



FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Projektarbeit

zum Thema:

Bestimmung von Biomassen mit linearen Modellen Projekt 3.2 - Buche II

vorgelegt an der Friedrich-Schiller-Universität Jena

am

28. September 2023

Betreuer:

Dr. Jens Schumacher

Bearbeitet von:
Matrikelnummer:

Eric Günl
199707

Sara Kuhnert
182707

Inhaltsverzeichnis

Abbildungsverzeichnis	iii
Tabellenverzeichnis	iv
Zusammenfassung	1
1. Einführung	2
2. Material und Methoden	3
2.1. Datensatz	3
2.2. Statistisches Modell	3
2.3. Schätzverfahren	4
2.4. Erstellung der Kandidatenmodelle	5
2.5. Modellwahlverfahren	8
2.6. Simulationsstudie	8
3. Ergebnisse	9
4. Diskussion	13
Literaturverzeichnis	14
A. Appendix	A-1
Selbstständigkeitserklärung	

Abbildungsverzeichnis

2.1.	Zusammenhang zwischen den einzelnen Einflussgrößen und der Biomasse für die einzelnen Autoren.	5
2.2.	Logarithmisch transformierter Zusammenhang zwischen Brusthöhendurchmesser und Biomasse.	6
2.3.	Plot der Residuen zur Beurteilung der Modelle.	7
2.4.	Beispielhafte Untersuchung der Interaktion zwischen age und height. . . .	7
3.1.	Auswertung der Kreuzvalidierung der Modelle für die Daten der einzelnen Autoren.	9
3.2.	Auswertung der Kreuzvalidierung der Modelle für die Daten aller Autoren.	10
3.3.	Auswertung der Simulationsstudie. Prozentuale Modellauswahlhäufigkeit in Abhängigkeit des Stichprobenumfangs für jedes Modell.	10
3.4.	Relative Auswahlhäufigkeit der Prädiktoren in Abhängigkeit des Stichprobenumfangs.	11
A.1.	A-1
A.2.	A-2

Tabellenverzeichnis

2.1. Erste Zeilen des verendeten Datensatzes.	3
3.1. Counter der Einflussparameter bei Stichprobenumfang 200.	11

Zusammenfassung

Dieses Paper beschäftigt sich mit dem von Th. Wutzler zusammengefassten Datensatz über Buchen (*Fagus Sylvatica*). Dabei soll ein geeignetes lineares Modell bestimmt werden, mit dem die Biomasse anhand spezifischer Einflussgrößen geschätzt wird. Der Datensatz umfasst die Daten von drei Autoren, welche an unterschiedlichen Standorten gemessen wurden. Die Ergebnisse zeigen, dass die logarithmische Transformation der Ziel- und Einflussgrößen die Varianzheterogenität verringerte und lineare Abhängigkeiten ermöglichte. Bei der Modellauswahl zeigte sich, dass das Hinzufügen von Interaktionen und zahlreichen Prädiktoren nach der Transformation oft zu "overfitting" führte. Zudem wurde festgestellt, dass Höhe und Brusthöhendurchmesser wesentliche Einflussgrößen für die Biomasseschätzung sind, während das Alter weniger relevant ist. Der Einfluss der Autoren auf die Ergebnisse ist signifikant, aufgrund der unterschiedlichen Standortbedingungen.

1. Einführung

In einem von Th. Wutzler (MPI Biochemie Jena) zusammengetragenen Datensatz sind Untersuchungen zur oberirdischen Biomasse von Buchen aus drei Studien zusammengefasst. Diese Daten wurden von verschiedenen Autoren an unterschiedlichen Standorten in Mitteleuropa aufgenommen und enthalten Informationen zu Eigenschaften der Bäume, welche deren Biomasse beeinflussen könnten. Von besonderem Interesse sind bei diesem Datensatz Alter, Höhe und Brusthöhendurchmesser der Buchen. Ziel dieser Arbeit ist die Herleitung einer Biomassefunktion für die Buche aus diesen Daten mit besonderem Fokus auf die Abhängigkeit vom Autor. Dazu sollen im ersten Teil die Teildatensätze unabhängig voneinander untersucht und anschließend ein gemeinsames lineares Modell mit dem Autor als kategorieller Einflussgröße erstellt werden. Zur Auswahl des besten Modells wird die k-fache Kreuzvalidierung genutzt. Anschließend an die Datenanalyse soll dieses Verfahren in einer Simulationsstudie genauer untersucht werden.

2. Material und Methoden

2.1. Datensatz

Der für diese Projektarbeit genutzte Datensatz besteht aus drei von Th. Wutzler zusammengetragenen Studien zur Biomasse von Buchen. Der Autor der jeweiligen Studie ist in der Variable *autor* gespeichert. Als weitere Einflussgrößen enthält der Datensatz die folgenden Variablen:

- *age* - Alter des Baumes
- *dbh* - Bruthöhendurchmesser (diameter at breast height), welcher als der Durchmesser des Baumstammes auf einer Höhe von 1,3 m definiert ist [3].
- *height* - Höhe des Baumes Die Zielgröße ist die gesamte oberirdische Biomasse des Baumes, welche in der Variable *biom* gespeichert ist.

Die Zielgröße ist die gesamte oberirdische Biomasse des Baumes, welche in der Variable *biom* gespeichert ist.

Die Variable *hsl* gibt die Höhe über dem Meeresspiegel an und wird im folgenden vernachlässigt, da dieser Wert sehr eng mit dem Autor verknüpft ist und kaum zusätzliche Information liefert.

In Tabelle 2.1 ist ein Auszug aus dem Datensatz mit allen enthaltenen Variablen dargestellt.

Tabelle 2.1.: Erste Zeilen des verwendeten Datensatzes.

author	hsl	age	dbh	height	biom
Bartelink	23	20	8.4	9.80	22.7
Bartelink	23	20	9.9	11.25	29.2
Bartelink	23	20	10.7	9.70	38.7
Bartelink	23	21	10.6	9.75	31.3
Bartelink	23	21	10.7	9.30	32.0
Bartelink	23	21	7.9	8.40	11.3

2.2. Statistisches Modell

Die klassische lineare Regressionsanalyse wird verwendet, um den unbekannten Zusammenhang zwischen Ziel- und Einflussgrößen zu modellieren. Sie besteht aus einem deterministischen und einem stochastischen Teil [1].

Im deterministischen Teil werden die normalverteilten Zielgrößen als Linearkombinationen der k unabhängigen Eingangsvariablen (x_1, \dots, x_k) mit festen, aber unbekannten Regressionsparametern $(\beta_0, \beta_1, \dots, \beta_k)$ dargestellt:

$$\mathbb{E}[Y_i] = \mu_i = \beta_0 1 + \beta_1 x_{i,1} + \dots + \beta_k x_{i,k}, \quad i = 1, \dots, n \quad (2.1)$$

Im Folgenden wird die Gleichung 2.1 durch $\underline{\beta} \underline{x}_i$ in Vektorschreibweise dargestellt. Wobei $\underline{\beta}$ die Regressionsparameter beschreibt und \underline{x}_i für die Prädiktoren steht.

Im stochastischen Teil werden die Beobachtungen der Zielgrößen $(Y_i, i = 1, \dots, n)$ als Realisierungen von Zufallsvariablen betrachtet, deren Erwartungswert $\mathbb{E}[Y_i] = \mu_i$ durch den deterministischen Teil des Modells angegeben wird. Für normalverteilte Zufallsgrößen gilt

$$Y_i \sim \mathcal{N}(\mu_i, \sigma^2), \quad i = 1, \dots, n. \quad (2.2)$$

Mithilfe der Daten werden Schätzungen für die linearen Regressionsparameter β_i erzeugt, mit dem Ziel, die Prognosen möglichst gut an die Beobachtungen anzunähern.

2.3. Schätzverfahren

Die Parameter β_i werden mithilfe der Maximum-Likelihood-Methode abgeschätzt. Diese maximiert die Wahrscheinlichkeit der Realisierung aus den Erwartungswerten μ_i [1]. Der Schätzer ist gegeben durch

$$\beta = \arg \max L(\beta) \quad (2.3)$$

mit der Likelihood-Funktion

$$L(\beta) = \prod_i \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{y_i - \mu_i}{\sigma} \right)^2} \quad (2.4)$$

Die Regressionsparameter sind darin, wie in Gleichung 2.1 zu sehen, in μ_i enthalten [1].

2.4. Erstellung der Kandidatenmodelle

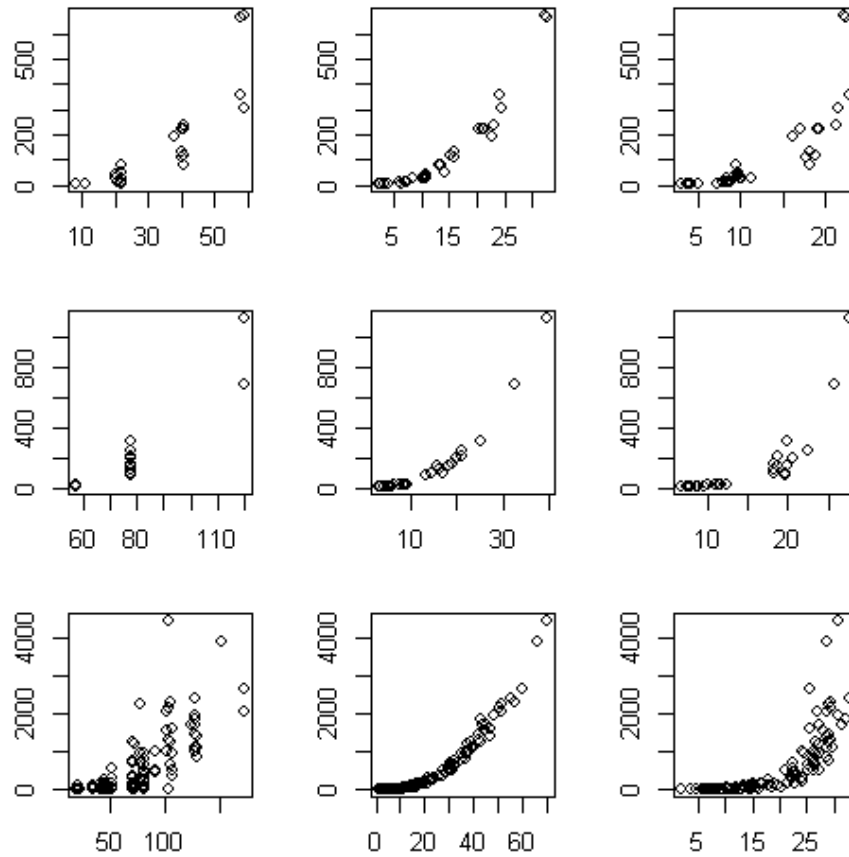


Abbildung 2.1.: Zusammenhang zwischen den einzelnen Einflussgrößen und der Biomasse für die einzelnen Autoren.

Eine Zeile bezieht sich hier immer auf den Datensatz von einem Autor (von oben nach unten: Bartelink, Heller, Joosten) und jede Spalte zeigt den Zusammenhang für eine Variable (von links nach rechts: age, dbh, height).

Wie Abbildung 2.1 vermuten lässt, besteht ein exponentieller Zusammenhang zwischen den Einflussgrößen age, dbh, height und der Zielgröße biom. Um geeignete lineare Modelle für die Biomassefunktion zu erstellen, müssen die Daten logarithmiert werden [2]. Die Kandidatenmodelle haben dann allgemein die Form:

$$\log(\text{biom}) \sim 1 + \log(x_{j,i}) + \dots \quad i = 1, \dots, n \quad (2.5)$$

Damit ergibt sich, wie in Abb. 2.2 zu sehen, ein annähernd linearer Zusammenhang. Bei genauerer Betrachtung zeigt sich allerdings, dass es bei allen Einflussvariablen Nichtlinearitäten gibt. Außerdem liegt die Vermutung nahe, dass auch Interaktionen zwischen den einzelnen Einflussgrößen existieren und diese auch die Biomasse beeinflussen. Im folgenden Abschnitt wird beschrieben, wie die Auswirkungen dieser Zusammenhänge genauer untersucht wurden.

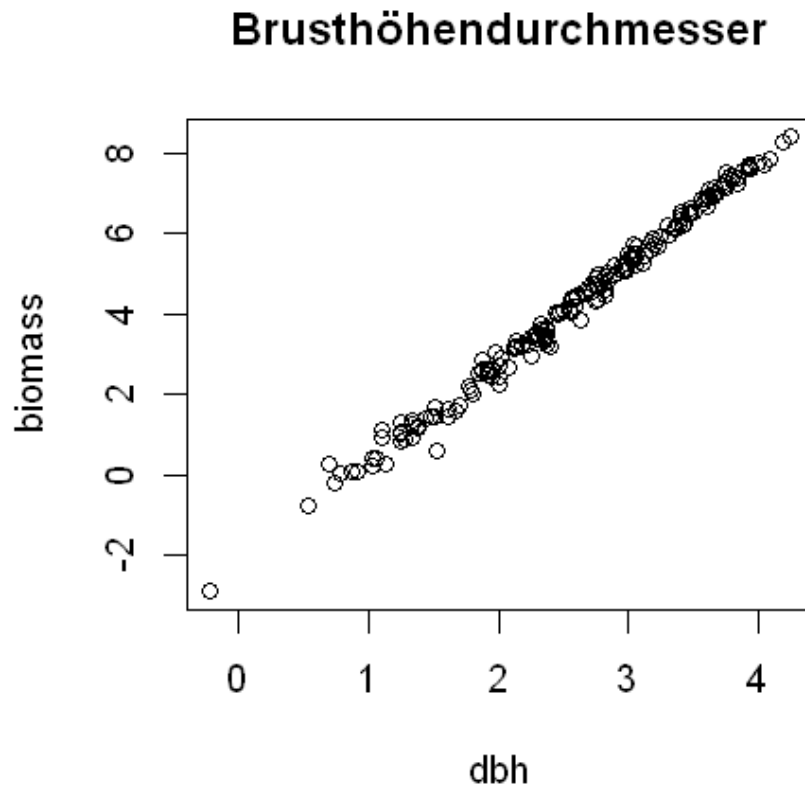


Abbildung 2.2.: Logarithmisch transformierter Zusammenhang zwischen Brusthöhendurchmesser und Biomasse.

Hier wird beispielhaft der auf den ersten Blick lineare Zusammenhang nach der Transformation sichtbar. Ein ähnlicher Verlauf ergibt sich auch für alle anderen Einflussgrößen.

Im ersten Teil werden die Datensätze der Autoren unabhängig voneinander betrachtet. Danach wird der Vorgang für die gesamten Daten zusammen wiederholt. Dadurch kommt der kategorielle Einfluss durch den Autor hinzu.

Als erstes wurde, um einen Überblick zu bekommen, jede Einflussgröße einzeln untersucht. Dafür werden einfache Modelle erstellt, die neben dem linearen Zusammenhang auch höhere Potenzen einbeziehen. Dabei wurde für jeden Autor und jede Einflussvariable das Modell gesucht, welches den Zusammenhang am besten beschreibt. Zur Beurteilung wurden die Residuen geplottet, wie beispielsweise in Abb. 2.3.

Hier ist das Vorgehen am Beispiel des Brusthöhendurchmessers aus dem Datensatz von Bartelink dargestellt. Links wurde nur der lineare Term genutzt. Deutlich wird dabei eine klare Abweichung, welche einen quadratischen Zusammenhang vermuten lässt. Dieser wurde im mittleren Bild hinzugefügt. Trotz deutlicher Verbesserung ist die Abweichung noch zu groß und kann durch das Hinzufügen von einem kubischen Einfluss weiter ausgeglichen werden. Noch höhere Potenzen verbessern das Modell nicht weiter.

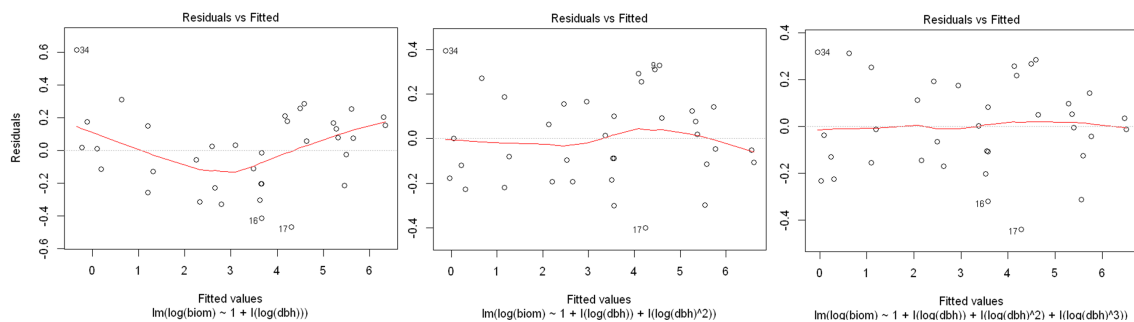


Abbildung 2.3.: Plot der Residuen zur Beurteilung der Modelle.

Anschließend wurden alle drei möglichen Interaktionen (age-height, height-dbh, dbh-age) betrachtet. Dazu wurde jeweils eine Größe in drei Bereiche geteilt und dann geschaut, wie sich die zweite Größe in den jeweiligen Bereichen verhält. Gibt es starke Veränderungen des Anstiegs, liegt eine deutliche Interaktion zwischen beiden Einflussgrößen nahe. In Abb. 2.4 sind die drei verschiedenen Bereiche dargestellt, in die das Alter geteilt wurde. Die unterschiedlichen Abhängigkeiten der Biomasse von der Höhe sind in den jeweiligen Feldern zu sehen.

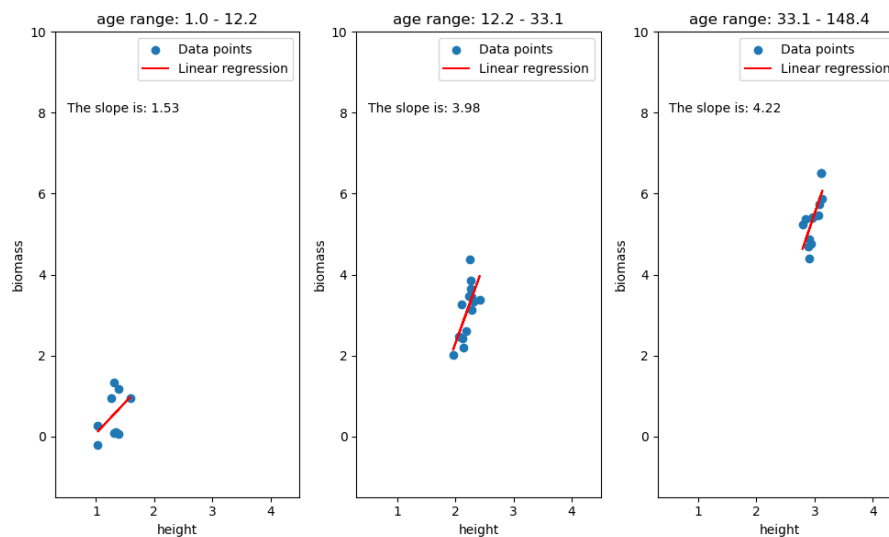


Abbildung 2.4.: Beispielhafte Untersuchung der Interaktion zwischen age und height.

Zum Schluss wurden jeweils verschiedene Modelle erstellt, welche unterschiedliche Nicht-linearitäten, beispielsweise eine quadratische Abhängigkeit, und Interaktionen kombinieren. Welche Prädiktoren dabei in die Modelle einbezogen wurden, hängt von den vorherigen Untersuchungen ab und unterscheidet sich für jeden Autor. Für das Auswahlverfahren wurden als erstes sehr einfache Modelle erstellt, welche dann zunehmend erweitert werden. Als grober Richtwert, ob eine Variable in ein Modell aufgenommen wird, kam der p -Wert $Pr(> |t|)$ zum Einsatz. Hierbei wurde der p -Wert für einzelne Prädiktoren in Modellen mit steigender Komplexität (höherer Potenz) verglichen um somit eine Aussage für die Relevanz des Prädiktors zu treffen.

Anschließend wurde der gesamte Vorgang für alle kombinierten Datensätze wiederholt.

Um den Einfluss des Autors zu untersuchen, wurden Modelle erstellt, welche diesen als zusätzliche Variable und kategoriale Einflussgröße einbeziehen. Zu Referenzzwecken gibt es auch hier ein Modell (Nummer 8) ohne diesen Einfluss.

2.5. Modellwahlverfahren

Nachdem sämtliche Kandidatenmodelle erstellt wurden, muss das beste ausgewählt werden. In dieser Arbeit wurde dafür die k -fache Kreuzvalidierung gewählt. Die vorhandenen Messwerte werden dafür in k Gruppen eingeteilt. k ergibt sich aus dem Minimum von 10 und der Wurzel der Anzahl der Beobachtungen. $k - 1$ Gruppen werden anschließend als Trainingsdaten genutzt. Mit diesen werden die Parameter bestimmt. Die Letzte Gruppe sind die Testdaten. Mit diesen wird die Abweichung durch die berechneten Parameter ermittelt. Nacheinander wird jede der k Gruppen einmal als Testgruppe genutzt und über die gesamte Abweichung summiert. Das Modell mit dem geringsten Gesamtfehler wird als das beste gesehen.

2.6. Simulationsstudie

Als Grundlage für die Simulationsstudie dient der Datensatz von Joosten. Mit dem besten Modell aus den vorherigen Betrachtungen (bei Joosten Modell 7) und der Datenmatrix, wurden neue Beobachtungswerte generiert. Auf diese Werte wird ein normalverteiltes Rauschen mit entsprechender Varianz gelegt. Mit diesen generierten Pseudodaten werden die Regressionsparameter aller Modelle berechnet. Anschließend wird die k -fache Kreuzvalidierung durchgeführt und bestimmt, wie oft das „Originalmodell“ als das beste ausgewählt wurde. Um bessere Aussagen zum Verhalten der Kreuzvalidierung treffen zu können, wurde der Stichprobenumfang variiert und jede Simulation 10.000 Mal wiederholt.

3. Ergebnisse

Die Tabelle A.2 zeigt eine Übersicht der gewählten Prädiktoren für jeden Teildatensatz der Autoren. Beispielhaft würde das Modell M6 des Autors Joosten folgendermaßen aussehen:

$$\mathbb{E}[Y_i] = \beta_0 + \beta_1 age + \beta_2 height + \beta_3 (age \times height) + \beta_4 (dbh \times age) \quad (3.1)$$

Im Folgenden sind die Ergebnisse der Kreuzvalidierung für alle Modelle der jeweiligen Autoren zusehen.

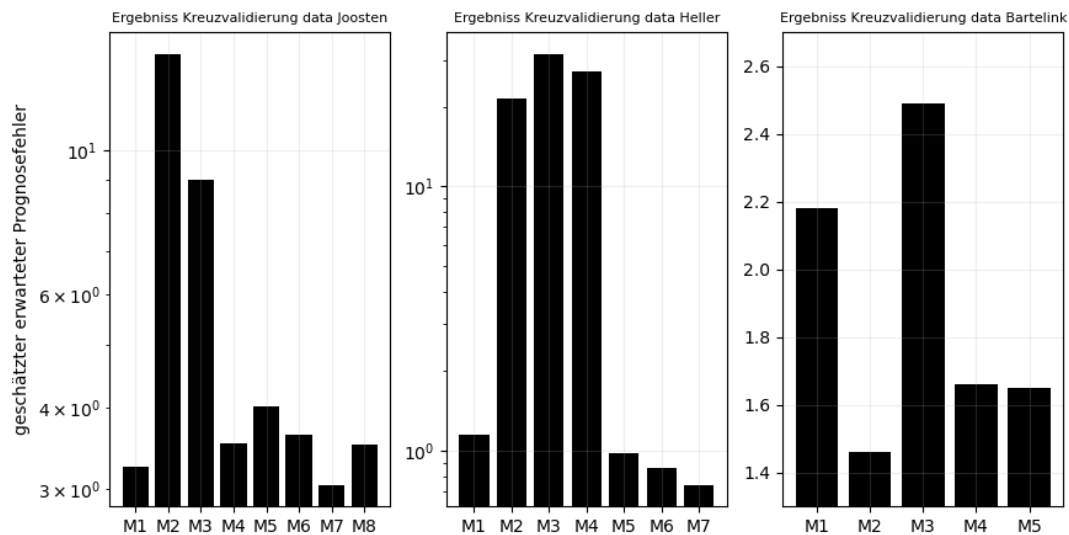


Abbildung 3.1.: Auswertung der Kreuzvalidierung der Modelle für die Daten der einzelnen Autoren.

Zu beachten ist, dass die Achseneinteilungen bei den in Abbildung 3.1 gezeigten Balkendiagrammen unterschiedlich sind, um die Unterschiede der einzelnen Modelle bestmöglich zu visualisieren. Die quantitativen Werte des geschätzten erwarteten Prognosefehlers zwischen den Modellen unterschiedlicher Autoren kann nicht verglichen werden, da die Anzahl der Datenpunkte bei jedem Autor unterschiedlich ist. Den kleinsten geschätzten erwarteten Prognosefehler haben die Modelle 7 der Autoren Joosten und Heller (beachte: die Modelle haben dennoch unterschiedliche Einflussgrößen). Im Fall der Daten von Autor Bartelink hat das Modell 2 den geringsten geschätzten erwarteten Prognosefehler.

Für die Betrachtung der Modelle, die für die Daten aller Autoren erstellt wurden, ergibt sich folgendes Ergebnis.

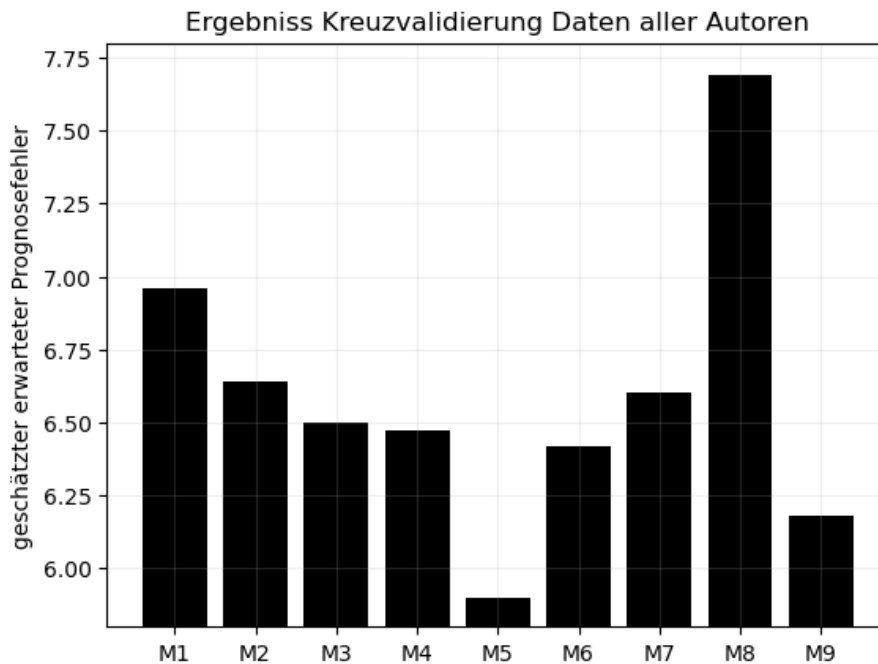


Abbildung 3.2.: Auswertung der Kreuzvalidierung der Modelle für die Daten aller Autoren.

Das Modell 5 hat den geringsten geschätzten erwarteten Prognosefehler. Das Modell 8 ist ein Referenzmodell ohne Betrachtung der Autoren als Einflussgröße und hat den höchsten geschätzten erwarteten Prognosefehler. Die Prädiktoren für jedes Modell sind in Abbildung A.1 aufgeführt.

Die Ergebnisse der Simulationsstudie können in 2 Abbildungen aufgezeigt werden.

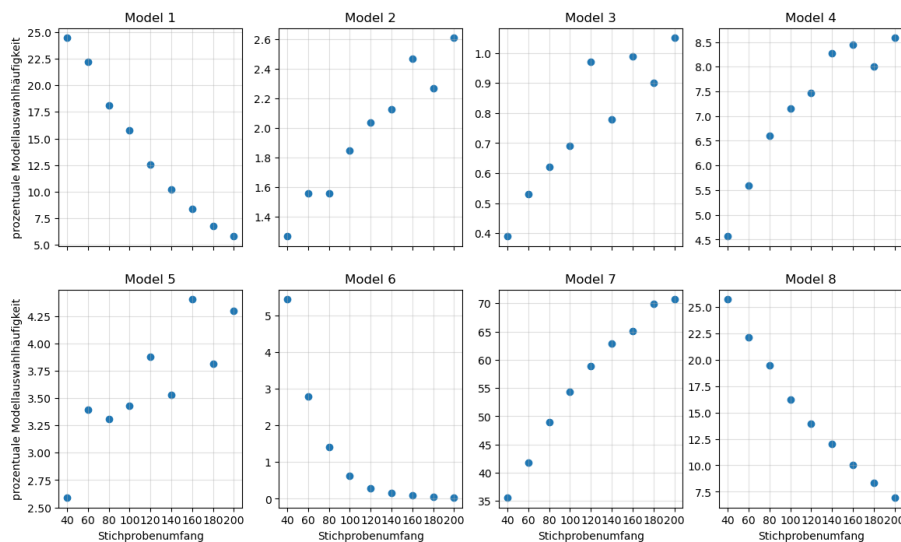


Abbildung 3.3.: Auswertung der Simulationsstudie. Prozentuale Modellauswahlhäufigkeit in Abhängigkeit des Stichprobenumfangs für jedes Modell.

Abbildung 3.3 zeigt die Auswahlhäufigkeit in Prozent für jedes Modell. Es stellt dar wie

häufig jedes Modell als bestes Modell nach der Kreuzvalidierung ausgewählt wurde, in Abhängigkeit des Stichprobenumfangs. Die Simulation wurde mit 10.000 Wiederholungen für jeden Stichprobenumfang ausgeführt. Als Simulationsmodell wurde das Modell 7 genutzt. Bei jedem einzelnen Durchlauf wurde die Zielgröße, der Varianz entsprechend, verrauscht. Die Modelle 2, 3, 4, 5 und 7 weisen eine Steigerung der Auswahlhäufigkeit mit steigendem Stichprobenumfang auf, während die anderen Modelle einen negativen Anstieg aufweisen. Insgesamt wird das Modell 7 am häufigsten über alle Stichprobenumfänge ausgewählt.

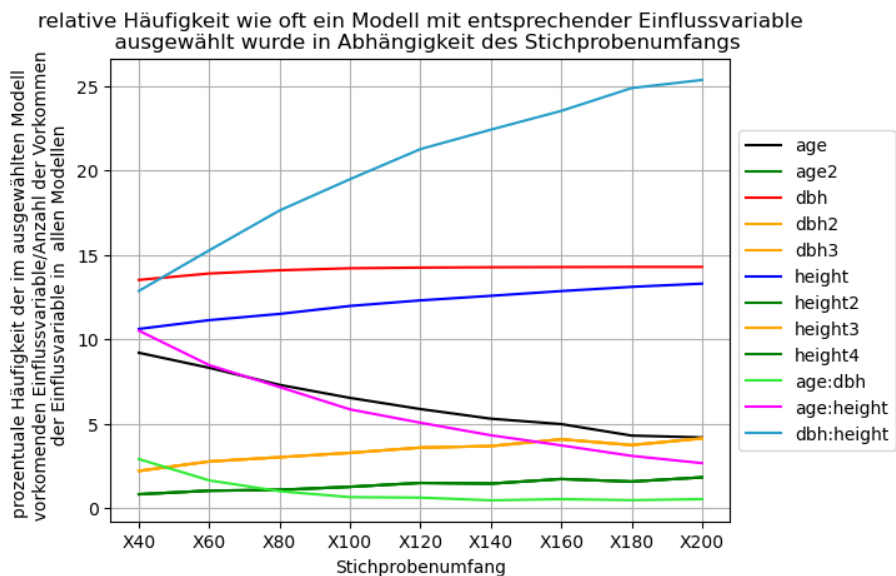


Abbildung 3.4.: Relative Auswahlhäufigkeit der Prädiktoren in Abhängigkeit des Stichprobenumfangs.

Um eine zweite Auswertungsmöglichkeit für die Simulation zu haben, wurde jedem Prädiktor ein Counter zugewiesen. Wenn ein Prädiktor in einem Modell vorkam, welches durch Kreuzvalidierung als bestes Modell ausgewählt wurde, erhöhte sich der Counter um 1. Da nicht alle Prädiktoren gleichhäufig in den Modellen vorkommen, wurde der Counter jedes Prädiktors um die Anzahl der Vorkommen von diesem in allen Modellen geteilt. Somit erhält man die relative Häufigkeit der genutzten Prädiktoren 3.4.

Tabelle 3.1.: Counter der Einflussparameter bei Stichprobenumfang 200.

	age	age2	dbh	dbh2	dbh3	height	height2	height3	height4	age:dbh	age:height	dbh:height
E1	583	NA	583	NA	NA	583	NA	NA	NA	NA	NA	NA
E2	261	261	261	261	261	261	261	261	261	NA	NA	NA
E3	105	105	105	105	105	105	105	105	105	105	105	105
E4	858	NA	858	858	858	858	NA	858	NA	NA	NA	NA
E5	430	NA	430	430	430	430	NA	430	NA	NA	NA	430
E6	3	NA	NA	NA	NA	3	NA	NA	NA	3	3	NA
E7	NA	NA	7066	NA	NA	7066	NA	NA	NA	NA	NA	7066
E8	694	NA	694	NA	NA	NA	NA	NA	NA	NA	694	NA

Die Tabelle 3.1 zeigt die Häufigkeit wie oft jedes Modell (E^*) für einen Stichprobenumfang von 200 ausgewählt wurde. Jede Spalte ist einem Prädiktor zugehörig. Summiert man eine Spalte auf erhält man die Summe wie häufig ein Prädiktor insgesamt über alle Modelle bei der Kreuzvalidierung ausgewählt wurde. Da zum Beispiel Modell 1 (E1) nur die Prädiktoren age, dbh, und height hat sind alle anderen Zellen dieser Spalte auf NA gesetzt.

4. Diskussion

Die logarithmische Transformation der Ziel- und Einflussgrößen hat die Varianzheterogenität vermindert und die Abhängigkeiten in erster Näherung linear werden lassen.

Die Ergebnisse der Auswahl (Kreuzvalidierung) der Modelle für die jeweiligen Autoren hat gezeigt, dass das Hinzufügen aller Interaktionen und vieler Prädiktoren, nach der logarithmischen Transformation der Ziel- und Einflussgrößen, meist einen negativen Einfluss auf die Ergebnisse der Kreuzvalidierung hatten, da dies zu „overfitting“ der Modelle führt.

Ebenso wurde festgestellt, dass die Einflussparameter eine starke Korrelation aufweisen und es deshalb nicht zwingend notwendig ist alle 3 Einflussparameter in das Modell aufzunehmen. Auch die Interaktionen zwischen den 3 Einflussgrößen ist nicht zu vernachlässigen. So wurde zum Beispiel festgestellt, dass es eine Interaktion zwischen dem Alter und der Höhe der Bäume gibt.

Zudem weisen die gewählten Modelle für die Datensätze der verschiedenen Autoren große Unterschiede auf. Dies ist auch gut nachzuvollziehen, wenn die Daten aller Autoren zusammen genommen werden, um ein Modell zu erstellen. Die Interaktion der quantitativen Einflussvariablen mit der Einflussvariable Autor ist nicht zu vernachlässigen. Natürlich dürfte der Autor an sich keinen Einfluss haben. Da die Autoren ihre Daten jedoch an unterschiedlichen Standort aufgenommen haben, an denen unterschiedliche Bodenverhältnisse und Klimabedingungen herrschen ist der Einfluss des „Autors“ immens.

Durch die Simulation konnte festgestellt werden, dass die Einflussgrößen Höhe und Brusthöhendurchmesser wichtiger für ein akkurates Modell sind als das Alter. Bei steigendem Stichprobenumfang erhöhte sich die Auswahl der Modelle mit den Prädiktoren Höhe und Brusthöhendurchmesser während die Auswahl von Modellen mit Alter sank. In absoluten Zahlen wurden ebenfalls Modelle häufiger gewählt, welche Höhe und Brusthöhendurchmesser miteinbezogen, als jene die alter einbezogen. Dies ergibt Sinn, da ein Baum in erster Näherung ein Zylinder ist. Die Biomasse ist somit das Volumen des Zylinders multipliziert mit der Dichte. Das Volumen des Baumes hängt direkt von der Höhe und dem Brusthöhendurchmesser ab 4.1.

$$mass = \rho \cdot V = \rho \cdot height \cdot \pi \cdot \left(\frac{dbh}{2}\right)^2 \quad (4.1)$$

Natürlich verjüngt sich der Baum zur Spitze hin, hat Äste etc. Daher ist dieses Modell nur eine grobe Näherung. Das Alter des Baumes korreliert natürlich stark mit den anderen beiden Parametern und kann deshalb auch Informationen über die Biomasse bereitstellen,

hängt jedoch nicht direkt mit der Biomasse zusammen. Je älter der Baum, desto größer und dicker ist er. Je größer und dicker er ist, umso schwerer ist er.

Abbildung 3.3 zeigt, dass das Modell 7 am häufigsten durch die Kreuzvalidierung ausgewählt wurde. Dies ist nicht unerwartet, da Modell 7 das Simulationsmodell war und auch schon mit den Originaldaten den besten geschätzten erwarteten Prognosefehler hatte. Mit steigenden Stichprobenumfang steigt die Auswahlhäufigkeit des Modells 7 an. Es kann davon ausgegangen werden, dass mit steigendem Stichprobenumfang die Ergebnisse der Kreuzvalidierung verlässlicher werden. Der Abstieg der Auswahlhäufigkeit der Modelle 1, 6, und 8 kann mit einer geringen Anzahl an Prädiktoren aber dem Vorhandensein des Prädiktors Alter erklärt werden. Somit hat das Alter einen hohen Einfluss auf diese Modelle, wobei schon festgestellt wurde, dass Höhe und Brusthöhendurchmesser verlässlichere Einflussgrößen sind.

Literaturverzeichnis

- [1] Jens Schumacher. Vorlesung - Statistische Verfahren, 2022.
- [2] Jens Schumacher. Aufgabenblatt (projekt 3.2 - buche ii). 2023.
- [3] Wikipedia. Stammdurchmesser. *de.wikipedia.org*, 8 2005.

A. Appendix

```

M1=lm(log(biom) ~ (1 + l(log(age)) + log(dbh) + l(log(height))) * as.factor(author)
M2=lm(log(biom) ~ (1 + l(log(age)) + l(log(age)^2) + log(dbh) + l(log(dbh)^2) + l(log(height)) + l(log(height)^2)) * as.factor(author)
M3=lm(log(biom) ~ (1 + l(log(age)) + l(log(age)^2) + log(dbh) + l(log(height)) + l(log(height)^2)) * as.factor(author)
M4=lm(log(biom) ~ (1 + l(log(age)) + log(dbh) + l(log(height)) + log(dbh).log(age)+log(dbh).log(height)+log(age)).log(age)) * as.factor(author)
M5=lm(log(biom) ~ (1 + log(dbh) + l(log(height)) + log(dbh).log(height)) * as.factor(author)
M6=lm(log(biom) ~ (1 + l(log(age)) + l(log(age)^2)+ log(dbh) + l(log(dbh)^2) + l(log(height)) + l(log(height)^2) + log(age).log(height)) * as.factor(author)
M7=lm(log(biom) ~ 1 + l(log(age)) + log(dbh) + log(height) + as.factor(author) + as.factor(author).log(dbh)
M8=lm(log(biom) ~ 1 + l(log(age)) + log(dbh) + l(log(height))
M9=lm(log(biom) ~ 1 + log(dbh) + l(log(height)) + log(dbh).log(height) + as.factor(author) + as.factor(author).log(dbh)

```

Abbildung A.1.

	Joosten								Heller					Bartelink						
	M1	M2	M3	M4	M5	M6	M7	M8	M1	M2	M3	M4	M5	M6	M7	M1	M2	M3	M4	M5
intersection	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
age	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓			✓	✓	✓	✓	
age2		✓	✓							✓	✓	✓								
dbh	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓
dbh2		✓	✓	✓	✓					✓	✓	✓	✓	✓			✓	✓	✓	✓
dbh3		✓	✓	✓	✓					✓	✓	✓	✓	✓			✓	✓	✓	✓
dbh4											✓	✓	✓							
height	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
height2		✓	✓							✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
height3		✓	✓	✓	✓					✓	✓	✓	✓	✓	✓					
height4		✓	✓							✓	✓	✓	✓	✓	✓					
height5			✓								✓	✓	✓	✓	✓					
age:height			✓			✓		✓			✓	✓		✓				✓	✓	
dbh:age			✓		✓	✓					✓	✓		✓				✓	✓	
dbh:height			✓		✓		✓				✓		✓	✓	✓			✓		✓

Abbildung A.2.

Selbstständigkeitserklärung

Wir erklären, dass wir die vorliegende Arbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel angefertigt haben.

Jena, den 28. September 2023



Eric Günl



Sara Kuhnert