

Produktivitätsmodell
"Fällen und Vorrücken von Vollbäumen 2017"

Teil B:Analyse der Datensätze und Diskussion
der Modellierung

Dario Pedolin, WSL Birmensdorf

19. Dezember 2017

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

Zusammenfassung

gfhftugjhgjg

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	8
3.1	Diskussion der Datengrundlage	8
4	Modelle einsetzen	11
4.1	Formel, Koeffizienten und Systemgrenzen	11
4.2	Sensitivitätsanalysen	12
5	Anhang	13
5.1	Testen auf Verletzungen der Annahmen für multiple Regression .	13
5.2	Beschreibung der 'Ausreisser'	20
5.3	Kreuzvalidierung	22
5.4	Beschreibung der im Modell benutzten Messgrößen	25

1 Einleitung

Ein Datensatz bestehend aus 51 Vorliefer-Arbeiten aus den Jahren 2014 bis 2015 aus zentraleuropäischen Forstbetrieben wurde benutzt, um ein Modell für die Produktivität des Vorrückens 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 4 bis 7). Abschliessend wurde eine Sensitivitätsanalyse für die Parameter durchgeführt. Die Qualität des Modells wurde ausserdem mit einer Kreuzvalidierung getestet (Anhang Abbildung 10).

2 Analyse Produktivitäten

2.1 Beschreibung der Prozesse

Der Prozess 'Fällen und Vorrücken' umfasst alle Arbeitsschritte des Fällens und Vorrückens von Vollbäumen.

2.2 Beschreibung der Einheiten und Masse

Alle Zeiten (Stunden [h]) sind [PSH15] (Produktive Arbeitszeit = MAS, vergleiche Anhang Abbildung 17) und beinhalten alle Zeitaufwände des Fällens und Vorrückens von Vollbäumen. Die Mengenangaben [m³] sind Kubikmeter Derbholz Vollbaum in Rinde.

2.3 Erstellen und Testen des Modells

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

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.8390	0.9379	1.96	0.0557
Vmit	6.4133	1.0980	5.84	0.0000
MaschineSchlepper	1.4059	0.8959	1.57	0.1232

Tabelle 1 zeigt die Koeffizienten und p-Werte der ersten Analyse des Modells. Das korrigierte R² für die Modellbildung beträgt 0.39.

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 entfernt¹ (siehe Anhang Tabelle 5 für Beschreibung der Ausreisser).

Für die nachfolgenden Auswertungen wurden die in Abbildung 1 identifizierten Ausreisser entfernt. Es werden noch 50 Fälle benutzt.

¹Die 1 Ausreisser entsprechen 2% der Fälle der Grundgesamtheit.

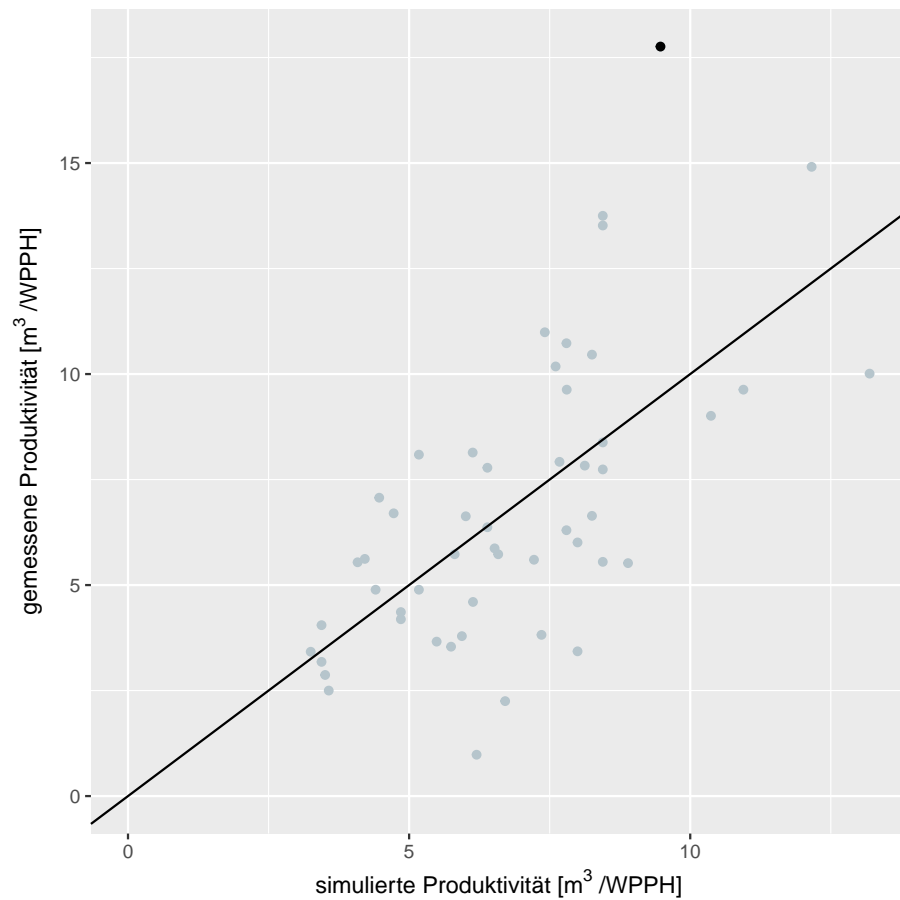


Abbildung 1: Vergleich der mittels des Modells berechneten mit der gemessenen Produktivität [m³ /PSH15]. Messwerte deren Residuen (= gemessener Wert - durch das Modell geschätzter Wert) mehr als zwei Standardabweichungen über dem Mittelwert der Residuen liegen, sind schwarz eingefärbt.

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)	2.1003	0.8438	2.49	0.0164
Vmit	5.7734	1.0003	5.77	0.0000
MaschineSchlepper	1.5344	0.8038	1.91	0.0624

Tabelle 2 zeigt, dass vor allem die Variable 'BHD' für die Produktivität des Forwarders massgebend ist. Das korrigierte R^2 für die Modellbildung beträgt 0.4. Zusätzlich wurden die in Anhang Abbildungen 4 bis 7 gezeigten Tests durchgeführt, um auf Verletzungen der Voraussetzungen der Annahmen für einen linearen Zusammenhang zu testen. Weiter wurde für das Modell eine Kreuzvalidierung durchgeführt (siehe Anhang Tabelle 10).

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 1.8 m^3 /WPPH (Median: 1.5 m^3 /WPPH). Um ein Mass für die relative Genauigkeit der durch das Modell berechneten Werte zu bekommen, wurden auch die relative Abweichung der berechneten Werte von den gemessenen Werten untersucht (d.h. (gemessener Wert - berechneter Wert) / gemessener Wert). Die mittlere relative Abweichung beträgt 40% (Median: 24%). Vergleiche auch Anhang Abbildung 7 und 9 für einen Überblick über die Verteilung der Residuen.

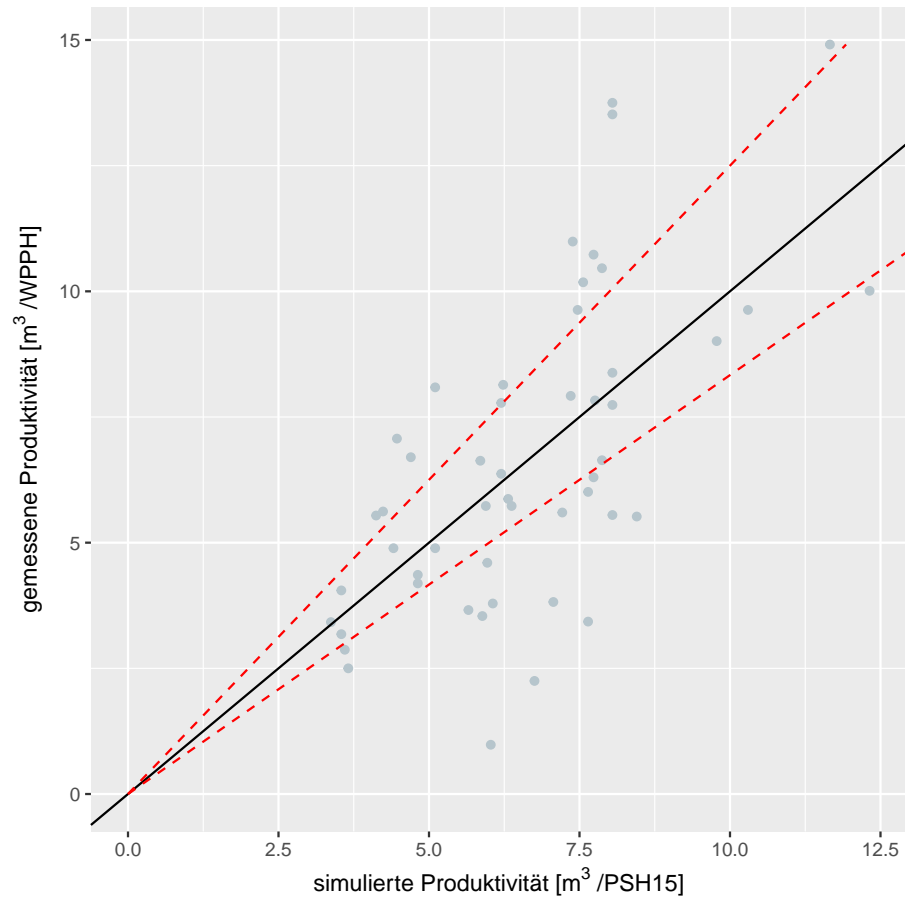


Abbildung 2: Vergleich der mittels des Modells berechneten Produktivität mit der gemessenen Produktivität. Die roten gestrichelten Linien grenzen den Bereich von 20% Abweichung ein.

3 Diskussion

3.1 Diskussion der Datengrundlage

Die Datengrundlage umfasst sowohl kombinierte Vorruecke / Zufaell Arbeiten als auch reines Vorruecken. Dazu wurden Schlepper und Raupen eingesetzt.

benutzt	Maschine	Maschinentyp	N	mean_vmit	min_vmit	max_vmit	min_totFM	max_totFM
ja	Raupe	WC	17	0.66	0.22	1.19	6.80	141.61
ja	Raupe	Wicki	22	0.84	0.25	1.77	15.68	72.42
ja	Schlepper	Same	9	0.64	0.35	1.39	45.60	587.97
ja	Schlepper	WFTrec	3	0.52	0.39	0.62	70.98	195.30
nein	Schlepper	Welte	4	0.85	0.57	1.42	29.07	73.84

Tabelle 3: Maschinen im originalen Datenset

benutzt	Maschine	Maschinentyp	N	mean_vmit	min_vmit	max_vmit	min_totFM	max_totFM
ja	Raupe	WC	16	0.62	0.22	1.10	6.80	84.00
ja	Raupe	Wicki	22	0.84	0.25	1.77	15.68	72.42
ja	Schlepper	Same	9	0.64	0.35	1.39	45.60	587.97
ja	Schlepper	WFTrec	3	0.52	0.39	0.62	70.98	195.30

Tabelle 4: Maschinen im benutzten Datenset

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 'Vorrücken' lautet:

$$\begin{aligned} \text{Produktivität } \left[\frac{m^3}{PSH15} \right] = & a \\ & + b * \text{Mittelstammvolumen } [m^3] \\ & + c_{\text{Maschine}} \end{aligned} \quad (1)$$

Formel 1 zeigt das Modell aus Tabelle 2 in mathematischer Form (die Zeiten in Stunden sind als 'Produktive arbeitseit' (PSH15 in Abbildung 17)). Das Modell umfasst alle Teilprozesse des Vorrückens. Die Koeffizienten dazu sind:

	Koeffizient	Wert
(Intercept)	a	2.10
Vmit	b	5.77
MaschineSchlepper	c	1.53
MaschineRaupe	c	0.00

4.2 Sensitivitätsanalysen

Die nachfolgenden Abbildung 3 der Sensitivitätsanalyse zeigt die Abhängigkeit der durch das Modell berechneten Produktivität vom Mittelstammvolumen und der eingesetzten Maschine. Die Punkte markieren die für die gemessenen Produktivitäten, die Linien die durch das Modell vorhergesagten Produktivitäten. Die Sensitivitätsanalyse zeigt wie sich die berechneten Werte in Abhängigkeit der Modellparameter verändern.

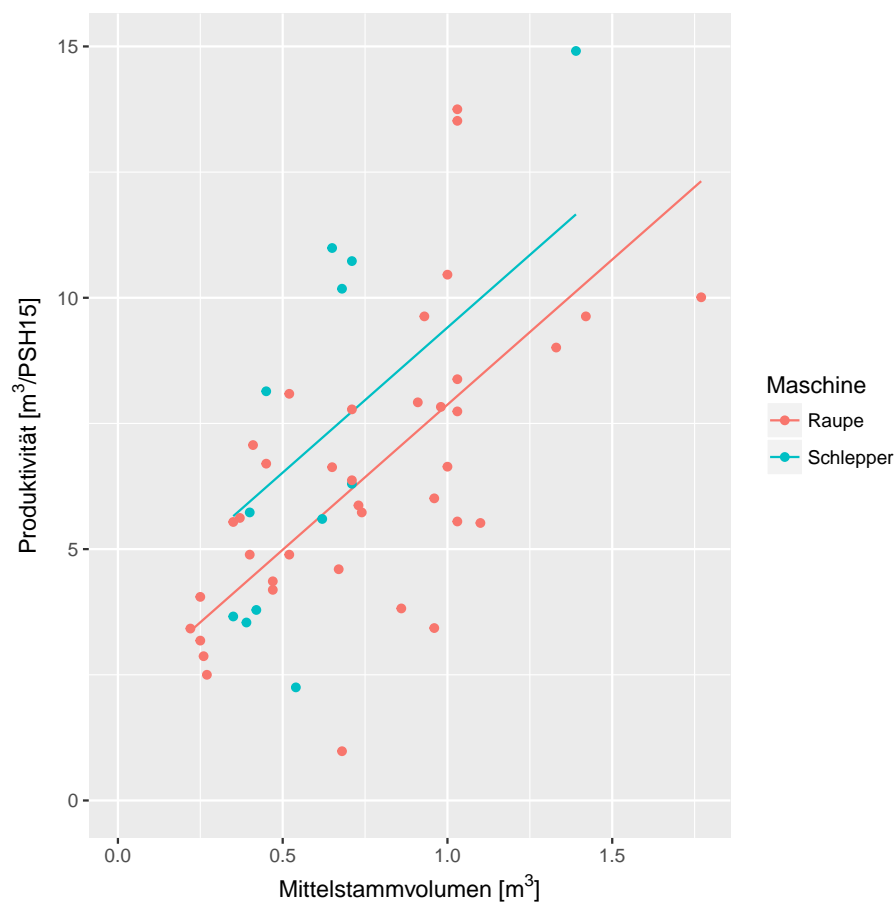


Abbildung 3: Sensitivitätsanalyse für das Modell Vorrücken

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. (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 8).)
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.

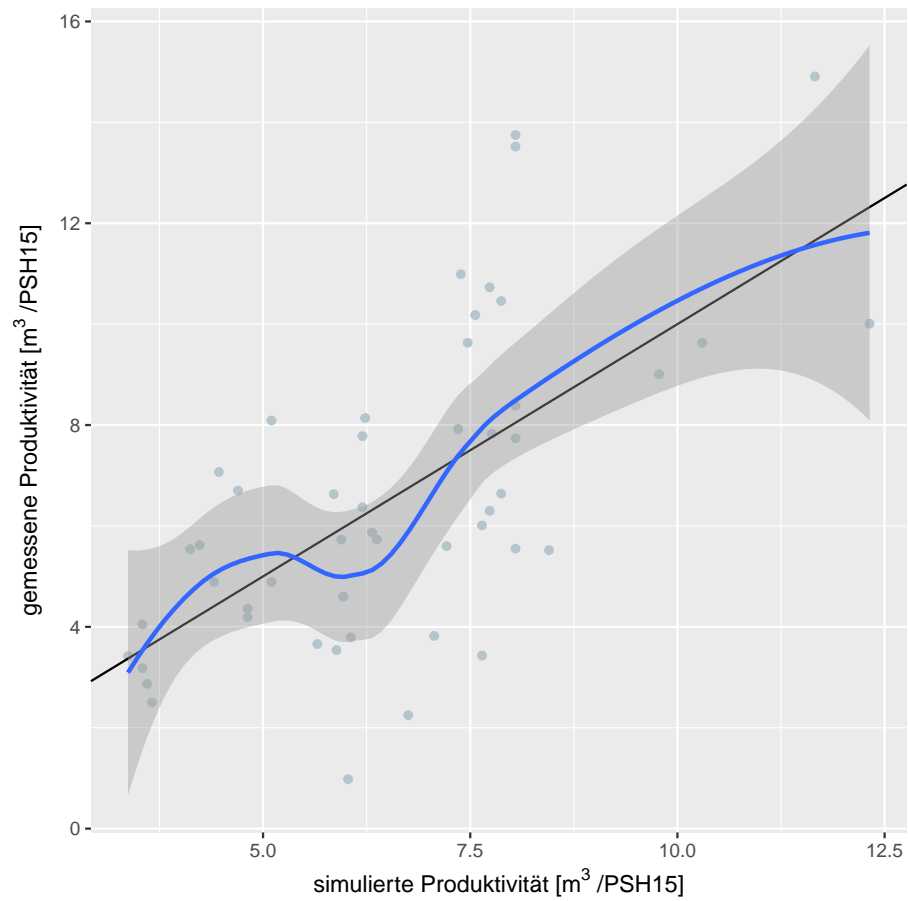


Abbildung 4: Test Modellannahmen 1: Linearität. Die Punkte streuen um die Diagonale, es gibt keine Anzeichen für einen systematischen Fehler.

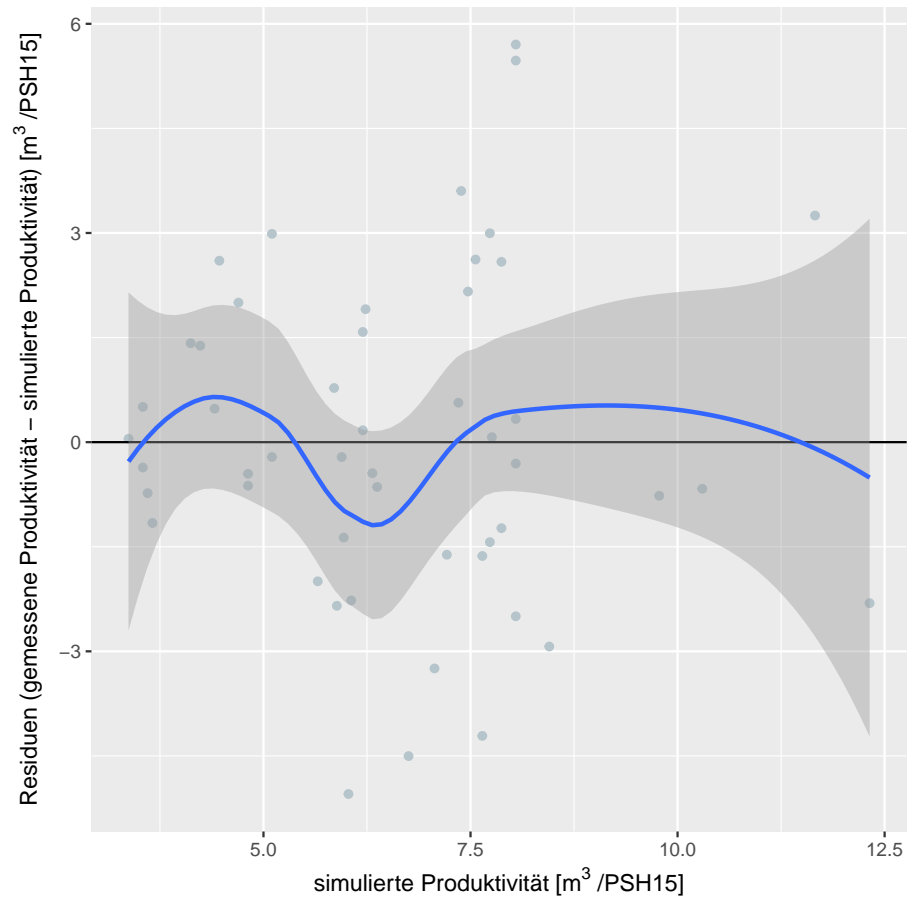


Abbildung 5: Modellannahme 1: Normalverteilung der Residuen. Tukey-Anscombe Plot für das Modell aus Tabelle 1. Der p-Wert der ANOVA der Mittelwertunterschiede der Residuen über die vier Quartile beträgt 0.985

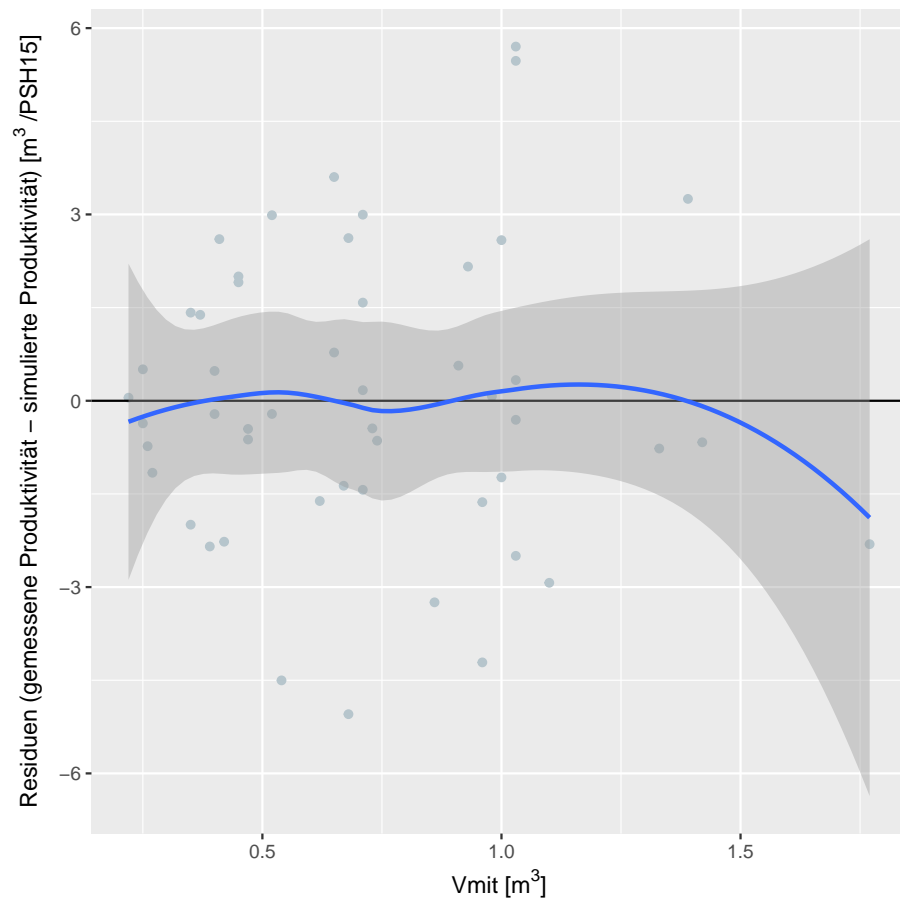


Abbildung 6: Test Modellannahmen 2: Korrelation der Residuen mit unabhängiger Variable 'Vmit [m³]'. Die blaue Linie steht für das gleitende Mittel. Der p-Wert der ANOVA der Mittelwertunterschiede der Residuen über die vier Quartile beträgt 0.589

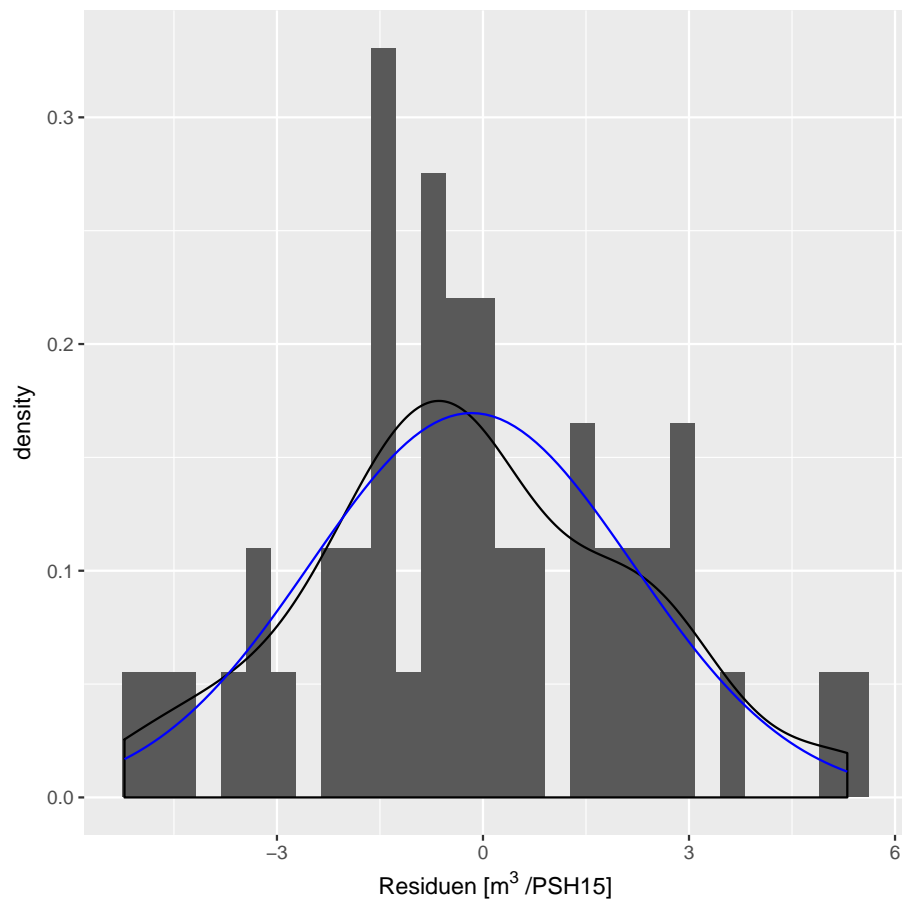


Abbildung 7: Test Modellannahmen 4: Normalverteilung der Residuen: Dichtefunktion der Residuen. Als Referenz ist die Dichtefunktion einer normalverteilten Variable blau eingezeichnet. Die Residuen sind annähernd normalverteilt.

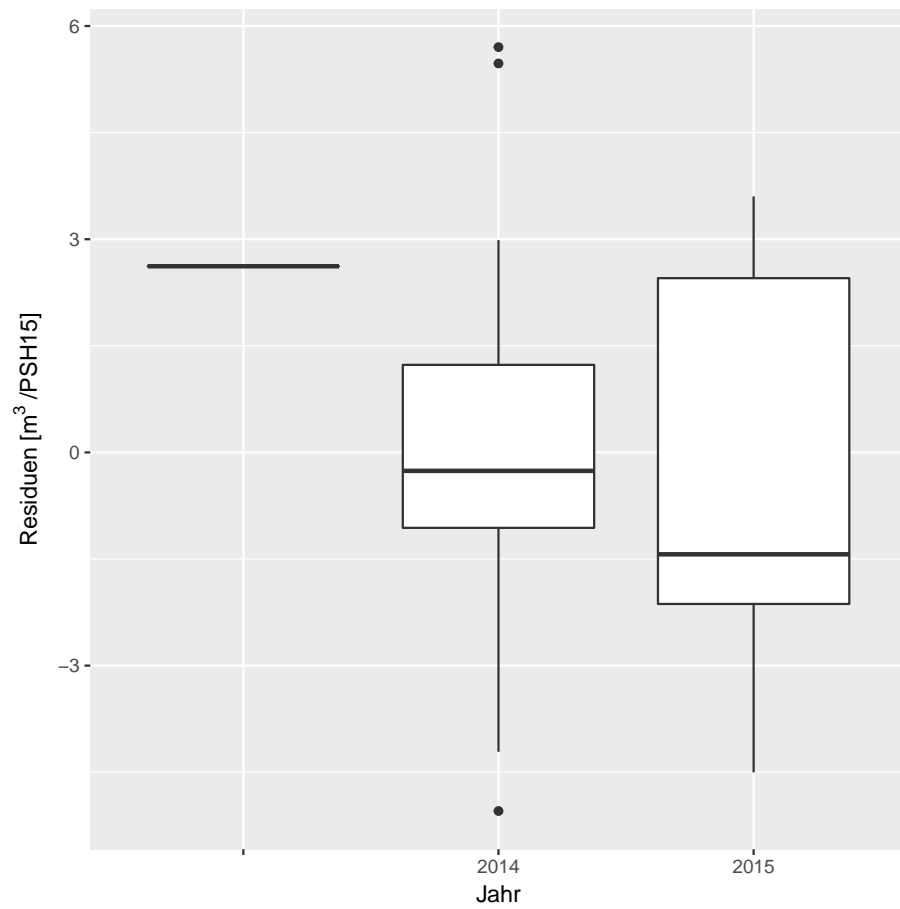


Abbildung 8: Boxplot der Residuen ueber die Jahre

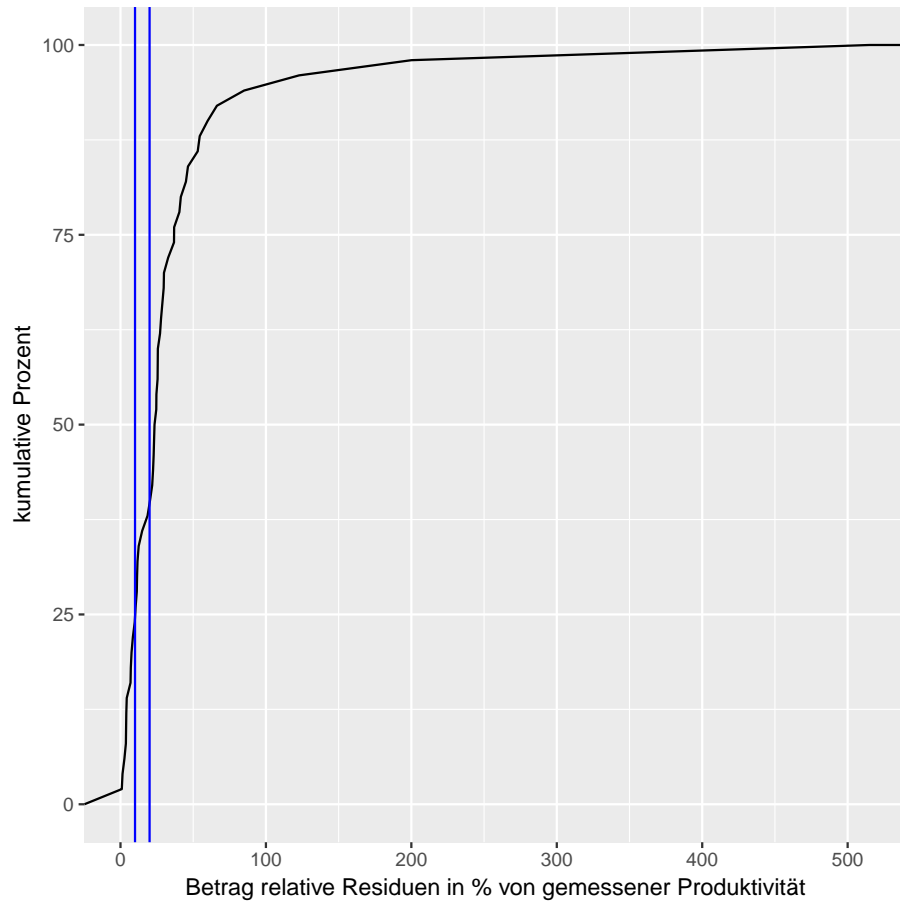


Abbildung 9: Kumulative Verteilung der Beträge der relativen Residuen. Die vertikalen blauen Linie markieren 10% bzw. 20% Abweichung. Bei 24% der Fälle ist der relative Schätzfehler kleiner gleich 10%.

5.2 Beschreibung der 'Ausreisser'

Tabelle 5: Beschreibung der Ausreisser			
Jahr	Anmerkung	Maschine	finPROmas
2014		Raupe	17.76

5.3 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 6: Resultat der Analyse des linearen Zusammenhangs und Schätzung der Modellkoeffizienten für den Trainingsdatensatz

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.726	0.815	3.344	0.002
Vmit	4.460	0.916	4.868	0.000
MaschineSchlepper	1.443	0.740	1.949	0.060

Tabelle 6 gibt die Koeffizienten-Schätzungen für das Trainingsdatenset wieder. Das korrigierte R^2 für die Modellbildung beträgt 0.4.

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

	p-Wert	alle Daten	trainings Daten	% Diff
(Intercept)	0.016	2.100	2.726	-30.0
Vmit	0.000	5.773	4.460	23.0
MaschineSchlepper	0.062	1.534	1.443	6.0

Tabelle 7 zeigt die Unterschiede für die Koeffizientenschätzung aus allen Daten (N= 50) vs. Koeffizientenschätzung aus Trainingsdaten (N= 34). Der p-Wert bezieht sich auf die Schätzung aus allen Daten.

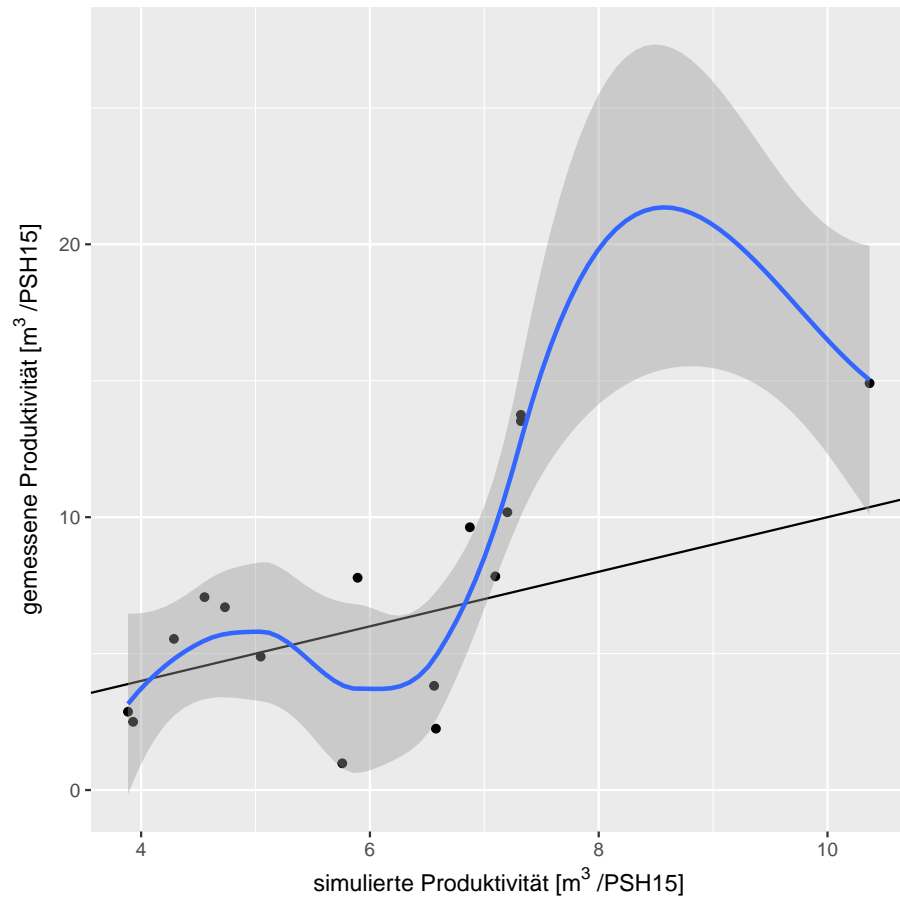


Abbildung 10: Der Korrelationskoeffizient zwischen simulierter Produktivität [$\text{m}^3 / \text{PSH15}$] und gemessener Produktivität [$\text{m}^3 / \text{PSH15}$] für das Validierungsset beträgt 0.72.

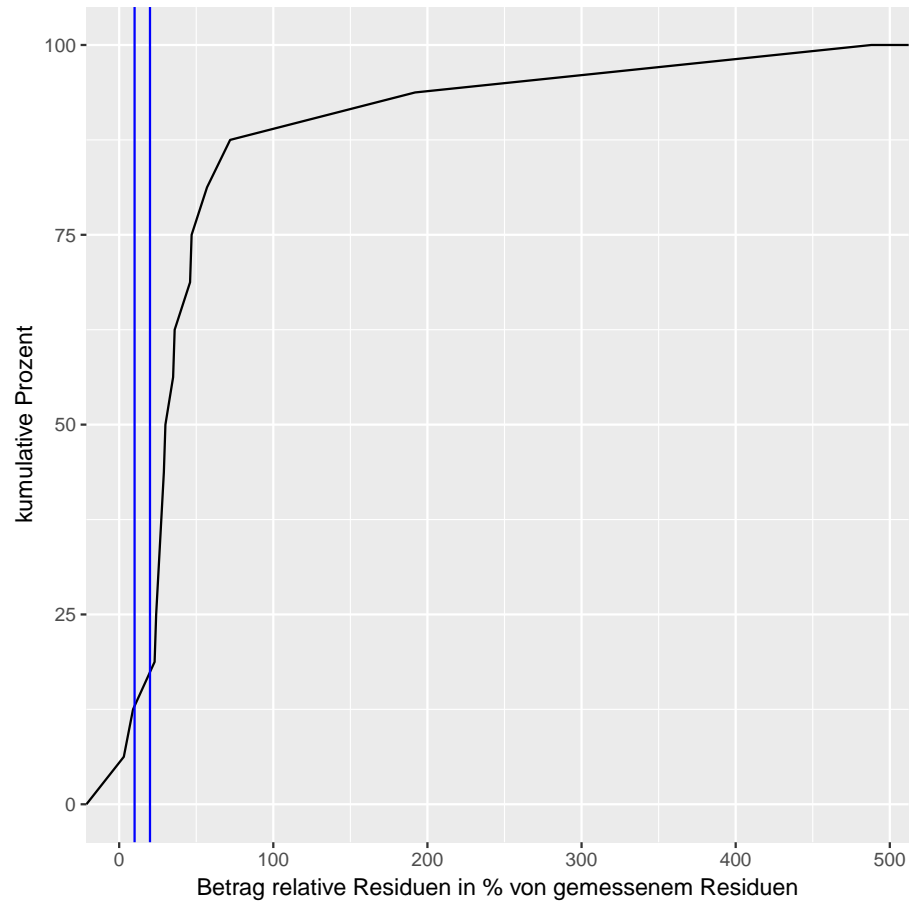


Abbildung 11: Kumulative Verteilung der Beträge der relativen Residuen aus der Kreuzvalidierung. Die vertikalen blauen Linie markieren 10% bzw. 20% Abweichung.

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 = standard Fehler des Mittelwerts.

Tabelle 8: Beschreibung der Produktivität [m^3 /PSH15]

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	50	6.60	3.08	5.94	0.98	14.91	0.69	0.44
Trainingsdaten	34	6.35	2.30	5.80	3.18	10.99	0.49	0.40
Validierungsdaten	16	7.14	4.33	6.88	0.98	14.91	0.37	1.08

Tabelle 9: Beschreibung der Variable 'Kubikmeter liegend' [m^3]

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	50	80	105	45	7	588	3	15
Trainingsdaten	34	73	71	46	18	303	2	12
Validierungsdaten	16	94	158	33	7	588	2	39

Tabelle 10: Beschreibung der Variable Vmit' [m^3]

Daten	n	mean	sd	median	min	max	skew	se
Alle Daten	50	1	0	1	0	2	1	0
Trainingsdaten	34	1	0	1	0	2	1	0
Validierungsdaten	16	1	0	1	0	1	0	0

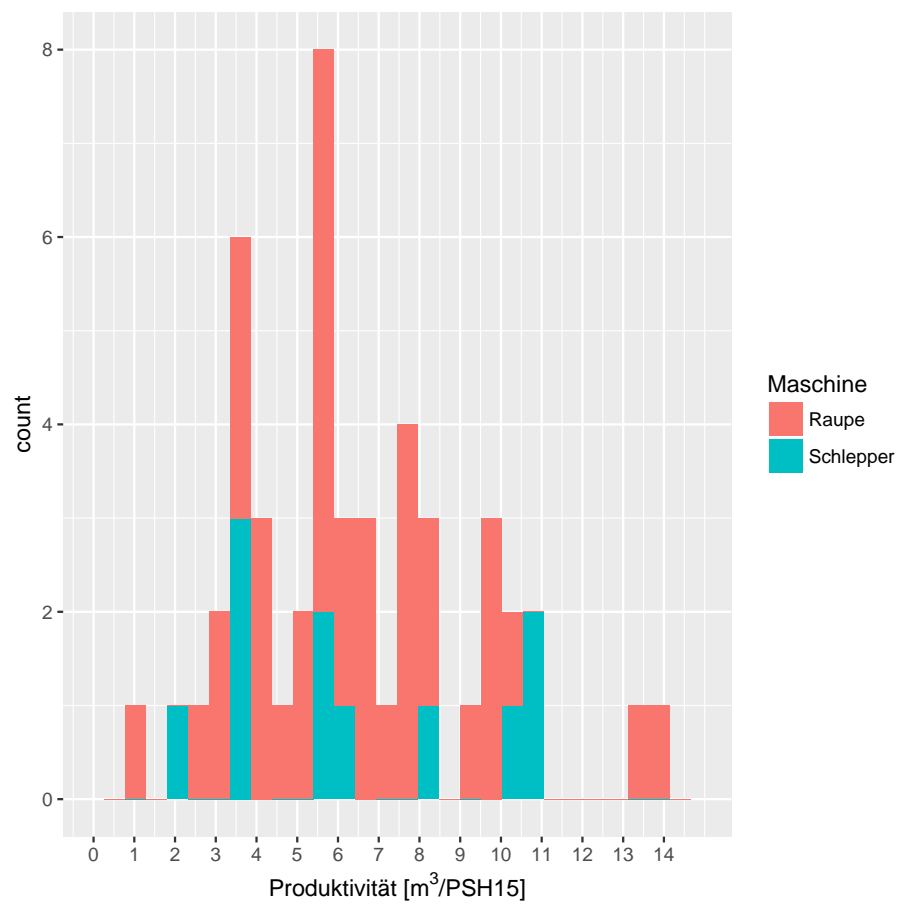


Abbildung 12: Verteilung der Variable 'totale Produktivität [m³/PSH15]'.

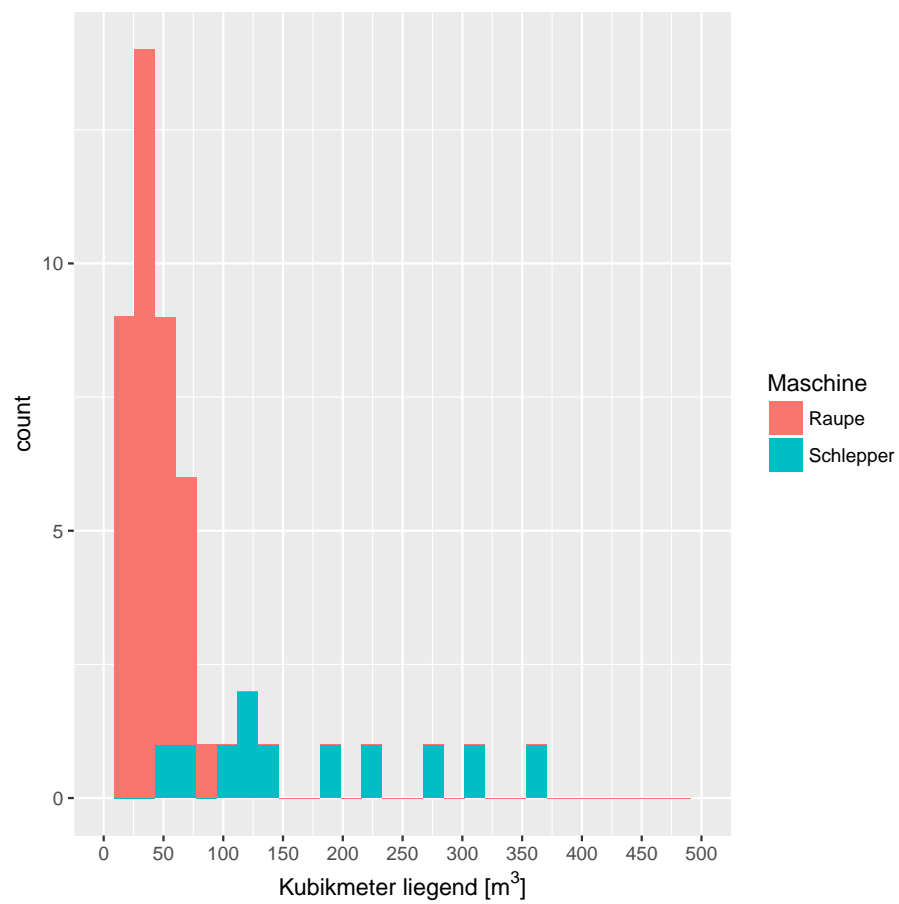


Abbildung 13: Verteilung der Variable 'Kubikmeter liegend [m³]'.

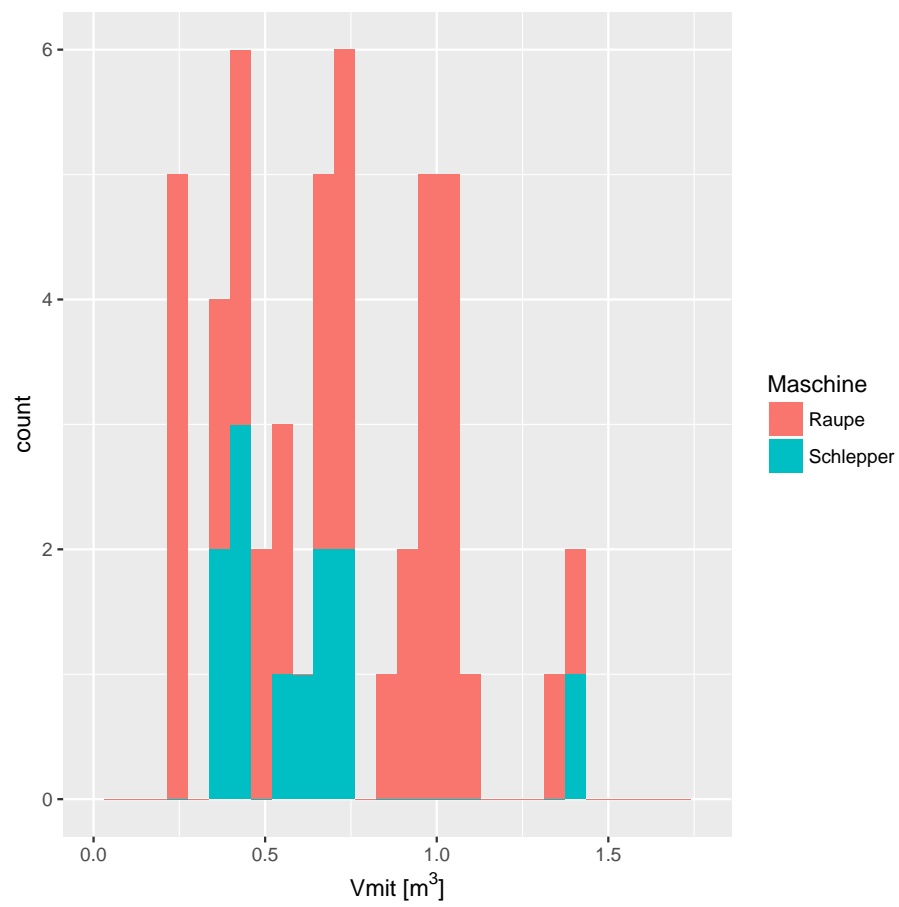


Abbildung 14: Verteilung der Variable 'Vmit [m³]'.

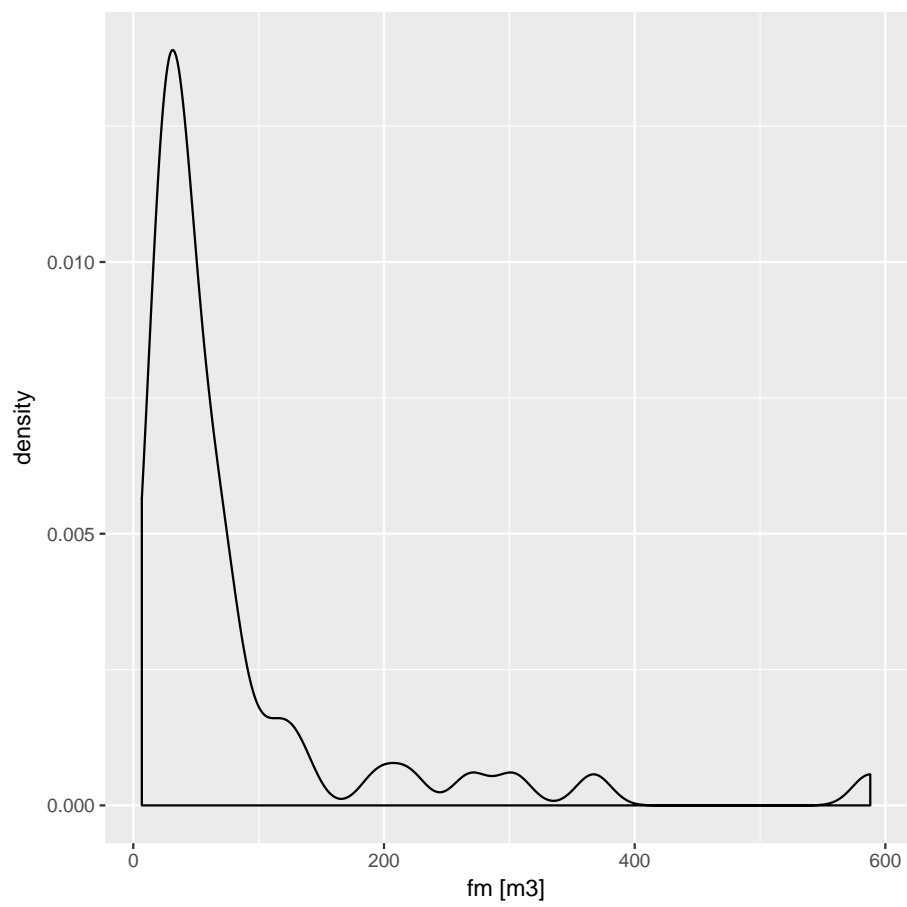


Abbildung 15: Verteilung der Variable 'fm [m³]'.

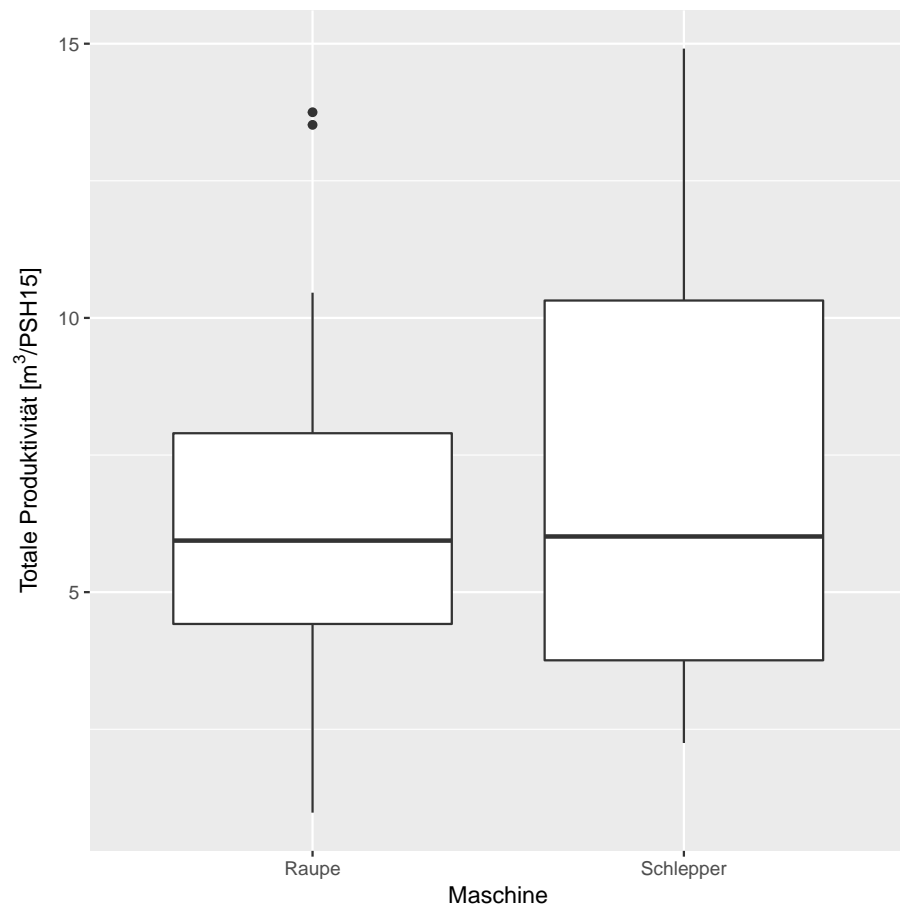


Abbildung 16: Boxplot der gemessene totalen Produktivitäten für die verschiedenen Maschinen.

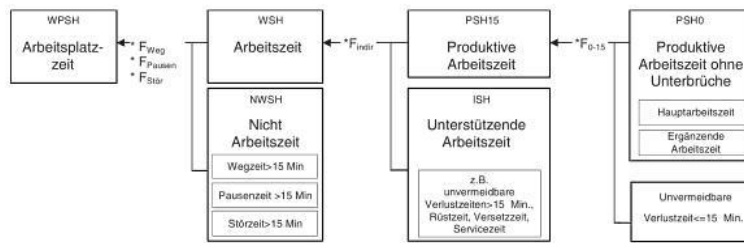


Abbildung 17: Zeitsystem

Version: HeProMo`Vorliefern`neu