

**Anhang zu Teil E des Schlussberichts**  
**"Entwicklung einer effizienten forstlichen Betriebsinventur":**

Schlussbericht  
Betriebsinventur  
Analyse Lidar Daten Vegetationshoeohenmodell

WSL Birmensdorf

23. Februar 2018

Co-Autoren: Dario Pedolin  
Dr. Leo Bont

Leitung: Dr. Leo Bont

**Zusammenfassung**

Ziel: Herstellen, Testen und Beschreiben von Regressionsmodellen fuer verschiedene waldbaulich Interessante Parameter wie Vorratsvolumen, Stammzahl und Basalflaeche aufgrund von Eigenschaften des CHM (Vegetationshoeohenmodell).

Aus einem aus LIDAR Daten erstellten Vegetations-Hoeohenmodell (CHM) soll mittels gemessenen Stichprobendaten des Forstbetriebs Wagenrain (Kanton AG, Schweiz) ein Modell fuer Waldbaulich interessante Parameter erstellt und mittels Testdaten aus Kanton FR validiert werden. Die Daten werden dazu in R bearbeitet (R Development Core Team (2006)).

Der Datensatz umfasst 337 Faelle fuer das Trainings-/Modellier Datenset aus dem Kanton AG und 118 Faelle im Testdatenset aus dem Kanton FR. Die 337 vermessene Stichproben im Trainingsdatenset beinhalten im Schnitt  $265.3 \text{ m}^3/\text{ha}$  Holz pro Stichprobe ("TS\_TotalVol", bzw.  $146.6 \text{ m}^3/\text{ha}$  Nadelholz ("TS\_NdH\_Vol") und  $118.7 \text{ m}^3/\text{ha}$  Laubholz ("TS\_LbH\_Vol") pro Stichprobe (Stichprobenflaechengroesse =  $400\text{m}^2$ ). Die in diesem Bericht praesentierten Modelle haben ein kor. R-Quadrat von 0.316 (Stammzahl) bis 0.587 (TotalVol). Zu jedem Modell wurde ein Test mit unabhaengigen Daten (aus Kanton FR), eine Kreuzvalidierung, Bootstrap sowie eine Residuenanalyse durchgefuehrt.

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>4</b>
1.1	Beschreibung der benutzten Daten . . . . .	4
1.2	Methoden . . . . .	5
1.2.1	Koeffizienten schätzen . . . . .	5
1.2.2	Test mit unabhaengigen Daten . . . . .	5
1.2.3	Bootstrap . . . . .	5
1.2.4	Kreuzvalidierung . . . . .	5
1.2.5	Residuenanalyse . . . . .	6
1.2.6	Alternative Modelle . . . . .	6
1.2.7	Ausreisser . . . . .	6
<b>2</b>	<b>Resultate</b>	<b>7</b>
2.1	Volumen Vorrat [ $\text{m}^3/\text{ha}$ ] . . . . .	7
2.1.1	Modell . . . . .	7
2.1.2	Test mit unabhaengigen Daten . . . . .	9
2.1.3	Bootstrap . . . . .	11
2.1.4	Kreuzvalidierung . . . . .	13
2.1.5	Residuenanalyse . . . . .	16
2.1.6	Alternative Modelle . . . . .	21
2.1.7	Ausreisser . . . . .	23
2.2	Basalflaeche [ $\text{m}^2/\text{ha}$ ] . . . . .	25
2.2.1	Modell . . . . .	25
2.2.2	Test mit unabhaengigen Daten . . . . .	27
2.2.3	Bootstrap . . . . .	29
2.2.4	Kreuzvalidierung . . . . .	31
2.2.5	Residuenanalyse . . . . .	35
2.2.6	Alternative Modelle . . . . .	40
2.2.7	Ausreisser . . . . .	42
2.3	Stammzahl [ $\text{ha}^{-1}$ ] . . . . .	44
2.3.1	Modell . . . . .	44
2.3.2	Test mit unabhaengigen Daten . . . . .	46
2.3.3	Bootstrap . . . . .	48
2.3.4	Kreuzvalidierung . . . . .	50
2.3.5	Residuenanalyse . . . . .	53
2.3.6	Alternative Modelle . . . . .	58
2.3.7	Ausreisser . . . . .	60
<b>3</b>	<b>Diskussion der Resultate</b>	<b>62</b>
3.1	Modell fuer Vorrat . . . . .	62
3.2	Modell fuer Basalflaeche . . . . .	62
3.3	Modell fuer Stammzahl . . . . .	62
3.4	Uebersicht ueber die drei Modelle . . . . .	62
3.4.1	Fazit . . . . .	65

<b>4</b>	<b>Anhang</b>	<b>68</b>
4.1	Beschreibende Analysen . . . . .	69
4.1.1	Volumen . . . . .	75
4.1.2	Laubholz Volumen . . . . .	77
4.1.3	Nadelholz Volumen . . . . .	81
4.1.4	Basalflaeche . . . . .	83
4.1.5	LbH Anteil und MW BHD . . . . .	85
4.2	Korrelationen im Trainingsdatenset . . . . .	87
4.3	Details zu den Resultaten . . . . .	89
4.3.1	Code der Filter . . . . .	89
4.4	Dokumente . . . . .	90

# 1 Einleitung

## 1.1 Beschreibung der benutzten Daten

### Einheiten und Variablen

- Die Volumen sind in Rinde d.h.  $[\text{m}^3/\text{ha}] = [\text{m}^3/\text{ha i. Rinde}]$ .
- Die LIDAR Rohdaten werden mit den LAS\_ Parametern beschrieben (Kennzahlen). Dazu wurden nur Werte ueber 1.37m benutzt.
- Aus den Lidar Rohdaten und Orthophotos<sup>1</sup> wird ein CHM (canopy height model) erstellt. Dazu wurden alle LIDAR Daten benutzt.
- Die TS\_ Daten(terrestrische sample) beschreiben die gemessenen (wahren) Volumen  $[\text{m}^3/\text{ha}]$ .

### Manipulationen<sup>2</sup>

- Fuer die Daten aus Kanton FR wurden die TS\_Total\_Vol mit 10000 multipliziert in  $\text{m}^3/\text{ha}$  umgerechnet
- Fuer die Daten aus Kanton AG wurden die TS\_Total\_Vol mit 10 multipliziert in  $\text{m}^3/\text{ha}$  umgerechnet
- Fuer die Daten aus FR wurden die X und Y Koordinaten (sowie X\_sys und Y\_sys) von Format "CH 1903+ / LV95" ins Format "CH 1903 / LV03" (wie in AG Daten) umgerechnet
- Fuer die Umrechnung der CH 1903 Daten in Laengen- und Breitengrad wurden die Variablen X\_sys und Y\_sys benutzt.
- Die Variable Stammzahl wurde mit dem Faktor 10 multipliziert

### Filter<sup>3</sup>

- Es werden nur die 455 Faelle mit gemessenen Werten (aus terrestrischen Aufnahmen) benutzt. (Kanton AG N = 337, FR = 118) (0 Faelle ausgeschlossen).
- 38 Faelle mit Waldrand = 1 wurden von der Analyse ausgeschlossen.
- 8 Faelle mit  $\text{chm\_surf\_int} < 10$  und  $\text{TS\_Total\_Vol} > 2\text{m}^3/\text{ha}$  wurden von der Analyse ausgeschlossen.

### Benutztes Datenset

Das benutzte Datenset umfasste 455 Faelle.

---

<sup>1</sup>Geo-Referenzierte Luftbilder

<sup>2</sup>Ueberschneidungen von Kriterien moeglich

<sup>3</sup>Fuer detailliertere Beschreibung der Filter siehe Code Abbildung 58 im Anhang

## 1.2 Methoden

### 1.2.1 Koeffizienten schätzen

Fuer die Parameter bei denen ein log-linearer Zusammenhang zu Canopy Height Modell (CHM) Werten erwartet wird, wurde die invers-hyperbolische-sinus-Transformation (ihst) (Oyamakin and Chukwu (2014)) benutzt (In den unabhaengigen Daten des CHM sind Werte von 0 erlaubt und sinnvoll).

Fuer Laub- und Nadelholz Baeume wurde erwartet, dass der Zusammenhang CHM und Volumen fuer die Baumarten unterschiedlich ist. Deshalb wurden der Parameter `chm_surf_int_Laub_Anteil`<sup>4</sup> berechnet und im Modell benutzt. Diese (relativen) Anteile wurden mit "arcsine-square-root" Transformatiert (Sokal and Rohlf (1995))

### 1.2.2 Test mit unabhaengigen Daten

Mit unabhaengigen Daten aus Kanton Fribourg ( $N = 118$ ) wurde ein Test der geschaeztzten Koeffizienten durchgefuehrt. Zudem wurde Koeffizienten fuer das selbe Modell mit den FR Daten neu geschaeztzt und mit den Werten fuer AG verglichen.

### 1.2.3 Bootstrap

Fuer jedes der Modelle wurden 10000 Bootstraps durchgefuehrt (Davison and Hinkley (1997)). Fuer jede "Bootstrap-Wiederholung" wird eine zufaellig gewaehlte Stichprobe (mit Zuruecklegen) aus dem Datensets benutzt um die Modell-Koeffizienten zu schaezten<sup>5</sup>. Die resultierende Verteilung der geschaeztzten Koeffizienten (die hier gezeigten Standardfehler bzw. 95% Vertrauensintervalle) geben Aufschluss darueber wie sensitv das Modell auf das Ein-/Ausschliessen einzelner Faelle reagiert. Je robuster das Modell desto kleiner sind die Standardfehler und damit die Vertrauensintervalle fuer die Koeffizientenschaetzungen.

### 1.2.4 Kreuzvalidierung

Fuer jedes der Modelle wurden 10 Kreuzvalidierungen durchgefuehrt Maindonald and Braun (2015)). Fuer jeden Durchgang der Kreuzvalidierung (in Abbildung 4, 18 und 32 "Fold" genannt) wird das Datenset zufaellig zweigeteilt. Aus dem einen Teil der Daten werden Koeffizienten geschaeztzt und diese werden dann mit dem zweiten Teil der Daten getestet (kreuzvalidiert).

Zudem wurden 1000 Modellierungen mit jeweils 10% der Daten gemacht und mit den restlichen Daten getestet. Dieses Vorgehen wurde gewaehlt, um den "realen" Einsatz der Modelle ab zu bilden. In der Praxis sieht das Vorgehen so aus, dass die Modelle anhand einiger weniger gemessenen Werte parametrisiert werden, und dann auf die restlichen Flaechen angewendet werden (analog dazu

---

<sup>4</sup>Verhaeltnis `chm_surf_int_Laub` zu `chm_surf_int_Laub` + `chm_surf_int_Nadel`

<sup>5</sup>Fuer "Bootstrapping" wird aus der Stichprobe mit  $N$  Beobachtungen eine "Stichproben-Stichprobe" von  $N$  Samples MIT zuruecklegen gezogen.

werden hier 10% fuer die "Parametrisierung" benutzt und dann das so parametrisierte Modell auf die restlichen 90% angewendet. Als Ergebnis werden die R-Quadrate und der Median der absoluten relativen Residuen, d.h. die Differenz jeder einzelnen Messung - geschaetzter Wert wird durch den gemessenen Wert geteilt und ergibt die Prozentuale Abweichung vom "wahren" Wert gezeigt.

### 1.2.5 Residuenanalyse

Fuer die lineare Regression werden folgende Voraussetzungen ueberprueft (aus Stahel (2012), Buehlmann (2008) und Fox (2009)):

1. Modellannahme 1: Es gibt einen Linearen Zusammenhang zwischen gemessenen Werten und mit linearem Modell simulierten Werten (Linearitaet).
2. Modellannahme 2: Die Residuen sind unabhaengig von der abhaengigen und den unabhaengigen Variablen (sog. Heteroskedastie). Der Erwartungswert der Residuen ist fuer jeden Wert der Variable = 0 (in den Grafiken durch die blaue Linie (gleitendes Mittel) dargestellt). Die Residuen sind unabhaengig vom simulierten Wert.
3. Modellannahme 3: Falls die Residuen geordnet sind (zBsp. bei Zeitreihen), duerfen sie nicht voneinander abhaengig sein (keine Autokorrelation, zBsp. auf grosse Residuen folgen kleine, etc). Da unsere Daten keine natuerliche Reihenfolge haben, spielt diese Annahme keine Rolle.
4. Modellannahme 4: Die Residuen sind normalverteilt um den Mittelwert 0.

### 1.2.6 Alternative Modelle

Die im Voraus identifizierten "besten" Parameter wurden in diversen linearen Modellen auf korr. R-Quadrat und AIC<sup>6</sup> getestet (siehe Abschnitt 2.1.6, 2.2.6 und 2.3.6).

### 1.2.7 Ausreisser

Das Datenset wurde fuer jedes der drei Modelle auf Ausreisser<sup>7</sup> untersucht. Dazu wurden zwei Methoden angewendet: Bonferroni Outlier Test und Cook's Distanz Plot. (Cook and Weisberg (1982)).

Cook's Distanz untersucht die Daten auf Faelle deren "Anwesenheit" im Datenset fuer die Schaetzung der Koeffizienten "uebermaessig" grossen Einfluss hat. Der Bonferroni Outlier Test untersucht die Residuen auf Ausreisser. In der Tabelle werden die wichtigsten Parameter der so identifizierten Faelle gezeigt sowie ob der Fall auch im Cook's Distance Test aufgefallen ist.

---

<sup>6</sup>Akaike information criterion Mass fuer die relative Guete eines Modells

<sup>7</sup>Faelle deren Werte in Abhaengigen oder Unabhaengigen Parametern auffaellig vom Rest der Daten abweichen

## 2 Resultate

### 2.1 Volumen Vorrat [m<sup>3</sup>/ha]

#### 2.1.1 Modell

Tabelle 1: Formel fuer Modell  
modell

---

TS\_Total\_Vol ~  
 ihst(chm\_mean\_sum\_of\_squares)  
 + chm\_mean\_sum\_of\_squares  
 + asins-  
 qrt(chm\_mean\_sum\_of\_squares\_Laub\_Anteil)

---

Tabelle 2: Koeffizienten Formel fuer Modell

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.0937	45.3338	0.02	0.9808
ihst(chm_mean_sum_of_squares)	36.1647	9.1269	3.96	0.0001
chm_mean_sum_of_squares	0.4392	0.0527	8.33	0.0000
asinsqrt(chm_mean_sum_of_squares_Laub_Anteil)	-89.3289	14.2742	-6.26	0.0000

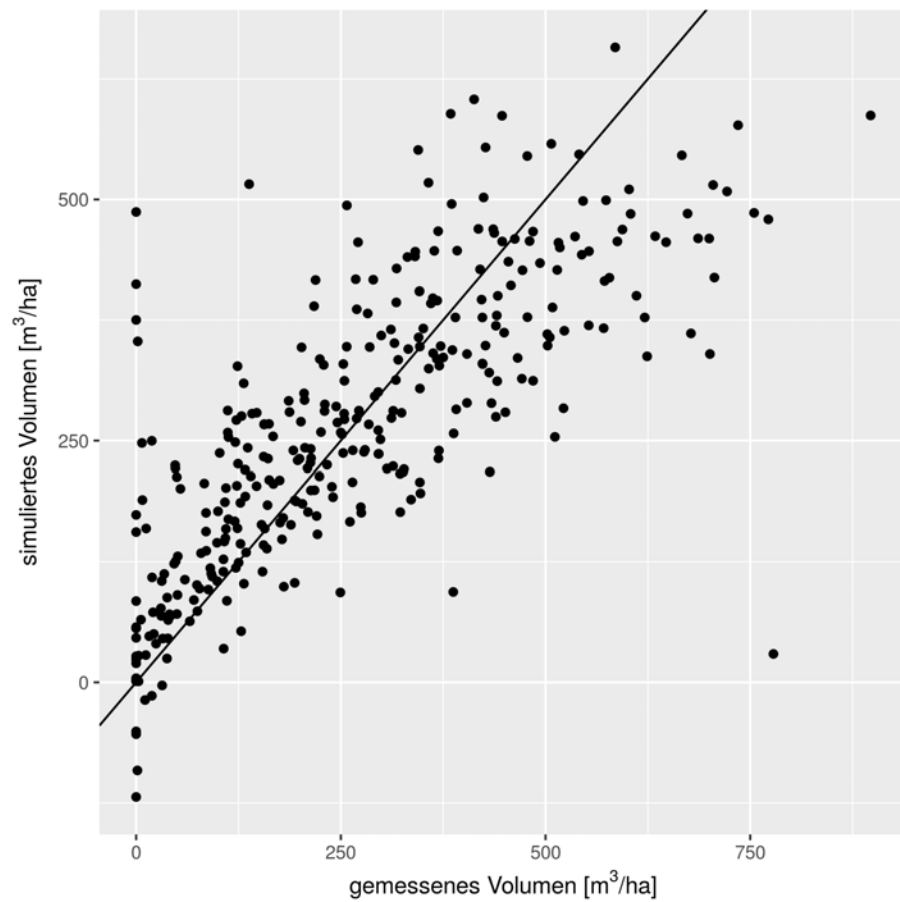


Abbildung 1: Modell fuer TS Volumen [m³/ha]. Gemessene Werte vs. simulierte Werte. korr.R-Quadrat: 0.59. Liegen die Punkte auf der diagonalen schwarzen Linie erklart das Modell 100% der Varianz. (Liegen die Punkte unterhalb der Linie unterschaezt das Modell, liegen die Punkte ueber der Linie liegt eine Ueberschaetzung vor)



### 2.1.2 Test mit unabhaengigen Daten

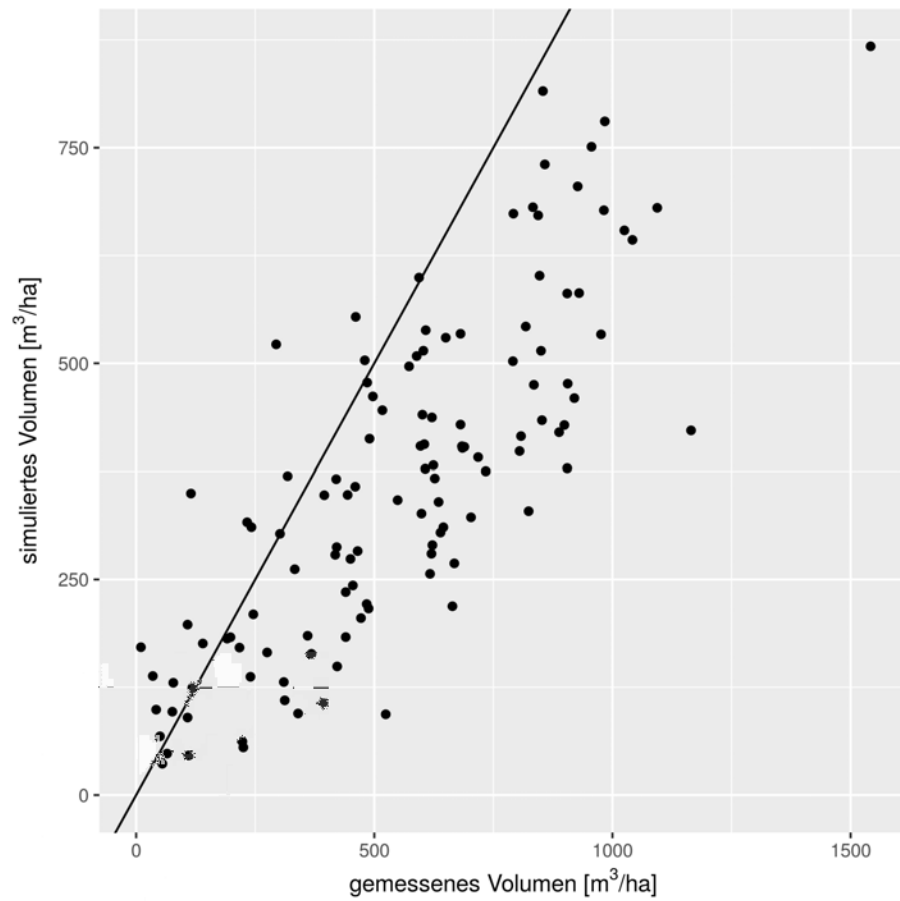


Abbildung 2: Unabhaengige Daten: Modell fuer TS Volumen [ $\text{m}^3/\text{ha}$ ] aus AG Daten angewendet auf FR Daten. Gemessene Werte vs. simulierte Werte.  $\text{corr.R-Quadrat}$ : 0.18. (Liegen die Punkte unterhalb der Linie unterschaezt das Modell, liegen die Punkte ueber der Linie liegt eine Ueberschaetzung vor)

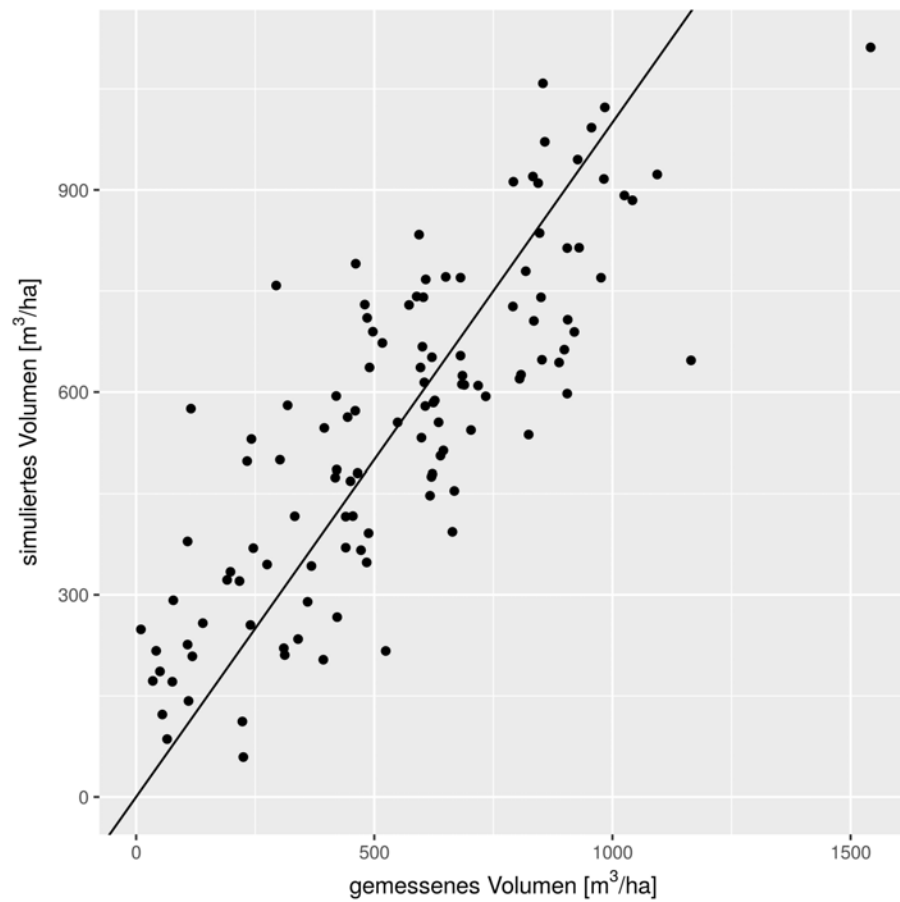


Abbildung 3: Unabh angige Daten: Modell fuer TS Volumen [m<sup>3</sup>/ha] aus FR Daten. Gemessene Werte vs. simulierte Werte. korr.R-Quadrat: 0.65. (Liegen die Punkte unterhalb der Linie untersch etzt das Modell, liegen die Punkte ueber der Linie liegt eine Uebersch etzung vor)

Tabelle 3: vergleich Koeffizienten fuer Modell

koeffs	wert_ag	wert_fr	przt_diff
(Intercept)	1.09	-186.85	17184.00
ihst(chm_mean_sum_of_squares)	36.16	98.58	-173.00
chm_mean_sum_of_squares	0.44	0.39	11.00
asinsqrt(chm_mean_sum_of_squares.Laub_Anteil)	-89.33	-87.57	2.00

### 2.1.3 Bootstrap

Tabelle 4: Bootstrap Modell

Parameter	Koeffizient	se	unteres_95pct_VI	oberes_95pct_VI
(Intercept)	1.09	32.71	-58.11	69.75
ihst(chm_mean_sum_of_squares)	36.16	7.22	21.55	50.23
chm_mean_sum_of_squares	0.44	0.06	0.33	0.55
asinsqrt(chm_mean_sum_of_squares_Laub_Anteil)	-89.33	13.23	-114.94	-62.25

### 2.1.4 Kreuzvalidierung

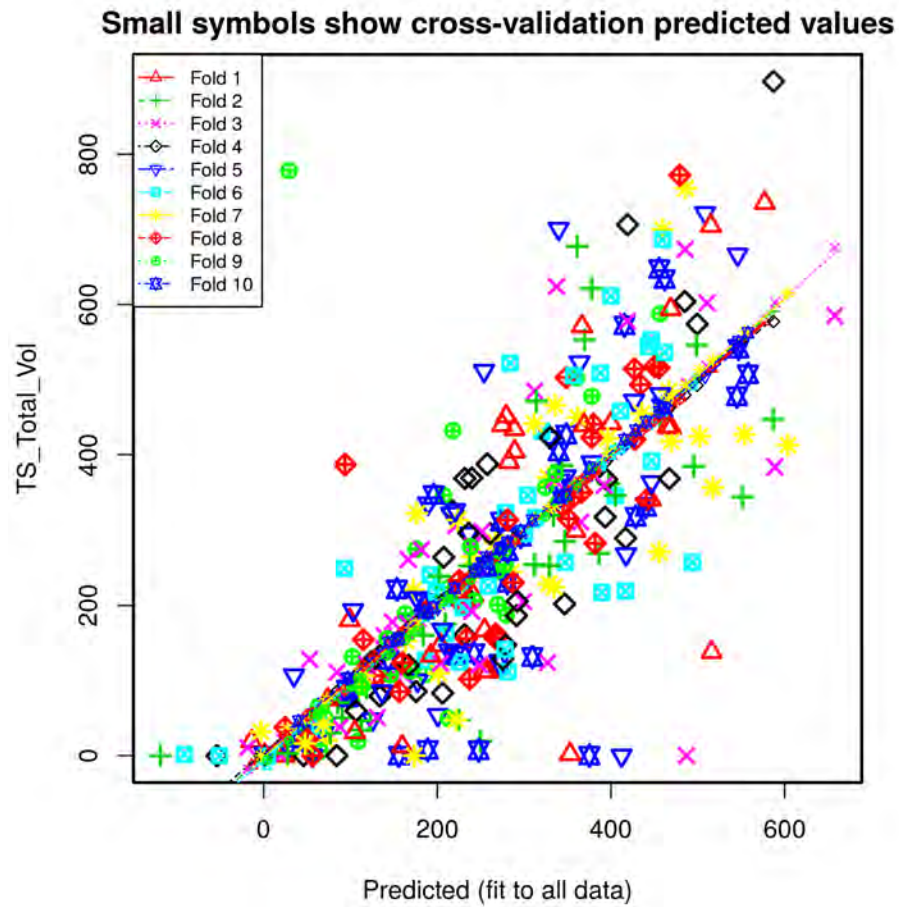


Abbildung 4: Ergebnis von 10 Kreuzvalidierungs Durchläufen (je 34 Faelle im Testset). Das Kreuzvalidierte  $R^2$  betraegt 0.562. Das Kreuzvalidierte RMSE betraegt 126[m<sup>3</sup>/ha].

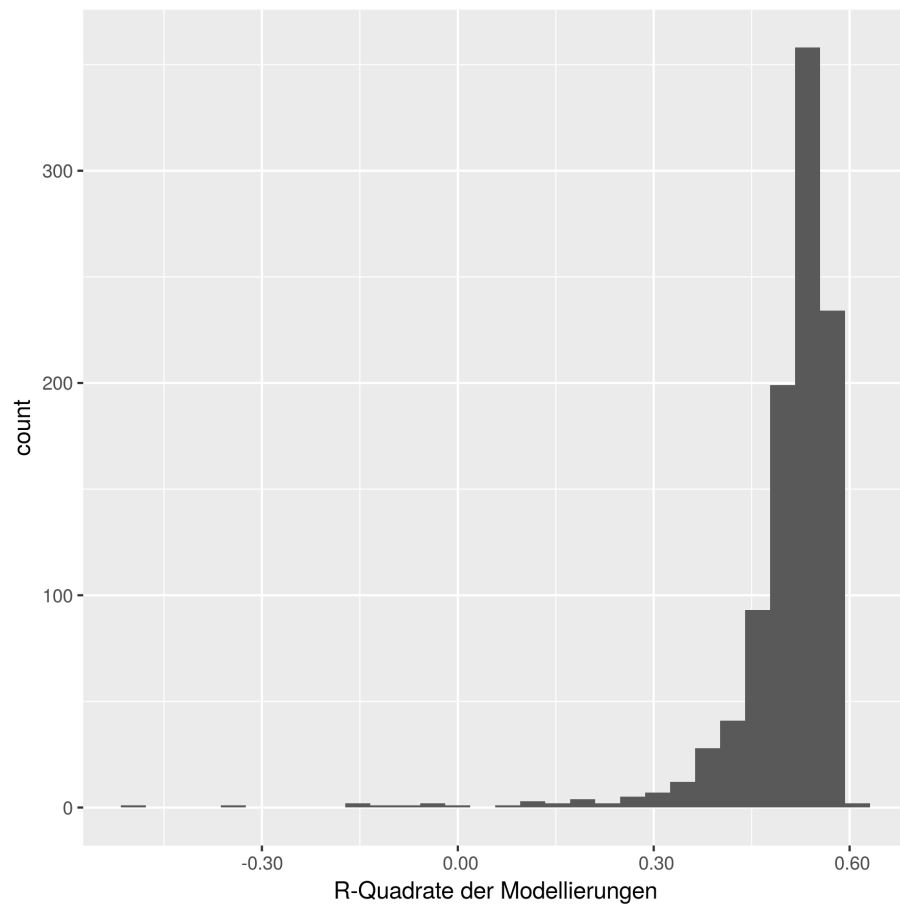


Abbildung 5: Ergebnis von 1000 Kreuzvalidierungen-Wiederholungen. Das Korrigiert R-Quadrat wurde jeweils aus den Simulationen mit den 90% der nicht fuer die Modellschaetzung verwendeten Daten berechnet.

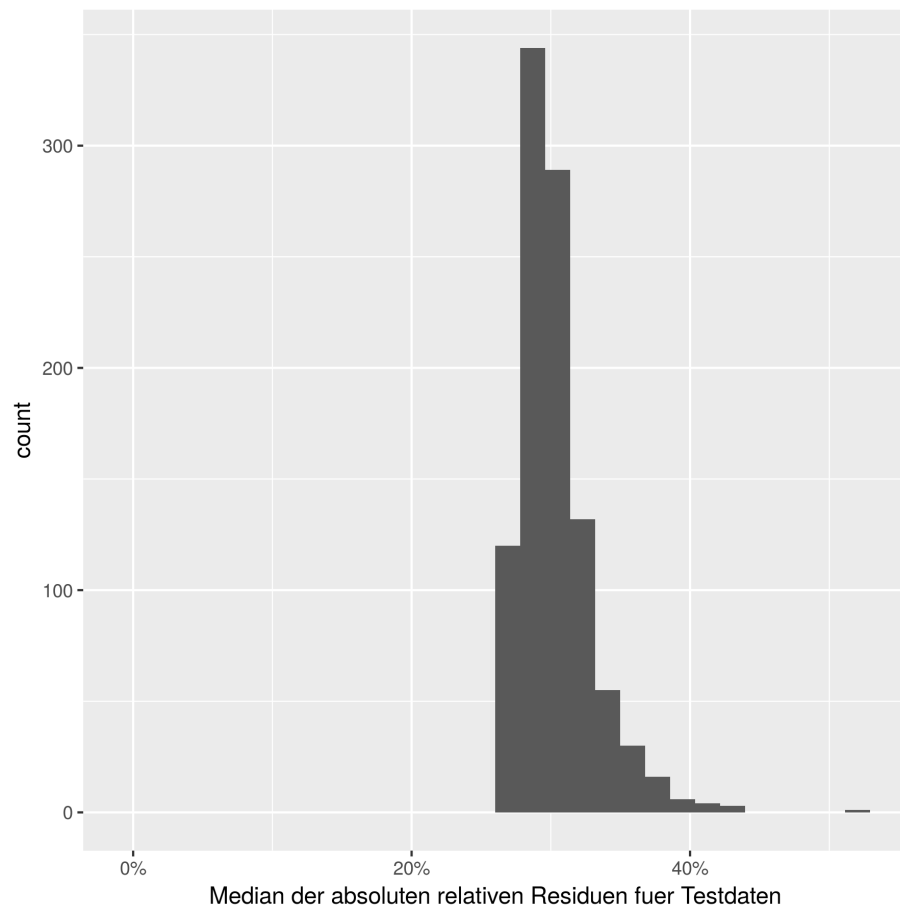


Abbildung 6: Ergebnis von 1000 Kreuzvalidierungen-Wiederholungen. Die absoluten relativen Residuen wurde jeweils aus den Simulationen mit den 90% der nicht fuer die Modellschaetzung verwendeten Daten berechnet. Gezeigt ist die Verteilung der Mediane der relativen Residuen.

### 2.1.5 Residuenanalyse

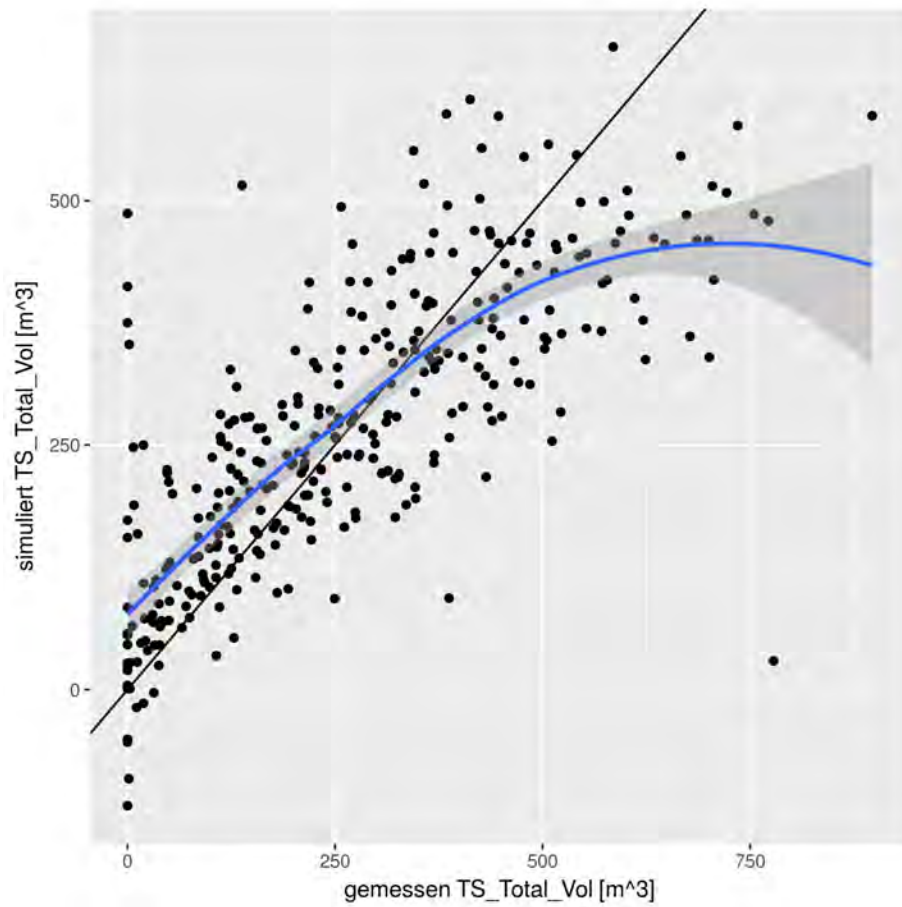


Abbildung 7: Modellannahme 1: Linearer Zusammenhang gemessene Werte und simulierte Werte



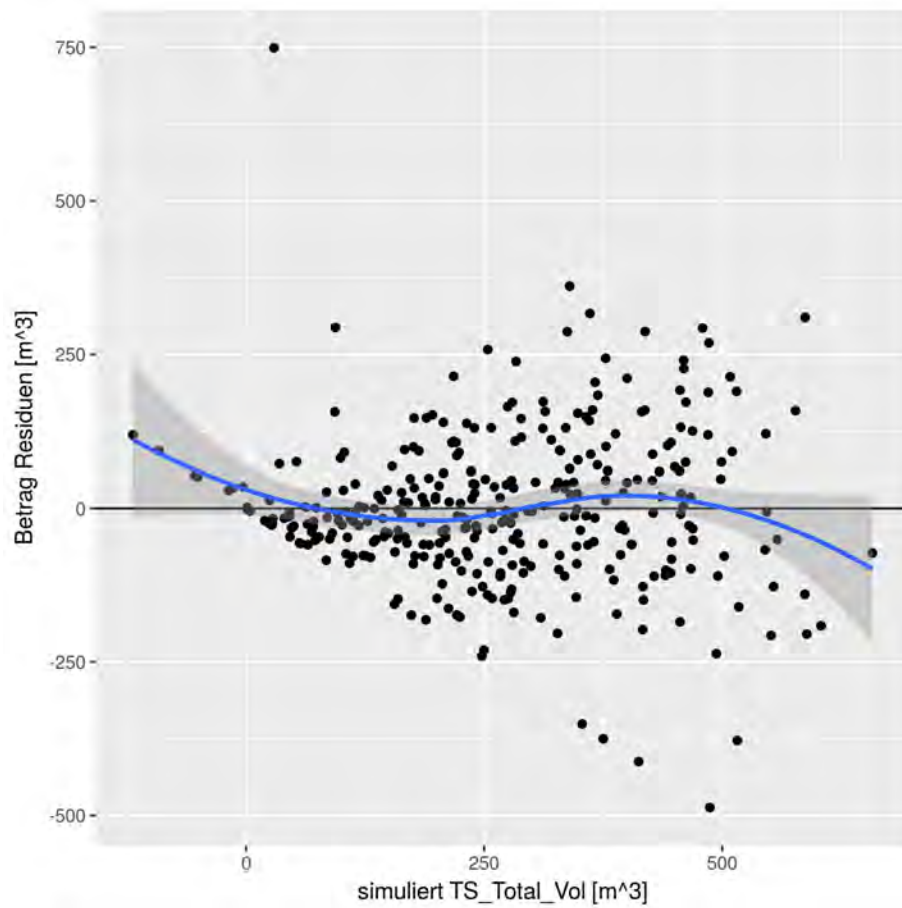


Abbildung 8: Modellannahme 2: Tukey-Anscombe Plot. Die Residuen sind unabhängig von den simulierten Werten.

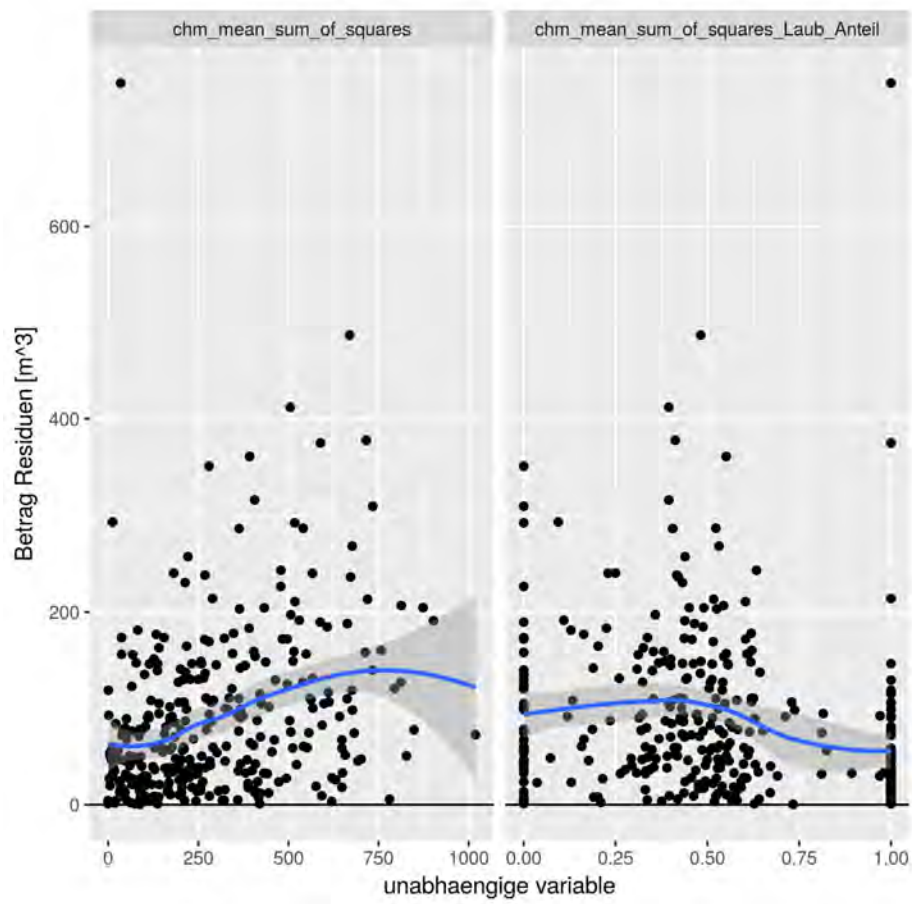


Abbildung 9: Modellannahme 2: Die Residuen sind unabhängig von den den unabhängigen Werten.

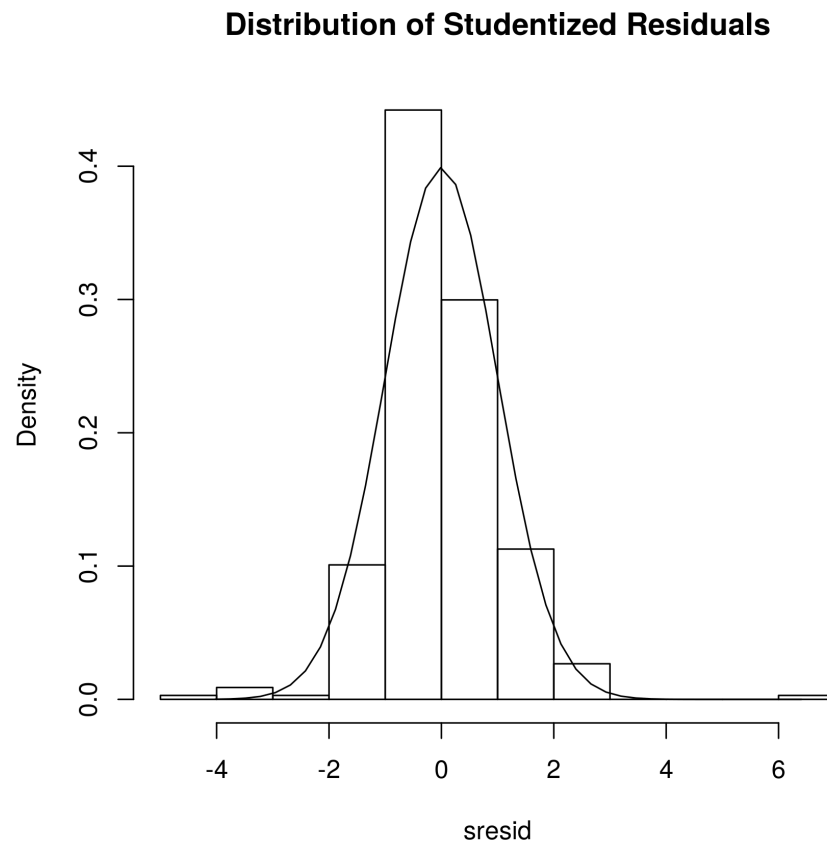


Abbildung 10: Modellannahme 4: Die Residuen sind normalverteilt um 0.

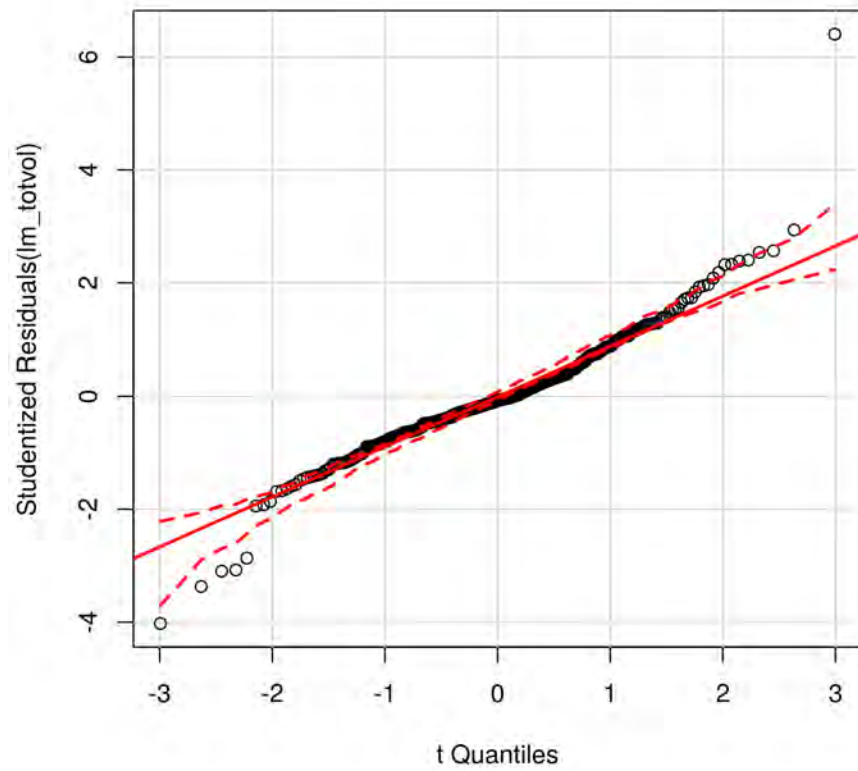


Abbildung 11: Modellannahme 4: Die Residuen sind normalverteilt.

### 2.1.6 Alternative Modelle

Tabelle 5: Beschreibung diverse Modelle fuer Total Vol. Das hoechste korr.R-Quadrat bzw. der tiefste AIK Wert sind rot markiert.

Modell	Formel	korr.RQuadrat	AIC
predict_dlm1	TS_Total_Vol ~ chm_mean_sum_of_squares * asins- qrt(chm_mean_sum_of_squares_Laub_Anteil)	0.58	4217.02
predict_dlm2	TS_Total_Vol ~ chm_mean_sum_of_squares + asins- qrt(chm_mean_sum_of_squares_Laub_Anteil)	0.57	4227.95
predict_dlm3	TS_Total_Vol ~ I(chm_mean_sum_of_squares_Laub^2) + I(chm_mean_sum_of_squares_Nadel^2) + chm_mean_sum_of_squares_Laub + chm_mean_sum_of_squares_Nadel	0.44	4319.36
predict_dlm4	TS_Total_Vol ~ ihst(chm_mean_sum_of_squares_Laub) + ihst(chm_mean_sum_of_squares_Nadel) + chm_mean_sum_of_squares_Laub + chm_mean_sum_of_squares_Nadel	0.44	4317.4
predict_dlm5	TS_Total_Vol ~ ihst(chm_mean_sum_of_squares) + chm_mean_sum_of_squares + asins- qrt(chm_mean_sum_of_squares_Laub_Anteil)	0.59	4214.42
predict_dlm6	TS_Total_Vol ~ ihst(chm_c15_Laub) + ihst(chm_NumberOfValidValues_Nadel) + chm_c15_Laub + chm_NumberOfValidValues_Nadel	0.46	4303.58
predict_dlm7	TS_Total_Vol ~ ihst(chm_mean_sum_of_squares_Nadel) + chm_mean_sum_of_squares_Nadel + chm_mean_sum_of_squares_Laub	0.4	4338.09

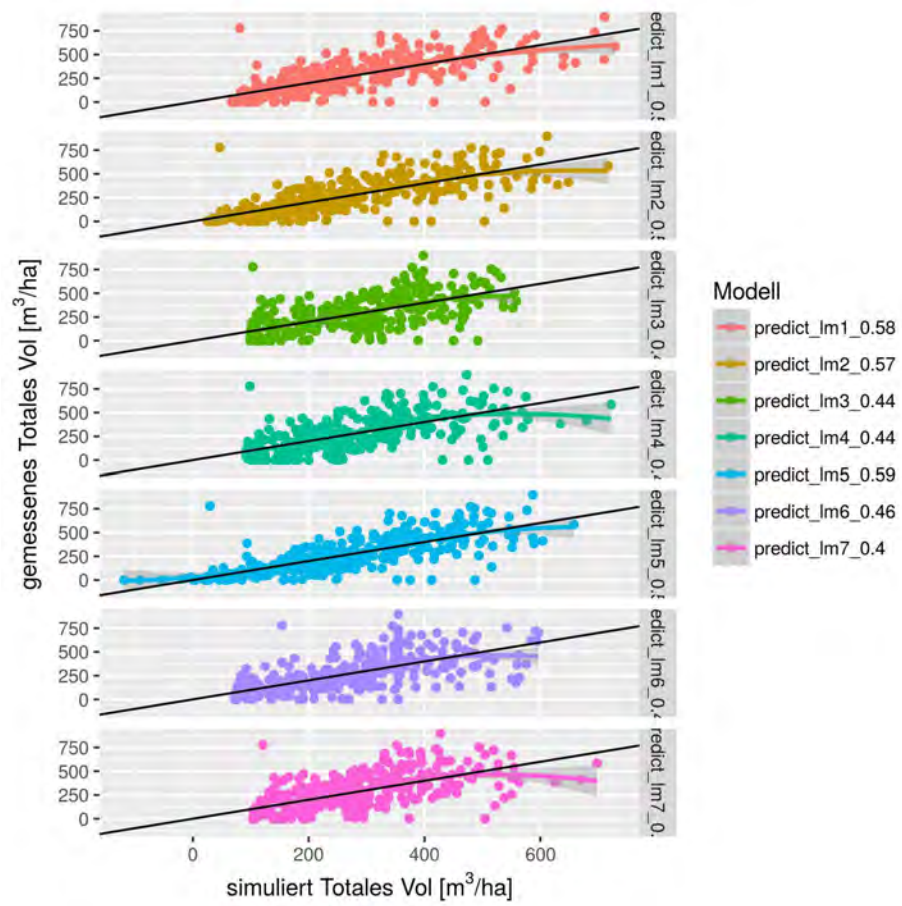


Abbildung 12: Vorhersagegenauigkeit fuer unterschiedliche Modelle (siehe Tabelle 5 fuer Beschreibung der Modelle). Die Vorhersagen von lm\_5 sind aus dem diskutierten Modell erstellt.

### 2.1.7 Ausreisser

NULL

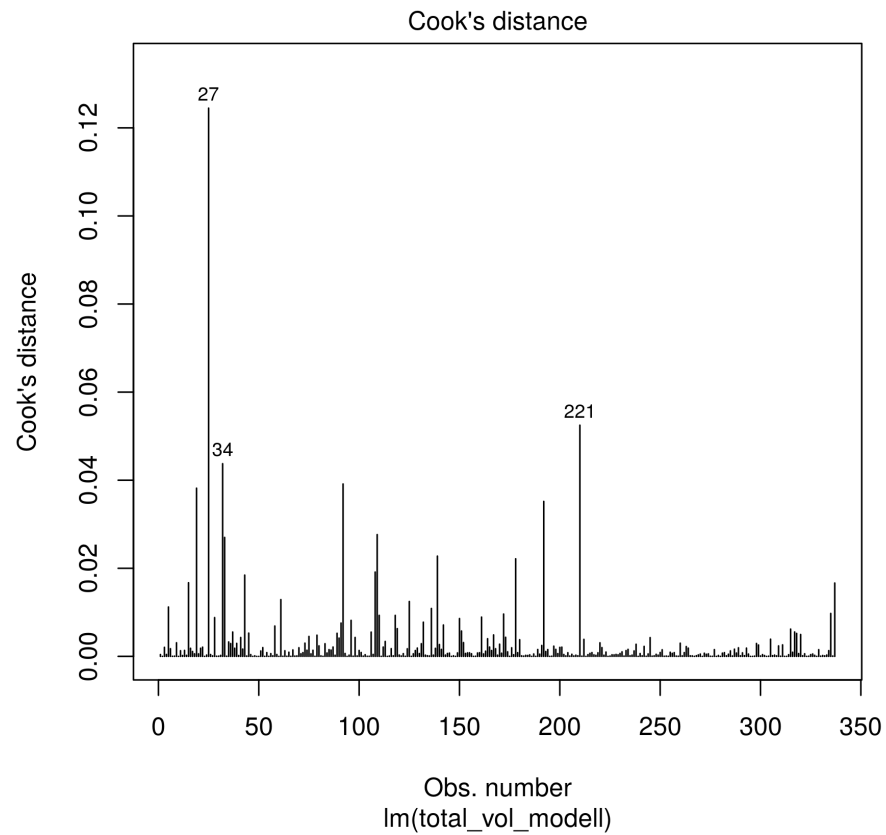


Abbildung 13: Ausreisser der Modellierung: Cook's D Plot. Cooks Distance bewertet den Einfluss einer Beobachtung auf das Ergebnis der Koeffizientenschätzung.

Tabelle 6: Die durch Bonferroni Outlier Test identifizierten Ausreisser

Zeilen Num- mer	Plot ID	TS BHDN mean	TS To- tal Vol	TS LbH Vol	TS NdH Vol	chm mean sum of squares Laub	chm mean sum of squares Nadel	CooksD
27	236662	33.95	778.32	754.83	23.50	137	0	TRUE
233	249662	34.86	230.45	0.00	230.45	0	172	FALSE

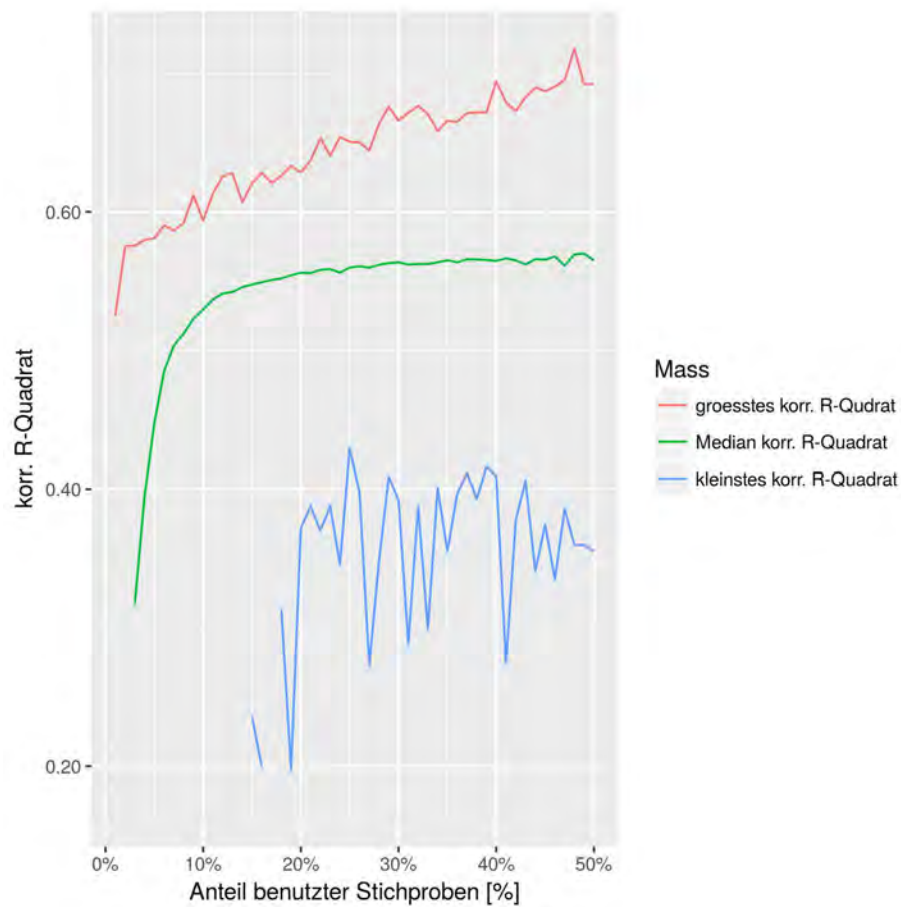


Abbildung 14: Parameter TS\_TotalVol [ $\text{m}^3/\text{ha}$ ]. Entwicklung der korr. R-Quadrat in Abhängigkeit vom Anteil der zur Parameter Schätzung benutzten Stichprobenflächen. Bei  $x=1$  wurde 1% der Stichproben zur Schätzung der Koeffizienten benutzt, bei  $x=2$  2% etc. Gezeigt sind Max, Min und Median aus jeweils 1000 Wiederholungen. Werte unter 0.001 werden nicht gezeigt.



## 2.2 Basalflaeche [m<sup>2</sup>/ha]

### 2.2.1 Modell

Tabelle 7: Formel fuer Modell  
modell

TS\_BasalAreaha ~  
I(chm\_mean\_sum\_of\_squares^2)  
+ chm\_mean\_sum\_of\_squares  
+ asins-  
qrt(chm\_mean\_sum\_of\_squares\_Laub\_Anteil)

Tabelle 8: Koeffizienten Formel fuer Modell

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	103.0640	16.2326	6.35	0.0000
I(chm_mean_sum_of_squares^2)	-0.0006	0.0001	-5.96	0.0000
chm_mean_sum_of_squares	0.8272	0.0765	10.82	0.0000
asinsqrt(chm_mean_sum_of_squares_Laub_Anteil)	-63.8929	10.8101	-5.91	0.0000

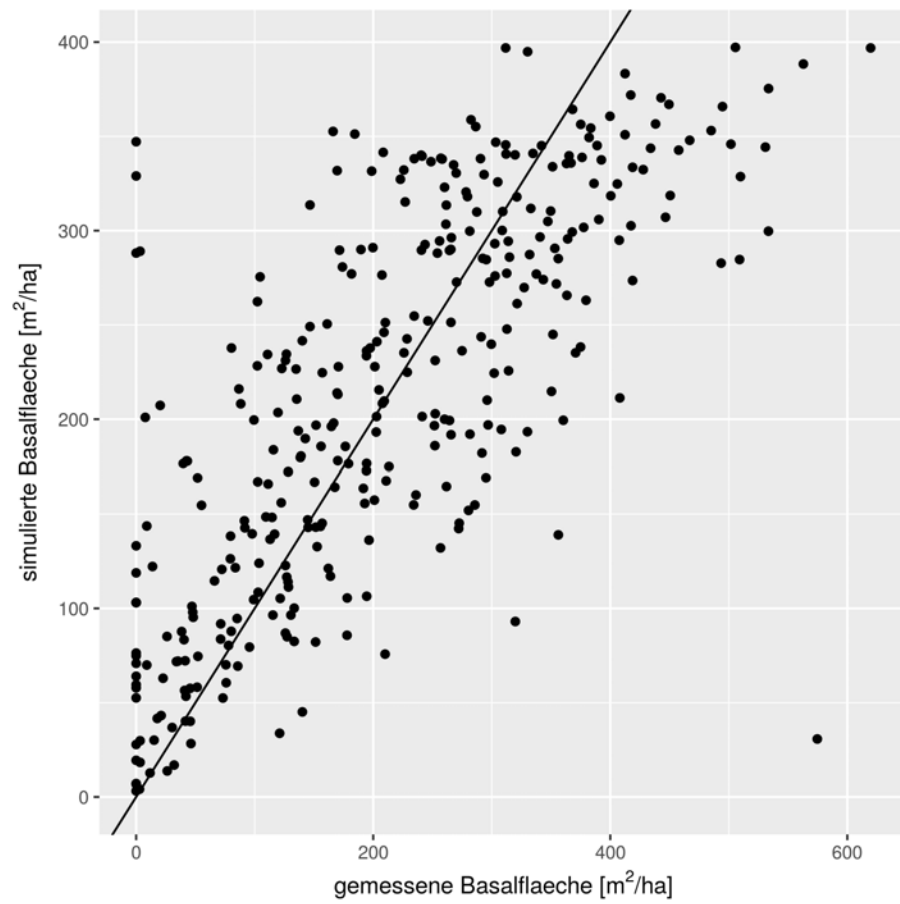


Abbildung 15: Modell fuer Basalflaeche [ $\text{m}^2/\text{ha}$ ]. Gemessene Werte vs. simulierte Werte.  $\text{corr.R-Quadrat}$ : 0.55. (Liegen die Punkte unterhalb der Linie unterschaezt das Modell, liegen die Punkte ueber der Linie liegt eine Ueberschaetzung vor)

### 2.2.2 Test mit unabhängigen Daten

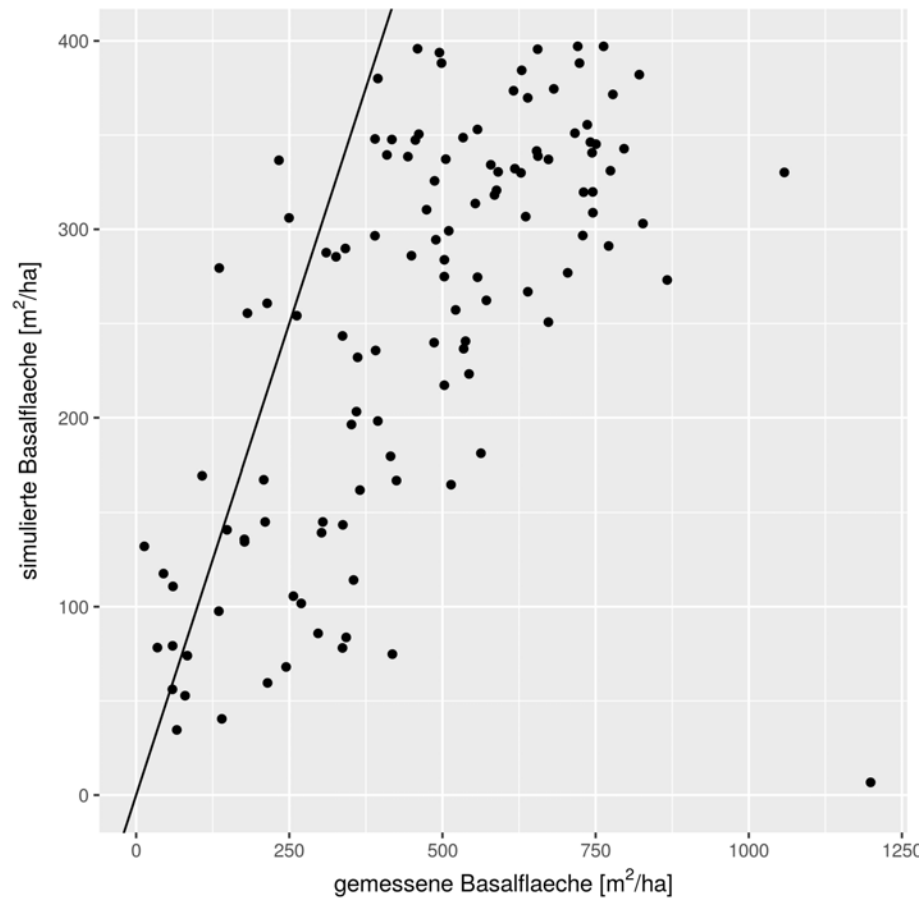


Abbildung 16: Unabhängige Daten: Modell fuer Basalflaeche [m<sup>2</sup>/ha] aus AG Daten angewendet auf FR Daten. Gemessene Werte vs. simulierte Werte.  $\text{corr.R-Quadrat}$ : -0.53. (Liegen die Punkte unterhalb der Linie unterschätzt das Modell, liegen die Punkte ueber der Linie liegt eine Ueberschaetzung vor)

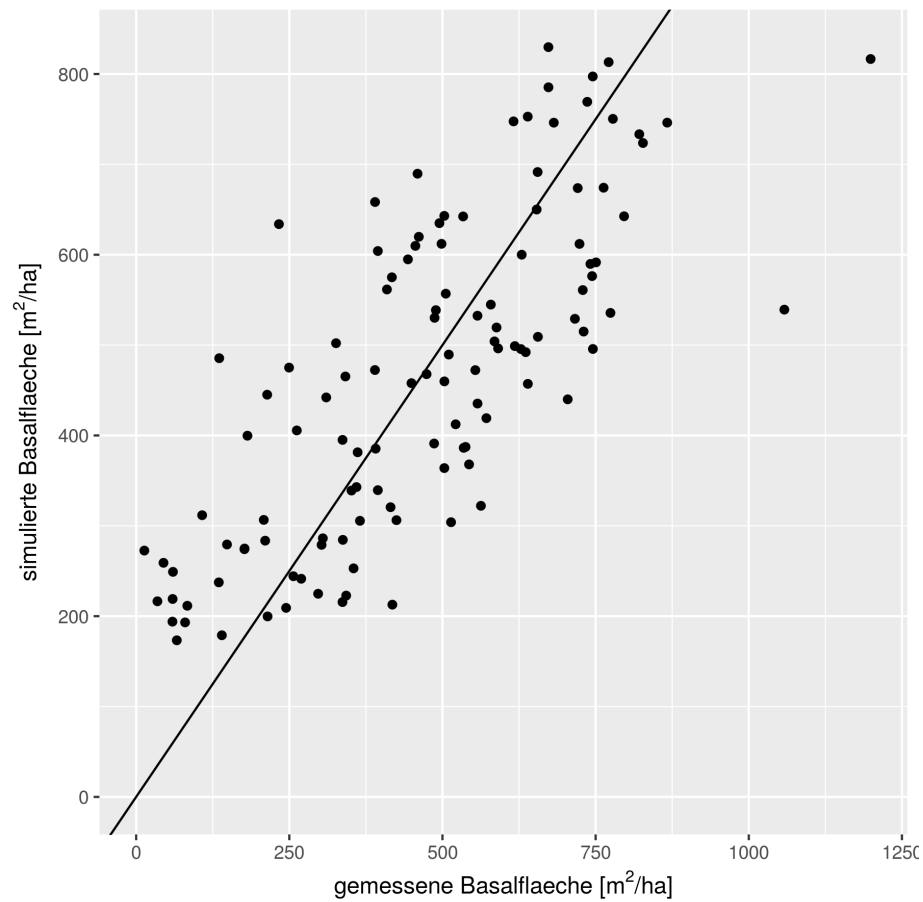


Abbildung 17: Unabh angige Daten: Modell fuer TS BasalAreaha [m<sup>2</sup>/ha] aus FR Daten. Gemessene Werte vs. simulierte Werte. korr.R-Quadrat: 0.56. (Liegen die Punkte unterhalb der Linie untersch etzt das Modell, liegen die Punkte ueber der Linie liegt eine Uebersch etzung vor)

Tabelle 9: vergleich Koeffizienten fuer Modell

koeffs	wert_ag	wert_fr	przt_diff
(Intercept)	103.06	245.67	-138.00
I(chm_mean_sum_of_squares^2)	-0.00	-0.00	61.00
chm_mean_sum_of_squares	0.83	0.76	8.00
asinsqrt(chm_mean_sum_of_squares.Laub_Anteil)	-63.89	-64.97	-2.00

### 2.2.3 Bootstrap

Tabelle 10: Bootstrap Modell

Parameter	Koeffizient	se	unteres_95pct_VI	oberes_95pct_VI
(Intercept)	103.06	16.55	72.94	137.23
I(chm_mean_sum_of_squares^2)	-0.00	0.00	-0.00	-0.00
chm_mean_sum_of_squares	0.83	0.08	0.67	0.97
asinsqrt(chm_mean_sum_of_squares_Laub_Anteil)	-63.89	10.58	-84.89	-41.88

## 2.2.4 Kreuzvalidierung

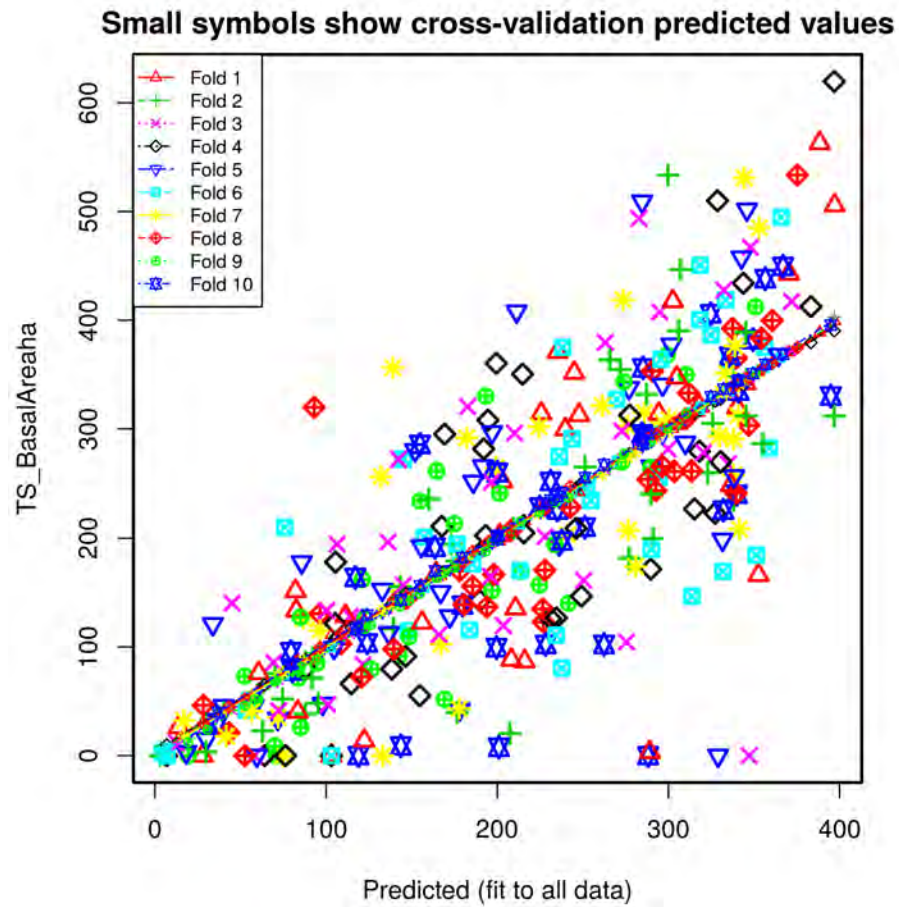


Abbildung 18: Ergebnis von 10 Kreuzvalidierungs Durchläufen (je 34 Fälle im Testset). Das Kreuzvalidierte  $R^2$  beträgt 0.492. Das Kreuzvalidierte RMSE beträgt 143[m<sup>2</sup>/ha].

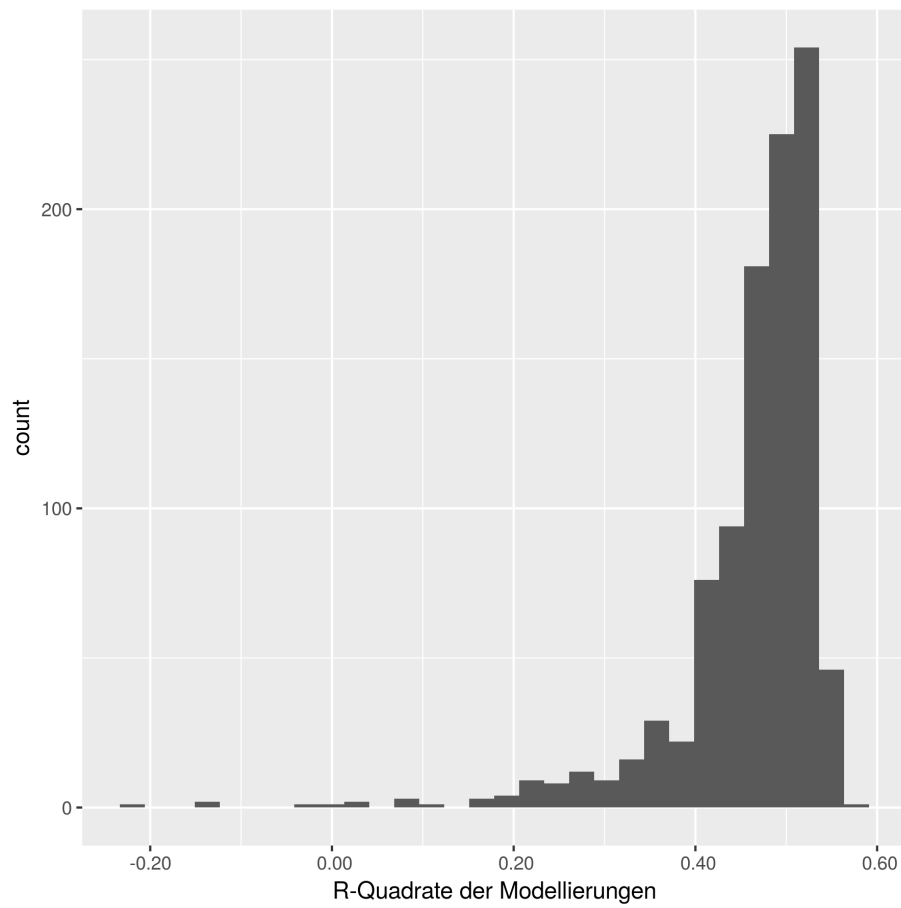


Abbildung 19: Ergebnis von 1000 Kreuzvalidierungen-Wiederholungen. Das Korrigiert R-Quadrat wurde jeweils aus den Simulationen mit den 90% der nicht fuer die Modellschaetzung verwendeten Daten berechnet.



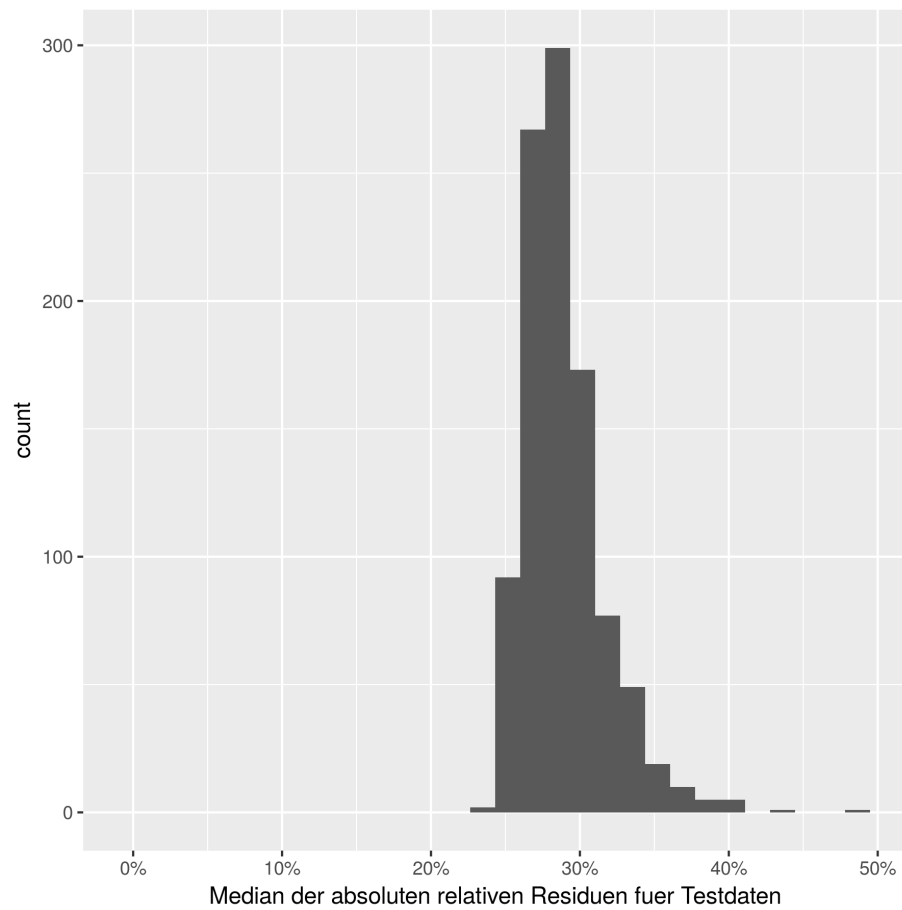


Abbildung 20: Ergebnis von 1000 Kreuzvalidierungen-Wiederholungen. Die absoluten relativen Residuen wurde jeweils aus den Simulationen mit den 90% der nicht fuer die Modellschaetzung verwendeten Daten berechnet. Gezeigt ist die Verteilung der Mediane der relativen Residuen.

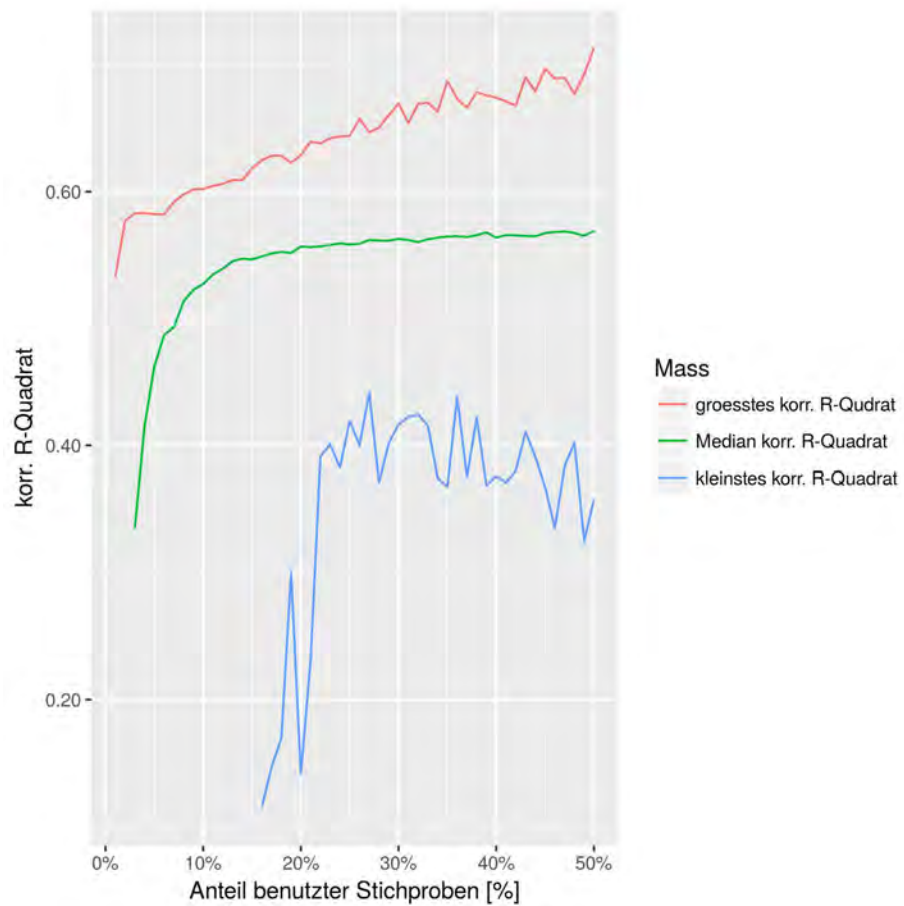


Abbildung 21: Parameter TS\_Basalarea\_ha [ $\text{m}^2/\text{ha}$ ]. Entwicklung der korr. R-Quadrat in Abhängigkeit vom Anteil der zur Parameter Schätzung benutzten Stichprobenflächen. Bei  $x=1$  wurde 1% der Stichproben zur Schätzung der Koeffizienten benutzt, bei  $x=2$  2% etc. Gezeigt sind Max, Min und Median aus jeweils 1000 Wiederholungen. Werte unter 0.001 werden nicht gezeigt

### 2.2.5 Residuenanalyse

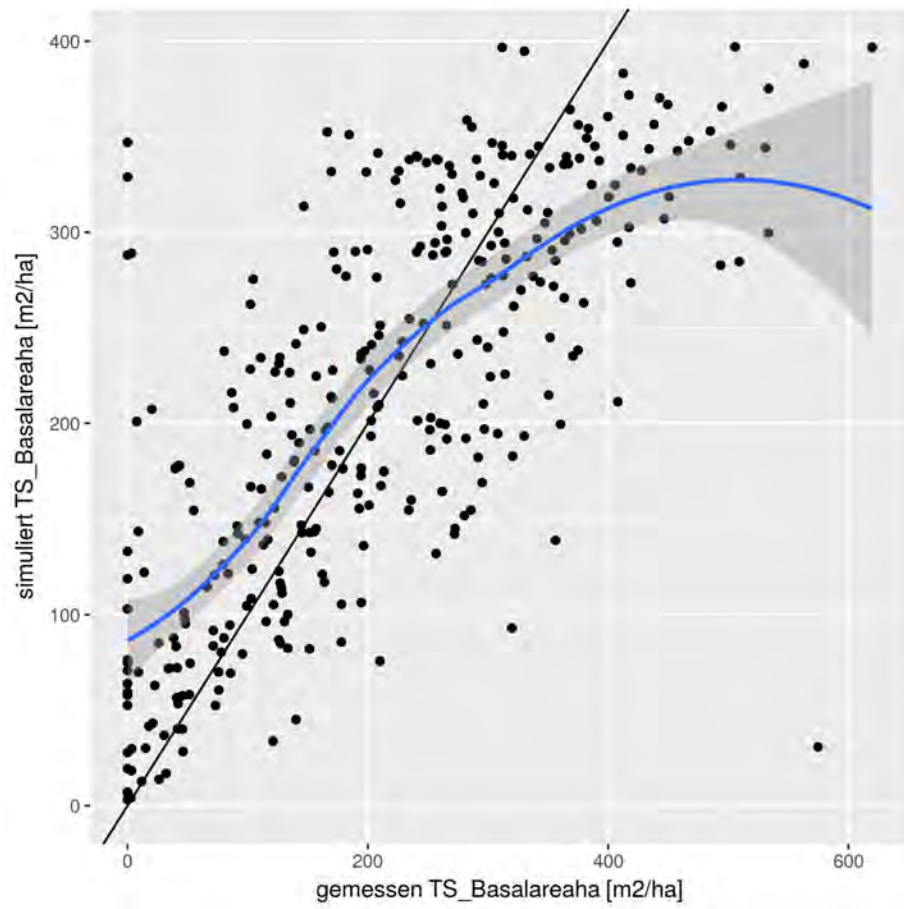


Abbildung 22: Modellannahme 1: Linearer Zusammenhang gemessene Werte und simulierte Werte

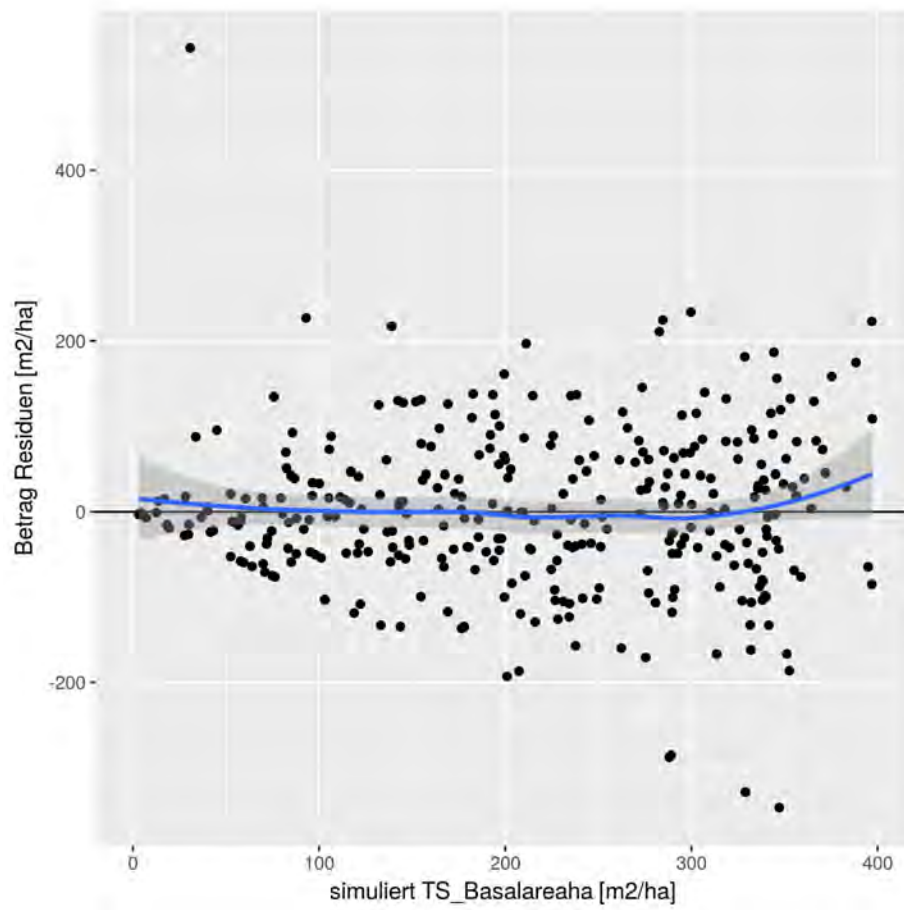


Abbildung 23: Modellannahme 2: Tukey-Anscombe Plot. Die Residuen sind unabhangig von den den simulierten Werten.

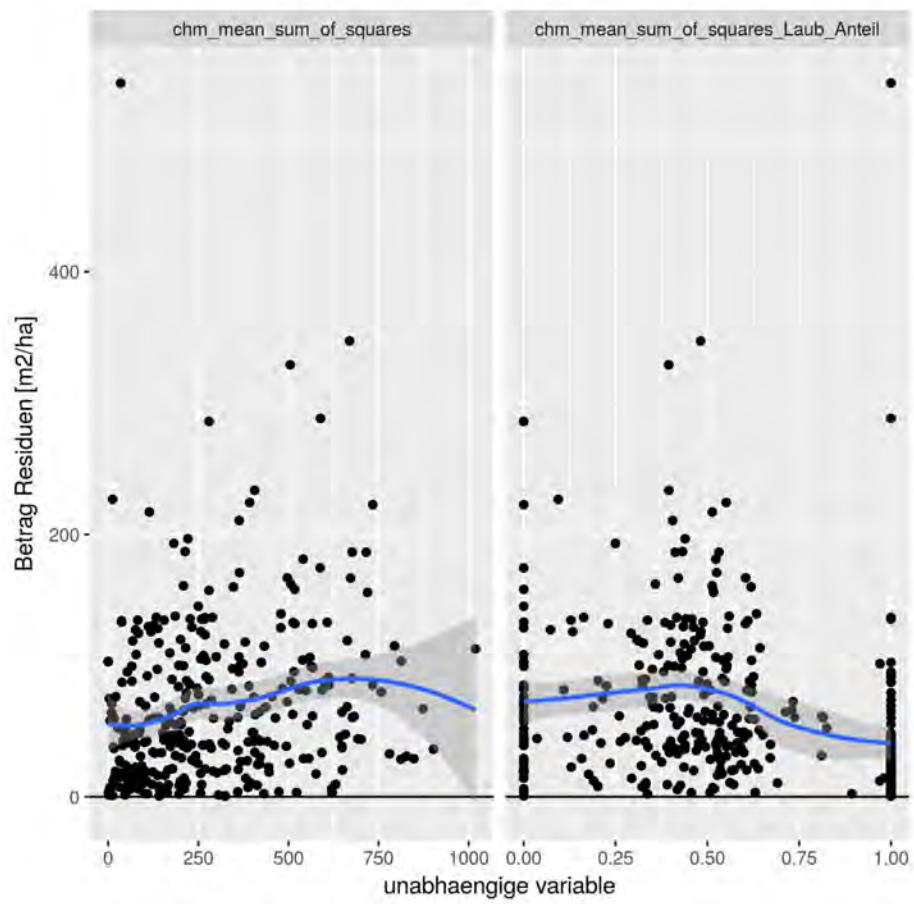


Abbildung 24: Modellannahme 2: Die Residuen sind unabhaengig von den den unabhaengigen Werten.

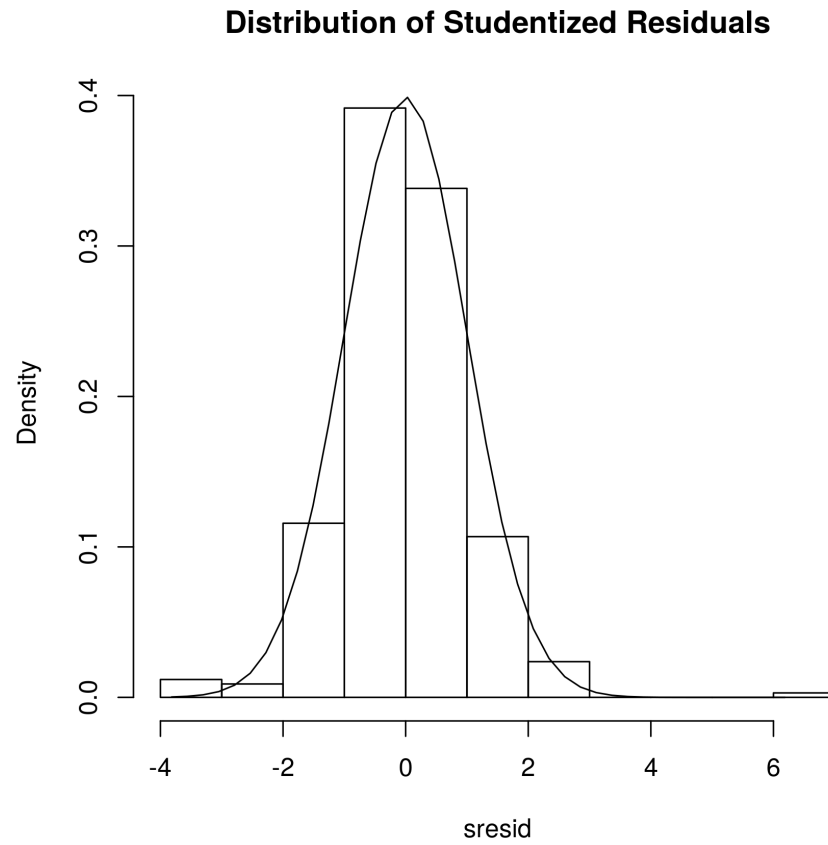


Abbildung 25: Modellannahme 4: Die Residuen sind normalverteilt um 0.

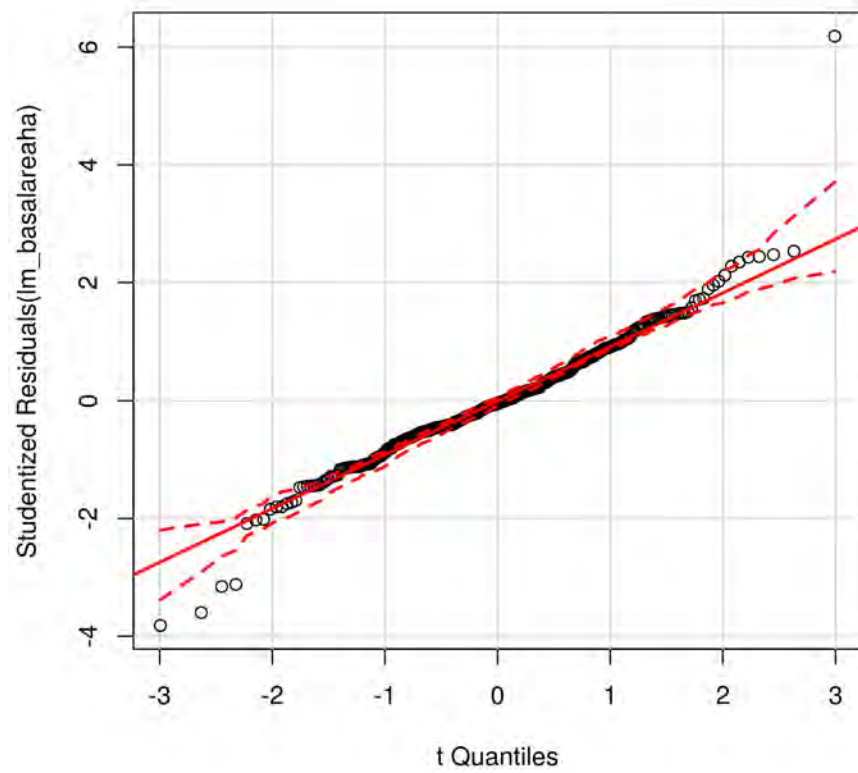


Abbildung 26: Modellannahme 4: Die Residuen sind normalverteilt.

## 2.2.6 Alternative Modelle

Tabelle 11: Beschreibung diverse Modelle fuer TS BasalAreaha. Das hoechste korr.R-Quadrat bzw. der tiefste AIK Wert sind rot markiert.

Modell	Formel	korr.RQuadrat	AIC
predict_dlm1	TS_BasalAreaha ~ chm_mean_sum_of_squares * asins- qrt(chm_mean_sum_of_squares_Laub_Anteil)	0.51	4047.63
predict_dlm2	TS_BasalAreaha ~ chm_mean_sum_of_squares + asins- qrt(chm_mean_sum_of_squares_Laub_Anteil)	0.5	4052.39
predict_dlm3	TS_BasalAreaha ~ I(chm_mean_sum_of_squares_Laub^2) + I(chm_mean_sum_of_squares_Nadel^2) + chm_mean_sum_of_squares_Laub + chm_mean_sum_of_squares_Nadel	0.38	4126.4
predict_dlm4	TS_BasalAreaha ~ ihst(chm_mean_sum_of_squares_Laub) + ihst(chm_mean_sum_of_squares_Nadel) + chm_mean_sum_of_squares_Laub + chm_mean_sum_of_squares_Nadel	0.39	4124.29
predict_dlm5	TS_BasalAreaha ~ I(chm_mean_sum_of_squares^2) + chm_mean_sum_of_squares + asins- qrt(chm_mean_sum_of_squares_Laub_Anteil)	0.55	4020.21
predict_dlm6	TS_BasalAreaha ~ ihst(chm_c15_Laub) + ihst(chm_NumberOfValidValues_Nadel) + chm_c15_Laub + chm_NumberOfValidValues_Nadel	0.47	4074.19
predict_dlm7	TS_BasalAreaha ~ ihst(chm_mean_sum_of_squares) + chm_mean_sum_of_squares + asins- qrt(chm_mean_sum_of_squares_Laub_Anteil)	0.54	4025.77



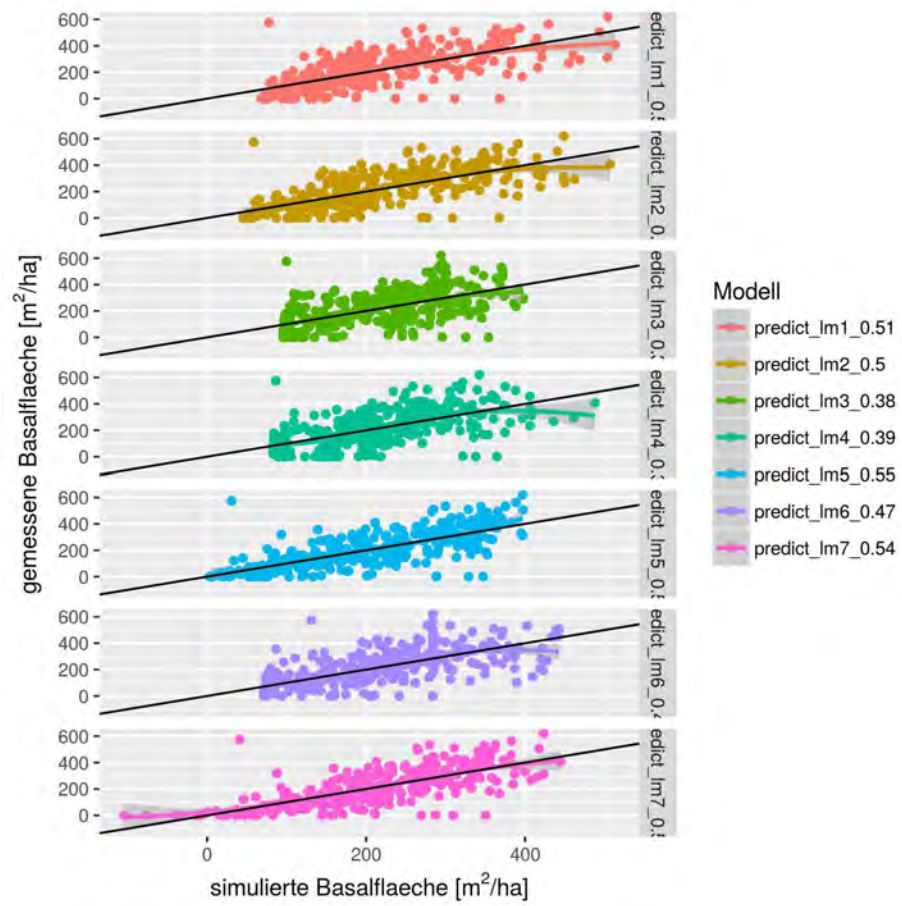


Abbildung 27: Vorhersagegenauigkeit fuer unterschiedliche Modelle (siehe Tabelle 11 fuer Beschreibung der Modelle). Die Vorhersagen von lm\_5 sind aus dem diskutierten Modell erstellt.

### 2.2.7 Ausreisser

NULL

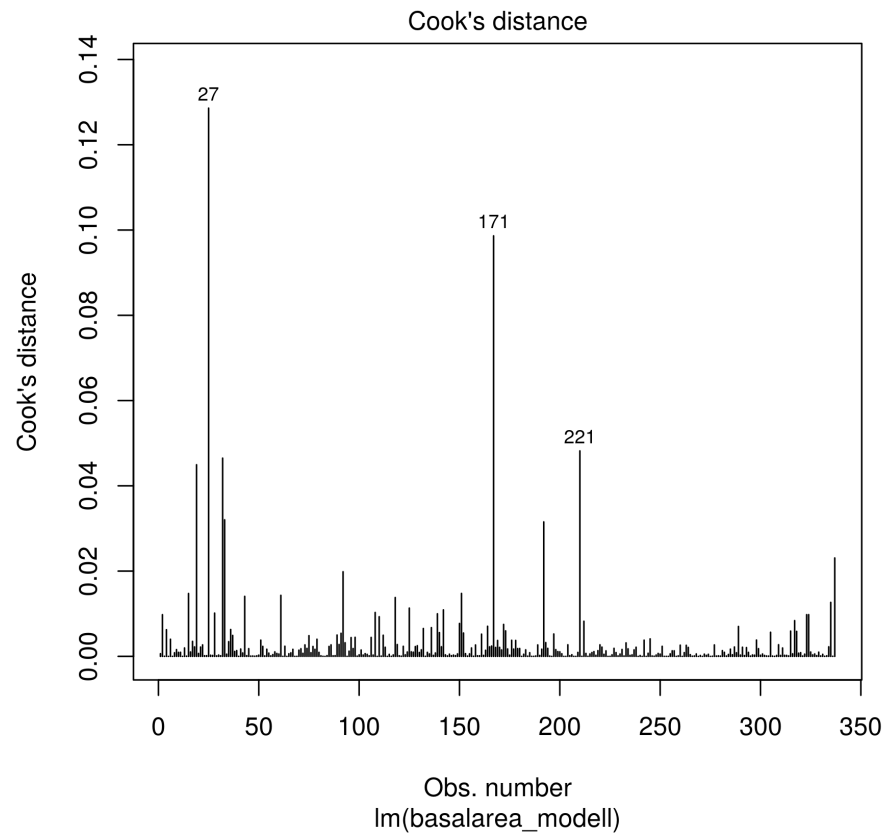


Abbildung 28: Ausreisser der Modellierung: Cook's D Plot. Cooks Distance bewertet den Einfluss einer Beobachtung auf das Ergebnis der Koeffizientenschätzung.

Tabelle 12: Die durch Bonferroni Outlier Test identifizierten Ausreisser								
Zeilen	Plot ID	TS	TS To-	TS	TS	chm	chm	CooksD
Num-		BHDN	tal Vol	LbH	NdH	mean	mean	
mer		mean		Vol	Vol	sum of	sum of	
						squares	squares	
						Laub	Nadel	
27	236662	33.95	778.32	754.83	23.50	137	0	TRUE

## 2.3 Stammzahl [ha<sup>-1</sup>]

### 2.3.1 Modell

Tabelle 13: Formel fuer Modell  
modell

$$\begin{aligned} & \text{TS\_Stammzahl} \\ & \sim \\ & \text{I(chm\_stddev}^2) \\ & + \text{chm\_stddev} + \\ & \text{I(chm\_mean\_sum\_of\_squares}^2) \\ & + \text{chm\_mean\_sum\_of\_squares} \\ & + \text{asins-} \\ & \text{qrt(chm\_mean\_sum\_of\_squares\_Laub\_Anteil)} \end{aligned}$$

Tabelle 14: Koeffizienten Formel fuer Modell

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	181.1906	21.8047	8.31	0.0000
I(chm_stddev^2)	-0.6591	0.3839	-1.72	0.0869
chm_stddev	-10.7109	6.2215	-1.72	0.0861
I(chm_mean_sum_of_squares^2)	-0.0009	0.0001	-8.65	0.0000
chm_mean_sum_of_squares	0.7888	0.0888	8.88	0.0000
asinsqrt(chm_mean_sum_of_squares_Laub_Anteil)	-44.8063	10.0335	-4.47	0.0000

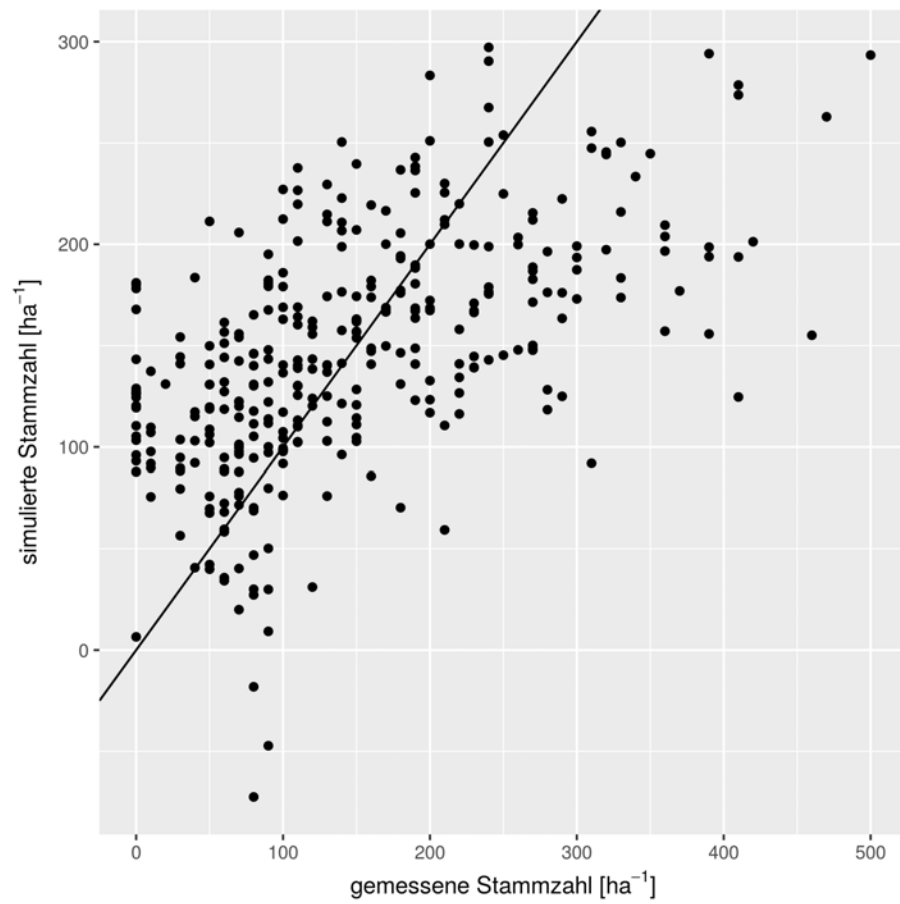


Abbildung 29: Modell fuer TS Stammzahl [ $\text{ha}^{-1}$ ]. Gemessene Werte vs. simulierte Werte. korr.R-Quadrat: 0.32. (Liegen die Punkte unterhalb der Linie unterschätzt das Modell, liegen die Punkte ueber der Linie liegt eine Ueberschaetzung vor)

### 2.3.2 Test mit unabhaengigen Daten

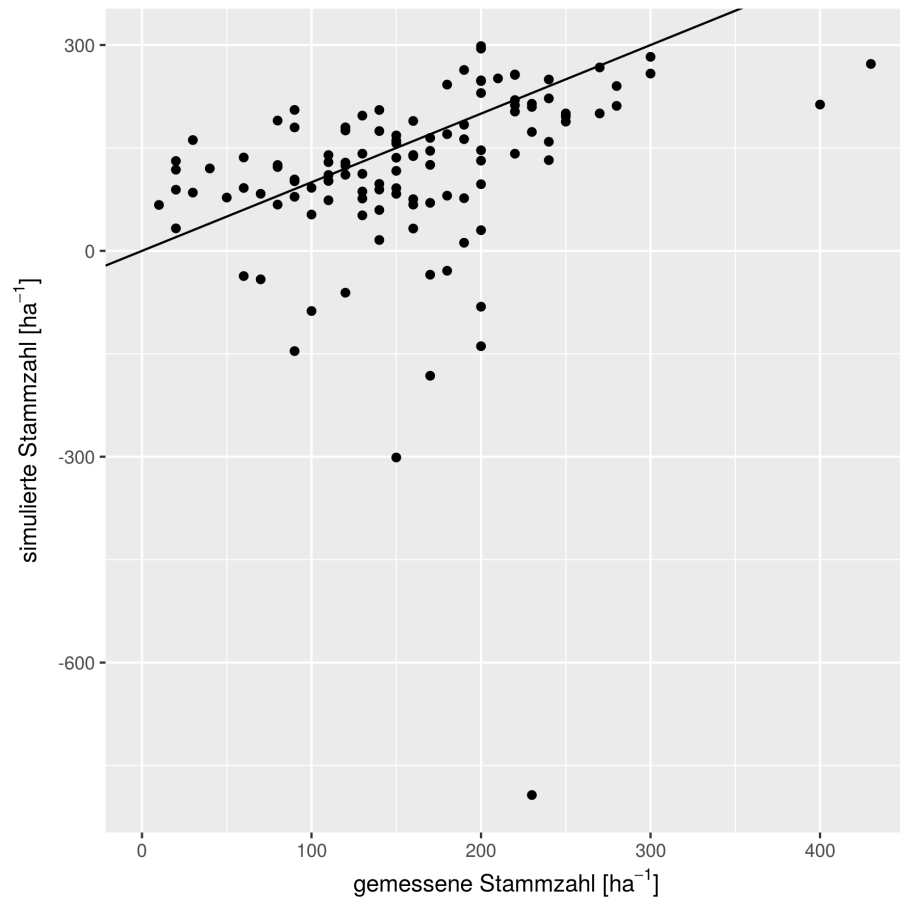


Abbildung 30: Unabhaengige Daten: Modell fuer TS Stammzahl [ha<sup>-1</sup>] aus AG Daten angewendet auf FR Daten. Gemessene Werte vs. simulierte Werte.  $\text{corr.R-Quadrat}$ : -2.6. (Liegen die Punkte unterhalb der Linie unterschaezt das Modell, liegen die Punkte ueber der Linie liegt eine Ueberschaetzung vor)

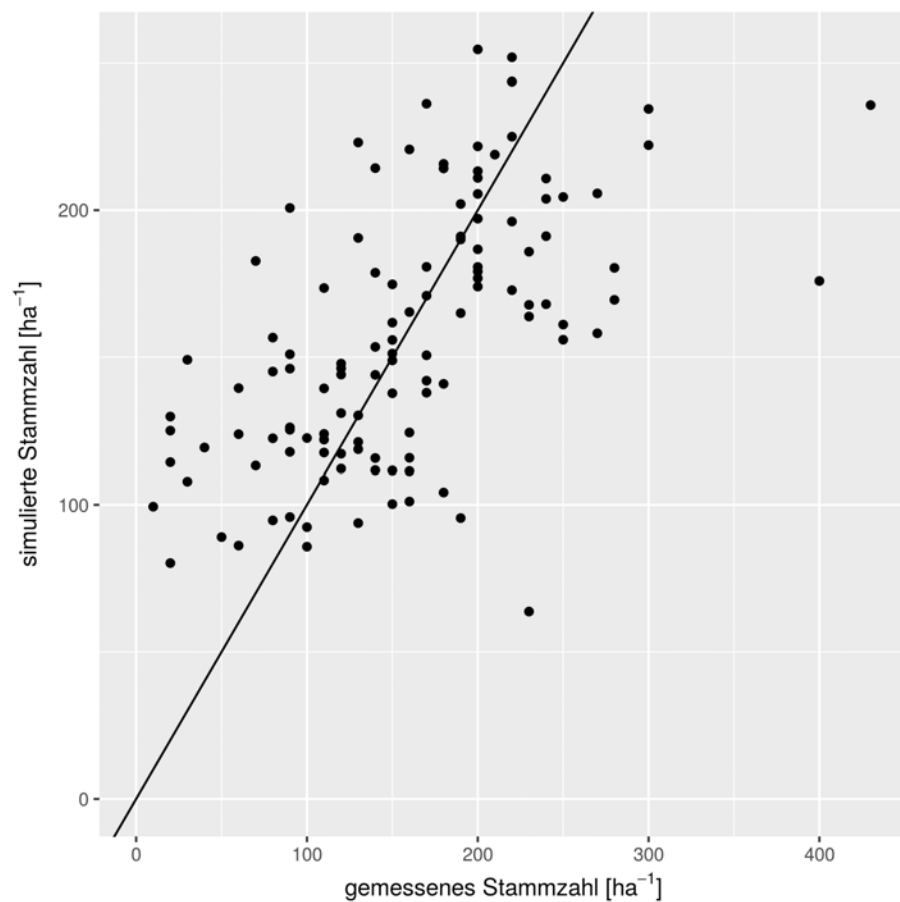


Abbildung 31: Unabh angige Daten: Modell fuer TS Stammzahl [ $\text{ha}^{-1}$ ] aus FR Daten. Gemessene Werte vs. simulierte Werte.  $\text{corr.R-Quadrat}$ : 0.32. (Liegen die Punkte unterhalb der Linie untersch etzt das Modell, liegen die Punkte ueber der Linie liegt eine Uebersch etzung vor)

Tabelle 15: vergleich Koeffizienten fuer Modell

koeffs	wert_ag	wert_fr	przt_diff
(Intercept)	181.19	170.85	6.00
I(chm_stddev^2)	-0.66	-0.12	83.00
chm_stddev	-10.71	-9.81	8.00
I(chm_mean_sum_of_squares^2)	-0.00	-0.00	77.00
chm_mean_sum_of_squares	0.79	0.31	61.00
asinsqrt(chm_mean_sum_of_squares.Laub_Anteil)	-44.81	-13.19	71.00

### 2.3.3 Bootstrap



Tabelle 16: Bootstrap Modell

Parameter	Koeffizient	se	unteres_95pct_VI	oberes_95pct_VI
(Intercept)	181.19	31.99	118.70	245.20
I(chm_stddev^2)	-0.66	0.46	-1.76	0.10
chm_stddev	-10.71	7.92	-23.46	7.64
I(chm_mean_sum_of_squares^2)	-0.00	0.00	-0.00	-0.00
chm_mean_sum_of_squares	0.79	0.10	0.58	0.97
asinsqrt(chm_mean_sum_of_squares_Laub_Anteil)	-44.81	11.14	-65.35	-20.89

### 2.3.4 Kreuzvalidierung

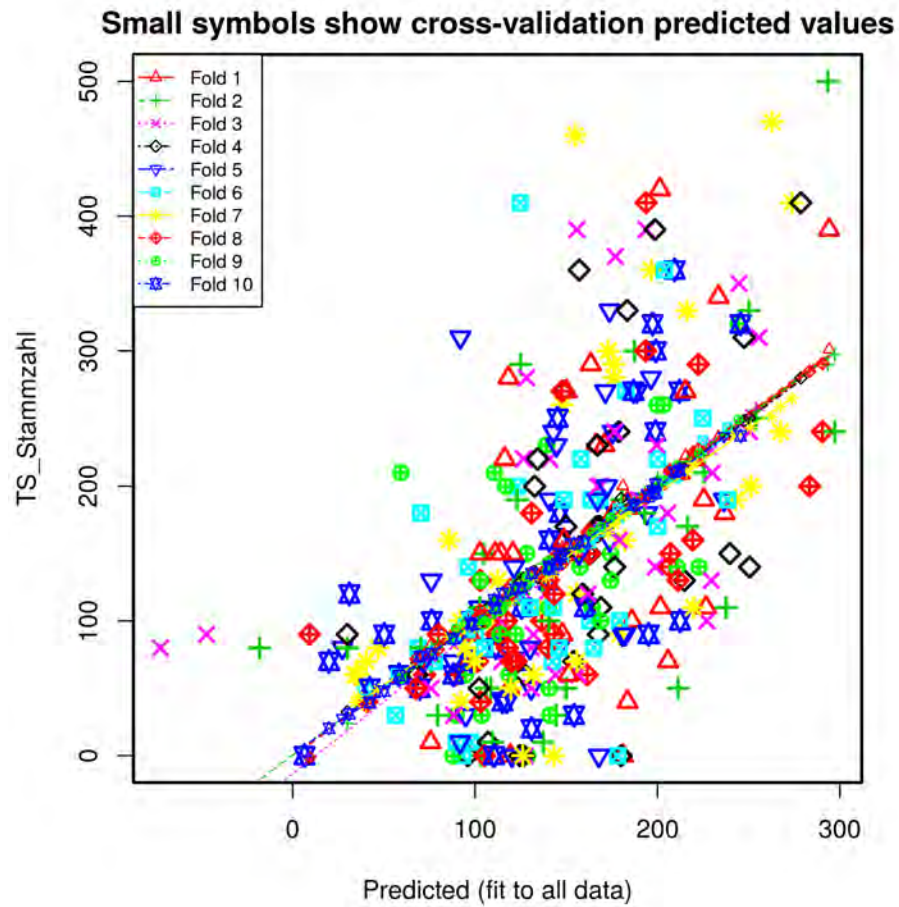


Abbildung 32: Ergebnis von 10 Kreuzvalidierungs Durchläufen (je 34 Faelle im Testset). Das Kreuzvalidierte  $R^2$  betraegt 0.0111. Das Kreuzvalidierte RMSE betraegt 236[m<sup>3</sup>/ha].

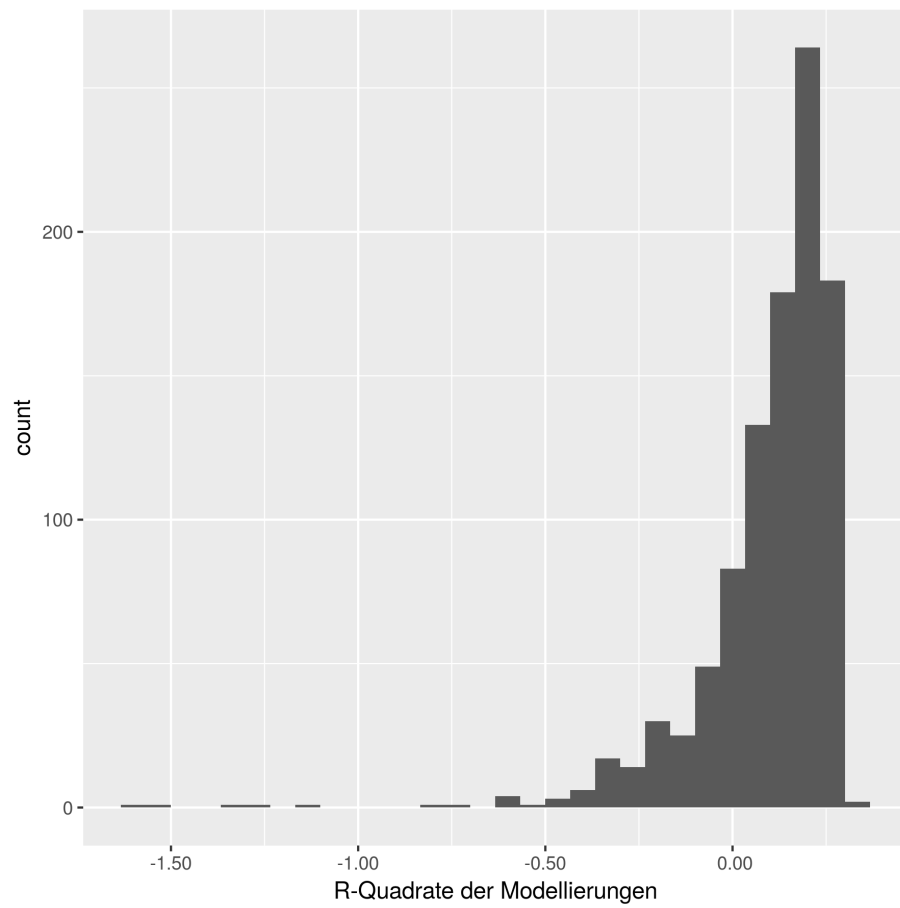


Abbildung 33: Ergebnis von 1000 Kreuzvalidierungen-Wiederholungen. Das Korrigiert R-Quadrat wurde jeweils aus den Simulationen mit den 90% der nicht fuer die Modellschaetzung verwendeten Daten berechnet.

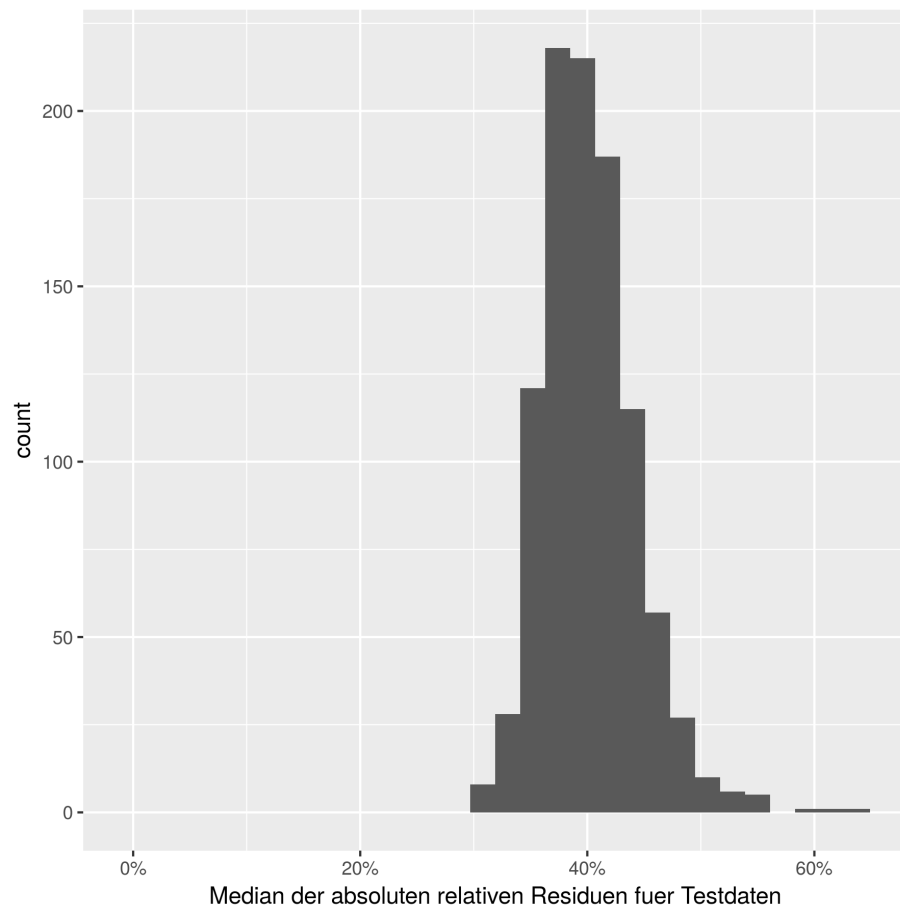


Abbildung 34: Ergebnis von 1000 Kreuzvalidierungen-Wiederholungen. Die absoluten relativen Residuen wurde jeweils aus den Simulationen mit den 90% der nicht fuer die Modellschaetzung verwendeten Daten berechnet. Gezeigt ist die Verteilung der Mediane der relativen Residuen.

### 2.3.5 Residuenanalyse

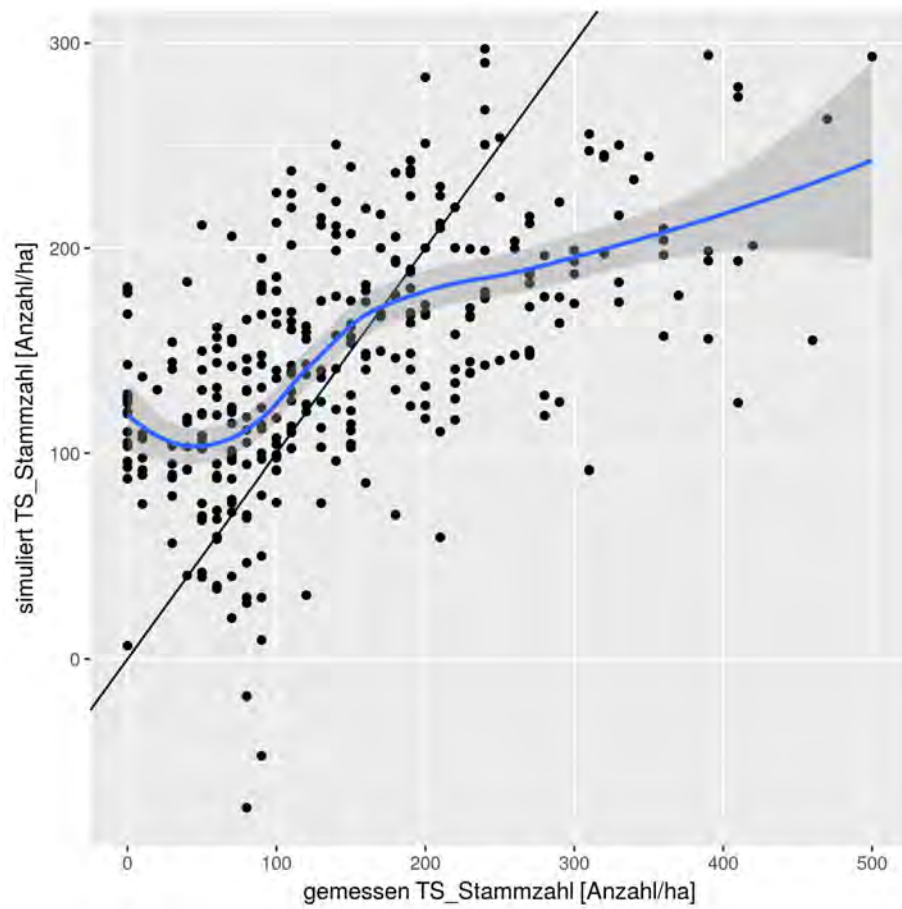


Abbildung 35: Modellannahme 1: Linearer Zusammenhang gemessene Werte und simulierte Werte

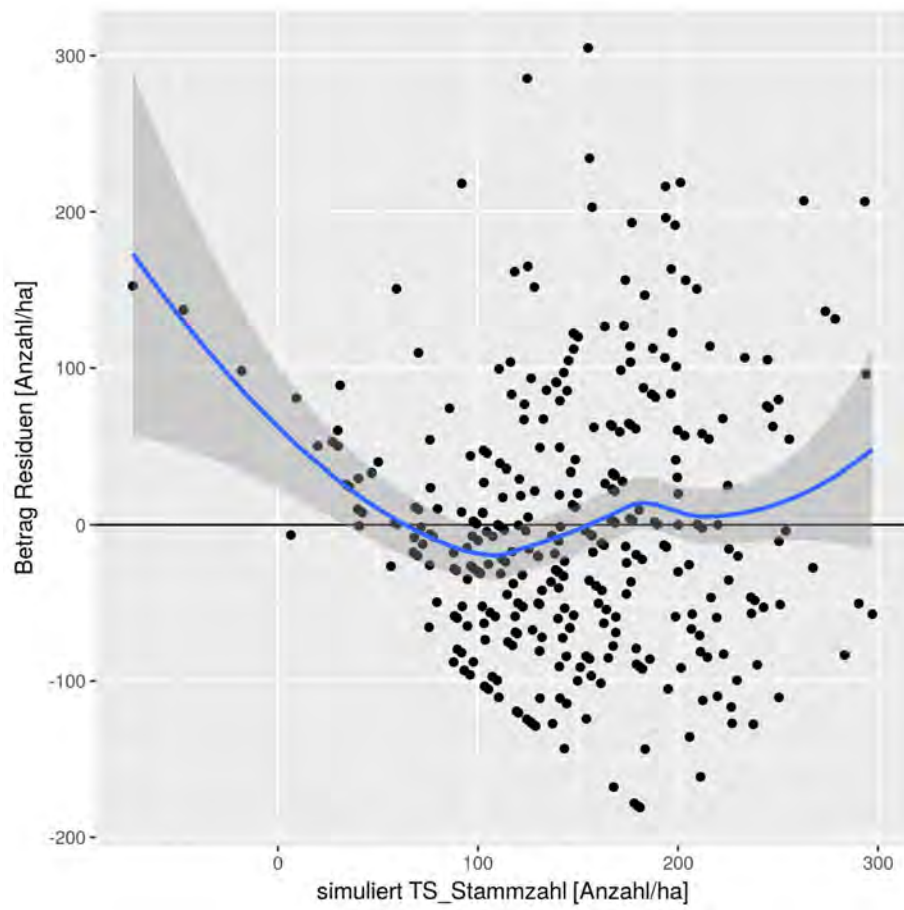


Abbildung 36: Modellannahme 2: Tukey-Anscombe Plot. Die Residuen sind unabhangig von den den simulierten Werten.

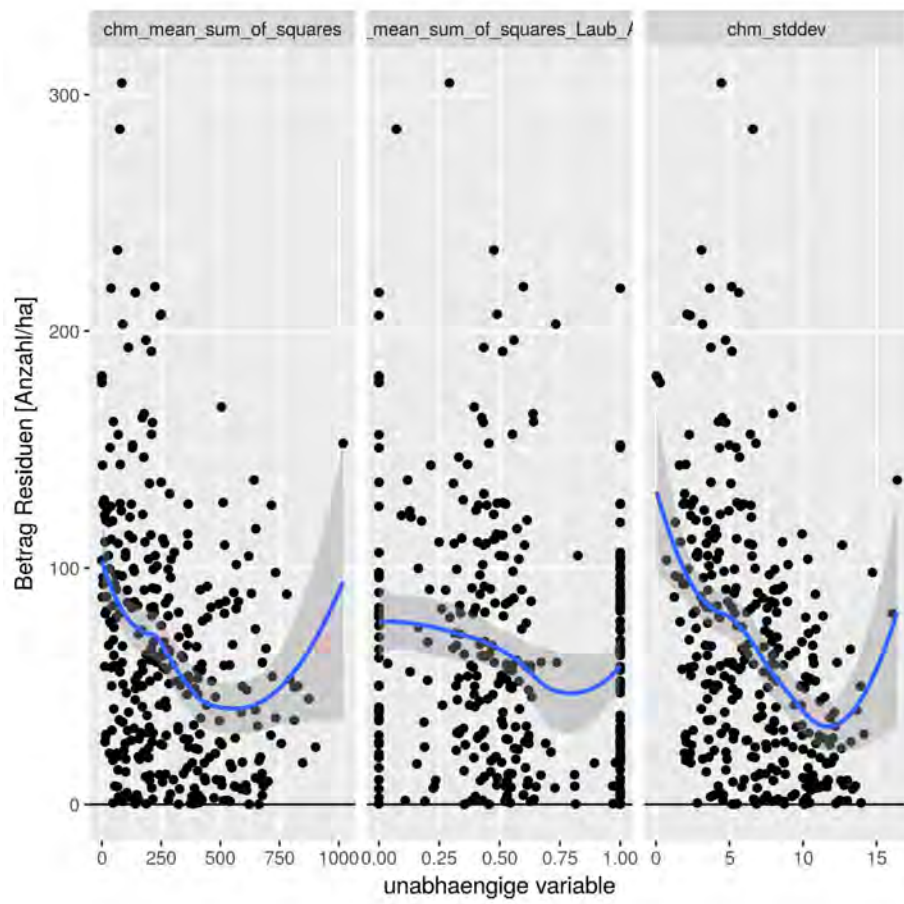


Abbildung 37: Modellannahme 2: Die Residuen sind unabhaengig von den den unabhaengigen Werten.

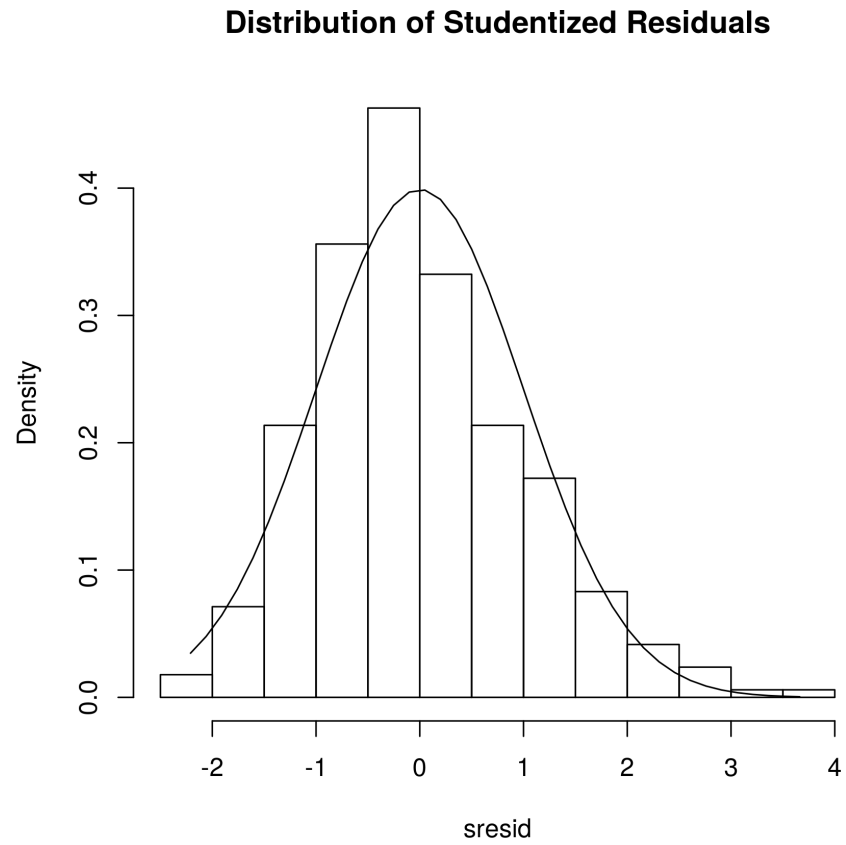


Abbildung 38: Modellannahme 4: Die Residuen sind normalverteilt um 0.



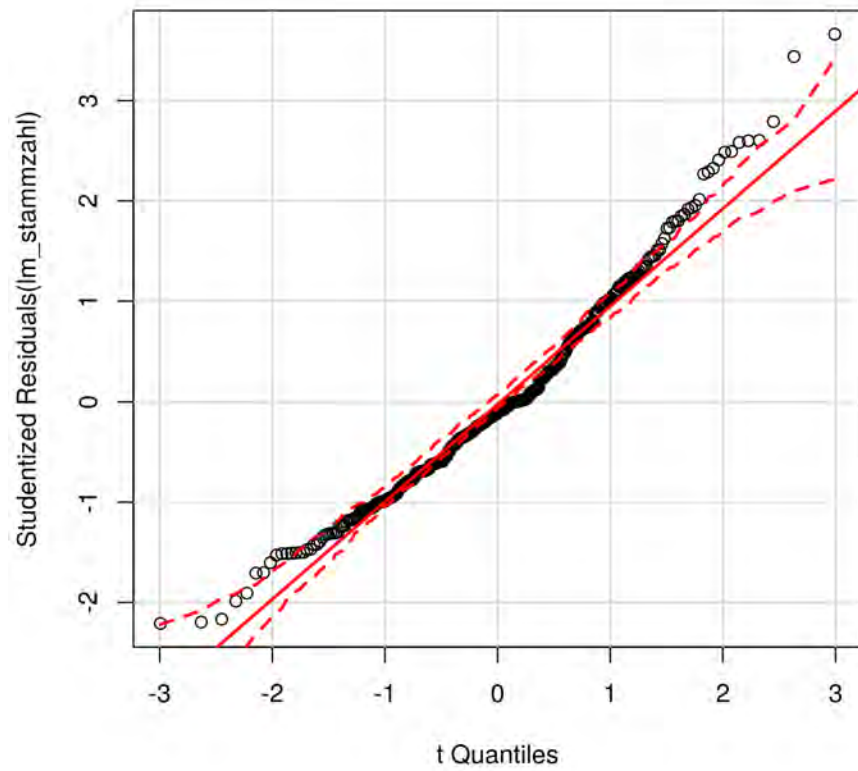


Abbildung 39: Modellannahme 4: Die Residuen sind normalverteilt.

### 2.3.6 Alternative Modelle

Tabelle 17: Beschreibung diverse Modelle fuer TS Stammzahl. Das hoechste korr.R-Quadrat bzw. der tiefste AIK Wert sind rot markiert.

Modell	Formel	korr.RQuadrat	AIC
predict_dlm1	TS_Stammzahl ~ chm_mean_sum_of_squares * asins- qrt(chm_mean_sum_of_squares_Laub_Anteil)	0.02	4077.83
predict_dlm2	TS_Stammzahl ~ chm_mean_sum_of_squares + asins- qrt(chm_mean_sum_of_squares_Laub_Anteil)	0.02	4076.5
predict_dlm3	TS_Stammzahl ~ I(chm_mean_sum_of_squares_Laub^2) + I(chm_mean_sum_of_squares_Nadel^2) + chm_mean_sum_of_squares_Laub + chm_mean_sum_of_squares_Nadel	0.05	4070.05
predict_dlm4	TS_Stammzahl ~ ihst(chm_mean_sum_of_squares_Laub) + ihst(chm_mean_sum_of_squares_Nadel) + chm_mean_sum_of_squares_Laub + chm_mean_sum_of_squares_Nadel	0.08	4059.46
predict_dlm5	TS_Stammzahl ~ I(chm_stddev^2) + chm_stddev + I(chm_mean_sum_of_squares^2) + chm_mean_sum_of_squares + asins- qrt(chm_mean_sum_of_squares_Laub_Anteil)	0.32	3960.12
predict_dlm6	TS_Stammzahl ~ ihst(chm_c15_Laub) + ihst(chm_NumberOfValidValues_Nadel) + chm_c15_Laub + chm_NumberOfValidValues_Nadel	0.06	4065.76
predict_dlm7	TS_Stammzahl ~ ihst(chm_mean_sum_of_squares_Nadel) + chm_mean_sum_of_squares_Nadel + chm_mean_sum_of_squares_Laub	0.08	4057.69

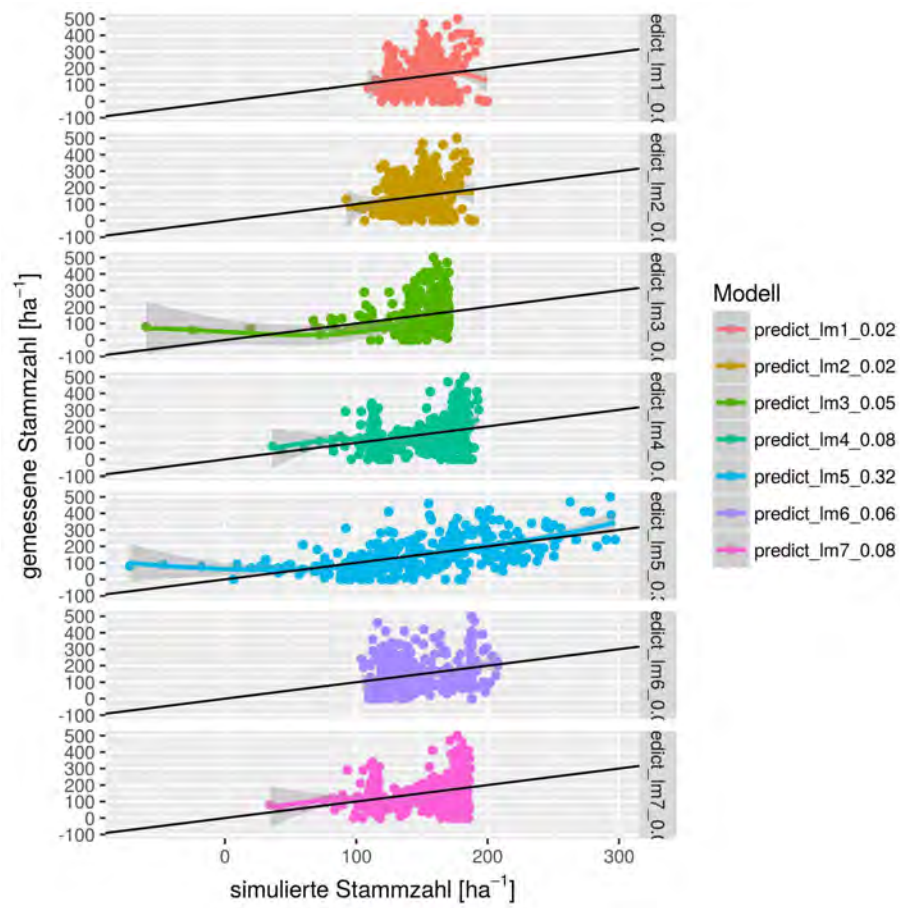


Abbildung 40: Vorhersagegenauigkeit fuer unterschiedliche Modelle (siehe Tabelle 5 fuer Beschreibung der Modelle). Die Vorhersagen von lm\_5 sind aus dem diskutierten Modell erstellt.

### 2.3.7 Ausreisser

NULL

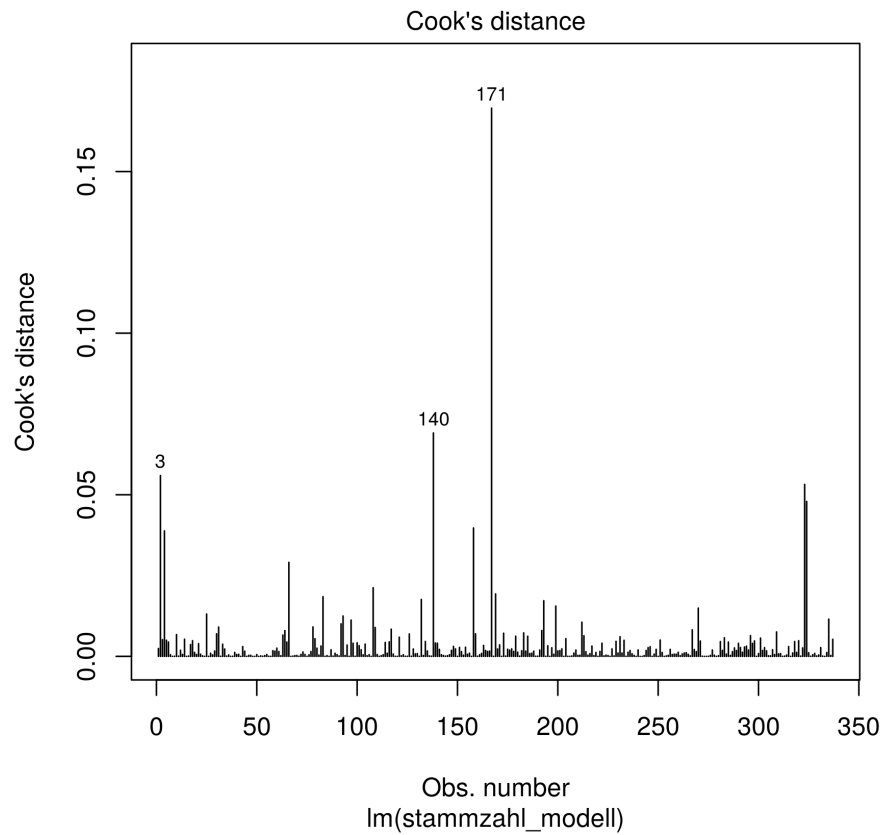


Abbildung 41: Ausreisser der Modellierung: Cook's D Plot. Cooks Distance bewertet den Einfluss einer Beobachtung auf das Ergebnis der Koeffizientenschätzung.

Tabelle 18: Die durch Bonferroni Outlier Test identifizierten Ausreisser								
Zeilen	Plot ID	TS	TS To-	TS	TS	chm	chm	CooksD
Num-		BHDN	tal Vol	LbH	NdH	mean	mean	
mer		mean		Vol	Vol	sum of	sum of	
						squares	squares	
						Laub	Nadel	
175	246670	19.50	313.85	116.53	197.33	157	211	FALSE

## 3 Diskussion der Resultate

### 3.1 Modell fuer Vorrat

Das hier gezeigte Modell aus CHM Daten fuer totalen stehenden Vorrat [ $\text{m}^3$ ] hat ein korrigiertes R-Quadrat von 0.59. Aus 1000 Wiederholungen betraegt der relative Fehler<sup>8</sup> im Median 30% und das korrigierte R-Quadrat im Median 0.53.

### 3.2 Modell fuer Basalflaeche

Das hier gezeigte Modell aus CHM Daten fuer Basalflaeche [ $\text{m}^2/\text{ha}$ ] hat ein korrigiertes R-Quadrat von 0.55. Aus 1000 Wiederholungen betraegt der relative Fehler im Median 28% und das korrigierte R-Quadrat im Median 0.48.

### 3.3 Modell fuer Stammzahl

Das hier gezeigte Modell aus CHM Daten fuer Stammzahl [ $\text{ha}^{-1}$ ] hat ein korrigiertes R-Quadrat von 0.32. Aus 1000 Wiederholungen betraegt der relative Fehler im Median 40% und das korrigierte R-Quadrat im Median 0.15.

### 3.4 Uebersicht ueber die drei Modelle

Die Tabelle 19 und Grafiken 42 und 43 zeigen, dass die Modelle fuer Volumen und Basalflaeche mit einem R-Quadrat zwischen 0.5 und 0.7 nuetzliche Schaetzungen liefern, mit relativen Schaetzfehlern um 30% vom gemessenen Wert. Das Modell fuer Stammzahl erklart deutlich weniger Varianz und auch die relativen Schaetzfehler sind mit ca 40% hoeher. Gezeigt ist auch der Median des relativen RSE (Residual Standard Error)<sup>9</sup> fuer jede der 1000 Koeffizienten Schaetzungen (im Gegensatz zum relativen Schaetzfehler wird der RSE aus den Werten berechnet die fuer die Schaetzung der Koeffizienten benutzt werden. Die relativen Schaetzfehler werden aus den Testdaten berechnet).

Dieses Ergebnis deckt sich mit der Erwartung, dass das Hoehenmodell, aggregiert auf Stichprobenebene nur beschaenkt in der Lage ist einzelne Baeume voneinander ab zu grenzen.

---

<sup>8</sup>Die Differenz jeder einzelnen Messung - geschaetzter Wert wird durch den gemessenen Wert geteilt und ergibt die Prozentuale Abweichung vom "wahren" Wert

<sup>9</sup>rel. RSE: Residual Standard Error (RSE) der Koeffizientenschaetzung / Mittelwert der "wahren" Werte

Tabelle 19: Die drei Modelle im Vergleich

Modell	adj.R Quadrat	Median R Quadrat	Median abs relative Re- siduen	Median rel RSE
	(direkt)	(1000 Wiederholungen)		
Volumen_m3proha	0.59	0.53	30%	32%
Basalarea_m2proha	0.55	0.48	28%	31%
Stammzahl_proha	0.32	0.15	40%	41%

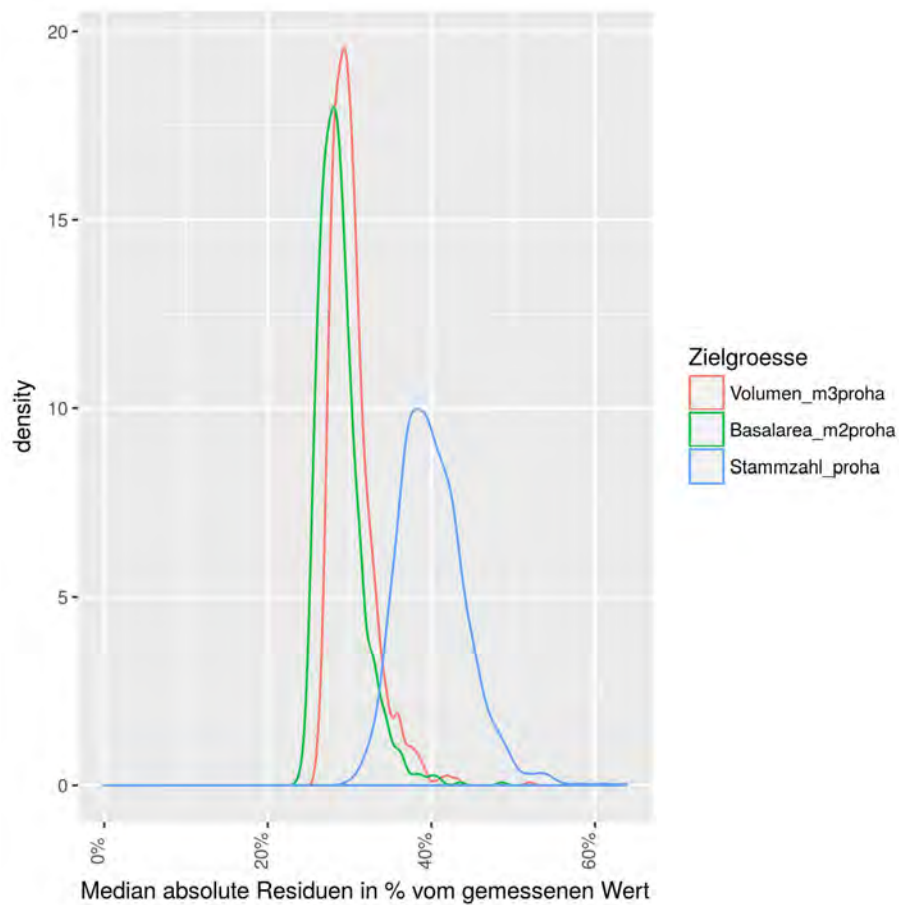


Abbildung 42: Vergleich der Verteilungen der Mediane der absoluten relativen Residuen fuer die 1000 Kreuzvalidierung Wiederholungen. Tief Werte markieren kleine Unsicherheiten, grosse Werte grosse Unsicherheit.

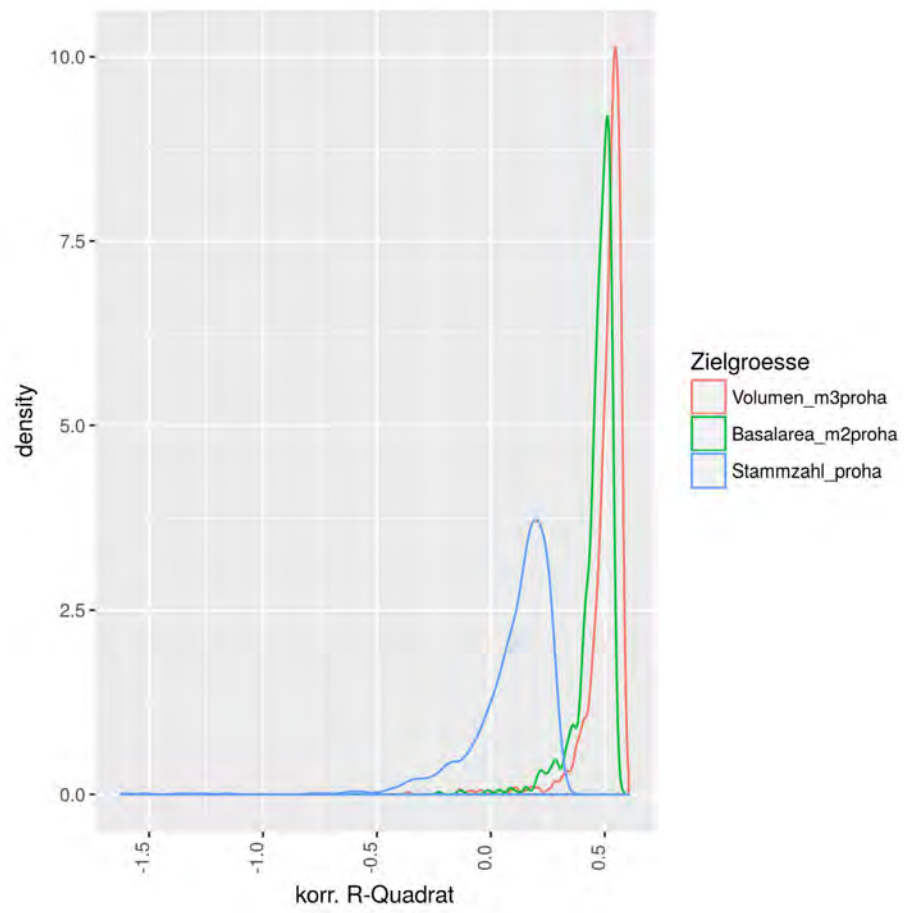


Abbildung 43: Vergleich der Verteilungen der korr. R-Quadrate fuer die 1000 Kreuzvalidierung Wiederholungen. Hohe Werte markieren grosse Anteile erklarter Varianz.



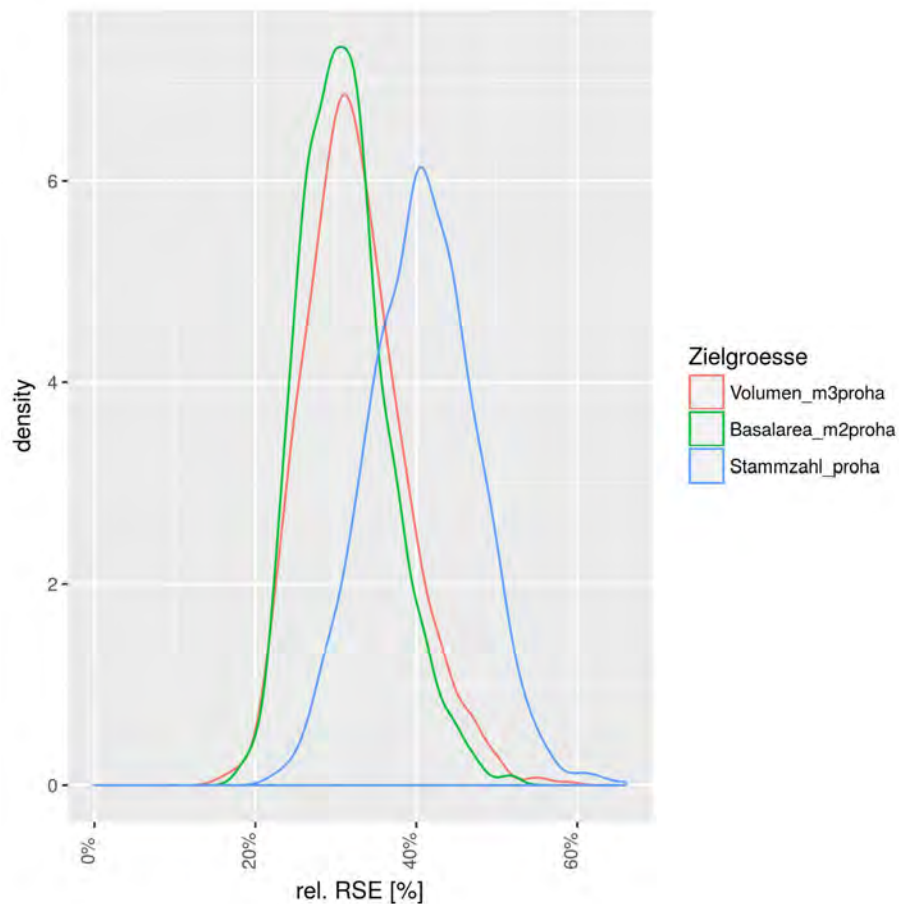


Abbildung 44: Vergleich der Verteilungen relativen RSE fuer die 1000 Kreuzvalidierung Wiederholungen. Hohe Werte stehen fuer grosse Standardfehler, tiefere Werte fuer tiefe Standardfehler.

### 3.4.1 Fazit

Das hier gezeigte Vorgehen sieht vor, dass um aus dem Canopy-Height-Model Schaetzungen fuer Vorratsvolumen und Basalflaeche zu erhalten, ein Modell aus CHM Paramtern vor Ort mit gemessenen Werten parametrisiert wird.

Die Modelle fuer Vorratsvolumen und Basalflaeche lassen sich gut mit einem kleinen vermessenen Teil des Bestandes Parametrisieren (hier im Beispiel Tabelle 19 10% der Stichproben). Diese Modelle koennen dann auf den Rest der Stichproben angewendet werden. Die Genauigkeit der Schaetzungen variiert fuer die Modelle, betraegt aber fuer Vorratsvolumen und Basalflaeche ca  $\pm 30\%$  (vergleiche Tabelle 19 und Abbildung 42). Im Gegensatz dazu liefert das Modell fuer Stammzahl Schaetzungen mit Schaetzfehlern um  $\pm 40\%$ .

Wie in den Kapiteln "Kreuzvalidierung" am Beispiel Bremgarten gezeigt kann mit einem zufaellig gewaehlten Subset (hier 10%) der Messpunkte eine robuste Schaetzung der Koeffizienten fuer ein einfaches (lineares) Modell erreicht werden. Fuer die Parameter "Vorratsvolumen [ $\text{m}^3/\text{ha}$ ]" und "Basalflaeche [ $\text{m}^2/\text{ha}$ ]" wurde in 70% (Vorrat) bzw. 38% (Basalflaeche) der Wiederholungen ein korr. R-Quadrat von  $> 0.5$  gefunden. In 2% (Vorrat) bzw. 2% (Basalflaeche) der Wiederholungen wurde ein korr. R-Quadrat von  $< 0.2$  gefunden. Abbildung 14 und 21 zeigen die Entwicklung der hoechsten und tiefsten gefundenen korr. R-Quadrate aus 1000 Wiederholten Versuchen mit unterschiedlichen Anteilen der Grundgesamtheit die Koeffizienten zu schaetzen. Dabei wurden jeweils 1%, 2% ... bis 50% der Faelle benutzt. Die korr. R-Quadrate stabilisieren sich bei ca. 20% benutzte Faelle.

## Literatur

- Martin Buehlmann, Peter Maechler. *Computational Statistics*. ETHZ, Seminar fuer Statistik, 2 2008.
- R.D. Cook and S. Weisberg. *Residuals and Influence in Regression*. Monographs on statistics and applied probability. Chapman & Hall, 1982. ISBN 9780412242809. URL <https://books.google.ch/books?id=MVSqAAAAIAAJ>.
- A. C. Davison and D. V. Hinkley. *Bootstrap Methods and Their Applications*. Cambridge University Press, Cambridge, 1997. URL <http://statwww.epfl.ch/davison/BMA/>. ISBN 0-521-57391-2.
- John Fox. Regression diagnostics. electronic, November 2009. URL <http://socserv.socsci.mcmaster.ca/jfox/Courses/Brazil-2009/index.html>.
- John H. Maindonald and W. John Braun. *DAAG: Data Analysis and Graphics Data and Functions*, 2015. URL <https://CRAN.R-project.org/package=DAAG>. R package version 1.22.
- Samuel Oyamakin and Angela Chukwu. On hyperbolic sine function in height-diameter modeling of gmelina arborea. 12:139–156, 12 2014.
- R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2006. URL <http://www.R-project.org>. ISBN 3-900051-07-0.
- Robert R. Sokal and F. James Rohlf. *Biometry : the principles and practice of statistics in biological research*. W. H. Freeman and Company, New York, 1995.
- Werner Stahel. *Lineare Regression*. ETHZ, Seminar fuer Statistik, May 2012.

## 4 Anhang

## 4.1 Beschreibende Analysen

Tabelle 20: Beschreibung ausgewaehlter numerischen Parameter Teil 1

	n	mean	sd	median	trimmed	min	max	skew	kurtosis	se
chm_mean_sum_of_squares_Nadel	337	260.51	246.35	226.03	231.40	0.00	1138.52	0.83	0.09	13.42
chm_sum_of_squares_Nadel	337	486729.89	627210.66	173908.28	370246.79	0.00	2636989.92	1.37	1.04	34166.34
chm_surf_int_Nadel	337	486729.89	627210.66	173908.28	370246.79	0.00	2636989.92	1.37	1.04	34166.34
chm_coverage_Nadel	337	0.67	0.40	0.88	0.71	0.00	1.00	-0.95	-0.87	0.02
chm_average_Nadel	337	11.78	8.61	13.20	11.49	0.00	32.70	0.01	-1.04	0.47
chm_stddev_Nadel	337	5.40	4.32	5.04	5.15	0.00	16.43	0.28	-1.06	0.24
chm_maximum_Nadel	337	21.17	14.04	23.83	21.32	0.00	47.46	-0.34	-1.15	0.76
chm_minimum_Nadel	337	0.89	3.16	0.00	0.08	-0.28	23.88	4.82	25.71	0.17
chm_NumberOfValidValues_Nadel	337	1323.89	1333.19	901.00	1208.79	0.00	3593.00	0.59	-1.19	72.62
chm_NumberOfNoDataValues_Nadel	337	0.00	0.00	0.00	0.00	0.00	0.00			0.00
chm_NumberOfValuesInclNoData_Nadel	337	0.00	0.00	0.00	0.00	0.00	0.00			0.00
chm_p99_Nadel	337	20.19	13.51	22.31	20.29	0.00	44.51	-0.31	-1.16	0.74
chm_p95_Nadel	337	18.96	12.86	20.75	19.00	0.00	42.48	-0.26	-1.19	0.70
chm_p90_Nadel	337	17.89	12.33	19.91	17.87	0.00	40.77	-0.22	-1.22	0.67
chm_p80_Nadel	337	16.47	11.72	18.08	16.36	0.00	38.78	-0.15	-1.28	0.64
chm_p75_Nadel	337	15.88	11.43	17.49	15.74	0.00	38.03	-0.13	-1.30	0.62
chm_p70_Nadel	337	15.28	11.12	16.92	15.10	0.00	37.39	-0.11	-1.30	0.61
chm_p60_Nadel	337	14.02	10.51	15.45	13.73	0.00	36.09	-0.03	-1.29	0.57
chm_p50_Nadel	337	12.61	9.91	13.36	12.20	0.00	34.49	0.08	-1.27	0.54
chm_p40_Nadel	337	10.84	9.16	11.61	10.22	0.00	33.23	0.26	-1.12	0.50
chm_p30_Nadel	337	8.96	8.34	8.11	8.12	0.00	31.50	0.53	-0.76	0.45
chm_p25_Nadel	337	7.88	7.86	6.76	6.92	0.00	30.39	0.73	-0.36	0.43
chm_p20_Nadel	337	6.68	7.39	4.66	5.58	0.00	29.74	1.01	0.22	0.40
chm_p10_Nadel	337	4.31	6.24	0.33	3.03	0.00	28.62	1.62	2.09	0.34
chm_p05_Nadel	337	2.77	5.03	0.01	1.61	0.00	27.69	2.27	5.45	0.27
chm_p01_Nadel	337	1.41	3.76	0.00	0.43	0.00	23.89	3.74	15.82	0.20
chm_c00_Nadel	337	0.72	0.41	0.95	0.78	0.00	1.00	-1.15	-0.58	0.02
chm_c02_Nadel	337	0.66	0.39	0.85	0.70	0.00	1.00	-0.91	-0.90	0.02
chm_c05_Nadel	337	0.61	0.39	0.79	0.64	0.00	1.00	-0.70	-1.19	0.02
chm_c10_Nadel	337	0.51	0.38	0.65	0.52	0.00	1.00	-0.31	-1.55	0.02
chm_c15_Nadel	337	0.39	0.35	0.41	0.37	0.00	1.00	0.14	-1.52	0.02

Tabelle 21: Beschreibung ausgewaehlter numerischen Parameter Teil 2

	n	mean	sd	median	trimmed	min	max	skew	kurtosis	se
chm_c20_Nadel	337	0.26	0.30	0.10	0.22	0.00	1.00	0.75	-0.81	0.02
chm_c25_Nadel	337	0.15	0.24	0.00	0.10	0.00	0.98	1.52	1.32	0.01
chm_c30_Nadel	337	0.06	0.14	0.00	0.03	0.00	0.77	2.59	6.88	0.01
chm_c35_Nadel	337	0.02	0.05	0.00	0.00	0.00	0.47	4.74	26.35	0.00
chm_c40_Nadel	337	0.00	0.01	0.00	0.00	0.00	0.12	6.76	48.09	0.00
chm_c45_Nadel	337	0.00	0.00	0.00	0.00	0.00	0.00	10.93	128.22	0.00
chm_c50_Nadel	337	0.00	0.00	0.00	0.00	0.00	0.00			0.00
chm_mean_sum_of_squares_Laub	337	236.51	231.54	166.88	204.01	0.00	953.87	1.06	0.24	12.61
chm_sum_of_squares_Laub	337	532555.55	622318.11	311692.27	417175.44	0.00	3046620.11	1.52	1.80	33899.83
chm_surf_int_Laub	337	532555.55	622318.11	311692.27	417175.44	0.00	3046620.11	1.52	1.80	33899.83
chm_coverage_Laub	337	0.78	0.32	0.93	0.85	0.00	1.00	-1.71	1.42	0.02
chm_average_Laub	337	11.78	7.86	11.52	11.47	0.00	30.36	0.28	-0.71	0.43
chm_stddev_Laub	337	4.96	3.39	4.54	4.81	0.00	14.32	0.36	-0.73	0.18
chm_maximum_Laub	337	21.44	11.65	22.88	21.95	0.00	47.22	-0.40	-0.67	0.63
chm_minimum_Laub	337	0.87	3.47	0.00	0.01	-2.94	23.71	4.37	19.23	0.19
chm_NumberOfValidValues_Laub	337	1983.26	1353.69	2254.00	2028.75	0.00	3593.00	-0.25	-1.50	73.74
chm_NumberOfNoDataValues_Laub	337	0.00	0.00	0.00	0.00	0.00	0.00			0.00
chm_NumberOfValuesInclNoData_Laub	337	0.00	0.00	0.00	0.00	0.00	0.00			0.00
chm_p99_Laub	337	20.21	11.25	21.29	20.62	0.00	41.91	-0.35	-0.79	0.61
chm_p95_Laub	337	18.65	10.78	19.63	18.92	0.00	38.58	-0.24	-0.91	0.59
chm_p90_Laub	337	17.54	10.40	18.29	17.70	0.00	37.35	-0.16	-0.96	0.57
chm_p80_Laub	337	15.93	9.90	16.28	15.93	0.00	35.63	-0.02	-0.99	0.54
chm_p75_Laub	337	15.25	9.71	14.92	15.16	0.00	34.71	0.06	-0.99	0.53
chm_p70_Laub	337	14.69	9.53	14.17	14.53	0.00	34.12	0.11	-0.99	0.52
chm_p60_Laub	337	13.55	9.18	12.52	13.25	0.00	33.26	0.24	-0.96	0.50
chm_p50_Laub	337	12.23	8.88	10.68	11.75	0.00	32.26	0.38	-0.90	0.48
chm_p40_Laub	337	11.05	8.55	9.42	10.43	0.00	30.98	0.47	-0.80	0.47
chm_p30_Laub	337	9.55	8.01	8.17	8.75	0.00	29.29	0.66	-0.50	0.44
chm_p25_Laub	337	8.55	7.66	6.95	7.62	0.00	28.39	0.82	-0.21	0.42
chm_p20_Laub	337	7.44	7.38	5.66	6.40	0.00	27.61	0.98	0.11	0.40
chm_p10_Laub	337	4.96	6.49	2.19	3.61	0.00	26.35	1.57	1.75	0.35

Tabelle 22: Beschreibung ausgewählter numerischen Parameter Teil 3

	n	mean	sd	median	trimmed	min	max	skew	kurtosis	se
chm_p05_Laub	337	3.36	5.60	0.28	2.01	0.00	25.89	2.11	4.00	0.31
chm_p01_Laub	337	1.70	4.40	0.00	0.48	0.00	24.37	3.26	10.56	0.24
chm_c00_Laub	337	0.84	0.33	0.98	0.92	0.00	1.00	-2.04	2.35	0.02
chm_c02_Laub	337	0.77	0.33	0.90	0.83	0.00	1.00	-1.63	1.18	0.02
chm_c05_Laub	337	0.69	0.33	0.82	0.74	0.00	1.00	-1.10	-0.18	0.02
chm_c10_Laub	337	0.50	0.36	0.55	0.50	0.00	1.00	-0.18	-1.49	0.02
chm_c15_Laub	337	0.36	0.35	0.25	0.32	0.00	1.00	0.46	-1.29	0.02
chm_c20_Laub	337	0.23	0.31	0.04	0.17	0.00	1.00	1.14	-0.13	0.02
chm_c25_Laub	337	0.14	0.25	0.00	0.08	0.00	0.97	1.79	2.02	0.01
chm_c30_Laub	337	0.04	0.12	0.00	0.01	0.00	0.66	3.07	9.20	0.01
chm_c35_Laub	337	0.01	0.03	0.00	0.00	0.00	0.23	5.38	32.45	0.00
chm_c40_Laub	337	0.00	0.00	0.00	0.00	0.00	0.04	8.79	83.30	0.00
chm_c45_Laub	337	0.00	0.00	0.00	0.00	0.00	0.00	18.19	330.02	0.00
chm_c50_Laub	337	0.00	0.00	0.00	0.00	0.00	0.00			0.00
TS_ddom	336	36.37	17.51	35.50	36.46	0.00	75.33	-0.03	-0.53	0.96
TS_Stammzahl	337	146.68	102.97	120.00	137.27	0.00	500.00	0.85	0.30	5.61
TS_BHDN_mean	336	25.44	12.27	23.00	25.04	0.00	64.00	0.33	0.35	0.67
TS_BHDN_std	336	8.49	6.40	6.69	7.92	0.00	26.82	0.77	-0.26	0.35
TS_BHDN_max	336	42.44	21.59	42.00	42.28	0.00	112.00	0.14	-0.36	1.18
TS_BasalArea	336	8428.58	5497.21	8078.21	8206.19	0.00	24786.38	0.32	-0.58	299.90
TS_PBF	337	245898.14	6620.71	246660.00	245842.43	234664.00	259680.00	0.01	-1.24	360.65
TS_X_Koord	337	666771.59	666.18	666725.33	666766.70	665197.15	668559.38	0.14	-0.04	36.29
TS_Y_Koord	337	245034.97	989.04	245133.08	245026.94	243355.12	247100.00	0.01	-1.25	53.88
TS_Total_Vol	337	265.28	193.79	244.33	250.78	0.00	897.05	0.57	-0.32	10.56
TS_NdH_Vol	337	146.57	186.50	60.32	114.40	0.00	897.05	1.30	0.87	10.16
TS_Winterkahl_Vol	337	14.66	18.65	6.03	11.44	0.00	89.70	1.30	0.87	1.02
TS_BasalAreaha	337	211.53	138.92	202.53	205.75	0.00	619.66	0.32	-0.59	7.57
TS_Stemnumber	337	14.61	10.32	12.00	13.66	0.00	50.00	0.85	0.30	0.56
TS_STZha	337	370.20	274.68	300.00	342.71	0.00	2071.43	1.37	3.90	14.96
TS_Immergruen_Vol	337	11.87	13.00	7.92	9.80	0.00	75.48	1.35	1.85	0.71
TS_LbH_Vol	337	118.71	129.99	79.25	98.04	0.00	754.83	1.35	1.85	7.08



Tabelle 23: Beschreibung ausgewählter numerischen Parameter Teil 4

		n	mean	sd	median	trimmed	min	max	skew	kurtosis	se
73	TS_Anteil_NdH	317	0.46	0.41	0.39	0.45	0.00	1.00	0.12	-1.68	0.02
	TS_Anteil_Immergruen	317	0.54	0.41	0.61	0.55	0.00	1.00	-0.12	-1.68	0.02
	TS_Anteil_Immergruen_rd	317	0.54	0.41	0.60	0.55	0.00	1.00	-0.13	-1.67	0.02
	TS_Anteil_NdH_rd	317	0.46	0.41	0.40	0.45	0.00	1.00	0.13	-1.67	0.02
	chm_mean_sum_of_squares	337	286.47	217.96	232.21	265.63	0.00	1018.42	0.76	-0.20	11.87
	chm_sum_of_squares	337	1029281.01	783145.66	834330.51	954397.60	1.42	3659175.99	0.76	-0.20	42660.66
	chm_surf_int	337	1029281.01	783145.66	834330.51	954397.60	1.42	3659175.99	0.76	-0.20	42660.66
	chm_coverage	337	0.85	0.19	0.92	0.88	0.00	1.00	-2.18	5.22	0.01
	chm_average	337	13.67	6.62	13.58	13.58	0.00	31.18	0.11	-0.58	0.36
	chm_stddev	337	6.66	3.41	6.24	6.51	0.02	16.41	0.37	-0.70	0.19
	chm_maximum	337	26.65	9.31	27.40	26.87	0.39	47.46	-0.23	-0.50	0.51
	chm_minimum	337	0.22	1.52	0.00	-0.00	-2.94	15.01	7.54	62.42	0.08
	chm_NumberOfValidValues	337	3593.00	0.00	3593.00	3593.00	3593.00	3593.00			0.00
	chm_NumberOfNoDataValues	337	0.00	0.00	0.00	0.00	0.00	0.00			0.00
	chm_NumberOfValuesInclNoData	337	3593.00	0.00	3593.00	3593.00	3593.00	3593.00			0.00
	chm_p99	337	24.87	9.31	25.70	25.09	0.09	44.49	-0.21	-0.62	0.51
	chm_p95	337	22.90	9.10	22.97	23.12	0.00	42.16	-0.19	-0.69	0.50
	chm_p90	337	21.28	9.06	21.01	21.46	0.00	40.26	-0.16	-0.77	0.49
	chm_p80	337	19.32	8.84	19.00	19.48	0.00	36.93	-0.12	-0.84	0.48
	chm_p75	337	18.46	8.78	18.26	18.58	0.00	35.65	-0.08	-0.90	0.48
	chm_p70	337	17.71	8.67	17.37	17.82	0.00	34.77	-0.07	-0.91	0.47
	chm_p60	337	16.12	8.37	15.79	16.15	0.00	33.98	0.01	-0.92	0.46
	chm_p50	337	14.43	8.20	14.27	14.37	0.00	32.80	0.08	-0.91	0.45
	chm_p40	337	12.62	7.85	12.06	12.44	0.00	31.70	0.17	-0.82	0.43
	chm_p30	337	10.36	7.41	9.71	9.89	0.00	29.86	0.42	-0.59	0.40
	chm_p25	337	9.00	7.10	8.10	8.36	0.00	28.38	0.62	-0.30	0.39
	chm_p20	337	7.59	6.83	6.28	6.75	0.00	27.61	0.85	0.04	0.37
	chm_p10	337	4.43	5.72	2.12	3.29	0.00	25.99	1.55	1.90	0.31
	chm_p05	337	2.54	4.38	0.16	1.51	0.00	24.19	2.28	5.47	0.24
	chm_p01	337	0.94	2.76	0.00	0.21	0.00	20.89	4.34	22.19	0.15
	chm_c00	337	0.92	0.13	0.97	0.95	0.02	1.00	-3.66	17.16	0.01

Tabelle 24: Beschreibung ausgewaehlter numerischen Parameter Teil 5

	n	mean	sd	median	trimmed	min	max	skew	kurtosis	se
chm_c02	337	0.83	0.20	0.90	0.87	0.00	1.00	-2.04	4.38	0.01
chm_c05	337	0.76	0.24	0.84	0.81	0.00	1.00	-1.53	1.70	0.01
chm_c10	337	0.59	0.31	0.69	0.61	0.00	1.00	-0.64	-0.86	0.02
chm_c15	337	0.43	0.32	0.45	0.42	0.00	1.00	0.00	-1.39	0.02
chm_c20	337	0.28	0.30	0.15	0.24	0.00	1.00	0.70	-0.86	0.02
chm_c25	337	0.16	0.24	0.02	0.12	0.00	0.93	1.32	0.55	0.01
chm_c30	337	0.06	0.13	0.00	0.03	0.00	0.69	2.23	4.28	0.01
chm_c35	337	0.01	0.04	0.00	0.00	0.00	0.28	3.89	16.57	0.00
chm_c40	337	0.00	0.01	0.00	0.00	0.00	0.11	7.28	60.06	0.00
chm_c45	337	0.00	0.00	0.00	0.00	0.00	0.00	10.60	116.61	0.00
chm_c50	337	0.00	0.00	0.00	0.00	0.00	0.00			0.00
chm_mean_sum_of_squares_Laub_Anteil	337	0.52	0.32	0.50	0.53	0.00	1.00	0.09	-0.90	0.02

#### 4.1.1 Volumen

Tabelle 25: Beschreibung gemessenes Volumen

n	mean	sd	median	min	max	skew
337	265.28	193.79	244.33	0.00	897.05	0.57

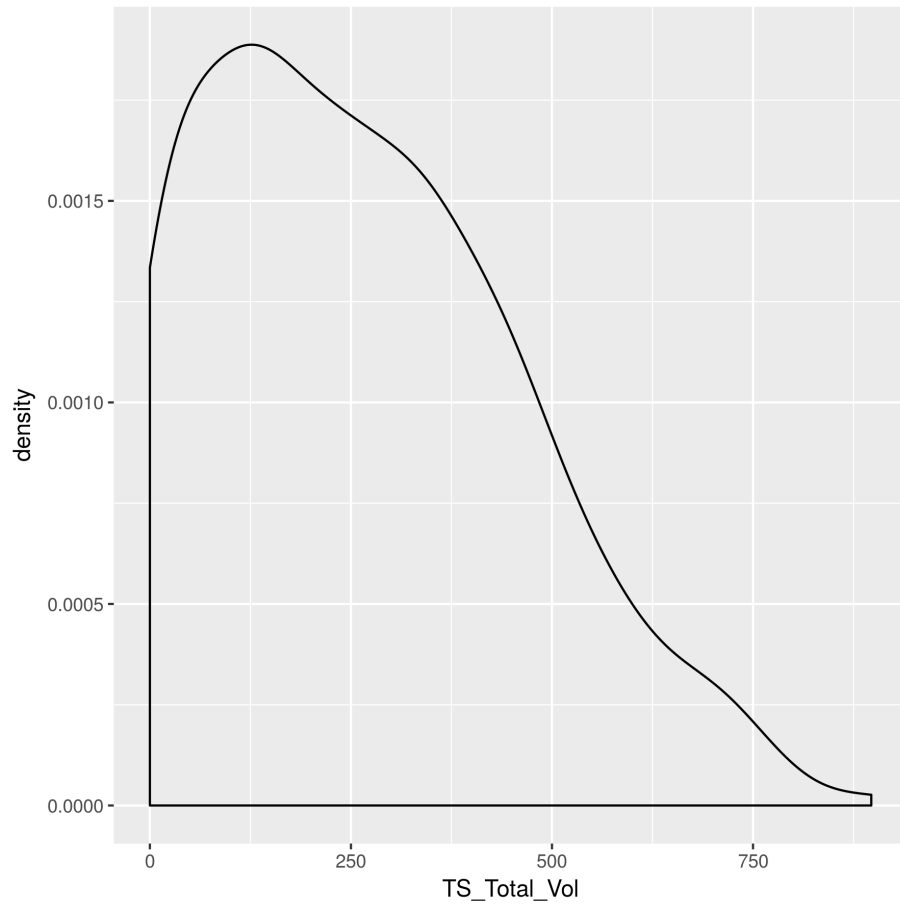


Abbildung 45: Verteilung Variable gemessenes Volumen (TS\_Total\_Vol) [ $\text{m}^3/\text{ha}$ ]

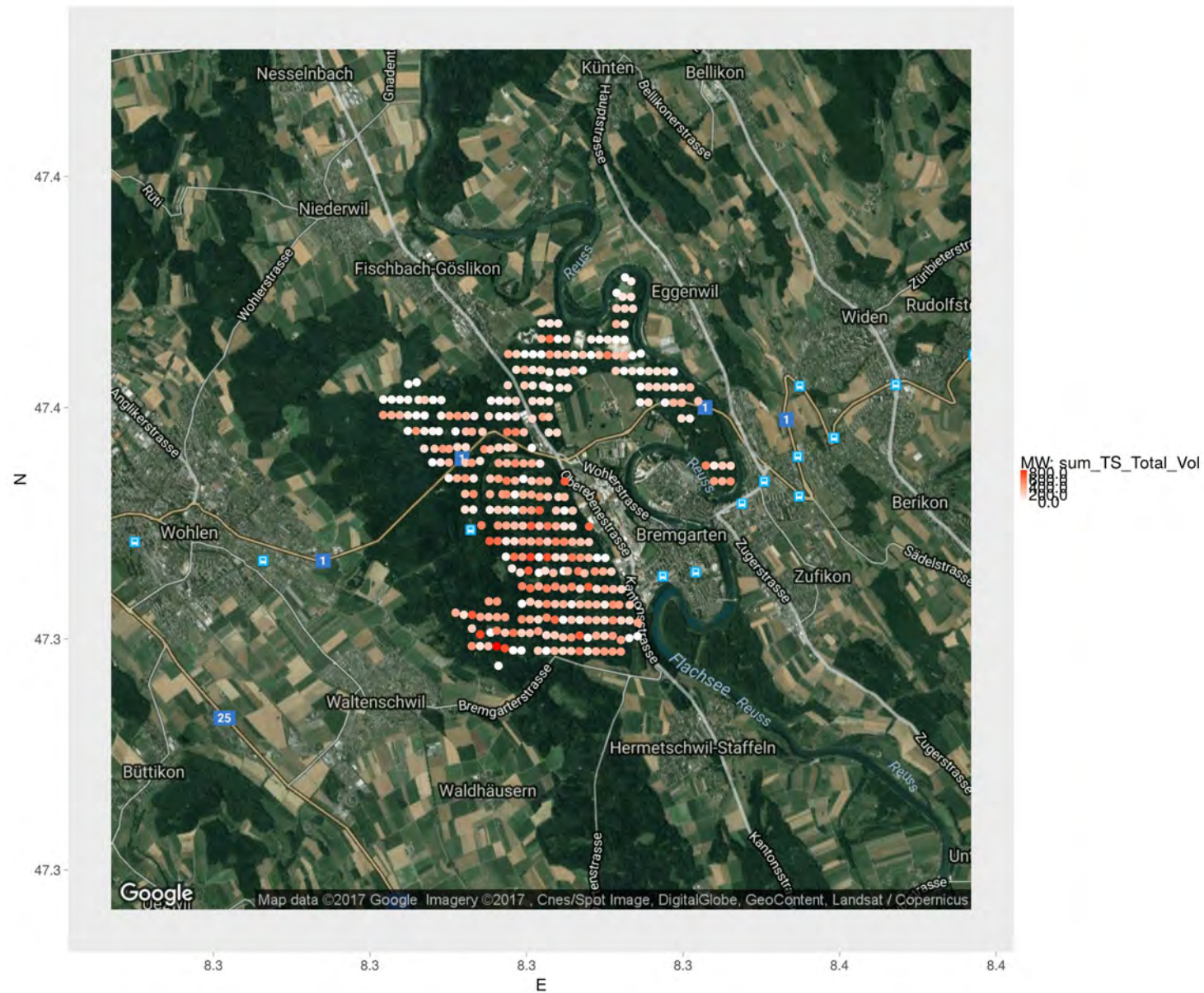


Abbildung 46: Decaden: Volumen nach Stichprobenflächen.

#### 4.1.2 Laubholz Volumen

Tabelle 26: Beschreibung Prozent Laubholz Volumen

n	mean	sd	median	min	max	skew
317	54.31	40.64	61.00	0.00	100.00	-0.12

Tabelle 27: Beschreibung gemessenes Laubholz Volumen

n	mean	sd	median	min	max	skew
337	14.66	18.65	6.03	0.00	89.70	1.30

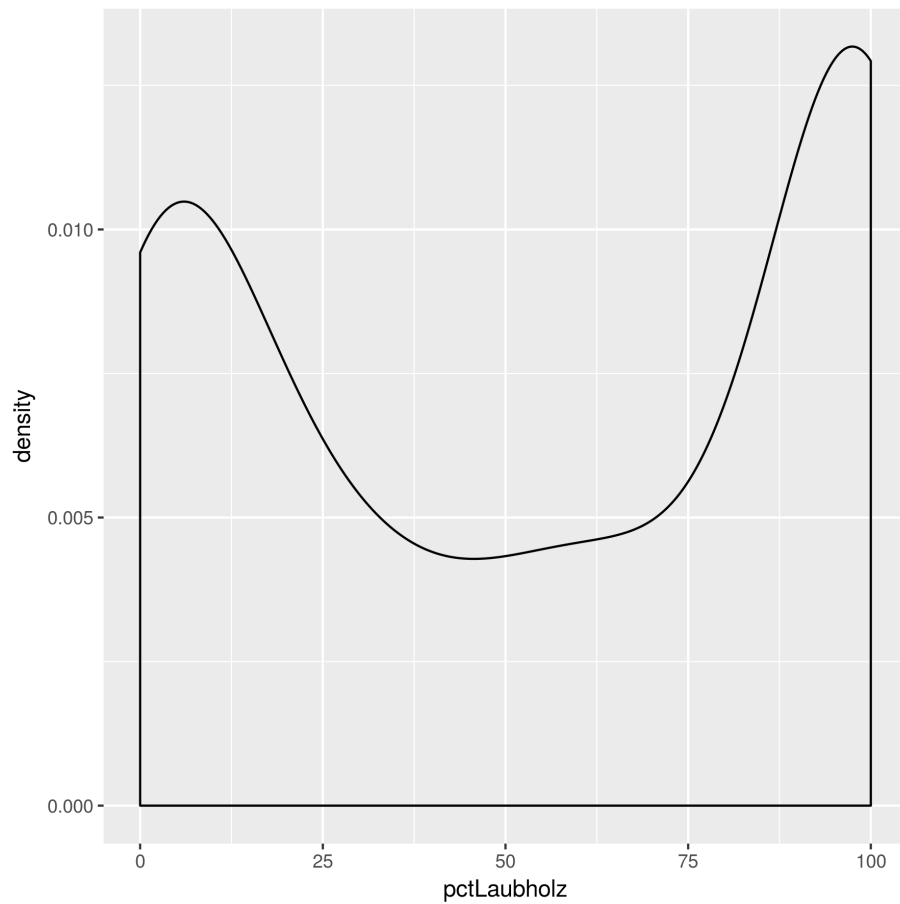


Abbildung 47: Verteilung Variable Prozent Laubholz Volumen (pctLaubholz) [%]

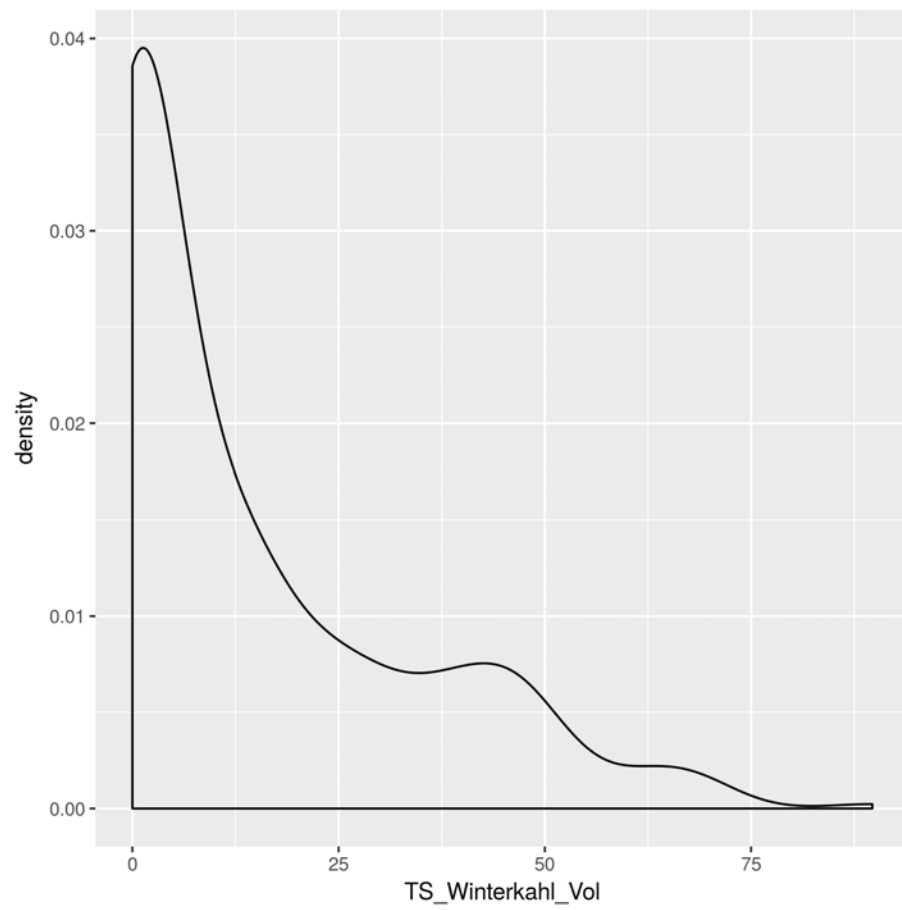


Abbildung 48: Verteilung Variable gemessenes Laubholz Volumen  
(TS\_WKahl\_Vol) [m<sup>3</sup>/ha]



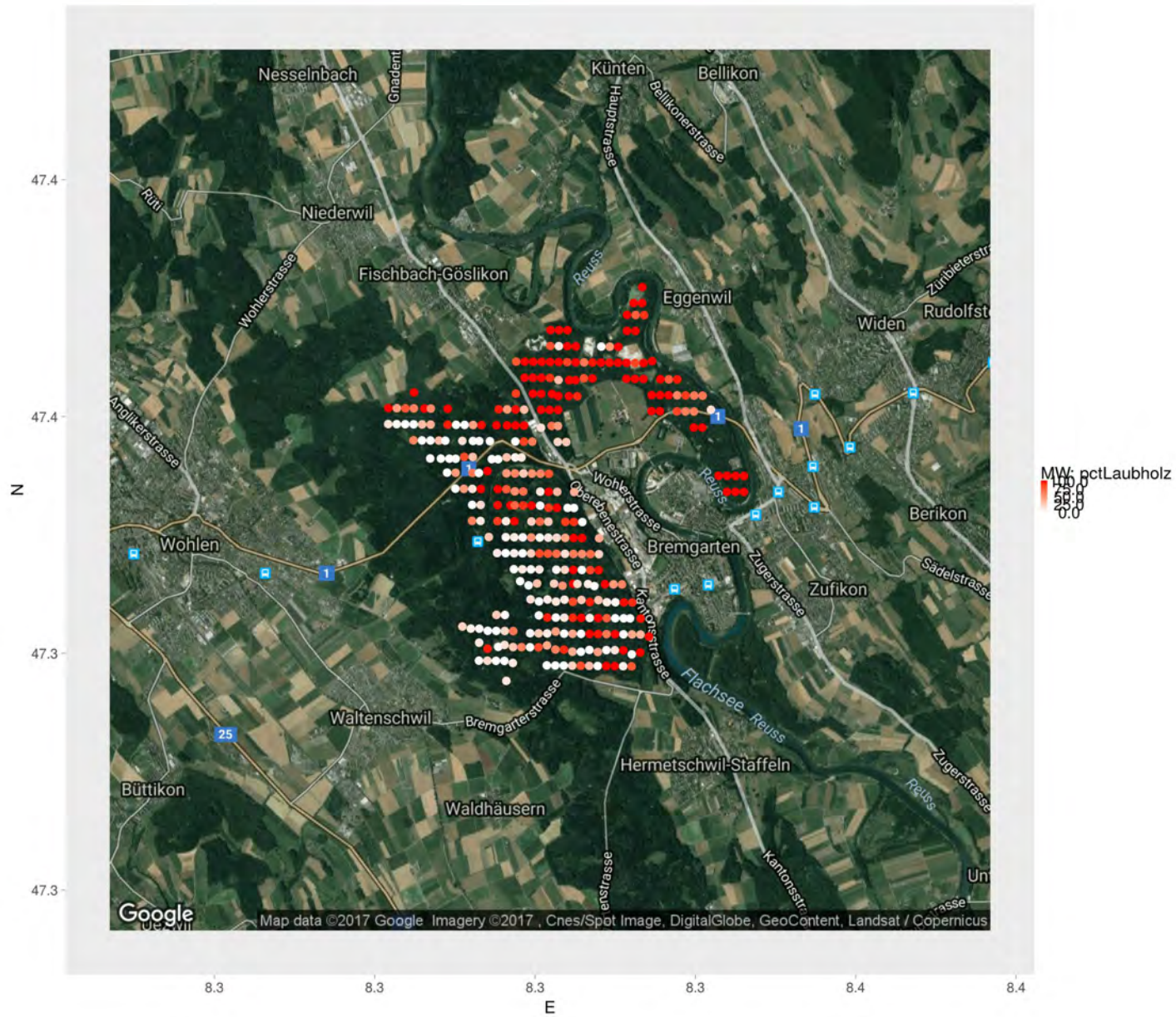


Abbildung 49: Decaden: Prozentualer Anteil Laubholz nach Stichprobeflächen.

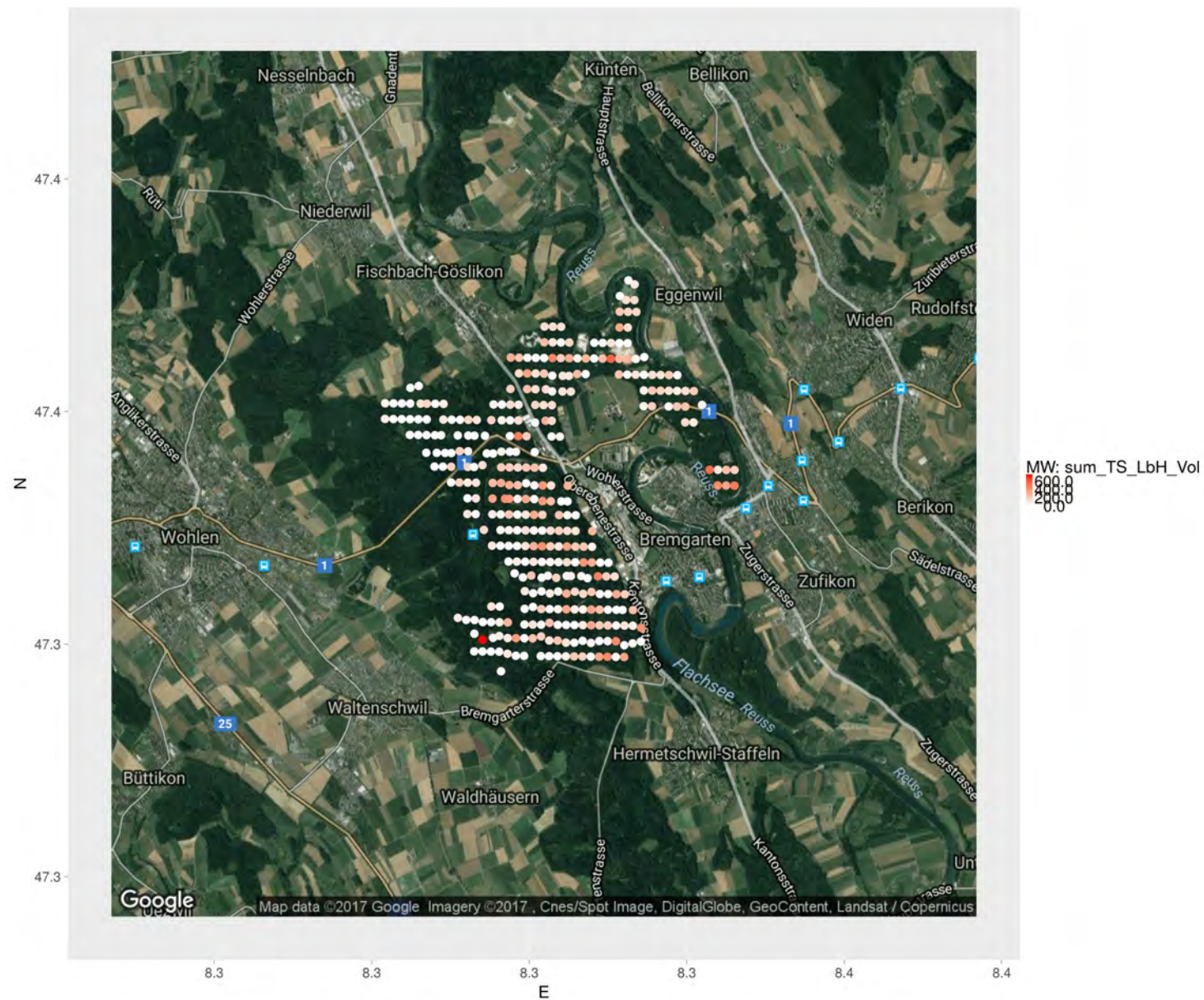


Abbildung 50: Decaden: Volumen Laubholz nach Stichprobeflächen.



### 4.1.3 Nadelholz Volumen

Tabelle 28: Beschreibung gemessenes Nadelholz Volumen

n	mean	sd	median	min	max	skew
337	11.87	13.00	7.92	0.00	75.48	1.35

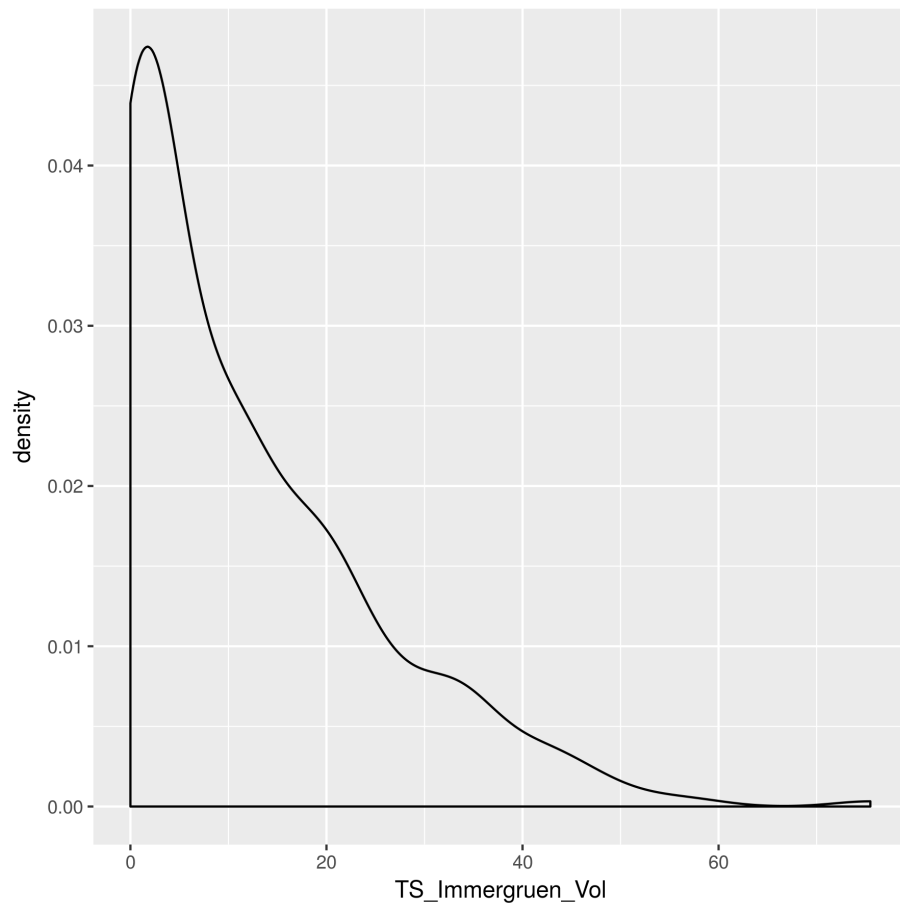


Abbildung 51: Verteilung Variable gemessenes Nadelholz Volumen (TS\_Immergruen\_Vol) [ $\text{m}^3/\text{ha}$ ]

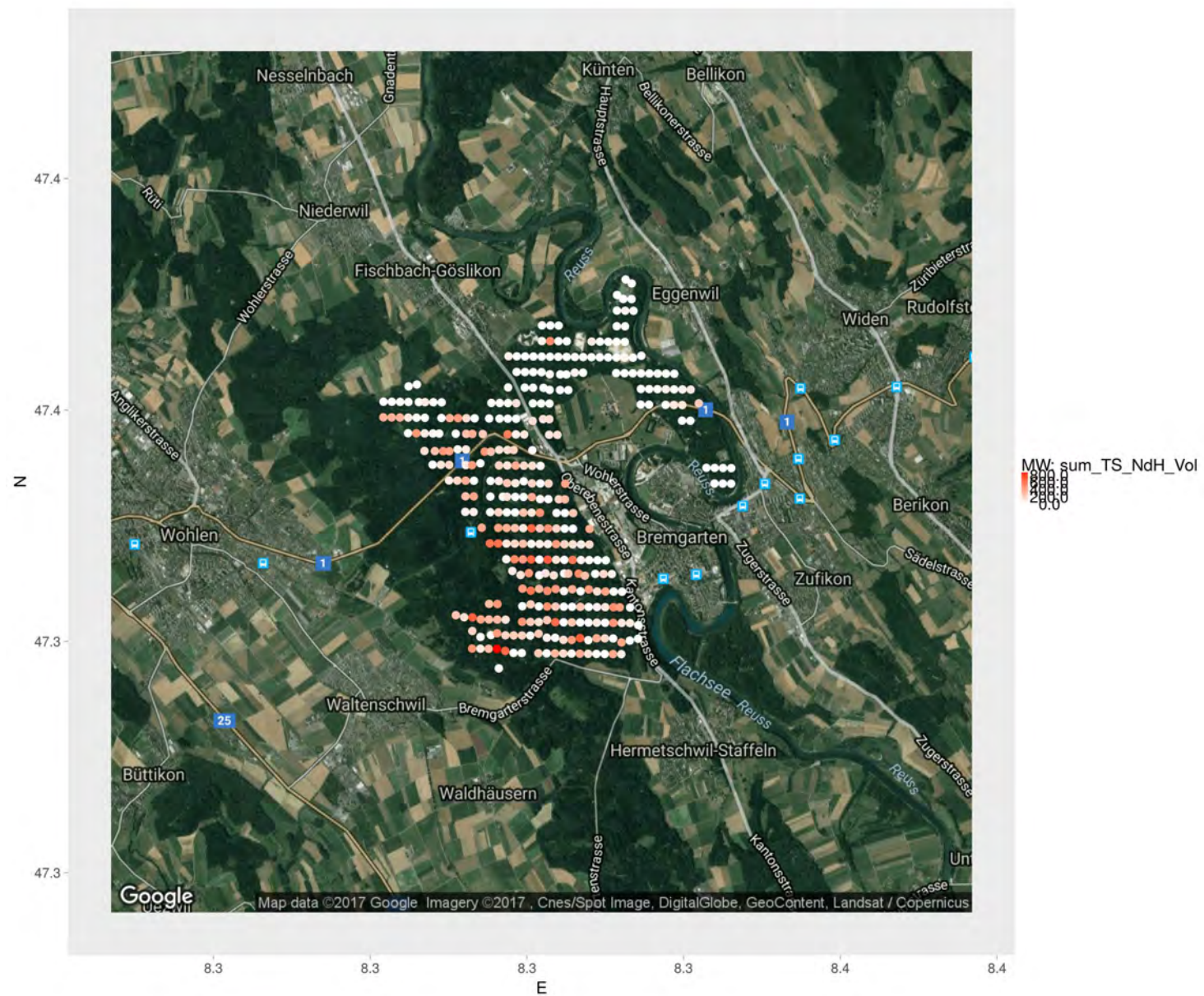


Abbildung 52: Decaden: Volumen Nadelholz nach Stichprobeflächen.

#### 4.1.4 Basalflaeche

Tabelle 29: Beschreibung gemessenes Basalflaeche pro ha

n	mean	sd	median	min	max	skew
337	211.53	138.92	202.53	0.00	619.66	0.32

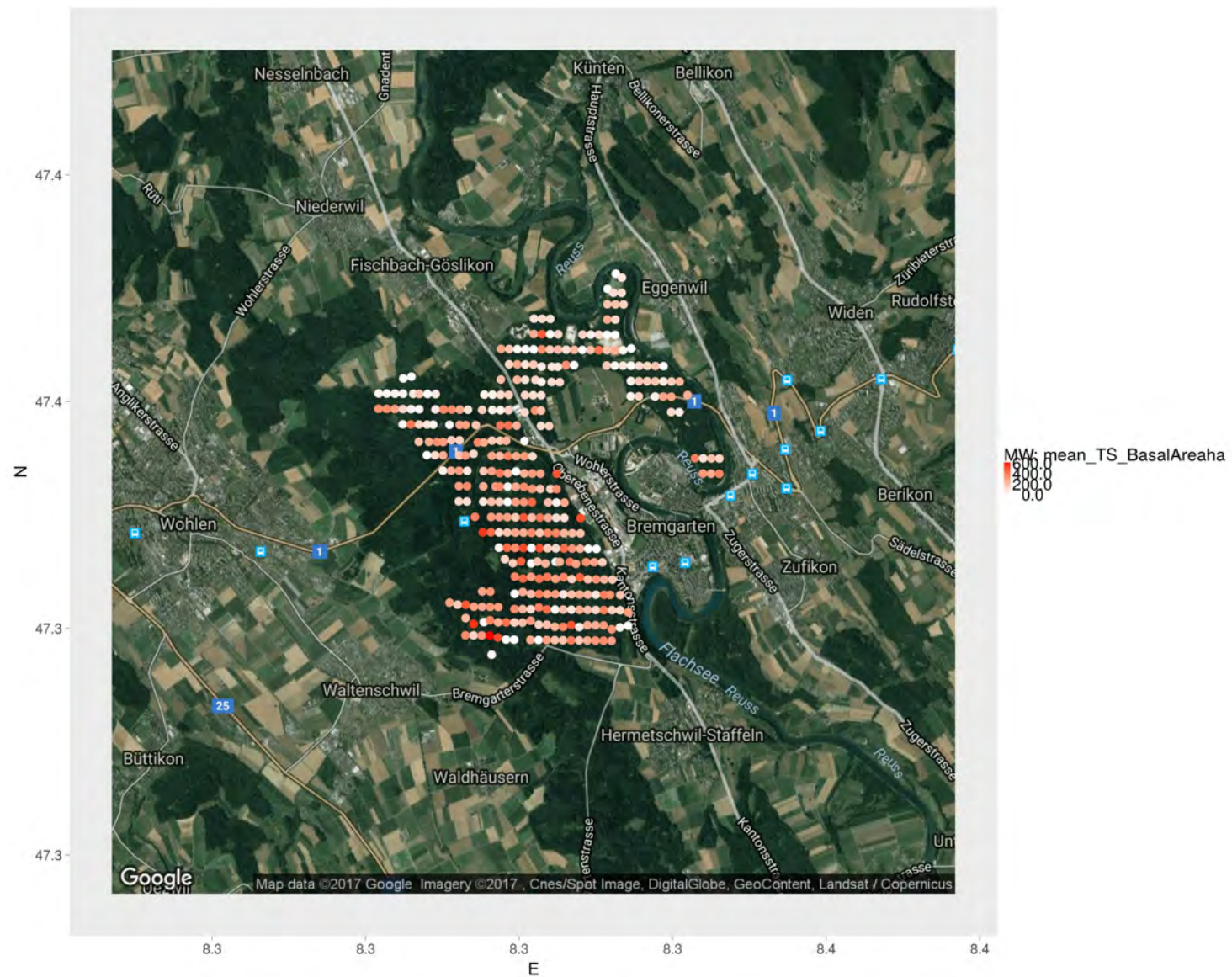


Abbildung 53: Decaden: Basalflaeche nach Stichprobeflaechen.



#### 4.1.5 LbH Anteil und MW BHD

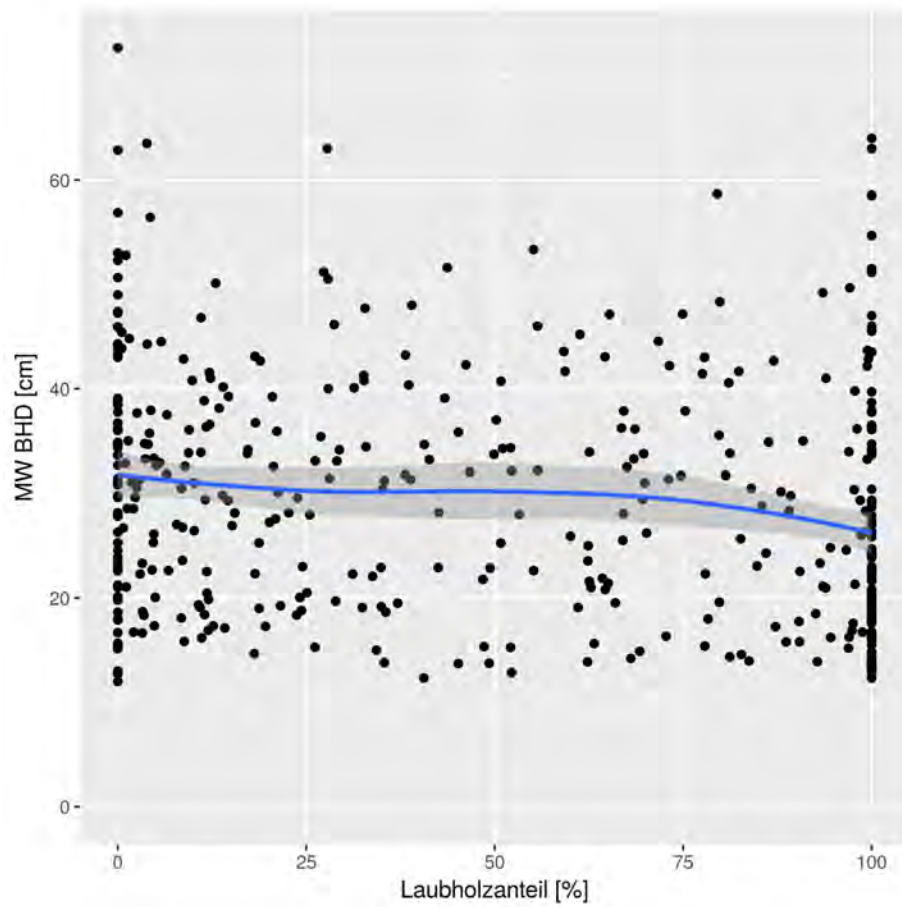


Abbildung 54: Zusammenhang Laubholz Anteil [%] und MW BHD [cm]

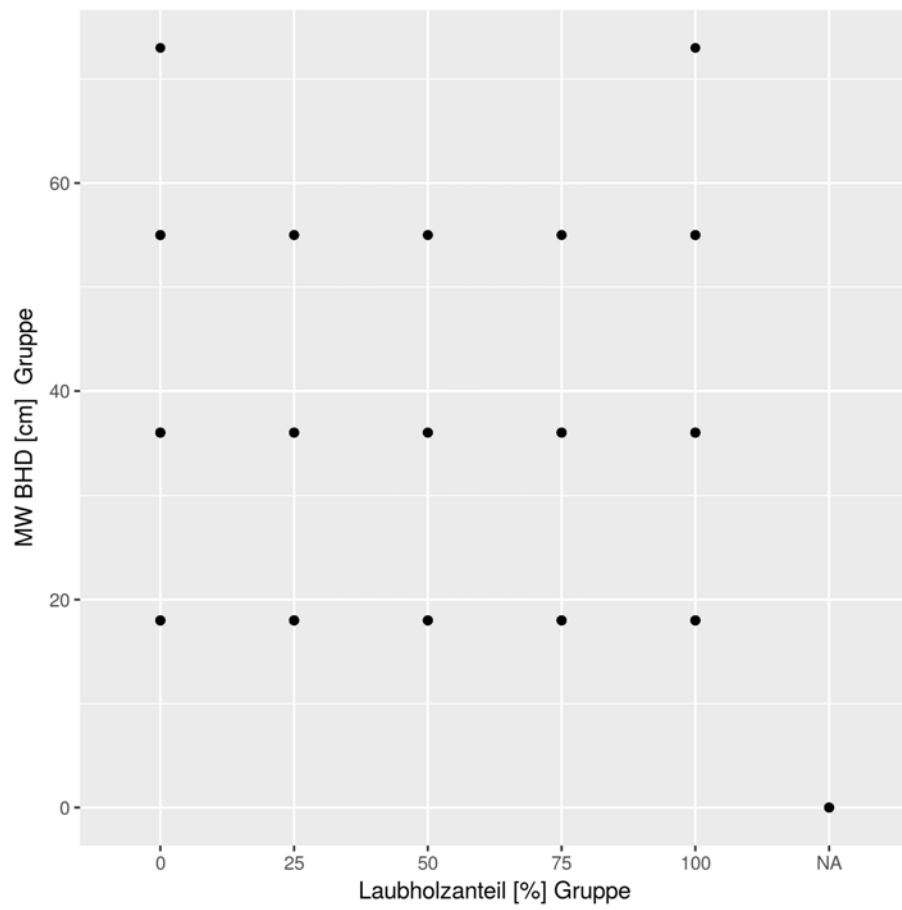


Abbildung 55: Zusammenhang Laubholz Anteil [%] GRUPPE und MW BHD [cm] GRUPPE

## 4.2 Korrelationen im Trainingsdatenset

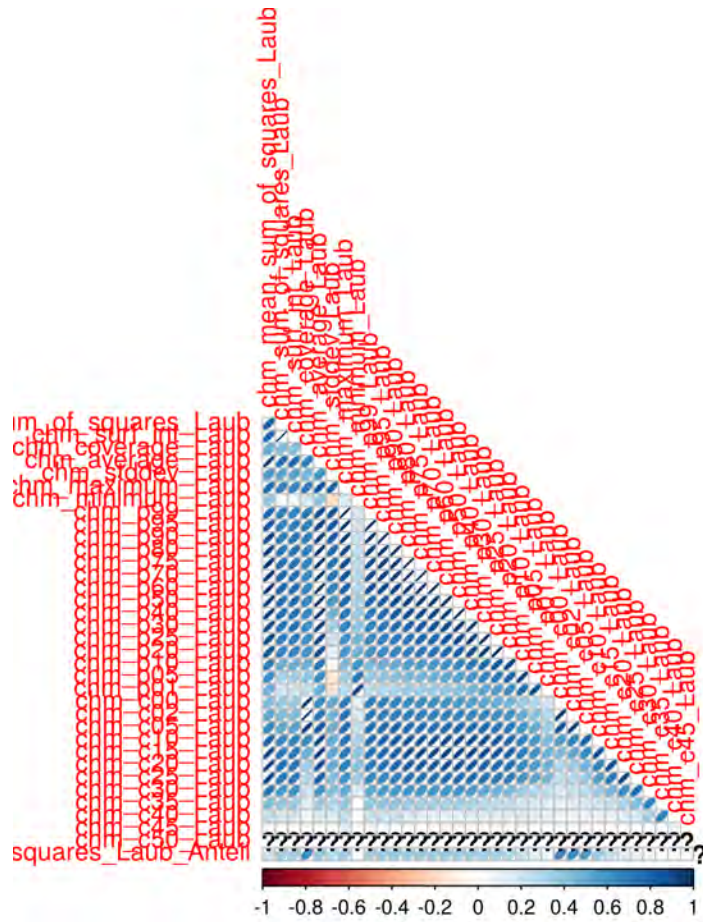


Abbildung 56: Korrelationen im Laubholzdatenset





## 4.3 Details zu den Resultaten

### 4.3.1 Code der Filter

```
function(alldata) {  
  return(alldata[alldata$TerrSample == 1, ])  
}  
<bytecode: 0xf557010>  
  
function(alldata) {  
  return(alldata[alldata$Waldrand == 0, ])  
}  
<bytecode: 0xed7aad8>  
  
function(alldata) {  
  return(alldata[!(alldata$chm_mean_sum_of_squares < filter3_chm_surf_limit & alldata$TS_Tot  
])  
}  
<bytecode: 0xdce4018>
```

Abbildung 58: Benutzte Filter

## 4.4 Dokumente

	Beschreibung	Unit	Kommentar
Unnamed: 0			
average_CHM<3m			
p90_CHM<3m			
p75_CHM<3m			
stddev_CHM<3m			
maximum_CHM<3m			
minimum_CHM<3m			
NumberOfValidValues_CHM<3m			
NumberOfNoDataValues_CHM<3m			
NumberOfValuesIncNoData_CHM<3m			
NumberOfTrees_CHM<3m			
NumberOfTreesPer_ha_CHM<3m			
average_Winterkahl			
p90_Winterkahl			
p75_Winterkahl			
stddev_Winterkahl			
maximum_Winterkahl			
minimum_Winterkahl			
NumberOfValidValues_Winterkahl			
NumberOfNoDataValues_Winterkahl			
NumberOfValuesIncNoData_Winterkahl			
NumberOfTrees_Winterkahl			
NumberOfTreesPer_ha_Winterkahl			
average_Immergruen			
p90_Immergruen			
p75_Immergruen			
stddev_Immergruen			
maximum_Immergruen			
minimum_Immergruen			
NumberOfValidValues_Immergruen			
NumberOfNoDataValues_Immergruen			
NumberOfValuesIncNoData_Immergruen			
NumberOfTrees_Immergruen			
NumberOfTreesPer_ha_Immergruen			
PLOT_ID	Probeflächen Nummer (= PBF)	[]	index
TerrSample	Terrestrische Stichprobenfläche? [0 = nein, 1 = ja]	[0, 1]	gemessener Wert
Area	Fläche des Probekreises inkl. Berücksichtigung Waldrand (aufgrund Walddaske)	[m2]	gemessener Wert
Boundary_Weights	Tatsächliche Probekreisfl. / maximal mögliche Pkfl.	[0 ... 1]	kleiner 1 wenn Waldrand
Pointer_Polygon	FID des Polygons (Polygon_of_Samples.shp)	[]	index
Waldrand	Waldrandenfluss [0 = nein, 1 = ja]	[0, 1]	1 wenn Waldrand ( + Bound_weight < 1
Small_Area	Nummer des Kleingebiets	[]	index
phase_id_2p	1 = keine terr. Aufnahme vorhanden, 2 = terr. Aufnahme vorhanden		umcodiertes TerrSample
index_ge	Generelle Index Nr. Gleiches Attribut findet sich ebenfalls in den Polygons	[]	index
X	X - Koordinate des Stichprobenzentrums (tatsächliche Werte)	[]	
Y	Y - Koordinate des Stichprobenzentrums	[]	
X_sys	X - Koordinate des Stichprobenzentrums (schematisches Raster)		
Y_sys	Y - Koordinate des Stichprobenzentrums (schematisches Raster)		
NO DATA			
TS_Total_Vol	Terrestrial Sample: Totales Volumen	dm3	gemessener Wert
TS_Immergruen_Vol	Terrestrial Sample: Immergrünes Volumen	dm3	gemessener Wert
TS_WKahl_Vol	Terrestrial Sample: Winterhahes Volumen	dm3	gemessener Wert
TS_Anteil_Immergruen	Terrestrial Sample: Anteil Immergrüner Baumarten am Volumen	0..1	abgeleiteter Wert
average		[m]	
p90		[m]	
p75		[m]	
stddev		[m]	
maximum		[m]	
minimum		[m]	
NumberOfValidValues			
NumberOfNoDataValues			
NumberOfValuesIncNoData			
NumberOfTrees	aufgrund von watershed algorithmus abgeleitet		
NumberOfTreesPer_ha	aufgrund von watershed algorithmus abgeleitet		
ind_pgon	Polygon Index		index wert
las_min_x	Koordinate		
las_min_y	Koordinate		
las_max_x	Koordinate		
las_max_y	Koordinate		
las_p01	percentil	[m]	gemessen aus LIDAR Rohdaten
las_p05	percentil	[m]	gemessen aus LIDAR Rohdaten
las_p10	percentil	[m]	gemessen aus LIDAR Rohdaten
las_p25	percentil	[m]	gemessen aus LIDAR Rohdaten
las_p50	percentil	[m]	gemessen aus LIDAR Rohdaten
las_p75	percentil	[m]	gemessen aus LIDAR Rohdaten
las_p90	percentil	[m]	gemessen aus LIDAR Rohdaten
las_p95	percentil	[m]	gemessen aus LIDAR Rohdaten
las_p99	percentil	[m]	gemessen aus LIDAR Rohdaten
las_b10	bincentiles		
las_b20	bincentiles		
las_b30	bincentiles		
las_b40	bincentiles		
las_b50	bincentiles		
las_b60	bincentiles		
las_b70	bincentiles		
las_b80	bincentiles		
las_b90	bincentiles		
las_int_min	intensity , wohl eher ohne nutzen		

mehr der LAS Tools, Dokumentation siehe LAS Tools

	las_int_max	Intensity, wohl eher ohne nutzen
	las_int_avg	Intensity, wohl eher ohne nutzen
	las_int_qav	Intensity, wohl eher ohne nutzen
	las_int_std	Intensity, wohl eher ohne nutzen
	las_int_ske	Intensity, wohl eher ohne nutzen
	las_int_kur	Intensity, wohl eher ohne nutzen
	las_cov	Canopy Cover
	las_dns	Canopy Density



Version: 2017\_Schlussbericht\_Betriebsinventur