

Produktivitätsmodell "Radharvester 2014"

Teil B: Analyse der Datensätze und Diskussion der Modellierung

Dario Pedolin, WSL Birmensdorf

10. Dezember 2014

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

Zusammenfassung

Anhand eines Datensatzes über die Leistungen von Holzschlägen mit Radharvester soll ein Modell für die Berechnung der Zeitaufwände für die Holzhauerei erstellt werden. Das Modell erklärt 62% der im Datensatz ($N = 928$ Holzschläge, total Holzmenge = 532'157 m³ o.R., 9 verschiedene Radharvester) gefundenen Varianz.

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	9
4	Modelle einsetzen	12
4.1	Formel, Koeffizienten und Systemgrenzen	12
4.2	Sensitivitätsanalysen	13
5	Anhang	15
5.1	Testen auf Verletzungen der Annahmen für multiple Regression .	15
5.2	Beschreibung der 'Ausreisser'	22
5.3	Kreuzvalidierung	23
5.4	Beschreibung der im Modell benutzten Messgrößen	26
5.5	Anhang zur Diskussion	35
5.6	Vergleich mit unabhängigen Testdaten	42
5.7	Erweiterung des Modells für kleine und grosse BHD_{mit}	50

1 Einleitung

Ein Datensatz bestehend aus 974 Holzschlägen aus den Jahren 2010, 2011 und 2012 zweier mitteleuropäischer Forstbetriebe wurde benutzt, um ein Modell für die Produktivität der Holzhauerei mit Radharvester 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. Das Modell wurde in der Residuenanalyse auf Verletzungen der Voraussetzungen für lineare Modellierung untersucht (Anhang Abbildung 5 bis 9). Abschliessend wurde eine Sensitivitätsanalyse für die Parameter (Abschnitt 'Diskussion'), eine Kreuzvalidierung (Anhang Abbildung 11), sowie ein Test des Modells mit einem unabhängigen Datensatz durchgeführt.

2 Analyse Produktivitäten

2.1 Beschreibung der Prozesse

Der Prozess 'Fällen mit Radharvester' umfasst alle Arbeitsschritte des Fahrens auf der Rückegasse, des Fällens der Bäume in Kranreichweite, des Aufarbeitens (Entasten, Vermessen, Einschneiden, Zopfen) und Ablegens. Der Output des Prozesses 'Fällen mit Radharvester' sind die Rundholzabschnitte auf Rohpolter entlang der Rückegasse (aus Lemm et al. (2003)).

2.2 Beschreibung der Einheiten und Masse

Alle Zeiten (Stunden [h]) sind PMH15 (Produktive Arbeitszeit mit Unterbrechungen bis 15 Minuten, vergleiche Anhang Abbildung 27). Die Mengenangaben [m^3 o.R.] sind Kubikmeter Holz ohne Rinde.

2.3 Erstellen und Testen des Modells

Die Messgrößen wurden mittels der R Methode 'lm' ('Fitting Linear Models') auf ihren Einfluss auf die abhängige Variable 'Kubikmeter o.R. pro Stunde MAS [Maschinenstunde = PMH15]' untersucht. Dabei wurden nur die 974 Fälle benutzt deren gemessene Produktivität mehr als 2 m^3 o.R./PMH15 und weniger als 50 m^3 o.R./PMH15 betrug¹ (siehe Anhang Tabelle 8 für Beschreibung der Variable 'Produktivität').

Nachfolgend werden die Ergebnisse dieser Analyse vorgestellt.

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

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-34.9401	3.9073	-8.94	0.0000
Maschinegross	-8.2555	5.9861	-1.38	0.1682
log(bhd)	14.9536	1.2068	12.39	0.0000
Maschinegross:log(bhd)	4.4151	1.8242	2.42	0.0157

Tabelle 1 zeigt die geschätzten Koeffizienten und deren p-Werte. Das korrigierte R^2 für diese erste Modellbildung beträgt 0.47.

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 die eingesetzten Maschinen (Farbe der Messpunkte) und Gesamtmenge in Kubikmeter o.R. (Grösse der Messpunkte) dargestellt. Fälle deren Residuen (Differenz gemessene Werte - geschätzte Werte)

¹Fälle mit Produktivitäten kleiner als 2 m^3 o.R./PMH15 und grösser als 50 m^3 o.R./PMH15 werden als Ausreisser betrachtet, da diese Produktivitäten im 'normalen' Betrieb nicht erreicht werden.

mehr als zwei Standardabweichungen (der Residuen) vom Mittelwert der Residuen abweichen, sind als Ausreisser (schwarz) markiert.

Die mittlere Abweichung der berechneten Werte von den gemessenen beträgt 3.8 Kubikmeter o.R./Stunde (Median: 3 Kubikmeter o.R./Stunde). 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 29% (Median: 20%) . Es gibt 46 Ausreisser² (siehe Anhang Abschnitt 5.2 für eine Beschreibung der Ausreisser). Diese Datensätze wurden von den nachfolgenden Auswertungen ausgeschlossen.

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

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-34.4624	3.0544	-11.28	0.0000
Maschinegross	-24.9268	4.8241	-5.17	0.0000
log(bhd)	14.7029	0.9446	15.57	0.0000
Maschinegross:log(bhd)	9.5333	1.4717	6.48	0.0000

Tabelle 2 zeigt, dass vor allem die Variablen 'Maschine' und mittlerer 'BHD' (nachfolgend BHD_{mit}) für die Produktivität des Fällens mit Radharvester massgebend sind. Das korrigierte R^2 für die Modellbildung beträgt 0.62. Zusätzlich wurden die in Anhang Abbildung 5 bis 9 gezeigten Tests durchgeführt, um auf Verletzungen der Voraussetzungen der Annahmen für multiple Regression zu testen. Weiters wurde für das Model eine Kreuzvalidierung (siehe Anhang Abschnitt 5.3) sowie ein Test mit unabhängigen Daten (siehe Anhang Abschnitt 5.6) durchgeführt .

Die Abbildung 2 zeigt den Vergleich der berechneten mit den gemessenen Werten für den Datensatz nach Ausschluss der 46 Ausreisser. 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 vom Modell erklärte Varianz (R^2) beträgt 62%. Zusätzlich sind die eingesetzten Maschinen (Farbe/ Form der Messpunkte) und der mittlere BHD (Grösse der Messpunkte) eingezeichnet.

²Die 46 Ausreisser machen 5% der Fälle aus

Die mittlere Abweichung³ der berechneten Werte von den gemessenen beträgt 3.2 Kubikmeter o.R./Stunde (Median: 2.6 Kubikmeter o.R./Stunde). 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 von den Modelldaten beträgt 23% (Median: 18%) .

³Die Abweichungen/ Schätzfehler werden aufgrund der zum Modellieren benutzten Daten berechnet und nicht mittels unabhängigen Testdaten. Eine robustere Schätzung ergibt sich aus der Kreuzvalidierung (Anhang Tabelle 12 und dem Vergleich mit unabhängigen Testdaten)

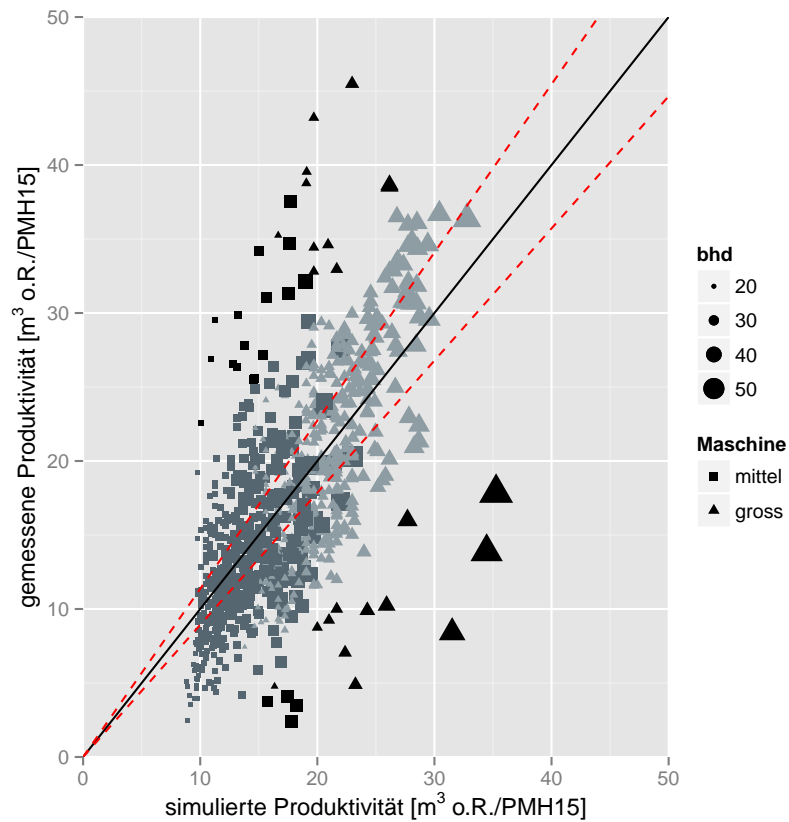


Abbildung 1: Vergleich der mittels des Modells berechneten Werte für 'Kubikmeter o.R. pro Maschinenstunde' mit den gemessenen Werten. Die roten gestrichelten Linien grenzen den Bereich von 12% Abweichung ein. Datensätze mit Residuen, die mehr als 2 Standardabweichungen über dem Mittelwert der Residuen liegen, sind als schwarze Punkte eingezeichnet.

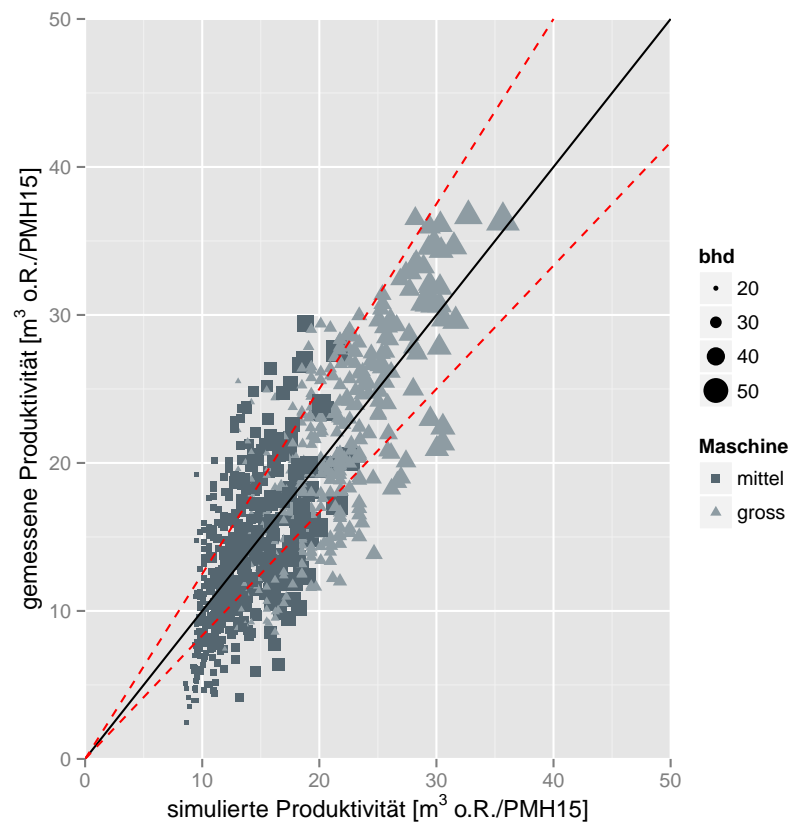


Abbildung 2: Vergleich der mittels des Modells berechneten Werte für 'Kubikmeter o.R. pro Maschinenstunde' mit den gemessenen Werten für die Daten ohne Ausreisser. Die roten gestrichelten Linien grenzen den Bereich von 20% Abweichung ein.

3 Diskussion

3.1 Diskussion der Datengrundlage

Das Datenset besteht aus Datensätzen von zwei verschiedenen Betrieben (vergleiche auch Abbildung 3. Aus einem der beiden Betriebe sind nur die Mittleren Volumen der Bäume (statt der mittleren BHDs) der bearbeiteten Stämme bekannt. Damit für das Modell der mittlere BHD benutzt werden kann wurde für diese Fälle der BHD_{mit} aus dem Massenmittelstamm berechnet. Dazu wurde der Luzerner Tarif III verwendet (siehe auch Anhang Abbildung 28).

Die eingesetzten Maschinen Typen sind in Tabelle 12 aufgeführt. Diese Maschinen wurden gemäss der neuen Klasseneinteilung der Radharvester für die FPA Prüfung des KWF in zwei Klassen eingeteilt ⁴. Die gleiche Gruppierung wurde auch mittels einer Clusteranalyse der Maschinendaten nach Leistung [kW] und maximaler Fälldurchmesser [mm] erreicht (vergleiche Abbildung 17). Allerdings wird die Maschine "Ponsse Ergo" mittels dieser Methode nicht korrekt eingeteilt - nach KWF wäre der Ponsse der Kategorie 'gross' zuzuordnen, in Wahrheit hat die Maschine aber alle Merkmale einer Maschine der Kategorie "mittel". Grundsätzlich ist eine Maschine 'gross' wenn a) alle Merkmale (Leistung, maximaler Fälldurchmesser, Kran Länge, Hydraulik auf einander abgestimmt und für grosse Bäume ausgelegt sind und b) die Maschine auch für grosse Bäume eingesetzt wird.

Das in der Analyse benutzte kombinierte Datenset umfasst 1040 Holzschläge mit dem Radharvester. Nach Entfernung der Fälle mit offensichtlich zu kleinen ($< 2 \text{ m}^3 \text{ o.R./PMH15}$) bzw. zu grossen ($> 50 \text{ m}^3 \text{ o.R./PMH15}$) Produktivitäten bleiben noch 974 Fälle für die erste Analyse. Dabei wurden 46 Ausreisser (d.h. übermässig grosse Abweichung der gemessenen von der durch das erste Modell geschätzten Produktivitäten) identifiziert. Die restlichen 928 Fälle wurden für die eigentliche Schätzung der Modellkoeffizienten benutzt.

Die exakte Anzahl der an den Arbeiten beteiligten Personen ist nicht bekannt. Es wird von einem Normalfall mit einem Maschinenführer und einer Maschine ausgegangen.

3.2 Diskussion des Modells

Das in Tabelle 2 gezeigte Modell für die Produktivität des Fällens und Aufarbeitens mit Radharvester beinhaltet neben den Grössen für den eingesetzten Maschinentyp den mittleren BHD der bearbeiteten Bäume. Die Arbeitsproduktivität der Holzerei mit Radharvester hängt in erster Linie von der Grösse (bzw. dem BHD_{mit}) der Bäume ab. Die Anteile Nadelholz bzw. Laubholz sowie die Hangneigung fliessen nicht in das Modell ein. Wie im Anhang Abbildung 19 bis Tabelle 20 gezeigt, gibt es keine grossen Unterschiede der Produktivitäten für Fälle mit fast nur Nadel- bzw. Laubholz hat die Hangneigung keinen grossen

⁴Die drei Klassen nach KWF richten sich nach der Leistung der Maschine: Klasse 'klein' bis 120 kW, Klasse 'mittel' bis 180 kW und Klasse 'gross' ab 180 kW

Einfluss auf die Produktivität (der Einfluss Hangneigung ist mehr als eine Größenordnung kleiner als der Einfluss des BHD_{mit} . Da die Hangneigung zudem nur schwierig im Voraus zu erfassen ist wurde in diesem Modell auf diese Einflussgrösse verzichtet). Die vom Modell berechneten Produktivitäten beziehen sich auf Kubikmeter o.R. pro Maschinenstunde.

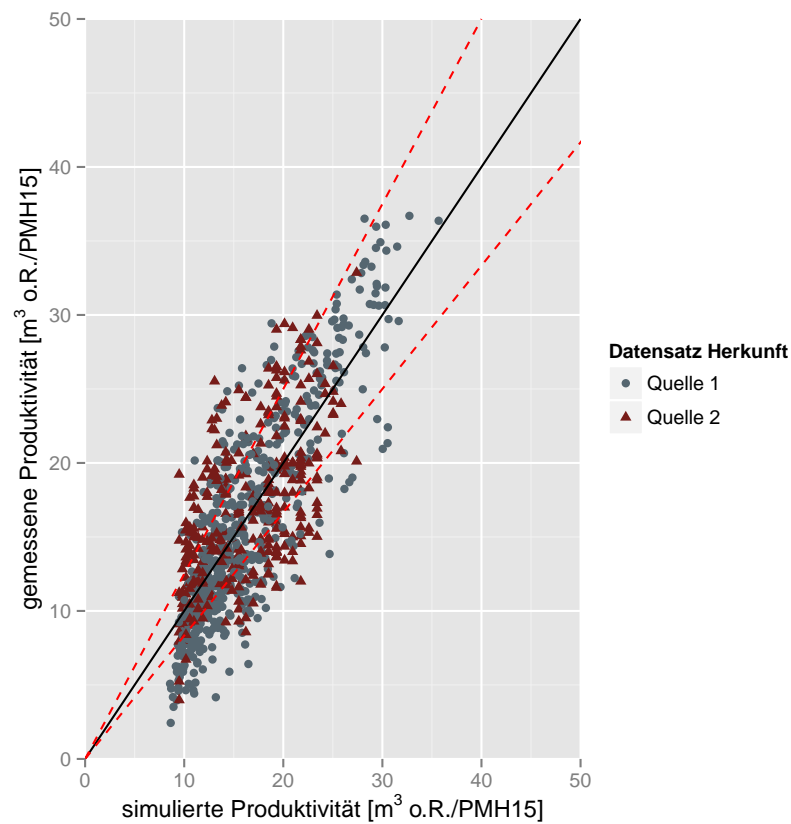


Abbildung 3: Vergleich der mittels des Modells berechneten Werte für 'Kubikmeter o.R. pro Maschinenstunde' mit den gemessenen Werten für die Daten ohne Ausreisser. Für jeden Datensatz ist dessen Herkunft angegeben. Die roten gestrichelten Linien grenzen den Bereich von 20% Abweichung ein.

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 geschehen. 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 'Fällen mit Radharvester' lautet:

$$\begin{aligned}
 [m^3_{o.R.}/PMH15] = & a \\
 & + b_{\text{Maschine}} \\
 & + c * \ln(\text{BHD}_{\text{mit}} [cm]) \\
 & + d_{\text{Maschine}} * \ln(\text{BHD}_{\text{mit}} [cm])
 \end{aligned} \tag{1}$$

Formel 1 zeigt das Modell aus Tabelle 2 in mathematischer Form . Das Modell umfasst alle Prozesse des Fällens und Aufarbeitens mit Radharvester. Die Koeffizienten dazu sind:

Tabelle 3: Koeffizienten für Modell Produktivität 'Fällen mit Radharvester'. Die Min. und Max. Werte geben die Grenzen, für die das System einsetzbar (bzw. durch die Datengrundlage definiert) ist, an.

Koeff.	Kategorie	Wert	Einheiten	Min.	Max.
a	Konstante	-34.462	$[\frac{m^3_{o.R.}}{h}]$		
b_{Maschine}	falls Maschine mittel	0	$[\frac{m^3_{o.R.}}{h}]$		
b_{Maschine}	falls Maschine gross	-24.927	$[\frac{m^3_{o.R.}}{h}]$		
c	BHD_{mit}	14.7	$[\ln(\frac{1}{cm}) * \frac{m^3_{o.R.}}{h}]$	19	51
d_{Maschine}	falls Maschine mittel: BHD_{mit}	0	$[\ln(\frac{1}{cm}) * \frac{m^3_{o.R.}}{h}]$	19	48
d_{Maschine}	falls Maschine gross: BHD_{mit}	9.5333	$[\ln(\frac{1}{cm}) * \frac{m^3_{o.R.}}{h}]$	19	51

4.2 Sensitivitätsanalysen

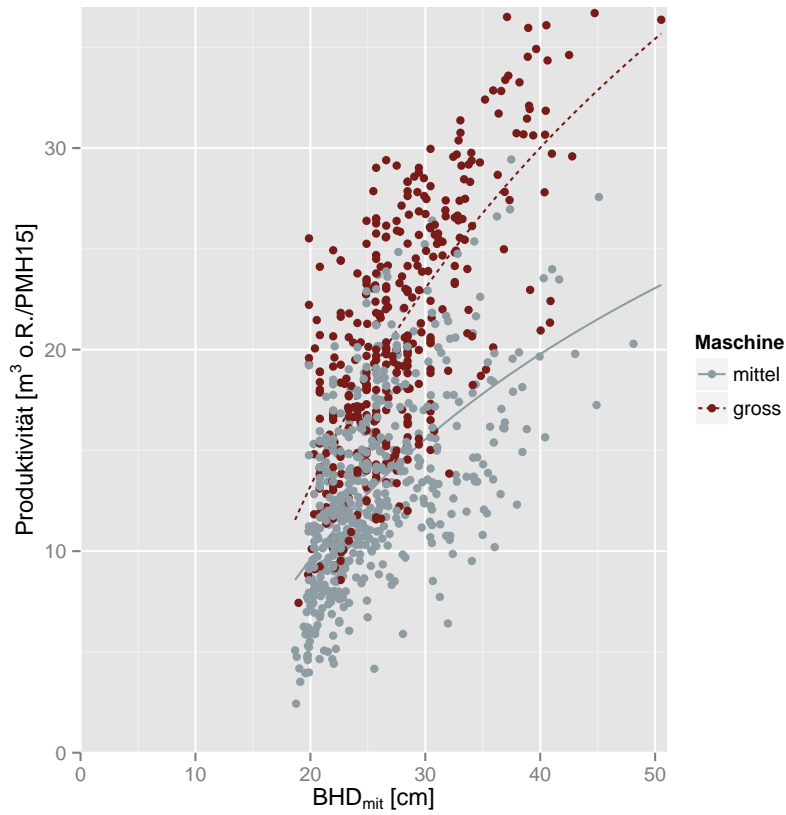


Abbildung 4: Sensitivitätsanalyse für Variable 'BHD_{mit} [cm]'. Der lineare Zusammenhang gilt für BHD_{mit} kleiner als der maximale Fälldurchmesser des Harvesterkopfes. Für den Zusammenhang von BHD_{mit} und 'm³ o.R. pro Baum' siehe Anhang Abbildung 28. Die Punkte markieren die gemessenen Produktivitäten.

Abbildung 4 zeigt die Abhängigkeit der durch das Modell berechneten Produktivitäten vom mittleren BHD für die verschiedenen Maschinen. Diesem Zusammenhang werden durch die Kapazitäten des Harvester-Kopfes (maximale Baumdurchmesser, die vom Harvester noch bearbeitet werden können) Grenzen gesetzt.

Literatur

- 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 zukunftsfähige 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 der 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.
4. Modellannahme 4: Die Residuen sind normalverteilt um den Mittelwert 0.

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

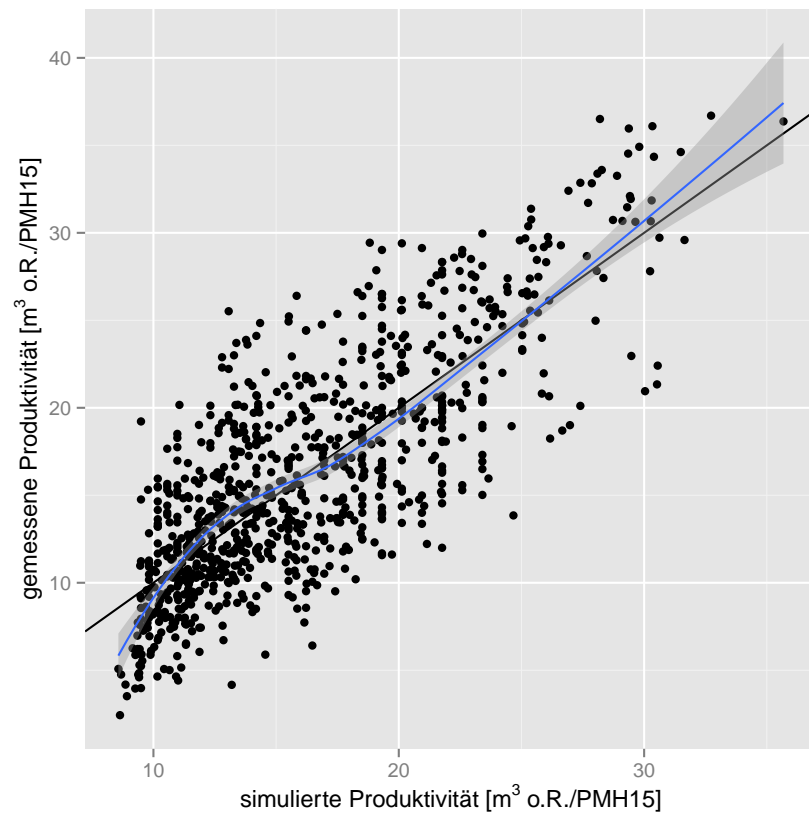


Abbildung 5: Test Modellannahmen 1: Linearität. Die Punkte sind annähernd ideal um die Diagonale verteilt.

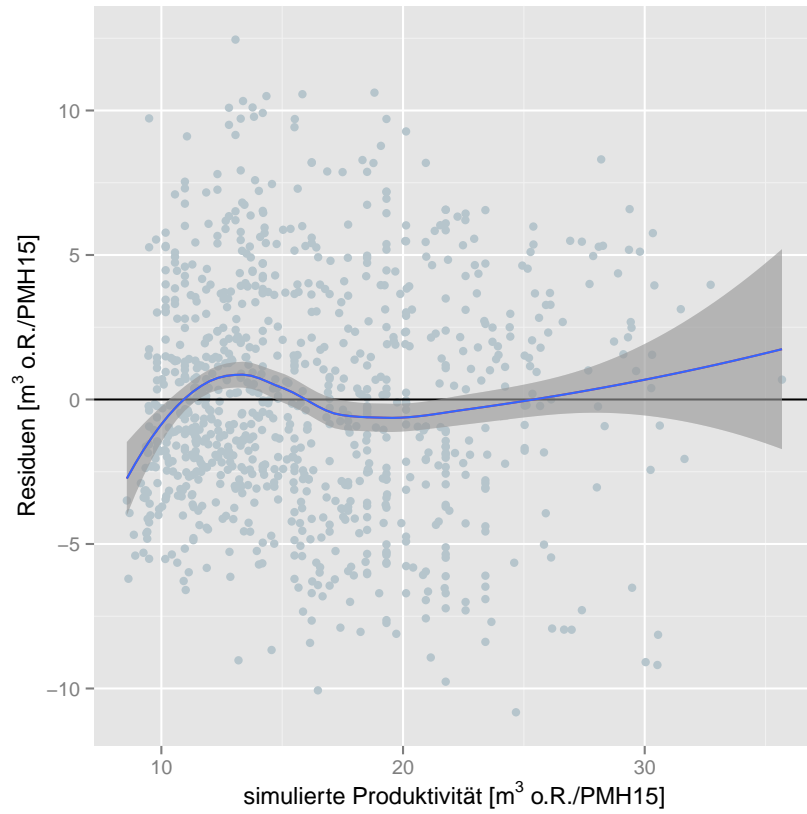


Abbildung 6: Modellannahme 1: Normalverteilung der Residuen. Tukey-Anscombe Plot für das Modell aus Tabelle 2. Die blaue Linie markiert das gleitende Mittel der Messpunkte. Der p-Wert der ANOVA der Mittelwertunterschiede der Residuen über die vier Quartile beträgt 0.752.

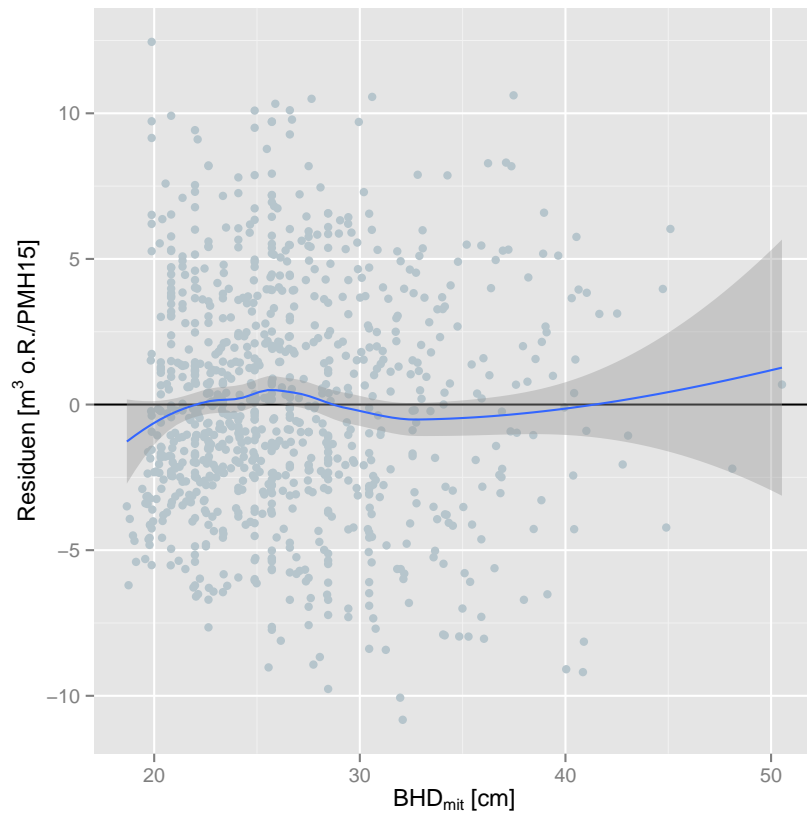


Abbildung 7: Test Modellannahmen 2: Korrelation der Residuen mit unabhängiger Variable ' BHD_{mit} '. Die blaue Linie steht für das gleitende Mittel. Der p-Wert der ANOVA der Mittelwertunterschiede der Residuen über die vier Quartile beträgt 0.946.

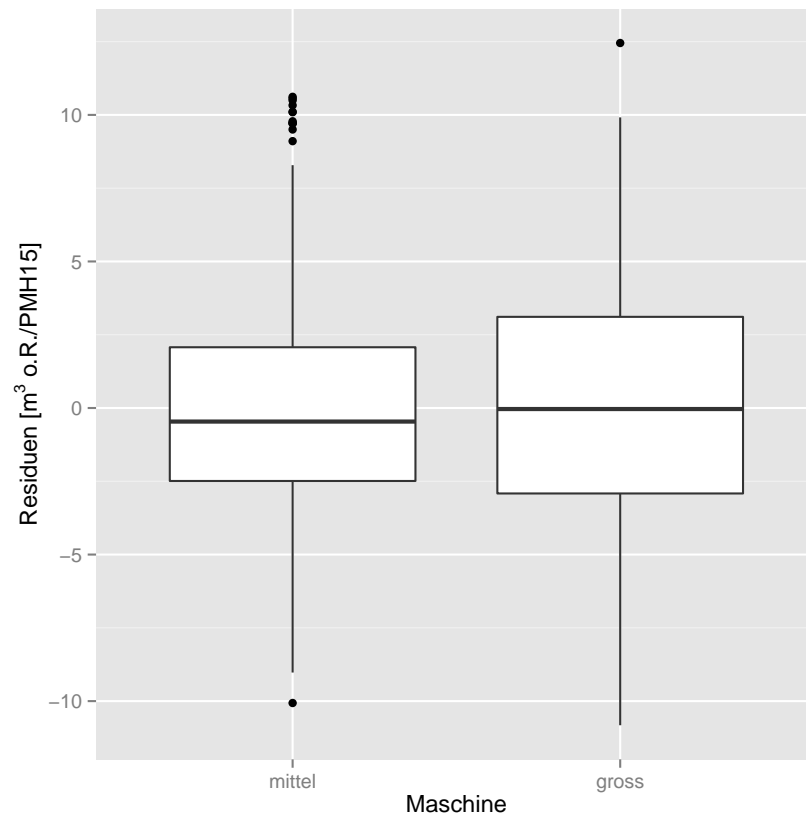


Abbildung 8: Test Modellannahmen 2: Korrelation der Residuen mit unabhängiger Variable 'Maschine'. Der Median der Residuen ist bei ähnlicher Verteilung für beide Maschinenklassen nahe bei 0.

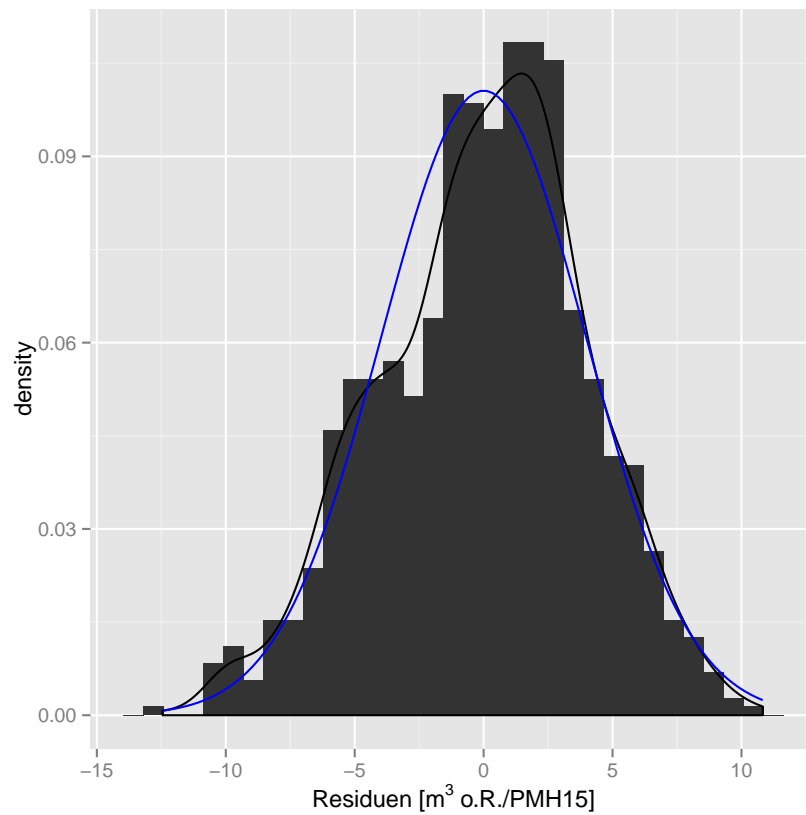


Abbildung 9: Test Modellannahmen 4: Normalverteilung der Residuen: Dichtefunktion der Residuen. Als Referenz ist die Dichtefunktion einer normalverteilten Variable blau eingezeichnet. Die Residuen sind annähernd normalverteilt.

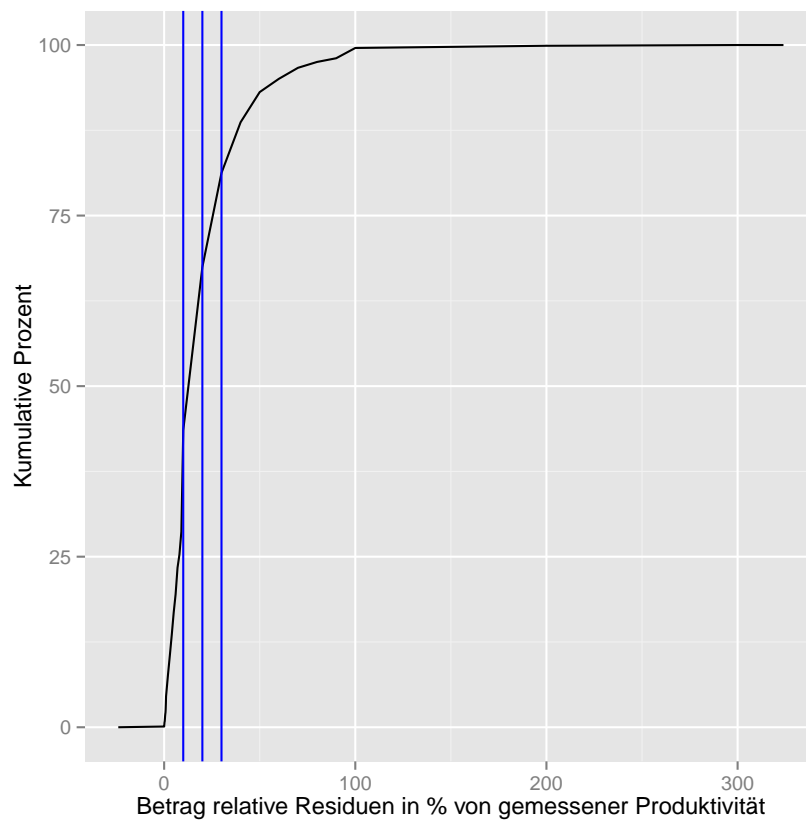


Abbildung 10: Kumulative Verteilung der absoluten relativen Residuen. Die vertikalen blauen Linien grenzen den Bereich von bis 10% und 20% bzw. 30% Abweichung ein.

5.2 Beschreibung der 'Ausreisser'

Tabelle 4: Häufigkeiten der Maschinen bei den Ausreissern

Maschine	N Ausreisser	N Total	% Ausreisser	% total Ausr.
Tiger.	12	138	9	1.2
Rott.	3	43	7	0.3
Erg.	11	185	6	1.1
JD	10	181	6	1.0
Timb.	3	49	6	0.3
Val.	6	161	4	0.6
HSM	1	47	2	0.1
Cob.	0	141	0	0.0
Val.C	0	29	0	0.0

Tabelle 5: Häufigkeiten der konstruierten Maschinen bei den Ausreissern

konstr. Maschine	N Ausreisser	N Total	% Ausreisser
mittel	20	565	4
gross	26	409	6

5.3 Kreuzvalidierung

Tabelle 6: Resultat der Analyse des linearen Zusammenhangs und Schätzung der Modellkoeffizienten

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-36.3312	3.4588	-10.50	0.0000
Maschinegross	-22.5106	5.6674	-3.97	0.0001
log(bhd)	15.2700	1.0682	14.29	0.0000
Maschinegross:log(bhd)	8.8191	1.7284	5.10	0.0000

Tabelle 6 gibt die Koeffizienten aus der Modellbildung mit dem Trainingsset wieder. Das korrigierte R^2 für die Modellbildung beträgt 0.62.

Tabelle 7: Resultat der Analyse des linearen Zusammenhangs und Schätzung der Modellkoeffizienten für den Trainingsdatensatz - Unterschiede der Koeffizientenschätzungen

	p-Wert	alle Daten	Trainingsset	% Diff
(Intercept)	0.00	-34.46	-36.33	-5.00
Maschinegross	0.00	-24.93	-22.51	10.00
log(bhd)	0.00	14.70	15.27	-4.00
Maschinegross:log(bhd)	0.00	9.53	8.82	7.00

Tabelle 7 zeigt die Unterschiede für die Koeffizientenschätzung aus allen Daten (N= 928) vs. Koeffizientenschätzung aus Trainingsdaten (N= 663). Der p-Wert bezieht sich auf die Schätzung aus allen Daten.

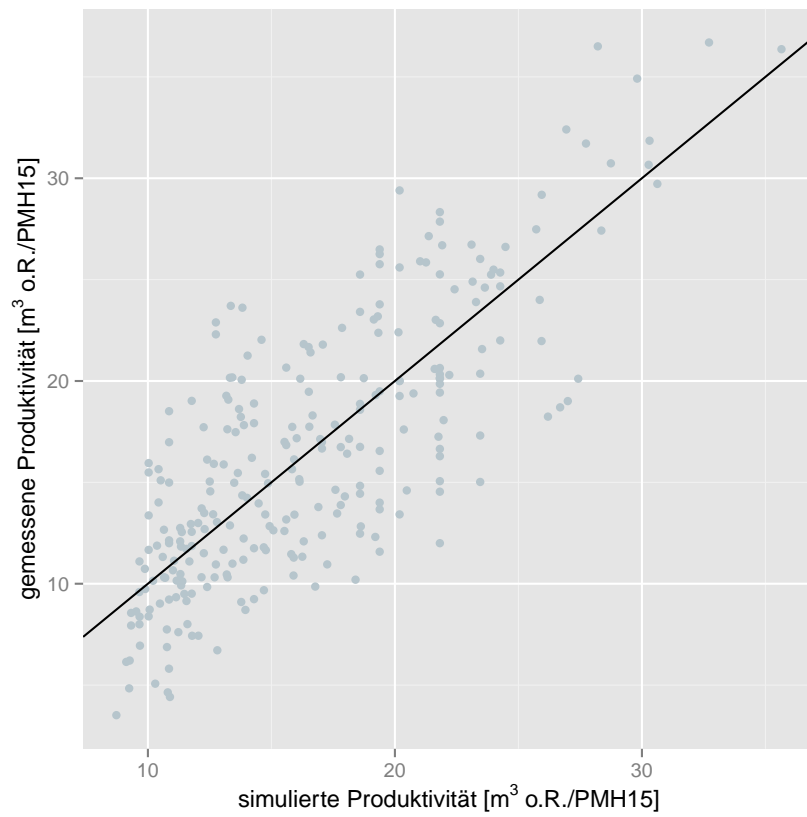


Abbildung 11: Die Korrelation zwischen simulierter und gemessener Produktivität für das Validierungsset beträgt 0.79.

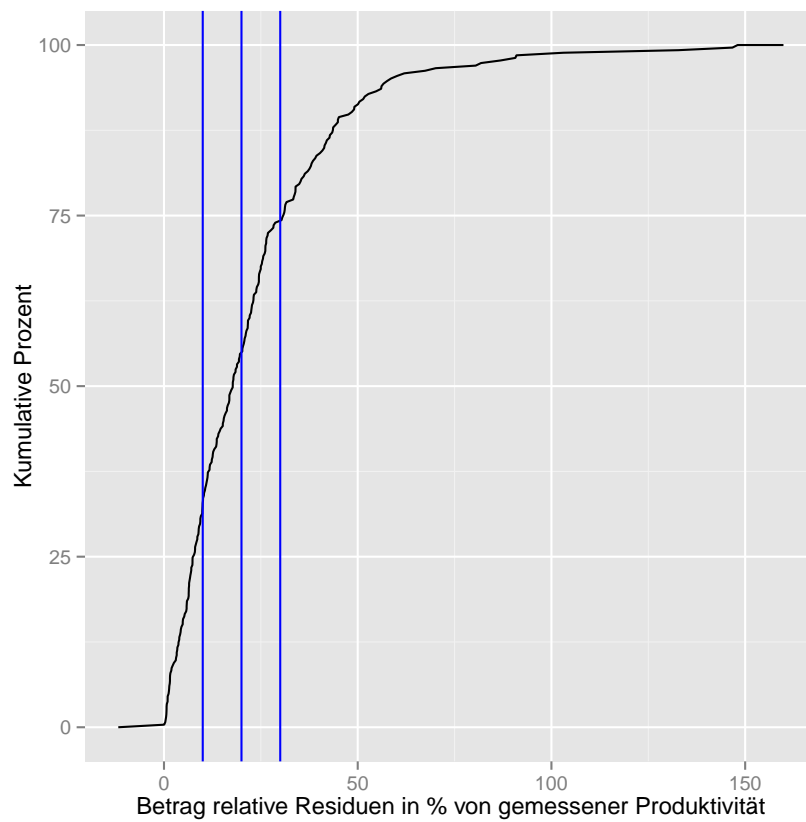


Abbildung 12: Kumulative Verteilung der absoluten relativen Residuen der Kreuzvalidierung. Die vertikalen blauen Linien grenzen den Bereich von bis 10% und 20% bzw. 30% Abweichung ein.

5.4 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 8: Beschreibung der Produktivitäten [m3/PMH15] der Trainings- und Validierungsdaten

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	928	16.18	6.48	15.00	2.43	36.70	0.64	0.21
Trainingsdaten	663	16.05	6.46	14.86	2.43	36.09	0.65	0.25
Validierungsdaten	265	16.49	6.54	15.49	3.52	36.70	0.63	0.40

Tabelle 9: Beschreibung der Variable 'Totale m3 inkl. x' [m3] der Trainings- und Validierungsdaten

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	928	573.45	606.96	354.07	9.68	4170.84	1.97	19.92
Trainingsdaten	663	567.00	614.12	339.31	10.31	4170.84	2.07	23.85
Validierungsdaten	265	589.56	589.48	398.82	9.68	3230.38	1.71	36.21

Tabelle 10: Beschreibung der Variable 'Durchschnittliche m3 pro Baum' [m3] der Trainings- und Validierungsdaten

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	928	0.48	0.29	0.43	0.08	2.03	1.34	0.01
Trainingsdaten	663	0.48	0.29	0.42	0.08	1.84	1.23	0.01
Validierungsdaten	265	0.48	0.29	0.43	0.09	2.03	1.60	0.02

Die Maschinentypen werden durch folgende Kennzahlen charakterisiert (Tabelle 12):

Wie in Tabelle 12 aufgezeigt, unterscheiden sich die im Datensatz erhobenen Maschinen nicht grundlegend. Die bei der Modellbildung gefundenen unterschiedlichen Produktivitäten der Maschinen könnten auch durch den selektiven Einsatz bestimmter Maschinen für Starkholz, Schwachholz erklärt werden.

Tabelle 11: Beschreibung der Variable 'mittlerer BHD' [cm] der Trainings- und Validierungsdaten

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	928	26.50	5.19	25.67	18.67	50.53	1.09	0.17
Trainingsdaten	663	26.50	5.23	25.56	18.67	48.12	1.03	0.20
Validierungsdaten	265	26.51	5.09	25.72	18.97	50.53	1.26	0.31

Tabelle 12: Beschreibung Maschinen

Bezeichnung	Abk.	Grösse	[t]	[kW]	Räder	Kran [m]	Fäll.Durch. [mm]	Herk.	minBHD	maxBHD	meanBHD	N
John Deere 1470D	JD	gross	20	180	6	7.9	710	RIP	18.97	50.53	30.05	171
Ponsse Ergo H7	Erg.	mittel	19	205	8	10	750	RIP	18.81	48.12	26.96	174
Valmet 911.3	Val.	mittel	17	170	6	11	650	RIP	19.12	41.04	26.11	155
Ponsse Cobra H60	Cob.	mittel	17	157	8	8	640	RIP	18.67	37.98	25.31	141
Rottne H20	Rott.	gross	21	187	6	10	750	BW	20.82	33.66	27.04	40
Tigercat H16 HP	Tiger.	gross	20	181	6	10	710	BW	19.88	35.91	25.08	126
Valmet 921C	Val.C	mittel	19	155	6	9	650	BW	20.82	30.45	24.37	29
Timberjack 1270 A	Timb.	mittel	16-17	152	6	10	660	BW	19.88	25.72	21.32	46
HSM 405H3 8 WD	HSM	gross	23	260	8	11	710	BW	20.82	30.45	26.50	46

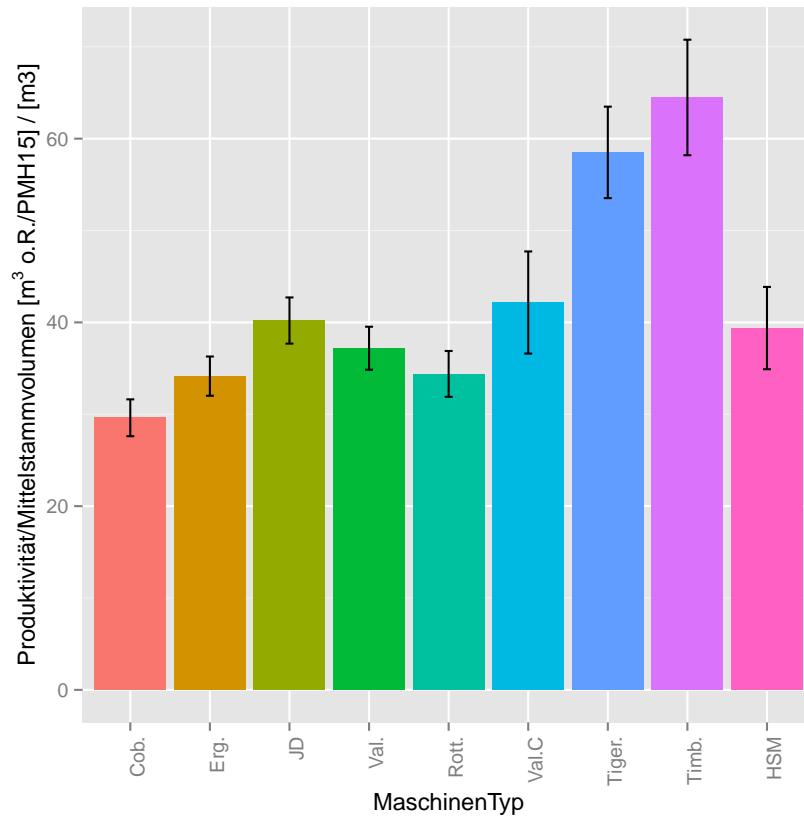


Abbildung 13: Vergleich der durchschnittlichen, vom Mittelstammvolumen unabhängigen Produktivitäten (= Produktivität/ Mittelstammvolumen) für die erfassten Maschinentypen. Gezeigt sind die Mittelwerte und die 95% Konfidenzintervalle der Mittelwerte. Der p-Wert der ANOVA beträgt: < 0.000.

Abbildung 13 zeigt, dass es Unterschiede bei den gemessenen, für das Mittelstammvolumen bereinigten Produktivitäten für die erfassten Maschinen gibt. Um die Ursache dieser Unterschiede ab zu schätzen wurden die durchschnittlichen Kubikmeter o.R. pro Baum für die Maschinen untersucht (Anhang Abbildung 14.)

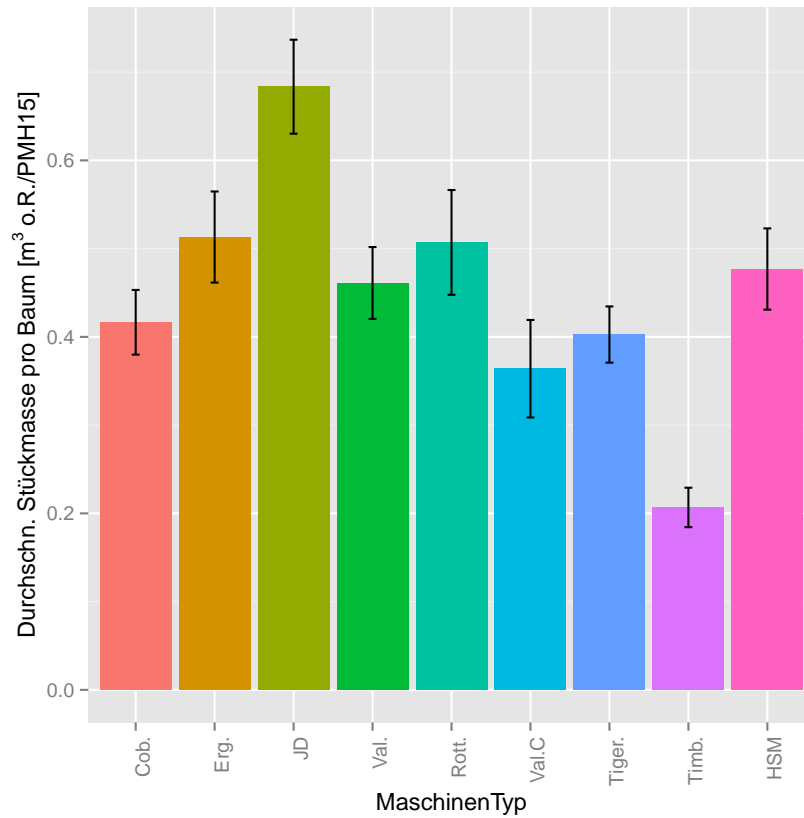


Abbildung 14: Vergleich der durchschnittlichen Stückmassen pro Baum für die erfassten Maschinentypen. Gezeigt sind die Mittelwerte und die 95% Konfidenzintervalle. Der p-Wert der ANOVA beträgt: < 0.000.

Die erfassten Maschinen wurden für unterschiedliche Stückvolumen eingesetzt (vergl. Anhang Abbildung 14). Insbesondere die Maschine 'JD' wurde bei größeren Stückvolumen benutzt, während die Maschine 'Cob.' bei den kleinsten Stückvolumen eingesetzt worden ist. Der Zusammenhang zwischen Produktivität und Stückvolumen ist in Anhang Abbildung 15 dargestellt. Für die Auswertung wurden die Maschinen in 'gross' (JD) und 'mittel' (Val., Erg. & Cob.) eingeteilt (siehe auch Anhang Tabelle 12).

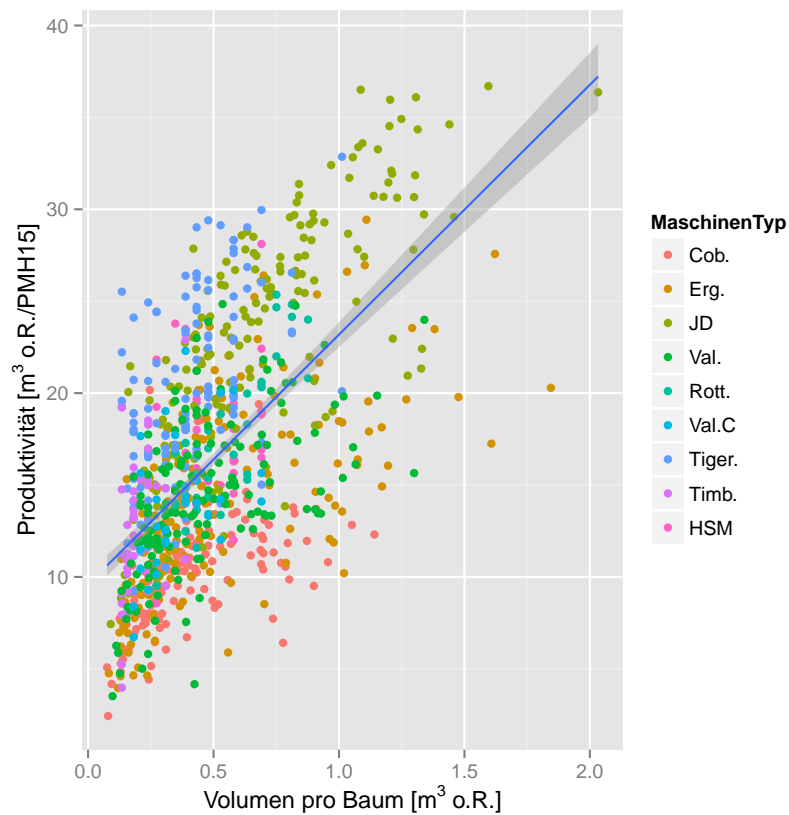


Abbildung 15: Lineare Regression der Produktivität eines Holzschlages nach dem durchschnittlichen Volumen pro Baum. Die Abbildung zeigt, dass höhere durchschnittliche Stückvolumen (Kubikmeter o.R. pro Baum) mit höheren Produktivitäten korrelieren. Der Regressionskoeffizient R^2 beträgt 37%.

Tabelle 13: Beschreibung (Häufigkeiten) der konstruierten Variable 'Maschinen Art'

	mittel	gross
Alle Daten	545	383
Trainingsdaten	400	263
Validierungsdaten	145	120

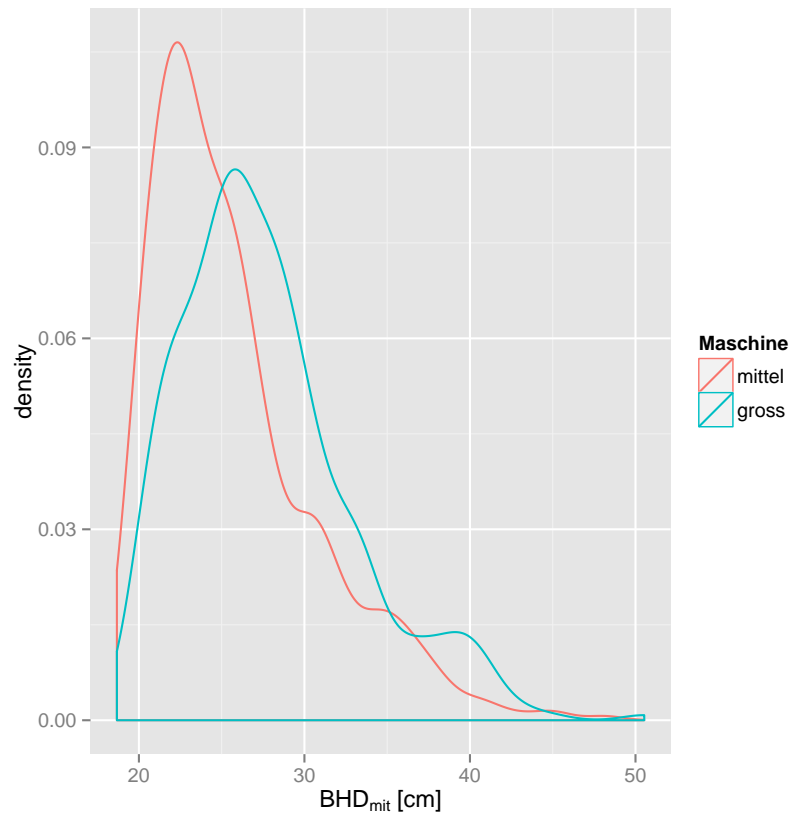


Abbildung 16: Verteilung der durchschnittlichen BHD_{mit} für die konstruierten Maschinen Arten.

Tabelle 14: Beschreibung der Variable 'Kubikmeter pro Stück'

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	606	0.13	0.10	0.10	0.02	0.76	2.10	0.00
Trainingsdaten	436	0.13	0.10	0.10	0.02	0.59	1.80	0.00
Validierungsdaten	170	0.13	0.11	0.10	0.03	0.76	2.77	0.01

Tabelle 15: Beschreibung der Variable 'Total Bäume'

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	606	866.75	888.54	585.00	17.00	5621.00	2.03	36.09
Trainingsdaten	436	839.47	869.47	565.00	17.00	5621.00	2.18	41.64
Validierungsdaten	170	936.71	934.69	646.00	17.00	5171.00	1.69	71.69

Tabelle 16: Beschreibung der Variable 'Nadelholzanteil der liegenden Stämme'

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	821	0.90	0.20	0.99	0.00	1.00	-2.70	0.01
Trainingsdaten	589	0.90	0.19	0.99	0.00	1.00	-2.81	0.01
Validierungsdaten	232	0.89	0.21	0.99	0.01	1.00	-2.45	0.01

5.5 Anhang zur Diskussion

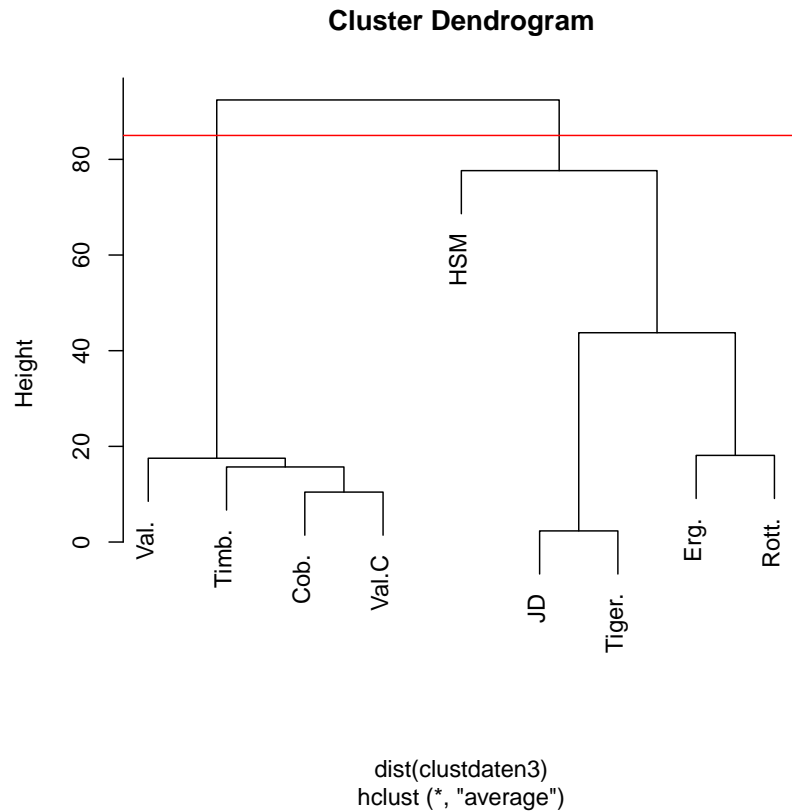


Abbildung 17: Baumdiagramm der Clusterung der Maschinen. Die hier gezeigte Clusterung erfolgte über die Variablen 'Leistung [kW]' und 'maximaler Fälldurchmesser [mm]'. Die rote Linie zeigt die gewählte Gruppierungshöhe. Die beiden so gefundenen Clustergruppen entsprechen (bis auf die Maschine Erg. (Ponsse Ergo)) den KWF Gruppen 'mittel' und 'gross'.

Tabelle 17: Mittelwerte der MaschinenKlassen			
	Leistung [kW]	Kran [m]	Fälldurchmesser [mm]
Klasse mittel	168	9.60	670
Klasse gross	202	9.72	720

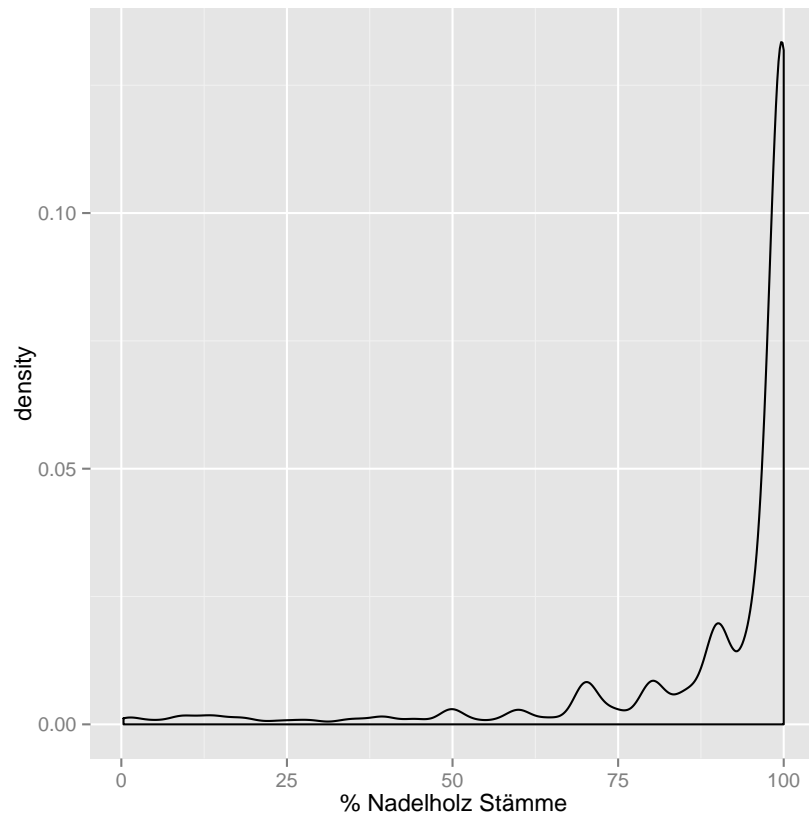


Abbildung 18: Verteilung der Variable 'Prozent Nadelholz Stämme liegend'.

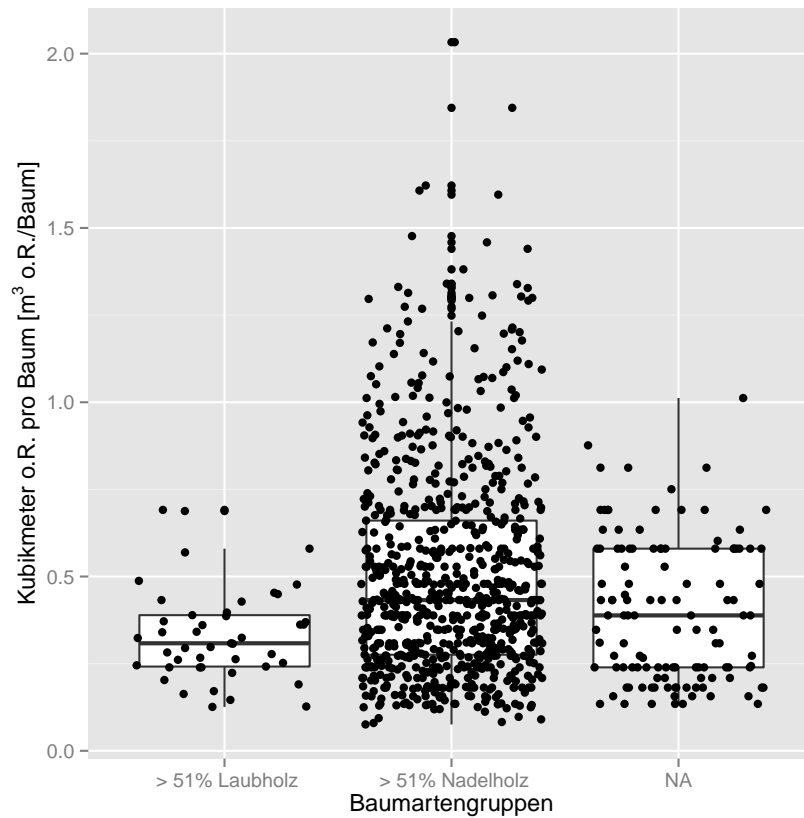


Abbildung 19: Laubholz vs. Kubikmeter o.R. pro Baum. Der p-Wert der ANOVA beträgt 0.00014, der p Wert des t-Test für den Mittelwert der Produktivität für die Gruppen '> 51% Laubholz' vs '> 51% Nadelholz' beträgt < 0.000 , $Mw > 51\% \text{ LbH} = 0.33 \text{ [fm/Baum]}$, $Mw > 51\% \text{ NdH} = 0.51 \text{ [fm/Baum]}$. Um die Lesbarkeit zu erhöhen sind die eingezeichneten Messwerte (Punkte) auf der x-Achse versetzt gezeichnet.

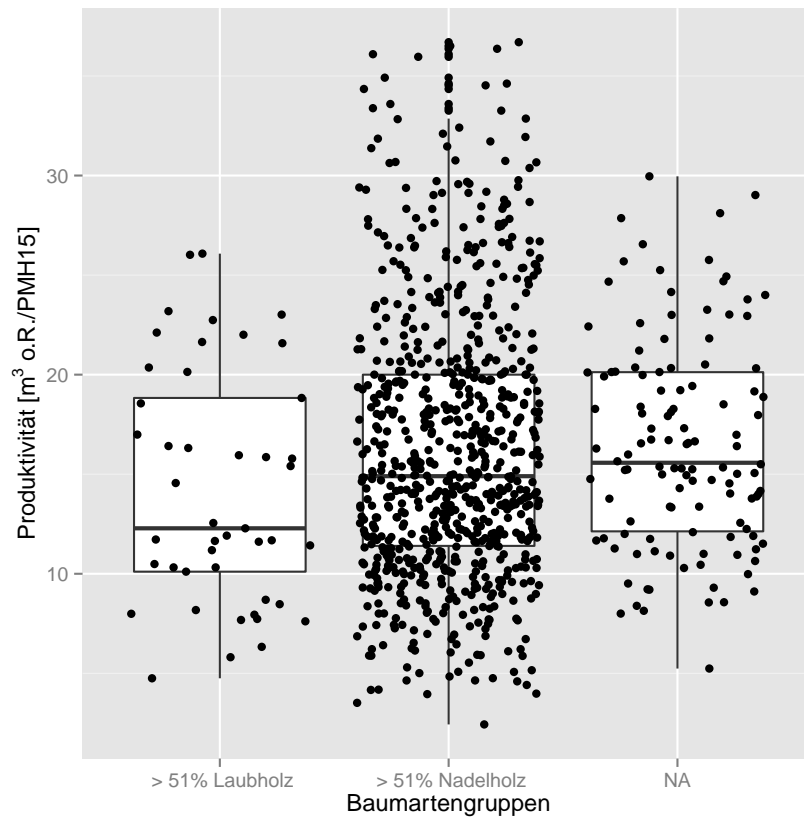


Abbildung 20: Laubholz vs. Produktivität. Der p-Wert der ANOVA beträgt 0.0532, der p-Wert des t-Test für den Mittelwertsunterschied der Produktivität für die Gruppen '> 51% Laubholz' vs '> 51% Nadelholz' beträgt 0.0354, $Mw > 51\% \text{ LbH} = 14 \text{ [m}^3 \text{ o.R./PMH15]}$, $Mw > 51\% \text{ NdH} = 16 \text{ [m}^3 \text{ o.R./PMH15]}$. Um die Lesbarkeit zu erhöhen sind die eingezeichneten Messwerte (Punkte) auf der x-Achse versetzt gezeichnet.

Tabelle 18: Beschreibung der Variable 'Mittlere Hangneigung [Grad]'

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	257	15.10	7.25	15.00	2.00	35.00	0.43	0.45
Trainingsdaten	183	14.50	7.05	10.00	5.00	30.00	0.60	0.52
Validierungsdaten	74	16.58	7.57	20.00	2.00	35.00	0.01	0.88

Tabelle 19: Resultat der Analyse des linearen Zusammenhangs und Schätzung der Modellkoeffizienten für Modell mit Hangneigung (lineare Abhängigkeit vom mittleren BHD!)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.0060	2.4597	1.22	0.2227
Maschinemittel	2.5818	3.3491	0.77	0.4414
Mittl.BHD	0.8623	0.0990	8.71	0.0000
Mittl.Hangneigung	-0.1409	0.0359	-3.92	0.0001
Maschinemittel:Mittl.BHD	-0.3426	0.1399	-2.45	0.0149

Tabelle 20: Resultat der Analyse des linearen Zusammenhangs und Schätzung der Modellkoeffizienten für Modell wie Original aber mit Hangneigung

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-37.9792	7.5489	-5.03	0.0000
Maschinemittel	18.2346	10.4050	1.75	0.0807
log(Mittl.BHD)	19.5208	2.3854	8.18	0.0000
Mittl.Hangneigung	-0.1431	0.0364	-3.93	0.0001
Maschinemittel:log(Mittl.BHD)	-7.5481	3.3058	-2.28	0.0231

Tabelle 19 und 20 zeigt die Schätzungen der Koeffizienten und p-Werte für das Modell mit Abhängigkeit von der Hangneigung. Damit der Einfluss der Hangneigung besser abgeschätzt werden kann, benutzt das Modell für Tabelle 19 eine lineare Abhängigkeit vom BHD_{mit} (statt Abhängigkeit vom $\ln(BHD_{mit})$). Der mittlere Beitrag der Hangneigung berechnet sich aus dem Median der Hangneigung (15) und deren Koeffizient (-0.14) ($= -2.1 \text{ m}^3 \text{ o.R./PMH15}$, versus Einfluss des BHD_{mit} $62 \text{ m}^3 \text{ o.R./PMH15}$).

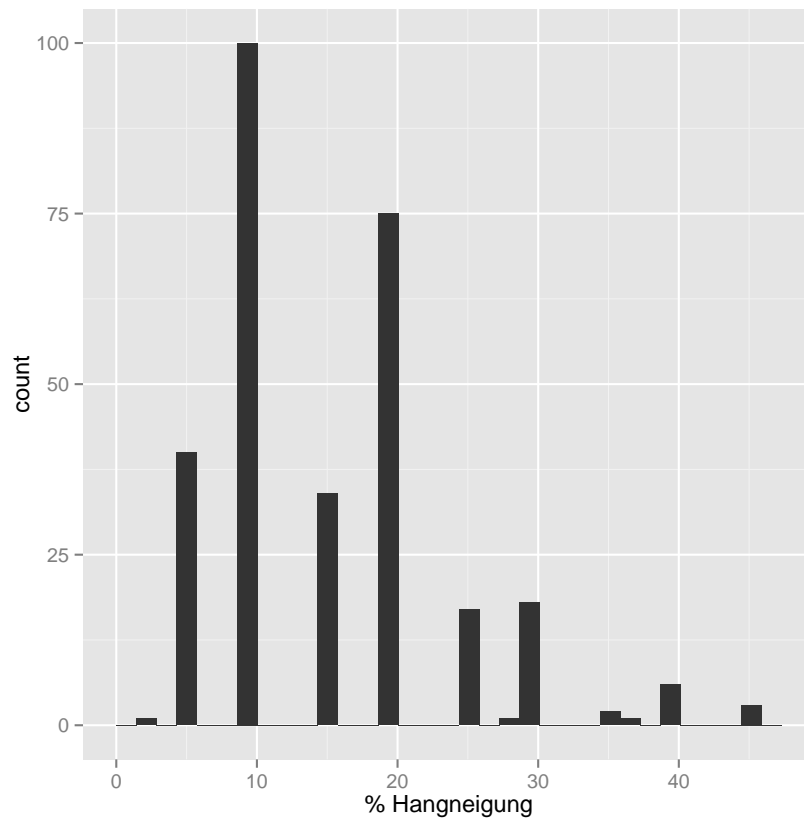


Abbildung 21: Verteilung der Variable 'mittlere Hangneigung [%]'. Der Einfluss der Hangneigung kann nur bis Hangneigung 30% geschätzt werden. Für grössere Hangneigungen hat es zu wenig Fälle im Datenset.

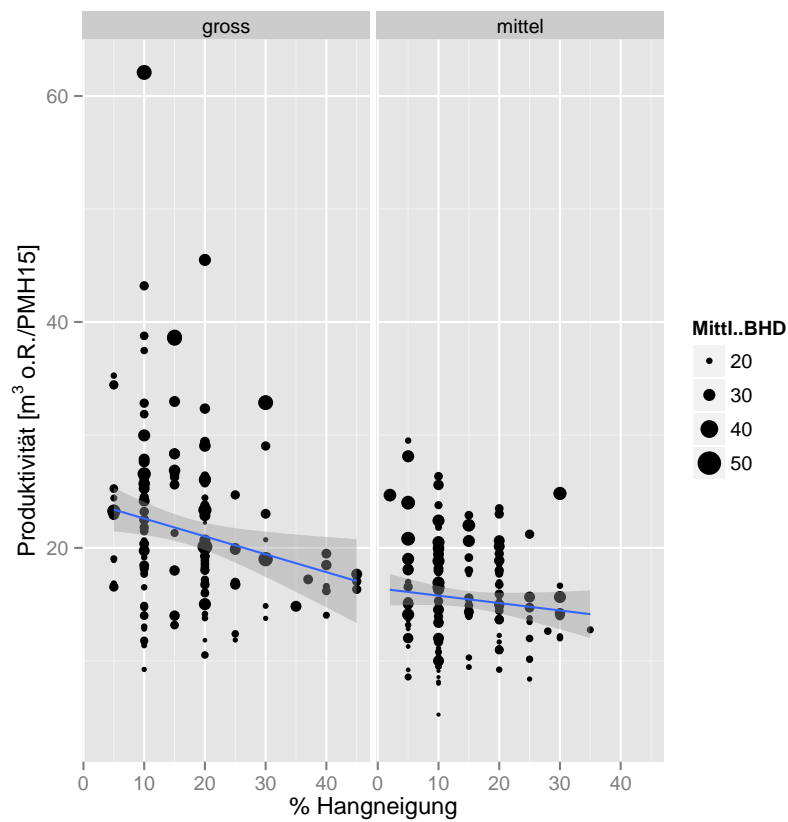


Abbildung 22: Einfluss der Variable 'mittlere Hangneigung [%]' auf die gemessene Produktivität [m³ o.R./PMH15] für die eingesetzten Maschinen und $BHD_{mit.}$. Der p-Wert des Koeffizienten ($-0.14 \text{ m}^3 \text{ o.R./PMH15/Grad}$) für die Hangneigung im Modell beträgt < 0.001 (siehe Tabelle 19).

5.6 Vergleich mit unabhängigen Testdaten

Anhand eines Datensets mit 134 Aufträgen und 4 verschiedenen Maschinen (vergl. Tabelle 21) konnte das neue Modell getestet werden. Die Verteilung der absoluten Schätzfehler ist in Abbildung 26 dargestellt.

Tabelle 21: Beschreibung der Vergleichsdaten

MaschinenOriginal	N	MW BHD	Total m3	MW m3/PMH15	Einteilung
John Deere 1270 E	33	24.00	18574.67	13.08	gross
Ponsse Ergo	37	25.56	50863.87	13.03	mittel
timberjack 1270 D	16	22.62	5607.10	10.28	mittel
valmet 911.3	48	21.87	29908.50	11.11	mittel

Die Verteilung der Schätzfehler für die Testdaten ist ähnlich wie für die Modellierdaten. Der Median der absoluten Schätzfehler beträgt 19% (Mittelwert: 31%).

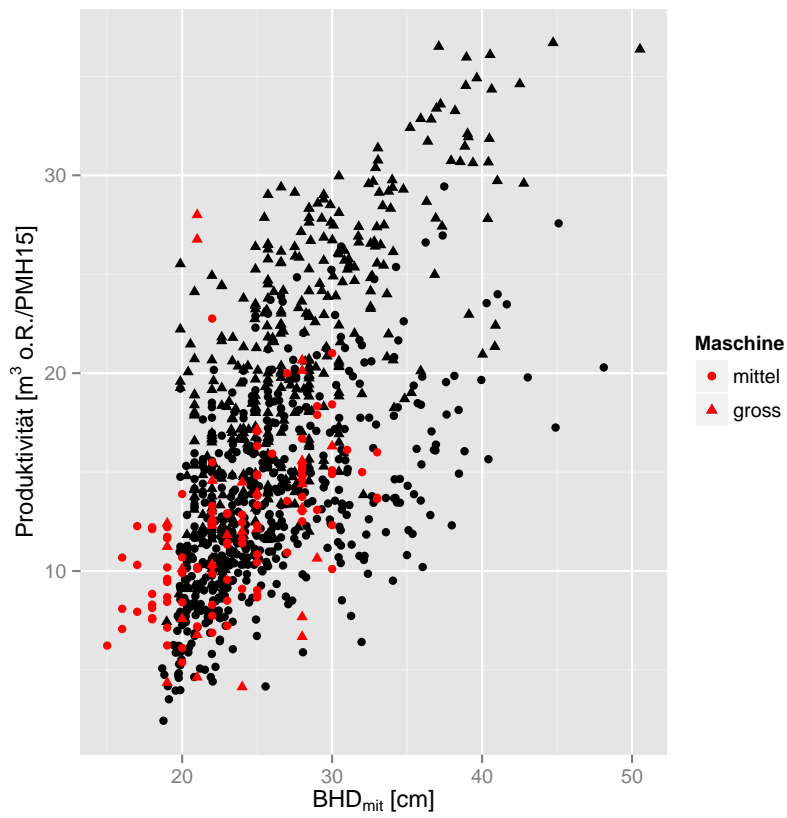


Abbildung 23: Lage der Vergleichs/ Testdaten (rot) im Vergleich zu den Modellierdaten (schwarz). Testdaten deren BHD_{mit} kleiner als 18cm ist wurden aus den nachfolgenden Tests ausgeschlossen.

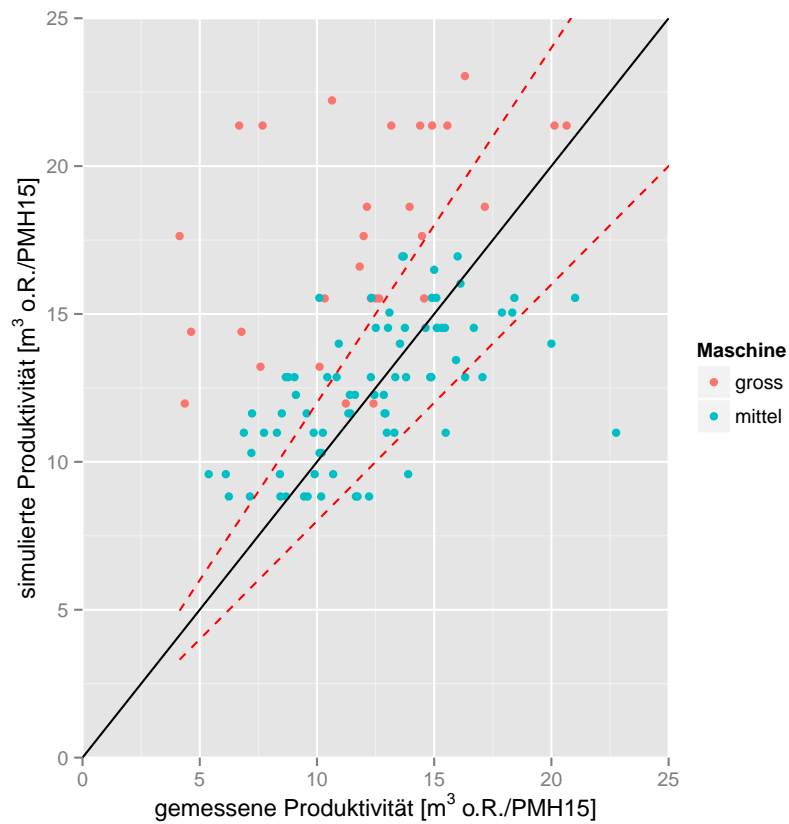


Abbildung 24: Vergleich gemessene Produktivität vs. berechnete Produktivität für die Vergleichs/ Testdaten. Die gestrichelten roten Linien grenzen den Bereich von 20% Abweichung ein. Die Schätzungen für die grossen Maschinen weisen die grössten Schätzfehler auf. Der mittlere Schätzfehler beträgt 33% (Median: 21%).

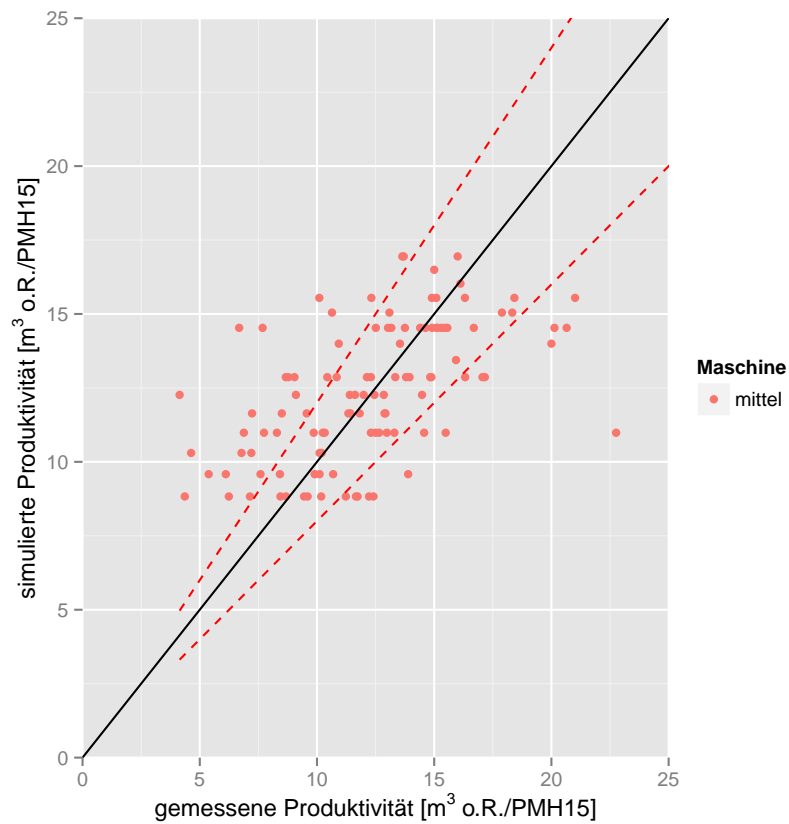


Abbildung 25: Vergleich gemessene Produktivität vs. berechnete Produktivität für die Vergleichs/ Testdaten, alle Maschinen auf 'mittel' gesetzt. Die gestrichelten roten Linien grenzen den Bereich von 20% Abweichung ein. Der mittlere Schätzfehler beträgt 24% (Median: 15%).

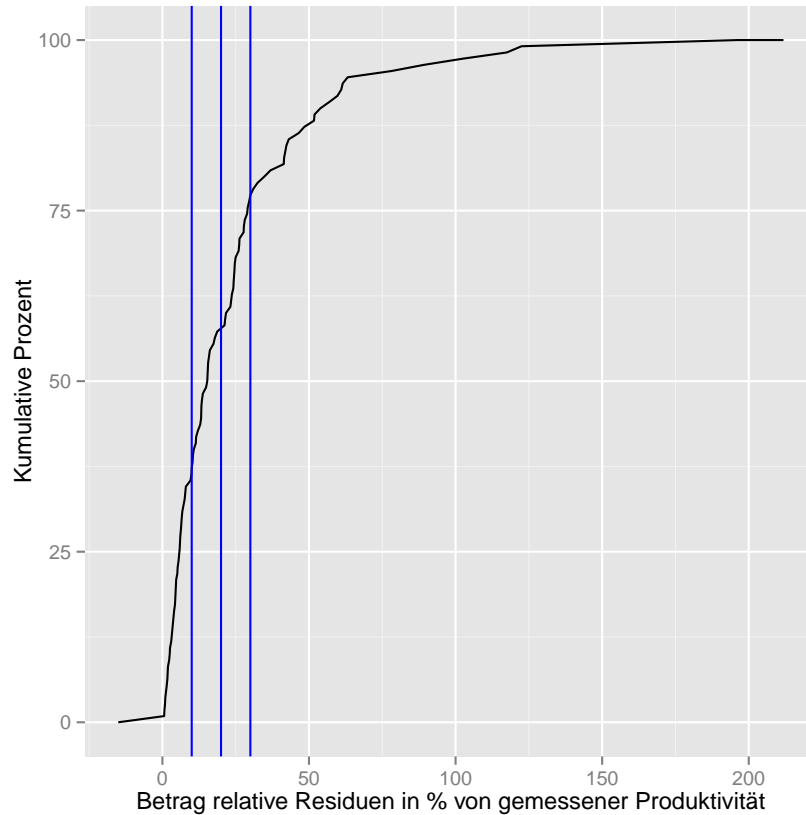


Abbildung 26: Kumulative Verteilung der absoluten relativen Schätzfehler für die Vergleichsdaten (alle Maschinen = 'mittel', wie in Abbildung 25 dargestellt). Die vertikalen blauen Linien grenzen den Bereich von bis 10% und 20% bzw. 30% Abweichung ein.

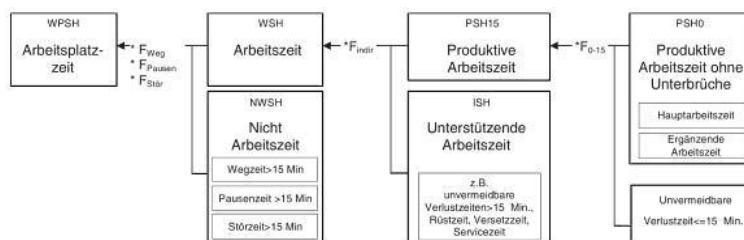


Abbildung 27: Zeitsystem (übernommen aus Thees and Lemm (2009))

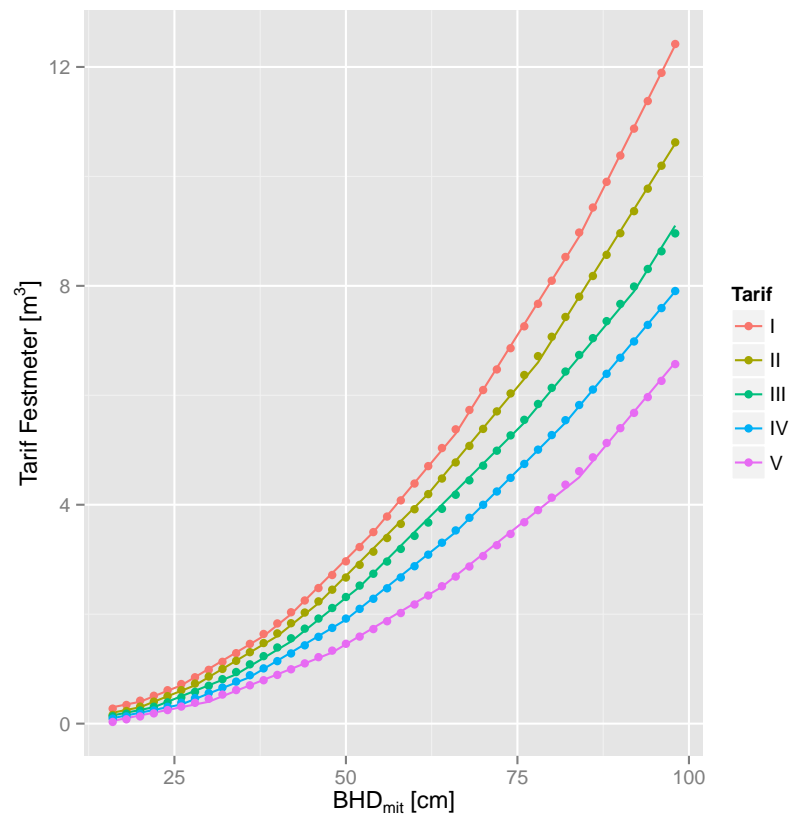


Abbildung 28: Verwendeter Tarif (Luzerner Tarif I bis V) zur Berechnung des Volumens [m³] pro Baum aus BHD_{mit} [cm] (Schweizer (2011)).

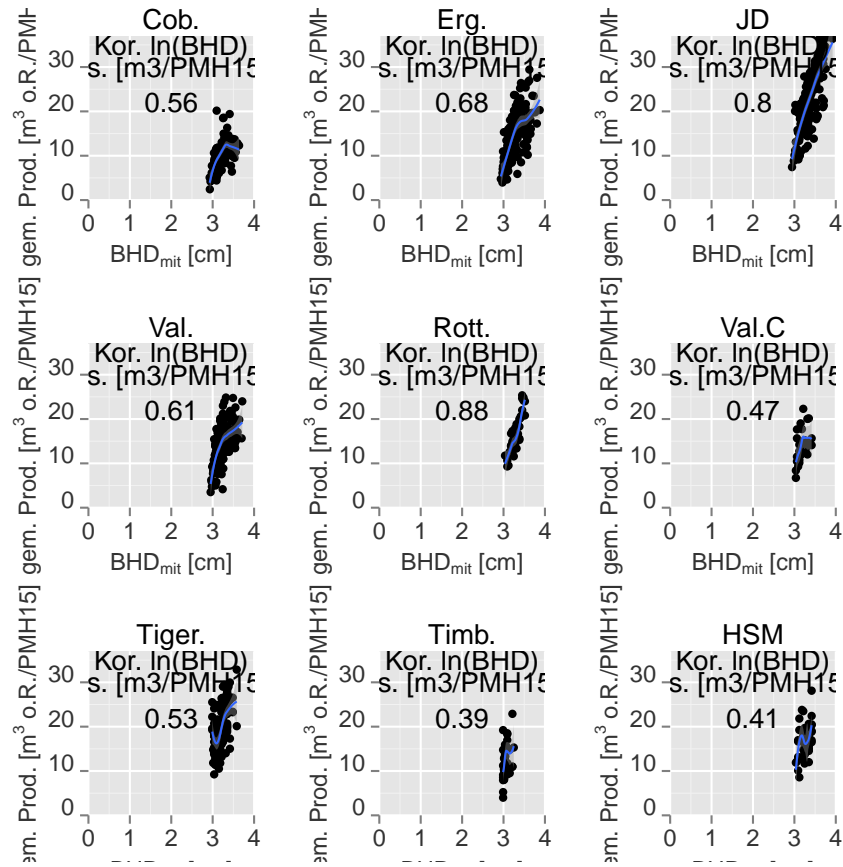


Abbildung 29: Abhängigkeiten der gemessenen Produktivitäten $[m^3 \text{ o.R./PMH15}]$ vom $\ln(\text{BHD}_{\text{mit}})$ für die erfassten Maschinentypen.

Tabelle 22: Berechnungsbeispiele

Maschine	Total m3 [m3]	BHD [cm]	m3 pro Baum	[m3]	Hangneigung	berechnet m3/PMH15	gemessen m3/PMH15	% Differenz
gross	915	26	0.43	25	19.31	16.95	14	
gross	469	30	0.64		22.73	21.31	7	
mittel	496	26	0.42		13.17	13.41	2	
mittel	702	22	0.26		11.28	15.95	29	
mittel	137	20	0.13	10	9.49	8.56	11	
gross	1677	25	0.39		18.51	12.47	48	
mittel	2409	35	0.94		17.72	22.62	22	
mittel	111	21	0.18		10.16	4.64	119	
mittel	964	32	0.77		16.41	21.67	24	
gross	411	29	0.63	10	22.59	25.69	12	
gross	2806	28	0.58	10	21.76	14.54	50	

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

Da das Modell im Bereich $BHD_{mit} < 20\text{cm}$ und $BHD_{mit} > 50\text{cm}$ nicht durch gemessene Werte gestützt werden kann (vergleiche Verteilung der BHDs in den Daten, Abbildung 30), wurden für diese Bereiche Sonderlösungen implementiert. Für Schläge mit $BHD_{mit} > 50\text{cm}$ (bei Maschine 'gross' ab $BHD_{mit} > 60\text{cm}$) und $BHD_{mit} < 75\text{cm}$ (grösster Durchmesser den der Prozessorkopf bearbeiten kann = 75cm) ist die Produktivität konstant. Für Schläge mit $BHD_{mit} < 12\text{cm}$ (bei Maschine 'gross' ab $BHD_{mit} < 15\text{cm}$) wird keine Produktivität berechnet (Einsatz des Radharvesters nicht sinnvoll, bzw. unkontrollierbar teuer), bzw. wird für $BHD_{mit} > 75\text{cm}$ keine Produktivität berechnet. Dieses Erweiterte Modell ist für BHD_{mit} 12cm bis 75cm definiert (vergleiche Abbildung 31).

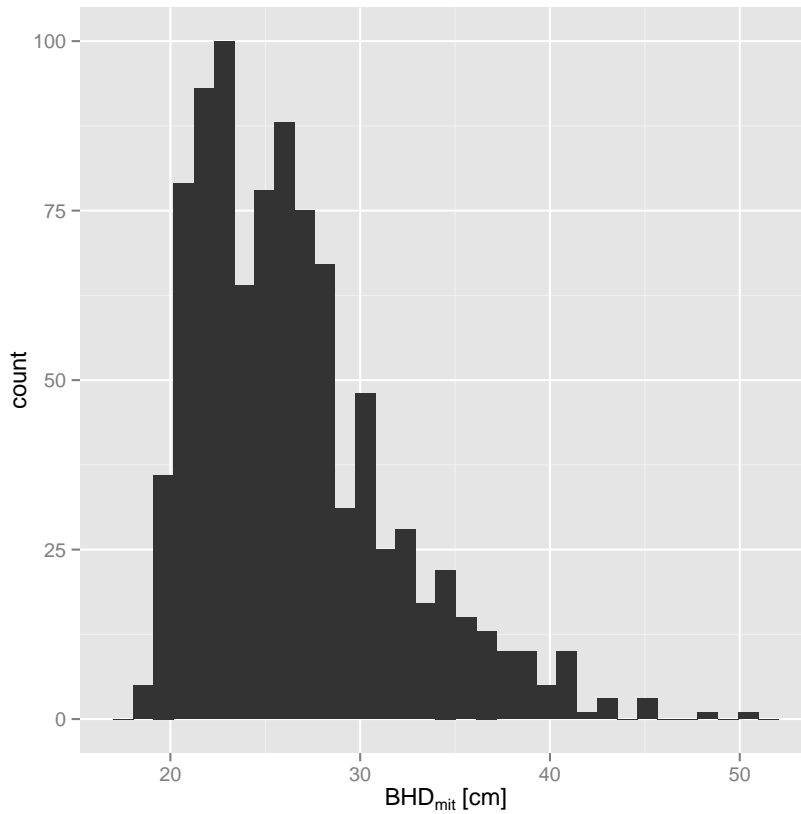


Abbildung 30: Histogramm der BHD_{mit} im zur Modellierung benutzten Datenset.

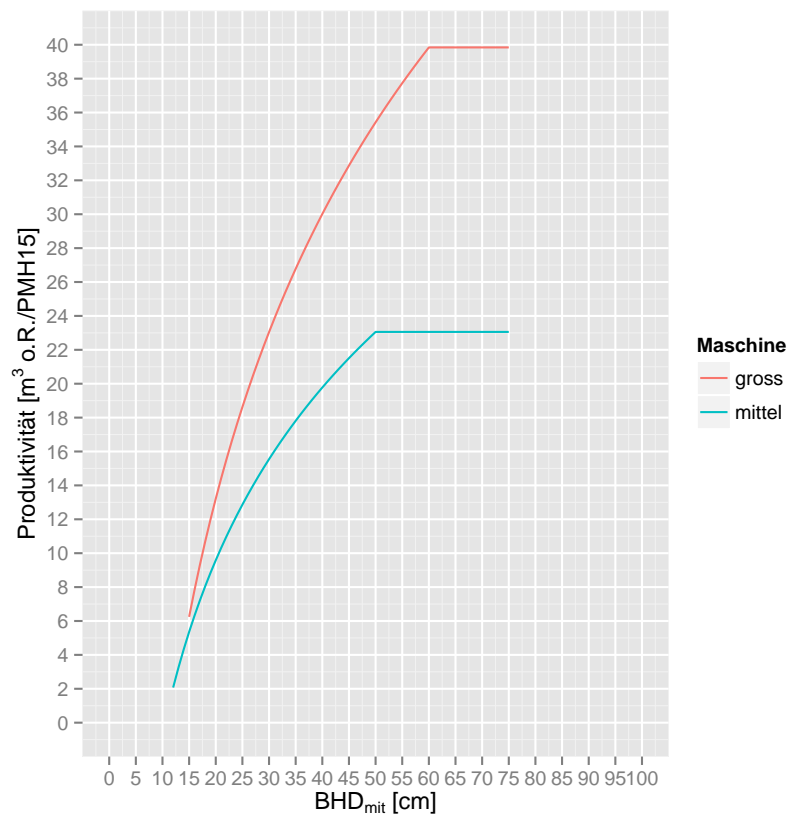


Abbildung 31: Zusammenhang BHD_{mit} und Produktivität mit Speziallösungen für Fälle mit sehr kleinem oder grossem BHD_{mit} .

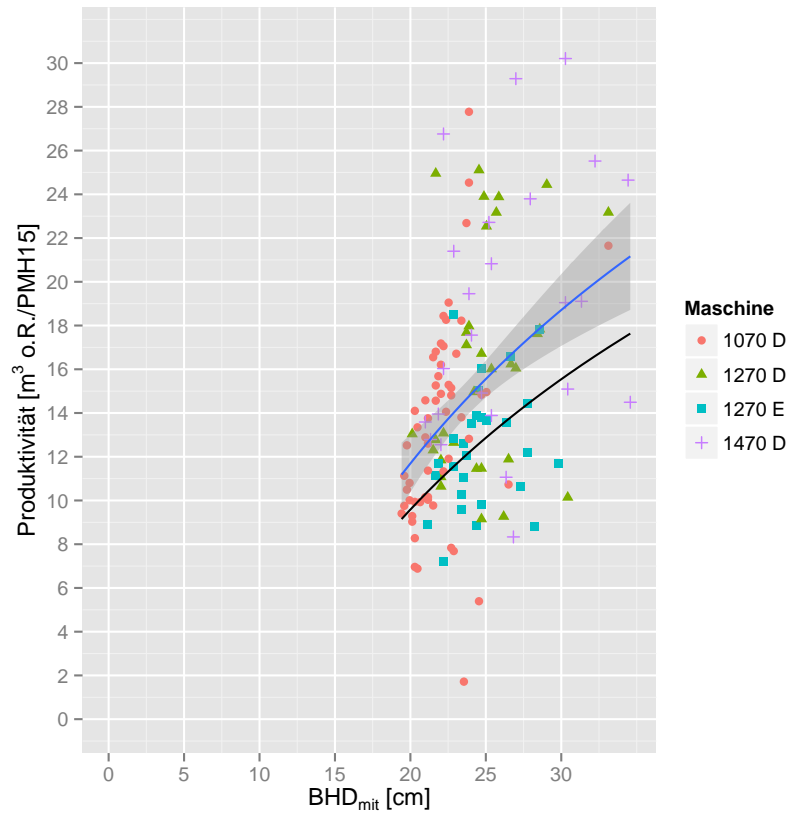


Abbildung 32: Vergleich mit Schweizer Testdaten (N= 143, aus den Jahren 2010 bis 2014). Die farbigen Punkte stehen für die Messwerte (mit der Trendlinie und 95% Vertrauensintervall in blau bzw. grau), die schwarze durchgezogene Linie für die Vorhersagen des Modells (Maschine = mittel, BHD_{mit} [cm] aus Mittelstammvolumen geschätzt (via Luzerner Tarif III)). Der mittlere relative Schätzfehler beträgt 29% (Median = 22%).

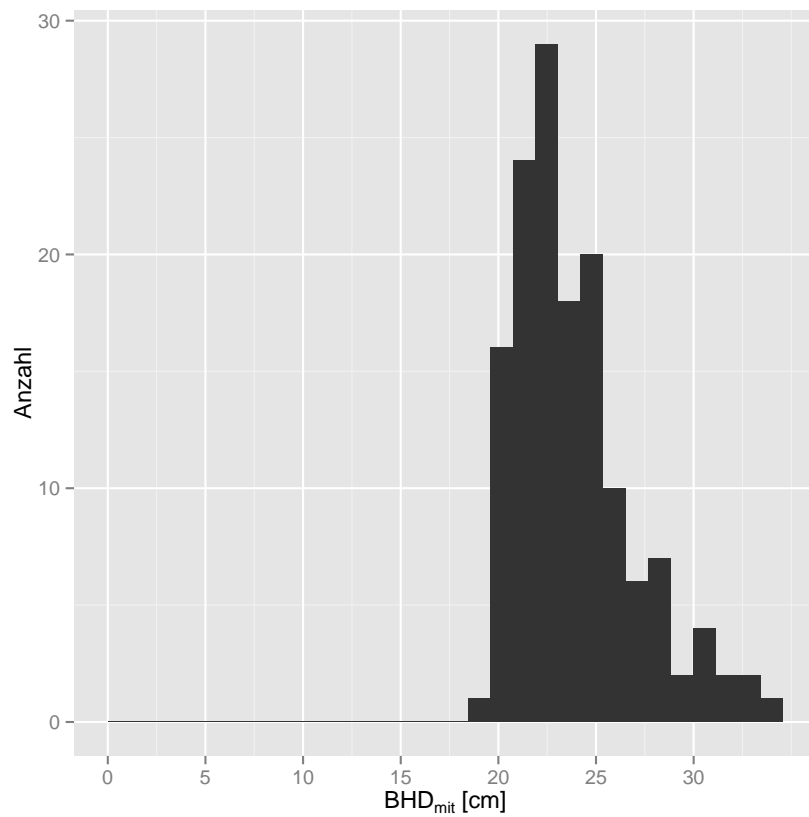


Abbildung 33: Histogramm für die BHD_{mit} der Schweizer Vergleichsdaten. Die BHD_{mit} der CH Vergleichsdaten sind ähnlich verteilt wie in den originalen Modellierdaten (vergleiche Abbildung 30).