



Berner
Fachhochschule

Berner Fachhochschule
Technik und Informatik
Institut für Optimierung und Datenanalyse



Rauheit des SBB-Schienenennetzes

Verfeinerung der Modelle anhand neuer Messdaten

Zweite Iteration

Im Auftrag der Schweizerischen Eidgenossenschaft

Biel, Februar 2022

Impressum

Auftraggeberin

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).

Auftragnehmerin

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

Autoren

Dr. Niki Zumbrunnen

Begleitung BAFU / BAV

Dr. Sandro Ferrari, Franz Kuster

Dr. 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 Auftragnehmerin verantwortlich.

Titelbild

Messgerät und Fahrspiegel

Quelle: SBB

Inhaltsverzeichnis

Zusammenfassung	1
1. Einleitung	3
1.1. Ausgangslage	3
1.2. Grundlage	4
1.3. Ziele der aktuellen Projektphase	4
2. Beschreibung der vorhandenen Daten	5
2.1. Beschreibung der einzelnen Variablen	5
2.1.1. Rauheit	5
2.1.2. Tonnage	7
2.1.3. Angewandte Technik beim letzten Schleifen (Maschine)	7
2.1.4. Stahlqualität	8
2.1.5. Schienenprofil	8
2.1.6. Schwellentyp	9
2.1.7. Schienenbefestigungstyp	9
2.1.8. Schienenzwischenlage	10
2.1.9. Anteil Güterverkehr	11
2.1.10. Streckengeometrie	12
2.1.11. Ort	12
2.1.12. Weitere Variablen	12
3. Statistische Modelle	15
3.1. Verwendete Modelle	15
3.1.1. Multivariate adaptive Regressionssplines (MARS)	15
3.1.2. Random Forest	15
3.1.3. Kriterien für die Wahl des besten Modells	16
3.2. Modellierung der neuen Daten	17
3.3. Weitere Überlegungen	18
3.3.1. Effekte der erklärenden Variablen	18
3.3.2. Abhängigkeiten zwischen den Messungen	19
4. Schätzung der Rauheit	21
5. Messstrategie und Ausblick	25
A. Anhang	27
A.1. Beschreibung der einzelnen Variablen	27
A.2. Details zum MARS-Modell ohne Schienenprofil	33
A.3. Prädiktionsbänder	34

Zusammenfassung

Gemäss Artikel 7 der Verordnung über die Lärmsanierung der Eisenbahn (VLE) sind die Bahninfrastukturbetreiberinnen vom Bund beauftragt, die Schienenrauheit in ihrem Netz zu kennen und sind verpflichtet, die Rauheit unter einem bestimmten Grenzwert zu halten.

Im Jahr 2020 wurde in einer Vorstudie die Schienenrauheit anhand vorhandener Messdaten geschätzt. Es lagen zu diesem Zeitpunkt jedoch nur sehr wenige Messungen vor. In der aktuellen Projektphase soll nun die Datengrundlage durch zusätzliche Messungen erweitert und die Modelle für die Schätzung der Rauheit damit verfeinert werden. Dies geschieht in drei Iterationen und dauert bis Ende 2022. In jeder Iteration werden zusätzliche Messungen durchgeführt und anschliessend die Modelle an die neuen Daten angepasst. Basierend auf den daraus gewonnenen Erkenntnissen werden dann neue Messstandorte für die nächste Iteration definiert. Dieser Bericht beschreibt die Datenanalyse der zweiten Iteration.

In der zweiten Iteration wurden 53 zusätzliche Messungen gemacht. Insgesamt konnten für die Modellierung nun 164 Messungen verwendet werden.

Die Vorhersagen sind mit multivariaten adaptiven Regressionssplines (MARS) gemacht worden, welche für die vorliegenden Daten stückweise lineare Regressionskurven geliefert haben. Die Schienenrauheit nimmt nach dem Schleifen zuerst sehr stark ab bis zu einer Tonnage von 0.26 Mt. Danach ist die Abnahme deutlich geringer bis zu einer Tonnage von 4.7 Mt. Ab diesem Tonnagewert nimmt die Rauheit wieder leicht zu. Die Zunahme ist jedoch so gering, dass der Verlauf im Wesentlichen als konstant betrachtet werden kann. Das MARS-Modell verwendet neben der Tonnage die kategoriellen Variablen Stahlqualität, Schienenprofil und Anteil Güterzüge. Je nach Ausprägungskombinationen dieser Variablen wird die Regressionskurve vertikal verschoben.

Die geschätzte Rauheit liegt für vier der acht Kombinationen bereits bei Tonnage 0 unter dem kritischen Grenzwert von 10 dB. Auch für die anderen Kombinationen fällt die Rauheit ziemlich bald, spätestens bei 0.23 Mt, unter den Grenzwert.

Die Unsicherheit der Schätzungen quantifizieren wir mit 95%-Prädiktionsintervallen. Für eine neue Beobachtung mit den gegebenen Ausprägungen gehen wir davon aus, dass sie mit einer Sicherheit von 95% in diesem Intervall enthalten ist. Die Abbildung 0.1 zeigt die Regressionskurve (blau) und das Prädiktionsband (hellblau) für die Kombination Stahl R260, Profil UIC 60 und wenig Güterzüge.

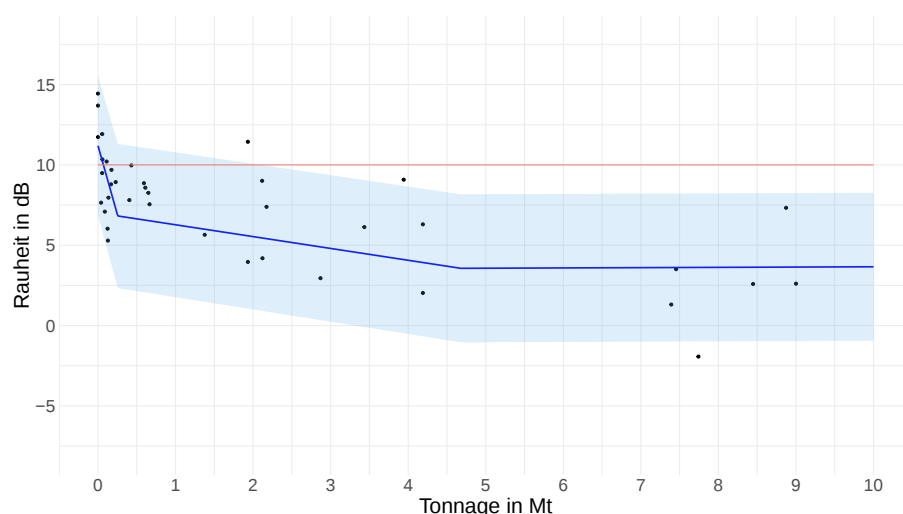


Abbildung 0.1.: Prädiktionsband für Stahl R260, Profil UIC 60 und wenig Güterzüge.

Die obere Grenze des Prädiktionsintervalls fällt für vier der acht Kombinationen spätestens bei 3.18 Mt unter 10 dB, wobei es für die meisten Kombinationen bereits deutlich früher geschieht. Die Kombination Stahl R350, Profil UIC 60 und wenig Güterzüge ist die einzige, für die das Prädiktionsband den Wert 10 dB für alle Tonnagewerte enthält.

In der dritten Iteration sollen vor allem Messungen an Stellen mit den im SBB-Netz am weitesten verbreiteten Oberbaukombinationen durchgeführt werden. Dabei werden Kombinationen bevorzugt, für die einerseits bisher nur wenige Messungen vorhanden sind und die andererseits eine hohe geschätzte Rauheit aufweisen.

1. Einleitung

1.1. Ausgangslage

Gemäss Artikel 7 der Verordnung über die Lärmsanierung der Eisenbahn (VLE) sind die Bahninfrastrukturbetreiberinnen vom Bund beauftragt, die Schienenrauheit in ihrem Netz zu kennen und sind verpflichtet, die Rauheit unter einem bestimmten Grenzwert zu halten.

In der im nächsten Abschnitt beschriebenen Vorstudie wurde die Schienenrauheit anhand der vorhandenen Messdaten geschätzt. In der aktuellen Projektphase soll nun die Datengrundlage durch zusätzliche Messungen erweitert und die Modelle für die Schätzung der Rauheit damit verfeinert werden. Dies geschieht in drei Iterationen und dauert bis Ende 2022. In jeder Iteration werden zusätzliche Messungen durchgeführt und anschliessend die Modelle an die neuen Daten angepasst. Basierend auf den daraus gewonnenen Erkenntnissen werden dann neue Messstandorte für die nächste Iteration definiert. Dieser Bericht beschreibt die Datenanalyse der zweiten Iteration.

Vorstudie

Im Jahr 2020 wurde eine Vorstudie gemacht mit dem Ziel, erste statistische Modelle zu rechnen, um die Schienenrauheit zu schätzen und die damit verbundene Schätzunsicherheit zu quantifizieren. Dies basierend auf den vorhandenen Daten und folgenden Einflussfaktoren:

- a) Summe der gefahrenen Radsatzlasten seit dem letzten Schleifen
- b) Art des Schleifens, insbesondere kurz nach dem Schleifen (akustisches Schleifen). Konkret sollen die verschiedenen Schleifmaschinen miteinander verglichen werden.
- c) Stahlqualität
- d) Schienenprofil (UIC 54 oder UIC 60)
- e) Schwellentyp
- f) Besohlung der Schwelle
- g) Schienenbefestigungstyp
- h) Schienenzwischenlage
- i) Schotterdicke
- j) Gewichtsanteil von Güter-, respektive Personenverkehr seit dem letzten Schleifen
- k) Streckentyp: Kurve oder Gerade

Als Datengrundlage dienten Messdaten, welche die SBB durch punktuelle Messungen an verschiedenen Standorten im Netz erhoben hatten. Zudem wurden Daten der jährlichen Messungen an den Monitoringstationen des BAV verwendet, soweit das Datum des letzten Schleifens bekannt war. Insgesamt standen nur 84 Beobachtungen zur Verfügung. Diese geringe Datenmenge führte im Modell mit allen oben aufgeführten Einflussfaktoren zu einer Überanpassung, weil zu wenige Beobachtungen pro zu schätzendem Parameter zur Verfügung standen. Deshalb wurde die Anzahl verwendeter Einflussfaktoren reduziert. Die Variablenselektion wurde einerseits basierend auf Fachwissen der SBB und andererseits datenbasiert durchgeführt. Es ist klar, dass man mit der sehr limitierten Anzahl Messungen die Rauheit des gesamten Schienennetzes nicht mit einer hohen Genauigkeit schätzen kann. Hierzu bräuchte es viel mehr Daten. Im Idealfall 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 zurzeit 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 dieser Vorstudie 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.

Auf Basis der gerechneten Modelle wurden Standorte für zukünftige Messungen vorgeschlagen. Es wurden Parameterkombinationen identifiziert, mit hoher geschätzter Rauheit oder hoher Schätzunsicherheit.

1.2. Grundlage

Als Datenbasis dienen einerseits die bereits in der Vorstudie verwendeten Messdaten der SBB. Dazu kommen die neuen Messungen, welche die SBB in den ersten beiden Iterationen dieser Projektphase durchgeführt haben.

Ursprünglich war angedacht auch die Daten des ETH-Projekts zu verwenden. Diese stehen jedoch bis dato noch nicht zur Verfügung und werden voraussichtlich auch für die dritte Iteration nicht zur Verfügung stehen. Falls wir für die dritte Iteration rechtzeitig Daten erhalten, werden wir sie in die Analyse miteinbeziehen.

1.3. Ziele der aktuellen Projektphase

In der zweiten Projektphase (bis Ende 2022) sollen folgende Ziele erreicht werden:

1. Verfeinerung der in der Vorstudie erstellten Modelle mit zusätzlichen Messdaten und eine entsprechende Reduktion der Schätzunsicherheit für die Rauheit.
2. Eruierung der Robustheit der resultierenden Modelle anhand einer Simulationsstudie.
3. Erstellung einer Schweizerkarte mit der Darstellung der geschätzten Rauheit auf dem gesamten SBB-Schienennetz.

Die Schweizerkarte wird im Rahmen dieses Projekts voraussichtlich nicht erstellt werden können. Um die Karte zu erstellen, bräuchte es detaillierte Daten des ganzen Schienennetzes zum Oberbau und zum letzten Schleifen. Insbesondere die Daten des Schleifens stehen zur Zeit nur dezentral in den Schleifprotokollen zur Verfügung. Eine zentrale Aufbereitung der Daten ist vor Ende 2022 nicht geplant. Der Code für die Schätzung der Rauheit wird der SBB dennoch zur Verfügung gestellt. Sobald dann eine entsprechende Datengrundlage vorhanden wäre, könnte das GIS-Team der SBB wie geplant die Schweizerkarte erstellen.

2. Beschreibung der vorhandenen Daten

In der zweiten Iteration wurden 53 Messungen gemacht, davon 36 Messungen an vier Standorten in Alznach. Diese Messungen sind nicht statistisch unabhängig voneinander. Auf die damit verbundenen Probleme gehen wir im Abschnitt 3.3.2 ein.

Neben den neuen Messungen verwenden wir auch die bis zur ersten Iteration erhobenen 123 Datenpunkte. Diese bestehen aus 109 Messungen, welche die SBB an verschiedenen Orten und zu verschiedenen Zeiten durchgeführt hat. Dazu kommen 6 Rauheitsmessungen an den Monitoringstationen des BAV und 8 Messungen aus einer Studie mit dem Ziel, den Einfluss von Schienendämpfern auf die Rauheit zu ermitteln. Diese Messungen wurden von der Firma Müller-BBM durchgeführt. Deshalb wird der entsprechende Datensatz im Folgenden mit MBBM bezeichnet.

Wie in der ersten Iteration schliessen wir 12 Beobachtungen aus. Ausgeschlossen werden die Beobachtungen der Linie 460: Grauenstein, Trubschachen, Hasle, Gigen, Wigen. Aus Sicherheitsgründen wurde systematisch in der Nähe eines Bahnübergangs gemessen und es wurde dabei bemerkt, dass am Bahnübergang das Schleifen kurz unterbrochen wurde. Dies führt zu einer wesentlichen Unsicherheit bezüglich der Tonnage. In Grauenstein hat das Fräsen sogar mitten in der gemessenen Strecke angefangen. Zusätzlich gibt es zu 3 dieser 5 Beobachtungen keine Angaben zur angewandten Schleiftechnik.

Zwei Messungen wurden an Stellen mit der Stahlqualität R320C gemacht. Auch diese Beobachtungen werden entfernt. Da es zu wenige Beobachtungen mit diesen Ausprägungen hat, würde das sonst zu Fehlern in der Kreuzvalidierung führen.

Beobachtungen, bei welchen mit der Fräsmaschine gearbeitet wurde, werden aus dem gleichen Grund ebenfalls ausgeschlossen, dadurch werden nach den vorangegangenen Ausschlüssen zwei weitere Beobachtungen entfernt.

Weiter werden drei Messungen ausgeschlossen, weil die Angaben zur Schleiftechnik fehlen.

Es wurden also 164 Messungen für die Modellierung verwendet. Im folgenden Abschnitt werden diese Daten beschrieben.

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 die Veränderung der Datenlage hervorzuheben, sind die in der ersten Iteration verwendeten und die seither erhobenen Messungen unterschiedlich eingefärbt.

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. In Tabelle 2.1 sind die wichtigsten Kennzahlen für die Rauheit aufgeführt.

Auf dem Boxplot in Abbildung 2.1 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. Die neuen Messungen weisen im Mittel tiefere Rauheiten aus.

Tabelle 2.1.: Kennzahlen für die Rauheit

Rauheit	alle, N = 164	alt, N = 111	neu, N = 53
Minimum	-1.9	-1.9	-1.6
1. Quartil	2.6	2.8	2.4
Median	5.0	5.1	4.9
Mittelwert	5.3	5.4	5.1
3. Quartil	7.8	8.0	7.4
Maximum	15.7	15.7	12.8

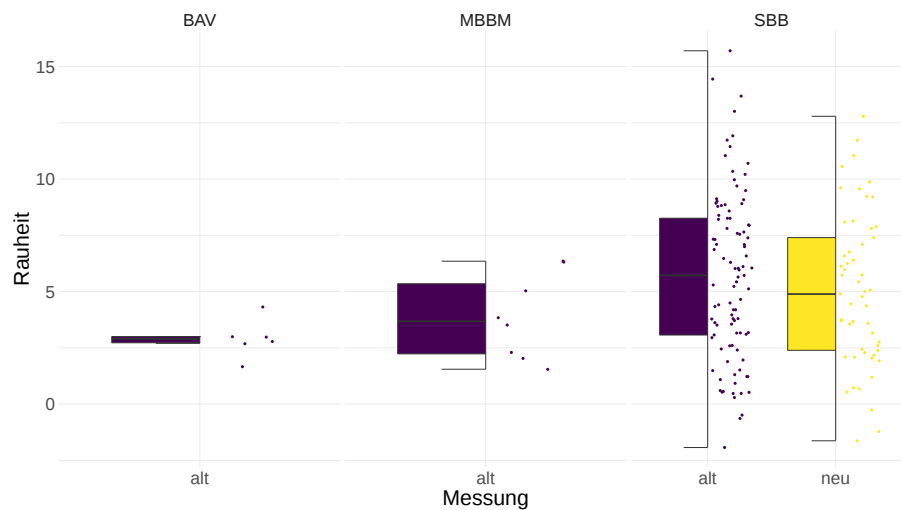


Abbildung 2.1.: Boxplot der Rauheit pro Datensatz und Messung

Tabelle 2.2.: Kennzahlen für die Tonnage

Tonnage	alle, N = 164	alt, N = 111	neu, N = 53
Minimum	0	0	0
1. Quartil	135'411	216'706	115'704
Median	2'024'383	2'173'440	1'217'916
Mittelwert	8'408'052	10'704'250	3'599'035
3. Quartil	8'873'015	13'625'460	4'939'672
Maximum	64'541'062	64'541'062	24'935'420

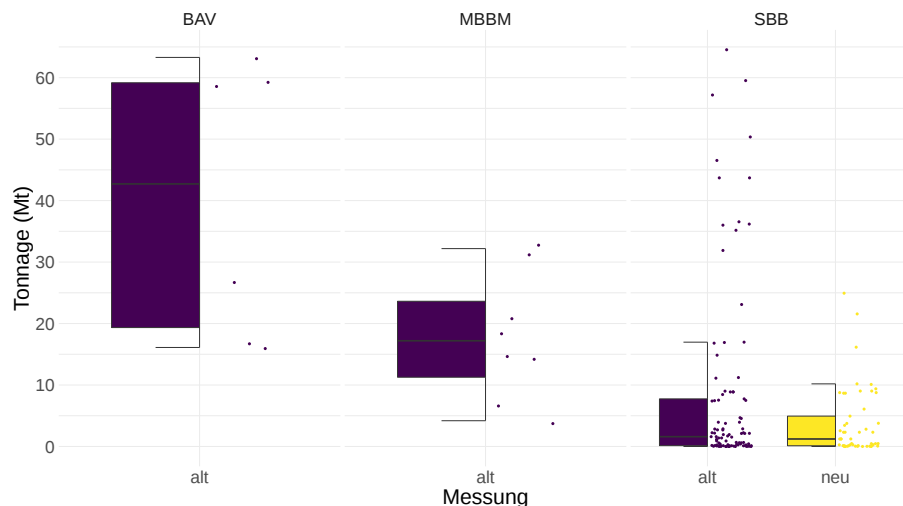


Abbildung 2.2.: Boxplot der Tonnage pro Datensatz und Messung

2.1.2. Tonnage

Die Variable Tonnage beschreibt die Summe der Radsatzlasten seit dem letzten Schleifen in Tonnen bzw. Megatonnen (Mt). Sie ist numerisch und ihre Verteilung ist ziemlich schief. Es hat ein paar Ausreisser gegen oben. Eine Messung (in Wichtrach) mit 115 Megatonnen (Mt) sticht besonders heraus. Wegen fehlenden Angaben zur Schleiftechnik wurde sie jedoch nicht für die Modelle und auch nicht für die Grafiken und die Berechnung der Kennzahlen verwendet.

In Tabelle 2.2 sind die wichtigsten Kennzahlen für die Tonnage aufgeführt. Die Mehrheit der Beobachtungen im SBB Datensatz weisen tiefe Tonnagen aus, es wird aber der gesamte relevante Tonnagebereich abgedeckt. In der zweiten Iteration wurden vor allem Messungen an Stellen mit tiefen Tonnagen gemacht. In den anderen beiden Datensätzen kommen keine ganz tiefen Tonnagewerte vor. Dies ist auf dem Boxplot in Abbildung 2.2 ersichtlich.

Abbildung 2.3 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 scheint für kleine Tonnagewerte viel grösser als für grosse.

2.1.3. 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.

Maschine ist kategoriell mit 4 Ausprägungen. Die Angabe zur Maschine fehlte in den Daten der ersten Iteration für 7 Beobachtungen. Für diese Messungen fehlen die Schleifprotokolle, welche unter anderem Informationen zur Schleiftechnik enthalten. Es handelt sich dabei um Messungen mit grossen bis sehr grossen Tonnagen. Für eine dieser Beobachtungen konnte die Angabe der Maschine (B) ergänzt werden. Es fehlen also noch 6 Werte für die Variable Maschine. Diese Beobachtungen wurden für die Modellierung nicht verwendet und sind hier auch nicht aufgeführt. Auch die Messungen

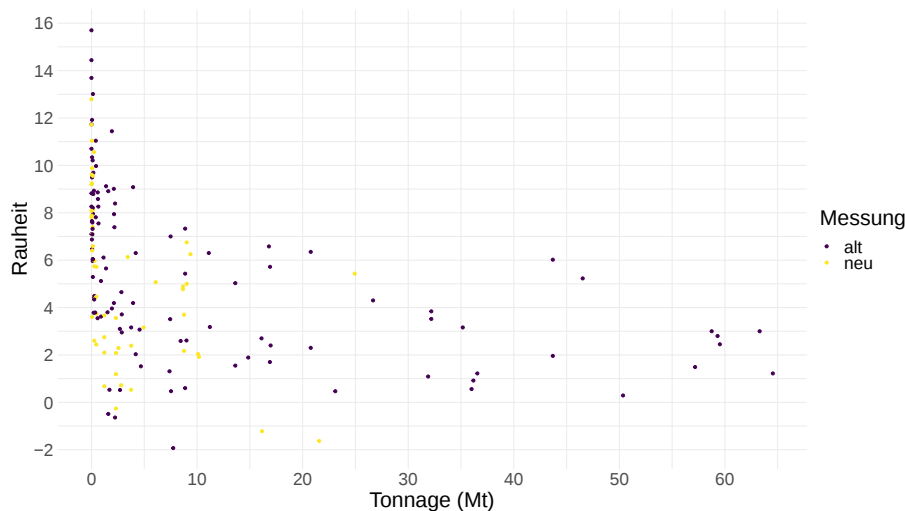


Abbildung 2.3.: Streudiagramm zur Tonnage und Rauheit der Messungen.

Tabelle 2.3.: Häufigkeitstabelle für die Schleiftechnik.

Maschine	alle, N = 164	alt, N = 111	neu, N = 53
A	84 (51%)	42 (38%)	42 (79%)
B	25 (15%)	17 (15%)	8 (15%)
C	20 (12%)	20 (18%)	0 (0%)
D	35 (21%)	32 (29%)	3 (5.7%)

an Stellen, die mit einer Fräsmaschine bearbeitet wurden, verwenden wir nicht für die Modelle und führen sie hier auch nicht auf. Für eine Messung in Aarau wurde seit der letzten Iteration der Wert für die Maschine von A auf B korrigiert.

Die Verteilung der Variable Maschine ist in der Häufigkeitstabelle in Tabelle 2.3 und im Balkendiagramm in Abbildung A.1 im Anhang ersichtlich. Der Maschinentyp C kommt bei den neuen Beobachtungen nicht mehr vor. In Abbildung A.2 ist der Zusammenhang zwischen der Rauheit und der Maschine in Boxplots dargestellt.

2.1.4. Stahlqualität

Die Variable Stahl beschreibt die Stahlqualität. Sie ist kategoriell mit den 2 Ausprägungen R260 und R350. Die meisten Messungen wurden an Stellen mit dem Typ R260 gemacht. Der Typ R350 ist härter als R260 und wird in Kurven verwendet. Für Kurven mit einem Radius bis 1200 m wird R350 eingesetzt. Ab einem Radius von 1800 m wird R260 eingesetzt. Dazwischen kommt je nach Belastung der Strecke einer der beiden Typen zum Einsatz.

In den Daten kommt zusätzlich R320Cr als dritter Typ vor. Er liegt bzgl. Härte zwischen den anderen beiden Typen und kommt auf dem SBB-Schienennetz nur selten vor. An Stellen mit diesem Typ wurden nur zwei Messungen gemacht. Diese werden für die Modelle nicht verwendet und hier auch nicht aufgeführt.

Die Verteilung ist in Tabelle 2.4 und in Abbildung A.3 aufgetragen. In Abbildung 2.4 ist der Zusammenhang zwischen der Rauheit und der Stahlqualität in Boxplots dargestellt. Der weiche Stahl hat tendenziell eine tiefere Rauheit als der harte.

2.1.5. Schienenprofil

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

Tabelle 2.4.: Häufigkeitstabelle für die Stahlqualität

Stahlqualität	alle, N = 164	alt, N = 111	neu, N = 53
R260	143 (87%)	95 (86%)	48 (91%)
R350	21 (13%)	16 (14%)	5 (9.4%)

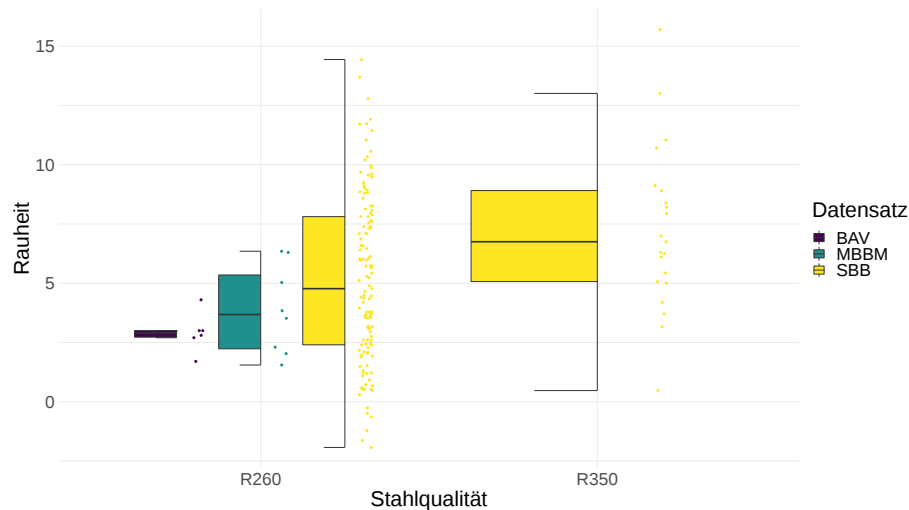


Abbildung 2.4.: Boxplots der Rauheit pro Stahlqualität und Datensatz.

UIC 54 haben, ist die Verteilung nun ausgeglichener.

Die Verteilung ist in Tabelle 2.5 und in Abbildung A.4 dargestellt. In Abbildung 2.5 ist der Zusammenhang zwischen der Rauheit und dem Schienenprofil in Boxplots dargestellt.

2.1.6. Schwellentyp

Bei Betonschwellen beschreibt die Variable Schwellentyp, ob es sich um den alten Schwellentyp B70 oder den neuen Schwellentyp B91 handelt. Bei Holz- und Metallschwellen werden keine Typen unterschieden. In diesen Fällen enthält die Variable Schwellentyp das Schwellenmaterial.

Schwellentyp ist kategoriell mit 4 verschiedenen Ausprägungen. Die meisten Messungen wurden an Stellen mit B91-Schwellen gemacht. Dies gilt für die neuen Messungen noch mehr als für die bisherigen. Dadurch wird die Verteilung noch unausgeglichener – sie ist in Tabelle 2.6 und in Abbildung A.5 aufgetragen. In Abbildung A.6 ist der Zusammenhang zwischen der Rauheit und dem Schwellentyp in Boxplots dargestellt.

2.1.7. Schienenbefestigungstyp

Der Schienenbefestigungstyp setzt sich aus der Winkelführungsplatte und der Spannklemme zusammen. Für die verschiedenen Schwellenmaterialien werden unterschiedliche Schienenbefestigungen verwendet. Deshalb wäre es nicht sinnvoll, den Schienenbefestigungstyp und den Schwellentyp im

Tabelle 2.5.: Häufigkeitstabelle für das Schienenprofil

Schienenprofil	alle, N = 164	alt, N = 111	neu, N = 53
UIC 54	61 (37%)	19 (17%)	42 (79%)
UIC 60	103 (63%)	92 (83%)	11 (21%)

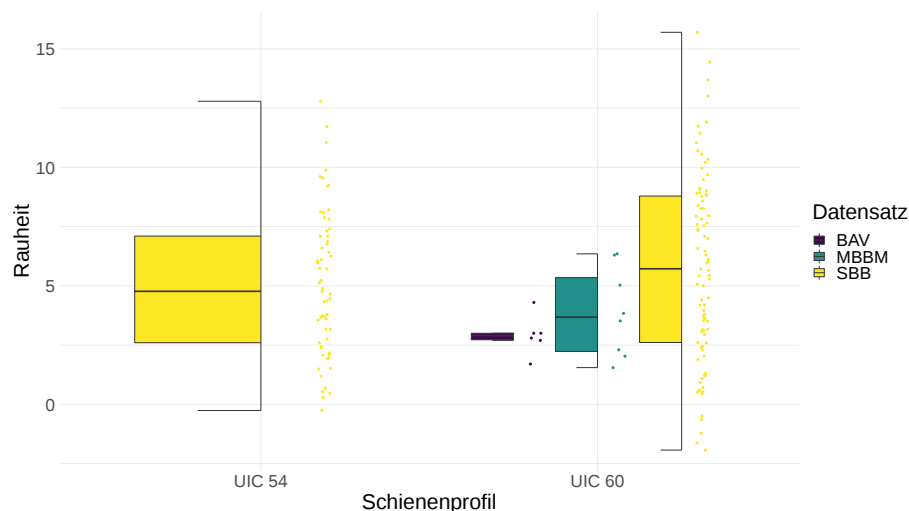


Abbildung 2.5.: Boxplots der Rauheit pro Schienenprofil und Datensatz.

Tabelle 2.6.: Häufigkeitstabelle für den Schwellentyp

Schwellentyp	alle, N = 164	alt, N = 111	neu, N = 53
B70	15 (9.1%)	14 (13%)	1 (1.9%)
B91	122 (74%)	79 (71%)	43 (81%)
Holz	12 (7.3%)	7 (6.3%)	5 (9.4%)
Stahl	15 (9.1%)	11 (9.9%)	4 (7.5%)

gleichen Modell zu verwenden. Wir haben sie deshalb zur Variable Schienenbefestigung kombiniert. Diese ist kategoriell mit 10 verschiedenen Ausprägungen. Die Verteilung ist in Tabelle 2.7 und in Abbildung A.7 aufgetragen. Viele Ausprägungen haben jedoch nur sehr wenige Beobachtungen. Deshalb werden wir die Schienenbefestigung für die Modelle nicht verwenden. In Abbildung A.8 ist der Zusammenhang zwischen der Rauheit und der Schienenbefestigung in Boxplots dargestellt.

2.1.8. Schienenzwischenlage

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

Tabelle 2.7.: Häufigkeitstabelle für die Schienenbefestigung

Schienenbefestigung	alle, N = 164	alt, N = 111	neu, N = 53
B70_W_Skl1	10 (6.1%)	9 (8.1%)	1 (1.9%)
B70_W14	1 (0.6%)	1 (0.9%)	0 (0%)
B70_Ws_Skl1	4 (2.4%)	4 (3.6%)	0 (0%)
B91_W_Skl1	8 (4.9%)	5 (4.5%)	3 (5.7%)
B91_W14	52 (32%)	51 (46%)	1 (1.9%)
B91_Ws_Skl1	62 (38%)	23 (21%)	39 (74%)
Holz_K	2 (1.2%)	2 (1.8%)	0 (0%)
Holz_Ke_Skl12	1 (0.6%)	1 (0.9%)	0 (0%)
Holz_Ke_Skl3	9 (5.5%)	4 (3.6%)	5 (9.4%)
Stahl_Aek	15 (9.1%)	11 (9.9%)	4 (7.5%)

Tabelle 2.8.: Häufigkeitstabelle für die Schienenzwischenlage

Zwischenlage	alle, N = 164	alt, N = 111	neu, N = 53
hart	122 (74%)	73 (66%)	49 (92%)
keine	2 (1.2%)	2 (1.8%)	0 (0%)
weich	40 (24%)	36 (32%)	4 (7.5%)

Tabelle 2.9.: Häufigkeitstabelle für den Anteil Güterverkehr

Anteil Güterzüge	alle, N = 164	alt, N = 111	neu, N = 53
viel	42 (26%)	32 (29%)	10 (19%)
wenig	122 (74%)	79 (71%)	43 (81%)

unterschieden. Die Zwischenlagehärte ist folgendermassen definiert:

$$\text{Zwischenlagehärte(Steifigkeit)} = \begin{cases} \text{hart} & \text{falls Steifigkeit} \geq 500 \text{ kN/mm} \\ \text{weich} & \text{sonst} \end{cases}$$

Zwischenlage ist kategoriell mit 3 verschiedenen Ausprägungen. Ihre Verteilung ist in Tabelle 2.8 und in Abbildung A.9 aufgetragen. In Abbildung A.10 ist der Zusammenhang zwischen der Rauheit und der Schienenzwischenlage in Boxplots dargestellt.

Der Wert der Zwischenlage kann bei den vorhandenen Daten fast perfekt anhand der anderen Erklärenden vorhergesagt werden. D.h. die vorkommenden Ausprägungen der Zwischenlage stimmen (im Wesentlichen) mit Ausprägungskombinationen der anderen Erklärenden überein. Deshalb bringt sie keinen Mehrwert für die Vorhersage und macht die Modelle instabiler. Aus diesem Grund verwenden wir die Zwischenlage nicht für die Modellierung.

2.1.9. 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 „wenig“ zu, den Anteilen grösser als 30% den Wert „viel“. Diese Einteilung wurde von Experten der SBB definiert.

Die resultierende Variable AnteilGz ist kategoriell mit 2 verschiedenen Ausprägungen und enthält keine fehlenden Werte. Die Verteilung ist in Tabelle 2.9 und in Abbildung A.11 aufgetragen. Es fällt auf, dass der Anteil Güterverkehr an den BAV-Messstationen gross und an den MBBM-Messstellen klein ist.

In der zweiten Iteration wurden vor allem Messungen an Standorten mit wenig Güterverkehr gemacht. Die Verteilung ist somit noch weniger ausgeglichen als nach der ersten Iteration. In Abbildung 2.6 ist der Zusammenhang zwischen der Rauheit und dem Anteil Güterverkehr in Boxplots dargestellt.

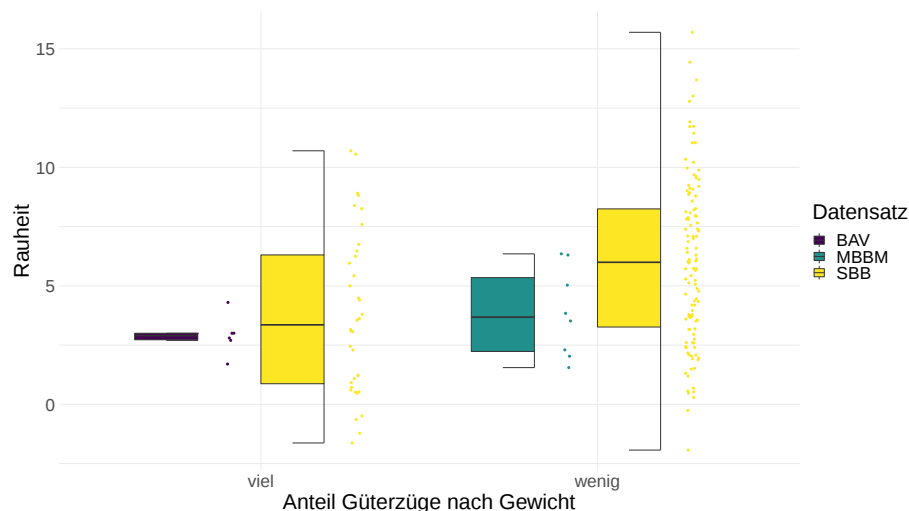


Abbildung 2.6.: Boxplots der Rauheit pro Anteil Güterzüge und Datensatz.

Tabelle 2.10.: Häufigkeitstabelle für die Streckengeometrie

Kurve	alle, N = 164	alt, N = 111	neu, N = 53
ja	15 (9.1%)	10 (9.0%)	5 (9.4%)
nein	149 (91%)	101 (91%)	48 (91%)

2.1.10. Streckengeometrie

Der Datensatz enthält die Variable Radius in Meter. Wir kategorisieren sie und berechnen die neue Variable Kurve:

$$\text{Kurve} = \begin{cases} \text{ja} & \text{falls Radius} < 1000m \\ \text{nein} & \text{sonst} \end{cases}$$

Der Grenzwert von 1000m wurde von Experten der SBB definiert. Die Verteilung ist in Tabelle 2.10 und in Abbildung A.12 aufgetragen. In Abbildung A.13 ist der Zusammenhang zwischen der Rauheit und der Variable Kurve in Boxplots dargestellt.

2.1.11. Ort

Der Datensatz enthält immer auch den Ort der Messung. In dieser Iteration verwenden wir den Ort jedoch nicht für die Modelle. Auf die damit verbundenen Probleme gehen wir im Abschnitt 3.3.2 ein.

Die Verteilung der Anzahl Messungen pro Ort ist in Tabelle 2.11 abgebildet. Für rund 52% der Orte liegt genau eine Messung vor. Für die restlichen 48% wurden mehrere Messungen durchgeführt. Am häufigsten wurde in Aarau gemessen. Dort sind es 13 Messungen.

2.1.12. Weitere Variablen

Im Schlussbericht 2020 wurden unter anderem die beiden Variablen Besohlung der Schwelle und Schotterdicke, entweder aus fachlicher Sicht oder aufgrund zu vieler fehlender Werte, als unwichtig eingestuft. Wir verwenden Sie auch in dieser Iteration nicht.

Zusätzlich wird im Datensatz das Datum der Messung angegeben. Auch diese Information verwenden wir nicht. Insbesondere berücksichtigen wir keine möglichen saisonalen Effekte.

Tabelle 2.11.: Häufigkeitstabelle für Ort.

Anzahl Messungen pro Ort	Anzahl Orte
1	32
2	9
3	6
4	3
5	2
6	3
7	1
8	1
9	2
10	1
13	1

3. Statistische Modelle

Im Abschnitt 3.1 werden die Modelle kurz erklärt, welche dann im Abschnitt 3.2 verwendet werden, um die neuen Daten zu modellieren.

Im Abschnitt 3.3 werden weitere Überlegungen zu den Daten und ihrer Modellierung ausgeführt. Insbesondere wird das Problem der abhängigen Daten behandelt.

3.1. Verwendete Modelle

In der ersten Projektphase haben wir lineare Regression verwendet, weil die bescheidene Datengrundlage bei komplexeren Modellen zu einer Überanpassung geführt hätte. Der Zusammenhang zwischen der Tonnage und der Rauheit ist jedoch offensichtlich nicht linear. Deshalb haben wir den Wertebereich der Tonnage unterteilt und dann separate lineare Modelle für kleine und grosse Tonnagewerte gerechnet.

In dieser Iteration möchten wir für die Vorhersage nur noch ein Modell verwenden, das den gesamten Wertebereich der Tonnage abdeckt. Wegen dem nichtlinearen Zusammenhang verwenden wir keine linearen Modelle mehr, sondern nur noch Regressionssplines (MARS) und Random Forest. Diese beiden Modelle können gut mit der Nichtlinearität umgehen.

3.1.1. Multivariate adaptive Regressionssplines (MARS)

MARS generiert abschnittsweise lineare Funktionen durch schrittweises Durchsuchen aller möglichen (univariaten) Haltepunkten (Breakpoints) und über Interaktionen zwischen allen Variablen. Ein adaptiver Regressionsalgorithmus wird für die automatische Auswahl der Haltepunkte verwendet. Der MARS-Algorithmus umfasst eine Vorwärtsphase und eine Rückwärtsphase.

In der Vorwärtsphase werden an zufälligen Positionen innerhalb des Wertebereichs jeder kontinuierlichen Variablen Haltepunkte platziert, um zwei lineare Funktionen zu definieren. Bei jedem Schritt passt das Modell den Knoten und die entsprechenden zwei Funktionen an, um die maximale Reduzierung des Fehlers zu erreichen. Dieser Prozess des Hinzufügens von linearen Funktionen wird wiederholt, bis die maximale Anzahl erreicht ist. Dies führt normalerweise zu einem sehr komplizierten und überangepassten Modell.

In der Rückwärtsphase werden die redundanten Funktionen gelöscht, die am wenigsten dazu beigetragen haben den Fehler zu reduzieren. Ausserdem wird am Ende jeder Iteration die Implementierung eines Interaktionseffekts auf mögliche Modellverbesserungen überprüft.

Dieses schrittweise Zerlegen in separate Regionen mit unterschiedlichen linearen Beziehungen wird in unserem Fall nur für die Tonnage vorgenommen, da sie die einzige kontinuierliche Erklärende ist. Da Interaktionseffekte höherer Ordnung schwer zu interpretieren sind, werden nur Interaktionseffekte 1. Ordnung (i.e. zwischen zwei Variablen) erlaubt.

Der Mars Algorithmus wird in [2] ausführlicher beschrieben. Details zum verwendeten R Package können in [4] nachgeschlagen werden.

3.1.2. Random Forest

Eine andere nichtlineare Methode sind Regressionsbäume, die vereinfacht gesagt eine Reihe von Prädiktorvariablen verwenden, um Entscheidungsbäume zu erstellen, die den Wert einer Zielvariablen vorhersagen. Der Vorteil von Entscheidungsbäumen besteht darin, dass sie gut mit nichtlinearen Zusammenhängen umgehen können. Ein Nachteil besteht darin, dass sie tendenziell unter einer

hohen Variabilität leiden und je nach Datensatz sehr unterschiedliche Ergebnisse generieren. Ein Random Forest löst das Problem der Variabilität, indem eine Vielzahl an Regressionsbäumen gebildet und die Ergebnisse kombiniert werden. Das resultierende Modell ist dann viel stabiler als ein einzelner Baum. Für diese Stabilität bezahlt man jedoch mit einer geringeren Interpretierbarkeit der Effekte der erklärenden Variablen.

Siehe [1] für eine ausführliche Beschreibung des Algorithmus.

3.1.3. 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 Zielvariablen möglichst gut durch die Erklärenden beschreiben? Soll bestimmt werden, welche Erklärenden einen Einfluss auf die Zielvariable haben? Oder möchte man den Wert der Zielvariablen für zukünftige Beobachtungen möglichst gut vorhersagen?

Nach Absprache mit den Experten der SBB betrachten wir den Vorhersagefehler, die Stabilität der Vorhersage und die Interpretierbarkeit.

Vorhersagefehler

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. Dafür unterteilen wir den Datensatz in 10 ungefähr gleich grosse Gruppen. Dann nehmen wir eine Gruppe aus dem Datensatz heraus und rechnen ein Modell mit den übrigen Daten. Dann verwenden wir dieses Modell um Schätzungen für die entfernten Beobachtungen zu machen. Dies wiederholen wir für alle Gruppen. So erhalten wir für jede Beobachtung einen unverfälschten Schätzwert und können damit den RMSE berechnen.

Stabilität der Vorhersage

Ein weiteres Gütekriterium ist die Stabilität eines Modells. Wenn man die Daten ein wenig abändert, sollte sich der RMSE nicht zu stark verändern. Mit der Kreuzvalidierung erhalten wir 10 Modelle, für die wir jeweils den RMSE berechnen können. Die Standardabweichung dieser 10 Werte ist ein Mass für die Stabilität der Vorhersage.

Tabelle 3.1.: Vorhersagefehler für MARS und Random Forest.

Modell	RMSE _{CV}	SD(RMSE _{CV})
MARS	2.13	0.29
Random Forest	2.17	0.33

Tabelle 3.2.: Koeffizienten für das MARS-Modell.

	Koeffizient
(Intercept)	-73.397
h(4.67842-MTonnage)	17.135
h(MTonnage-0.254677)	16.398
h(MTonnage-4.67842)	-16.380
Stahl: R350	2.923
Anteil Gz: wenig	2.296
Profil: UIC 60	2.123

Interpretierbarkeit

Ein weiteres Kriterium ist die Interpretierbarkeit. Die Effekte der einzelnen Variablen sind im MARS-Modell viel einfacher zu interpretieren als im Random Forest. Dies spricht für das MARS-Modell, welches deshalb bei ähnlichem Vorhersagefehler zu bevorzugen ist.

Zudem wird man in der Regel ein Modell bevorzugen, das man fachlich besser interpretieren kann. Der Zusammenhang zwischen den verwendeten erklärenden Variablen und der Zielvariable soll aus fachlicher Sicht möglichst gut erklärbar sein.

3.2. Modellierung der neuen Daten

Die verwendeten Modelle machen zwar automatisch eine Variablenselektion. Wenn ihnen aber zu viele Variablen zur Verfügung stehen, werden sie instabil. Insbesondere, wenn die Erklärenden korreliert sind oder wenn die in den Daten vorkommenden Ausprägungen einer kategoriellen Variablen (im Wesentlichen) mit Ausprägungskombinationen der anderen Erklärenden übereinstimmen. In den vorhandenen Daten ist das, wie bereits in Abschnitt 2.1.8 erwähnt, für die Zwischenlage der Fall. Deshalb verwenden wir sie nicht für die Modelle.

Wie bereits in Abschnitt 2.1 erwähnt, verwenden wir auch die Variablen Schienenbefestigungstyp, Besohlung der Schwelle und Schotterdicke nicht für die Modellierung.

Tabelle 3.1 zeigt den RMSE und seine Standardabweichung für das MARS-Modell und den Random Forest (RF). MARS liefern einen leicht kleineren Fehler als RF. Sie weisen auch eine tiefere Standardabweichung auf, was auf eine stabilere Vorhersage hinweist. MARS haben zusätzlich den Vorteil, dass die Effekte der erklärenden Variablen besser interpretierbar sind. Aus diesen drei Gründen verwenden wir für die Vorhersage der Rauheit das MARS-Modell.

Die Koeffizienten des MARS-Modells sind in Tabelle 3.2 dargestellt. MTonnage bezeichnet die Tonnage in Mt. Der Buchstabe h steht für Hinge function:

$$h(x) := \max(0, x)$$

Die einzelnen Effekte für die Tonnage sind also auf der einen Seite des entsprechenden Knotens gleich Null und auf der anderen Seite linear im Abstand zum Knoten. Aus den drei Effekten resultiert eine stückweis lineare Funktion mit den in Tabelle 3.3 angegebenen Steigungen. In Abbildung 3.1 ist die Regressionskurve für die Kombination Stahl R260, Profil UIC 60 und wenig Güterzüge aufgezeichnet. Für andere Ausprägungskombinationen wird die Kurve den Koeffizienten entsprechend vertikal verschoben.

Tabelle 3.3.: Steigungen der stückweis linearen Regressionskurve des MARS-Modells.

Tonnageabschnitt	Steigung
0 bis 0.255	-17.135
0.255 bis 4.678	-0.737
> 4.678	0.018

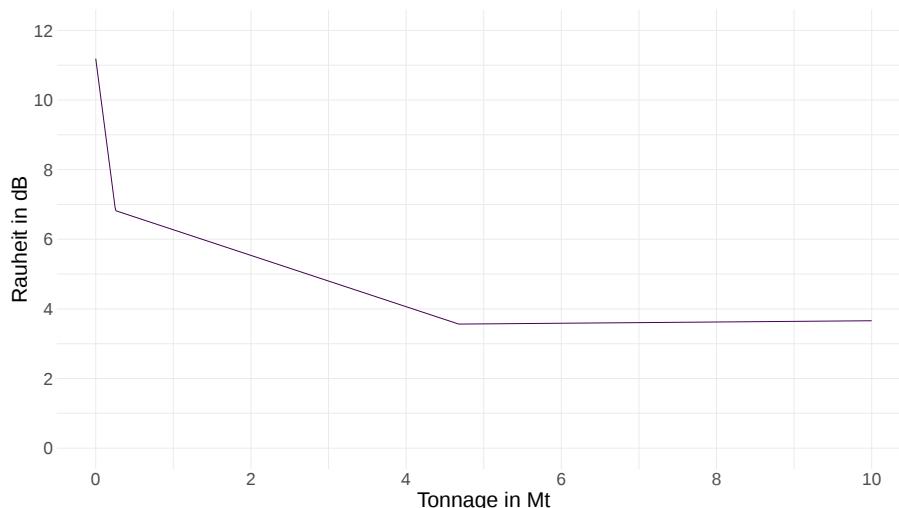


Abbildung 3.1.: Regressionskurve für Stahl R260, Profil UIC 60 und wenig Güterzüge.

Die Haltepunkte wurden, verglichen mit dem Modell der ersten Iteration, verschoben. Der untere Knoten lag bei 0.206 und ist jetzt leicht höher bei 0.255. Der obere Knoten lag bei 7.495 und wurde jetzt deutlich nach unten verschoben zum Punkt 4.678. Das erste Stück mit der starken negativen Steigung ist also ein wenig länger geworden und das letzte Stück, in dem die Tonnage fast keinen Einfluss mehr auf die Rauheit hat, beginnt nun früher.

Das Modell verwendet neben der Tonnage nur die Stahlqualität, das Schienenprofil und den Anteil Güterzüge. Interaktionseffekte werden keine ins Modell genommen, es ist also ein additives Modell. Auf die Effekte der erklärenden Variablen gehen wir im Abschnitt 3.3.1 genauer ein.

Verglichen mit dem Modell der ersten Iteration wurde die Vorhersage nicht deutlich besser, aber das Modell wurde deutlich stabiler, die Standardabweichung des RMSE ist nun deutlich tiefer.

3.3. Weitere Überlegungen

3.3.1. Effekte der erklärenden Variablen

Tonnage

Wie erwartet nimmt die Rauheit mit steigender Tonnage am Anfang sehr stark ab. Die Kurve flacht dann ab und für grosse Tonnagen ist die Steigung wieder positiv, aber sehr nahe bei Null. Die Steigung ist so klein, dass wir mit den verfügbaren Daten keinen Hinweis auf einen relevanten „Badewanneneffekt“ erkennen können, also darauf, dass die Rauheit mit sehr grossen Tonnagewerten wieder zunimmt. Um eine verlässliche Aussage dazu machen zu können, bräuchten wir aber viel mehr Messungen mit sehr grossen Tonnagewerten.

In der ersten Iteration lag die Steigung der Rauheit für grosse Tonnagen bei 0.0356. Mit 0.018 wird sie mit dem aktuellen Modell nur noch rund halb so gross geschätzt.

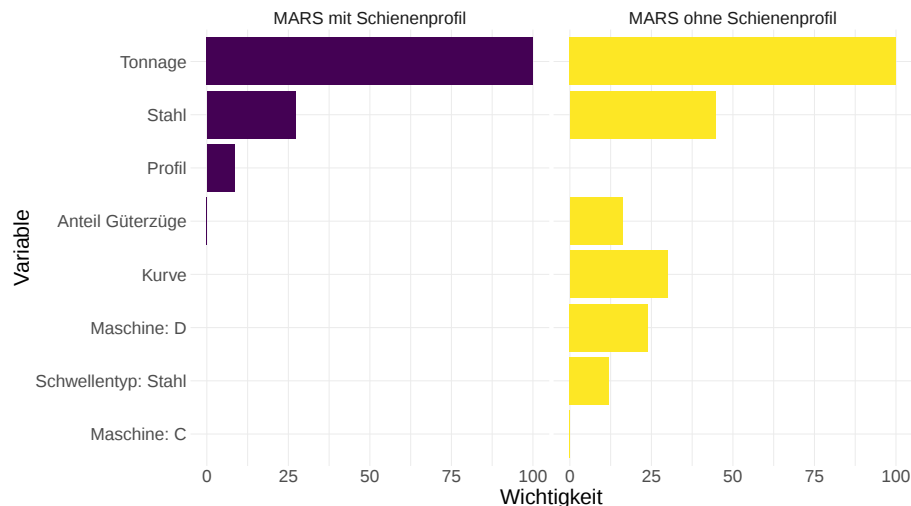


Abbildung 3.2.: Wichtigkeit der Variablen basierend auf der Auswirkung auf den Vorhersagefehler, wenn Variablen hinzugefügt werden.

Die Variablen Zwischenlage, Schienenbefestigungstyp, Besohlung der Schwelle und Schotterdicke wurden nicht für die Modellierung verwendet.

Variablenwichtigkeiten

In Abbildung 3.2 sind die Variablenwichtigkeiten für die Erklärung der Rauheit aufgeführt. Vereinfacht gesagt, wird die Wichtigkeit von Variablen berechnet, indem die durchschnittliche Verringerung des Vorhersagefehlers berechnet wird, wenn erklärende Variablen hinzugefügt werden. Alle Werte sind so skaliert, dass sie einen maximalen Wert von 100 haben, welcher sich hier auf die Variable Tonnage bezieht. Die Werte zwischen den Modellen können nicht verglichen werden. Vielmehr kann die Reihenfolge der Variablen innerhalb eines Modells einen Aufschluss über die Wichtigkeit geben.

Auf der linken Seite der Abbildung sind die Variablenwichtigkeiten für das Modell aufgeführt, das wir für die Vorhersagen verwenden. Neben der Tonnage ist vor allem die Stahlqualität wichtig für die Vorhersage. Das ist aus fachlicher Sicht einleuchtend. Danach kommt noch das Schienenprofil. Dabei ist es aus fachlicher Sicht nicht klar, ob das Profil wirklich einen Einfluss hat. Es könnte sein, dass sich die Stellen an denen unterschiedliche Profile verbaut wurden, durch andere relevante Faktoren unterscheiden, die jedoch nicht erhoben wurden.

Deshalb haben wir ein MARS-Modell gerechnet, bei dem das Profil als Erklärende explizit ausgeschlossen wurde. Der Vorhersagefehler dieses Modells ist nur leicht grösser als derjenige des Modells mit Profil. Die Stabilität ist jedoch deutlich tiefer. Das Modell verwendet nun zusätzlich die Variablen Kurve, Maschine und Schwellentyp. Die Variablenwichtigkeiten für dieses Modell sind auf der rechten Seite der Abbildung 3.2 aufgezeichnet. Die Vorhersagefehler und Koeffizienten sind in den Tabellen A.1 und A.2 im Anhang aufgeführt.

3.3.2. Abhängigkeiten zwischen den Messungen

Das MARS-Modell setzt spätestens bei der Berechnung von Prädiktionsintervallen voraus, dass die Beobachtungen statistisch unabhängig voneinander sind. Dies ist bei den vorliegenden Daten nur bedingt gegeben. Zum Teil wurden mehrere Messungen am gleichen Standort durchgeführt. Dies 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 pro Ausprägungskombination 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.

Eine andere Möglichkeit wäre, pro Ort nur eine Messung zu verwenden. Dann müssten wir jedoch auf

103 der 164 Messungen verzichten, was zu deutlich schlechteren Vorhersagen führen würde. Weil unser primäres Ziel die Vorhersage der Rauheit ist, verwenden wir alle verfügbaren Messungen.

4. Schätzung der Rauheit

In diesem Abschnitt stellen wir die Schätzung der Rauheit dar. 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 10 dB ist.

Alle Schätzungen in diesem Abschnitt sind mit Vorsicht zu interpretieren. Auch wenn die Datenbasis mit jeder Iteration solider und somit die Modelle stabiler werden, bestehen weiterhin grosse Unsicherheiten in den Schätzungen, vor allem in Bereichen mit wenigen Messungen.

Zudem benutzen wir zum Teil mehrere Messungen pro Standort. Die für die Berechnung der Prädiktionsintervalle eigentlich notwendige Unabhängigkeit der einzelnen Beobachtungen ist also nur bedingt gegeben.

Für die Schätzung der Rauheit verwenden wir das MARS-Modell, welches neben der Tonnage die kategoriellen Variablen Stahlqualität, Profil und Anteil Güterzüge berücksichtigt. Es liefert also für jede der acht Ausprägungskombinationen dieser kategoriellen Variablen eine Regressionskurve.

In Abbildung 4.1 sind die Kurven abgebildet und in Tabelle 4.1 sind zu jeder Kombination diverse Informationen aufgeführt. Die geschätzte Rauheit liegt für vier der acht Kombinationen bereits bei Tonnage 0 unter dem kritischen Grenzwert von 10 dB. Auch für die anderen Kombinationen fällt die Rauheit ziemlich bald unter den Grenzwert. Zu sehen ist das an der „Tonnagegrenze Schätzung“. Sie bezeichnet den Tonnagewert, an dem die geschätzte Rauheit unter 10 dB fällt. Dies ist für alle Kombinationen spätestens bei 0.23 Mt der Fall.

Die Schätzungen sind natürlich mit einer gewissen Unsicherheit behaftet. Wir quantifizieren sie mit 95%-Prädiktionsintervallen. 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%-Konfidenzintervallen, die den wahren Mittelwert mit einer Sicherheit von 95% enthalten, geben die Ausprägungen der Erklärenden.

Die Abbildungen 4.2 bis 4.4 und A.14 bis A.18 im Anhang zeigen die Regressionskurven (blau) und Prädiktionsbänder (hellblau) für alle möglichen Kombinationen von Stahlqualität, Profil und Anteil Güterzüge. Der kritische Grenzwert von 10 dB ist jeweils als rote horizontale Linie eingezeichnet.

Die in Tabelle 4.1 aufgeführte „Tonnagegrenze Prädiktionsband“ bezeichnet den Tonnagewert, an dem die obere Grenze des Prädiktionsintervalls unter 10 dB fällt. Die Kombination Stahl R350, Profil UIC 60 und wenig Güterzüge ist die einzige, für die das nie passiert (vgl. Abbildung 4.3). Für alle anderen Kombinationen liegt die Tonnagegrenze höchstens bei 3.18 Mt, wobei sie für die meisten

Tabelle 4.1.: Geschätzte Rauheit und Tonnagegrenzen für alle acht möglichen Kombinationen von Stahlqualität, Profil und Anteil Güterzüge.

Stahl	Profil	Anteil Güterzüge	Relevante Strecke in km	Rauheit für Tonnage 0	Tonnagegrenze Schätzung	Tonnagegrenze Prädiktionsband
R350	UIC 60	wenig	204	14.109	0.23	NA
R350	UIC 54	wenig	111	11.985	0.11	3.18
R350	UIC 60	viel	112	11.813	0.10	2.95
R260	UIC 60	wenig	911	11.186	0.06	2.09
R350	UIC 54	viel	63	9.690	NA	0.24
R260	UIC 54	wenig	406	9.063	NA	0.21
R260	UIC 60	viel	578	8.890	NA	0.19
R260	UIC 54	viel	28	6.767	NA	0.07

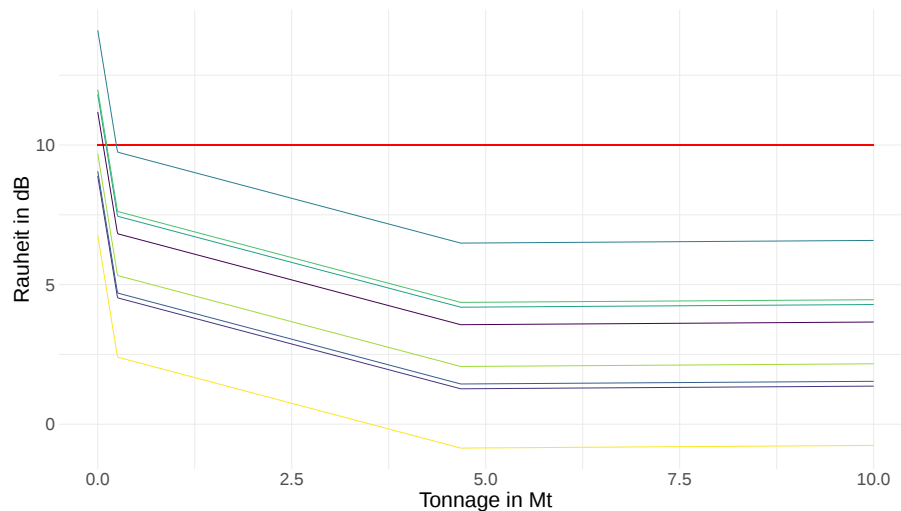


Abbildung 4.1.: Regressionskurven für alle acht möglichen Kombinationen von Stahlqualität, Profil und Anteil Güterzüge. Der kritische Grenzwert von 10 dB ist als rote horizontale Linie eingezeichnet.

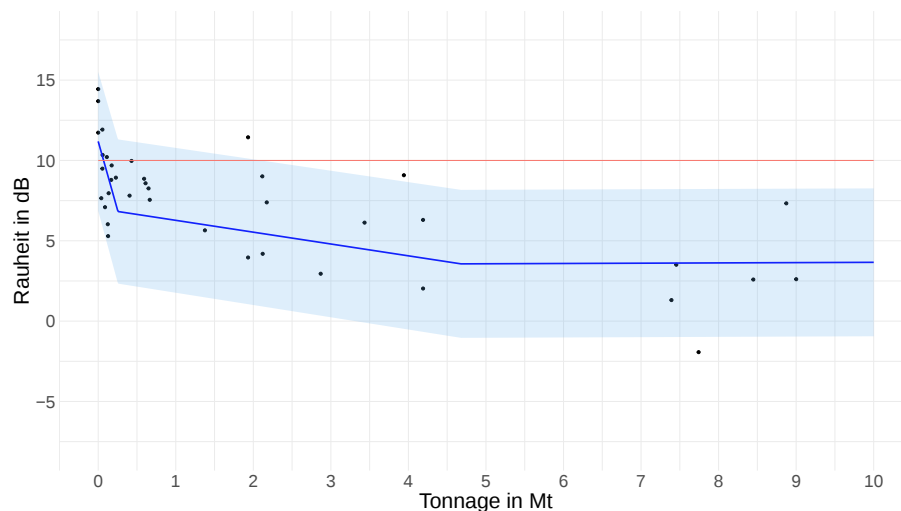


Abbildung 4.2.: Prädiktionsband für Stahl R260, Profil UIC 60 und wenig Güterzüge.

Kombinationen noch deutlich tiefer liegt. Abbildung 4.3 zeigt das am tiefsten liegende Prädiktionsband. Es gehört zur Kombination Stahl R260, Profil UIC 54 und viel Güterzüge. Abbildung 4.2 zeigt die Kombination Stahl R260, Profil UIC 60 und wenig Güterzüge. Sie ist am weitesten verbreitet im SBB-Schienennetz in dem Sinne, dass die Summe der Länge der relevanten Strecken am grössten ist.

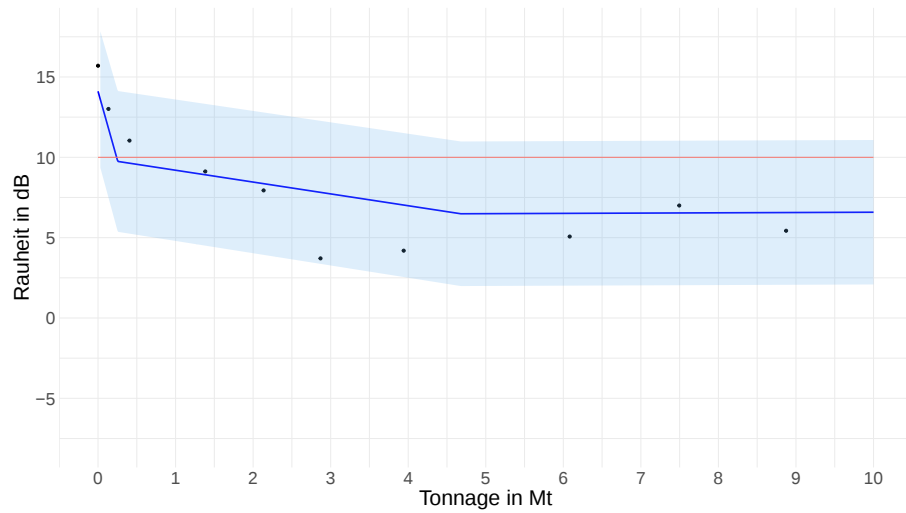


Abbildung 4.3.: Prädiktionsband für Stahl R350, Profil UIC 60 und wenig Güterzüge.

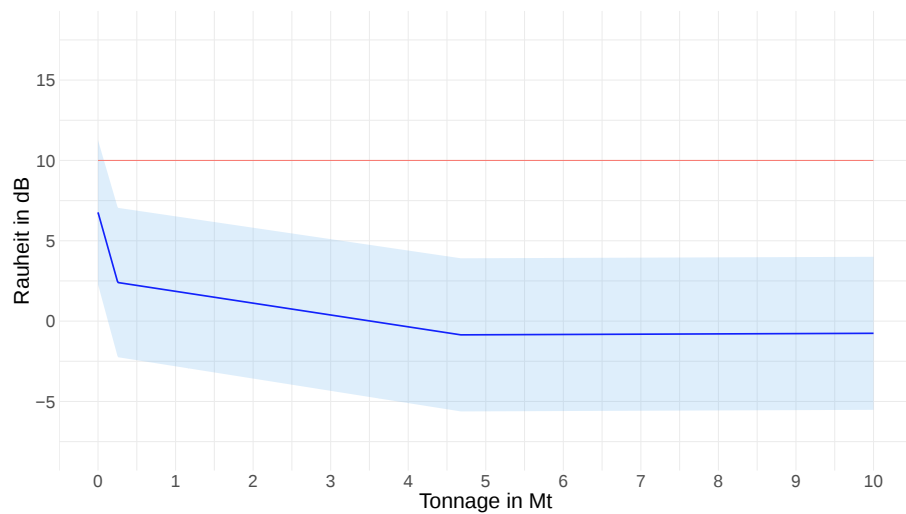


Abbildung 4.4.: Prädiktionsband für Stahl R260, Profil UIC 54 und viel Güterzüge.

5. Messstrategie und Ausblick

Die SBB haben eine Liste mit den 14 wichtigsten Oberbaukombinationen zusammengestellt. Es sind die Kombinationen, die im SBB-Schienennetz auf relevanten Strecken von total mindestens 50km Länge verbaut sind. Dabei wurden nur Strecken mit einer Belastung von mindesten 15'000 t pro Tag berücksichtigt. Ausgeschlossen wurden zudem alle Tunnelstrecken und die Neubaustrecke Mattstetten-Rothrist. Da der Oberbau nicht überall vollständig dokumentiert ist, handelt es sich bei den verwendeten Streckenlängen um approximative Werte.

In der dritten Iteration sollen vor allem Messungen an Stellen mit diesen Oberbaukombinationen durchgeführt werden. Die Kombinationen sind in Tabelle 5.1 aufgeführt.

Diese 14 Kombinationen priorisieren wir nun nach den folgenden beiden Kriterien:

1. Kombinationen mit bisher wenigen Messungen
Die Beobachtungen sollen möglichst gleichmässig im relevanten Teil des Parameterraums verteilt sein. Je ausgeglichener die Verteilung ist, desto genauer und stabiler werden die Vorhersagen im Allgemeinen.
2. Kombinationen mit hoher geschätzter Rauheit
Interessant sind vor allem Kombinationen mit hoher geschätzter Rauheit. Für diese möchten wir die Genauigkeit der Schätzung erhöhen und so eine bessere Aussage darüber machen, ab welchem Tonnagewert der Grenzwert von 10 dB unterschritten wird.

Die Kombination 9 erhält so die höchste Priorität. Die entsprechende obere Grenze des Prädiktionsbands unterschreitet den Grenzwert nie. Zudem steht uns nur eine Beobachtung von dieser Oberbaukombination zur Verfügung.

Danach kommen die Kombinationen 13, 4 und 7, die auch hohe geschätzte Rauheiten und wenige Messungen aufweisen.

Mögliche Kandidaten wären auch die Kombinationen 12, 14, 3 und 11, die zwar eher tiefe geschätzte Rauheiten aufweisen, für die uns aber nur wenige Messungen zur Verfügung stehen.

Für die Wahl der Tonnage ist die Ausgangslage ähnlich wie nach der ersten Iteration. Interessant sind vor allem die kleinen bis mittleren Tonnagen im Bereich von 0 bis 5 Mt. Nach dem zweiten Haltepunkt des aktuellen MARS-Modells bei 4.678 passiert nicht mehr viel. Die Steigung ist zwar leicht positiv, aber so nahe bei Null, dass man nicht von einem „Badewanneneffekt“ sprechen kann. Das heisst aber nicht, dass wir gezeigt hätten, dass es keinen „Badewanneneffekt“ gibt. Für eine verlässliche Aussage dazu bräuchten wir mehr Messungen mit sehr grossen Tonnagewerten.

Die Tonnagewerte zwischen 0 und 5 Mt sollten so gewählt werden, dass alle Messungen einer Kombination möglichst gut auf dem Intervall von 0 bis 5 Mt verteilt sind. Da es aber nicht einfach ist, für eine bestimmte Kombination aus Oberbau und Tonnage eine geeignete Messstelle zu finden, beschränken wir uns darauf, grobe Bereiche für die Tonnage anzugeben. In Tabelle 5.2 sind die empfohlenen Kombinationen für zusätzliche Messungen noch einmal zusammengefasst. Für jede dieser Kombinationen ist ein Tonnagebereich angegeben, in dem idealerweise Messungen durchgeführt werden sollten.

In der dritten Iteration werden wir mit den neu gemessenen Daten ein letztes Mal die Modelle anpassen und die Schätzung der Rauheit aktualisieren. Auf dieser Grundlage werden wir dann eine Simulationsstudie durchführen. Mit dieser Studie soll gezeigt werden, ob diese einfachen statistischen Modelle genügend robust sind und ob trotz geringer Datenmengen solch pragmatische Modelle ausreichen können, die Schienenrauheit mit einer akzeptablen Unsicherheit genügend genau zu schätzen.

Tabelle 5.1.: Die wichtigsten Oberbaukombinationen

Kombination	Anzahl Messungen	Stahlqualität	Schienenprofil	Anteil Gz	Schwellentyp	Schienenbefestigung	Kurve	Zwischenlage
1	5	R260	UIC 60	wenig	B91	W_Skl1	nein	hart
2	7	R260	UIC 60	wenig	B91	Ws_Skl1	nein	hart
3	3	R260	UIC 60	viel	B91	W_Skl1	nein	hart
4	1	R260	UIC 60	wenig	B70	W_Skl1	nein	hart
5	14	R260	UIC 54	wenig	Stahl	Aek	nein	weich
6	9	R260	UIC 60	viel	B70	W_Skl1	nein	hart
7	4	R260	UIC 60	wenig	Holz	Ke_Skl3	nein	hart
8	15	R260	UIC 60	viel	B91	Ws_Skl1	nein	hart
9	1	R350	UIC 60	wenig	B91	Ws_Skl1	ja	hart
10	9	R260	UIC 60	wenig	B91	W14	nein	hart
11	3	R260	UIC 60	viel	Holz	Ke_Skl3	nein	hart
12	2	R260	UIC 54	wenig	Holz	K	nein	keine
13	1	R350	UIC 60	viel	B91	Ws_Skl1	ja	hart
14	2	R350	UIC 54	viel	Holz	Ke_Skl3	ja	hart

Tabelle 5.2.: Empfohlene Kombinationen für zusätzliche Messungen.

Kombination	Tonnagebereich in Mt
9	0 bis 4
13	0 bis 5
4	0 bis 2.5 oder 4.5 bis 5
7	0 bis 3
3	0 bis 1.5 oder 4 bis 5
11	1.5 bis 5
12	1.5 bis 5
14	0 bis 5

A. Anhang

A.1. Beschreibung der einzelnen Variablen

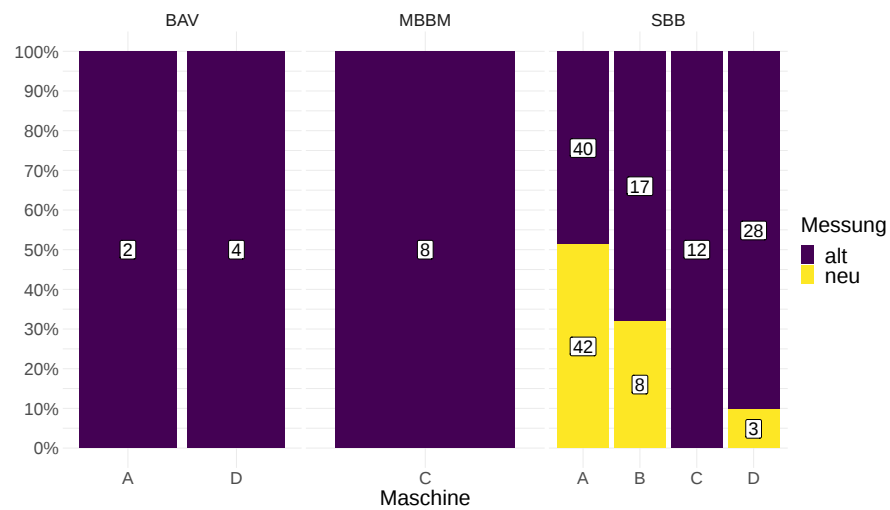


Abbildung A.1.: Balkendiagramm der Schleiftechnik pro Datensatz und Messung.

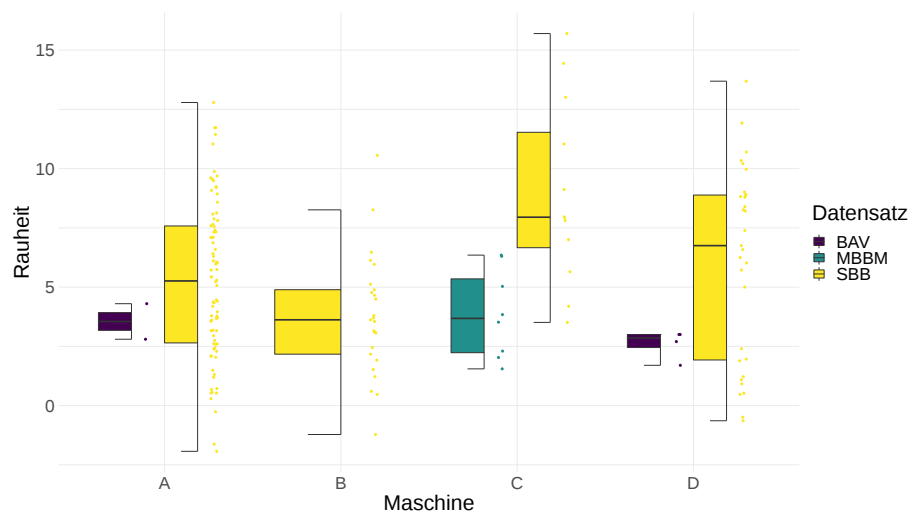


Abbildung A.2.: Boxplots der Rauheit pro Schleiftechnik und Datensatz.

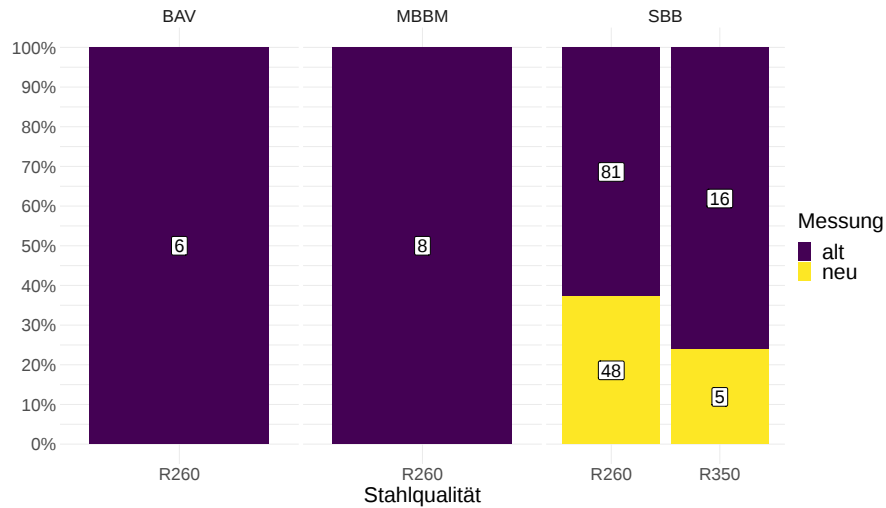


Abbildung A.3.: Balkendiagramm der Stahlqualität pro Datensatz und Messung.

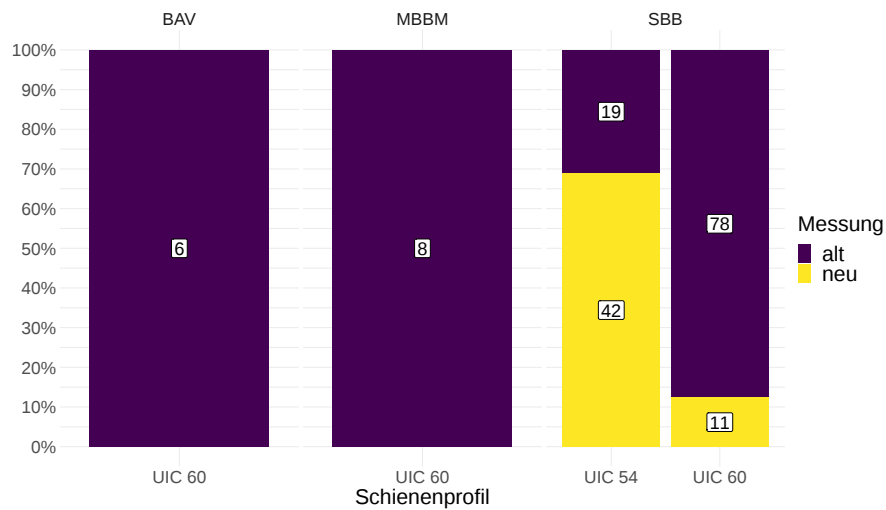


Abbildung A.4.: Balkendiagramm für das Schienenprofil pro Datensatz und Messung.

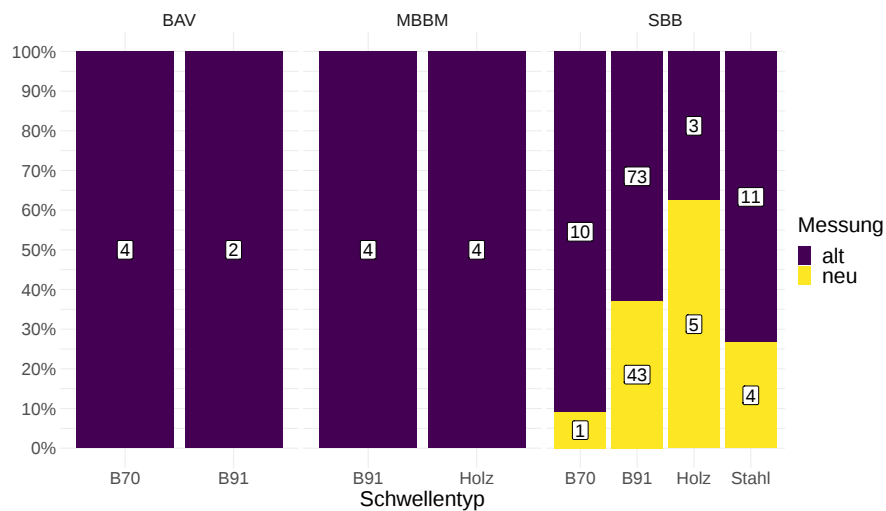


Abbildung A.5.: Balkendiagramm des Schwellentyps pro Datensatz und Messung.

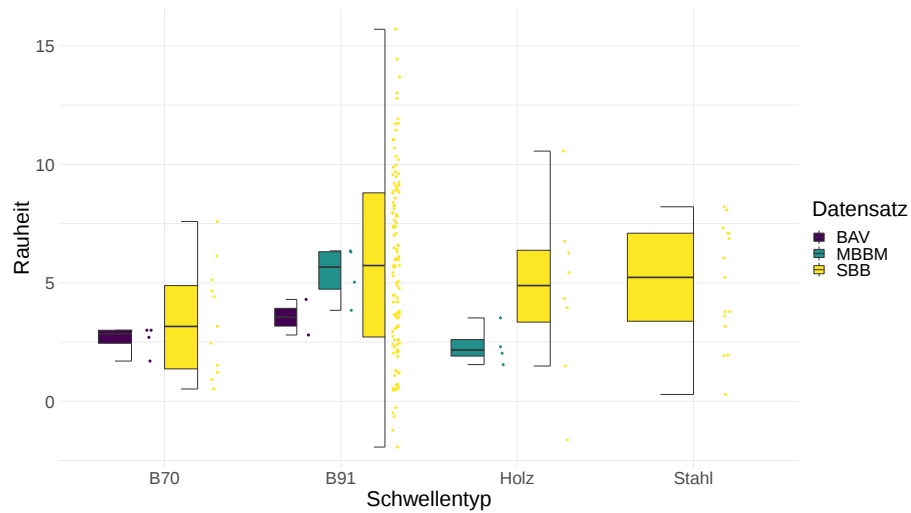


Abbildung A.6.: Boxplots der Rauheit pro Schwellentyp und Datensatz.

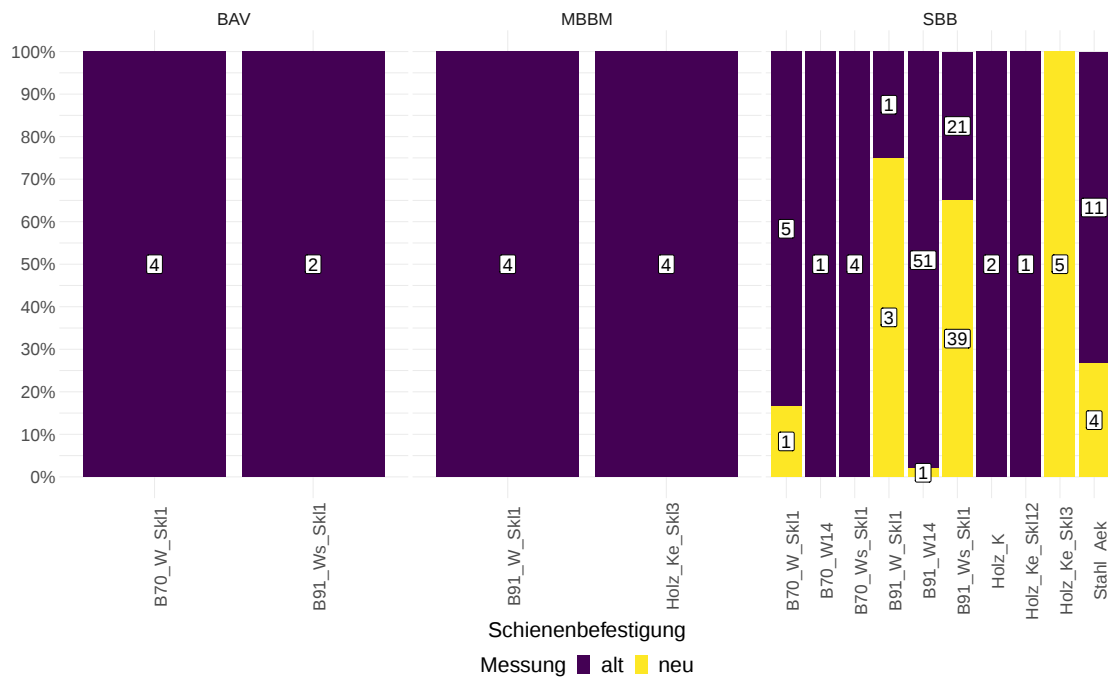


Abbildung A.7.: Balkendiagramm der Schienenbefestigung pro Datensatz und Messung.

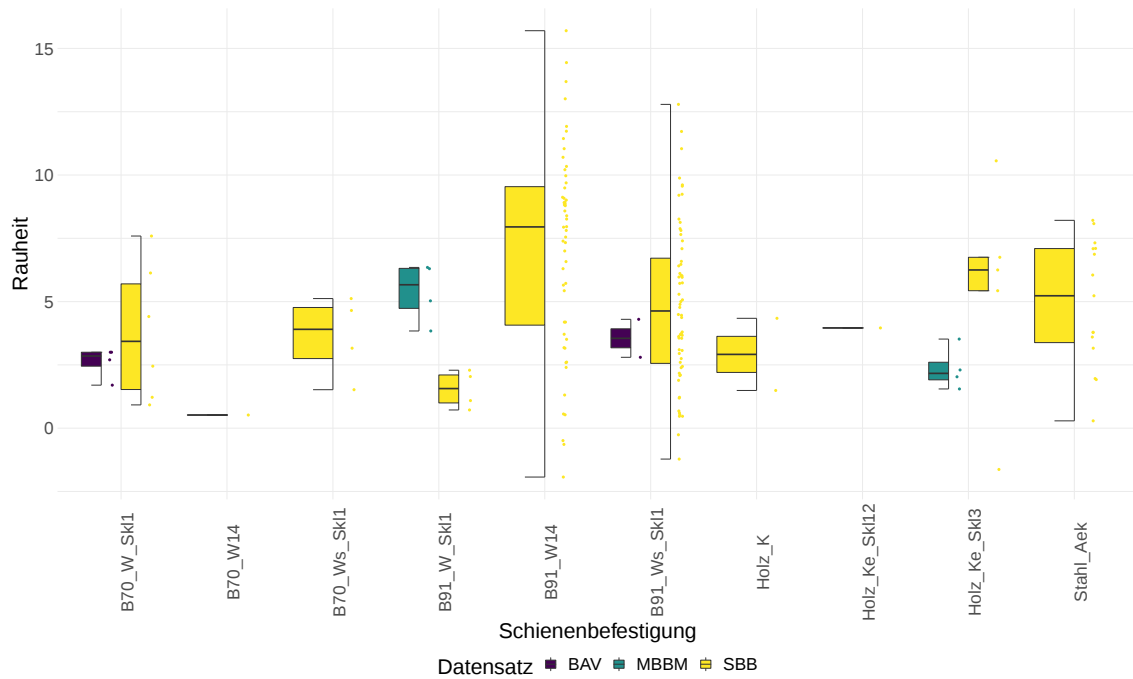


Abbildung A.8.: Boxplots der Rauheit pro Schienenbefestigungstyp und Datensatz.

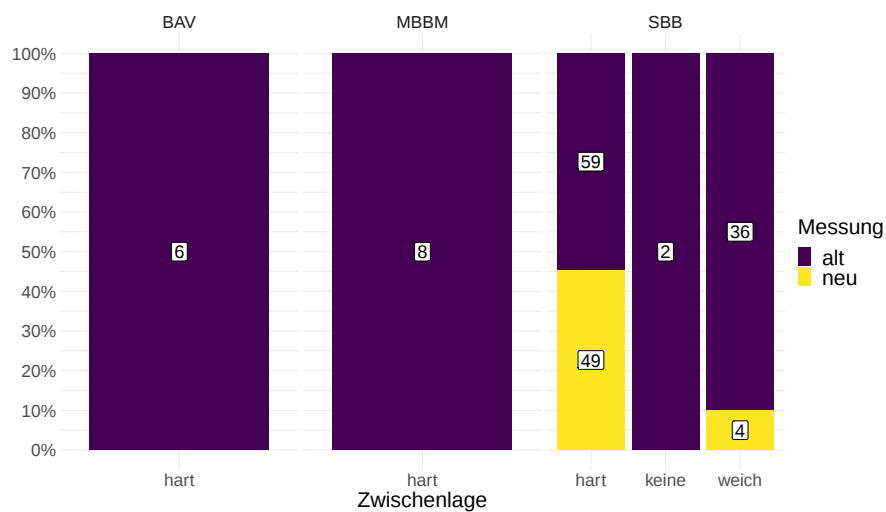


Abbildung A.9.: Balkendiagramm der Zwischenlage pro Datensatz und Messung.

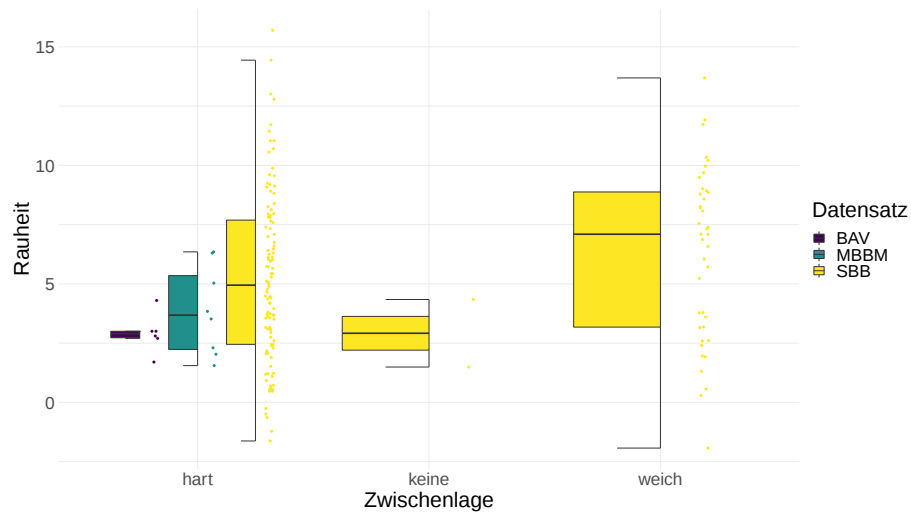


Abbildung A.10.: Boxplots der Rauheit pro Zwischenlage und Datensatz.

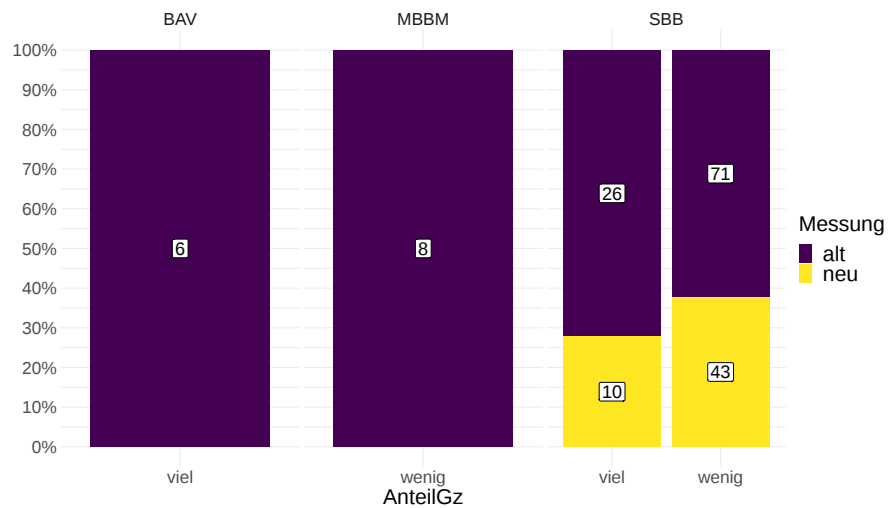
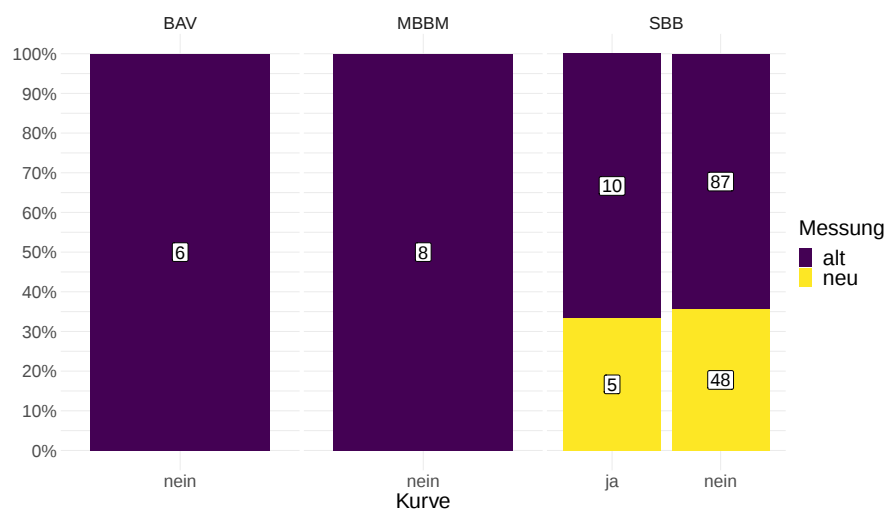


Abbildung A.11.: Balkendiagramm von Anteil Güterzüge pro Datensatz und Messung.



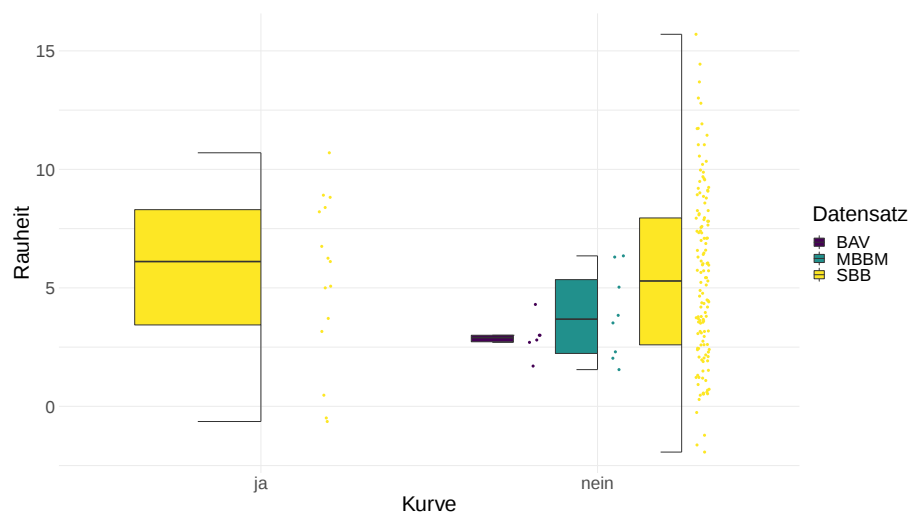


Abbildung A.13.: Boxplots der Rauheit pro Streckengeometrie und Datensatz.

A.2. Details zum MARS-Modell ohne Schienenprofil

Tabelle A.1.: Vorhersagefehler für MARS mit und ohne Schienenprofil.

Modell	RMSE _{CV}	SD(RMSE _{CV})
MARS mit Profil	2.13	0.29
MARS ohne Profil	2.16	0.41

Tabelle A.2.: Koeffizienten für MARS-Modell ohne Schienenprofil.

	Koeffizient
(Intercept)	-62.648
h(4.18906-MTonnage)	16.395
h(MTonnage-0.278741)	15.679
h(MTonnage-4.18906)	-15.675
Stahl: R350	3.807
Anteil Gz: wenig	1.330
Schwelle: Stahl	-1.807
Maschine: D	1.827
Maschine: C	1.330
Kurve: nein	2.559

A.3. Prädiktionsbänder

Die Abbildungen A.14 bis A.18 zeigen die Regressionskurven (blau) und Prädiktionsbänder (hellblau) für verschiedene Kombinationen von Stahlqualität, Profil und Anteil Güterzüge. Der kritische Grenzwert von 10 dB ist jeweils als rote horizontale Linie eingezeichnet.

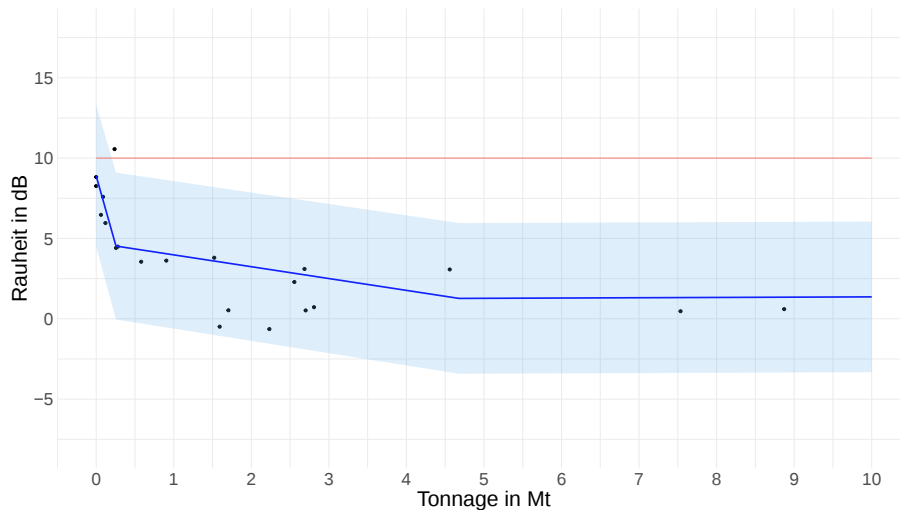


Abbildung A.14.: Prädiktionsband für Stahl R260, Profil UIC 60 und viel Güterzüge.

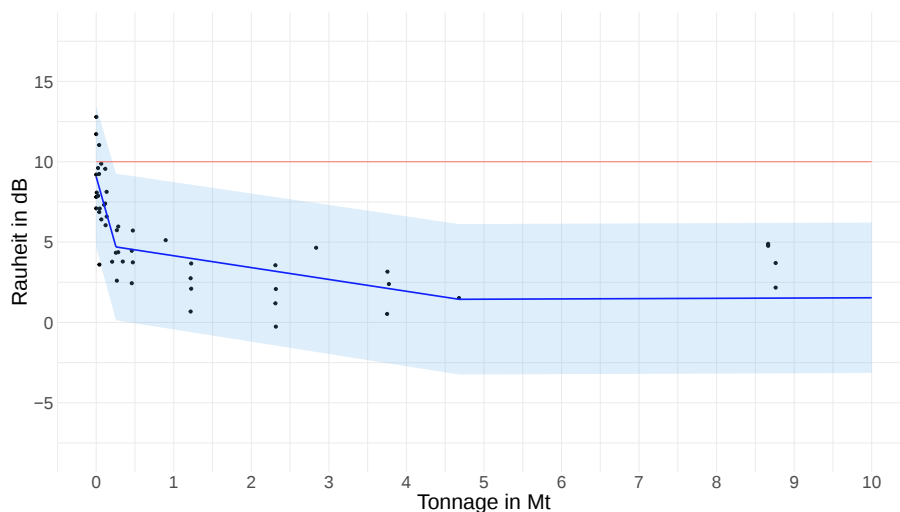


Abbildung A.15.: Prädiktionsband für Stahl R260, Profil UIC 54 und wenig Güterzüge.

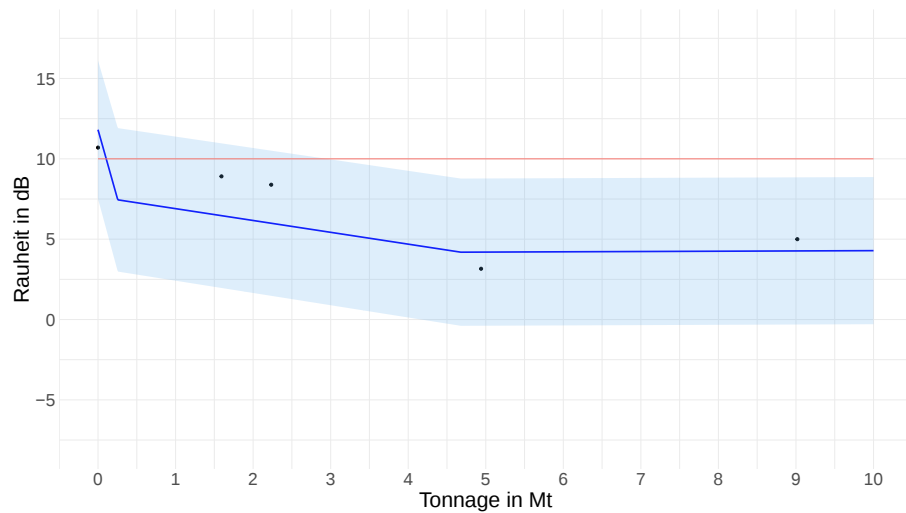


Abbildung A.16.: Prädiktionsband für Stahl R350, Profil UIC 60 und viel Güterzüge.

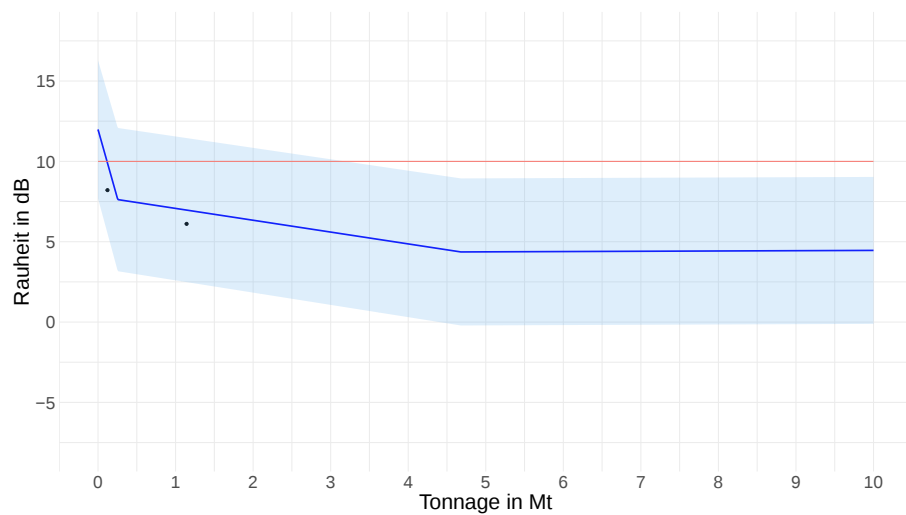


Abbildung A.17.: Prädiktionsband für Stahl R350, Profil UIC 54 und wenig Güterzüge.

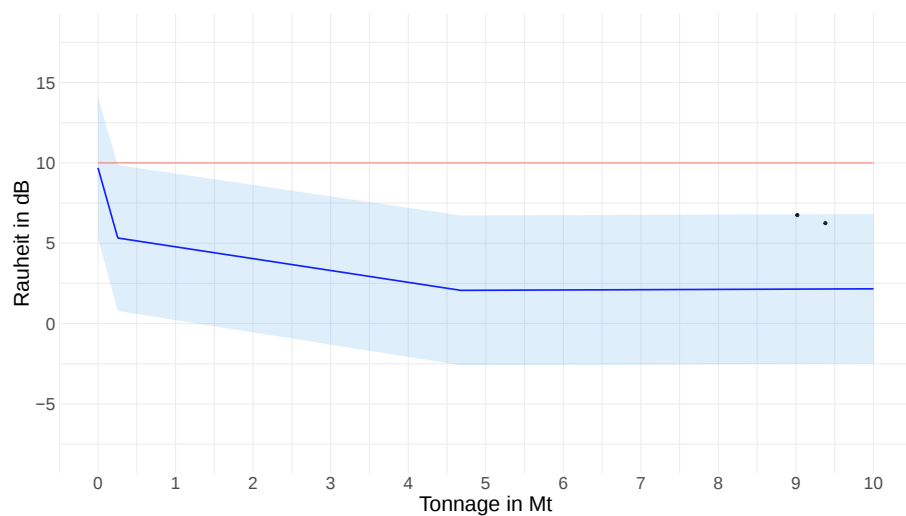


Abbildung A.18.: Prädiktionsband für Stahl R350, Profil UIC 54 und viel Güterzüge.

Literaturverzeichnis

- [1] L. Breiman, 2001, *Random Forests*, Machine Learning 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- [2] T. Hastie, R. Tibshirani, J. Friedman 2001, *The Elements of Statistical Learning*, Springer New York Inc., New York.
- [3] G. James, D. Witten, T. Hastie, R. Tibshirani, 2013, *An introduction to statistical learning – with applications in R*, Springer, New York.
- [4] S. Milborrow. Derived from mda:mars by T. Hastie and R. Tibshirani. Uses A. Miller’s Fortran utilities with T. Lumley’s leaps wrapper, 2021, *earth: Multivariate Adaptive Regression Splines*. R package version 5.3.1. <https://CRAN.R-project.org/package=earth>
- [5] 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*.
- [6] Europäisches Komitee für Normung, EN ISO 15610:2009: *Bahnanwendungen – Geräuschemission – Messung der Schienenrauheit im Hinblick auf die Entstehung von Rollgeräusch*.
- [7] 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