m^3 o.R./WPPH]

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	22525	2.82	0.55	2.80	0.31	5.25	-0.04	0.00
Trainigsdaten	15756	2.82	0.56	2.80	0.31	5.25	-0.04	0.00
Validierungsdaten	6769	2.83	0.55	2.81	0.42	5.00	-0.02	0.01

	Gebirge	Huegelland
mean TAP	1.88	2.88
sd TAP	0.46	0.50

Tabelle 11: Beschreibung der Produktivitäten der Regionen [m^3 o.R./WPPH]

Tabelle 12: Beschreibung der Variable 'Kubikmeter ohne Rinde' [m^3 o.R.] der Daten

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	22525	459.97	371.04	369.11	0.05	4540.35	1.82	2.47
Trainigsdaten	15756	457.50	368.90	367.29	0.05	4540.35	1.81	2.94
Validierungsdaten	6769	465.71	375.96	373.35	0.09	4258.75	1.83	4.57

Tabelle 13: Beschreibung der Variable mittlerer 'BHD' [cm] der Daten

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	22525	48.08	8.01	48.81	13.19	99.85	-0.12	0.05
Trainigsdaten	15756	48.09	8.02	48.73	13.19	99.85	-0.07	0.06
Validierungsdaten	6769	48.05	7.99	48.92	18.26	86.39	-0.22	0.10

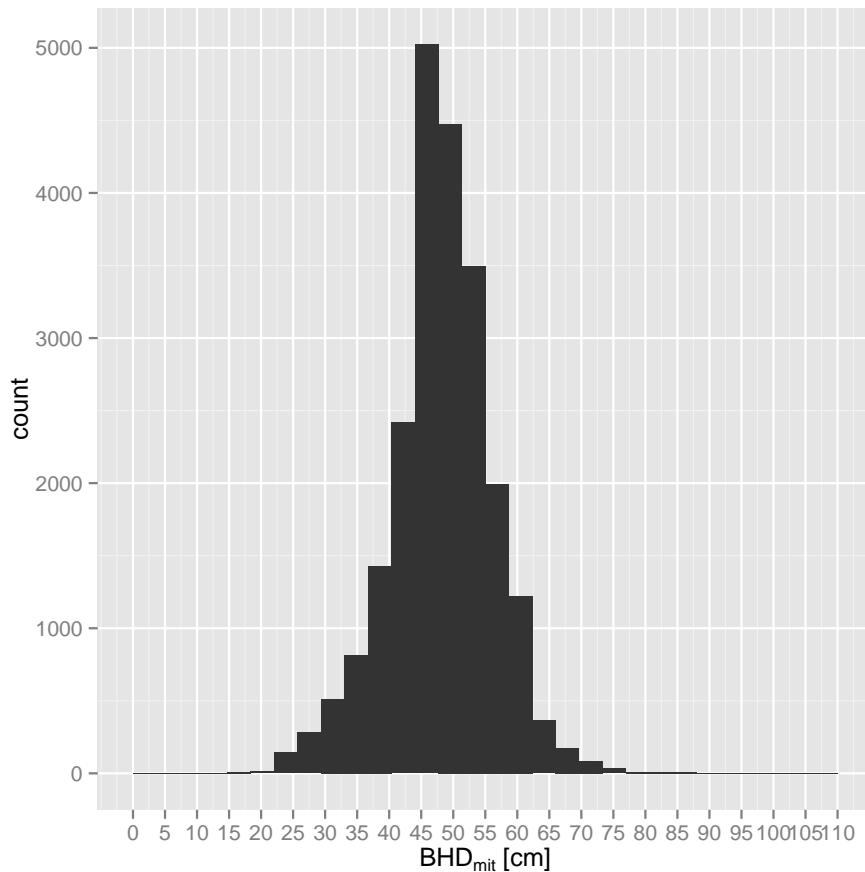


Abbildung 22: Häufigkeiten der BHD_{mit} im Datensatz. Unter BHD_{mit} = 30cm und über BHD_{mit} = 70cm sind fast keine Daten vorhanden.

Die Variable 'Region' wurde aus einer Clusteranalyse von Wuchsbezirksdaten abgeleitet. Dazu wurden Informationen aus 32 Gebieten benutzt (durchschnittliche Höhe über Meer, durchschnittliche Hangneigung, durchschnittliche Höhe der Fichten im Alter zwischen 75 und 85.). Die Ergebnisse dieser Clusteranalyse sind in Abbildung 23 dargestellt. Es wurden zwei Clusterlösungen (bzw. 'Gruppierungs-Höhen') getestet.

Tabelle 14: Beschreibung (Häufigkeiten) der Variable 'Region'

	Gebirge	Huegelland
Alle Daten	1411	21114
Trainingsdaten	992	14764
Validierungsdaten	419	6350

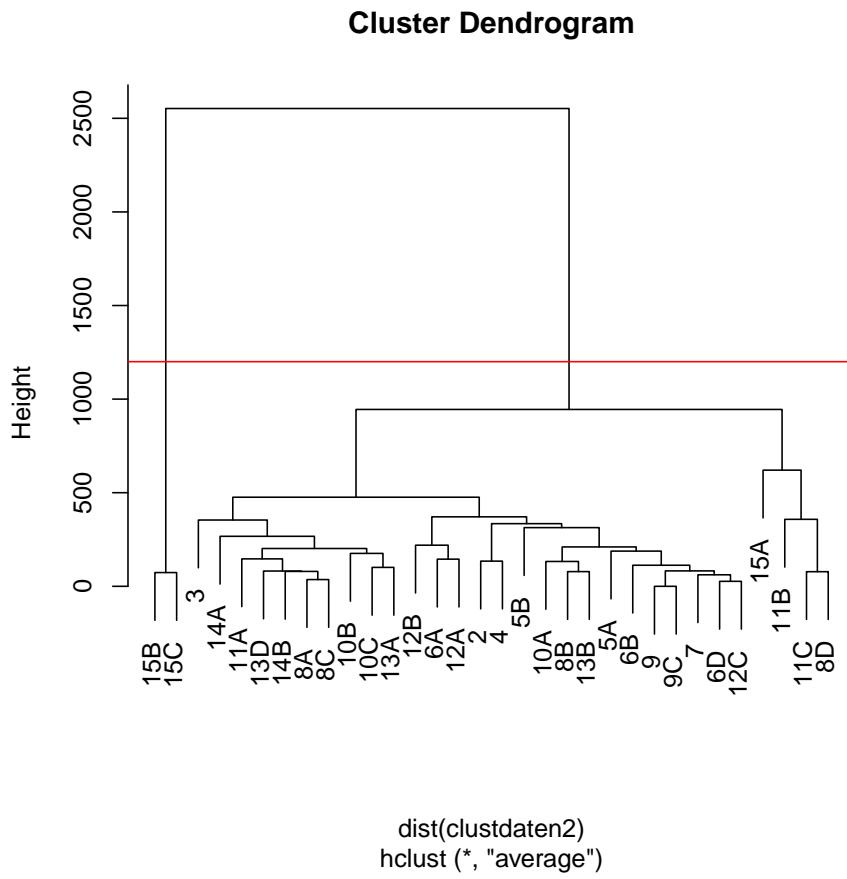


Abbildung 23: Baumdiagramm der Clusteranalyse der 32 Bezugsleistungsgruppen nach den Variablen 'Meter über Meer', 'Hangneigung in Grad' (jeweils: mean, sd, min, max) und Baumhöhe der Bäume im Alter 75 bis 85 (Baumhöhe für Fichte). Die rote Linie markiert die 'Gruppierungs Höhe', die für die Regionen-Gruppen Bildung benutzt wurde. Die 'Gruppierungs Höhe' wurde auf 1200 festgelegt da die so gebildeten Gruppen klar voneinander getrennt sind (siehe auch Tabelle 15). Tests mit 'Gruppierungs Höhen' von 900 und kleiner haben Gruppen ergeben, die zu wenig differenziert waren.

Tabelle 15: Mittelwerte Eigenschaften der Regionen

	MuM	Neig.	Fi_h80	Bu_h80	Fi_Vor.	Bu_Vor.	Ki_Vor.	Ei_Vor.
Gebirge	1124.73	21.73	253.78	236.44	432.18	274.04	140.47	231.17
Huegelland	527.36	5.33	285.17	259.43	458.74	281.56	376.38	310.58

Tabelle 15 beschreibt die beiden Regionen mit einigen Kennzahlen. Die Abkürzungen stehen für: Meter über Meer (MuM), Hangneigung in Grad (Neig.), Höhe m der Fichten (75 - 85 Jahre) (Fi_h80), Höhe m der Buchen (75 - 85 Jahre) (Bu_h80), Fichten Vorrat in m^3 o.R./ha (Fi_Vor.), Buchen Vorrat in m^3 o.R./ha (Bu_Vor.), Kiefer Vorrat in m^3 o.R./ha (Ki_Vor.), Eichen Vorrat in m^3 o.R./ha (Ei_Vor.).

Tabelle 16: Häufigkeiten der Hauptbaumarten (Anteil > 0.8) in den Regionen
(andere = kein Baumart Anteil > 0.8)

	Gebirge	Huegelland
andere	135	5910
Fichte	1237	12370
Kiefer	0	701
Laubholz	39	2133

Tabelle 17: Beschreibung der Variable 'Stunden'

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	22525	164.64	130.62	134.00	0.03	1781.50	1.90	0.87
Trainingsdaten	15756	163.97	129.79	133.00	0.03	1465.58	1.85	1.03
Validierungsdaten	6769	166.18	132.53	136.00	0.06	1781.50	2.00	1.61

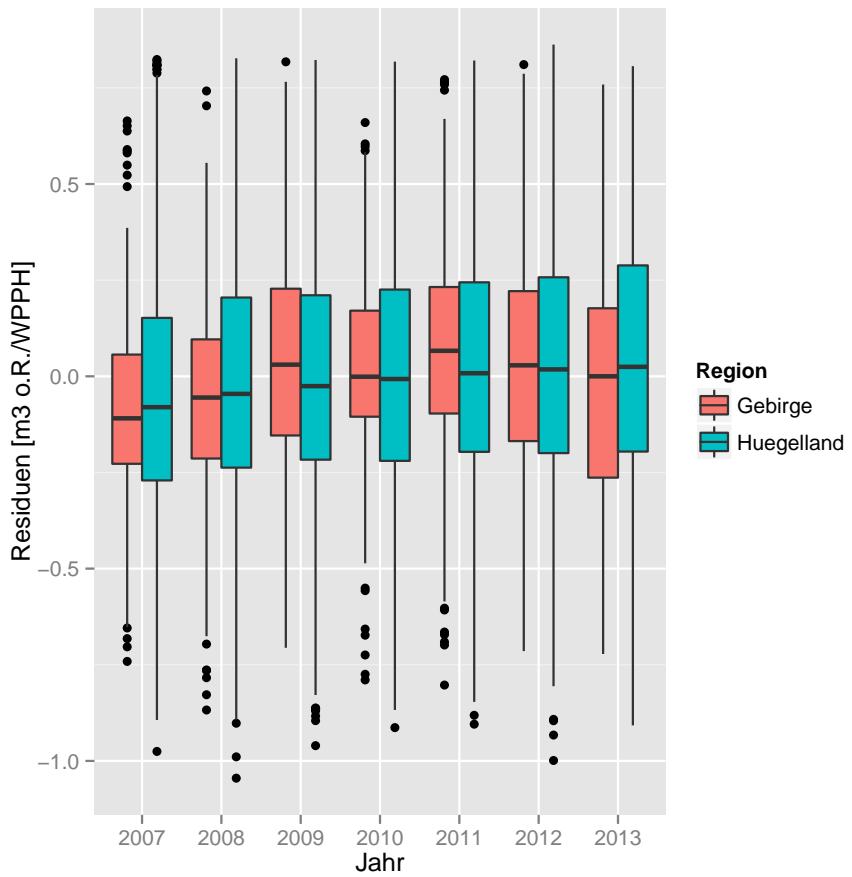


Abbildung 24: Mittlere Produktivität über die Jahre. Es gibt nur eine geringfügige Änderung der Produktivität.

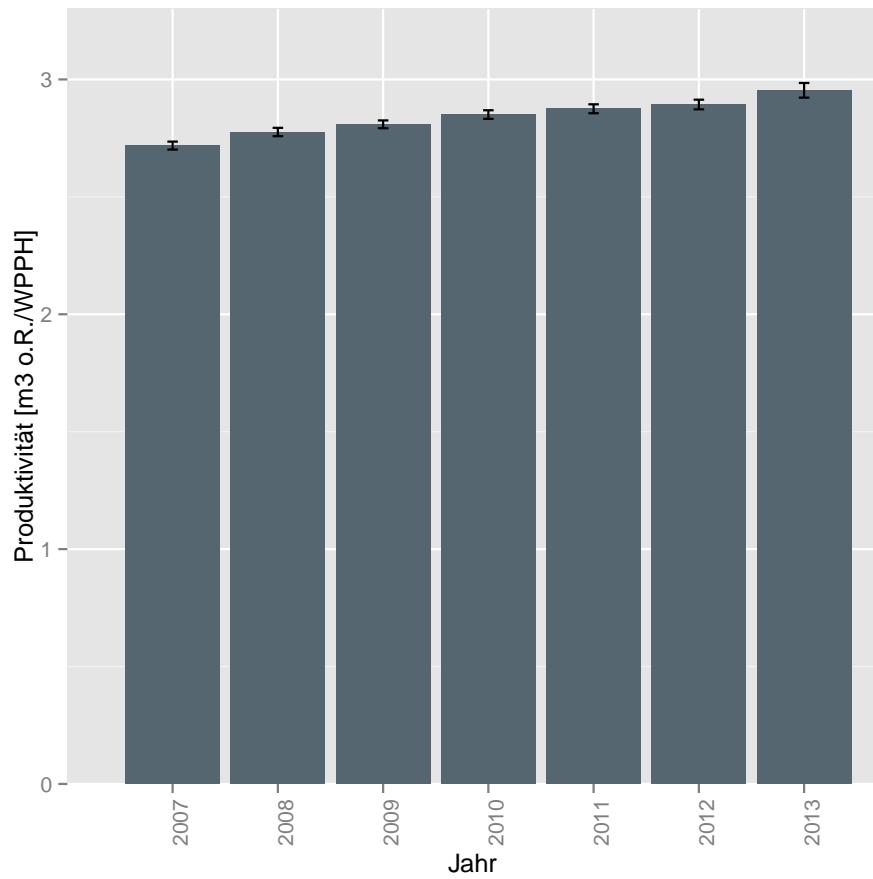


Abbildung 25: Mittlere Residuen aus der Modell-Schätzung über die Jahre. Es gibt nur eine geringfügige Änderung der Residuen. Der maximale Unterschied der gemittelten Produktivität beträgt 9% der tiefsten Produktivität.

Tabelle 18: Beschreibung der Variable 'Anteil Laubholz'

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	22525	0.15	0.32	0.00	0.00	1.00	1.93	0.00
Trainingsdaten	15756	0.15	0.31	0.00	0.00	1.00	1.96	0.00
Validierungsdaten	6769	0.16	0.32	0.00	0.00	1.00	1.86	0.00

Tabelle 19: Beschreibung der Variable 'Anteil Kiefer'

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	22525	0.07	0.20	0.00	0.00	1.00	3.46	0.00
Trainingsdaten	15756	0.07	0.20	0.00	0.00	1.00	3.39	0.00
Validierungsdaten	6769	0.06	0.19	0.00	0.00	1.00	3.63	0.00

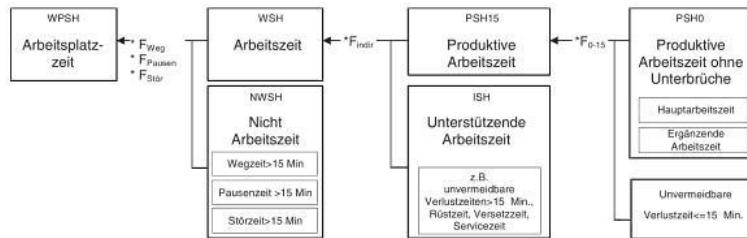


Abbildung 26: Zeitsystem (Übernommen aus Thees and Lemm (2009))

```

function(baumart, RD, Ant=0.1){
  # berechnet BHD in cm aus Rollendurchmesser in cm des ausscheidenden Stamms
  # Parameter:
  # baumart = Baumart
  # RD = Rollendurchmesser in cm
  # Ant = Anteil x-Holz (0 - 1)
  return(bhd)
}

bhd <- ifelse(baumart %in% c("Fichte", "Tanne", "Kiefer"), 4.34312 + 1.62839 * RD - 10 * 0.1, 5.620097 + 1.2679 * RD - 17.82271

```

Abbildung 27: Modell zur Berechnung des BHD aus dem Rollendurchmesser (übernommen aus Krüke und Möhring (2003)).
x-Holz = nicht verwertbares Holz

Tabelle 20: Berechnungsbeispiele

Region	Mittendurchmesser [cm]	BHD [cm]	Anteil Laubholz [%]	Anteil Kiefer [%]	Baumart	Prod. o.R./WPPH	Prod. [m ³ /WPPH]	simuliert [m ³]	Vmit
Gebirge	25	44.05	0	0	Fichte	1.45	1.50	1.74	
Huegelland	25	44.05	0	0	Fichte	2.77	2.45	1.74	
Gebirge	30	52.19	0	0	Fichte	1.79	1.97	2.54	
Huegelland	30	52.19	0	0	Fichte	2.98	2.89	2.54	
Huegelland	33	57.08	0	0	Fichte	3.43	3.12	3.09	
Huegelland	28	48.94	0	0	Fichte	3.23	2.72	2.21	
Huegelland	30	52.19	0	0	Fichte	3.23	2.89	2.54	
Huegelland	27	47.31	0	0	Fichte	2.91	2.64	2.05	
Huegelland	27	47.31	0	0	Fichte	3.44	2.64	2.05	
Huegelland	25	44.05	0	0	Fichte	2.87	2.45	1.74	
Huegelland	27	47.31	0	0	Fichte	2.24	2.64	2.05	
Huegelland	32	55.45	0	0	Fichte	3.31	3.04	2.90	
Huegelland	29	50.57	0	0	Fichte	3.30	2.81	2.37	
Huegelland	32	55.45	0	0	Fichte	3.14	3.04	2.90	
Huegelland	29	50.57	0	0	Fichte	3.11	2.81	2.37	
Gebirge	37	63.59	0	0	Fichte	2.27	2.51	3.87	

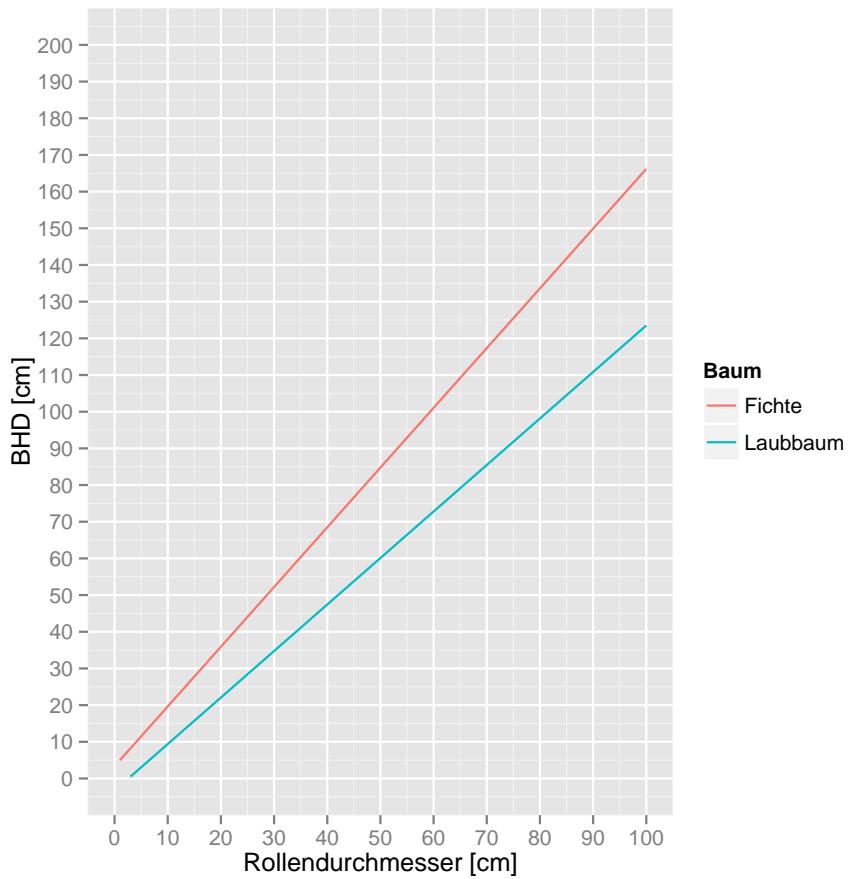


Abbildung 28: Der Zusammenhang zwischen Rollendurchmesser des liegenden Stammes und dessen BHD (nach Krüke and Möhring (2003)) für die beiden Baumarten-Gruppen (Fichte = Fichte, Kiefer und Tanne).

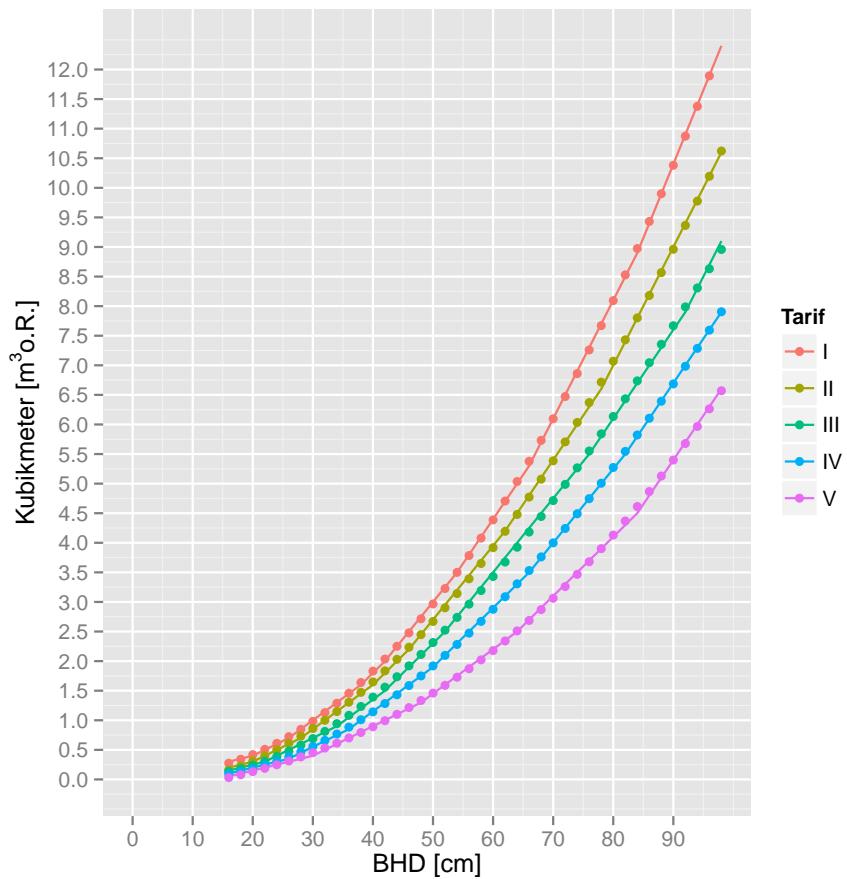


Abbildung 29: Verwendeter Tarif (Luzerner Tarif I bis V zur Berechnung des Volumens pro Baum aus dem BHD (Schweizer (2011))).

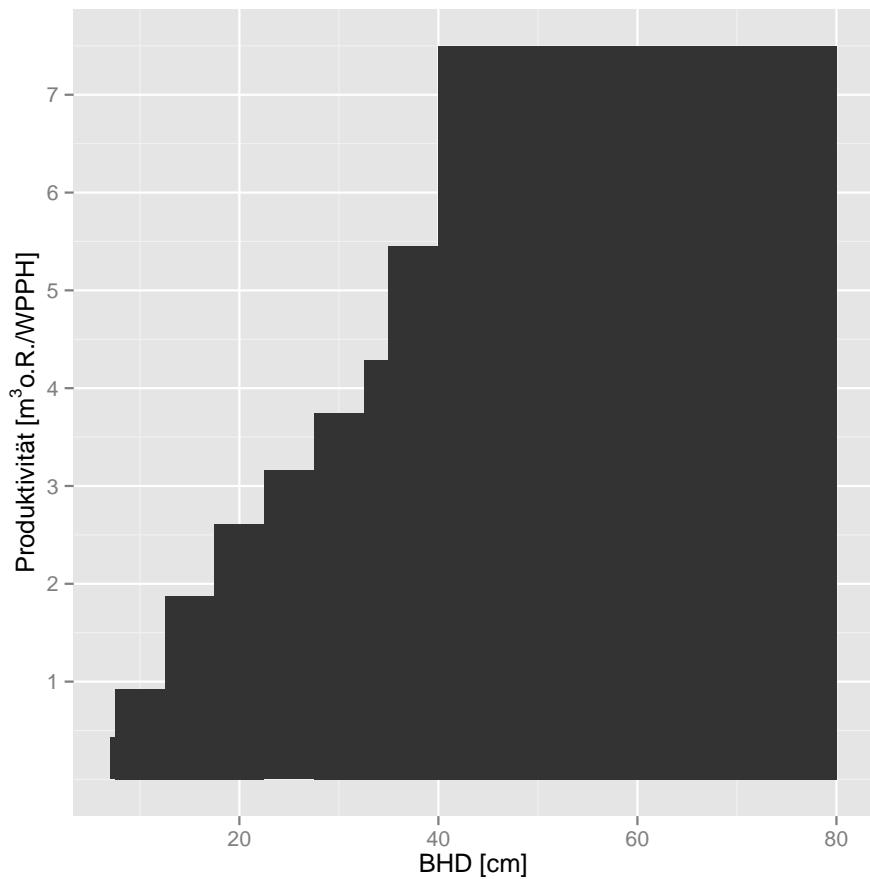


Abbildung 30: Der Erweiterte Sorten Tarif EST - Stammholz Fixlängen, abge-längt (AL) 3.51m - 6m, ohne Zuschläge.

5.6 Regionen vs. Hangneigung des Reviers

Zur Beschreibung der Standorte stehen neben der Variable Region (2 Ausprägungen: Gebirge und Hügelland) auch Informationen zu den Revieren (Hangneigung und Höhenlage) zur Verfügung. Wie Abbildungen 31 und 32 zeigen, gruppieren die Variable 'Region' die Hangneigung (bzw. Höhenlage) sehr gut. Der Einbezug der mittleren Hangneigung (bzw. Höhenlage) der Reviere würde das angepasste R-Quadrat von 0.67 auf 0.69 erhöhen. Dieser Gewinn an Erklärungskraft ist zu klein als dass sich die Erhöhung der Komplexität des Modells lohnen würde.

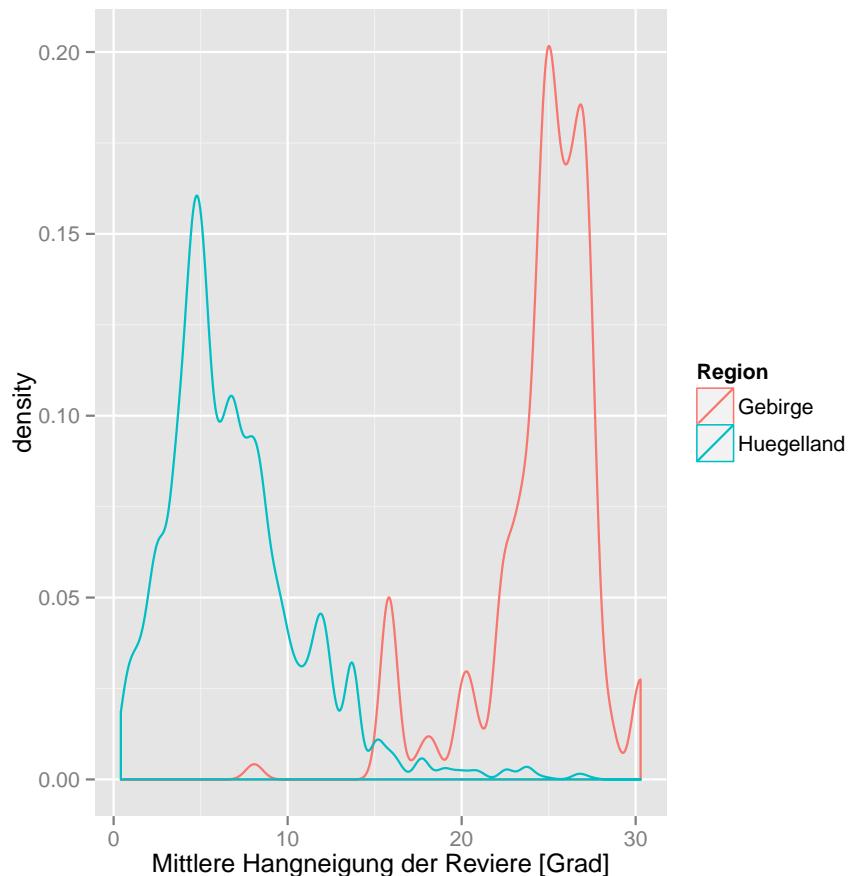


Abbildung 31: Verteilung der mittleren Hangneigung der Reviere für die beiden Regionen. Die Regionen gruppieren die Hangneigungen in zwei sinnvolle Gruppen.

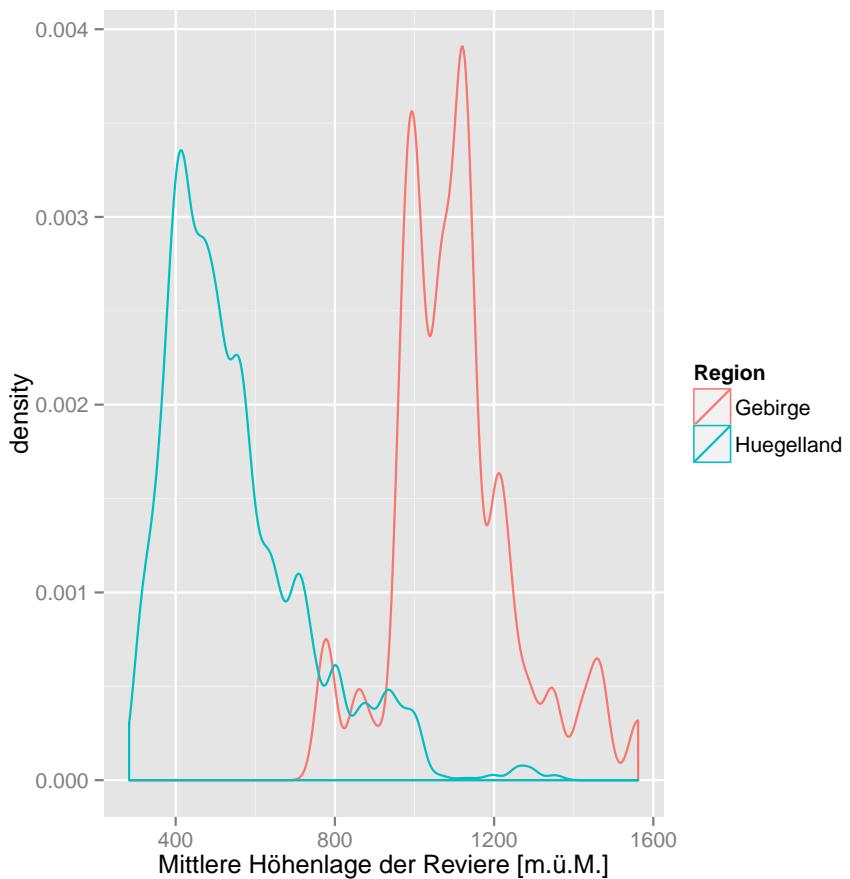


Abbildung 32: Verteilung der mittleren Höhenlage der Reviere für die beiden Regionen. Die Regionen gruppieren die Höhenlage in zwei sinnvolle Gruppen.

5.7 Erweiterung des Modells für kleine und grosse BHD_{mit}

Da das Modell im Bereich $BHD_{mit} < 35\text{cm}$ und $BHD_{mit} > 90\text{cm}$ (bzw. 75 im Gebirge) nicht durch gemessene Werte gestützt werden kann (vergleiche Verteilung der BHDs in den Daten, Abbildung 22), wurden für diese Bereiche Sonderlösungen implementiert. Für Schläge mit $BHD_{mit} > 90\text{cm}$ (im Gebirge ab BHD 75cm) ist die Produktivität konstant. Für Schläge mit $BHD_{mit} < 35\text{cm}$ wird die Produktivität linear aus der Produktivität bei $BHD_{mit} = 35\text{cm}$ und dem Ursprung extrapoliert. Dieses Erweiterte Modell ist für BHD_{mit} von 12cm bis 100cm definiert.

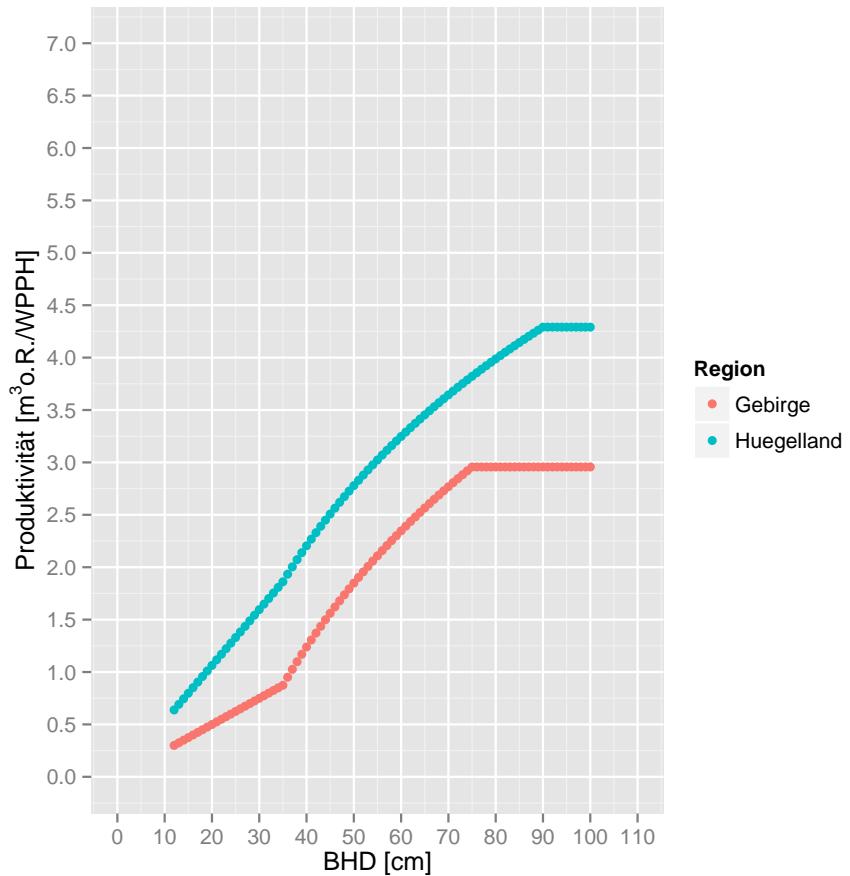


Abbildung 33: Beispiel für Produktivitätsberechnung mit Speziallösungen für grosse ($> 90\text{cm}$ im Hügelland, $> 75\text{cm}$ im Gebirge) und kleine ($< 35\text{cm}$) BHD_{mit}. Anteil Laubholz = 0 und Anteil Kiefer = 0.

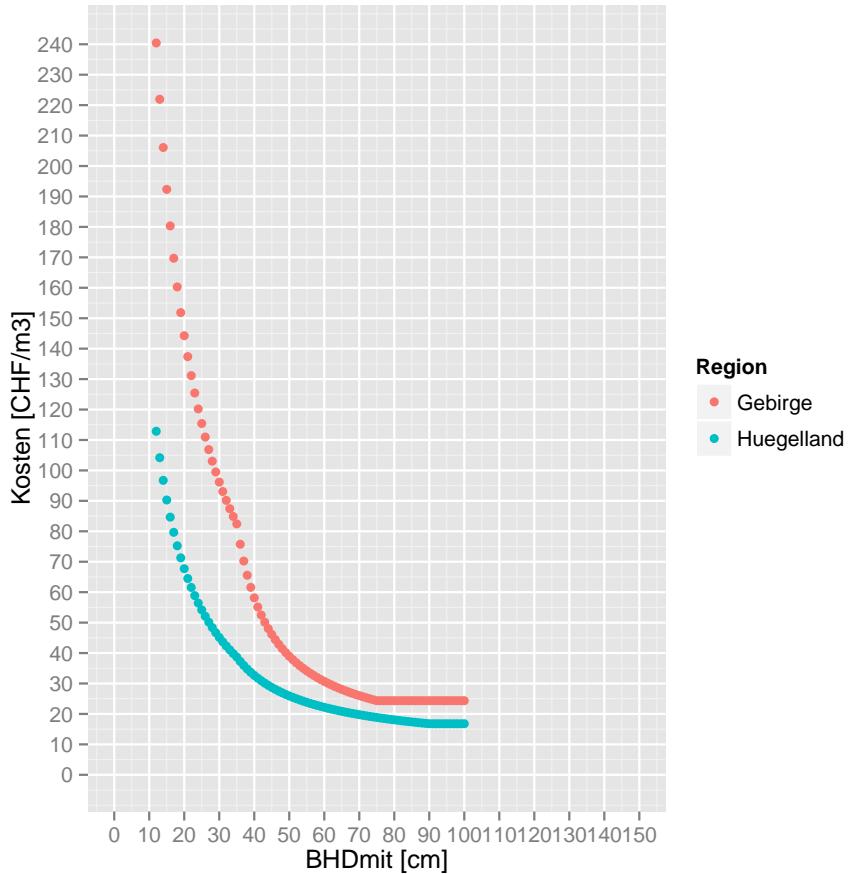


Abbildung 34: Beispiel für Kostenberechnung. Stundenansatz Personal inkl. Motorsäge (48% Maschinenlaufzeit) CHF 72.- pro Stunde. Anteil Kiefer = 0 und Anteil Laubholz = 0.

Tabelle 21: Berechnungsbeispiele und Testfälle

BHD	Region	pctLbh	pctKiefer	simProd
35.00	Hnegelland	0.00	0.00	1.86
35.00	Huegelland	0.00	0.00	1.86
10.00	Huegelland	0.00	0.00	0.53
15.00	Hnegelland	0.00	0.00	0.80
100.00	Huegelland	0.00	0.00	4.29
35.00	Huegelland	0.50	0.00	2.56
35.00	Huegelland	0.00	0.20	1.93
35.00	Hnegelland	0.20	0.20	2.21
35.00	Gebirge	0.00	0.00	0.87

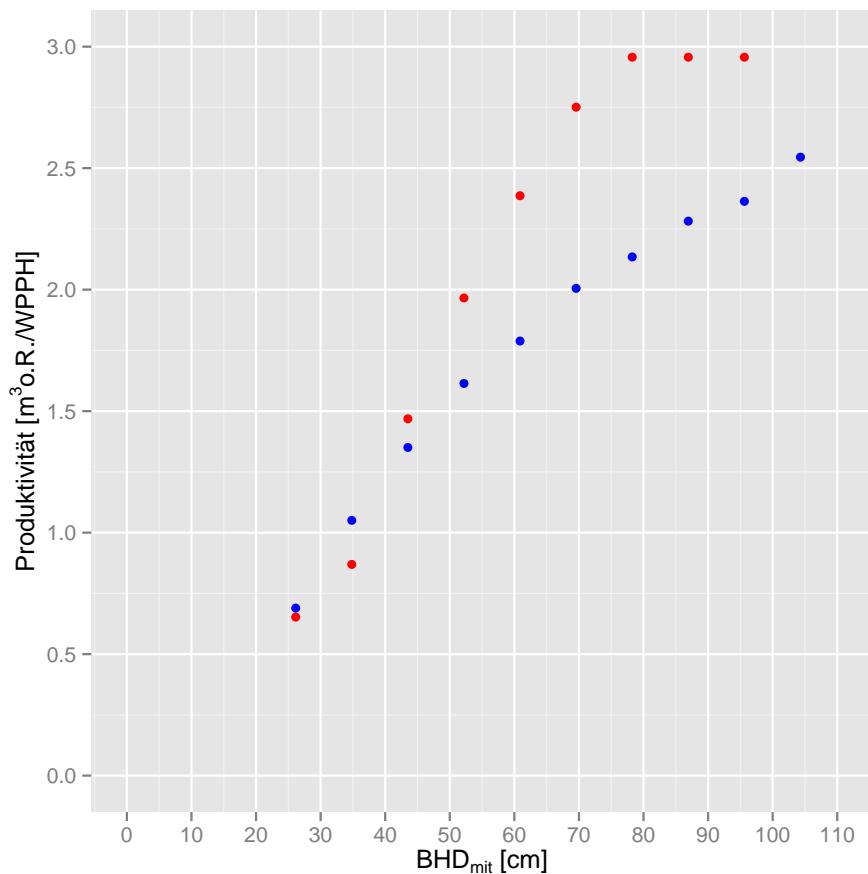


Abbildung 35: Vergleich Schätzungen des Modells mmNeu mit dem alten He-ProMo Motormanuell (nachfolgend: Modell mmAlt): **Gebirge**. Die roten Punkte markieren Schätzungen des Modells mmNeu für Fichte im Gebirge (Anteil Laubholz = 0%, Anteil Kiefer = 0%), die blauen Punkte markieren Schätzungen des Modells mmAlt (Fichte, bis 50% Hangneigung, Hindernisse = mässig, 100% Stammholz, 4- 6m Länge). Die Produktivitäten des mmAlt wurden auf WPSH umgerechnet.

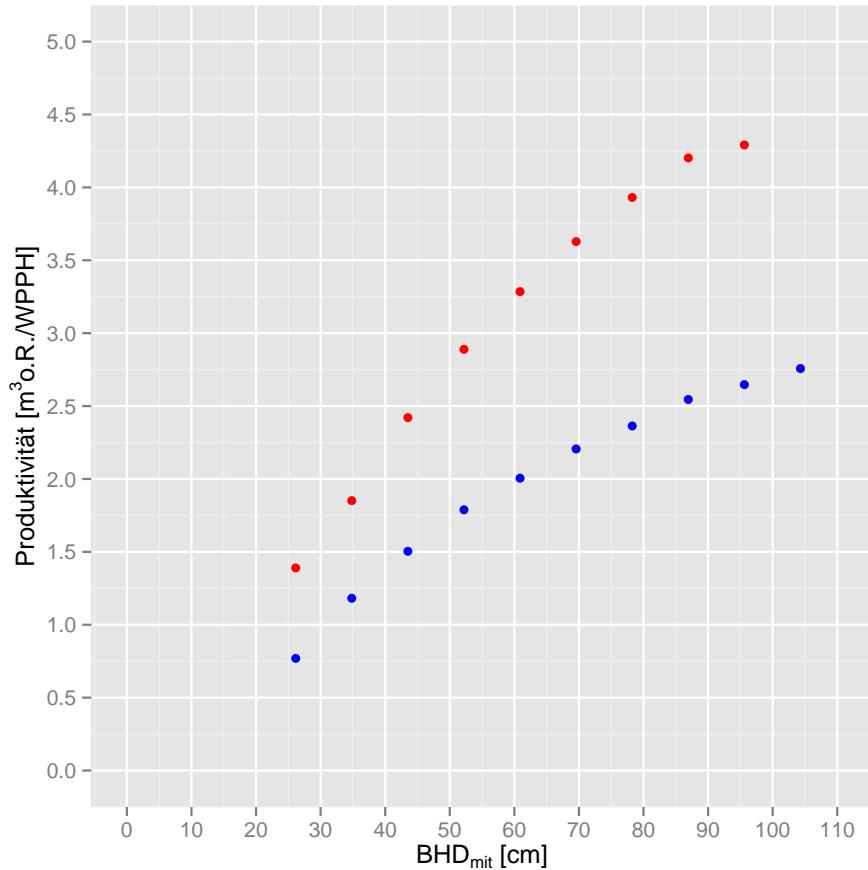


Abbildung 36: Vergleich Schätzungen des Modells mmNeu mit dem Modell mmAlt: **Hügelland**. Die roten Punkte markieren Schätzungen des Modells mm-Neu für Fichte im Hügelland (Anteil Laubholz = 0%, Anteil Kiefer = 0%), die blauen Punkte markieren Schätzungen des Modells mmAlt (Fichte, bis 30% Hangneigung, Hindernisse = gering, 100% Stammholz, 4- 6m Länge). Die Produktivitäten des mmAlt wurden auf WPSH umgerechnet.

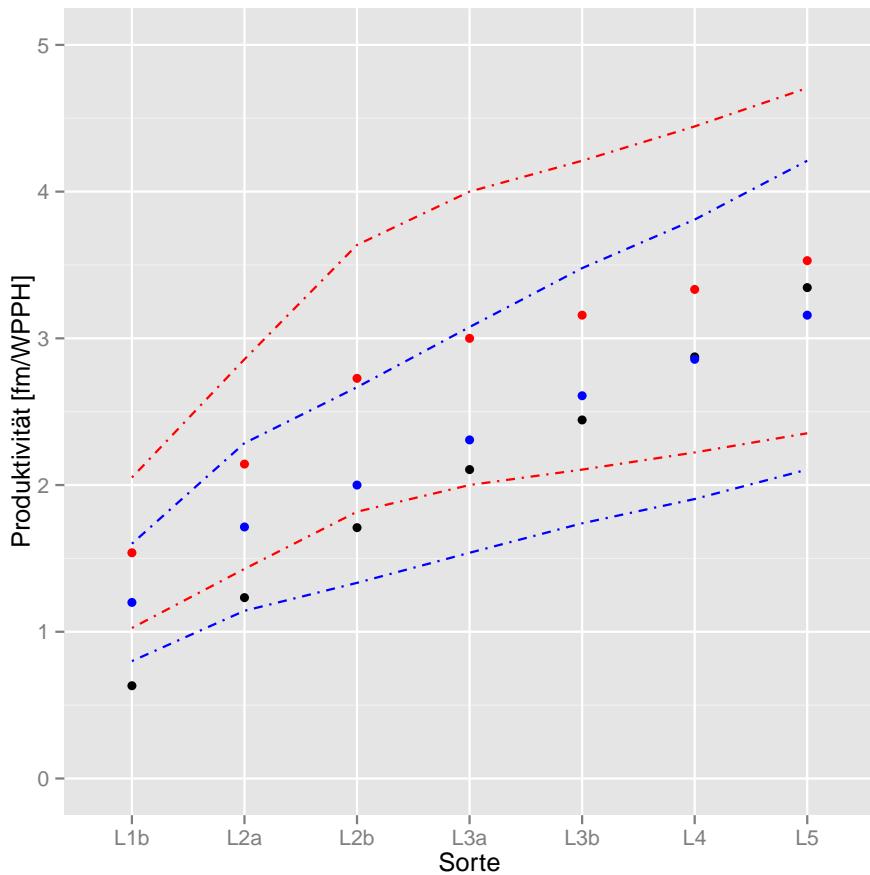


Abbildung 37: Vergleich des neuen Modells mit dem Hochgebirgstarif HGT. Modellwerte fürs Gebirge (schwarz) vs. HGT 5m (blau) und 7m (rot). Die gestrichelten Linien grenzen den Bereich +/- 20% ein.

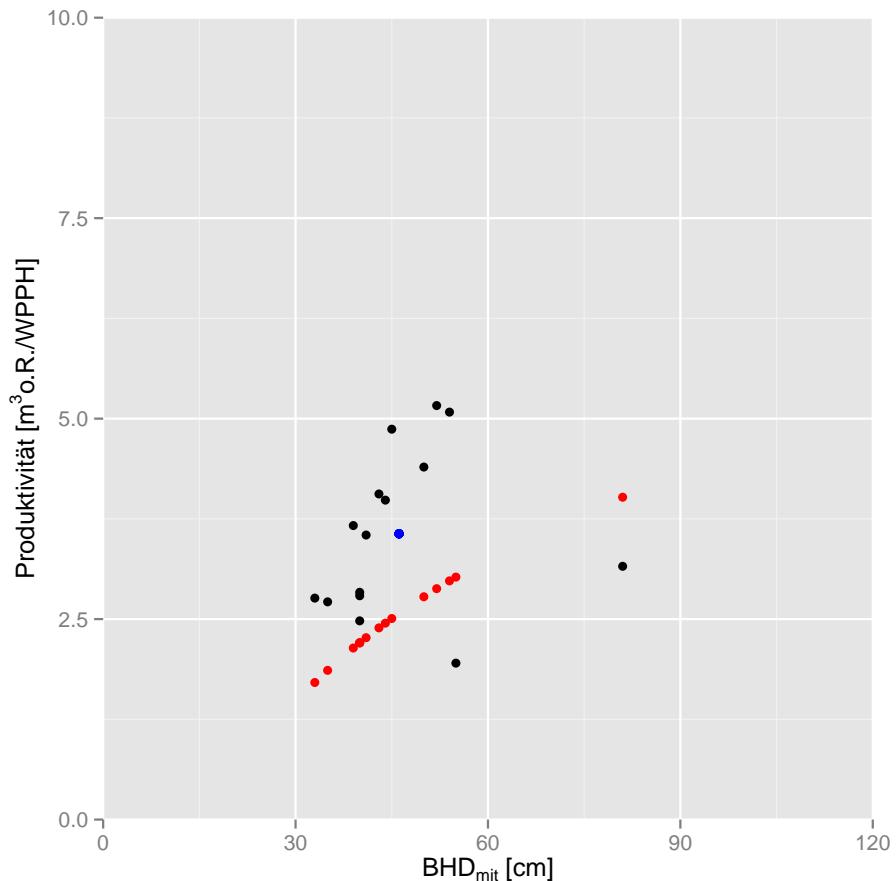


Abbildung 38: Vergleich der durch das neue Modell (nachfolgend Modell mm-Neu) berechneten Produktivitäten mit einigen Testdaten aus Schlägen im einfachen Gelände von Einzelbäumen aus dem Jahr 2013. Die schwarzen Punkte markieren die gemessenen Produktivitäten [m³ o.R./WPPH] (der blaue Punkt markiert den Durchschnitt der gemessenen Produktivitäten), die roten Punkte markieren die durch das Modell mmNeu berechneten Produktivitäten [m³ o.R./WPPH] (Region= Huegelland, Anteil Kiefer = 0, Anteil Laubholz = 0). In den Testdaten wurde nur der BHD [cm] erfasst. Die Mittellammvolumen wurden mittels Luzerner Tarif III (siehe Abbildung 29) berechnet. Die Zeiten der Testdaten wurden von PSH15 zu WPPH umgerechnet (eine PSH15 -Stunde entspricht 1.511 (WPPH) Stunden nach HeProMo).