



Berner
Fachhochschule

Berner Fachhochschule
Technik und Informatik
Institut für Optimierung und Datenanalyse



Rauheit des SBB-Schienenennetzes

Erste Modellierung der Schienenrauheit basierend auf
bisheriger Datenbasis

Schlussbericht

Im Auftrag der Schweizerischen Eidgenossenschaft

Burgdorf, Januar 2021

Impressum

Auftraggeber

Schweizerische Eidgenossenschaft; Bundesämter für Umwelt (BAFU) und Verkehr (BAV), CH-3003 Bern.

Das BAFU und das BAV sind Ämter des Eidg. Departements für Umwelt, Verkehr, Energie und Kommunikation (UVEK).

Auftragnehmer

Dr. Jasmin Wandel, Berner Fachhochschule (BFH), Departement Technik und Informatik (TI), Institut für Optimierung und Datenanalyse (IODA).

Autoren

Dr. Niki Zumbrunnen, Dr. Jasmin Wandel.

Begleitung BAFU / BAV

Franz Kuster, Fredy Fischer
Robert Attinger, Christoph Dürig

Hinweis

Dieser Bericht wurde im Auftrag der Bundesämter für Umwelt (BAFU) und Verkehr (BAV) verfasst. Für den Inhalt ist alleine der Auftragnehmer verantwortlich.

Titelbild

Messgerät und Fahrspiegel
Quelle: SBB

Inhaltsverzeichnis

1. Einleitung	1
1.1. Ausgangslage	1
1.2. Grundlage	2
1.3. Ziele der aktuellen Projektphase	2
1.4. Ziele des Auftrags	2
2. Beschreibung der vorhandenen Daten	3
2.1. Beschreibung der einzelnen Variablen	3
2.1.1. Rauheit	3
2.1.2. Tonnage	5
2.1.3. Anzahl Zugsüberfahrten	5
2.1.4. Angewandte Technik beim letzten Schleifen (Maschine)	6
2.1.5. Stahlqualität	7
2.1.6. Schienenprofil	7
2.1.7. Schwellentyp	8
2.1.8. Besohlung der Schwelle	8
2.1.9. Schienenbefestigungstyp	8
2.1.10. Schienenzwischenlage	9
2.1.11. Schotterdicke	9
2.1.12. Anteil Güterverkehr	9
2.1.13. Streckengeometrie	10
2.1.14. Weitere Variablen	10
3. Statistische Modelle	11
3.1. Modell mit allen erklärenden Variablen und Überanpassung	11
3.1.1. Zwischenbericht	11
3.2. Variablenselektion	11
3.2.1. Variablenselektion aus fachlicher Sicht	11
3.2.2. Datenbasierte Variablenselektion	12
3.3. Weitere Überlegungen zur Modellwahl	13
3.3.1. Modelle mit logarithmierter Zielgrösse	13
3.3.2. Modelle mit zufälligen Effekten	14
3.4. Modell über alle Tonnagewerte	14
3.5. Modelle für kleine bzw. grosse Tonnagewerte	15
3.5.1. Grenzwert für die Tonnage	17
3.5.2. Modell für kleine Tonnagen	17
3.5.3. Modell für grosse Tonnagen	18
4. Schätzung der Rauheit	21
5. Messstrategie für das Jahr 2021	27
5.1. Messstandorte für kleine Tonnagen	27
5.2. Messstandorte für grosse Tonnagen	28
5.3. Wiederholte Messungen am gleichen Standort	29
6. Messunsicherheit	31
7. Ausblick	33

A. Anhang	35
A.1. Abbildungen	35
A.1.1. Beschreibung der einzelnen Variablen	35
A.1.2. Diagnostische Plots	46

1. Einleitung

1.1. Ausgangslage

Gemäss Artikel 7 der Verordnung über die Lärmsanierung der Eisenbahn (VLE) sind die Bahninfrastrukturbetreiber vom Bund beauftragt die Schienenrauheit in ihrem Netz zu kennen und sind verpflichtet die Rauheit unter einem bestimmten Grenzwert zu halten. Zurzeit wird die Schienenrauheit einmal jährlich an den Monitoring Stationen des BAV gemessen. Zudem werden von der SBB punktuelle Messungen an verschiedenen Standorten im Netz durchgeführt. Zum Teil kann während Zugfolgepausen gemessen werden. Dies ist jedoch auf vielbefahrenen Strecken nicht möglich. Es können auch Streckensperrungen für die Messungen beantragt werden. Diese müssen aber mindestens drei Monate im Voraus bekannt sein und bedeuten in jedem Fall einen grösseren Aufwand. Aus diesem Grund werden vor allem bereits geplante Streckenunterbrüche für Unterhalt und Schleifarbeiten ausgenützt, um die Datenbasis zu erhöhen. Dies hat zur Folge, dass die Anzahl Messungen bzw. Messstandorte beschränkt sind und die momentane Datenbasis für eine ausreichend genaue statistische Modellierung der Rauheit des Schienennetzes nicht ausreicht.

Momentan werden die Schienen von der SBB ungefähr alle zwei bis sieben Jahre präventiv geschliffen, ohne die genaue Rauheit zu kennen. Die BLS auf der anderen Seite schleift nur dann, falls es nötig ist, um Schienenfehler zu eliminieren. Würde man die Rauheit des Schienennetzes kennen, so könnte unter Umständen die Schleifarbeit optimiert werden. Es gilt jedoch zu beachten, dass die Schleifstrategie nicht allein von der Rauheit der Schienen abhängig ist, sondern auch die Funktion einer vorausschauenden Wartung übernimmt. Durch das regelmässige Schleifen kann beispielsweise das Risiko von grösseren Rissen in den Gleisen verkleinert werden.

Um ein geeignetes statistisches Modell für die Schienenrauheit zu finden, ist es unumgänglich die Einflussfaktoren zu kennen. Die Rauheit einer Schiene könnte gemäss Experten von vielen Einflussfaktoren abhängen:

- a) Anzahl Zugüberfahrten seit dem letzten Schleifen
- b) Summe der gefahrenen Radsatzlasten seit dem letzten Schleifen
(Der Einfluss der Radsatzlast als solches kann nicht beurteilt werden, weil dazu keine Daten zur Verfügung stehen. Um zu bestimmen, ob der Anteil von schweren Zügen eine Rolle spielt und nicht nur das Gesamtgewicht selbst, wird der Anteil Güterzüge betrachtet.)
- c) Art des Schleifens, insbesondere kurz nach dem Schleifen (akustisches Schleifen). Konkret sollen die verschiedenen Schleifmaschinen miteinander verglichen werden.
- d) Stahlqualität und Profil (UIC 54 oder UIC 60) der Schiene
- e) Verwendeter Schwellentyp und Schotter
- f) Summenanteil von Güter-, respektive Personenverkehr seit dem letzten Schleifen
- g) Streckentyp: Kurve, Gerade, Nähe Bahnhof
- h) Nicht quantifizierbare und zum Teil unbekannte Einflüsse, z.B. Witterungseinflüsse

In einem ersten Schritt sollen ausschliesslich die ersten sechs Einflussfaktoren a) - f) berücksichtigt werden. Diese Faktoren sind für das gesamte Schienennetz gut bekannt und es wird angenommen, dass sie den stärksten Einfluss auf die Schienenrauheit haben. Die vorerst nicht genauer untersuchten Einflussfaktoren g) und h) resultieren in einer grösseren Unsicherheit der geschätzten Schienenrauheit auf dem gesamten Schienennetz, welche im Rahmen dieses Vorprojektes abgeschätzt werden soll.

1.2. Grundlage

Als Datenbasis dienen einerseits die jährlichen Messungen an den sechs stationären Monitoring Stationen des BAV, sowie die bereits vorhandenen Messungen der SBB, welche nicht systematisch durchgeführt wurden. Es wurden jeweils Messungen gemacht, wenn ein Streckenabschnitt aus anderen Gründen gesperrt wurde. Im Laufe des Jahres 2020 sollen noch weitere Messungen hinzukommen, welche bewusst an Standorten gewählt werden, welche im vorhandenen Datensatz noch unterrepräsentiert sind.

1.3. Ziele der aktuellen Projektphase

In der aktuellen Projektphase (2020) soll die Schienenrauheit und insbesondere deren Variabilität mit Hilfe eines ersten einfachen statistischen Modells für das Schienennetz grob geschätzt werden. Das Modell zusammen mit der Schätzunsicherheit der Rauheit dient als Grundlage des Berichts zur Schienenrauheit, welche die SBB bis Ende 2020 eingeben muss. In enger Zusammenarbeit mit der SBB sollen die Ergebnisse das weitere Vorgehen für eine genauere Modellierung der netzweiten Schienenrauheit definieren. Hierzu sind folgende Arbeitsschritte in dieser Projektphase nötig:

1. Zusammenstellung der vorhandenen Daten von Seiten BAV und SBB.
2. In einem ersten einfachen statistischen Modell wird die Schienenrauheit basierend auf den oben genannten Faktoren a) – f) geschätzt.
3. In Absprache mit der SBB werden zusätzliche Messungen an Standorten durchgeführt, welche bisher gar nicht erfasst wurden – insbesondere unter Berücksichtigung der genannten bekannten Einflussfaktoren.

1.4. Ziele des Auftrags

Mit dem Auftrag an die Berner Fachhochschule verfolgt der Projektleiter die folgenden Ziele:

- Erste Abschätzung der Rauheit für das SBB-Schienennetz.
- Erste Abschätzung der Unsicherheit in Bezug auf die Rauheit des SBB-Schienennetzes.
- Definition der weiteren Datenerhebung für das Jahr 2021.
- Definition des weiteren Vorgehens mit dem Ziel die Rauheit des gesamten SBB-Schienennetzes gemäss VLE mit einer gewissen Unsicherheit zu kennen.

2. Beschreibung der vorhandenen Daten

In diesem Abschnitt beschreiben wir die vorhandenen Daten. Es stehen drei Datensätze zur Verfügung:

1. Rauheitsmessungen SBB
2. Monitoringstationen BAV
3. Bümpliz MBBM

Der erste und für uns wichtigste Datensatz enthält Daten zu Messungen, welche von der SBB an verschiedenen Standorten durchgeführt wurden.

Der Datensatz enthält 70 Beobachtungen von Messungen an verschiedenen Orten und zu verschiedenen Zeiten. Seit dem Zwischenbericht konnten zusätzliche Messungen durchgeführt werden. Es stehen nun 22 zusätzliche Beobachtungen zur Verfügung.

Der zweite Datensatz enthält 198 Rauheitsmessungen an den Monitoringstationen des BAV. Für die meisten Messungen im BAV-Datensatz ist nicht bekannt, wann jeweils das letzte Mal geschliffen wurde. Deshalb kann die Tonnage für diese Messungen nicht berechnet werden. Ohne Tonnage sind die Daten nur bedingt aussagekräftig. Deshalb verwenden wir von diesem Datensatz nur die 6 Messungen mit bekannter Tonnage.

Der dritte Datensatz stammt aus einer Studie mit dem Ziel, den Einfluss von Schienendämpfern auf die Rauheit zu ermitteln. Da auf dem restlichen SBB-Netz keine Schienendämpfer verbaut sind, sind hier vor allem die Messungen der Kontrollstandorte ohne Dämpfer interessant. Wir verwenden von diesem Datensatz nur die 8 Beobachtungen ohne Dämpfer.

Total stehen uns 84 Beobachtungen zur Verfügung. Dies sind deutlich mehr als die 48 Beobachtungen, die wir für den Zwischenbericht verwendet haben. Damals hatten wir nur den SBB-Datensatz berücksichtigt.

2.1. Beschreibung der einzelnen Variablen

Im Folgenden werden die einzelnen Variablen kurz beschrieben und dargestellt. Zudem wird der Zusammenhang der erklärenden Variablen mit der Rauheit aufgezeichnet. Um den Lesefluss nicht zu stören werden die meisten Grafiken im Anhang aufgeführt.

2.1.1. Rauheit

Die Variable `Rauheit` beschreibt den auf 80km/h normalisierten Rauheitspegel in dB, also auf einer log-Skala. Die Bestimmung dieser Einzelwerte für die Rauheit nehmen wir als gegeben an und gehen nicht weiter auf deren Berechnung ein.

`Rauheit` ist die numerische Zielgrösse. Wir möchten sie also durch die anderen Variablen möglichst gut erklären. Die Variable `Rauheit` enthält keine fehlenden Werte. In Tabelle 2.1 sind die wichtigsten Kennzahlen für die Rauheit aufgeführt. Auf dem Histogramm in Abbildung 2.1 und dem Boxplot und in Abbildung 2.2 sieht man, dass der SBB-Datensatz den ganzen Wertebereich der Rauheit abdeckt. Wohingegen die beiden anderen Datensätze nur einen kleinen Bereich abdecken und einen deutlich tieferen Median der Rauheit aufweisen.

Tabelle 2.1.: Kennzahlen für die Rauheit

Minimum	-1.93
1. Quartil	3.00
Median	5.84
Mittelwert	6.00
3. Quartil	8.83
Maximum	15.70

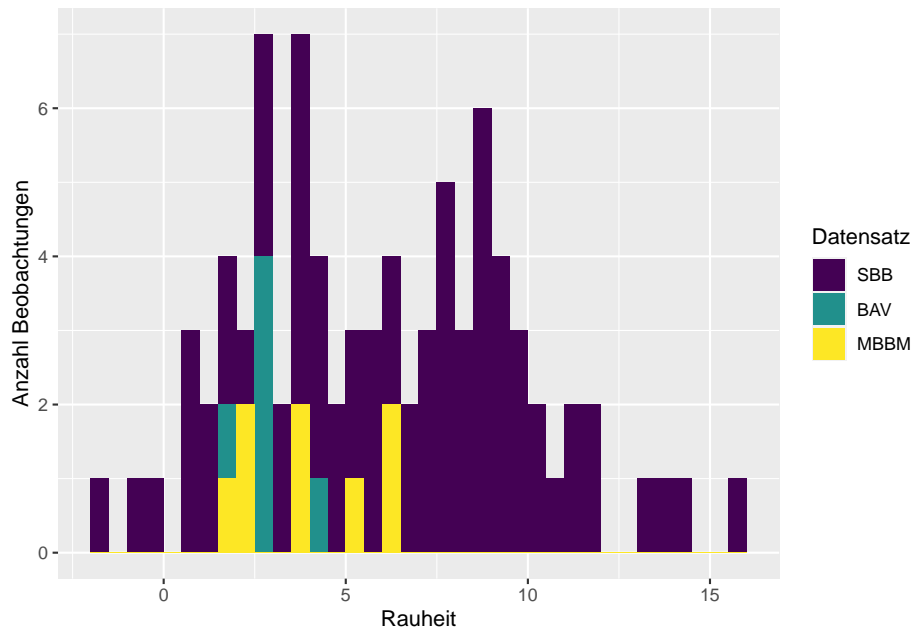


Abbildung 2.1.: Histogramm der Rauheit.

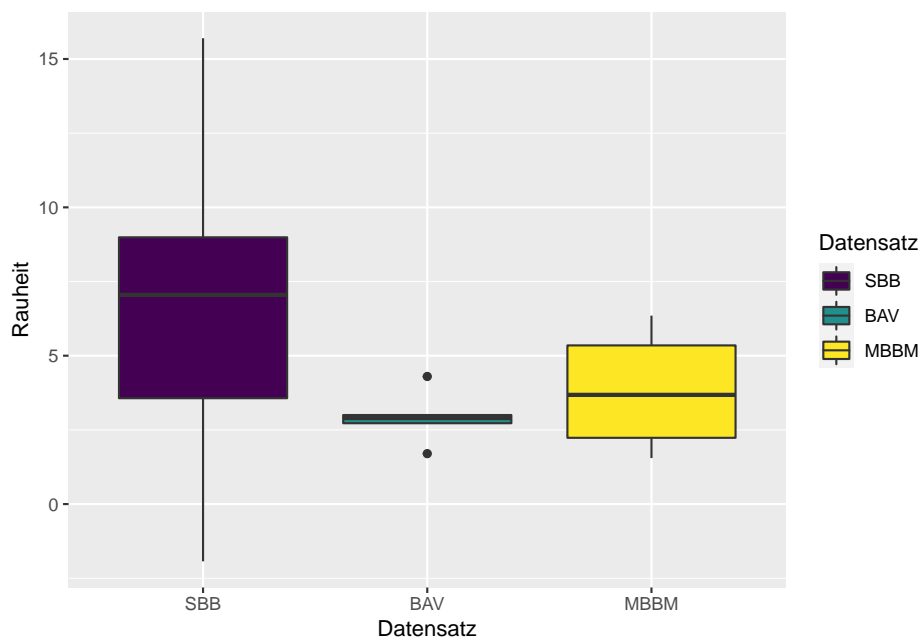


Abbildung 2.2.: Boxplot der Rauheit pro Datensatz.

Tabelle 2.2.: Kennzahlen für die Tonnage

Minimum	0
1. Quartil	373216
Median	2233360
Mittelwert	8728354
3. Quartil	11136045
Maximum	63292004

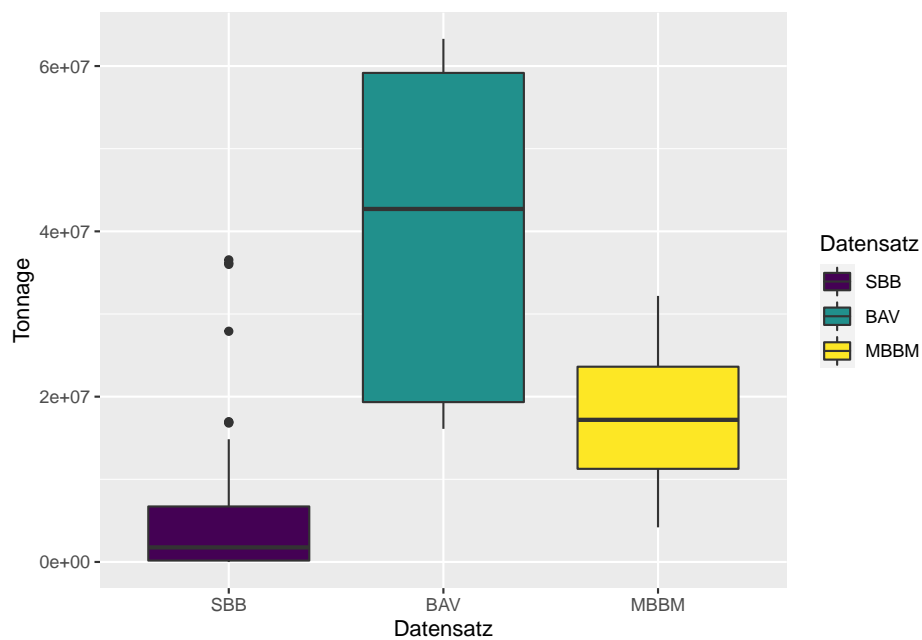


Abbildung 2.3.: Boxplot der Tonnage pro Datensatz.

2.1.2. Tonnage

Die Variable Tonnage beschreibt die Summe der Radsatzlasten seit dem letzten Schleifen in Tonnen.

Tonnage ist numerisch und enthält keine fehlenden Werte. Die Verteilung ist ziemlich schief. Es hat ein paar Ausreisser gegen oben. Tonnage enthält keine fehlenden Werte, weil wir alle Beobachtungen mit fehlenden Tonnage-Werten entfernt haben. In Tabelle 2.2 sind die wichtigsten Kennzahlen für die Tonnage aufgeführt. Im SBB Datensatz hat es viele Beobachtungen mit tiefer Tonnage, aber auch einzelne mit hoher Tonnage. In den anderen beiden Datensätzen kommen keine ganz tiefen Tonnagewerte vor. Dies ist auf dem Boxplot in Abbildung 2.3 und auf dem Histogramm in Abbildung A.1 ersichtlich.

Abbildung 2.4 zeigt ein Streudiagramm mit der Tonnage und der Rauheit. Die Rauheit nimmt mit wachsender Tonnage zuerst stark ab. Für grössere Tonnagewerte ist die Abnahme dann viel geringer. Auch die Streuung der Rauheit ist für kleine Tonnagewerte viel grösser als für grosse.

2.1.3. Anzahl Zugüberfahrten

Die Variable Anzahl .Zug beschreibt die Anzahl Zugüberfahrten seit dem letzten Schleifen.

Anzahl .Zug ist numerisch und enthält 8 fehlenden Werte. Wie zu erwarten sind die Tonnage und die Anzahl Zugüberfahrten hochkorreliert. Der Pearson-Korrelationskoeffizient beträgt 0.96. Wegen der hohen Korrelation ist es nicht sinnvoll, beide Variablen in den Modellen zu verwenden. Wir werden im Folgenden nur noch die Tonnage verwenden. Dies insbesondere auch, weil die Anzahl Zugüberfahrten im MBBM-Datensatz nicht vorkommt.

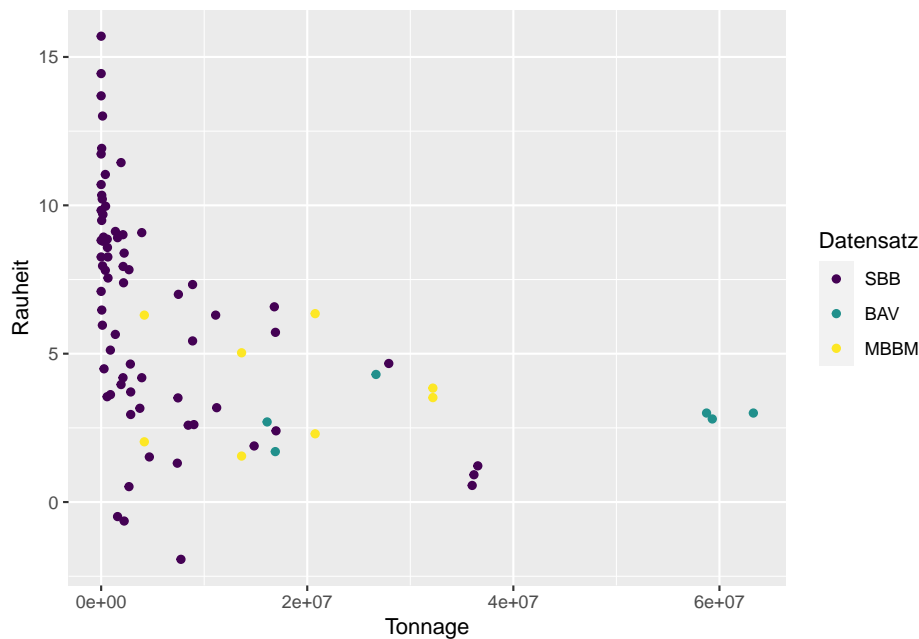


Abbildung 2.4.: Die Rauheit nach Tonnage.

Tabelle 2.3.: Kennzahlen für die Anzahl Zugüberfahrten

Minimum	0
1. Quartil	456
Median	4968
Mittelwert	15755
3. Quartil	21383
Maximum	106421

Abbildungen A.2 und A.3 zeigen ein Histogramm und einen Boxplot der Anzahl Zugüberfahrten. In Tabelle 2.3 sind die wichtigsten Kennzahlen für die Anzahl Zugüberfahrten aufgeführt.

2.1.4. Angewandte Technik beim letzten Schleifen (Maschine)

Die Variable Maschine enthält die beim letzten Schleifen verwendete Maschine und repräsentiert somit die angewandte Technik beim letzten Schleifen. Eine Messung wurde an einer Stelle gemacht, die mit einer Fräsmaschine bearbeitet wurde. Da wir nur eine solche Beobachtung haben, werden wir sie für die Modelle nicht verwenden. In den Daten hat diese Beobachtung den Wert F für die Maschine.

Maschine ist kategorial mit 5 Ausprägungen und enthält keine fehlenden Werte. Die Verteilung ist in der Häufigkeitstabelle in Tabelle 2.4 und im Säulendiagramm in Abbildung A.4 ersichtlich. In Abbildung A.5 ist der Zusammenhang zwischen der Rauheit und der Maschine in Boxplots dargestellt.

Tabelle 2.4.: Häufigkeitstabelle für die Maschine

A	25
B	10
C	20
D	28
F	1

Tabelle 2.5.: Häufigkeitstabelle für die Stahlqualität

R260	69
R320Cr	2
R350	13

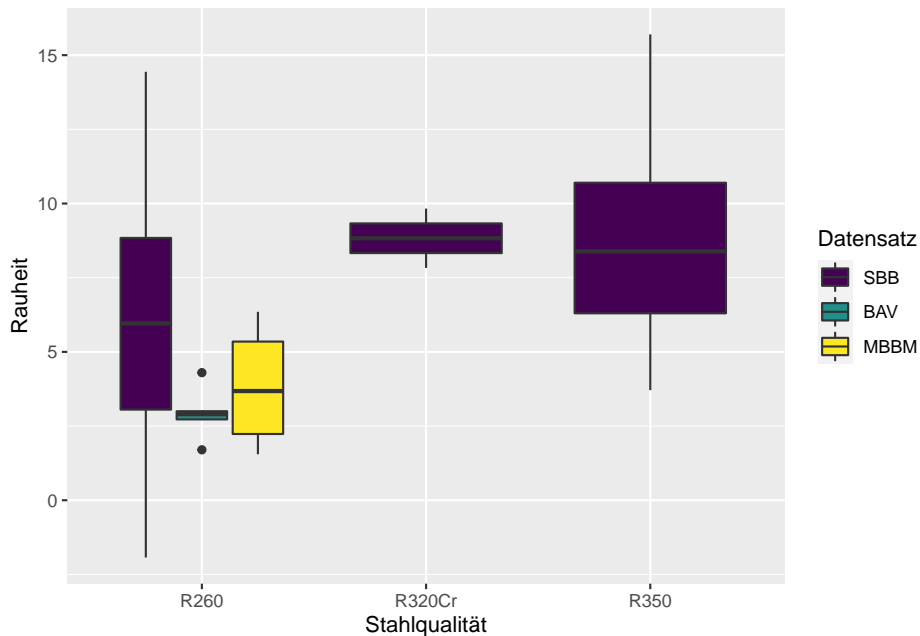


Abbildung 2.5.: Boxplots der Rauheit pro Stahlqualität und Datensatz.

2.1.5. Stahlqualität

Die Variable Stahl beschreibt die Stahlqualität. Sie ist kategoriell mit den 3 Ausprägungen R260, R350 und R320Cr. Der Typ R260 kommt mit Abstand am häufigsten vor. Die beiden Typen R350 und R320Cr sind härter als R260. Sie werden in engen Kurven verwendet.

Stahl enthält keine fehlenden Werte. Für die Kategorie R320Cr haben wir nur 2 Beobachtungen.

Die Verteilung ist in Tabelle 2.5 und in Abbildung A.6 aufgetragen. In Abbildung 2.5 ist der Zusammenhang zwischen der Rauheit und der Stahlqualität in Boxplots dargestellt. Der weiche Stahl hat tendenziell eine tiefere Rauheit als die härteren Stahlqualitäten.

2.1.6. Schienenprofil

Die Variable Schienenprofil enthält das Schienenprofil nach europäischer Notation. Sie ist kategoriell mit den 2 Ausprägungen UIC 60 und UIC 54.

Schienenprofil enthält keine fehlenden Werte. Für die Kategorie UIC 54 haben wir nur 6 Beobachtungen. Die Verteilung ist in Tabelle 2.6 und in Abbildung A.7 aufgetragen. In Abbildung A.8 ist der Zusammenhang zwischen der Rauheit und der Stahlqualität in Boxplots dargestellt.

Tabelle 2.6.: Häufigkeitstabelle für das Schienenprofil

UIC 54	6
UIC 60	78

Tabelle 2.7.: Häufigkeitstabelle für den Schwellentyp

B70	12
B91	65
Holz	5
Stahl	2

Tabelle 2.8.: Häufigkeitstabelle für den Schwellenbesohlung

ohne	69
steif	15

2.1.7. Schwellentyp

Die Variable Schwellentyp beschreibt den Schwellentyp. Bei Betonschwellen wird zwischen B70 (alte Schwelle) und B91 (neue Schwelle) unterschieden. Bei Holz- und Metallschwellen werden keine Typen unterschieden. In diesen Fällen enthält die Variable Schwellentyp das Schwellenmaterial.

Schwellentyp ist kategoriell mit 4 verschiedenen Ausprägungen und enthält keine fehlenden Werte. In den Daten kommen nur 5 Holzschwellen und 2 Stahlschwellen vor. Die Verteilung ist in Tabelle 2.7 und in Abbildung A.9 aufgetragen. In Abbildung A.10 ist der Zusammenhang zwischen der Rauheit und dem Schwellentyp in Boxplots dargestellt.

2.1.8. Besohlung der Schwelle

Die Variable Schwellenbesohlung beschreibt die Besohlung der Schwelle, also ob bei einer Betonschwelle ein Teil aus Gummi unter der Schwelle vorhanden ist. Zudem wird zwischen einer steifen und einer mittelsteifen Besohlung unterschieden. In den vorhandenen Daten kommen jedoch nur Beobachtungen mit einer steifen Besohlung oder ohne Besohlung vor.

Schwellenbesohlung ist kategoriell mit 2 Ausprägungen und enthält keine fehlenden Werte. Die Verteilung ist in Tabelle 2.8 und in Abbildung A.11 aufgetragen. In Abbildung A.12 ist der Zusammenhang zwischen der Rauheit und der Schwellenbesohlung in Boxplots dargestellt.

2.1.9. Schienenbefestigungstyp

Die Variable Schienenbefestigungstyp beschreibt den Schienenbefestigungstyp. Dieser setzt sich aus der Winkelführungsplatte und der Spannklemme zusammen.

Schienenbefestigungstyp ist kategoriell mit 6 verschiedenen Ausprägungen und enthält keine fehlenden Werte. Die Verteilung ist in Tabelle 2.9 und in Abbildung A.13 aufgetragen. In Abbildung A.14 ist der Zusammenhang zwischen der Rauheit und der Schwellenbesohlung in Boxplots dargestellt.

Tabelle 2.9.: Häufigkeitstabelle für den Schwellenbesohlung

Aek	2
Ke_Skl12	1
Ke_Skl3	4
W_Skl1	10
W14_Skl14	51
Ws_Skl1	16

Tabelle 2.10.: Häufigkeitstabelle für den Schienenzwischenlage

hart	57
weich	27

Tabelle 2.11.: Häufigkeitstabelle für den Schotterdicke

>35	34
21-35	29

2.1.10. Schienenzwischenlage

Die Variable Zwischenlage beschreibt die Schienenzwischenlage, also ob zwischen der Schiene und der Schwelle ein Teil aus Gummi vorhanden ist. Zudem wird zwischen einer harten und einer weichen Schienenzwischenlage unterschieden.

Zwischenlage ist kategoriell mit 2 verschiedenen Ausprägungen und enthält keine fehlenden Werte. Die Verteilung ist in Tabelle 2.10 und in Abbildung A.15 aufgetragen. In Abbildung A.16 ist der Zusammenhang zwischen der Rauheit und der Schwellenbesohlung in Boxplots dargestellt.

2.1.11. Schotterdicke

Die Variable Schotterdicke beschreibt, ob die Schotterdicke kleiner oder grösser als 35cm ist.

Schotterdicke ist kategoriell mit 2 verschiedenen Ausprägungen und enthält 21 fehlende Werte. Wegen der vielen fehlenden Werte verwenden wir die Schotterdicke in den Modellen nicht. Die Verteilung ist in Tabelle 2.11 und in Abbildung A.17 aufgetragen. In Abbildung A.18 ist der Zusammenhang zwischen der Rauheit und der Schwellenbesohlung in Boxplots dargestellt.

2.1.12. Anteil Güterverkehr

Der Datensatz enthält Angaben zu den Anteilen Güterzüge, Reisezüge und Dienstzüge. Einmal sind es die Anteile am totalen Gewicht und einmal die Anteile an der totalen Anzahl Züge. Total sind es also 6 Variablen.

Der Anteil Dienstzüge ist vernachlässigbar klein. Deshalb betrachten wir nur den Anteil Güterzüge. Zudem unterscheiden sich die Anteile gemessen am Gewicht und gemessen an der Anzahl Züge nicht wesentlich. Deshalb verwenden wir nur den Anteil nach Gewicht.

Im Folgenden verwenden wir den Anteil Güterzüge nicht als numerische Variable, sondern kategorisieren ihn. Den Anteilen kleiner oder gleich 30% weisen wir den Wert „klein“ zu, den Anteilen grösser als 30% den Wert „gross“.

Die resultierende Variable Anteil.GZ ist kategoriell mit 2 verschiedenen Ausprägungen und enthält keine fehlenden Werte. Die Verteilung ist in Tabelle 2.12 und in Abbildung A.19 aufgetragen. Es fällt auf, dass der Anteil Güterverkehr an den BAV-Messstationen gross und an den MBBM-Messstellen klein ist. In Abbildung A.20 ist der Zusammenhang zwischen der Rauheit und dem Anteil in Boxplots dargestellt.

Tabelle 2.12.: Häufigkeitstabelle für den Anteil Güterverkehr

gross	23
klein	61

Tabelle 2.13.: Häufigkeitstabelle für Kurve

ja	8
nein	76

2.1.13. Streckengeometrie

Der Datensatz enthält die Variable Radius in Meter. Wir kategorisieren den Radius und berechnen die neue Variable Kurve. Sie ist gleich 1, falls der Radius kleiner als 1000m ist, und 0, falls der Radius grösser oder gleich 1000m ist.

Die resultierende Variable Kurve ist kategoriell und enthält keine fehlenden Werte. Die Verteilung ist in Tabelle 2.13 und in Abbildung A.21 aufgetragen. In Abbildung A.22 ist der Zusammenhang zwischen der Rauheit und der Variable Kurve in Boxplots dargestellt.

2.1.14. Weitere Variablen

Zusätzlich werden im Datensatz der Ort und das Datum der Messung angegeben. Im Moment verwenden wir diese Variablen nicht. Im Wissen, dass das nicht ganz korrekt ist, gehen wir für die ersten Modelle naiv davon aus, dass der Ort und das Datum keinen Einfluss auf die Rauheit haben. Wenn wir diese Effekte berücksichtigen wollten, bräuchten wir deutlich mehr Daten.

3. Statistische Modelle

3.1. Modell mit allen erklärenden Variablen und Überanpassung

Zuerst haben wir versucht, ein Modell mit allen verfügbaren Variablen zu rechnen. Es stehen 10 Variablen zur Verfügung, um die Rauheit zu erklären. Eine Faustregel besagt, dass man pro erklärende Variable mindestens 10 Beobachtungen benötigt. Wenn wir also alle Variablen verwenden möchten, bräuchten wir ca. 100 Beobachtungen. Dies ist aber nur eine Faustregel. Sie berücksichtigt insbesondere nicht, wie viele mögliche Ausprägungen die Erklärenden haben und wie die Daten verteilt sind. Die Daten in unserer Stichprobe sind ziemlich unbalanciert. Das heisst die verschiedenen Ausprägungskombinationen der Erklärenden kommen sehr unterschiedlich häufig vor. Deshalb gehen wir davon aus, dass man noch deutlich mehr Beobachtungen bräuchte, um eine gute Schätzung zu erhalten. Mit den im Jahr 2021 geplanten Messungen könnte die Qualität der Schätzung verbessert werden.

Für die akute Studie stehen uns nur 84 Beobachtungen zur Verfügung. Dies führt dazu, dass sich ein Modell mit allen erklärenden Variablen zu stark an die Stichprobe anpasst. Wegen dieser Überanpassung (englisch *overfitting*) ist das Modell nicht oder nur bedingt verallgemeinerbar. Mit diesem Modell gemachte Vorhersagen der Rauheit für zukünftige Beobachtungen werden also nicht besonders gut sein. Eine weitere Folge der Überanpassung ist eine Instabilität der Modelle. Die Parameter des Modells verändern sich sehr stark, wenn man nur eine Beobachtung weglässt oder hinzufügt. Auch das Weglassen oder Hinzufügen einer erklärenden Variablen hat einen sehr grossen Einfluss auf die Parameterschätzung.

3.1.1. Zwischenbericht

Für den Zwischenbericht standen uns nur 48 Beobachtungen zur Verfügung. Deshalb war dort das Problem mit der Überanpassung noch viel grösser. Wegen der wenigen Beobachtungen traten seltsame Effekte auf und die Effekte der verschiedenen Erklärenden konnten zum Teil nicht getrennt werden. Diese seltsamen Effekte sind mit den zusätzlichen Beobachtungen verschwunden. Es bleibt jedoch dabei, dass es eher wenige Beobachtungen sind und deshalb die Modelle mit allen Erklärenden ziemlich instabil sind.

3.2. Variablenselektion

Um bessere und stabilere Modelle zu erhalten, müssen wir die Anzahl erklärender Variablen reduzieren.

3.2.1. Variablenselektion aus fachlicher Sicht

In einem ersten Schritt haben wir zusammen mit der SBB Variablen definiert, deren Einfluss auf die Rauheit aus fachlicher Sicht als eher gering eingeschätzt wird.

Die vier Variablen Besohlung der Schwelle, Schienenprofil, Schienenbefestigungstyp und Schotterdicke wurden dabei als unwichtig eingestuft. Für die Modelle in dieser Vorstudie werden wir sie also nicht mehr verwenden.

Es verbleiben also noch die 7 erklärenden Variablen Tonnage, Maschine, Stahlqualität, Schienenzwischenlage, Anteil Güterzüge, Schwellentyp und Kurve.

Für die Rauheit kurz nach dem Schleifen, also bei kleinen Tonnagewerten, wird zudem der Einfluss des Anteils Güterzüge als gering eingeschätzt. Diese Variable verwenden wir also nicht für Modelle, die sich auf kleine Tonnagewerte beschränken.

Analog verzichten wir bei Modellen für grosse Tonnagewerte auf die Variable Maschine. Es wird angenommen, dass die verwendete Maschine einen Einfluss auf die Rauheit nach dem Schleifen hat. Dieser Einfluss wird jedoch kleiner, je mehr Züge nach dem Schleifen über die Stelle gefahren sind. Für grosse Tonnagewerte wird der Einfluss als vernachlässigbar eingeschätzt.

3.2.2. Datenbasierte Variablenselektion

In einem zweiten Schritt haben wir einerseits das volle Modell mit allen im letzten Abschnitt gewählten Variablen gerechnet. Andererseits haben wir diverse reduzierte Modelle gerechnet. Dafür haben wir die Selektionsmethoden forward selection, backward elimination und lasso verwendet.

3.2.2.1. Forward selection und backward elimination

Für die forward selection beginnt man mit dem Nullmodell, also mit dem Modell ohne erklärende Variablen. Dann fügt man in jedem Schritt eine Variable hinzu bis man das volle Modell erreicht hat. In jedem Schritt fügt man diejenige Variable hinzu, welche die Modellanpassung am meisten verbessert. Man rechnet also alle Modelle mit einer zusätzlichen Variable und vergleicht sie mit dem Modell aus dem letzten Schritt. Für den nächsten Schritt verwendet man dann das beste Modell. Der Prozess wird gestoppt, wenn keines der Kandidatenmodelle besser ist als das aktuelle.

Die backward elimination funktioniert analog. Man startet aber mit dem vollen Modell, also mit dem Modell das alle Erklärenden enthält. In jedem Schritt wirft man eine Variable aus dem Modell bis man am Schluss beim Nullmodell landet.

Um zu bestimmen, welches Modell das beste ist, braucht man ein Vergleichskriterium. In der Literatur gibt es verschiedene Kriterien. Wir verwenden das Akaike-Informationskriterium (AIC). Es ist eines der am häufigsten verwendeten Kriterien. Das AIC belohnt die Anpassungsgüte und bestraft zu hohe Modellkomplexität.

3.2.2.2. Lasso

Das lasso (least absolute shrinkage and selection operator) ist eine weitere Selektionsmethode. Sie wurde für hochdimensionale Daten mit mehreren Tausend Erklärenden entwickelt und geht davon aus, dass die meisten Erklärenden keinen Einfluss auf die Zielvariable haben, dass also die meisten Parameter Null sind.

Diese Voraussetzung sind in unserem Fall zwar nicht gegeben. Dennoch kann man lasso verwenden, um das Problem der Überanpassung in den Griff zu kriegen.

Lasso und die anderen verwendeten Selektionsmethoden werden beispielsweise in [9] ausführlicher beschrieben.

3.2.2.3. Kriterien für die Wahl des besten Modells

Es gibt verschiedene Kriterien, um Modelle zu bewerten. Die Frage dabei ist, was das Ziel der Modellierung ist. Soll das Modell die Streuung der Zielvariable möglichst gut durch die Erklärenden beschreiben? Soll bestimmt werden, welche Erklärenden einen Einfluss auf die Zielvariable haben? Oder möchte man den Wert der Zielvariable für zukünftige Beobachtungen möglichst gut vorhersagen?

RMSE

Unser primäres Ziel ist, die Rauheit möglichst gut vorhersagen zu können. Deshalb ist ein naheliegendes Kriterium der mittlere quadrierte Vorhersagefehler (englisch mean squared error)

$$\text{MSE} := \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

Für die Berechnung des MSE werden die Differenzen zwischen den wahren und den geschätzten Werten quadriert und aufsummiert. Damit die Kenngrösse auf der gleichen Skala ist wie der Schätzer, betrachten wir die Wurzel des MSE, also den root mean squared error

$$\text{RMSE} := \sqrt{\text{MSE}}.$$

Damit wir den RMSE berechnen können, brauchen wir also Beobachtungen, von denen wir sowohl die wahren Werte als auch die geschätzten Werte der Rauheit kennen. Nun könnte man die Stichprobe für die Modellbildung und die Bewertung des Modells verwenden. Dies würde jedoch zu übermässig optimistischen Schätzungen des Fehlers führen. Die Vorhersagen eines Modells für die Daten, mit denen es trainiert wurde, sind typischerweise besser als für neue, zufällige Beobachtungen. Um diesen Effekt zu vermeiden, führen wir eine Kreuzvalidierung durch. Dabei nehmen wir eine Beobachtung aus dem Datensatz heraus und rechnen ein Modell mit den übrigen Daten. Dann verwenden wir dieses Modell um eine Schätzung für die entfernte Beobachtung zu machen. Dies wiederholen wir für alle Beobachtungen. So erhalten wir für jede Beobachtung einen unverfälschten Schätzwert und können damit den RMSE berechnen.

Das Bestimmtheitsmass R^2

Als Kennzahl für die Anpassungsgüte kann man das Bestimmtheitsmass R^2 verwenden. Es ist gegeben durch das Verhältnis der durch die Regression erklärten Streuung zur totalen Streuung der Zielvariablen. Als Anteil nimmt das R^2 Werte zwischen Null und Eins an.

Weitere Kriterien

Damit ein Modell als gut bewertet werden kann, müssen die Modellvoraussetzungen erfüllt sein. Die Verteilung der Residuen sollte unabhängig vom gefitteten Wert sein. Dies gilt insbesondere für ihre Streuung. Anpassungstests, Konfidenz- und Prädiktionsintervalle setzen zudem voraus, dass die Residuen normalverteilt sind. In der linearen Regression können einflussreiche Beobachtungen die Parameterschätzung massiv beeinflussen (Hebeleffekt). Ein gutes Modell sollte möglichst keine allzu einflussreichen Beobachtungen haben.

Anhand diagnostischer Plots können diese Voraussetzungen überprüft werden. Wenn ein Modell die Voraussetzungen besser erfüllt als ein anderes, wäre das ein Grund, es zu bevorzugen.

Ein weiteres Kriterium ist die Interpretierbarkeit. In der Regel wird man ein Modell bevorzugen, dass man fachlich besser interpretieren kann.

3.3. Weitere Überlegungen zur Modellwahl

3.3.1. Modelle mit logarithmierter Zielgrösse

Auf dem Streudiagramm mit der Rauheit und der Tonnage in Abbildung 2.4 sieht man, dass der Zusammenhang der beiden Variablen eher exponentiell als linear ist. Dies würde dafür sprechen, die Zielgrösse Rauheit zu logarithmieren. Wir haben auch solche Modelle gerechnet. Die entsprechenden diagnostischen Plots sehen jedoch deutlich schlechter aus als diejenigen der Modelle ohne Logarithmierung. Zudem müsste man für die Logarithmierung die Beobachtungen mit negativer Rauheit entfernen, weil der Logarithmus von negativen Zahlen nicht definiert ist. Deshalb betrachten wir nur Modelle ohne Logarithmierung.

Tabelle 3.1.: Modellvergleich für alle Tonnagewerte: verwendete Variablen

Modell							
voll	Tonnage	Maschine	Stahl	Zwischenlage	Anteil.GZ	Schwellentyp	Kurve
forward	Tonnage		Stahl	Zwischenlage			Kurve
backward	Tonnage	Maschine	Stahl			Schwellentyp	Kurve
lasso	Tonnage	Maschine*	Stahl	Zwischenlage	Anteil.GZ	Schwellentyp	Kurve

* Lasso setzt den Parameter für Maschine B auf Null.

Tabelle 3.2.: Modellvergleich für alle Tonnagewerte: RMSE und R^2

Modell	RMSE	R^2
voll	3.32	0.43
forward	3.28	0.34
backward	3.29	0.43
lasso	3.27	

3.3.2. Modelle mit zufälligen Effekten

Lineare Modelle setzen voraus, dass die Beobachtungen unabhängig sind. Dies ist bei den vorliegenden Daten nur bedingt gegeben. Zum Teil wurden mehrere Messungen am gleichen Standort durchgeführt. Die führt zu einer Abhängigkeit der Beobachtungen. Zwei Messungen am gleichen Standort werden sich zueinander nicht gleich verhalten, wie zwei Messungen (mit identischen Ausprägungen der Erklärenden) an zwei verschiedenen Standorten.

Diese Abhängigkeit könnte man mit gemischten Modellen (englisch mixed models) modellieren. Weil wir aber relativ wenige Beobachtungen und ein unbalanciertes Design haben, verzichten wir darauf, gemischte Modelle zu verwenden. Dies im Wissen, dass die Voraussetzungen für die verwendeten Modelle eigentlich nicht vollständig erfüllt sind.

3.4. Modell über alle Tonnagewerte

Aus fachlichen Überlegungen sollen zwei Modelle gerechnet werden. Das erste Modell für kleine Tonnagewerte, also für Messungen kurz nach dem Schleifen. Das zweite Modell für grosse Tonnagewerte. Es wird angenommen, dass manche Variablen nur kurz nach dem Schleifen einen Einfluss auf die Rauheit haben und andere erst später, für grössere Tonnagewerte eine Rolle spielen. Trotzdem haben wir zuerst Modelle für alle Beobachtungen gerechnet, also über den gesamten Wertebereich der Tonnage.

In den Tabellen 3.3 bis 3.6 sind die Parameterschätzwerte und die dazugehörigen Standardfehler angegeben. Für die Tonnage ist der Parameter für Tonnage in Megatonnen (Mt) angegeben. Er ist wie folgt zu interpretieren:

Wenn man alle anderen Erklärenden fixiert und die Tonnage um eine Million erhöht, sinkt die geschätzte mittlere Rauheit um 0.096 (im vollen Modell).

Für die kategoriellen Variablen wird jeweils eine Referenzkategorie gewählt und der Effekt relativ zu dieser Kategorie angegeben. Für die Maschine wird beispielsweise Maschine A als Referenz gewählt. Der Parameter der Maschine B wird wie folgt interpretiert:

Wenn man alle anderen Erklärenden fixiert, ist die geschätzte mittlere Rauheit für Maschine B um 0.216 höher als für Maschine A (im vollen Modell).

Das volle Modell

$$\text{Rauheit} \sim \text{Tonnage} + \text{Maschine} + \text{Stahl} + \text{Zwischenlage} + \text{Anteil.GZ} + \text{Schwellentyp} + \text{Kurve}$$

Tabelle 3.3.: Koeffizienten für das volle Modell für alle Tonnagewerte

	Schätzwert	Standardfehler
(Intercept)	0.340	2.416
Tonnage in Mt	-0.096	0.032
Maschine: B	0.216	1.530
Maschine: C	1.855	1.205
Maschine: D	1.932	1.081
Stahl: R320Cr	3.829	2.693
Stahl: R350	2.933	1.146
Zwischenschicht: weich	1.058	1.327
Anteil GZ: klein	0.141	1.319
Schwelle: B91	2.207	1.381
Schwelle: Holz	-0.774	1.991
Schwelle: Stahl	3.572	3.132
Kurve: nein	2.878	1.598

erreicht ein R^2 von 0.43 und einen RMSE von 3.32.

Mit der forward selection erhalten wir das Modell

$$\text{Rauheit} \sim \text{Tonnage} + \text{Stahl} + \text{Zwischenlage} + \text{Kurve}$$

mit einem R^2 von 0.34 und einem RMSE von 3.28.

Die backward elimination liefert das Modell

$$\text{Rauheit} \sim \text{Tonnage} + \text{Maschine} + \text{Stahl} + \text{Schwellentyp} + \text{Kurve}$$

mit einem R^2 von 0.43 und einem RMSE von 3.29.

Lasso wirft nur die Maschine B raus. Die anderen Maschinen und auch alle anderen Variablen bleiben im Modell. Der RMSE liegt bei 3.27, also nicht wesentlich tiefer als im vollen Modell. Aus diesen Gründen würden wir hier lasso nicht verwenden.

Wir wählen das Modell der backward elimination. Das R^2 ist fast gleich hoch wie im vollen Modell, obwohl es zwei erklärende Variablen weniger hat. Der RMSE ist vergleichbar mit dem der anderen Modelle.

Die diagnostischen Plots sind im Anhang in den Abbildungen A.23 bis A.28 aufgeführt. Sie sehen nicht schlecht aus. Es fällt einzig auf, dass bei den ganz tief geschätzten Werten alle Residuen positiv sind. Das ist ein wenig unschön, aber kein Grund, die Modelle zu verwerfen. Die Q-Q-Plots liefern keine Hinweise, dass die Normalverteilungsannahme verletzt sein könnte. Insgesamt sind die diagnostischen Plots für alle Modelle vergleichbar. Sie liefern also kein Argument, um ein Modell zu bevorzugen.

In Tabelle 3.1 sind alle Modelle mit den gewählten Variablen aufgeführt. Tabelle 3.2 fasst noch einmal den RMSE und das R^2 für alle Modelle zusammen.

3.5. Modelle für kleine bzw. grosse Tonnagewerte

Die Rauheit soll für Messungen mit kleinen bzw. grossen Tonnagewerten separat modelliert werden. A priori ist jedoch nicht klar, wo die Grenze zwischen klein und gross gesetzt werden soll. Deshalb suchen wir zuerst mit dem vollen Modell einen möglichst guten Grenzwert für die Tonnage. Danach rechnen wir möglichst gute Modelle für kleine und grosse Tonnagen.

Tabelle 3.4.: Koeffizienten für das Modell der forward selection für alle Tonnagewerte

	Schätzwert	Standardfehler
(Intercept)	3.800	1.307
Tonnage in Mt	-0.103	0.027
Stahl: R320Cr	3.228	2.344
Stahl: R350	3.706	1.069
Zwischenschicht: weich	1.860	0.800
Kurve: nein	2.022	1.287

Tabelle 3.5.: Koeffizienten für das Modell der backward elimination für alle Tonnagewerte

	Schätzwert	Standardfehler
(Intercept)	0.050	2.228
Tonnage in Mt	-0.107	0.028
Maschine: B	-0.397	1.328
Maschine: C	1.335	0.987
Maschine: D	2.043	0.963
Stahl: R320Cr	3.733	2.652
Stahl: R350	2.639	1.082
Schwelle: B91	2.663	1.257
Schwelle: Holz	-0.476	1.955
Schwelle: Stahl	4.758	2.863
Kurve: nein	3.579	1.396

Tabelle 3.6.: Koeffizienten für das lasso für alle Tonnagewerte

	Schätzwert
(Intercept)	1.409
Tonnage in Mt	-0.094
Maschine: B	0.000
Maschine: C	1.618
Maschine: D	1.504
Stahl: R320Cr	3.350
Stahl: R350	2.666
Zwischenschicht: weich	1.052
Anteil GZ: klein	0.097
Schwelle: B91	1.899
Schwelle: Holz	-0.863
Schwelle: Stahl	2.721
Kurve: nein	2.325

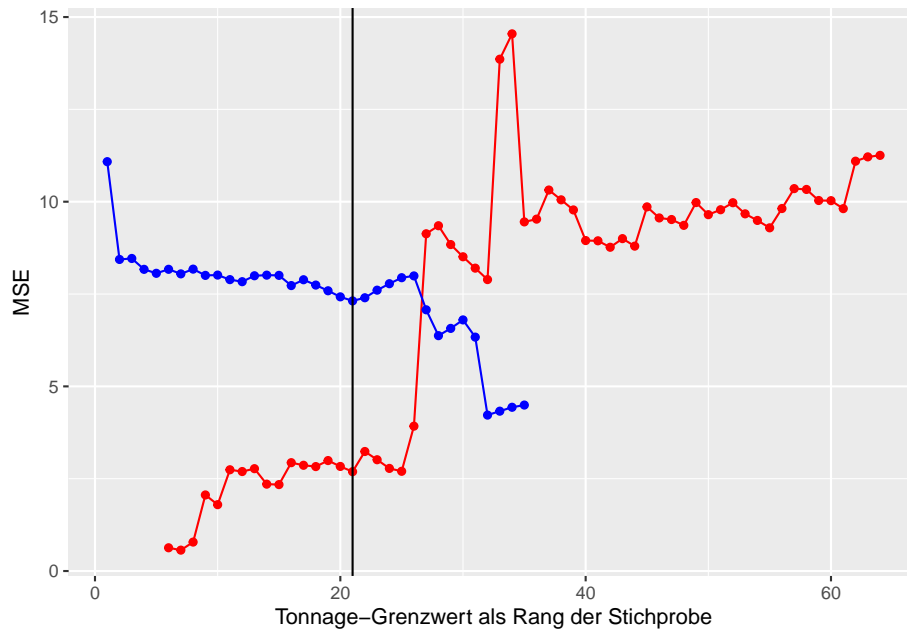


Abbildung 3.1.: Der quadratische Vorhersagefehler (MSE) für die Modelle mit Tonnagen kleiner als der Grenzwert (rot) bzw. mit Tonnagen grösser als der Grenzwert (blau).

3.5.1. Grenzwert für die Tonnage

Um den besten Tonnagegrenzwert zu finden, rechnen wir für verschiedene Grenzwerte je ein Modell mit Tonnage tiefer bzw. höher als der Grenzwert. Dann wählen wir den besten Grenzwert in dem Sinne, dass wir zwei Modelle mit möglichst kleinen RMSE erhalten.

Abbildung 3.1 zeigt den quadratische Vorhersagefehler (MSE) für die Modelle mit Tonnagen kleiner als der Grenzwert (rot) bzw. mit Tonnagen grösser als der Grenzwert (blau). Die schwarze Linie zeigt den gewählten Grenzwert. Für die bessere Lesbarkeit wird auf der X-Achse nicht die Tonnage des Grenzwerts aufgezeichnet, sondern deren Rang in der Stichprobe. Als Grenzwerte wurden alle Tonnagewerte betrachtet, die in der Stichprobe vorkommen.

Am linken Rand fehlen Punkte für die rote Kurve, am rechten für die blaue. Das liegt daran, dass es für diese Grenzwerte zu wenige Beobachtungen für die entsprechenden Modelle hätte und deshalb die Modellanpassung nicht möglich war.

Die rote Kurve macht einen Sprung zwischen Beobachtung 25 und 27. Die Grenze sollte sicher vor dem Sprung gewählt werden und nicht zu Nahe beim Sprung, damit man eine gewisse Stabilität hat. Zwischen Beobachtung 11 und 25 sind die beiden Kurven ziemlich stabil. Es wäre also sinnvoll, einen Wert in diesem Bereich zu wählen. Wir wählen die Beobachtung 21, da die Summe der beiden MSE an dieser Stelle minimal ist (innerhalb des stabilen Bereichs).

Innerhalb des Intervalls zwischen den Tonnagewerten für die Beobachtungen 21 und 22 spielt es keine Rolle, welchen Wert wir wählen, da sich die Modelle dadurch nicht ändern. Die beiden Beobachtungen haben die Werte 664860 und 896012. Wir setzen die Grenze bei 750000, damit wir eine einigermassen runde Zahl haben.

Der Median der Belastung aller Zugleise der SBB liegt bei rund 30000 Belastungstonnen pro Tag. Die Tonnage von 750000 wird an einer Stelle, an der die Belastung dem Median entspricht, nach 25 Tagen erreicht. Auf einer stark befahrenen Strecke mit 120000 Belastungstonnen pro Tag wird die Tonnage von 750000 bereits nach ca. einer Woche erreicht.

3.5.2. Modell für kleine Tonnagen

Für die Modelle für Tonnagen kleiner oder gleich 750000 entfernen wir alle Beobachtungen mit Stahlqualität R320Cr bzw. Schwellentyp Holz oder Stahl. Da es zu wenige Beobachtungen mit diesen Ausprägungen hat, würde das sonst zu Fehlern in der Kreuzvalidierung führen. Die Variable Anteil

Tabelle 3.7.: Koeffizienten für das volle Modell für kleine Tonnagen

	Schätzwert	Standardfehler
(Intercept)	7.594	1.329
Tonnage in Mt	-6.148	1.236
Maschine: B	-0.570	1.482
Maschine: C	3.744	1.370
Maschine: D	0.737	0.750
Stahl: R350	2.857	0.981
Zwischenschicht: weich	3.508	1.243

Tabelle 3.8.: Koeffizienten für das lasso für kleine Tonnagen

	Schätzwert
(Intercept)	7.756
Tonnage in Mt	-6.074
Maschine: B	-0.726
Maschine: C	3.578
Maschine: D	0.680
Stahl: R350	2.800
Zwischenschicht: weich	3.350

Güterzüge verwenden wir für diese Modelle nicht, da aus fachlicher Sicht vermutet wird, dass sie bei kleinen Tonnagen keinen wesentlichen Einfluss hat. Auch die Variable Kurve verwenden wir nicht, da es zu wenige Beobachtungen mit einer Kurve hat.

Forward selection und backward elimination resultieren beide im vollen Modell. Auch lasso ergibt keine sehr grossen Änderungen an den geschätzten Parametern. Insbesondere wird kein Parameter auf Null gesetzt. Wir wählen also das volle Modell

$$\text{Rauheit} \sim \text{Tonnage} + \text{Maschine} + \text{Stahl} + \text{Zwischenlage}$$

mit einem R^2 von 0.81 und einem RMSE von 1.64. Die geschätzten Parameter sind in Tabelle 3.7 für das volle Modell und in Tabelle 3.8 für das lasso aufgeführt.

Das Modell ist viel besser als dasjenige über den gesamten Wertebereich der Tonnage. Das R^2 und somit der Anteil der erklärten Streuung ist fast doppelt so hoch. Auch der Vorhersagefehler ist deutlich kleiner. Der RMSE ist nur rund halb so gross.

Die diagnostischen Plots finden sich im Anhang in den Abbildungen A.29 und A.30. Die beiden Plots sind nicht perfekt, aber weder der Residuenplot noch der Q-Q-Plot zeigen dramatische Abweichungen von den Annahmen.

3.5.3. Modell für grosse Tonnagen

Für die Modelle für Tonnagen grösser als 750000 entfernen wir alle Beobachtungen mit Stahlqualität R320Cr bzw. Schwellentyp Stahl. Da es wiederum zu wenige Beobachtungen mit diesen Ausprägungen hat. Die Variable Maschine verwenden wir für diese Modelle nicht, da aus fachlicher Sicht vermutet wird, dass sie bei grossen Tonnagen keinen wesentlichen Einfluss mehr hat.

Wir erhalten also das volle Modell

$$\text{Rauheit} \sim \text{Tonnage} + \text{Stahl} + \text{Zwischenlage} + \text{Anteil.GZ} + \text{Schwellentyp} + \text{Kurve}$$

mit einem R^2 von 0.33 und einem RMSE von 2.7. Die geschätzten Parameter sind in Tabelle 3.10 aufgeführt.

Tabelle 3.9.: Modellvergleich für grosse Tonnagen

Modell							RMSE	R^2
voll	Tonnage	Stahl	Zwischenlage	Anteil.GZ	Schwellentyp	Kurve	2.70	0.33
forward		Stahl		Anteil.GZ			2.60	0.24
backward	Tonnage	Stahl				Kurve	2.63	0.26
lasso	Tonnage	Stahl	Zwischenlage	Anteil.GZ	Schwellentyp*	Kurve	2.69	

* Lasso setzt den Parameter für die Holzschwelle auf Null.

Mit der forward selection erhalten wir das Modell

$$\text{Rauheit} \sim \text{Stahl} + \text{Anteil.GZ}$$

mit einem R^2 von 0.24 und einem RMSE von 2.6. Tabelle 3.11 zeigt die geschätzten Parameter.

Die backward elimination liefert das Modell

$$\text{Rauheit} \sim \text{Tonnage} + \text{Stahl} + \text{Kurve}$$

mit einem R^2 von 0.26 und einem RMSE von 2.63. Die geschätzten Parameter sind in Tabelle 3.12 aufgeführt.

Lasso wirft nur den Schwellentyp Holz raus, was aus fachlicher Sicht unschön ist. Die anderen Schwellentypen und auch alle anderen Variablen bleiben im Modell. Der RMSE liegt bei 2.69 also nicht wesentlich tiefer als im vollen Modell. Deshalb würden wir hier lasso nicht verwenden. Die geschätzten Parameter für das lasso sind in Tabelle 3.13 aufgeführt.

In Tabelle 3.9 sind alle Modelle für grosse Tonnagen mit den gewählten Variablen aufgeführt und die entsprechenden RMSE und R^2 zusammengefasst.

Die Kenngrößen für die Modelle mit forward bzw. backward selection sind sehr ähnlich. Das R^2 ist jeweils nicht viel tiefer als im vollen Modell, obwohl sie weniger Erklärende haben. Der RMSE ist leicht tiefer als im vollen Modell. Die im Anhang in den Abbildungen A.31 bis A.36 aufgeführten diagnostischen Plots sind für alle Modelle vergleichbar. Sie liefern also kein Argument, um ein Modell zu bevorzugen.

Für die Definition der Messstrategie in Abschnitt 5 verwenden wir das Modell der backward elimination. Man könnte sich jedoch auch für das Modell der forward selection entscheiden.

Beide Modelle sind aus fachlicher Sicht sehr interessant. Für grosse Tonnagen scheint vor allem die Stahlqualität eine wichtige Rolle zu spielen. Interessant ist, dass das Modell der forward selection die Tonnage nicht mehr enthält. Man könnte das so interpretieren, dass die Rauheit am Anfang mit steigender Tonnage stark abnimmt, aber nach einer gewissen Zeit zusätzliche Radsatzlasten keinen (grossen) Einfluss mehr auf die Rauheit haben. Andere Faktoren wie der Anteil Güterzüge spielen dann eine grössere Rolle. Das Modell der backward elimination suggeriert, dass die Tonnage auch bei grossen Werten weiterhin einen kleinen Einfluss auf die Rauheit hat.

Das Modell für grosse Tonnagen ist schlechter als dasjenige für kleine. Das R^2 ist deutlich tiefer und der RMSE deutlich höher. Zusammen mit der Tatsache, dass deutlich weniger Erklärende gewählt wurden, würden wir das so interpretieren, dass die Rauheit für grosse Tonnagen weniger gut mit den betrachteten Variablen erklärt werden kann, als für kleine Tonnagen. Es spielen also andere Faktoren eine Rolle, die wir bisher nicht betrachtet haben und die man eventuell nur schwer oder gar nicht kontrollieren kann.

Andererseits ist die geschätzte Rauheit für grosse Tonnagen viel kleiner als für kleine. Deswegen ist es auch weniger wichtig, sie gut vorhersagen zu können.

Die in diesem Abschnitt gemachten Aussagen sind sehr vorsichtig zu interpretieren. Die vorhandenen Daten suggerieren zwar, dass die genannten Variablen einen Einfluss auf die Rauheit haben. Mit der verwendeten Datenlage kann man dazu aber keine abschliessende Aussage machen. Insbesondere kann man mit diesen Daten sicher nicht zeigen, dass die anderen Variablen keinen Einfluss haben auf die Rauheit haben.

Tabelle 3.10.: Koeffizienten für das volle Modell für grosse Tonnagen

	Schätzwert	Standardfehler
(Intercept)	0.809	1.556
Tonnage in Mt	-0.025	0.028
Stahl: R350	2.534	1.098
Zwischenschicht: weich	-1.330	1.010
Anteil GZ: klein	0.936	1.189
Schwelle: B91	1.512	1.060
Schwelle: Holz	-0.672	1.517
Kurve: nein	1.965	1.617

Tabelle 3.11.: Koeffizienten für das Modell der forward selection für grosse Tonnagen

	Schätzwert	Standardfehler
(Intercept)	2.408	0.690
Stahl: R350	3.117	0.928
Anteil GZ: klein	1.609	0.790

Tabelle 3.12.: Koeffizienten für das Modell der backward elimination für grosse Tonnagen

	Schätzwert	Standardfehler
(Intercept)	2.000	1.291
Tonnage in Mt	-0.041	0.024
Stahl: R350	3.438	1.014
Kurve: nein	2.294	1.296

Tabelle 3.13.: Koeffizienten für das lasso für grosse Tonnagen

	Schätzwert
(Intercept)	2.910
Tonnage in Mt	-0.005
Stahl: R350	1.937
Zwischenschicht: weich	-0.015
Anteil GZ: klein	0.641
Schwelle: B91	0.653
Schwelle: Holz	0.000
Kurve: nein	0.014

4. Schätzung der Rauheit

In diesem Abschnitt stellen wir die Schätzung der Rauheit dar. Als Mass für die Unsicherheit geben wir 95%-Prädiktionsintervalle an. Für *eine* neue Beobachtung mit den gegebenen Ausprägungen gehen wir davon aus, dass sie mit einer Sicherheit von 95% in diesem Intervall enthalten ist. Im Gegensatz zu 95%-Konvergenzintervallen, die den wahren Mittelwert, gegeben die Ausprägungen der Erklärenden, mit einer Sicherheit von 95% enthalten.

Laut dem Artikel 7 der Verordnung über die Lärmsanierung der Eisenbahnen (VLE) Müssen die Infrastrukturbetreiberinnen die Schienenrauheit ihrer Anlagen überwachen und gewährleisten, dass die Rauheit nicht höher als 10dB ist. Deshalb geben wir zusätzlich zu den Prädiktionsintervallen die geschätzte Wahrscheinlichkeit an, dass eine zukünftige Beobachtung grösser als 10 ist.

Alle Schätzungen in diesem Abschnitt sind mit Vorsicht zu interpretieren. Dies ist eine explorative Vorstudie mit dem Ziel, Hypothesen zu generieren und neue Messstandorte zu definieren. Es soll eine erste Abschätzung der Rauheit für das SBB-Schiennetz und der diesbezüglichen Unsicherheit gemacht werden. Die angegebenen Prädiktionsintervalle und Wahrscheinlichkeiten sollen in erster Linie illustrieren, mit welcher Unsicherheit die Schätzungen behaftet sind. Sogar für eine Tonnage von 50 Million gibt es eine Parameterkombination, für die das Prädiktionsintervall die 10 enthält. Insbesondere empfehlen wir nicht, daraus abzuleiten, dass die VLE für bestimmte Ausprägungen der Erklärenden nicht erfüllt ist. Zum Teil bräuchte man mehr Beobachtungen, um eine verlässliche Aussage zu machen. Zudem müsste man berücksichtigen, ob die Parameterkombination überhaupt relevant ist. Es gibt Kombinationen, die in dicht besiedelten Gebieten gar nicht vorkommen oder nur auf einer vernachlässigbaren Schienenlänge.

Die Tabellen 4.1 bis 4.5 zeigen die Schätzwerte für bestimmte Tonnagewerte und alle möglichen Kombinationen der Ausprägungen für die kategorialen erklärenden Variablen. Die Schätzungen des Modells für kleine Tonnagen sind in den Tabellen 4.1 für Tonnage 0 und 4.2 für Tonnage 750000 abgebildet. Die Schätzungen des Modells für grosse Tonnagen sind in den Tabellen 4.3 für Tonnage 1 Million, 4.4 für Tonnage 10 Millionen und 4.5 für Tonnage 50 Millionen abgebildet.

In den Abbildungen 4.1 bis 4.4 sind Prädiktionsintervalle für ausgewählte Parameterkombinationen aufgezeichnet. Die ersten beiden Intervalle sind für kleine Tonnagen, die dritte und vierte für grosse Tonnagen. Abbildung 4.2 zeigt ein Prädiktionsintervall für eine Kombination, für die wir noch keine Messungen haben.

Tabelle 4.1.: Schätzungen für Tonnage 0

Maschine	Stahl	Zwischenlage	Rauheit					
			Schätzwert	P(Rauheit) > 10	Prädiktionsintervall			
					von	bis	Länge	
A	R260	hart	7.59	11.9%	3.59	11.60	8.01	
A	R260	weich	11.10	74.9%	7.89	14.31	6.42	
A	R350	hart	10.45	58.7%	6.44	14.46	8.01	
A	R350	weich	13.96	97.9%	10.15	17.77	7.62	
B	R260	hart	7.02	3.4%	3.81	10.24	6.43	
B	R260	weich	10.53	59.7%	6.32	14.75	8.43	
B	R350	hart	9.88	47.5%	6.07	13.69	7.62	
B	R350	weich	13.39	90.3%	8.28	18.50	10.23	
C	R260	hart	11.34	78.5%	8.02	14.66	6.64	
C	R260	weich	14.85	99.1%	10.81	18.88	8.07	
C	R350	hart	14.20	99.3%	10.87	17.52	6.64	
C	R350	weich	17.70	100.0%	13.18	22.23	9.05	
D	R260	hart	8.33	18.7%	4.64	12.02	7.38	
D	R260	weich	11.84	87.5%	8.70	14.98	6.28	
D	R350	hart	11.19	73.6%	7.50	14.88	7.38	
D	R350	weich	14.70	99.3%	10.95	18.44	7.50	

Tabelle 4.2.: Schätzungen für Tonnage 750'000

Maschine	Stahl	Zwischenlage	Rauheit					
			Schätzwert	P(Rauheit) > 10	Prädiktionsintervall			
					von	bis	Länge	
A	R260	hart	2.98	0.1%	-1.43	7.40	8.83	
A	R260	weich	6.49	2.0%	3.15	9.84	6.69	
A	R350	hart	5.84	3.2%	1.43	10.25	8.83	
A	R350	weich	9.35	37.3%	5.43	13.27	7.84	
B	R260	hart	2.41	0.0%	-1.05	5.88	6.93	
B	R260	weich	5.92	2.6%	1.82	10.03	8.21	
B	R350	hart	5.27	1.1%	1.25	9.29	8.05	
B	R350	weich	8.78	31.7%	3.76	13.80	10.05	
C	R260	hart	6.73	3.8%	3.13	10.33	7.20	
C	R260	weich	10.24	54.7%	6.28	14.19	7.91	
C	R350	hart	9.58	40.9%	5.98	13.19	7.20	
C	R350	weich	13.09	91.3%	8.64	17.55	8.91	
D	R260	hart	3.72	0.2%	-0.44	7.89	8.33	
D	R260	weich	7.23	5.1%	3.91	10.55	6.64	
D	R350	hart	6.58	5.4%	2.41	10.74	8.33	
D	R350	weich	10.09	51.8%	6.19	13.99	7.80	

Tabelle 4.3.: Schätzungen für Tonnage 1 Million

Stahl	Kurve	Rauheit		Prädiktionsintervall		
		Schätzwert	P(Rauheit) > 10	von	bis	Länge
R260	ja	1.96	0.3%	-3.75	7.67	11.42
R260	nein	4.25	1.5%	-0.94	9.44	10.38
R350	ja	5.40	5.5%	-0.23	11.03	11.26
R350	nein	7.69	20.3%	2.25	13.13	10.88

Tabelle 4.4.: Schätzungen für Tonnage 10 Millionen

Stahl	Kurve	Rauheit		Prädiktionsintervall		
		Schätzwert	P(Rauheit) > 10	von	bis	Länge
R260	ja	1.59	0.2%	-4.11	7.29	11.41
R260	nein	3.88	1.0%	-1.27	9.04	10.30
R350	ja	5.03	4.3%	-0.63	10.68	11.31
R350	nein	7.32	16.7%	1.89	12.75	10.86

Tabelle 4.5.: Schätzungen für Tonnage 50 Millionen

Stahl	Kurve	Rauheit		Prädiktionsintervall		
		Schätzwert	P(Rauheit) > 10	von	bis	Länge
R260	ja	-0.05	0.1%	-6.12	6.01	12.13
R260	nein	2.24	0.2%	-3.17	7.65	10.82
R350	ja	3.38	1.7%	-2.75	9.52	12.27
R350	nein	5.68	7.2%	-0.12	11.48	11.60

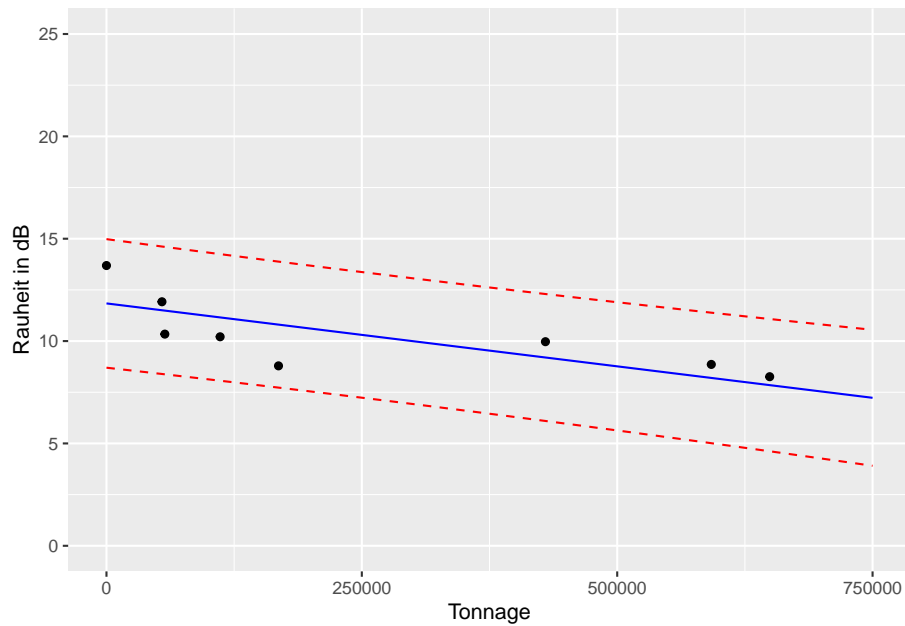


Abbildung 4.1.: Regressionsgerade (blau) und Prädiktionsintervall (rot) für kleine Tonnagen, Maschine D, Stahl R260 und eine weiche Zwischenlage.

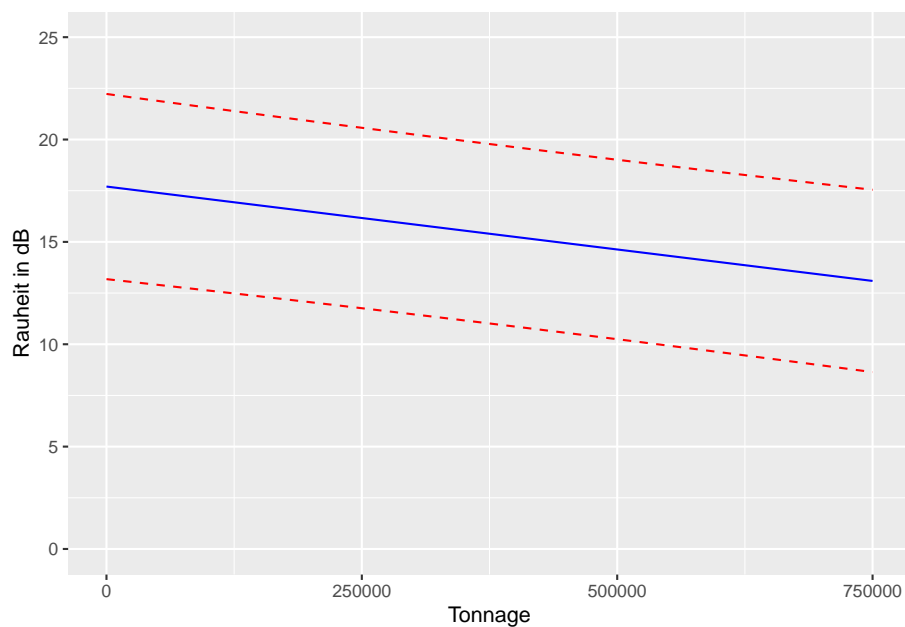


Abbildung 4.2.: Regressionsgerade (blau) und Prädiktionsintervall (rot) für kleine Tonnagen, Maschine C, Stahl R350 und eine weiche Zwischenlage.
Für diese Parameterkombination liegen noch keine Messwerte vor.

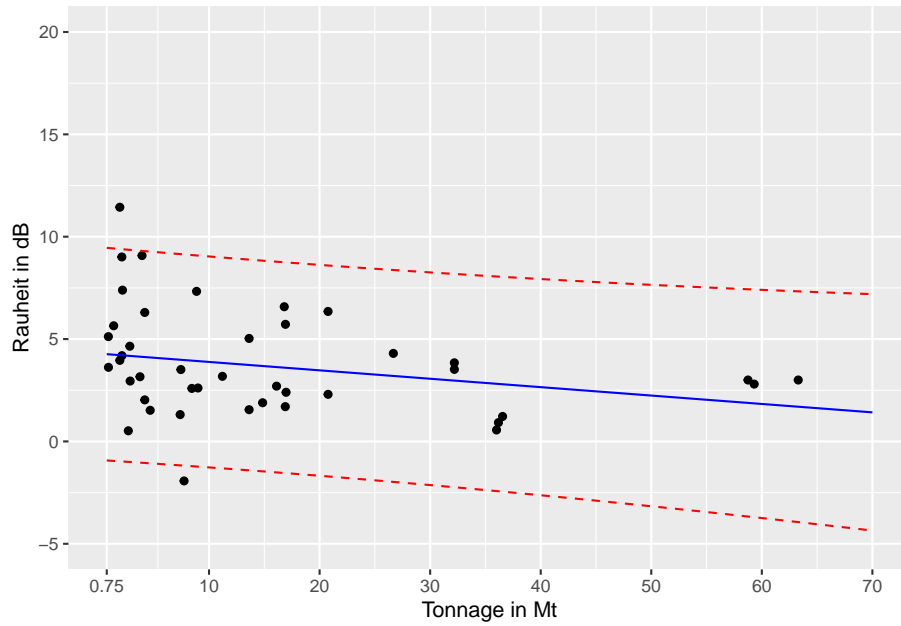


Abbildung 4.3.: Regressionsgerade (blau) und Prädiktionsintervall (rot) für grosse Tonnagen, Stahl R260 und ohne Kurve.

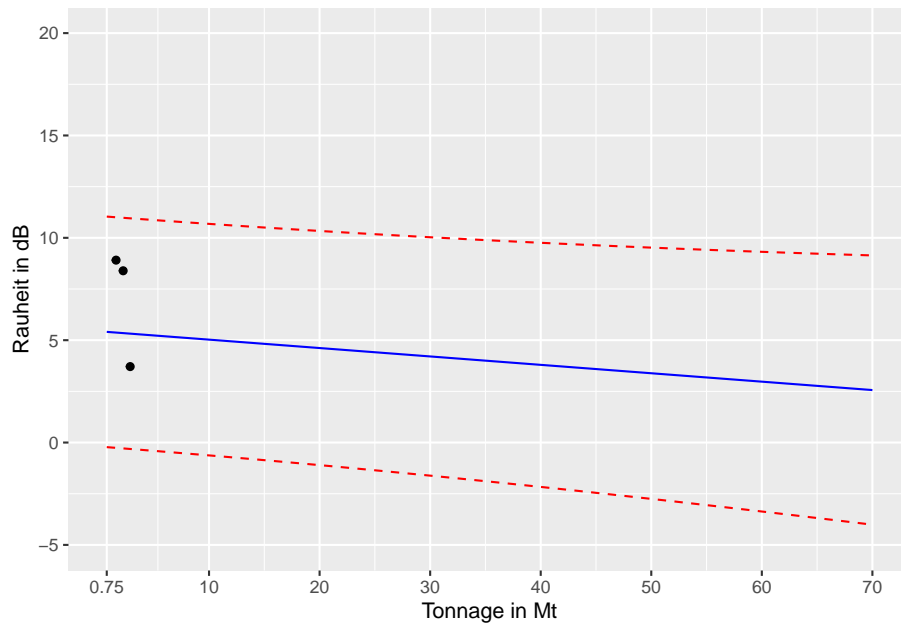


Abbildung 4.4.: Regressionsgerade (blau) und Prädiktionsintervall (rot) für grosse Tonnagen, Stahl R350 und mit Kurve.

5. Messstrategie für das Jahr 2021

In diesem Abschnitt skizzieren wir eine mögliche Messstrategie für das Jahr 2021. Wir sehen drei Varianten, wie man die zukünftigen Messstandorte auswählen und die Daten auswerten könnte:

1. Wir verwenden die bereits erhobenen Daten nur für das „design of experiments“. Das heisst wir definieren damit den Parameterraum und erstellen ein „space filling design“. Wir suchen also Punkte im Parameterraum, die den Raum möglichst gleichmässig abdecken. Die Anzahl Punkte wird dabei durch die Anzahl Messungen definiert, die im Jahr 2021 voraussichtlich gemacht werden können.
Für die Auswertung, also die Parameterschätzung und die Vorhersage der Rauheit, werden die bisherigen Daten dann nicht mehr verwendet. Die endgültigen Modelle und Vorhersagen basieren in diesem Fall also nur auf den Daten, die ab dem Jahr 2021 gemäss Versuchsplan erhoben werden.
2. Wir machen ein Design analog zur ersten Variante. Jedoch mit dem Unterschied, dass wir die bereits erhobenen Daten auch für die Auswertung weiterverwenden.
Das „space filling design“ berücksichtigt in diesem Fall die aktuellen Daten. Die neuen Messstandorte werden möglichst in Bereichen des Parameterraums gesetzt, in denen wir noch keine Beobachtungen haben.
3. Wir versuchen nicht, den ganzen Parameterraum abzutasten. Wir machen also kein „space filling design“. Die neuen Messungen machen wir an „interessanten“ Stellen, also in Bereichen des Parameterraums, an denen die Modelle hohe Schätzwerte für die Rauheit liefern oder an denen die Schätzunsicherheit gross ist. Als weiteres Kriterium könnten wir betrachten, wie gross der Anteil des Schienennetzes ist, der diesen Parametern entspricht.

Die erste Variante wäre ideal. Damit hätten wir eine prospektive Studie und die mit den Daten gemachten Aussagen wären entsprechend verlässlich. Da die Messungen sehr aufwendig und teuer sind, wird es aber wahrscheinlich nicht realistisch sein, dass wir genügend neue Daten erheben und auf die bereits erhobenen Daten verzichten können.

Die zweite Variante ist aus unserer Sicht nur bedingt sinnvoll. Sie liefert kein „sauberes“ Design und setzt die Messressourcen nicht optimal ein.

Wir empfehlen die dritte Variante. Sie führt zwar nicht zu einer prospektiven Studie, aber die verfügbaren Ressourcen können damit optimal eingesetzt werden. Mit einem iterativen Vorgehen können wir so die Schätzung der Rauheit laufend verbessern. Anfangs Jahr werden nach den weiter unten definierten Kriterien erste Messstandorte definiert und die entsprechenden Messungen durchgeführt. So bald die Messresultate vorliegen, können die Modelle angepasst und damit neue Messstandorte definiert werden. Dies kann dann wiederholt werden bis die Qualität der Schätzung den Erwartungen entspricht.

5.1. Messstandorte für kleine Tonnagen

Für kleine Tonnagen lautet die Modellgleichung

$$\text{Rauheit} \sim \text{Tonnage} + \text{Maschine} + \text{Stahl} + \text{Zwischenlage.}$$

Es sind also folgende Parameter zu betrachten:

- Tonnage hat als numerische Grösse nur einen Parameter. Wir betrachten die Werte von 0 bis 750000.

Tabelle 5.1.: Rangliste für kleine Tonnagen

Maschine	Stahl	Zwischenlage	Schätzwert	P(Rauheit) > 10	Rauheit		
					von	bis	Länge
C	R350	weich	17.70	100.0%	13.18	22.23	9.05
C	R260	weich	14.85	99.1%	10.81	18.88	8.07
D	R350	weich	14.70	99.3%	10.95	18.44	7.50
C	R350	hart	14.20	99.3%	10.87	17.52	6.64
B	R350	weich	13.39	90.3%	8.28	18.50	10.23
B	R260	weich	10.53	59.7%	6.32	14.75	8.43
A	R350	hart	10.45	58.7%	6.44	14.46	8.01
A	R260	hart	7.59	11.9%	3.59	11.60	8.01

- Maschine hat die vier Ausprägungen A, B, C und D. (Maschine F berücksichtigen wir nicht)
- Stahlqualität hat die drei Ausprägungen R260, R350 und R320Cr.
- Schienenzwischenlage hat die Ausprägungen keine, weich und hart

Als mögliche neue Messstandorte kommen Stellen infrage, die eine Parameterkombination mit hoher geschätzter Rauheit oder hoher Schätzunsicherheit, also langen Prädiktionsintervallen, aufweisen. Tabelle 5.1 zeigt solche Parameterkombinationen. Sie enthält alle Kombinationen mit einer geschätzten Rauheit grösser als 14 oder einem Prädiktionsintervall mit einer Länge grösser als 8. Die Werte sind absteigend nach geschätzter Rauheit sortiert. Die Tabelle kann als Rangliste aufgefasst werden. Wir würden also die erste Kombination als am interessantesten für zusätzliche Messungen einschätzen.

Die gezeigten Schätzwerte sind jeweils für Tonnage 0. Da für sehr kleine Tonnagen die geschätzte Rauheit und die Unsicherheit jeweils am grössten sind, wären Messungen mit sehr kleinen Tonnagen interessant. Es wäre aber trotzdem sinnvoll, auch Messungen mit höheren Tonnagen zu machen.

Stahlqualität R320Cr, „keine Schienenzwischenlage“, Stahl- und Holzschwellen werden wegen zu wenigen bzw. gar keinen Beobachtungen vom Modell nicht berücksichtigt. Falls man über diese Ausprägungen Aussagen machen möchte, müssten entsprechende Messungen durchgeführt werden.

Die aufgelisteten Kombinationen sind als Vorschläge zu verstehen, die jedoch noch fachlich beurteilt werden müssen. Es stellt sich die Frage, ob sie überhaupt relevant sind oder ob sie gar nicht oder nur sehr selten im SBB-Schienennetz vorkommen.

Für relevante Parameterkombinationen könnte die SBB in einem nächsten Schritt konkrete Standorte suchen, an denen Messungen durchgeführt werden könnten. Die gefundenen Standorte könnte dann das IODA zusammen mit der SBB priorisieren. Als Kriterien würden die Verteilungen der Tonnage und der anderen, im Modell nicht berücksichtigten, Erklärenden verwendet werden. So könnte man sicherstellen, dass die Tonnage und die eliminierten Variablen in den Messungen variieren.

5.2. Messstandorte für grosse Tonnagen

Für grosse Tonnagen lautet die Modellgleichung

$$\text{Rauheit} \sim \text{Tonnage} + \text{Stahl} + \text{Kurve}$$

Es sind also folgende Parameter zu betrachten:

- Tonnage hat als numerische Grösse nur einen Parameter. Wir betrachten die Werte von 750000 bis 50 Millionen.
- Stahlqualität hat die drei Ausprägungen R260, R350 und R320Cr.
- Kurve hat die Ausprägungen ja und nein

Tabelle 5.2.: Rangliste für grosse Tonnagen

Stahl	Kurve	Schätzwert	P(Rauheit) > 10	Rauheit		
				Prädiktionsintervall		
				von	bis	Länge
R350	nein	7.69	20.3%	2.25	13.13	10.88
R350	ja	5.40	5.5%	-0.23	11.03	11.26

Als mögliche neue Messstandorte kommen wiederum Stellen infrage, die eine Parameterkombination mit hoher geschätzter Rauheit oder hoher Schätzunsicherheit, also langen Prädiktionsintervallen, aufweisen. In diesem Modell unterscheiden sich die Schätzunsicherheiten jedoch nicht stark. Deshalb betrachten wir nur die geschätzte Rauheit.

Tabelle 5.2 zeigt Parameterkombinationen mit einer geschätzten Rauheit grösser als 5. Die Werte sind absteigend nach geschätzter Rauheit sortiert. Die Tabelle kann wiederum als Rangliste aufgefasst werden.

Stahlqualität R320Cr, „keine Schienenzwischenlage“ und Stahlschwellen werden wegen zu wenigen bzw. gar keinen Beobachtungen von diesem Modell nicht berücksichtigt. Falls man über diese Ausprägungen Aussagen machen möchte, müssten entsprechende Messungen durchgeführt werden.

Das Vorgehen wäre dann analog zu demjenigen für kleine Tonnagen. Es müsste noch entschieden werden, ob der Fokus auf kleine oder grosse Tonnagen gelegt werden soll, oder ob beide Bereiche weiter untersucht werden sollen.

Für kleine Tonnagen ist die Rauheit und ihre Abhängigkeit von der Tonnage bekanntlich viel grösser als für grosse Tonnagen. Die Rauheit scheint für kleine Tonnagen auch von mehr Erklärenden abzuhängen. Deshalb wäre es interessant, vor allem Messungen für kleine Tonnagen zu machen, um den wahren Mechanismus besser zu verstehen. Andererseits decken die grossen Tonnagen einen viel grösseren Zeitraum ab. Der grösste Teil des Schienennetzes ist also jeweils im Bereich grosser Tonnagen. Das spricht dafür, dass dieser Bereich wichtiger ist und die Messungen dementsprechend vor allem für grosse Tonnagen gemacht werden sollten. Die Streuung ist auch im Modell für grosse Tonnagen recht gross. Die Prädiktionsintervalle enthalten den Wert 10. Mit mehr Messungen könnten die Intervalle verkleinert werden, so dass sie den Wert 10 hoffentlich nicht mehr enthalten.

5.3. Wiederholte Messungen am gleichen Standort

Die bereits erhobenen Daten enthalten zum Teil mehrere Messungen am gleichen Standort. Dies führt dazu, dass die Beobachtungen nicht unabhängig sind und somit die Voraussetzungen des linearen Modells eigentlich nicht erfüllt. Für solche Daten sollten eigentlich gemischte Modelle verwendet werden, die mit diesen Varianzstrukturen umgehen können.

Für zukünftige Messungen ist es wichtig, dass an jedem Standort nur eine Messung durchgeführt wird, damit die Beobachtungen unabhängig sind.

Man könnte sich überlegen, von den bereits erhobenen Daten nur eine Beobachtung pro Messstandort zu verwenden. So wäre die Unabhängigkeit der Beobachtungen wieder gegeben.

6. Messunsicherheit

In diesem Abschnitt soll darauf hingewiesen werden, dass nebst der Unsicherheit des statistischen Modells, unter Umständen die Unsicherheit einer Messung an sich für die Praxis relevant sein kann. Insbesondere bei Lärmmessungen liegt die Vermutung nahe, dass eine grosse Unsicherheit bezüglich eines einzelnen Messwerts besteht und es sich lohnen würde, die Messunsicherheit von solchen Lärmmessungen zu bestimmen. In Bezug auf die Schienenrauheit wurden seitens SBB bereits Testmessungen durchgeführt um die Variabilität von Rauheitsmessungen grob abschätzen zu können. Für eine valide Schätzung der kompletten Messunsicherheit, müssten jedoch noch weitere systematische Messungen durchgeführt werden, als auch die sogenannte systematische Messunsicherheit, welche durch das Messsystem selbst definiert ist, genauer untersucht werden.

Es liegt in der Natur einer Messung, dass sie nicht beliebig genau sein kann. Es hat nicht unbedingt mit einem Fehlverhalten oder einer falschen Messung zu tun, dass ein Ergebnis vom „wahren“ Wert abweicht. Deshalb unterscheidet man bei **Messabweichungen** zwischen **Messfehlern** und **Messunsicherheiten**. Um einen Messfehler handelt es sich, wenn eine Abweichung vom wahren Wert auf Versagen des Messsystems zurückzuführen ist.

Systematische Messabweichungen (auch **Bias** genannt) haben einen konstanten Betrag mit einem bestimmten Vorzeichen, oder unterliegen nur einer sehr langsamen Veränderung aufgrund einer Ursache, welche die Messgrösse determiniert verändert. Sie kann daher einerseits zu einer immer gleichen, zeitlich konstanten Differenz des Messwerts vom wahren Wert führen, oder aber die Differenz kann auch abhängig von der Zeit sich verändern. Systematische Messfehler sind durch Wiederholungen der Messungen unter gleichen Bedingungen nicht erkennbar. Eine konstante systematische Abweichung kann aber mittels geeigneter Kalibrierung vernachlässigbar klein gemacht werden. Hingegen ist eine von der Zeit abhängige systematische Abweichung viel schwieriger in den Griff zu bekommen. Die Richtigkeit einer Messung ist also durch den systematischen Messfehler bestimmt.

Die Präzision einer Messung hingegen ist durch die **Messunsicherheit** gegeben. Sie besitzt gemäss [10] eine systematische, wie auch eine zufällige Komponente:

- Die **systematische Messunsicherheit** ist die sogenannte **Typ B Unsicherheit**. Diese Unsicherheit bei einer Messung entsteht durch die Ungenauigkeit und Bauart eines Messgeräts. Geht man von einem perfekt kalibrierten Messsystem aus, so ist die Typ B Unsicherheit lediglich von der Genauigkeit der Kalibrierung abhängig. Auch diese Unsicherheit lässt sich *nicht* statistisch messen. Die Typ B Unsicherheit lässt sich i.A. über die Gerätespezifikation bestimmen.
- Andererseits entstehen **zufällige Messunsicherheiten** aufgrund nicht beherrschbarer, nicht determinierter Einflüsse während der Messung (siehe [10]). Sie sind nicht vorausbestimmbar. Wird eine Messung theoretisch unter gleichen Bedingungen wiederholt, führen sie zu einer Streuung der Messwerte, d.h. zu einem unsicheren Messergebnis. Die sogenannte **Typ A Unsicherheit** einer Messung kann dann über die Variabilität der Messwerte geschätzt werden.

Diese beiden Messunsicherheiten werden gemäss Gauss'schem Fehlerfortpflanzungsgesetz wie folgt kombiniert:

$$\text{Messunsicherheit} = \sqrt{(\text{Typ A})^2 + (\text{Typ B})^2}$$

In der Praxis ist eine Trennung von zum Teil unbekanntem systematischen und zufälligen Abweichungen häufig sehr schwer bis unmöglich (siehe [11]).

Das Messergebnis zusammen mit der Messunsicherheit bildet das vollständige Messergebnis. Abbildung 6.1 illustriert den Einfluss der unterschiedlichen Messunsicherheiten auf das Messergebnis.

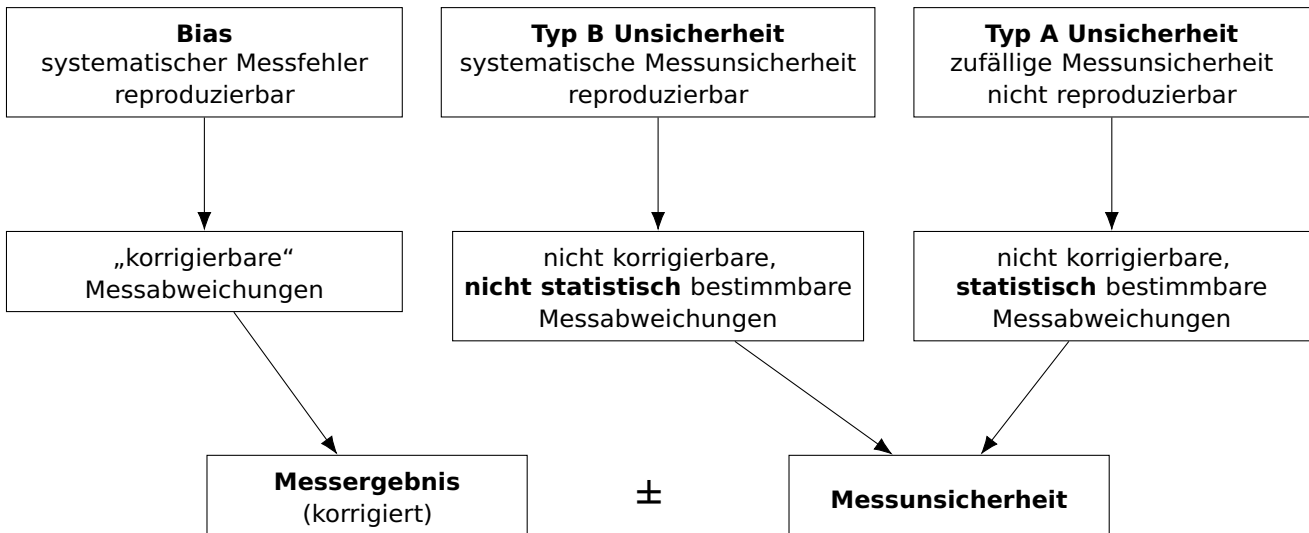


Abbildung 6.1.: Eine Zusammenstellung der wichtigsten Informationen bezüglich Messfehler und Messunsicherheiten. Das Messergebnis zusammen mit der Messunsicherheit bildet das vollständige Messergebnis.

Der Bias wird hierbei als „korrigierbar“ angegeben. Diese Aussage ist jedoch mit Vorsicht zu interpretieren. Einen konstanten Bias kann wie oben erwähnt mit Hilfe einer geeigneten Kalibrierung korrigiert werden. Bei einem Bias, welcher von äusseren Bedingungen (z.B. Temperatur, Wetter) abhängig ist, oder welcher sich über die Zeit verändert, ist diese Kalibrierung sehr anspruchsvoll bis unmöglich.

Wie zu Beginn dieses Abschnittes bereits erwähnt, hat die SBB bereits Messungen durchgeführt, um eine grobe Abschätzung der Typ A Messunsicherheit zu erhalten. Die SBB führte am 20. Oktober 2020 in Riedtwil 20 Messungen durch. Eine Messung besteht aus vier Teilmessungen à 15cm, welche dann zusammengesetzt werden zu einer Strecke von 60cm. Das Messgerät misst jeweils drei Taster gleichzeitig, welche circa 10mm Abstände haben (senkrecht zur Messrichtung). Die Variabilität der zwanzig Messungen betrug geschätzt ± 0.82 dB. Abbildung 6.2 zeigt zudem exemplarisch die Variabilität der drei Taster im Ortsraum für eine einzige Messung.

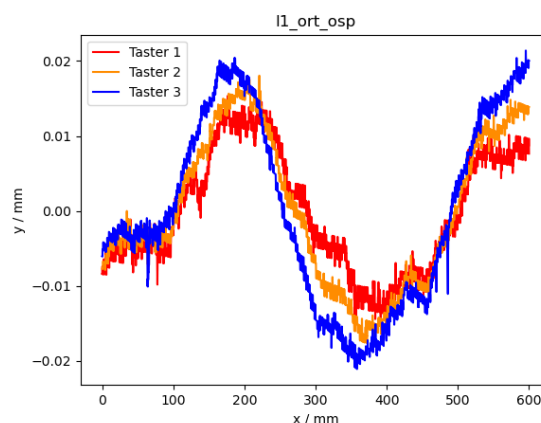


Abbildung 6.2.: Variabilität der drei Taster einer einzigen Rauheitsmessung

7. Ausblick

In diesem Abschnitt skizzieren wir, wie ein Folgeprojekt aussehen könnte. Dabei konzentrieren wir uns auf die dritte Variante aus dem Abschnitt 5. Dies bedeutet, dass die neuen Messungen an „interessanten“ Stellen durchgeführt werden sollen, d.h. in Bereichen des Parameterraums, an denen die Modelle hohe Schätzwerte für die Rauheit liefern oder an denen die Schätzunsicherheit gross ist.

Im Jahr 2021 hat die SBB die Kapazität, ca. 20 Messungen durchzuführen. Es ist klar, dass dies eine sehr limitierte Anzahl Messungen ist und man so nicht die Rauheit des gesamten Schienennetzes mit einer hohen Genauigkeit schätzen kann. Hierzu bräuchte es viel mehr Daten. Optimalerweise würde man die Schienenrauheit mit einem fahrenden Zug messen, so dass kein Streckenunterbruch nötig und der Aufwand für eine Messung minimal wäre. An dieser Möglichkeit wird zur Zeit mit einem ETH-Projekt zur Erfassung der Schienenrauheit mit optischen Methoden gearbeitet. Zum heutigen Zeitpunkt ist aber eine einzige Messung sehr aufwändig, weshalb die Anzahl Messungen limitiert ist. Dennoch scheinen die in diesem Projekt eruierten statistischen Modelle plausibel – auch aus Expertensicht. Es sind klar pragmatische, vereinfachte Modelle, welche aufgrund der kleinen Datenmenge nicht alle Einflussgrössen berücksichtigen können und zudem eine hohe Unsicherheit aufweisen.

In einem Folgeprojekt sollen einerseits zusätzliche Messungen durchgeführt und die neuen Daten analysiert werden. Andererseits soll mit einer Simulationsstudie gezeigt werden, dass diese einfachen statistischen Modelle dennoch genügend robust sind. So kann gezeigt werden, dass trotz geringer Datenmenge solch pragmatische Modelle ausreichen können, die Schienenrauheit mit einer gewissen Unsicherheit genügend genau zu schätzen. Hierzu müssen jedoch auch die im Modell bisher nicht berücksichtigten Einflussgrössen variiert werden.

Im vorliegenden Bericht wurden mögliche Messstandorte vorgeschlagen. Nebst dieser rein datenbasierten Selektion von optimalen Messstandorten, ist natürlich auch die Expertensicht unabdingbar. Aus diesem Grund wird die SBB jeweils die Relevanz der vorgeschlagenen Parameterkombinationen beurteilen und konkrete Standorte definieren, an denen Messungen durchgeführt werden sollen. Das IODA und die SBB priorisieren anschliessend gemeinsam die gefundenen Standorte basierend auf den Ergebnissen der Datenanalyse, sowie den Erfahrungswerten der Experten.

Es ist angedacht, die Messungen im Jahr 2021 in mehreren Iterationen durchzuführen. Nach jeder Iteration werden die Modelle den neuen Daten angepasst und basierend auf den daraus gewonnenen Erkenntnissen werden neue Messstandorte definiert. Nach der letzten Iteration wird, basierend auf allen gemessenen Daten, eine Simulationsstudie durchgeführt, um die Robustheit der Modelle abzuschätzen.

In Abbildung 7.1 ist eine möglicher Zeitplan aufgezeichnet. Dieser Zeitplan ist als erster Entwurf zu verstehen, der mit allen beteiligten verfeinert werden müsste.

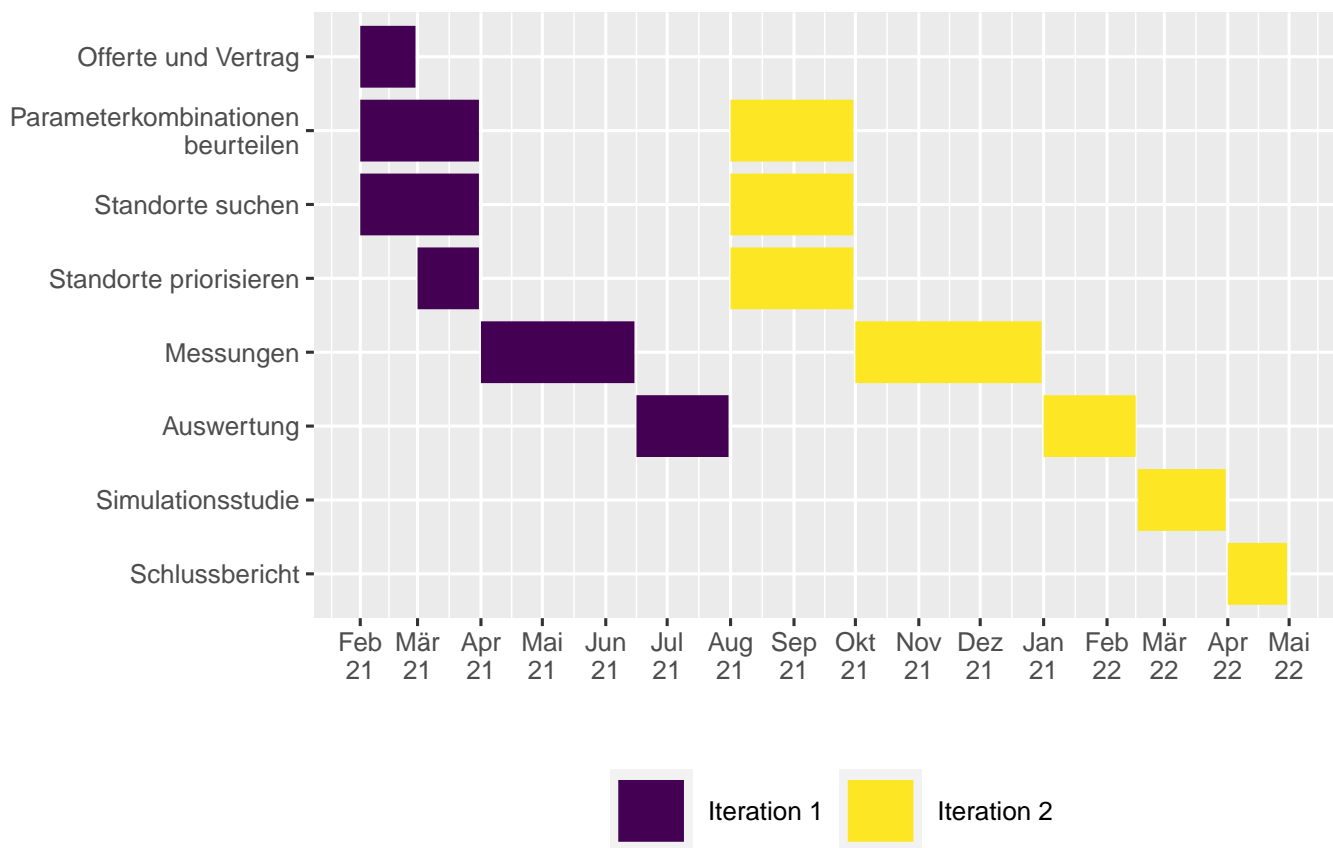


Abbildung 7.1.: Möglicher Zeitplan für ein Folgeprojekt.

A. Anhang

A.1. Abbildungen

A.1.1. Beschreibung der einzelnen Variablen

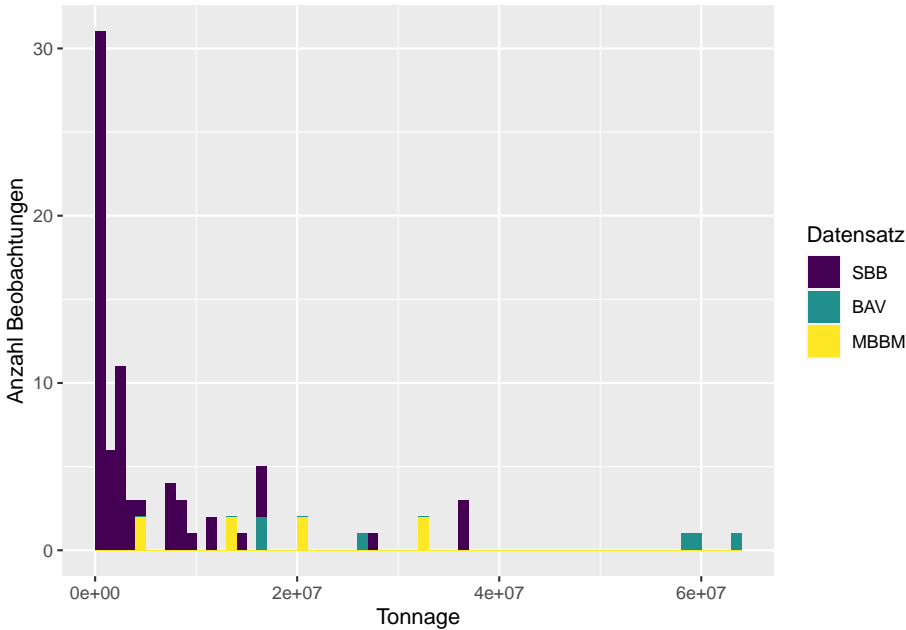


Abbildung A.1.: Histogramm der Tonnage.

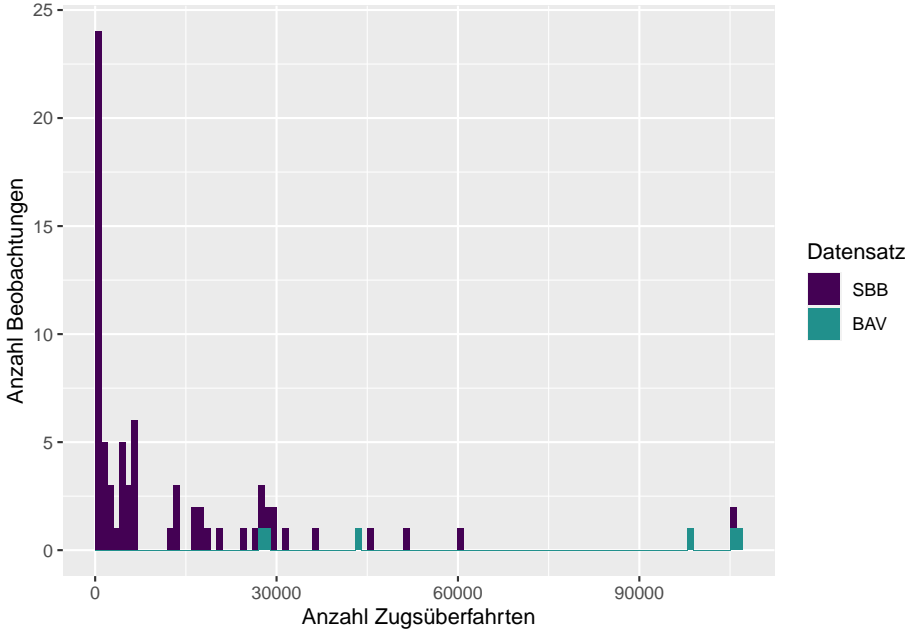


Abbildung A.2.: Histogramm der Anzahl Zugüberfahrten.

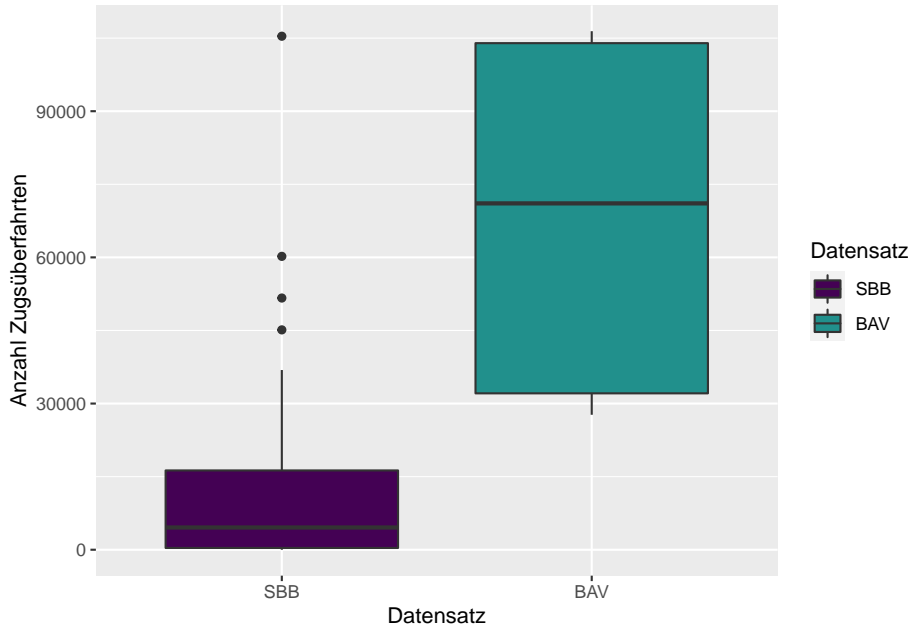


Abbildung A.3.: Boxplot der Anzahl Zugüberfahrten pro Datensatz.

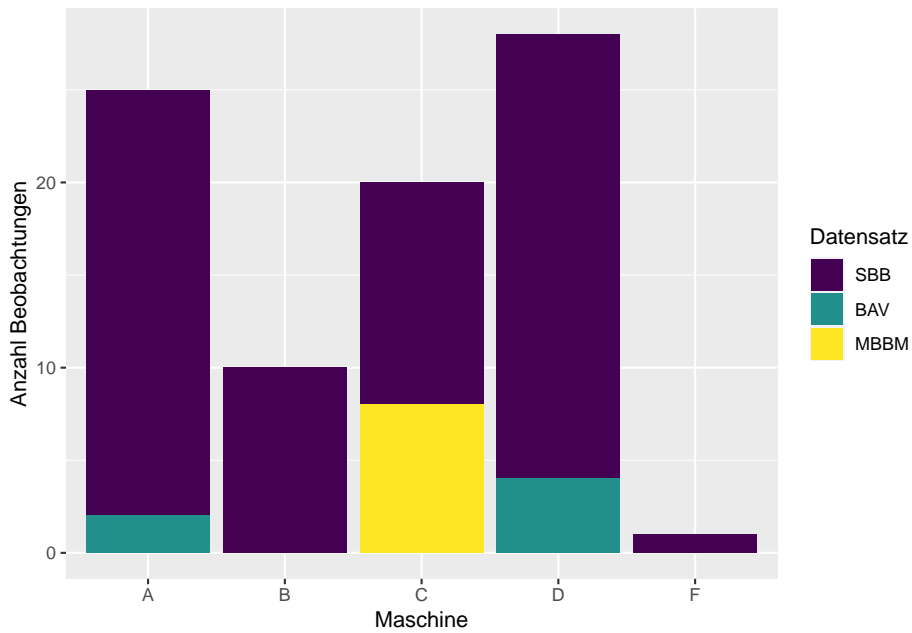


Abbildung A.4.: Säulendiagramm der Beobachtungen pro Maschine und Datensatz.

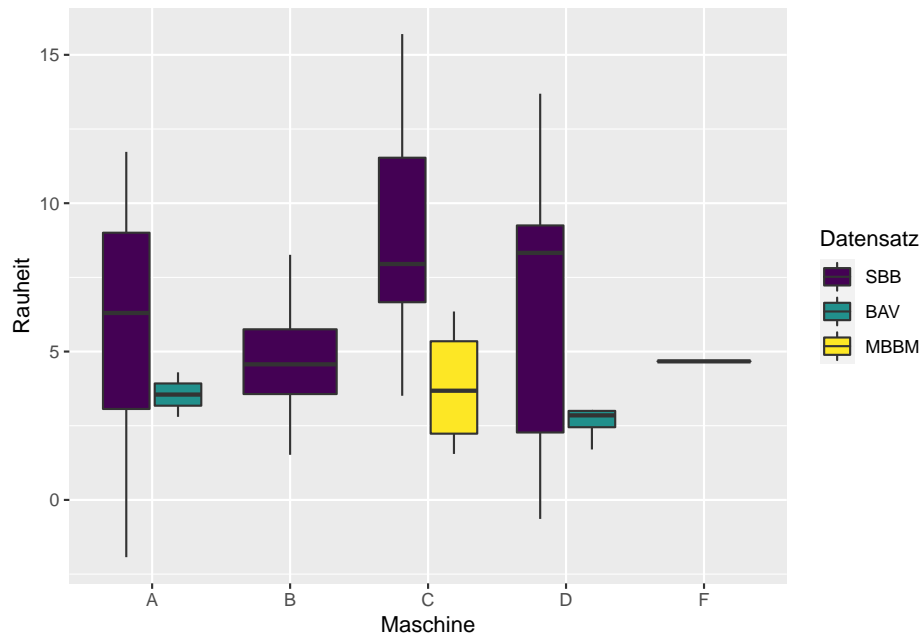


Abbildung A.5.: Boxplots der Rauheit pro Maschine und Datensatz.

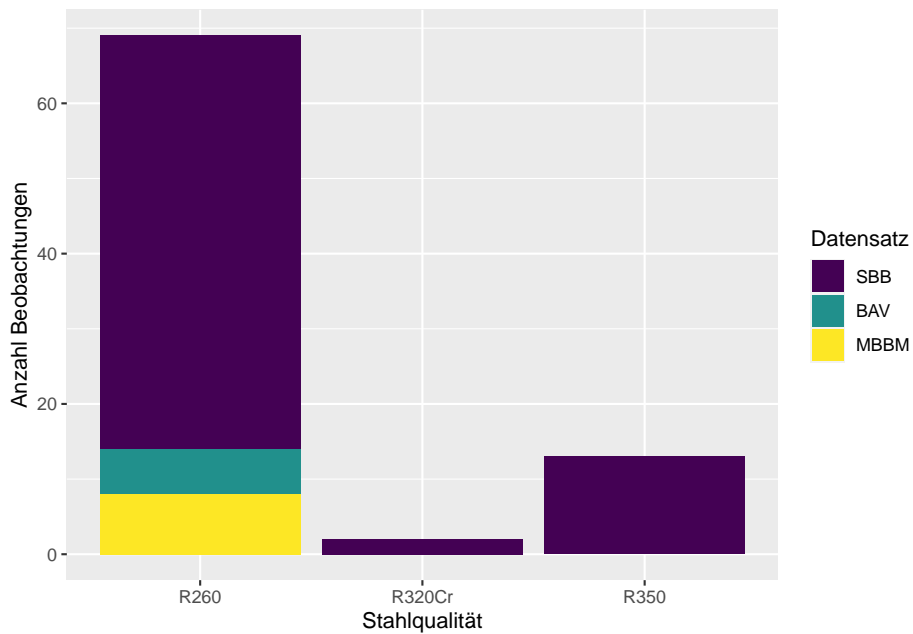


Abbildung A.6.: Säulendiagramm der Beobachtungen pro Stahlqualität und Datensatz.

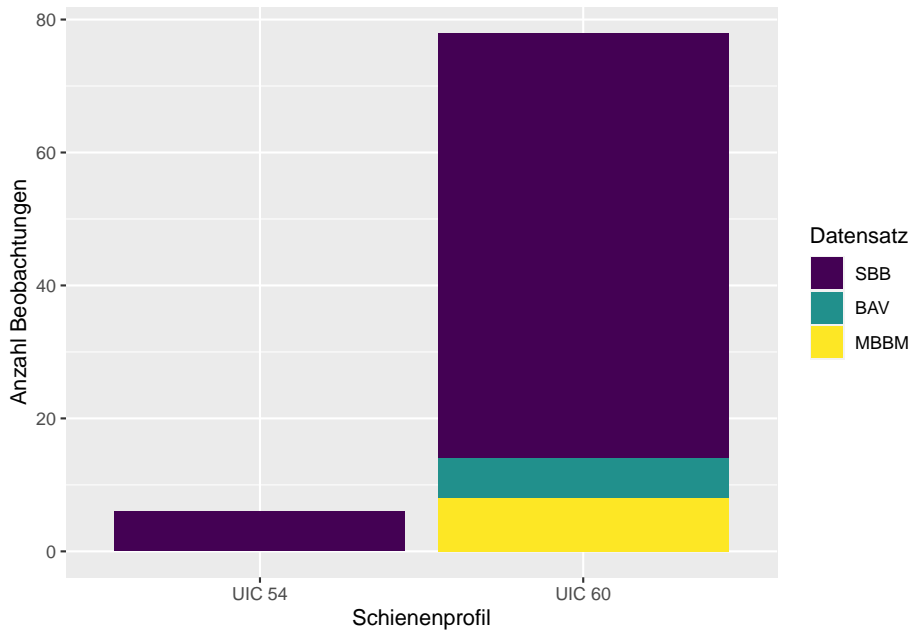


Abbildung A.7.: Säulendiagramm der Beobachtungen pro Schienenprofil und Datensatz.

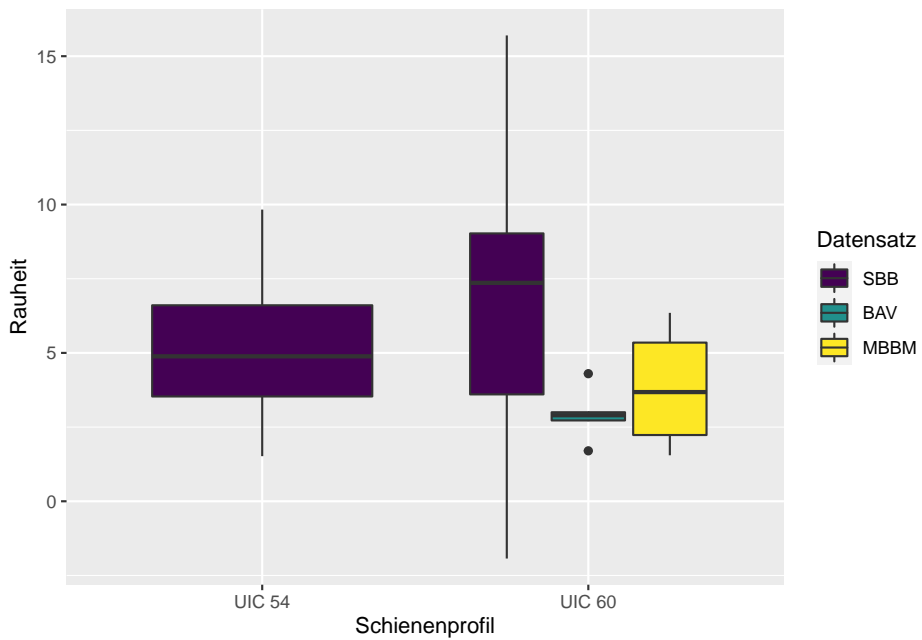


Abbildung A.8.: Boxplots der Rauheit pro Schienenprofil und Datensatz.

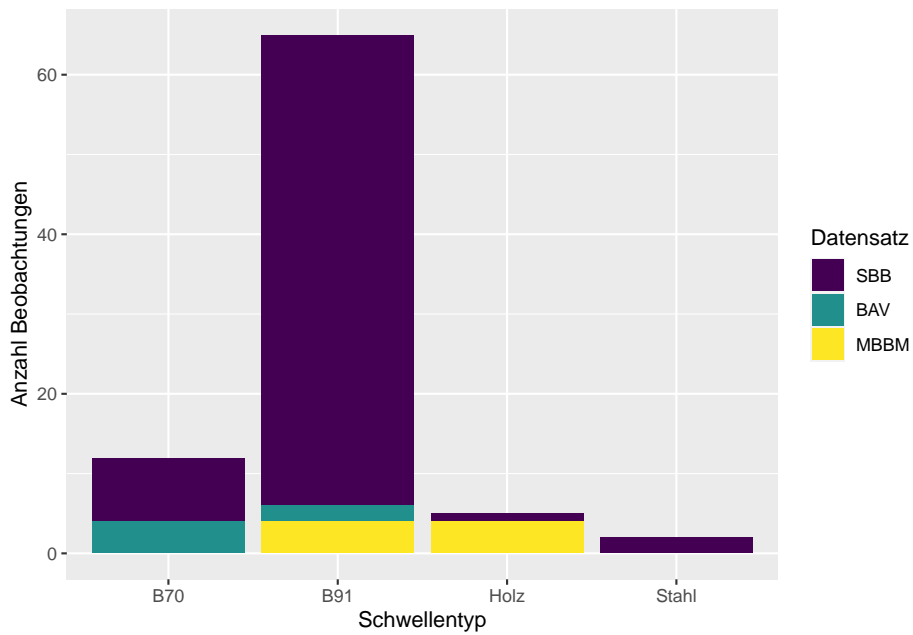


Abbildung A.9.: Säulendiagramm der Beobachtungen pro Schwellentyp und Datensatz.

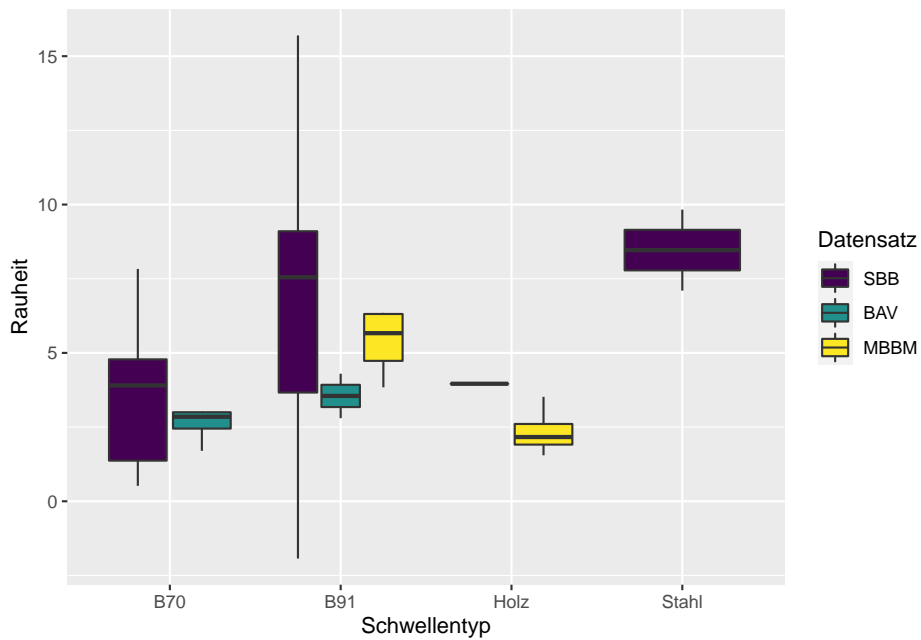


Abbildung A.10.: Boxplots der Rauheit pro Schwellentyp und Datensatz.

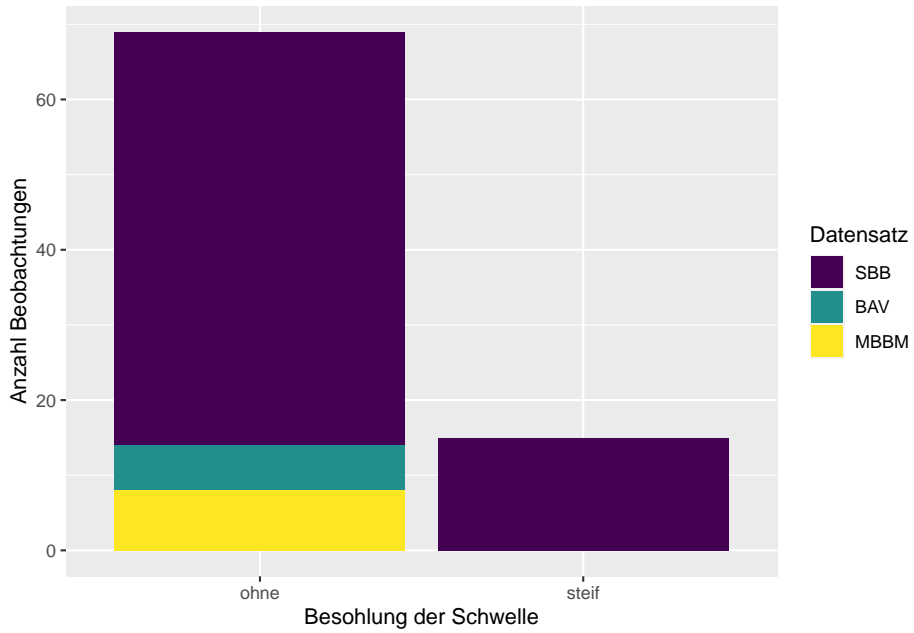


Abbildung A.11.: Säulendiagramm der Beobachtungen pro Besohlung und Datensatz.

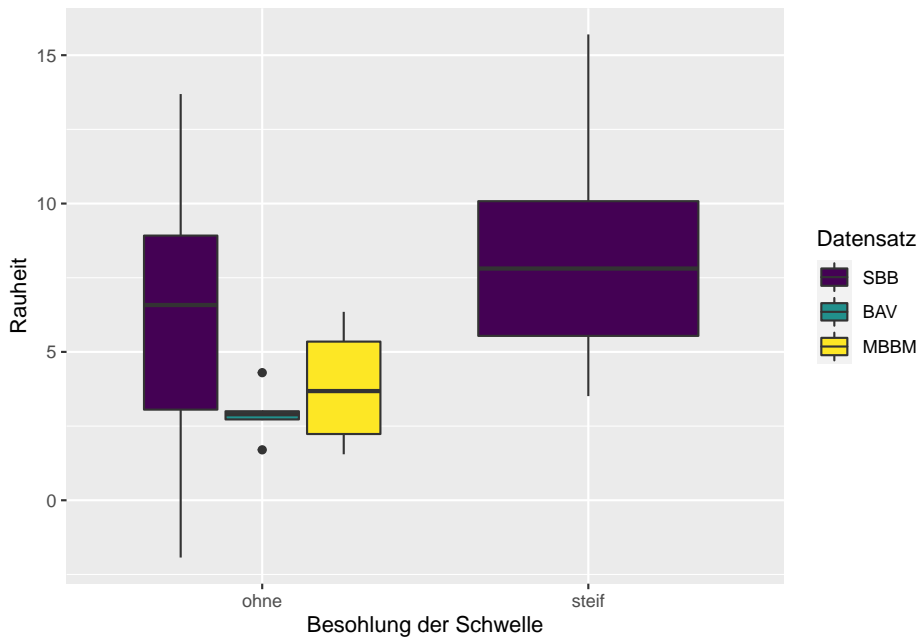


Abbildung A.12.: Boxplots der Rauheit pro Schwellenbesohlung und Datensatz.

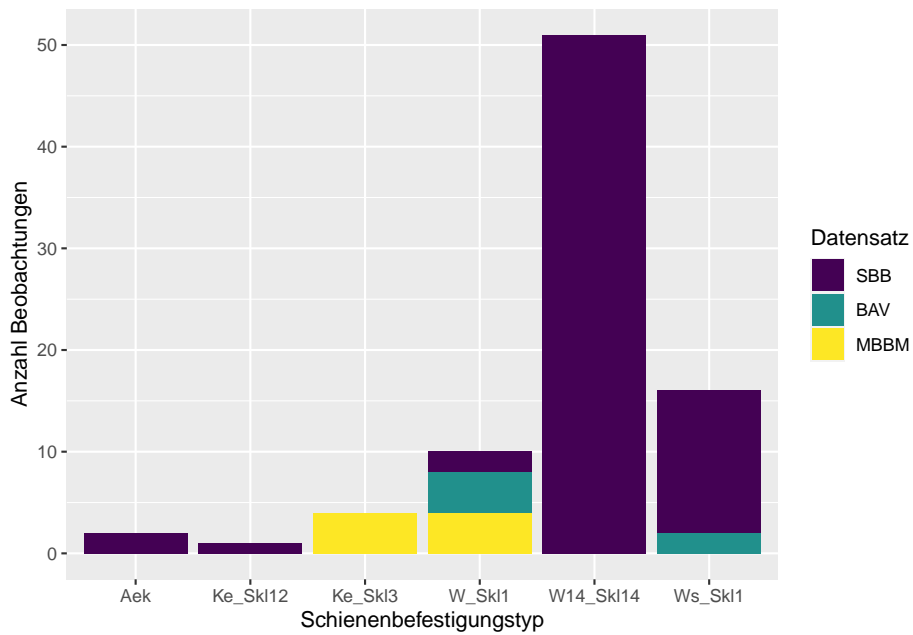


Abbildung A.13.: Säulendiagramm der Beobachtungen pro Schienenbefestigungstyp und Datensatz.

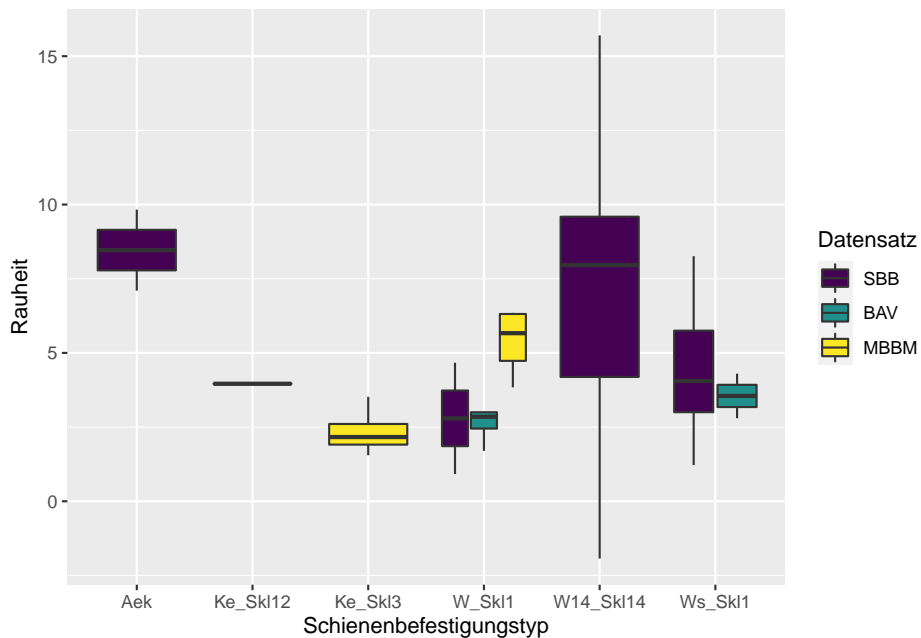


Abbildung A.14.: Boxplots der Rauheit pro Schienenbefestigungstyp und Datensatz.

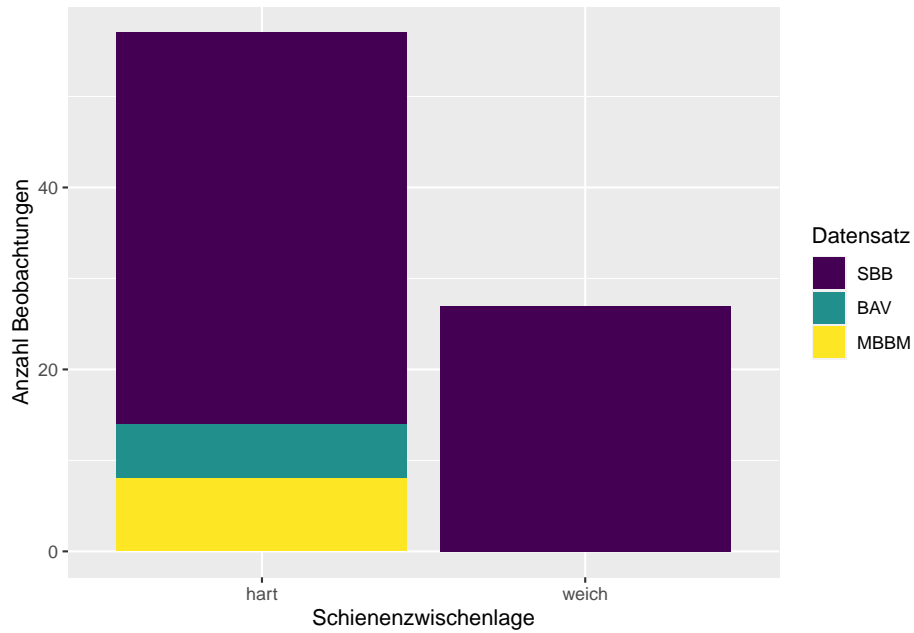


Abbildung A.15.: Säulendiagramm der Beobachtungen pro Schienenzwischenlage und Datensatz.

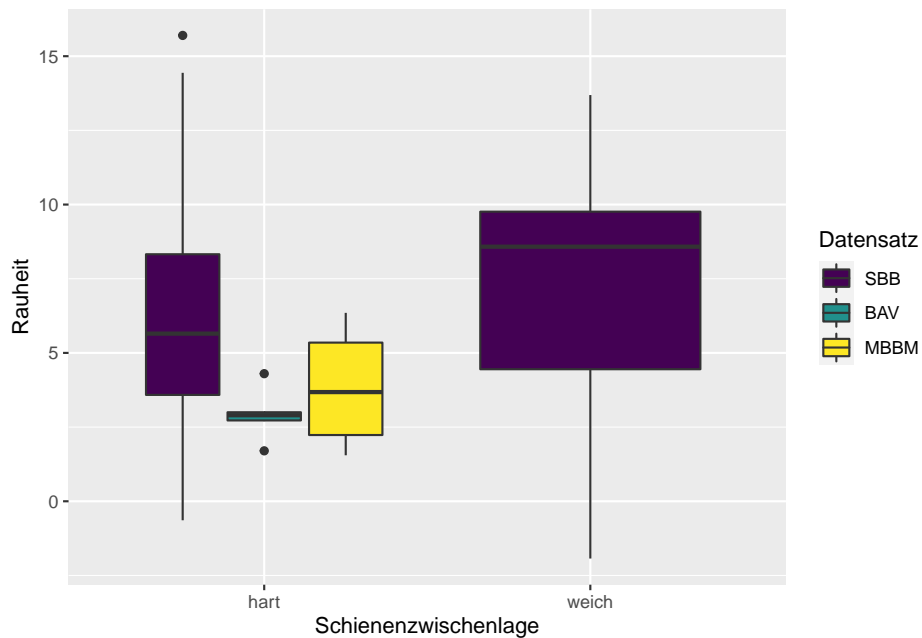


Abbildung A.16.: Boxplots der Rauheit pro Schienenzwischenlage und Datensatz.

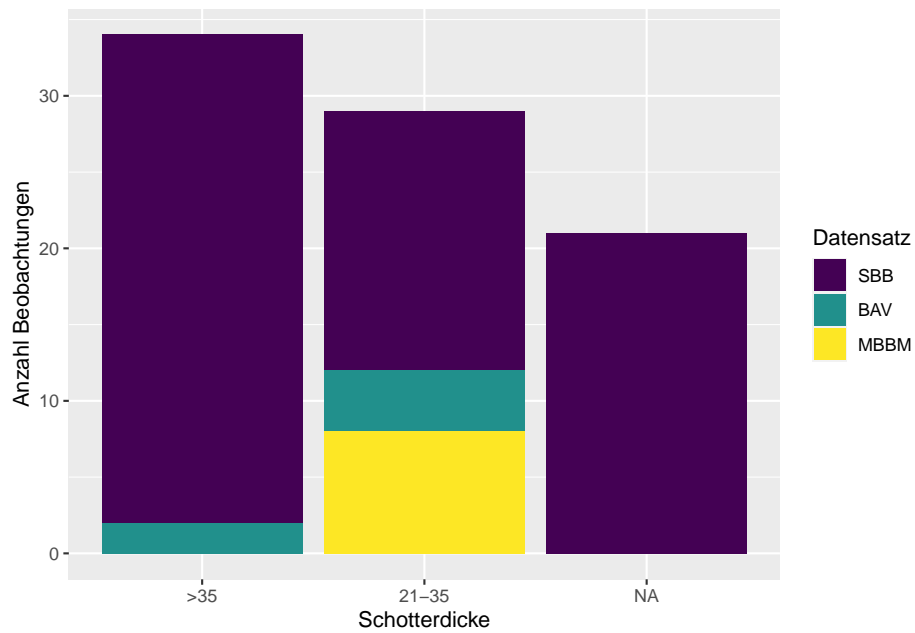


Abbildung A.17.: Säulendiagramm der Beobachtungen pro Schotterdicke und Datensatz.

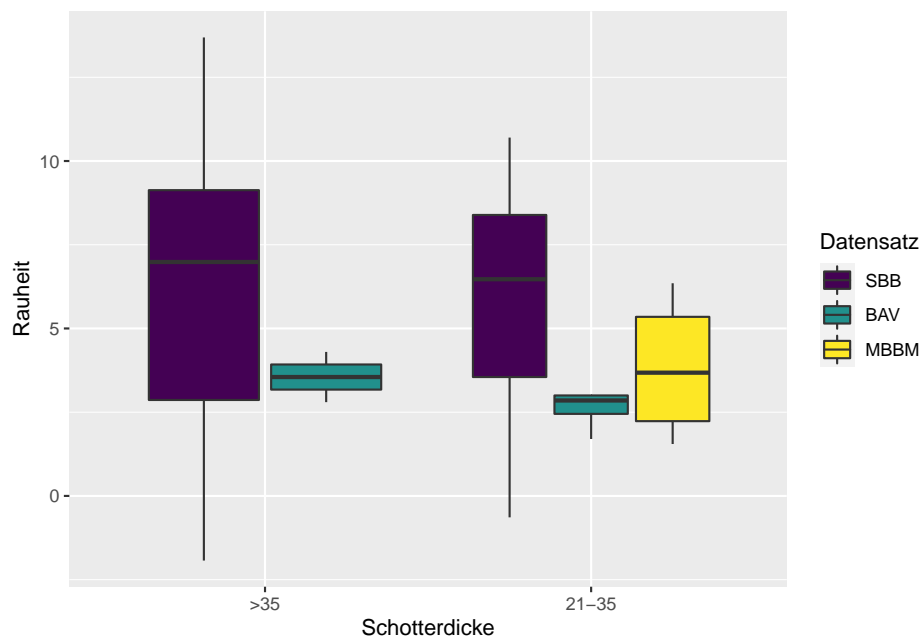


Abbildung A.18.: Boxplots der Rauheit pro Schotterdicke und Datensatz.

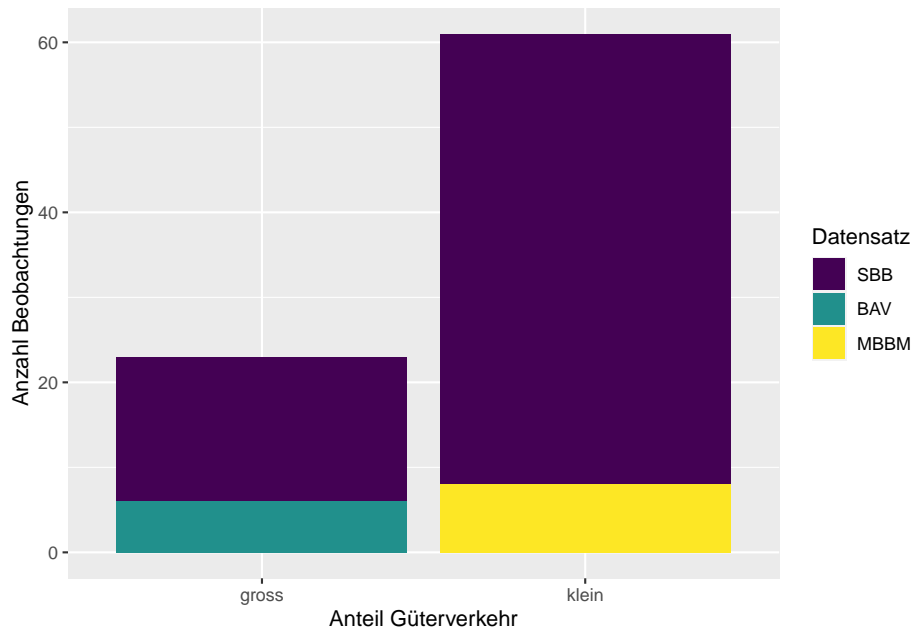


Abbildung A.19.: Säulendiagramm der Beobachtungen pro Anteil Güterverkehr und Datensatz.

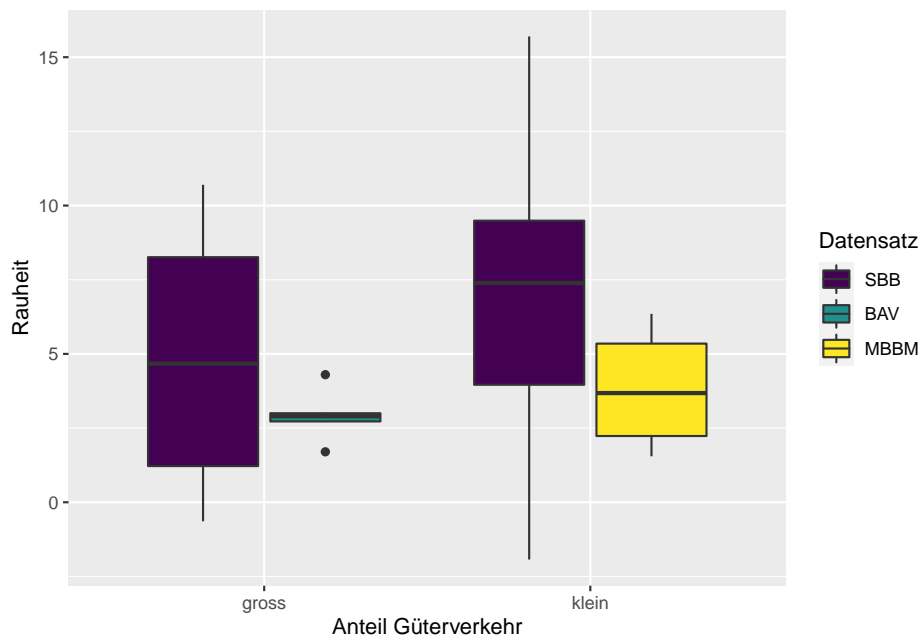


Abbildung A.20.: Boxplots der Rauheit pro Anteil Güterverkehr und Datensatz.

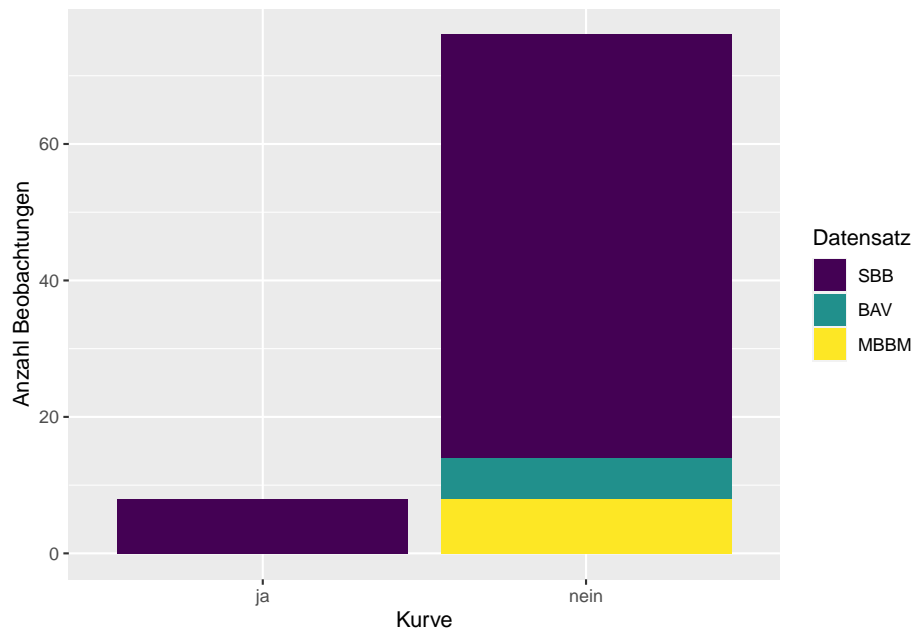


Abbildung A.21.: Säulendiagramm der Beobachtungen pro Kurve und Datensatz.

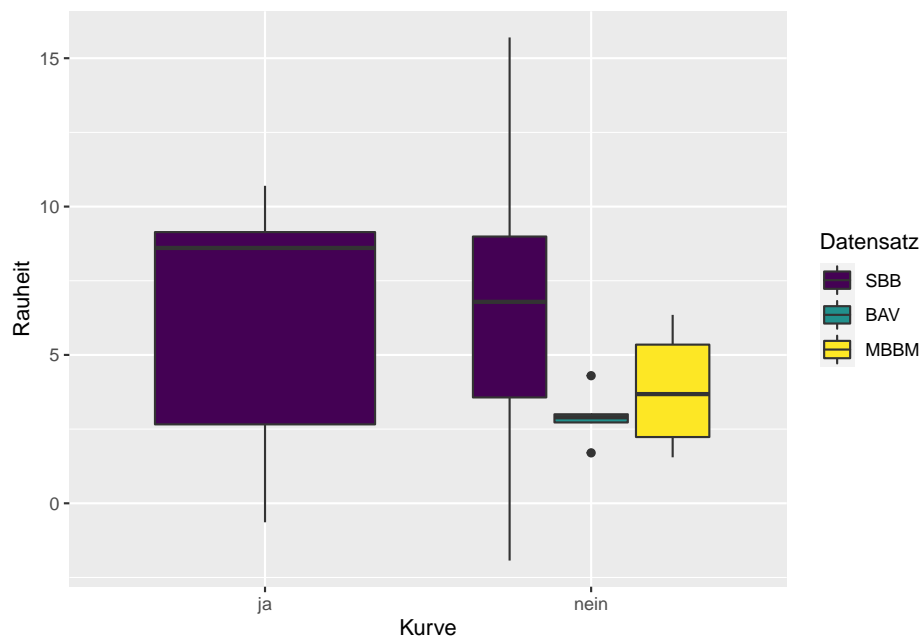


Abbildung A.22.: Boxplots der Rauheit pro Kurve und Datensatz.

A.1.2. Diagnostische Plots

A.1.2.1. Modelle über alle Tonnagewerte

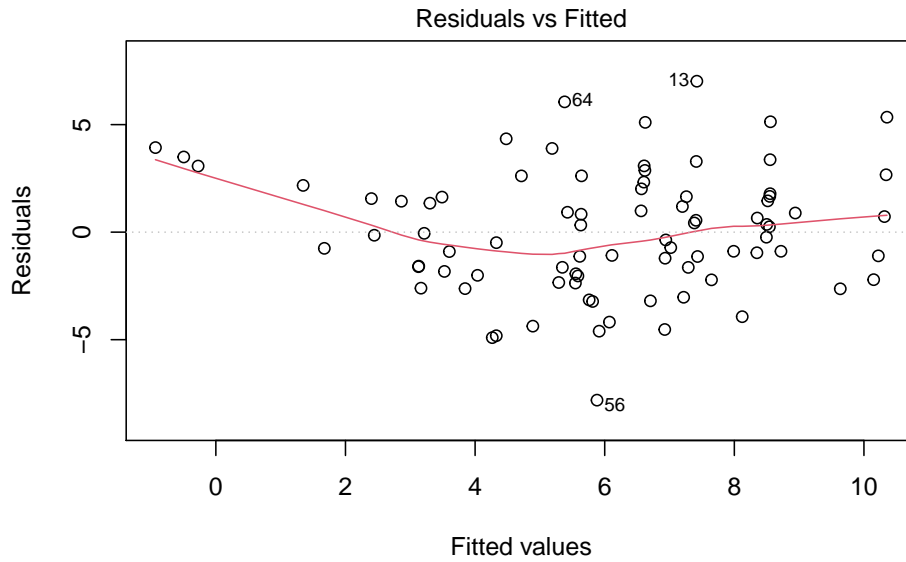


Abbildung A.23.: Residuals vs Fitted für das volle Modell

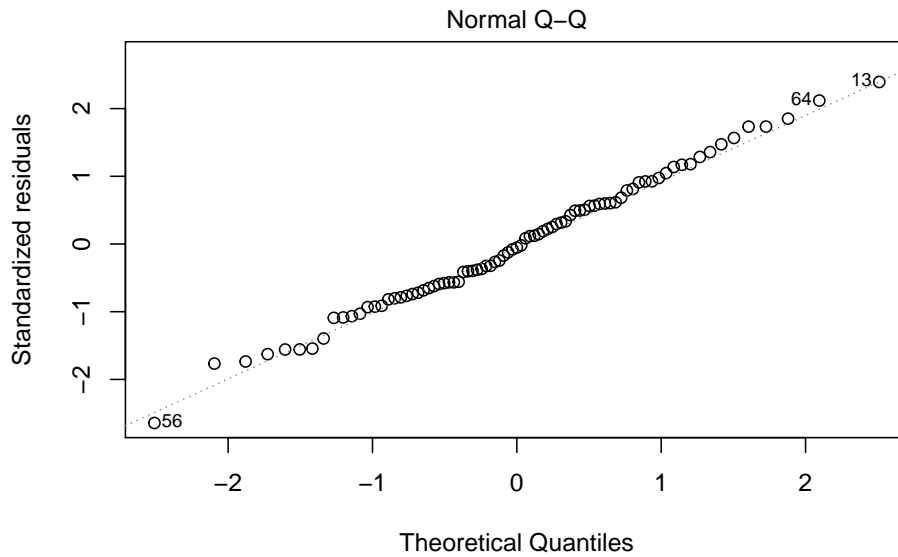


Abbildung A.24.: Q-Q-Plot für das volle Modell

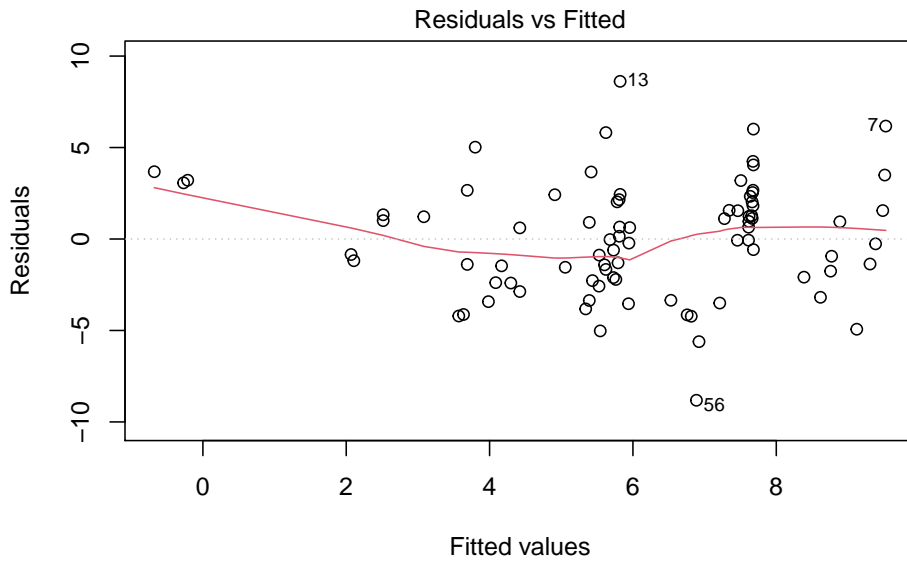


Abbildung A.25.: Residuals vs Fitted für das Modell der forward selection.

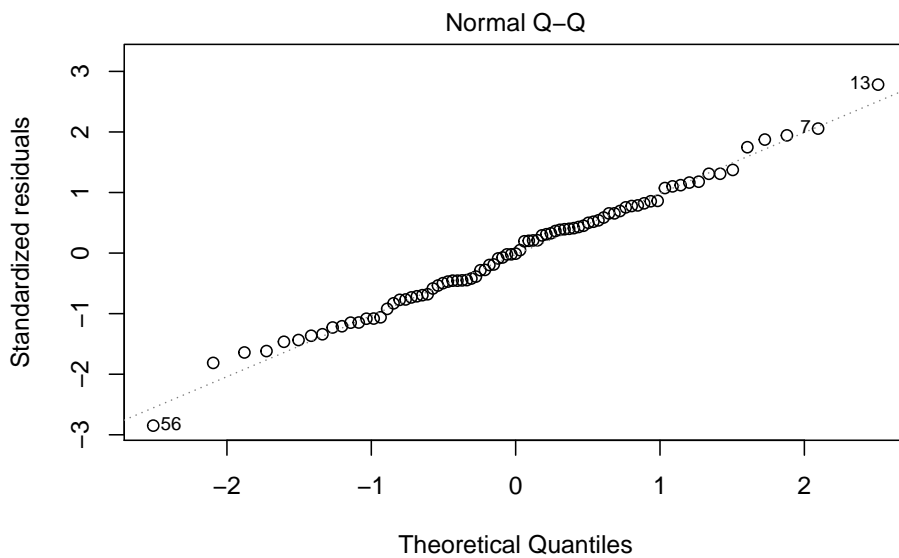


Abbildung A.26.: Q-Q-Plot für das Modell der forward selection.

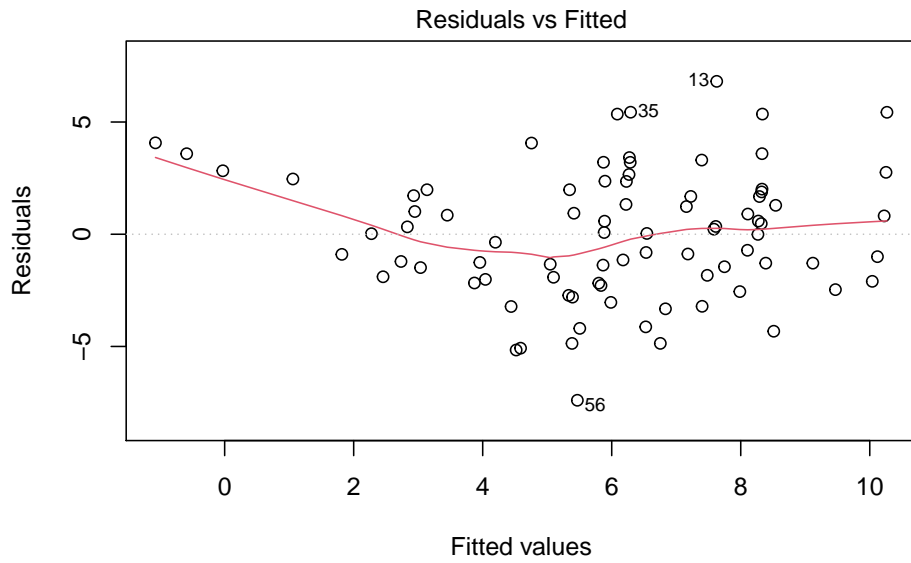


Abbildung A.27.: Residuals vs Fitted für das Modell der backward elimination

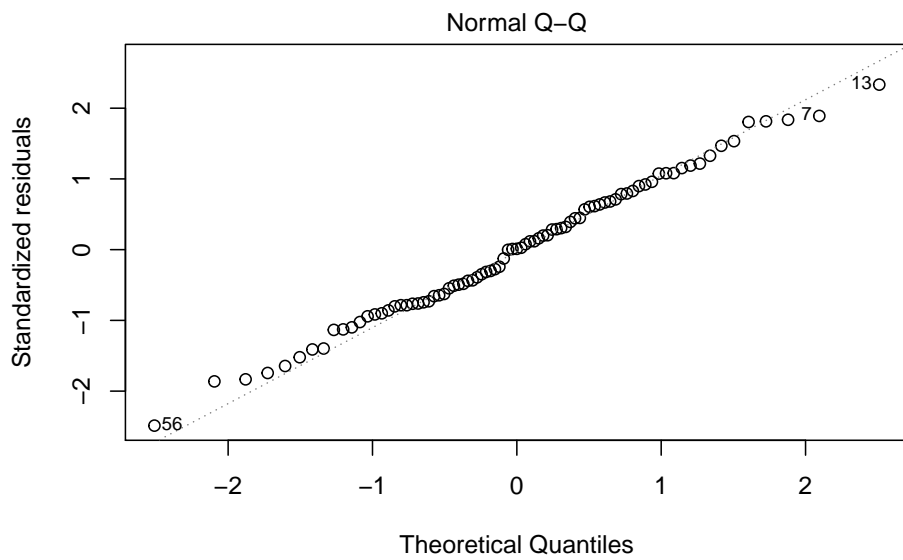


Abbildung A.28.: Q-Q-Plot für das Modell der backward elimination

A.1.2.2. Modell für kleine Tonnagen

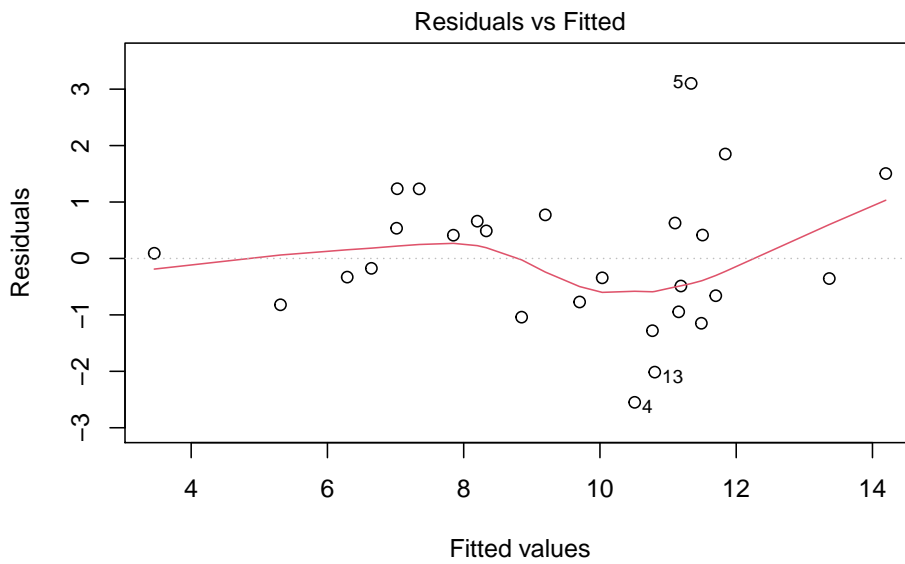


Abbildung A.29.: Residuals vs Fitted für das volle Modell

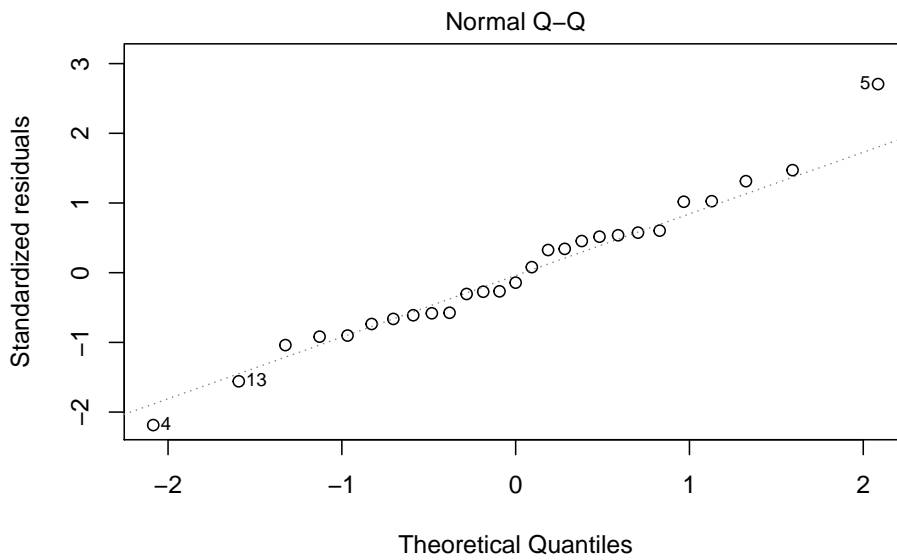


Abbildung A.30.: Q-Q-Plot für das volle Modell

A.1.2.3. Modelle für grosse Tonnagen

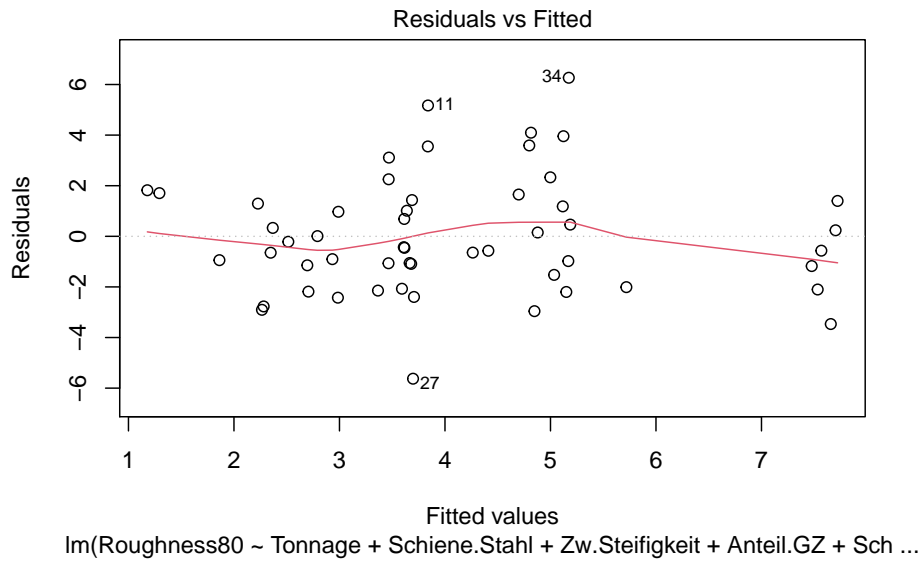


Abbildung A.31.: Residuals vs Fitted für das volle Modell

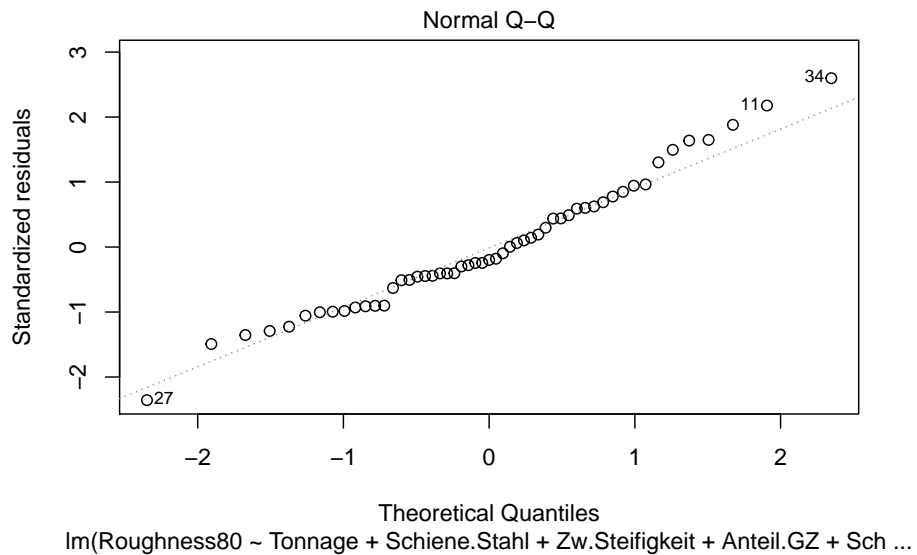


Abbildung A.32.: Q-Q-Plot für das volle Modell

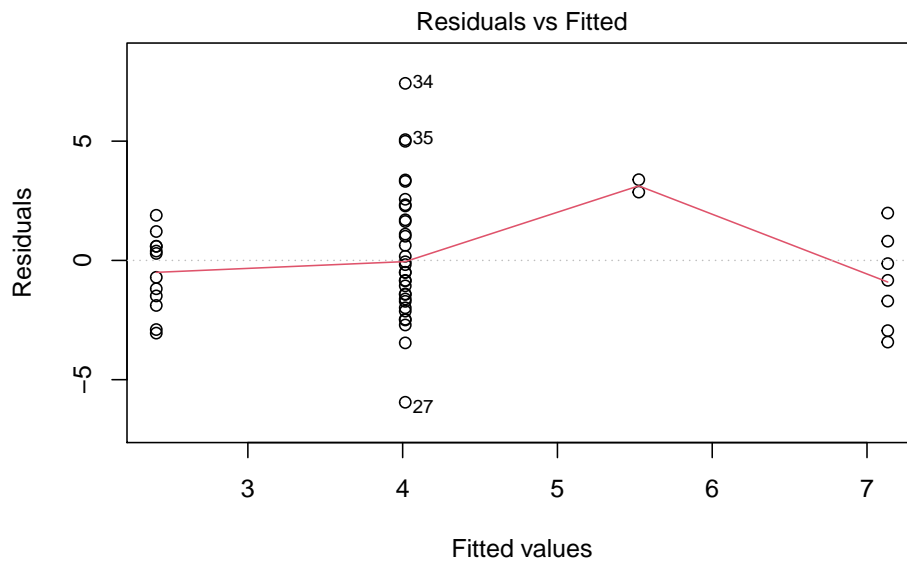


Abbildung A.33.: Residuals vs Fitted für das Modell der forward selection.

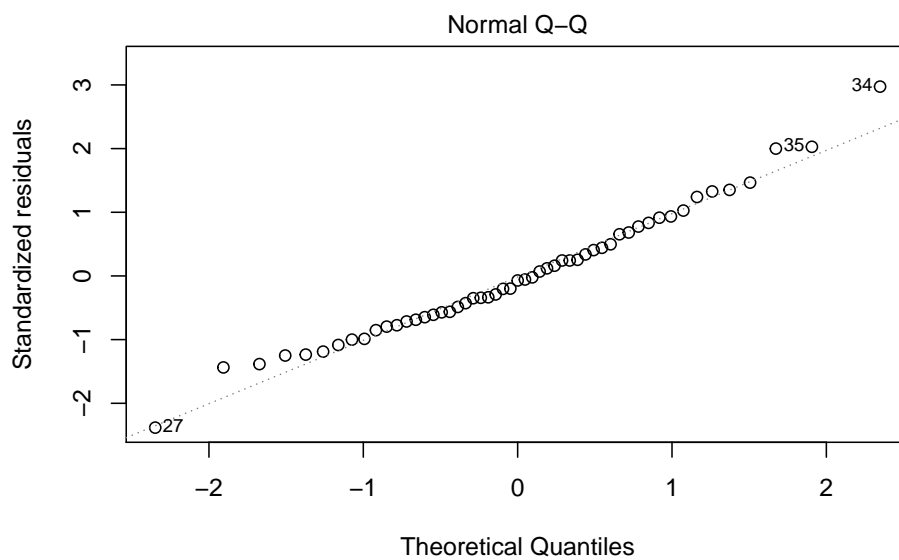


Abbildung A.34.: Q-Q-Plot für das Modell der forward selection.

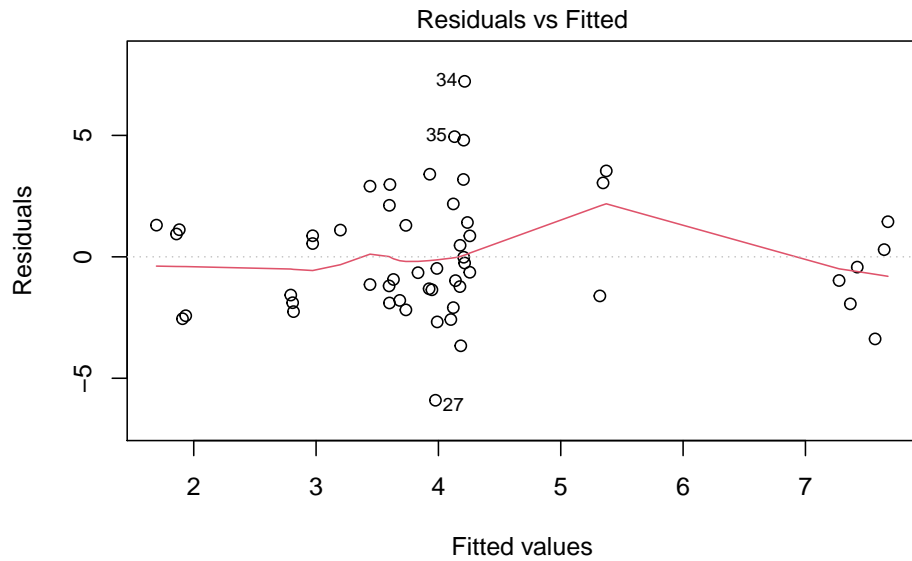


Abbildung A.35.: Residuals vs Fitted für das Modell der backward elimination

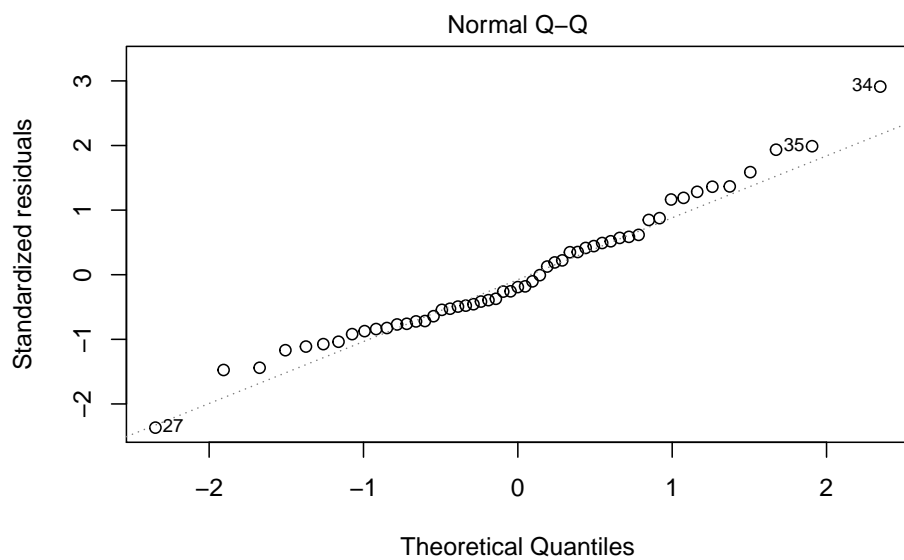


Abbildung A.36.: Q-Q-Plot für das Modell der backward elimination

Literaturverzeichnis

- [1] E. Freud, 2020, *Statistische Analyse von Schienenrauheit - Datenbeschreibung*.
- [2] Europäisches Komitee für Normung, EN ISO 15610:2009: *Bahnanwendungen – Geräuschemission – Messung der Schienenrauheit im Hinblick auf die Entstehung von Rollgeräusch*.
- [3] AEA Technology Rail BV, 2003, *Harmonised Accurate and Reliable Methods for the EU Directive on the Assessment and Management Of Environmental Noise – Definition Of Track Influence: Roughness In Rolling Noise*.
- [4] D. Belcher, C. Gutmann 2014, *Erprobung von Schienendämpfern – Schienenrauheitsmessung auf der Versuchsstrecke bei Bern-Bümpliz, 1. Messkampagne*, Bericht Nr. C89244/02.
- [5] D. Belcher, C. Gutmann 2014, *Erprobung von Schienendämpfern – Schienenrauheitsmessung auf der Versuchsstrecke bei Bern-Bümpliz, 2. Messkampagne*, Bericht Nr. C89244/03.
- [6] D. Belcher, 2015, *Erprobung von Schienendämpfern – Schienenrauheitsmessung auf der Versuchsstrecke bei Bern-Bümpliz, 3. Messkampagne im März 2015*, Bericht Nr. C89244/04.
- [7] D. Belcher, 2015, *Erprobung von Schienendämpfern – Schienenrauheitsmessung auf der Versuchsstrecke bei Bern-Bümpliz, 4. Messkampagne im Oktober 2015*, Bericht Nr. C89244/05.
- [8] D. Belcher, 2016, *Erprobung von Schienendämpfern – Schienenrauheitsmessung auf der Versuchsstrecke bei Bern-Bümpliz, 5. Messkampagne im August 2016*, Bericht Nr. C89244/06.
- [9] J. Gareth, D. Witten, T. Hastie, R. Tibshirani, 2013, *An introduction to statistical learning – with applications in R*, Springer, New York.
- [10] Internationale Organisation für Normung, ISO/IEC Guide 98-3:2008: *Uncertainty of measurement – Part 3: Guide to the expression of uncertainty in measurement*. ISO, Genf 2008, ISBN 92-67-10188-9
- [11] T. Mühl, 2014, *Einführung in die elektrische Messtechnik – Grundlagen, Messverfahren, Anwendungen*. 4. aktualisierte und erweiterte Auflage, Springer Vieweg