



Berner
Fachhochschule

Berner Fachhochschule
Technik und Informatik
Institut für Optimierung und Datenanalyse



Rauheit des SBB-Schienenennetzes

Verfeinerung der Modelle anhand neuer Messdaten

Schlussbericht nach der dritten Iteration

Im Auftrag der Schweizerischen Eidgenossenschaft

Biel, Dezember 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).

Autor

Dr. Niki Zumbrunnen

Begleitung BAFU / BAV

Dr. Sandro Ferrari, Franz Kuster

Dr. Robert Attinger, Christoph Dürig

Dank

Ich danke Emilie Freud, Dr. Jakob Oertli und Dr. Marc Schmid von den SBB für die fachliche Unterstützung sowie Pascal Bücheler für seine Mitarbeit in der zweiten Iteration und Dr. Jasmin Wandel für ihre wertvollen Inputs und Kommentare.

Hinweis

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

Titelbild

Messgerät und Fahrspiegel

Quelle: SBB

Inhaltsverzeichnis

Zusammenfassung	1
1. Einleitung	3
2. Beschreibung der vorhandenen Daten	5
2.1. Beschreibung der einzelnen Variablen	5
2.1.1. Rauheit	6
2.1.2. Tonnage	6
2.1.3. Stahlqualität	8
2.1.4. Angewandte Technik beim letzten Schleifen (Maschine)	9
2.1.5. Schienenprofil	9
2.1.6. Schwellentyp	10
2.1.7. Schienenbefestigungstyp	10
2.1.8. Schienenzwischenlage	11
2.1.9. Anteil Güterzüge	11
2.1.10. Streckengeometrie	12
2.1.11. Ort	12
2.1.12. Weitere Variablen	13
3. Statistische Modelle	15
3.1. Verwendete Modelle	15
3.1.1. Additive Modelle	15
3.1.2. Vorteile von Additiven Modellen	15
3.1.3. Multivariate adaptive Regressionssplines (MARS)	16
3.1.4. Kriterien für die Wahl des besten Modells	17
3.2. Modellierung der Daten	18
3.2.1. Ein Modell für die Stahlqualität R260	18
3.2.2. Modelle für die Stahlqualität R350	19
3.3. Vergleich mit dem Modell der zweiten Iteration	21
3.3.1. Evaluation des Modells der zweiten Iteration	21
3.3.2. Vergleich mit dem Modell der zweiten Iteration	21
4. Schätzung der Rauheit	23
4.1. Stahlqualität R260	23
4.2. Stahlqualität R350	25
5. Diskussion	29
A. Anhang	31
A.1. Beschreibung der einzelnen Variablen	31
A.2. Prädiktionsintervalle	38
A.3. Diagnostische Plots	45

Zusammenfassung

Anlass und Ziel dieser Studie

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 dem Grenzwert von 10 dB zu halten.

Das primäre Ziel dieser Studie ist die Vorhersage der Schienenrauheit auf dem Netz der SBB, basierend auf einzelnen Messungen, die von den SBB durchgeführt wurden. Ein sekundäres Ziel ist die Bestimmung von für die Schienenrauheit relevanten Einflussfaktoren. Damit könnten allenfalls Massnahmen definiert werden, um die Rauheit zu vermindern.

Vorgehen

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. Die aktuelle Projektphase hat im März 2021 begonnen und wird mit diesem Bericht abgeschlossen. In dieser Phase wurde in drei Iterationen die Datengrundlage durch zusätzliche Messungen erweitert und die Modelle für die Schätzung der Rauheit damit verfeinert.

Verwendete Modelle

Die einfachen Modelle aus der Vorstudie konnten mit den neuen Messungen deutlich verbessert werden. Die nun insgesamt 191 Beobachtungen ermöglichen eine stabile Schätzung von additiven Modellen mit sehr flexiblen Regressionskurven basierend auf kubischen Splines.

Die Flexibilität der Regressionskurven wird nur durch die Annahme der Additivität der Effekte der kategoriellen Erklärenden eingeschränkt. Es wird beispielsweise angenommen, dass sich für die beiden Stahlqualitäten die Verläufe der Schienenrauheit über die Tonnage nur um eine additive Konstante unterscheiden. Dies ist wahrscheinlich eine zu starke Vereinfachung. Ohne diese Annahme könnte das Modell aber nur mit deutlich mehr Beobachtungen einigermaßen stabil geschätzt werden.

Resultate

a) Faktoren, welche die Rauheit beeinflussen

Neben der Tonnage als wichtigstem Einflussfaktor wird die Schienenrauheit von der Stahlqualität, der verwendeten Schleifmaschine, dem Schienenprofil und dem Anteil Güterzüge beeinflusst.

b) Einhaltung Grenzwerte

Für die Schätzung der Schienenrauheit wurden die beiden Stahlqualitäten R260 und R350 separat betrachtet. Die Modelle liefern eine Regressionskurve für die Rauheit in Abhängigkeit von der Tonnage. Für jede Ausprägungskombination der anderen erklärenden Variablen wird die Kurve vertikal verschoben.

Für die Stahlqualität R260 liegt die geschätzte Rauheit für 12 der 16 Ausprägungskombinationen bereits bei Tonnage 0 unter dem kritischen Grenzwert von 10 dB. Für die anderen Kombinationen fällt die Rauheit spätestens bei 0.3 Mt unter den Grenzwert. Der Grenzwert wird von der geschätzten Rauheit für die meisten Kombinationen mit Stahlqualität R260 also nie überschritten und für alle höchstens während 0.6% der Zeit.

Für den Stahl R350 liegt die geschätzte Rauheit 3.1 dB höher als für den Stahl R260. Für alle Kombinationen mit Stahlqualität R350 fällt die geschätzte Rauheit spätestens nach 1 Mt unter den

Grenzwert. Die geschätzte Rauheit überschreitet den Grenzwert also für alle Kombinationen für weniger als 3.9% der Zeit.

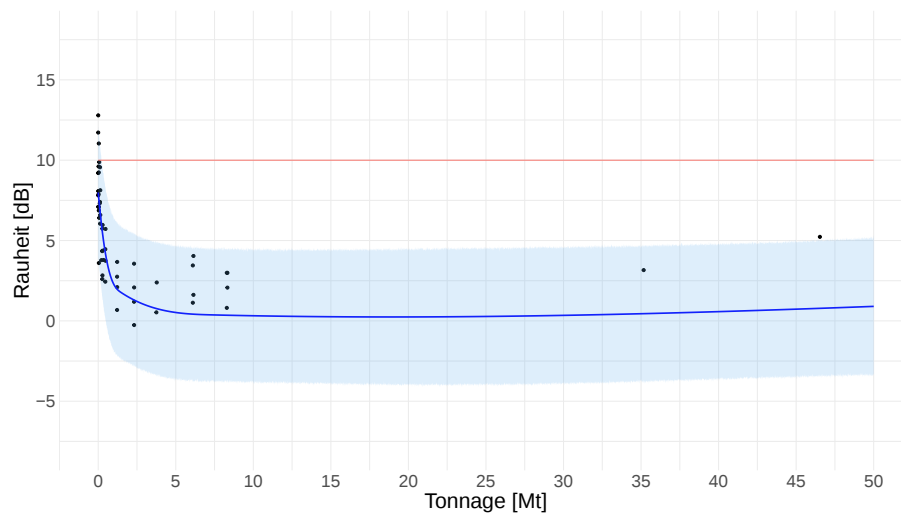


Abbildung 0.1.: 95%-Prädiktionsintervalle für Stahl R260, Maschine A, Profil UIC 54 und wenig Güterzüge.

Ausblick

Im Vergleich zur Vorstudie sind die aktuellen Modelle viel robuster und verlässlicher. Wenn man diese Modelle noch einmal deutlich verbessern möchte, müsste die Datengrundlage noch einmal deutlich erweitert werden. Eine Möglichkeit wären zusätzliche punktuelle Messungen mit Stahlqualität R350. Damit könnte man auf die sehr starke Annahme der Additivität des Effekts der Stahlqualität verzichten.

Sehr interessant wäre auch die Verwendung der Daten des ETH-Projekts [4]. Durch die Messung der Schienenrauheit mit einem fahrenden Zug wäre eine ganz andere Größenordnung der Anzahl Messungen möglich. Diese Datengrundlage würde eine deutlich differenziertere und genauere Schätzung der Schienenrauheit ermöglichen.

1. Einleitung

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.

Das primäre Ziel dieser Studie ist die Vorhersage der Schienenrauheit auf dem Netz der SBB, basierend auf einzelnen Messungen, die von den SBB durchgeführt wurden. Ein sekundäres Ziel ist die Bestimmung von für die Schienenrauheit relevanten Einflussfaktoren. Damit könnten allenfalls Massnahmen definiert werden, um die Rauheit zu vermindern.

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. Die aktuelle Projektphase hat im März 2021 begonnen und wird diesem Bericht abgeschlossen. In dieser Phase wurde in drei Iterationen die Datengrundlage durch zusätzliche Messungen erweitert und die Modelle für die Schätzung der Rauheit damit verfeinert. In jeder Iteration wurden zusätzliche Messungen durchgeführt und anschliessend die Modelle an die neuen Daten angepasst. Basierend auf den daraus gewonnenen Erkenntnissen wurden dann neue Messstandorte für die nächste Iteration definiert.

In der Vorstudie wurden erste Modelle gerechnet auf Basis der damals vorhanden Daten mit 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.

Jede einzelne Messung ist sehr aufwändig, weshalb die Anzahl Messungen limitiert 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. Dennoch scheinen die in dieser Vorstudie eruierten statistischen Modelle plausibel – auch aus Expertensicht.

Auf Basis der Modelle aus der Vorstudie wurden Standorte für zukünftige Messungen vorgeschlagen. Es wurden Parameterkombinationen identifiziert, mit hoher geschätzter Rauheit oder hoher Schätzunsicherheit. In der ersten Iteration wurden 39 Messungen durchgeführt. Für die Modellierung standen also 123 Beobachtungen zur Verfügung. Basierend auf den Modellen der ersten Iteration wurden wiederum Standorte für zukünftige Messungen vorgeschlagen. Nach den 53 Messungen der zweiten Iteration waren dann insgesamt 176 Messungen vorhanden, wovon 12 Beobachtungen ausgeschlossen wurden. Für die Modelle der zweiten Iteration wurden also 164 Messungen verwendet. Das Vorgehen wurde dann noch ein letztes Mal wiederholt mit 43 Messungen in der dritten Iteration. Die Resultate davon werden in diesem Bericht beschrieben.

Wie bereits erwähnt, wurden in der Vorstudie viele Variablen als potenzielle Einflussfaktoren berücksichtigt. Aufgrund der limitierten Datenmenge konnte damals nicht abschliessend beurteilt werden, welche davon wirklich einen Einfluss auf die Schienenrauheit haben. Deshalb musste man davon ausgehen, dass auch nach der letzten Iteration nicht genügend Daten vorhanden sein würden, um stabile Modelle mit allen relevanten Einflussfaktoren zu rechnen. Zumindest falls sich alle vorhandenen Variablen als relevant herausgestellt hätten. Aus diesem Grund wurde geplant, nach der letzten Iteration eine Simulationsstudie durchzuführen. Mit dieser Studie sollte untersucht werden, ob die statistischen Modelle genügend robust sind und ob trotz geringer Datenmengen solch pragmatische Modelle ausreichen können, die Schienenrauheit mit einer gewissen Unsicherheit genügend genau zu schätzen. Hierzu hätte man auch die im Modell bisher nicht berücksichtigten Einflussgrössen variiert. Mit den neuen Beobachtungen wurde dann aber klar, dass die meisten Variablen keinen Einfluss auf die Rauheit haben oder zumindest von den Modellen nicht für die Erklärung der Rauheit verwendet wurden. Die Fragestellung der Simulationsstudie ist deshalb nicht mehr so aktuell. Deshalb haben wir uns bei den Simulationen auf die Quantifizierung der Schätzunsicherheit konzentriert, insbesondere auf die Berechnung von Prädiktionsintervallen.

Die punktuellen Messungen der Schienenrauheit, wie sie von den SBB durchgeführt werden, sind sehr aufwendig. Eine Abdeckung des ganzen Schienennetzes ist damit nicht realistisch. 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 am Institut für Werkzeugmaschinen und Fertigung der ETH Zürich gearbeitet in einem Projekt zur Erfassung der Schienenrauheit mit optischen Methoden (vgl. [4]). Ursprünglich war angedacht, auch die Daten dieses ETH-Projekts für die vorliegende Studie zu verwenden. Diese stehen jedoch bis dato noch nicht zur Verfügung und wurden deshalb nicht verwendet. Das entsprechende optionale Arbeitspaket wurde nicht bearbeitet. Zu einem späteren Zeitpunkt wäre es interessant, die Modelle mit diesen Daten zu verfeinern oder Modelle zu rechnen, die nur auf diesen Daten beruhen.

Als Nebenprodukt dieser Studie war geplant, dass die SBB eine Schweizerkarte erstellen mit der geschätzten Rauheit auf dem gesamten SBB-Schienennetz. 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. Deshalb kann eine solche Karte momentan nicht erstellt werden. Der Code für die Schätzung der Rauheit wird den 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 dritten Iteration wurden 43 Messungen gemacht, davon 17 Messungen in Härkingen und 8 Messungen an vier Standorten in Alznach.

Neben den neu erhobenen Daten verwenden wir auch die bis zur zweiten Iteration von den SBB durchgeführten Messungen. Diese bestehen aus 162 Messpunkten und wurden bisher als „SBB“-Datensatz bezeichnet.

Die in den bisherigen Iterationen ebenfalls verwendeten Datensätze „BAV“ und „MBBM“ verwenden wir nicht mehr. Diese bestehen aus Rauheitsmessungen an den Monitoringstationen des BAV bzw. aus einer Studie mit dem Ziel, den Einfluss von Schienendämpfern auf die Rauheit zu ermitteln. Mit dem Verzicht auf die insgesamt 14 Messungen können wir sicherstellen, dass für alle verwendeten Beobachtungen die genau gleiche Methode für die Messung und die Berechnung des Einzahlwerts für die Schienenrauheit verwendet wurde. In den bisherigen Iterationen haben wir diese Messungen verwendet im Wissen, dass dies zu einer gewissen Inhomogenität führt. Weil uns aber insgesamt relativ wenige Beobachtungen zur Verfügung standen, hatten wir das in Kauf genommen. Mit den nun rund 200 Beobachtungen können wir in der dritten Iteration auf diese beiden Datensätze verzichten.

Zu den in den ersten beiden Iterationen ausgeschlossenen 12 Beobachtungen des „SBB“-Datensatzes kommen zwei Ausschlüsse hinzu. Wir schliessen also insgesamt 14 Beobachtungen aus.

Ausgeschlossen werden die Beobachtungen der Linie 460: Grauenstein, Trubschachen, Hasle, Gigen, Wiggen. 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 drei weitere Beobachtungen entfernt, eine davon ist eine neue Messung.

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

Eine neue Messung wurde in Flamatt an einer Stelle gemacht, an der zuvor eine Bremsung mit einer Magnetbremse stattgefunden hat (vgl. auch Abschnitt 2.1.3). Durch diese Bremsung wurde die Schienenrauheit massiv erhöht und der gemessene Wert ist unabhängig von der vorherigen Rauheit, das heisst insbesondere auch von der Tonnage zum Messzeitpunkt. Deshalb bringt diese Beobachtung keinen Mehrwert für das Modell und wird nicht für die Modellierung verwendet.

Es wurden also 191 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 zweiten Iteration verwendeten und die seither erhobenen Messungen separat dargestellt.

Um den Lesefluss nicht zu stören, sind 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 Median der Rauheit für neuen Messungen deutlich tiefer ist, als für die bisherigen. Vereinzelt wurden in der dritten Iteration aber auch sehr hohe Rauheitswerte gemessen. So liegt der maximale gemessene Wert nun bei 17.2 dB.

Tabelle 2.1.: Kennzahlen für die Rauheit

Rauheit [dB]	alle, N = 191	alt, N = 150	neu, N = 41
Minimum	-1.9	-1.9	-1.7
1. Quartil	2.4	2.6	1.2
Median	4.8	5.4	3.5
Mittelwert	5.3	5.5	4.6
3. Quartil	7.9	8.1	7.6
Maximum	17.2	15.7	17.2

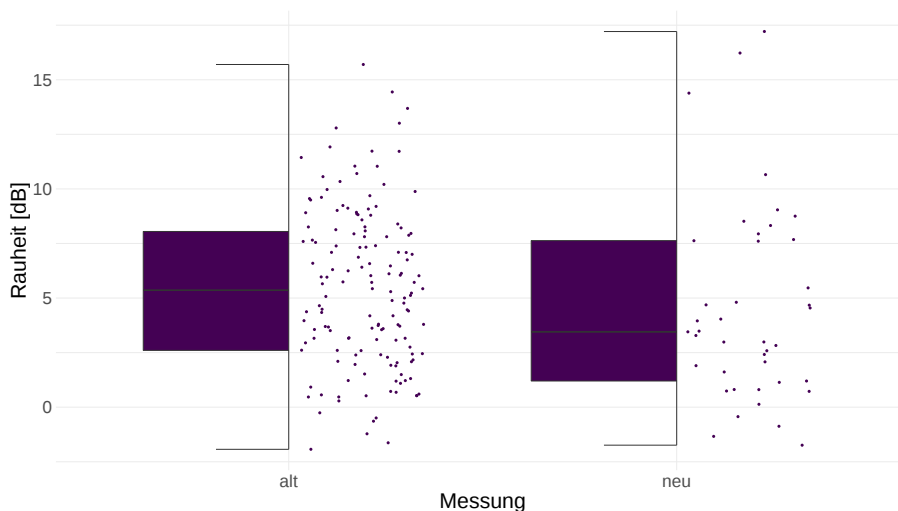


Abbildung 2.1.: Boxplot der Rauheit pro Messung.

2.1.2. Tonnage

Die Variable Tonnage beschreibt die Summe der Radsatzlasten seit dem letzten Schleifen in Tonnen [t] bzw. Megatonnen [Mt]. Die meisten Messungen wurden an Stellen mit kleinen Tonnagen gemacht, weil dann die erwartete Rauheit am höchsten ist. Dadurch ist in diesem Bereich die Gefahr am grössten, dass der Grenzwert für die Rauheit nicht eingehalten wird. Dementsprechend wichtig ist es, die Rauheit für diesen Bereich möglichst gut schätzen zu können.

Die Verteilung der numerischen Variablen Tonnage ist aus oben genannten Gründen schief mit 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 bisherigen Messungen weisen tiefe Tonnagen aus, es wird aber der gesamte relevante Tonnagebereich abgedeckt. Auch in der dritten Iteration wurden vor allem Messungen an

Stellen mit tiefen Tonnagen gemacht. Einige Messungen wurden aber auch an Stellen mit mittleren oder hohen Tonnagewerten durchgeführt. Beispielsweise wurde die Rauheit in Härkingen an zwei Stellen mit Tonnagewerten von knapp 50 Mt gemessen.

Tabelle 2.2.: Kennzahlen für die Tonnage

Tonnage	alle, N = 191	alt, N = 150	neu, N = 41
Minimum	0	0	0
1. Quartil	189'712	129'347	660'534
Median	1'592'093	1'379'506	3'097'807
Mittelwert	6'867'931	6'642'334	7'693'287
3. Quartil	7'638'850	7'525'199	8'292'014
Maximum	64'541'062	64'541'062	49'802'880

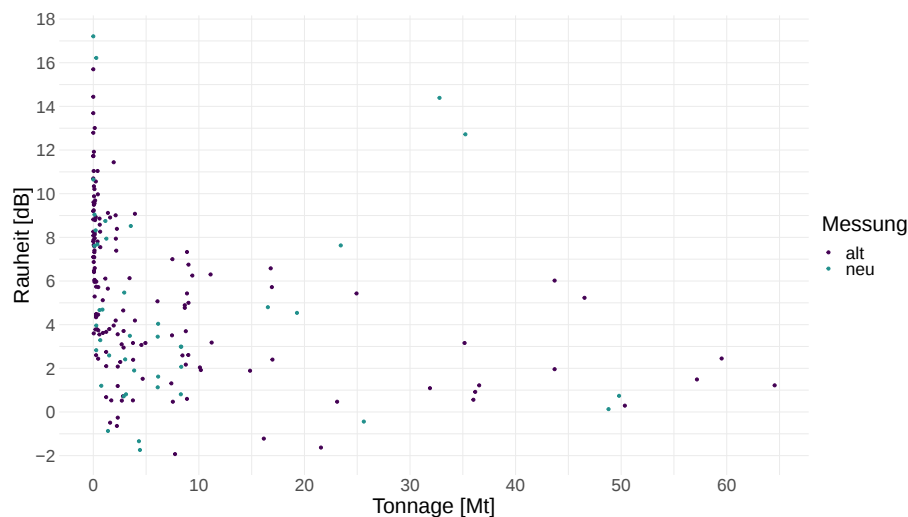


Abbildung 2.2.: Streudiagramm der Tonnage und Rauheit der Messungen. Die Beobachtung mit einer Tonnage von rund 35 Mt und einer Rauheit von 12.7 dB wurde für die Modellierung nicht verwendet, weil die Rauheit durch eine Bremsung mit einer Magnetbremse verfälscht wurde.

Abbildung 2.2 zeigt ein Streudiagramm mit der Tonnage und der Rauheit. Auf der Grafik fallen zwei Ausreisser auf mit Tonnagen im Bereich von 35 Mt und gemessenen Rauheiten von über 12 dB. Obwohl sie bezüglich Rauheit und Tonnage sehr ähnlich aussehen, weisen die beiden Ausreisser aus sehr unterschiedlichen Gründen so hohe Rauheitswerte auf.

Der erste Messwert stammt aus Vufflens. Die Schiene wies dort zum Messzeitpunkt eine Riffelung auf. Die Riffeln im Abstand von ca. 4 cm sind optisch gut erkennbar und auch auf dem Spektrum der gemessenen Rauheit sieht man einen Peak bei ca. 4 cm.

Der zweite Messwert stammt aus Flamatt. Die Ursache für die hohe Rauheit ist dort eine Bremsung mit einer Magnetbremse. Auch dies ist optisch gut erkennbar und das Spektrum weist gleichmässig über alle Frequenzen eine sehr hohe Rauheit aus.

Die Fotografien der gemessenen Schienen und die Rauheitsspektren werden hier nicht gezeigt.

Die Messung aus Flamatt wurde für die Modellierung nicht verwendet, weil durch diese Bremsung die Schienenrauheit massiv erhöht wurde und der gemessene Wert unabhängig von der vorherigen Rauheit ist, das heisst insbesondere auch von der Tonnage zum Messzeitpunkt.

2.1.3. Stahlqualität

Die Variable Stahl beschreibt die Stahlqualität. Sie ist kategoriell mit den 2 Ausprägungen R260 und R350. 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.

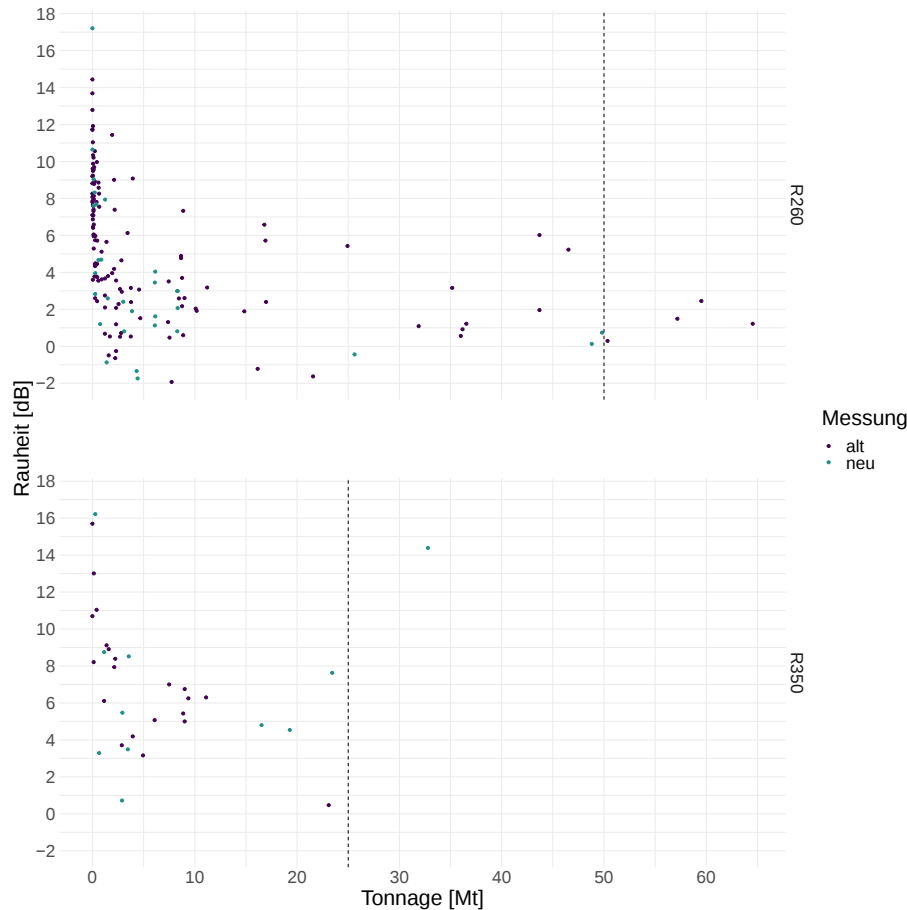


Abbildung 2.3.: Streudiagramm der Tonnage und Rauheit der Messungen, aufgeteilt nach Stahlqualität.

Tabelle 2.3.: Häufigkeitstabelle für die Stahlqualität

Stahlqualität	alle, N = 191	alt, N = 150	neu, N = 41
R260	159 (83%)	129 (86%)	30 (73%)
R350	32 (17%)	21 (14%)	11 (27%)

In den Daten kommt zusätzlich R320Cr als dritter Typ vor. Er liegt bezüglich Härte zwischen den anderen beiden Typen und kommt auf dem SBB-Schiennetz 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 der Stahlqualität ist in Tabelle 2.3 und in Abbildung A.1 aufgetragen. In Abbildung A.2 ist der Zusammenhang zwischen der Rauheit und der Stahlqualität in Boxplots dargestellt.

Aus fachlicher Sicht ist es sinnvoll, den Rauheitsverlauf für die beiden Stahlqualitäten separat zu betrachten. Der Stahl R350 weist eine deutlich höhere Rauheit auf als der Stahl R260. Zudem unterscheiden sich die zu betrachtenden Tonnagebereiche. Schienen der weichen Stahlqualität R260 werden normalerweise nach einer Belastung von rund 50 Mt geschliffen. Schienen mit der in Kurven verbauten härteren Stahlqualität R350 werden bereits nach rund 25 Mt geschliffen. Deshalb ist es für

den Stahl R350 nur bedingt sinnvoll, die Rauheit für Tonnagen zu schätzen, die deutlich grösser als 25 Mt sind.

Abbildung 2.3 zeigt für beide Stahlqualitäten je ein Streudiagramm der Tonnage und Rauheit der Messungen. Die vertikalen gestrichelten Linien zeigen an, wann üblicherweise geschliffen wird.

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.

Maschine ist kategoriell mit 4 Ausprägungen. Die Angabe zur Maschine fehlt für 6 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. Diese Beobachtungen wurden für die Modellierung nicht verwendet und sind hier auch nicht aufgeführt. Auch die Messungen an Stellen, die mit einer Fräsmaschine bearbeitet wurden, verwenden wir nicht für die Modelle und führen sie hier auch nicht auf. Dies betrifft zwei alte und eine neue Messung.

Die Verteilung der Variable Maschine ist in der Häufigkeitstabelle in Tabelle 2.4 und im Balkendiagramm in Abbildung A.3 im Anhang ersichtlich. Der Maschinentyp C kommt bei den neuen Beobachtungen nicht mehr vor. In Abbildung A.4 ist der Zusammenhang zwischen der Rauheit und der Maschine in Boxplots dargestellt.

Tabelle 2.4.: Häufigkeitstabelle für die Schleiftechnik.

Maschine	alle, N = 191	alt, N = 150	neu, N = 41
A	117 (61%)	82 (55%)	35 (85%)
B	26 (14%)	25 (17%)	1 (2.4%)
C	12 (6.3%)	12 (8.0%)	0 (0%)
D	36 (19%)	31 (21%)	5 (12%)

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. Rund zwei Drittel aller Messungen wurden an Stellen mit dem Profil UIC 60 gemacht. Die Verteilung ist in Tabelle 2.5 und in Abbildung A.5 dargestellt. In Abbildung 2.4 ist der Zusammenhang zwischen der Rauheit und dem Schienenprofil in Boxplots dargestellt.

Tabelle 2.5.: Häufigkeitstabelle für das Schienenprofil

Schienenprofil	alle, N = 191	alt, N = 150	neu, N = 41
UIC 54	70 (37%)	61 (41%)	9 (22%)
UIC 60	121 (63%)	89 (59%)	32 (78%)

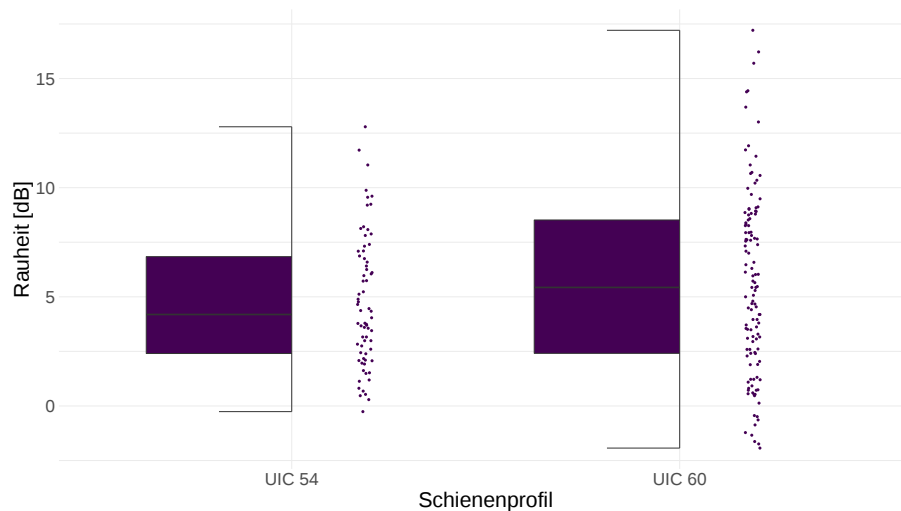


Abbildung 2.4.: Boxplots der Rauheit pro Schienenprofil.

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. Die Verteilung ist in Tabelle 2.6 und in Abbildung A.6 aufgetragen. Abbildung A.7 zeigt den Zusammenhang zwischen der Rauheit und dem Schwellentyp in Boxplots.

Tabelle 2.6.: Häufigkeitstabelle für den Schwellentyp

Schwellentyp	alle, N = 191	alt, N = 150	neu, N = 41
B70	19 (9.9%)	11 (7.3%)	8 (20%)
B91	146 (76%)	116 (77%)	30 (73%)
Holz	11 (5.8%)	8 (5.3%)	3 (7.3%)
Stahl	15 (7.9%)	15 (10%)	0 (0%)

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 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.8 aufgetragen. Viele Ausprägungen haben jedoch nur sehr wenige Beobachtungen. Deshalb wurde die Schienenbefestigung für die Modelle nicht verwendet.

In Abbildung A.9 ist der Zusammenhang zwischen der Rauheit und der Schienenbefestigung in Boxplots dargestellt.

Tabelle 2.7.: Häufigkeitstabelle für die Schienenbefestigung

Schienenbefestigung	alle, N = 191	alt, N = 150	neu, N = 41
B70_W_Skl1	6 (3.1%)	6 (4.0%)	0 (0%)
B70_W14	9 (4.7%)	1 (0.7%)	8 (20%)
B70_Ws_Skl1	4 (2.1%)	4 (2.7%)	0 (0%)
B91_W_Skl1	6 (3.1%)	4 (2.7%)	2 (4.9%)
B91_W14	55 (29%)	52 (35%)	3 (7.3%)
B91_Ws_Skl1	85 (45%)	60 (40%)	25 (61%)
Holz_K	3 (1.6%)	2 (1.3%)	1 (2.4%)
Holz_Ke_Skl12	1 (0.5%)	1 (0.7%)	0 (0%)
Holz_Ke_Skl3	7 (3.7%)	5 (3.3%)	2 (4.9%)
Stahl_Aek	15 (7.9%)	15 (10%)	0 (0%)

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 unterschieden. Die Zwischenlagenhärte ist folgendermassen definiert:

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

Zwischenlage ist kategoriell mit 3 verschiedenen Ausprägungen. Ihre Verteilung ist in Tabelle 2.8 und in Abbildung A.10 aufgetragen. In Abbildung A.11 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. Das heisst 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.

Tabelle 2.8.: Häufigkeitstabelle für die Schienenzwischenlage

Zwischenlage	alle, N = 191	alt, N = 150	neu, N = 41
hart	148 (77%)	108 (72%)	40 (98%)
keine	3 (1.6%)	2 (1.3%)	1 (2.4%)
weich	40 (21%)	40 (27%)	0 (0%)

2.1.9. Anteil Güterzüge

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 Expertinnen 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.12 aufgetragen. Rund zwei drittel der Messungen der dritten Iteration wurden an Standorten mit vielen Güterzügen gemacht.

Die Verteilung ist somit ein wenig ausgeglichener als nach der zweiten Iteration. In Abbildung A.13 ist der Zusammenhang zwischen der Rauheit und dem Anteil Güterzüge in Boxplots dargestellt.

Tabelle 2.9.: Häufigkeitstabelle für den Anteil Güterzüge

Anteil Güterzüge	alle, N = 191	alt, N = 150	neu, N = 41
viel	63 (33%)	36 (24%)	27 (66%)
wenig	128 (67%)	114 (76%)	14 (34%)

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.14 aufgetragen. In Abbildung A.15 ist der Zusammenhang zwischen der Rauheit und der Variable Kurve in Boxplots dargestellt.

Tabelle 2.10.: Häufigkeitstabelle für die Streckengeometrie

Kurve	alle, N = 191	alt, N = 150	neu, N = 41
ja	26 (14%)	15 (10%)	11 (27%)
nein	165 (86%)	135 (90%)	30 (73%)

2.1.11. Ort

Der Datensatz enthält den Ort, an dem die jeweilige Messung durchgeführt wurde. Uns interessiert nicht primär, wie sich die konkreten Orte bezüglich Schienenrauheit unterscheiden, weil wir die Messstellen als zufällige Stichprobe aus allen möglichen Messstellen auf dem Netz der SBB betrachten. Die Tatsache, dass an manchen Orten mehrere Messungen durchgeführt wurden, führt jedoch dazu, dass die Daten nicht statistisch unabhängig sind. Man kann davon ausgehen, dass sich beispielsweise zwei Messungen, die beide in Härkingen gemacht wurden, zueinander nicht gleich verhalten, wie eine Messung aus Härkingen zu einer Messung aus Aarau. Um diese Abhängigkeiten korrekt zu modellieren, nehmen wir den Ort als zufälligen Effekt ins Modell.

Tabelle 2.11.: Häufigkeitstabelle für Ort.

Anzahl Messungen pro Ort	Anzahl Orte
1	17
2	17
3	6
4	3
6	2
11	1
12	1
14	1
17	1
21	1
23	1

Die Verteilung der Anzahl Messungen pro Ort ist in Tabelle 2.11 abgebildet. Für rund 32% der Orte liegt genau eine Messung vor. Für die restlichen 68% wurden mehrere Messungen durchgeführt. Am häufigsten wurde in Alznach-Nord gemessen. Dort wurden 23 Messungen durchgeführt.

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.

3. Statistische Modelle

In Abschnitt 3.1 werden die Modelle kurz erklärt, welche dann in Abschnitt 3.2 verwendet werden, um die Daten zu modellieren. In Abschnitt 3.3 wird noch einmal das Modell der zweiten Iteration betrachtet und mit dem Modell verglichen, das in der dritten Iteration für die Vorhersage verwendet wird.

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 den ersten beiden Iterationen der zweiten Projektphase haben wir Regressionssplines (MARS) verwendet. Damit konnten wir den gesamten Wertebereich der Tonnage abdecken und die Nichtlinearität mit stückweise linearen Funktionen modellieren.

Es ist aber anzunehmen, dass der wahre Zusammenhang zwischen Rauheit und Tonnage auch nicht stückweise linear ist und dass die Annahme der stückweisen Linearität eine zu starke Vereinfachung ist. Die noch einmal erweiterte Datengrundlage in der dritten und letzten Iteration erlaubt es uns nun, noch flexiblere Modelle zu verwenden. Die additiven Modelle erlauben es uns, den Zusammenhang zwischen Rauheit und Tonnage mit kubischen Splines zu modellieren. Das Resultat ist eine glatte Regressionskurve, die eher dem vermuteten wahren Zusammenhang entspricht.

3.1.1. Additive Modelle

Additive Modelle sind eine Verallgemeinerung von klassischen linearen Modellen. Neben den parametrischen Termen enthalten sie eine Summe von glatten Funktionen der Prädiktorvariablen. In unserem Fall hat das Modell die Form

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} + f_1(x_{p+1i}) + f_2(x_{p+2i}) + \epsilon_i$$

wobei y_i die Zielvariable, $x_{1i}, \dots, x_{p+2i}$ die erklärenden Variablen und ϵ_i standardnormalverteilte Zufallsvariablen bezeichnen.

Mit der glatten Funktion f_1 wird die Abhängigkeit der Rauheit von der Tonnage modelliert. Dabei wird f_1 durch kubische Splines repräsentiert.

Man kann zeigen, dass konventionelle zufällige Effekte in einem additiven Modell auf die gleiche Art repräsentiert werden können, wie glatte Funktionen. Das verwenden wir, um mit f_2 einen zufälligen Effekt für den Ort in das Modell zu nehmen.

Eine ausführliche Beschreibung von (verallgemeinerten) additiven Modellen und dem dafür verwendeten R-Paket `mgcv` findet sich in [9].

3.1.2. Vorteile von Additiven Modellen

Additive Modelle haben unter anderem die folgenden Vorteile gegenüber anderen Modellen, weshalb wir in der dritten und letzten Iteration nur noch additive Modelle für die Vorhersage verwenden:

Flexible Modellierung durch glatte Funktionen

Den Zusammenhang zwischen der Rauheit und der Tonnage können wir mit einer glatten Funktion modellieren, die durch kubische Splines repräsentiert wird. Dadurch ist das Modell sehr flexibel. Insbesondere ist es nicht auf stückweise lineare Funktionen beschränkt.

Durch diese Flexibilität entsteht jedoch die Gefahr einer Überanpassung. Mittlerweile stehen uns aber recht viele Beobachtungen zur Verfügung, so dass eine stabile Schätzung des Modells möglich ist.

Additive Struktur

Die additive Struktur ist eine sehr starke Einschränkung für das Modell. Einerseits muss deshalb bei jeder Modellierung hinterfragt werden, ob diese Struktur für die gegebenen Daten plausibel ist. Andererseits können die Modelle durch diese Einschränkung besser und stabiler geschätzt werden.

Zufällige Effekte

In einem additiven Modell können zufällige Effekte auf die gleiche Art repräsentiert werden, wie glatte Funktionen. Mit einem zufälligen Effekt für den Ort können wir Abhängigkeiten zwischen den Messungen, die am gleichen Ort gemacht wurden, berücksichtigen.

Interpretierbarkeit

Abgesehen von den glatten Funktionen können die Effekte in einem additiven Modell gleich interpretiert werden, wie in einem klassischen linearen Modell. Es stehen auch Tests zur Verfügung, um zu prüfen, ob eine erklärende Variable einen signifikanten Effekt auf die Zielvariable hat. Dies ist ein grosser Vorteil zum Beispiel gegenüber Random Forests.

3.1.3. Multivariate adaptive Regressionssplines (MARS)

In der zweiten Iteration wurden für die Schätzung der Rauheit Multivariate adaptive Regressionssplines (MARS) verwendet. Aus den oben genannten Gründen wurden MARS in der dritten Iteration nicht mehr für die Schätzung verwendet. In Abschnitt 3.3 wird jedoch das MARS-Modell der zweiten Iteration mit Daten der dritten Iteration evaluiert und mit dem additiven Modell verglichen. Deshalb wird hier noch einmal kurz erklärt, wie MARS funktionieren:

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-Paket können in [5] nachgeschlagen werden.

3.1.4. Kriterien für die Wahl des besten Modells

Aus den in Abschnitt 3.1.2 genannten Gründen beschränken wir uns in dieser letzten Iteration auf additive Modelle. Wir müssen uns also nicht mehr zwischen unterschiedlichen Modellfamilien entscheiden, wie wir das in der zweiten Iteration zwischen Random Forest und MARS machen mussten.

Die Wahl reduziert sich so einerseits auf die Repräsentation der glatten Funktionen und andererseits auf die verwendeten erklärenden Variablen.

Für die Wahl der Repräsentation der glatten Funktion für die Tonnage verwenden wir restricted maximum likelihood (REML) als Kriterium, welches auch für die Anpassung des Modells inklusive automatischer Wahl der Glattheit verwendet wird.

Dabei schneiden kubische Splines mit $k = 5$ Knoten am besten ab. Die Zahl der Knoten ist relativ klein, was die Flexibilität des Modells beschränkt. Grössere Werte für k führen aber zu einer Überanpassung, was sich in einer wellenförmigen Regressionskurve zeigt.

REML kann nicht verwendet werden, um Modelle mit unterschiedlichen festen Effekten zu vergleichen. Deshalb verwenden wir für die Variablenselektion eine „backward selection“ basierend auf P-Werten. Dabei beginnt man mit dem vollen Modell, also mit dem Modell das alle Erklärenden enthält. Man definiert einen Schwellenwert α für die P-Werte. So lange mindestens eine Variable einen P-Wert grösser als α ausweist, wirft man Schritt für Schritt jeweils die Variable mit dem grössten P-Wert aus dem Modell. Dieses Selektionsverfahren resultiert in einem Modell, das nur noch Erklärende enthält, die einen signifikanten Einfluss auf die Zielvariable haben.

Vorhersagefehler

In der zweiten Iteration hatten wir den Vorhersagefehler als Kriterium für die Modellwahl verwendet. Dies weil wir Modelle aus verschiedenen Familien betrachtet hatten, die nicht direkt vergleichbar waren und weil unser primäres Ziel ist, die Rauheit möglichst gut vorhersagen zu können.

In Abschnitt 3.3 verwenden wir den Vorhersagefehler, genauer den RMSE, für den Vergleich der aktuellen Modelle mit denjenigen der zweiten Iteration. Deshalb gehen wir hier noch einmal auf seine Berechnung ein.

Für die Berechnung des mittleren quadrierten Vorhersagefehlers (englisch mean squared error)

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

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 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.

Auch die Stabilität eines Modells ist ein Gütekriterium. 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.

3.2. Modellierung der Daten

Wie bereits in Abschnitt 2.1.3 erwähnt, ist es aus fachlicher Sicht sinnvoll, den Rauheitsverlauf für die beiden Stahlqualitäten separat zu modellieren. Der Stahl R350 weist eine deutlich höhere Rauheit auf als der Stahl R260. Zudem unterscheiden sich die zu betrachtenden Tonnagebereiche. Schienen der weichen Stahlqualität R260 werden normalerweise nach einer Belastung von rund 50 Mt geschliffen. Schienen mit der in Kurven verbauten härteren Stahlqualität R350 werden bereits nach rund 25 Mt geschliffen. Deshalb ist es für den Stahl R350 nur bedingt sinnvoll, die Rauheit für Tonnagen zu schätzen, die deutlich grösser als 25 Mt sind.

In einem additiven Modell kann man beide Ausprägungen der Stahlqualität verwenden und sie als additiven Effekt ins Modell nehmen. Dies impliziert jedoch die Annahme, dass sie die Rauheit bis auf eine additive Konstante in beiden Stahltypen über die Zeit (d.h. Tonnage) gleich verhält. Damit wir diese ziemlich starke Annahme nicht treffen müssen, versuchen wir zuerst, die beiden Stahlqualitäten separat zu modellieren.

Für alle folgenden Modelle haben wir eine glatte Funktion der Tonnage geschätzt und einen zufälligen Effekt für den Ort verwendet. Zudem haben wir die Maschine, das Schienenprofil, den Schwellentyp, den Anteil Güterzüge und die Kurve (ja/nein) als feste Effekte in Betracht gezogen. Aus diesen Variablen haben wir dann mit „backward selection“ eine Auswahl getroffen.

Im Modell mit beiden Stahlqualitäten haben wir zusätzlich die Stahlqualität als festen Effekt integriert.

3.2.1. Ein Modell für die Stahlqualität R260

Die „backward selection“ wählt als feste Effekte die Maschine, das Profil und den Anteil Güterzüge. Die dazugehörigen Koeffizienten sind in Tabelle 3.1 aufgeführt. Abbildung 3.1 zeigt die geschätzte glatte Funktion der Tonnage. Die Residuen sind als Punkte aufgezeichnet und am unteren Rand ist jede Beobachtung als Strich eingezeichnet. Zudem sind punktweise 95%-Konfidenzintervalle für die Funktion gegeben.

Die Rauheit sinkt am Anfang sehr schnell und bleibt dann fast konstant. Die geschätzte Rauheit nimmt zwar wieder ein wenig zu. Wegen der wenigen Beobachtungen mit sehr grosser Tonnage, können wir darüber aber keine gesicherte Aussage machen.

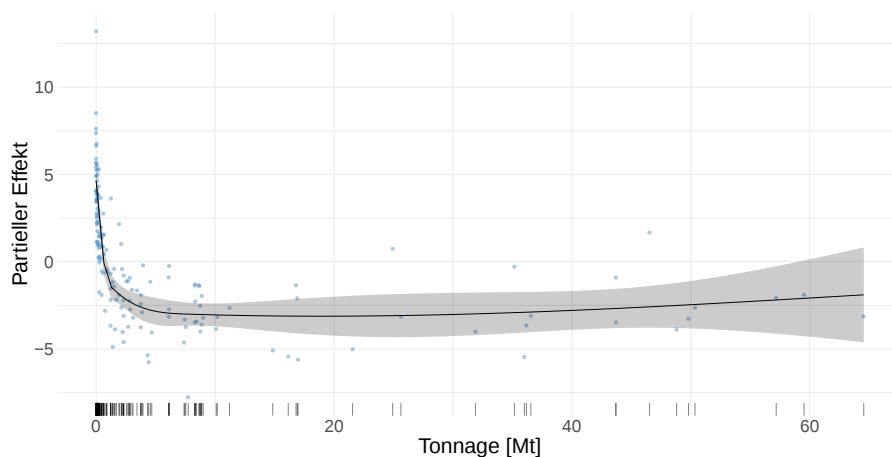


Abbildung 3.1.: Geschätzte glatte Funktion im Modell für die Stahlqualität R260.

Im Anhang in Abbildung A.30 sind die diagnostischen Plots für dieses Modell aufgezeichnet. Auf allen Plots ist ein Ausreisser deutlich erkennbar. Es handelt sich dabei um eine Messung in Härkingen direkt nach dem Schleifen mit einer Rauheit von 17.2. Glücklicherweise hat diese Beobachtung keinen grossen Einfluss auf die Schätzung der Parameter und der Rauheit. Abgesehen von diesem Ausreisser sehen die Plots akzeptabel aus.

Tabelle 3.1.: Koeffizienten der festen Effekte im Modell für die Stahlqualität R260.

	Schätzwert	Standardfehler
(Intercept)	0.421	1.001
Maschine: B	1.883	0.657
Maschine: C	0.946	1.693
Maschine: D	1.992	0.573
Profil: UIC 60	2.510	0.830
Anteil GZ: wenig	2.943	0.808

Abbildung A.31 im Anhang zeigt einen QQ-Plot der geschätzten zufälligen Effekte für die verschiedenen Orte, mit dem man die Normalverteilungsannahme überprüfen kann. Für den Ort Sempach 2 wird ein sehr grosser Effekt geschätzt. Dieser sticht auf dem QQ-Plot heraus. Abgesehen von diesem Ausreisser sieht der Plot akzeptabel aus.

Die P-Werte für die erklärenden Variablen geben wir hier nicht explizit an. Sie sind alle sehr klein (< 0.003).

3.2.2. Modelle für die Stahlqualität R350

Von der Stahlqualität R350 liegen uns nur 32 Messungen vor. Dies liegt einerseits daran, dass der Fokus nicht auf die Schienenrauheit in den Kurven gelegt wurde und andererseits an der Tatsache, dass die Messungen in Kurven viel aufwendiger sind auf Grund von zusätzlichen Sicherheitsmassnahmen. Die kleine Anzahl Messungen, erschwert eine sinnvolle Schätzung der glatten Funktion für die Tonnage. Trotzdem haben wir versucht, ein entsprechendes Modell zu rechnen. Als Alternative dazu haben wir beide Stahlqualitäten kombiniert und ein Modell gerechnet, dass von einem additiven Unterschied zwischen den beiden ausgeht.

3.2.2.1. Ein Modell nur mit R350

Das beste Modell für diese Daten enthält keine festen Effekte. Die Rauheit wird also nur abhängig von der Tonnage und vom Ort modelliert.

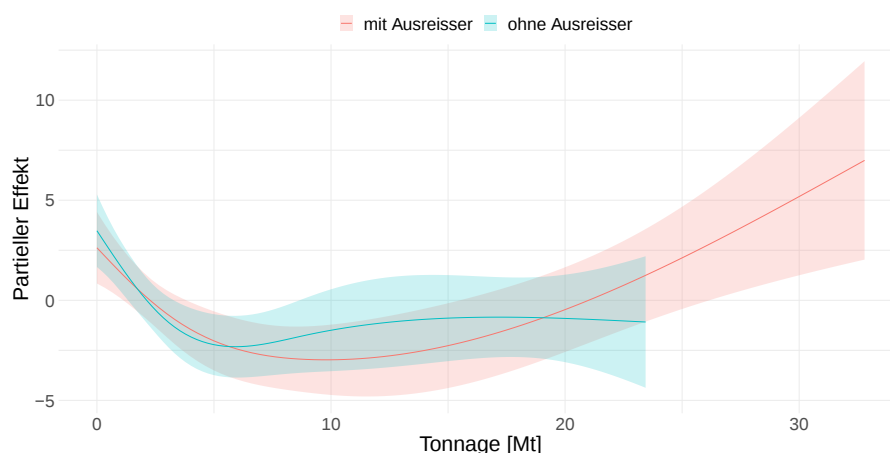


Abbildung 3.2.: Geschätzte glatte Funktionen im Modell für die Stahlqualität R350 mit und ohne Ausreisser.

Eine Messung in Vufflens mit einer Tonnage von rund 35 Mt und einer gemessenen Rauheit von 12.7 ist ein Ausreisser, der einen sehr grossen Einfluss auf das Modell hat. Dies wird in Abbildung 3.2 illustriert.

Dort sind die geschätzten Funktionen für die Tonnage mit punkweisen 95%-Konfidenzintervallen aufgezeichnet. Einmal gerechnet mit der Messung in Vufflens und einmal ohne. Auch wenn man den Ausreisser entfernen würde, wäre das Resultat wenig plausibel, wie die wellenförmige Funktion zeigt. Alle diese Punkte sprechen dafür, dieses Modell nicht zu verwenden. Die diagnostischen Plots in Abbildung A.32 bestätigen dieses Bild.

3.2.2.2. Ein Modell mit beiden Stahlqualitäten

Damit wir die Rauheit für die Stahlqualität R350 schätzen können, müssen wir auf die Additivität zurückkommen. Wenn wir akzeptieren, dass sich die Rauheit in beiden Stahltypen bis auf eine additive Konstante gleich verhält, können wir ein Modell mit beiden Stahlqualitäten rechnen. Das resultierende Modell ist sehr ähnlich wie das Modell für R260. Die wenigen Beobachtungen mit dem harten Stahl haben also keinen grossen Einfluss auf das Modell. Die Koeffizienten der festen Effekte sind in Tabelle 3.2 aufgeführt. Abbildung 3.3 zeigt die geschätzten glatten Funktionen der Tonnage mit punkweisen Konfidenzintervallen. Einmal für das Modell nur mit R260 und einmal für das kombinierte Modell.

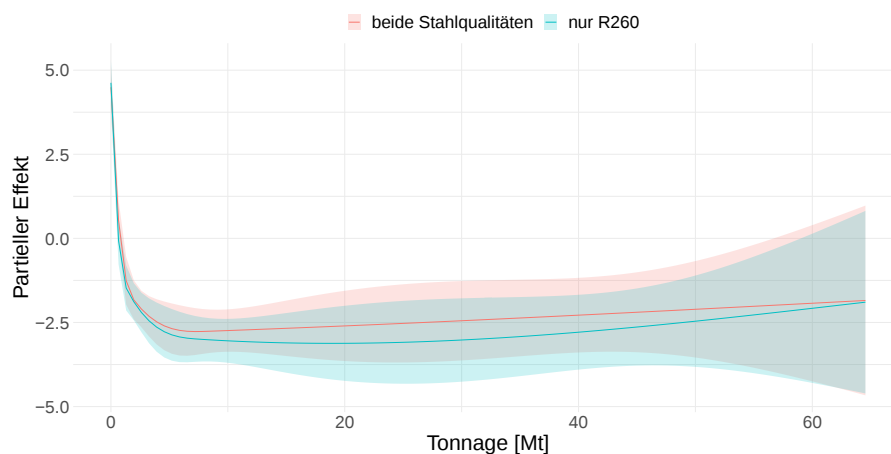


Abbildung 3.3.: Geschätzte glatte Funktionen für das kombinierte Modell und das Modell für die Stahlqualität R260.

Tabelle 3.2.: Koeffizienten der festen Effekte für das Modell mit beiden Stahlqualitäten

	Schätzwert	Standardfehler
(Intercept)	0.951	0.858
Stahl: R350	3.148	0.655
Maschine: B	1.645	0.663
Maschine: C	1.542	1.260
Maschine: D	2.185	0.538
Profil: UIC 60	2.133	0.700
Anteil GZ: wenig	2.134	0.671

Im Anhang in Abbildung A.30 sind die diagnostischen Plots für dieses Modell aufgezeichnet. Auf allen Plots sind zwei Ausreisser deutlich erkennbar. Der erste Ausreisser ist die Messung in Härkingen, die bereits im Modell für R260 aufgefallen ist. Der zweite Ausreisser ist die Messung in Vufflens, die wegen der Riffelung eine sehr hohe Rauheit aufweist. Glücklicherweise hat keine dieser Beobachtungen einen grossen Einfluss auf die Schätzung der Parameter und der Rauheit. Abgesehen von diesen Ausreissern sehen die Plots akzeptabel aus.

Auf dem QQ-Plot für den Ort in Abbildung A.35 fällt wieder Sempach 2 auf. Abgesehen von diesem

Ausreisser sieht auch dieser Plot akzeptabel aus.

Die P-Werte für die erklärenden Variablen geben wir hier nicht explizit an. Sie sind alle sehr klein (< 0.003).

Für die Schätzung der Rauheit verwenden wir im Folgenden das kombinierte Modell, akzeptieren also den additiven Effekt der Stahlqualität.

3.3. Vergleich mit dem Modell der zweiten Iteration

In diesem Abschnitt betrachten wir noch einmal das Modell der zweiten Iteration. Zuerst evaluieren wir, wie gut es die neuen Daten der dritten Iteration vorhersagen kann. Als Zweites passen wir das in der zweiten Iteration verwendete MARS-Modell noch einmal an. Diesmal jedoch mit allen Daten der dritten Iteration. Dann vergleichen wir den Vorhersagefehler mit demjenigen des additiven Modells, das wir in der aktuellen Iteration für die Vorhersage verwenden.

3.3.1. Evaluation des Modells der zweiten Iteration

Das Ziel der Modelle ist, die Rauheit auch an Stellen vorherzusagen, an denen noch keine Messungen gemacht wurden. Eine Möglichkeit, zu evaluieren wie gut sie das machen, besteht darin, mit dem Modell der vorherigen Iteration die neuen Messwerte vorherzusagen. Anschliessend werden die vom Modell zurückgegebenen Voraussagen mit den tatsächlichen Messungen verglichen. Schliesslich wird der Vorhersagefehler RMSE berechnet, welcher Aufschluss darüber gibt, wie gut die prognostizierten und die tatsächlichen Werte übereinstimmen.

Tabelle 3.3.: Vorhersagefehler für das MARS-Modell der zweiten Iteration.

Daten	RMSE
Daten der zweiten Iteration	2.13
neue Messungen der dritten Iteration	3.20

Tabelle 3.3 zeigt den RMSE für das MARS-Modell der zweiten Iteration, einerseits für die Daten der zweiten Iteration und andererseits für die neuen Messungen der dritten Iteration. Der RMSE für die Daten der zweiten Iteration wurde mit Kreuzvalidierung berechnet.

Die Vorhersage für die neuen Daten ist deutlich schlechter als diejenige für die alten Daten. Der RMSE ist rund 1.5 mal so gross. Dafür gibt es verschiedene Gründe. Die Rauheit der Messungen der dritten Iteration ist allgemein schwieriger vorherzusagen als die Rauheit der früheren Messungen. Das liegt unter anderem an den neu gemessenen Ausreissern mit sehr hoher Rauheit. Auch der RMSE der Modelle im Abschnitt 3.3.2, welche die Daten der dritten Iteration verwenden, bestätigen die erhöhte Schwierigkeit der Vorhersage.

Ein weiterer Grund für den hohen RMSE liegt im MARS-Modell, das in der zweiten Iteration verwendet wurde. Dieses Modell verwendet die Variable Ort nicht. Es unterscheidet also nicht, ob zwei Messungen an der gleichen Stelle oder an verschiedenen Orten gemacht wurden. Dadurch wird die Variabilität in den Daten unterschätzt, wenn mehrere Messungen vom gleichen Ort vorliegen. Das Modell passt sich zu sehr den Daten von den vorhandenen Orten an. Dies resultiert in einer schlechteren Vorhersage für neue Orte, an denen noch keine Messungen gemacht wurden.

3.3.2. Vergleich mit dem Modell der zweiten Iteration

Wie im Abschnitt 3.1.4 beschrieben, verwenden wir in dieser Iteration nur additive Modelle für die Vorhersage der Rauheit. Dennoch ist es interessant, die additiven Modelle mit den MARS-Modellen bezüglich Vorhersagefehler zu vergleichen.

Für den Vergleich haben wir das Modell für beide Stahlqualitäten verwendet und eine Kreuzvalidierung der Daten der dritten Iteration durchgeführt. Die Variable Ort haben wir für die Anpassung des

additiven Modells verwendet, jedoch nicht für die Vorhersage. Das entspricht der Situation, dass wir Vorhersagen für Orte machen, an denen noch keine Messung durchgeführt wurde. Die Vorhersage wird dadurch tendenziell ein wenig schlechter, entspricht dafür aber der realistischen Situation einer Vorhersage an einem neuen Ort.

Die Resultate sind in Tabelle 3.4 aufgeführt. Der RMSE ist für beide Modell fast identisch. Der Grund für den leicht höheren RMSE des additiven Modelles ist die Verwendung des Orts für die Anpassung, aber nicht für die Vorhersage. Für die Beobachtungen an einem Ort mit bereits vorhanden Messungen verwenden wir also nicht alle vorhanden Informationen. Dies führt tendenziell zu einem grösseren Vorhersagefehler, entspricht aber der realistischen Situation einer Vorhersage an einem neuen Ort. Es ist deshalb anzunehmen, dass das additive Modell für neue Orte bessere Vorhersagen liefert als das MARS-Modell.

Die Standardabweichung des RMSE ist für das MARS-Modell deutlich tiefer als für das additive Modell. Das MARS-Modell ist also stabiler, was ein Hinweis für eine leichte Überanpassung beim additiven Modell sein könnte.

Tabelle 3.4.: Vorhersagefehler für das additive Modell und das MARS-Modell basierend auf einer Kreuzvalidierung der Daten der dritten Iteration.

Modell	RMSE _{CV}	SD(RMSE _{CV})
additives Modell	2.48	0.546
MARS	2.47	0.445

4. Schätzung der Rauheit

In diesem Abschnitt stellen wir die Schätzung der Rauheit dar. Alle Schätzungen 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.

Für die Schätzung der Rauheit betrachten wir die beiden Stahlqualitäten separat, wie wir das bereits im Abschnitt 3.2 erwähnt haben.

4.1. Stahlqualität R260

Für die Schätzung der Rauheit für die Stahlqualität R260 verwenden wir das additive Modell, welches nur auf den Messungen mit R260 basiert. Neben der glatten Funktion der Tonnage und dem Ort als zufälligen Effekt enthält dieses Modell die Maschine, das Profil und den Anteil Güterzüge als feste Effekte. Es liefert also für jede der 16 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 Summe der Länge der relevanten Strecken auf dem SBB-Schienennetz bezieht sich nur auf den Oberbau und nicht auf die beim letzten Schleifen verwendete Maschine. Jede Strecke wird deshalb in der Tabelle mehrmals berücksichtigt.

Die geschätzte Rauheit liegt für 12 der 16 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 0.28 Mt der Fall. In der Spalte „Anteil über GW“ ist der Anteil der Zeit angegeben, während der der Grenzwert von 10 dB überschritten wird. Für den Stahl R260 ist dieser Anteil meisten Null und höchstens 0.56%. Für die Berechnung dieses Anteils gehen wir davon aus, dass bei der Stahlqualität R260 nach 50 Mt und bei R350 nach 25 Mt geschliffen wird.

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 und A.16 bis A.24 im Anhang zeigen die Regressionskurven (blau) und 95%-Prädiktionsintervalle (hellblau) für die Stahlqualität R260 und alle möglichen Kombinationen von Maschine, Profil und Anteil Güterzüge, von denen mindestens eine Messung zur Verfügung steht. Der kritische Grenzwert von 10 dB ist jeweils als rote horizontale Linie eingezeichnet.

Die in Tabelle 4.1 aufgeführte „Tonnagegrenze Prädiktionsintervalle“ bezeichnet den Tonnagewert, an dem die obere Grenze des 95%-Prädiktionsintervalls unter den Grenzwert von 10 dB fällt. Für eine neue Beobachtung mit den gleichen Ausprägungen der kategoriellen Variablen und diesem oder einem grösseren Tonnagenwert gehen wir mit einer Sicherheit von 95% davon aus, dass ihre Rauheit unter dem Grenzwert liegt. Diese Tonnagegrenze liegt für alle Kombinationen höchstens bei 2.09 Mt. Das heisst in höchstens 4.18% der Zeit liegt die obere Grenze des Prädiktionsintervalls über dem Grenzwert. Für die meisten Kombinationen ist der Anteil sogar deutlich kleiner.

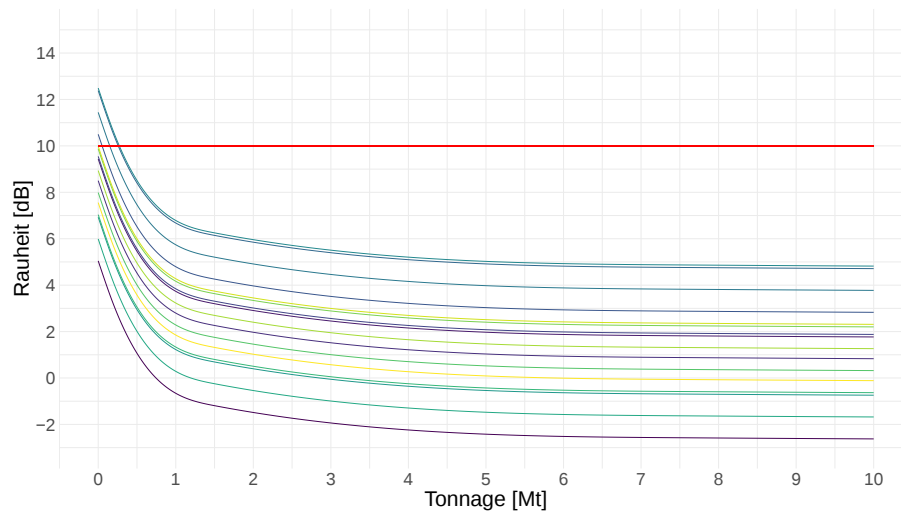


Abbildung 4.1.: Regressionskurven für die Stahlqualität R260 und alle 16 möglichen Kombinationen von Maschine, Profil und Anteil Güterzüge. Der kritische Grenzwert von 10 dB ist als rote horizontale Linie eingezeichnet.

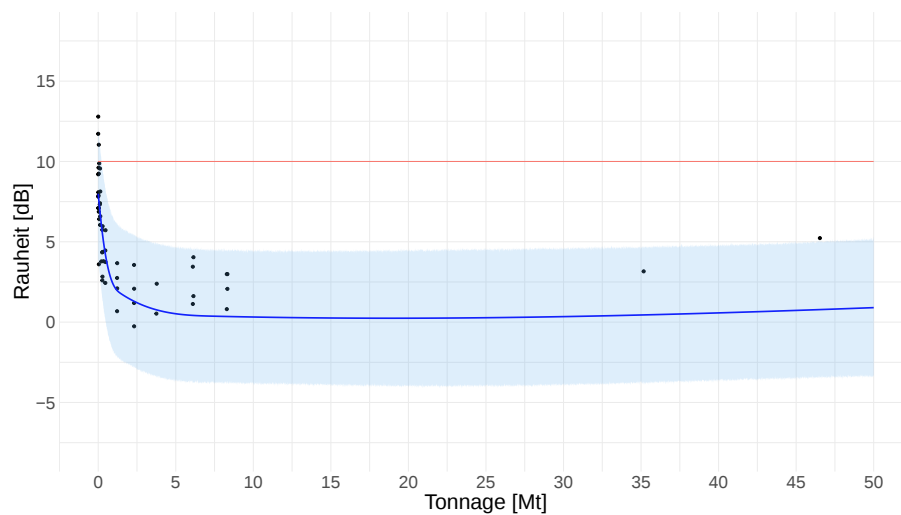


Abbildung 4.2.: 95%-Prädiktionsintervalle für Stahl R260, Maschine A, Profil UIC 54 und wenig Güterzüge.

Tabelle 4.1.: Geschätzte Rauheit, Tonnagegrenzen und der Anteil der Zeit, während der der Grenzwert überschritten wird, für die Stahlqualität R260 und alle 16 möglichen Kombinationen von Maschine, Profil und Anteil Güterzüge.

Maschine	Profil	Anteil Güterzüge	Relevante Strecke in km	Rauheit für Tonnage 0	Schätzung		Prädiktionsintervalle	
					Tonnagegrenze	Anteil über GW	Tonnagegrenze	Anteil über GW
D	UIC 60	wenig	911	12.492	0.28	0.56%	2.05	4.10%
B	UIC 60	wenig	911	12.383	0.26	0.52%	2.09	4.18%
C	UIC 60	wenig	911	11.445	0.15	0.30%	1.88	3.76%
A	UIC 60	wenig	911	10.500	0.06	0.12%	0.63	1.26%
D	UIC 54	wenig	406	9.981	NA	0.00%	0.56	1.12%
B	UIC 54	wenig	406	9.872	NA	0.00%	0.52	1.04%
D	UIC 60	viel	578	9.548	NA	0.00%	0.46	0.92%
B	UIC 60	viel	578	9.439	NA	0.00%	0.45	0.90%
C	UIC 54	wenig	406	8.935	NA	0.00%	0.52	1.04%
C	UIC 60	viel	578	8.502	NA	0.00%	0.48	0.96%
A	UIC 54	wenig	406	7.990	NA	0.00%	0.23	0.46%
A	UIC 60	viel	578	7.556	NA	0.00%	0.18	0.36%
D	UIC 54	viel	28	7.038	NA	0.00%	0.16	0.32%
B	UIC 54	viel	28	6.929	NA	0.00%	0.15	0.30%
C	UIC 54	viel	28	5.992	NA	0.00%	0.19	0.38%
A	UIC 54	viel	28	5.046	NA	0.00%	NA	0.00%

4.2. Stahlqualität R350

Wie in Abschnitt 3.2.2 ausgeführt, verwenden wir für die Schätzung der Rauheit für die Stahlqualität R350 das Modell, das auf den Messungen von beiden Stahlqualitäten basiert. Das heisst insbesondere, dass wir den additiven Effekt der Stahlqualität akzeptieren. Ohne diese Annahme wäre eine vernünftige Schätzung nicht möglich.

Dieses Modell verwendet die gleichen erklärenden Variablen wie das im vorherigen Abschnitt verwendete Modell für den Stahl R260, einzig erweitert um die Stahlqualität als festen Effekt.

In Abbildung 4.3 sind die Regressionskurven abgebildet und in Tabelle 4.2 sind zu jeder Kombination diverse Informationen aufgeführt. Die geschätzte Rauheit liegt deutlich höher als für den Stahl R260. Nur bei einer der 16 Kombinationen liegt sie bereits bei Tonnage 0 unter dem kritischen Grenzwert von 10 dB. Aber für alle Kombinationen fällt sie spätestens nach 1 Mt unter den Grenzwert. Die geschätzte Rauheit überschreitet den Grenzwert also für alle Kombinationen für weniger als 4% der Zeit.

Wenn man die Schätzunsicherheit mitberücksichtigen will, muss man die Prädiktionsintervalle betrachten. Diese enthalten den Grenzwert von 10 dB für einige Kombinationen über einen grossen Teil der Tonnagewerte; für drei Kombinationen sogar über den ganzen Tonnagebereich. Dort können wir also keine gesicherte Aussage darüber machen, dass eine zukünftige Beobachtung den Grenzwert nicht überschreitet. Aber die geschätzte Rauheit liegt wie erwähnt meistens unter 10 dB.

Die Regressionskurven und 95%-Prädiktionsintervalle für die Stahlqualität R350 sind in den Abbildungen 4.2 und A.16 bis A.24 im Anhang aufgezeichnet.

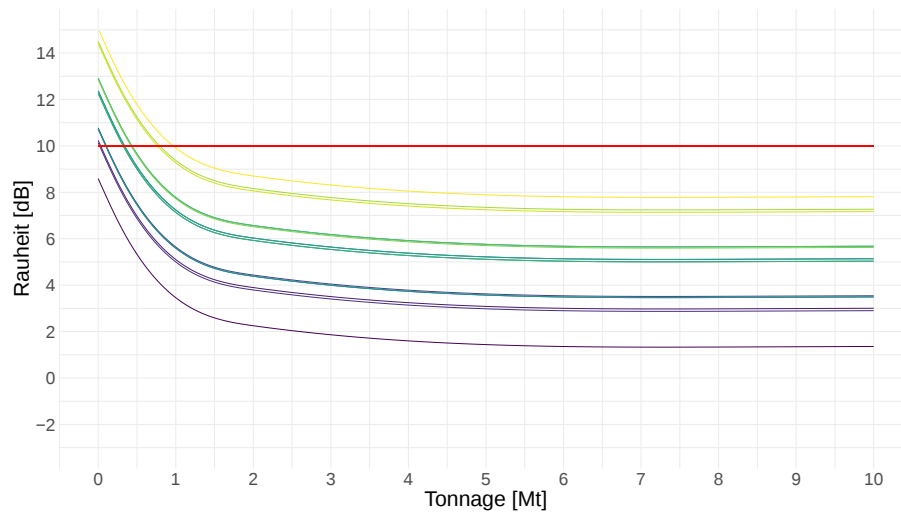


Abbildung 4.3.: Regressionskurven für die Stahlqualität R350 und alle 16 möglichen Kombinationen von Maschine, Profil und Anteil Güterzüge. Der kritische Grenzwert von 10 dB ist als rote horizontale Linie eingezeichnet.

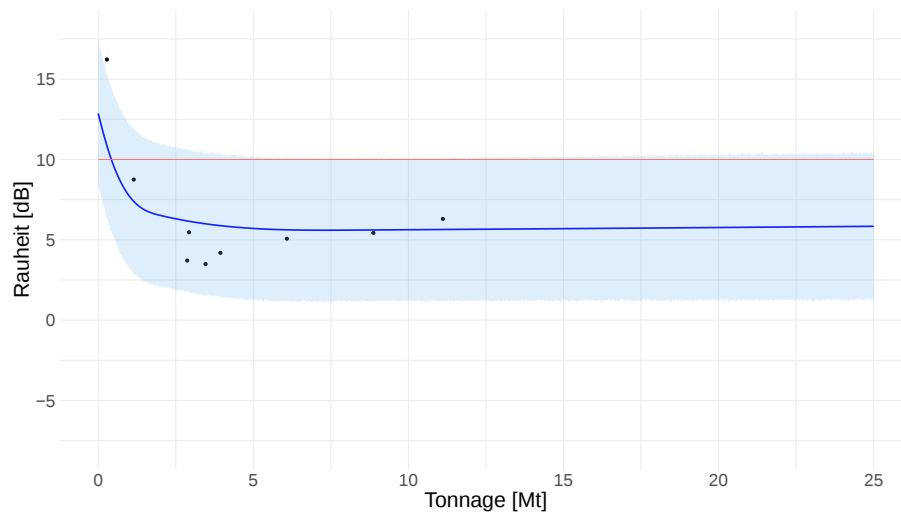


Abbildung 4.4.: 95%-Prädiktionsintervalle für Stahl R350, Maschine A, Profil UIC 60 und wenig Güterzüge.

Tabelle 4.2.: Geschätzte Rauheit, Tonnagegrenzen und der Anteil der Zeit, während der der Grenzwert überschritten wird, für die Stahlqualität R350 und alle 16 möglichen Kombinationen von Maschine, Profil und Anteil Güterzüge.

Maschine	Profil	Anteil Güterzüge	Relevante Strecke in km	Rauheit für Tonnage 0	Schätzung		Prädiktionsintervalle	
					Tonnagegrenze	Anteil über GW	Tonnagegrenze	Anteil über GW
D	UIC 60	wenig	204	15.044	0.97	3.88%	NA	100.00%
B	UIC 60	wenig	204	14.504	0.80	3.20%	NA	100.00%
C	UIC 60	wenig	204	14.401	0.77	3.08%	NA	100.00%
D	UIC 54	wenig	111	12.911	0.44	1.76%	8.24	32.96%
D	UIC 60	viel	112	12.909	0.44	1.76%	6.50	26.00%
A	UIC 60	wenig	204	12.859	0.43	1.72%	5.04	20.16%
B	UIC 54	wenig	111	12.371	0.34	1.36%	3.66	14.64%
B	UIC 60	viel	112	12.370	0.34	1.36%	3.16	12.64%
C	UIC 54	wenig	111	12.267	0.33	1.32%	5.67	22.68%
C	UIC 60	viel	112	12.266	0.33	1.32%	5.02	20.08%
D	UIC 54	viel	63	10.776	0.11	0.44%	1.09	4.36%
A	UIC 54	wenig	111	10.726	0.10	0.40%	1.03	4.12%
A	UIC 60	viel	112	10.724	0.10	0.40%	1.06	4.24%
B	UIC 54	viel	63	10.236	0.04	0.16%	0.96	3.84%
C	UIC 54	viel	63	10.133	0.02	0.08%	1.16	4.64%
A	UIC 54	viel	63	8.591	NA	0.00%	0.51	2.04%

5. Diskussion

Die einfachen Modelle aus der Vorstudie mit stückweise linearen Regressionskurven konnten mit den neuen Messungen deutlich verbessert werden. Die nun insgesamt 191 Beobachtungen ermöglichen eine stabile Schätzung von additiven Modellen mit sehr flexiblen Regressionskurven basierend auf kubischen Splines. Diese glatten Regressionskurven entsprechen viel eher dem vermuteten wahren Zusammenhang zwischen Schienenrauheit und Tonnage. Die additiven Modelle berücksichtigen zudem den Ort als zufälligen Effekt und können so die Abhängigkeitsstruktur zwischen den einzelnen Messungen besser abbilden.

Die Flexibilität der Regressionskurven wird nur durch die Annahme der Additivität der Effekte der kategoriellen Erklärenden eingeschränkt. Es wird beispielsweise angenommen, dass sich für die beiden Stahlqualitäten die Verläufe der Schienenrauheit über die Tonnage nur um eine additive Konstante unterscheiden. Dies ist wahrscheinlich eine zu starke Vereinfachung. Ohne diese Annahme hätte das Modell aber viel zu viele Freiheitsgrade und könnte nur mit deutlich mehr Beobachtungen einigermaßen stabil geschätzt werden.

Für die Stahlqualität R260 stehen mit 159 Messungen relativ viele Beobachtungen zur Verfügung. Mit dieser Datengrundlage können relativ stabile Modelle gerechnet werden. Diese Modelle ermöglichen eine ziemlich verlässliche Schätzung der Schienenrauheit. Das Modell verwendet die Maschine, das Profil und den Anteil Güterzüge als feste kategorielle Effekte. Daraus ergeben sich 16 Kombinationen, für die jeweils eine Regressionskurve angegeben werden kann. Die geschätzte Rauheit liegt für 12 der 16 Kombinationen bereits bei Tonnage 0 unter dem kritischen Grenzwert von 10 dB. Der Grenzwert wird für diese Kombinationen also nie überschritten. Auch für die anderen Kombinationen fällt die Rauheit ziemlich bald unter den Grenzwert. Dies ist für alle Kombinationen spätestens bei 0.28 Mt der Fall. Der Grenzwert wird von der geschätzten Rauheit für die meisten Kombinationen also nie überschritten und für alle höchstens während 0.56% der Zeit.

Die Schätzungen der Rauheit sind natürlich mit einer gewissen Unsicherheit behaftet. Diese wird durch eine reine Betrachtung der Schätzwerte nicht berücksichtigt. Mit 95%-Prädiktionsintervallen können diese Unsicherheiten quantifiziert werden. Eine neue Beobachtung mit den gegebenen Ausprägungen liegt mit einer Sicherheit von 95% in diesem Intervall. Für eine Kombination liegt die obere Grenze dieses Intervalls bereits bei Tonnage 0 unter dem Grenzwert. Für diese Kombination können wir also über den gesamten Tonnagebereich davon ausgehen, dass eine neue Beobachtung mit einer Sicherheit von 95% den Grenzwert nicht überschreitet. Für die meisten anderen Kombinationen unterschreitet die obere Grenze des Prädiktionsintervalls bereits bei weniger als einer Megatonne den Grenzwert von 10 dB. Bei einer Kombination ist dies erst bei 2.09 Mt der Fall. Das heisst für alle Kombinationen liegt die obere Grenze des Prädiktionsintervalls in höchstens 4.18% der Zeit über dem Grenzwert. Für die meisten Kombinationen ist der Anteil sogar deutlich kleiner.

Für die Stahlqualität R350 stehen nur 32 Messungen zur Verfügung. Mit dieser Datengrundlage ist es nicht möglich, ein einigermaßen stabiles und verlässliches Modell zu rechnen. Für die Schätzung der Rauheit für die Stahlqualität R350 wird deshalb das Modell verwendet, das auf den Messungen von beiden Stahlqualitäten basiert. Das heisst aber insbesondere, dass der additive Effekt der Stahlqualität akzeptiert werden muss. Ohne diese Annahme wäre eine vernünftige Schätzung nicht möglich.

Die geschätzte Rauheit liegt für den Stahl R350 3.15 dB höher als für den Stahl R260. Nur bei einer der 16 Kombinationen liegt sie bereits bei Tonnage 0 unter dem kritischen Grenzwert von 10 dB. Aber für alle Kombinationen fällt sie spätestens nach 1 Mt unter den Grenzwert. Die geschätzte Rauheit überschreitet den Grenzwert also für alle Kombinationen für weniger als 4% der Zeit.

Die Prädiktionsintervalle enthalten den Grenzwert von 10 dB für einige Kombinationen über einen grossen Teil der Tonnagewerte; für drei Kombinationen sogar über den ganzen Tonnagebereich. Dort können also keine gesicherte Aussage darüber gemacht werden, dass eine zukünftige Beobachtung

den Grenzwert nicht überschreitet. Aber die geschätzte Rauheit liegt wie erwähnt meistens unter 10 dB.

In der Vorstudie wurden elf potenziell relevante Einflussfaktoren berücksichtigt. Davon verwenden die aktuellen Modelle nur noch fünf. Neben der Tonnage sind dies die Stahlqualität, die Maschine, der Anteil Güterzüge und das Profil. Die Selektion dieser Variablen ist ziemlich stabil. Bereits in der zweiten Iteration wurden sie selektiert. Einzig die Maschine wurde damals nicht ins Modell genommen. Das liegt wahrscheinlich daran, dass die Maschine vor allem direkt nach dem Schleifen relevant ist und der Effekt mit zunehmender Tonnage abnimmt. Ein rein additiver Effekt der Maschine ist deshalb wahrscheinlich eine zu starke Annahme.

Laut den Expertinnen der SBB ist es aus fachlicher Sicht plausibel, dass die Tonnage, die Stahlqualität, die Maschine und der Anteil Güterzüge einen Einfluss auf die Schienenrauheit haben. Der Effekt des Schienenprofils ist jedoch schwieriger zu erklären. Ein direkter Effekt aufgrund von unterschiedlichen physikalischen Eigenschaften der beiden Profile ist eher unwahrscheinlich. Beispielsweise unterscheiden sie sich nicht bezüglich Kontaktfläche zwischen Schiene und Rad.

Die Profile UIC 54 und UIC 60 werden aber an unterschiedlichen Orten verbaut. Das Profil UIC 54 wird vor allem auf kurvenreichen Strecken mit kleinen Radii verwendet. Weil sich kurvenreiche Strecken nicht für IC-Züge eignen, unterscheiden sich die beiden Profile also nicht zuletzt bezüglich der Art des Personenverkehrs. Auf dem Profil UIC 60 verkehren Fern- und Regionalzüge, während auf UIC 54 vorwiegend Regionalzüge unterwegs sind.

Für eine zukünftige Studie wäre es interessant zu untersuchen, ob die Art des Personenverkehrs einen Einfluss auf die Schienenrauheit hat. Auch interessant wäre die Verwendung weiterer Variablen wie die Traktionskraft, mit denen der Effekt des Profils auf die Schienenrauheit allenfalls direkter erklärt werden könnte. Dies würde jedoch voraussetzen, dass die entsprechenden Informationen für alle Messpunkte verfügbar wären.

Im Vergleich zur Vorstudie sind die aktuellen Modelle viel robuster und verlässlicher. Wenn man diese Modelle noch einmal deutlich verbessern möchte, müsste die Datengrundlage noch einmal deutlich erweitert werden. Eine Möglichkeit wären zusätzliche punktuelle Messungen mit Stahlqualität R350. Damit könnte man auf die sehr starke Annahme der Additivität des Effekts der Stahlqualität verzichten.

Sehr interessant wäre auch die Verwendung der Daten des ETH-Projekts [4]. Durch die Messung der Schienenrauheit mit einem fahrenden Zug wäre eine ganz andere Größenordnung der Anzahl Messungen möglich. Diese Datengrundlage würde eine deutlich differenziertere und genauere Schätzung der Schienenrauheit ermöglichen.

A. Anhang

A.1. Beschreibung der einzelnen Variablen

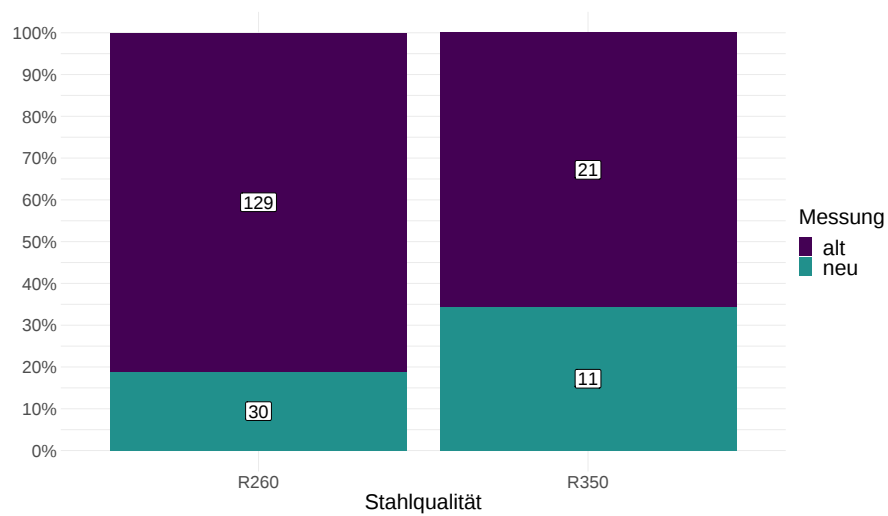


Abbildung A.1.: Balkendiagramm der Stahlqualität pro Messung.

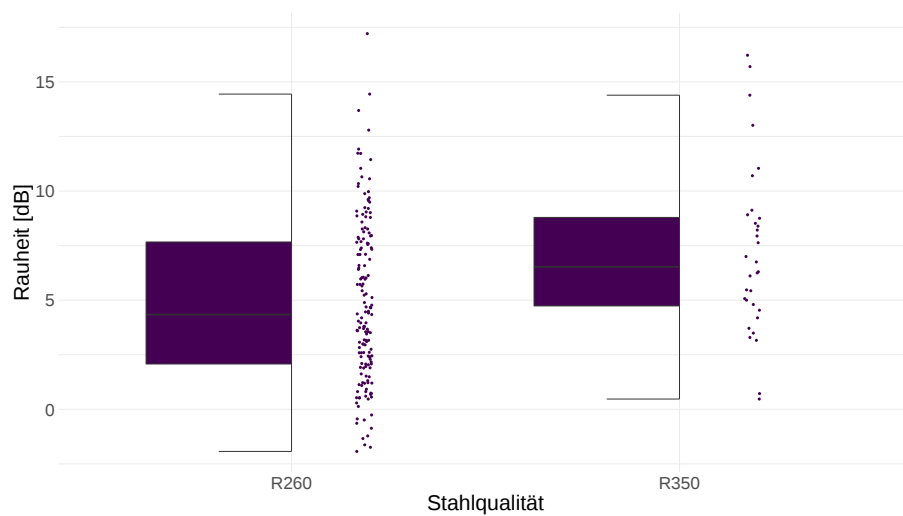


Abbildung A.2.: Boxplots der Rauheit pro Stahlqualität.

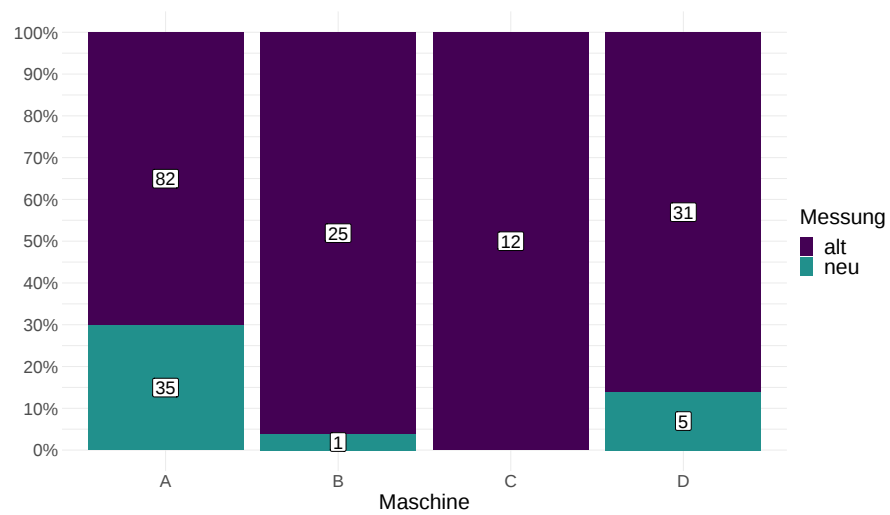


Abbildung A.3.: Balkendiagramm der Schleiftechnik pro Messung.

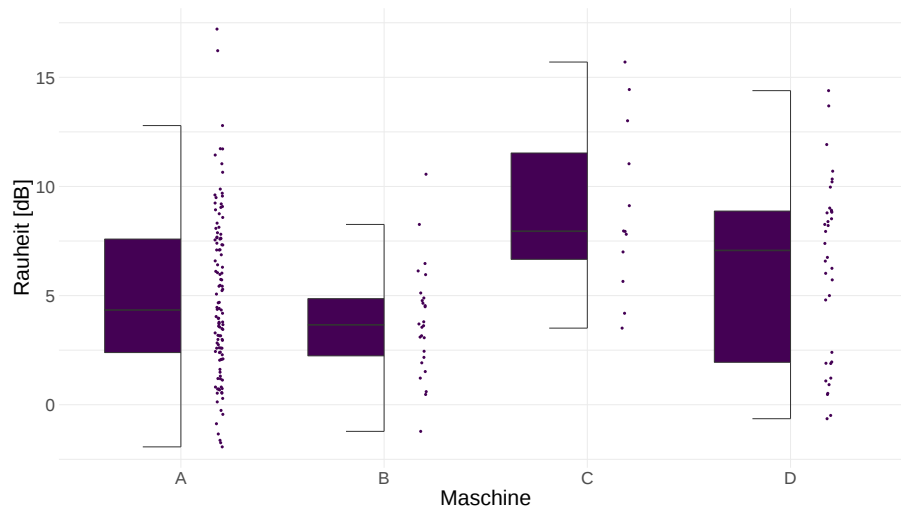


Abbildung A.4.: Boxplots der Rauheit pro Schleiftechnik.

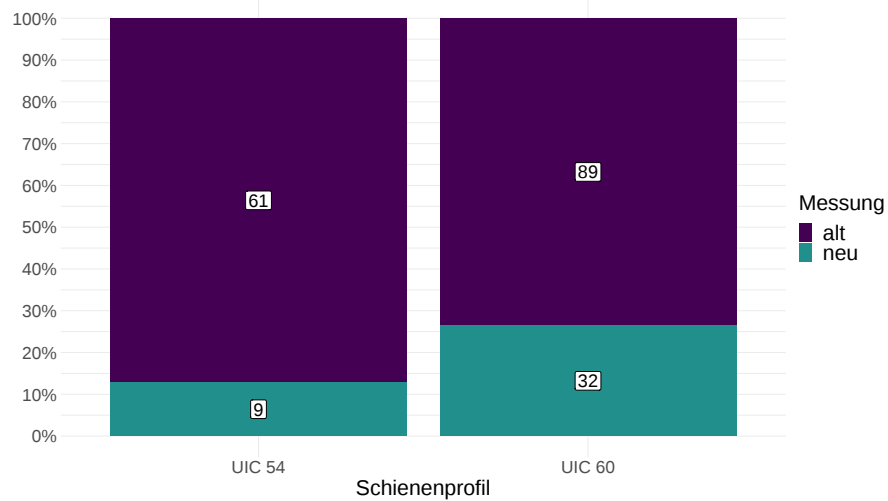


Abbildung A.5.: Balkendiagramm für das Schienenprofil pro Messung.

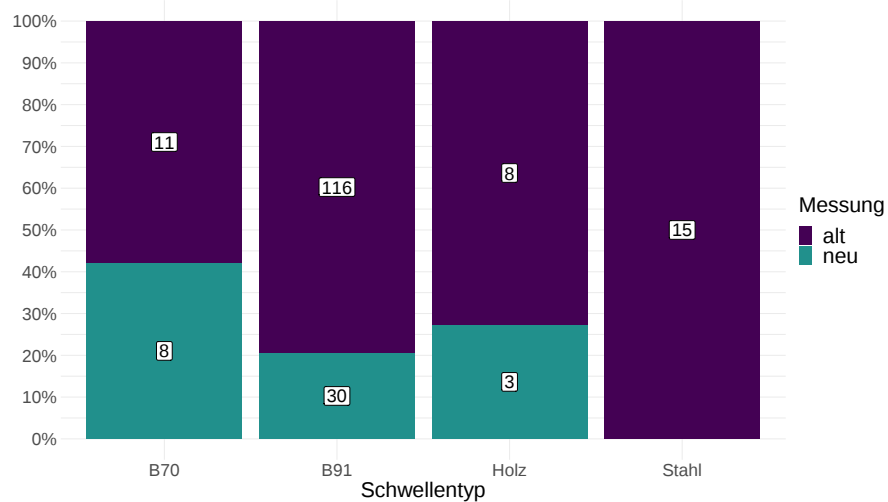


Abbildung A.6.: Balkendiagramm des Schwellentyps pro Messung.

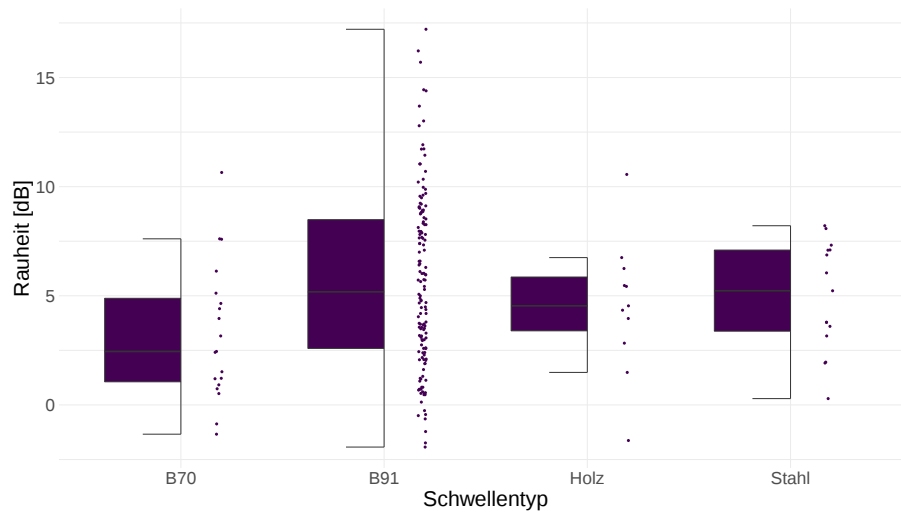


Abbildung A.7.: Boxplots der Rauheit pro Schwellentyp.

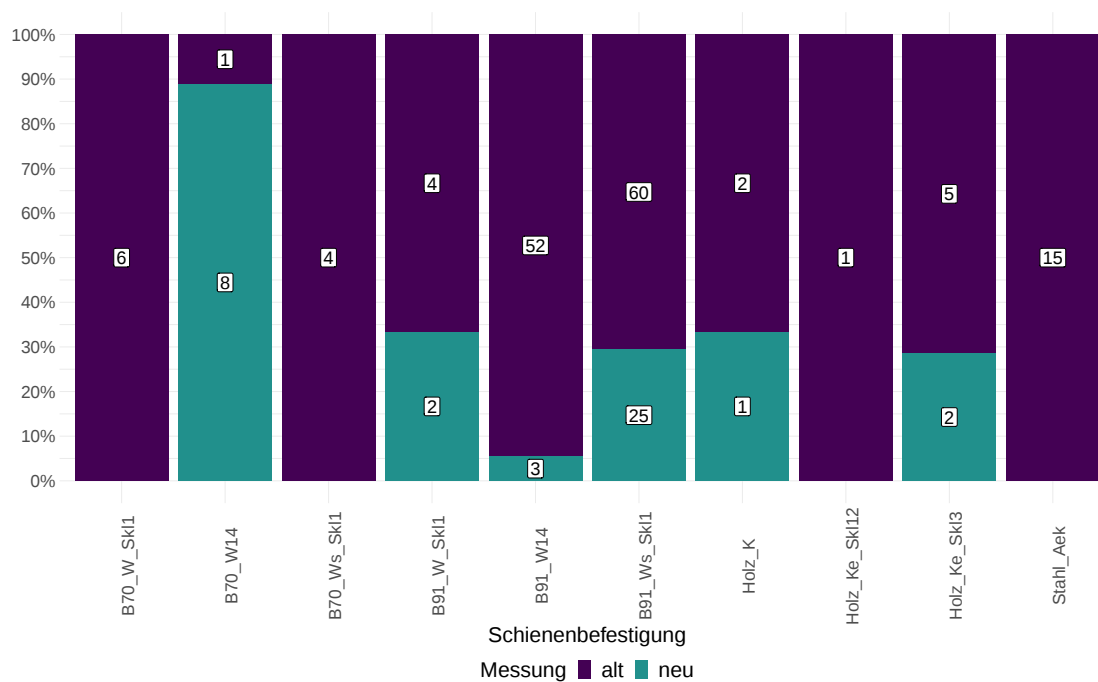


Abbildung A.8.: Balkendiagramm der Schienenbefestigung pro Messung.

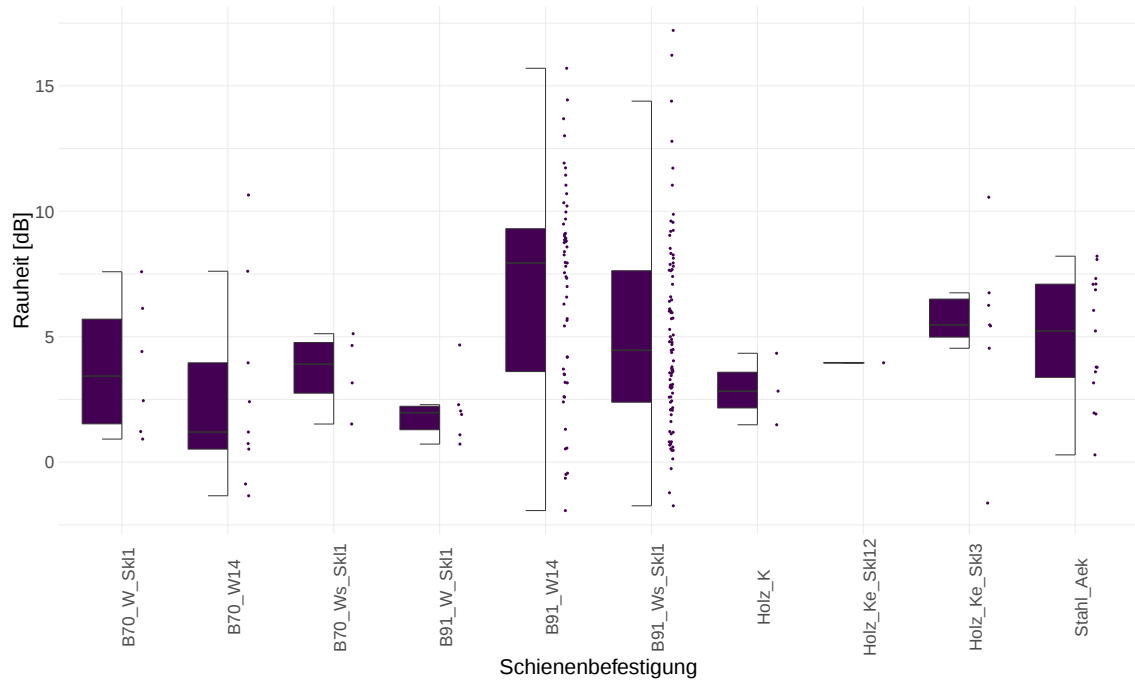


Abbildung A.9.: Boxplots der Rauheit pro Schienenbefestigungstyp.

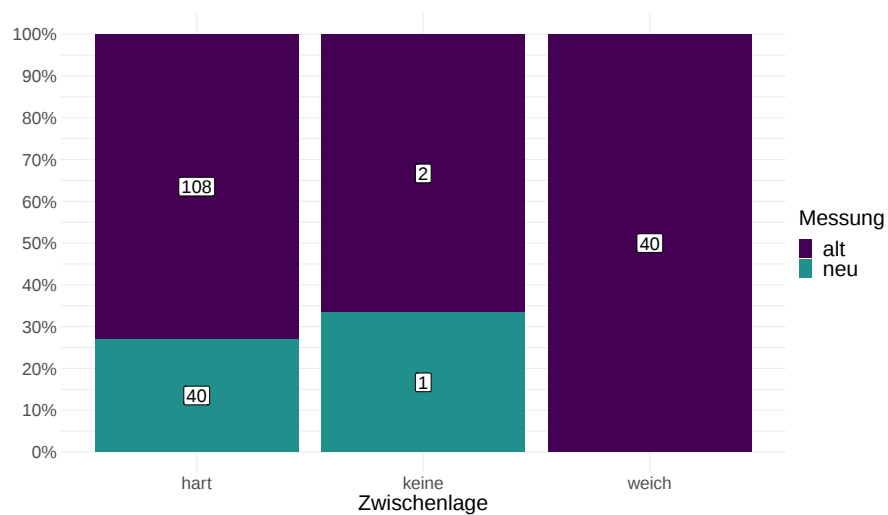


Abbildung A.10.: Balkendiagramm der Zwischenlage pro Messung.

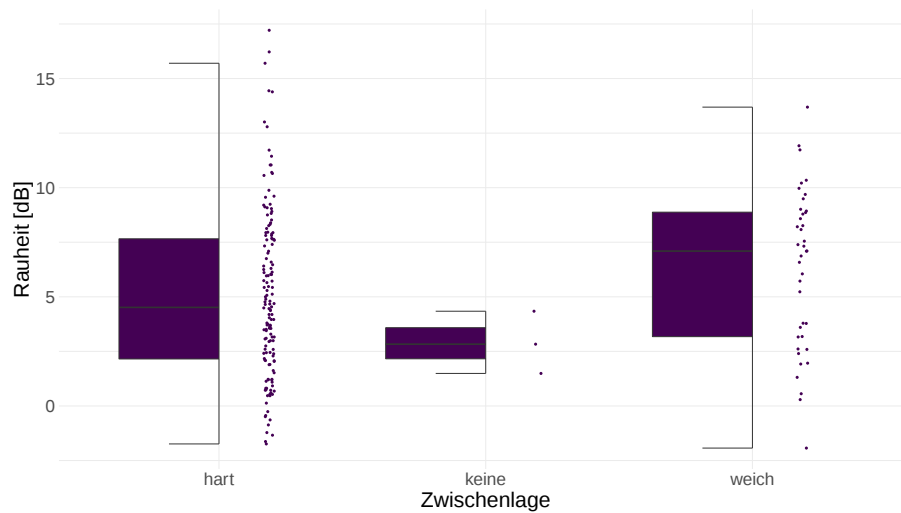


Abbildung A.11.: Boxplots der Rauheit pro Zwischenlage.

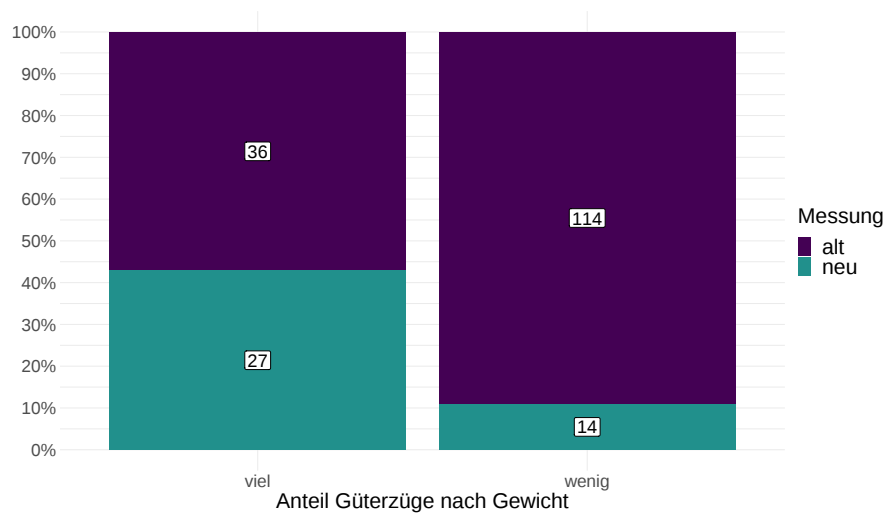


Abbildung A.12.: Balkendiagramm von Anteil Güterzüge pro Messung.

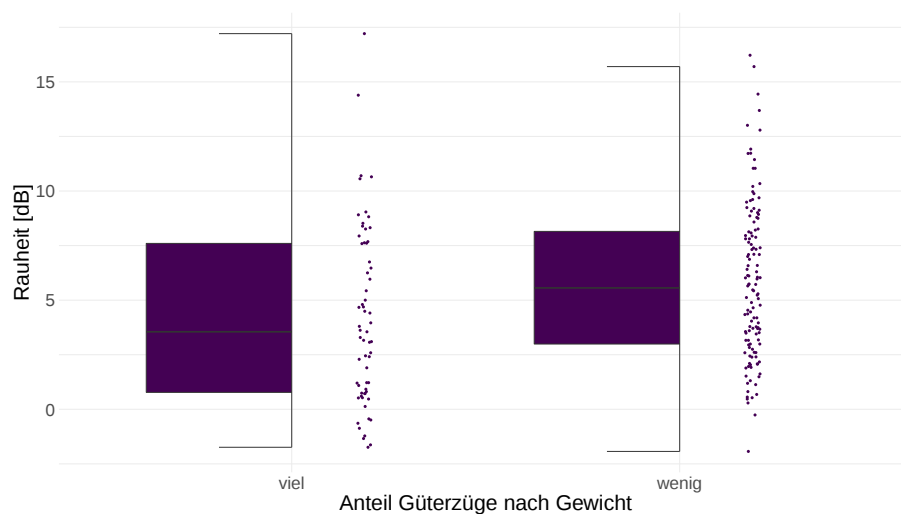


Abbildung A.13.: Boxplots der Rauheit pro Anteil Güterzüge.

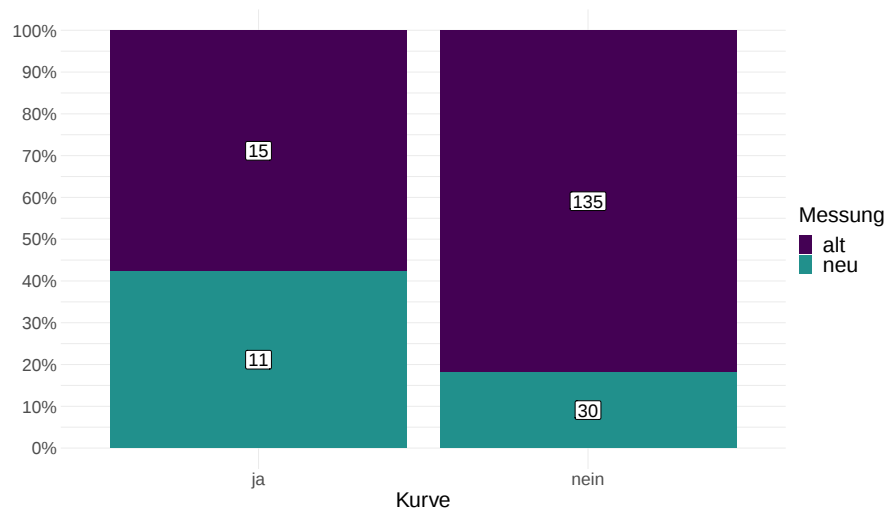


Abbildung A.14.: Balkendiagramm der Streckengeometrie Messung.

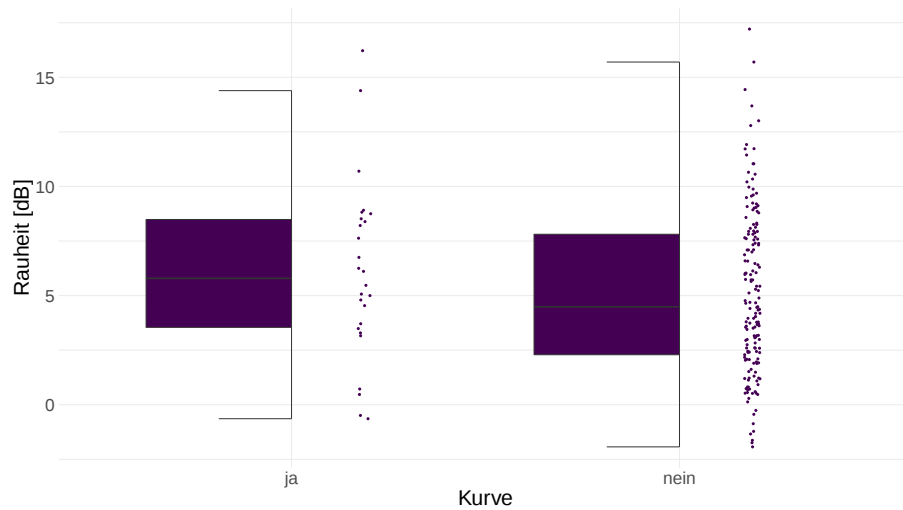


Abbildung A.15.: Boxplots der Rauheit pro Streckengeometrie.

A.2. Prädiktionsintervalle

Die Abbildungen A.16 bis A.28 zeigen die Regressionskurven (blau) und 95%-Prädiktionsintervalle (hellblau) für verschiedene Kombinationen von Stahlqualität, Maschine, Profil und Anteil Güterzüge. Der kritische Grenzwert von 10 dB ist jeweils als rote horizontale Linie eingezeichnet.

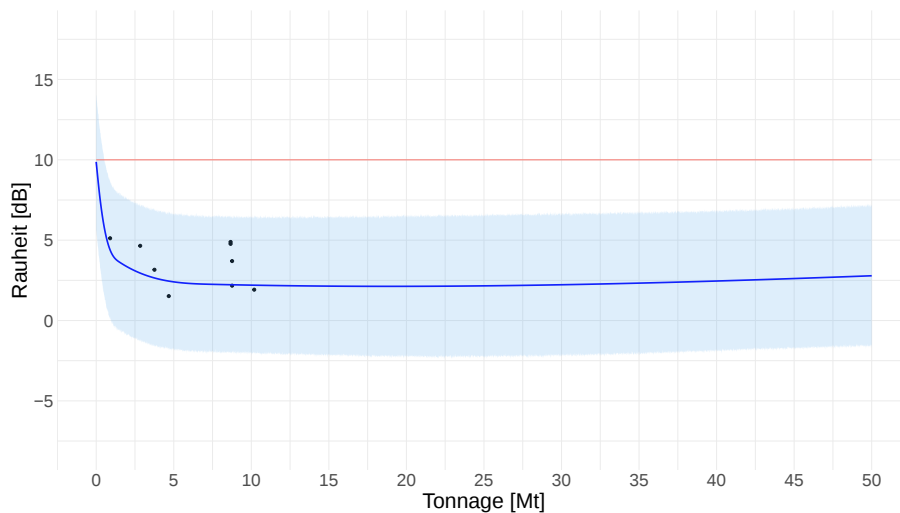


Abbildung A.16.: 95%-Prädiktionsintervalle für Stahl R260, Maschine B, Profil UIC 54 und wenig Güterzüge.

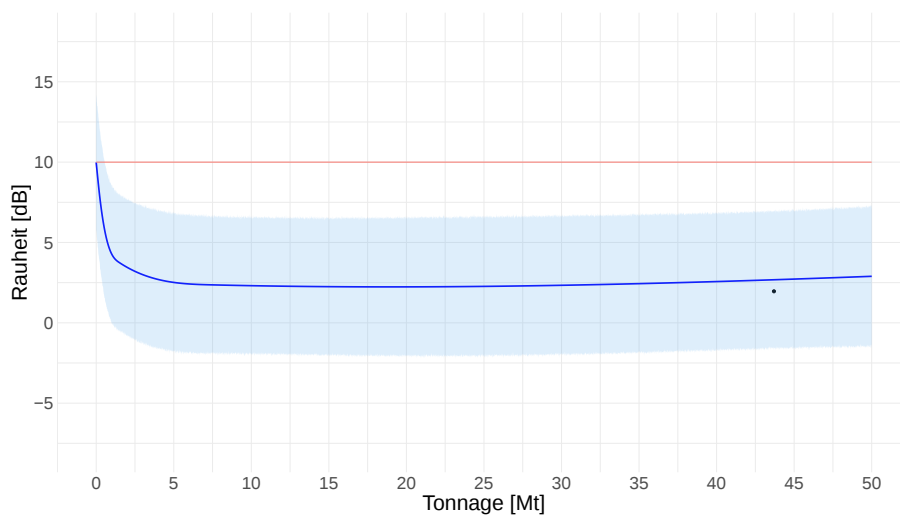


Abbildung A.17.: 95%-Prädiktionsintervalle für Stahl R260, Maschine D, Profil UIC 54 und wenig Güterzüge.

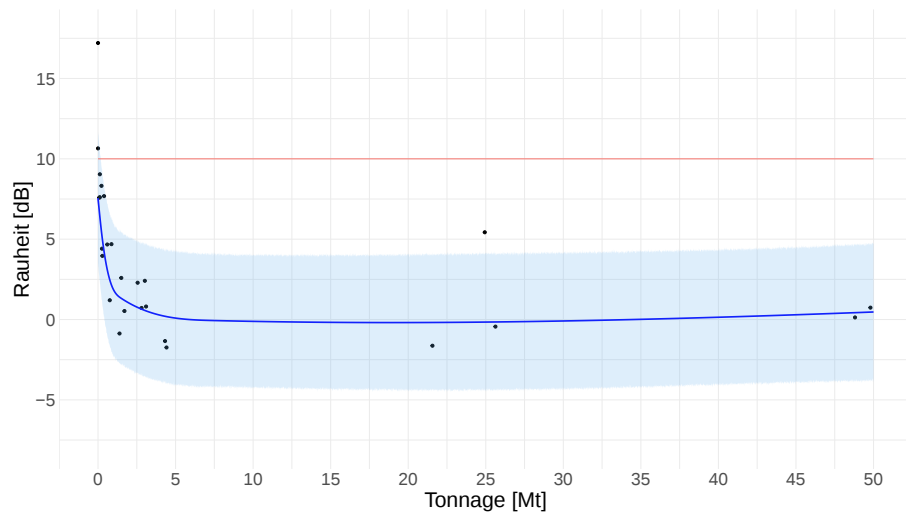


Abbildung A.18.: 95%-Prädiktionsintervalle für Stahl R260, Maschine A, Profil UIC 60 und viel Güterzüge.

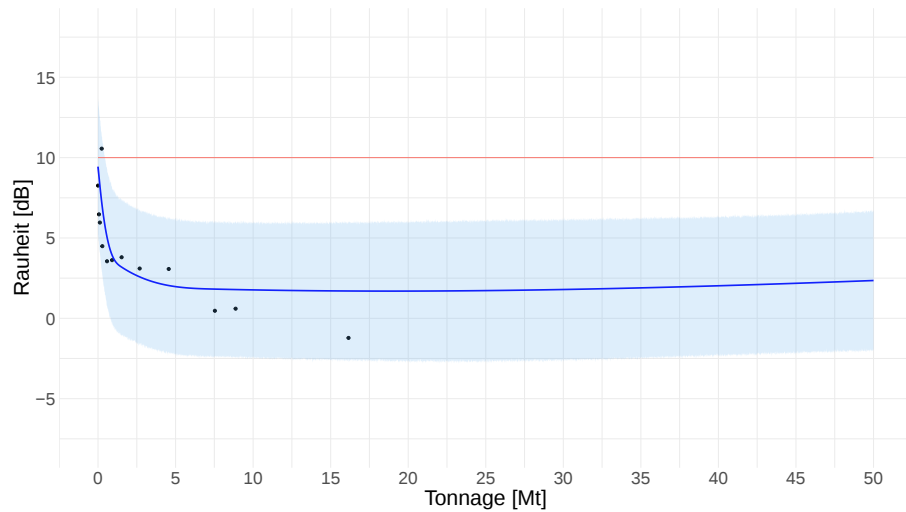


Abbildung A.19.: 95%-Prädiktionsintervalle für Stahl R260, Maschine B, Profil UIC 60 und viel Güterzüge.

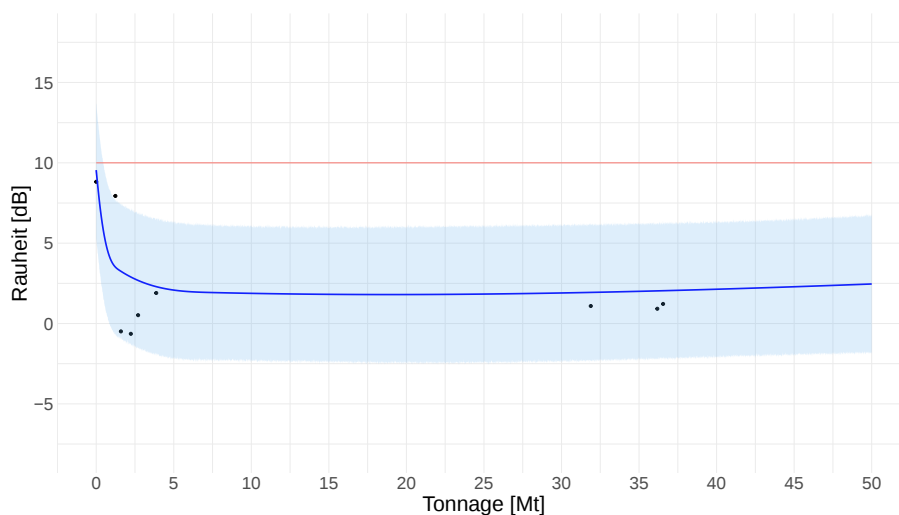


Abbildung A.20.: 95%-Prädiktionsintervalle für Stahl R260, Maschine D, Profil UIC 60 und viel Güterzüge.

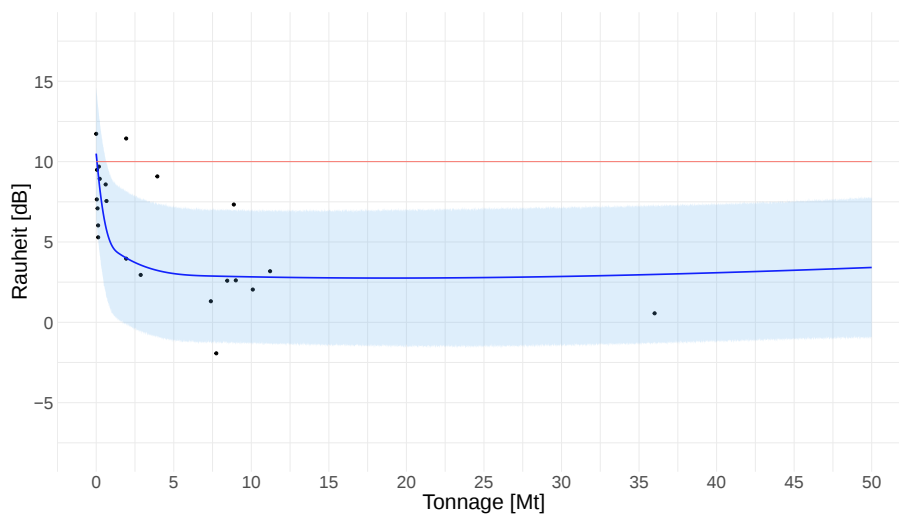


Abbildung A.21.: 95%-Prädiktionsintervalle für Stahl R260, Maschine A, Profil UIC 60 und wenig Güterzüge.

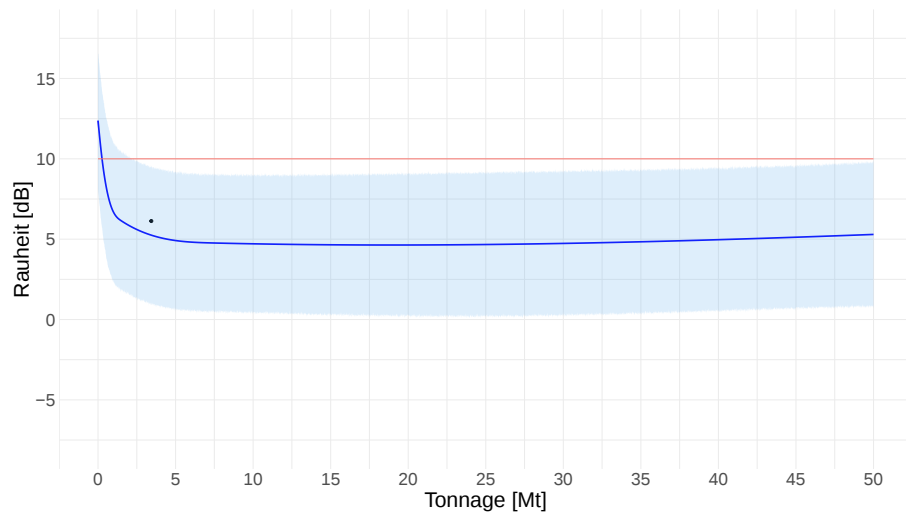


Abbildung A.22.: 95%-Prädiktionsintervalle für Stahl R260, Maschine B, Profil UIC 60 und wenig Güterzüge.

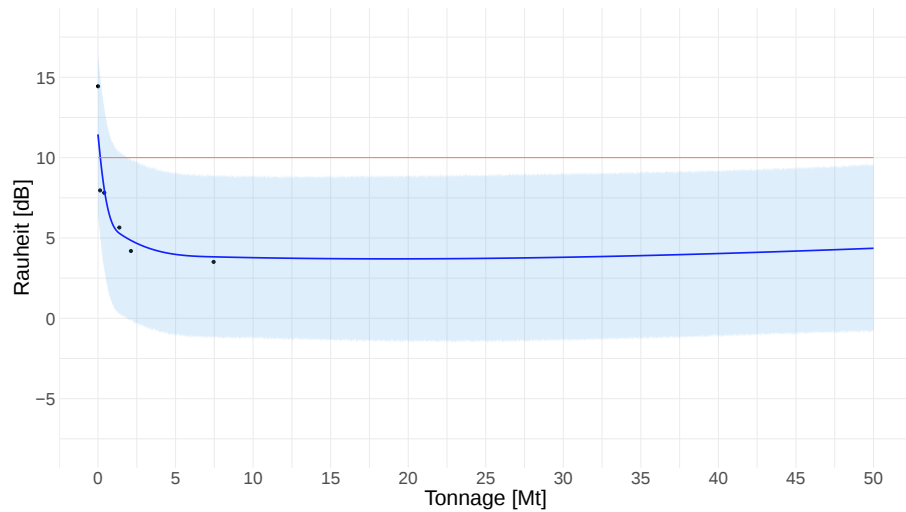


Abbildung A.23.: 95%-Prädiktionsintervalle für Stahl R260, Maschine C, Profil UIC 60 und wenig Güterzüge.

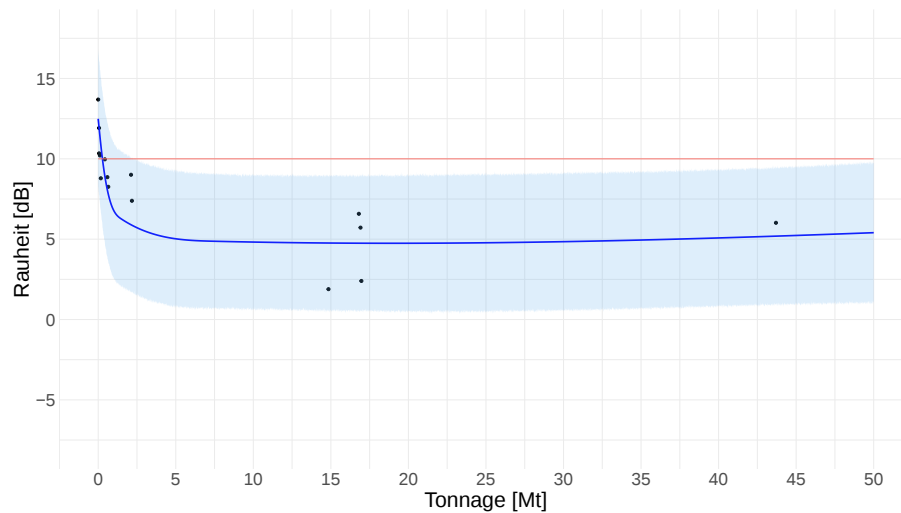


Abbildung A.24.: 95%-Prädiktionsintervalle für Stahl R260, Maschine D, Profil UIC 60 und wenig Güterzüge.

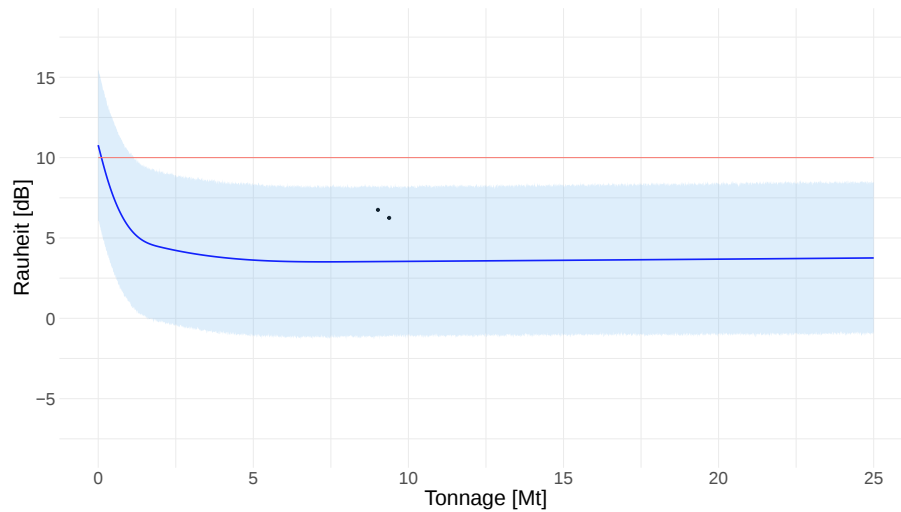


Abbildung A.25.: 95%-Prädiktionsintervalle für Stahl R350, Maschine D, Profil UIC 54 und viel Güterzüge.

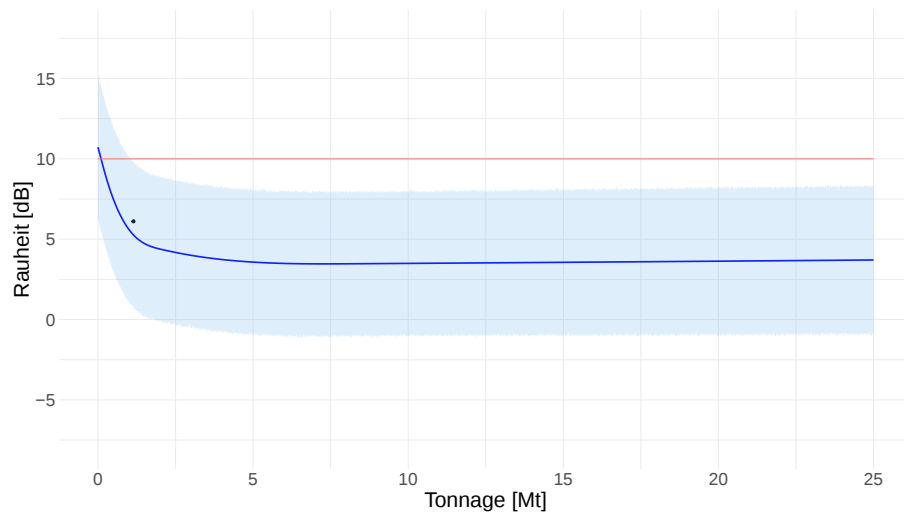


Abbildung A.26.: 95%-Prädiktionsintervalle für Stahl R350, Maschine A, Profil UIC 54 und wenig Güterzüge.

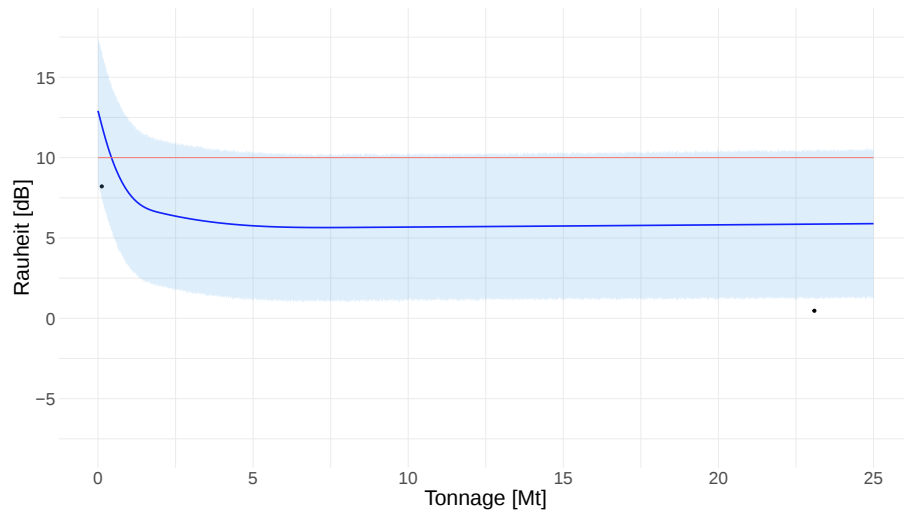


Abbildung A.27.: 95%-Prädiktionsintervalle für Stahl R350, Maschine D, Profil UIC 54 und wenig Güterzüge.

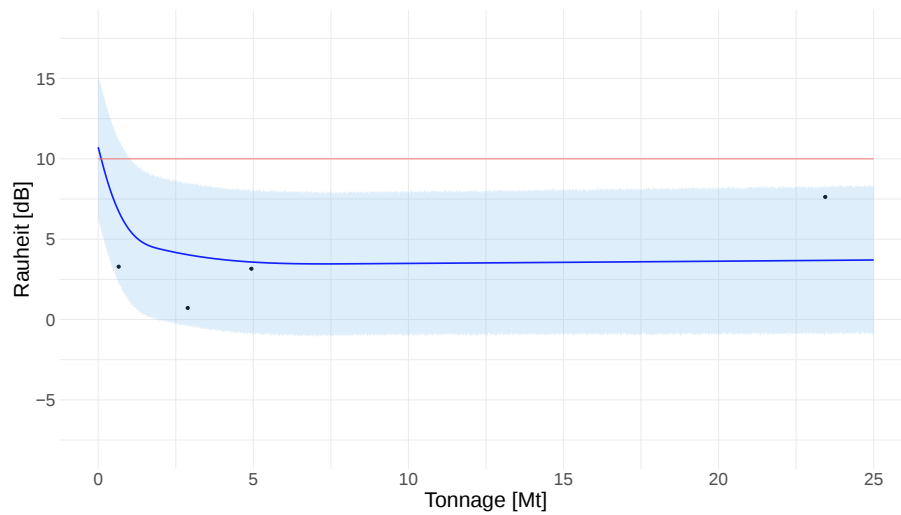


Abbildung A.28.: 95%-Prädiktionsintervalle für Stahl R350, Maschine A, Profil UIC 60 und viel Güterzüge.

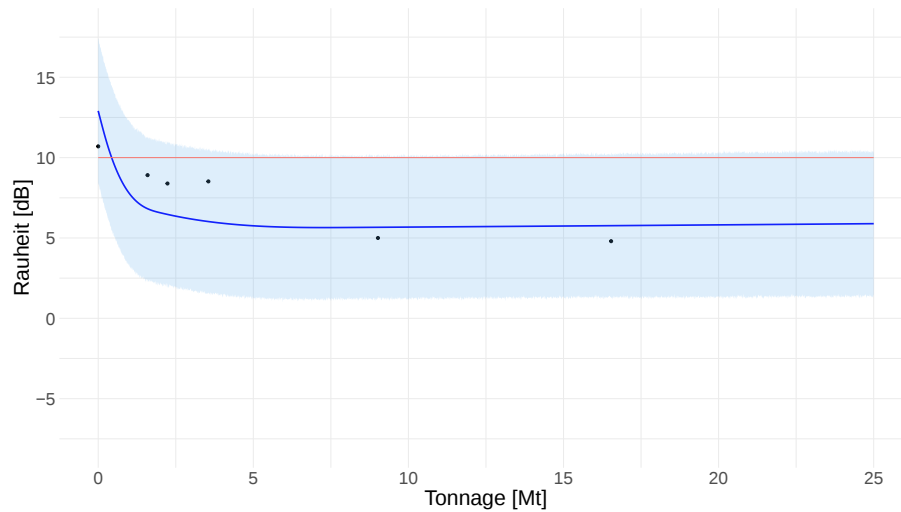


Abbildung A.29.: 95%-Prädiktionsintervalle für Stahl R350, Maschine D, Profil UIC 60 und viel Güterzüge.

A.3. Diagnostische Plots

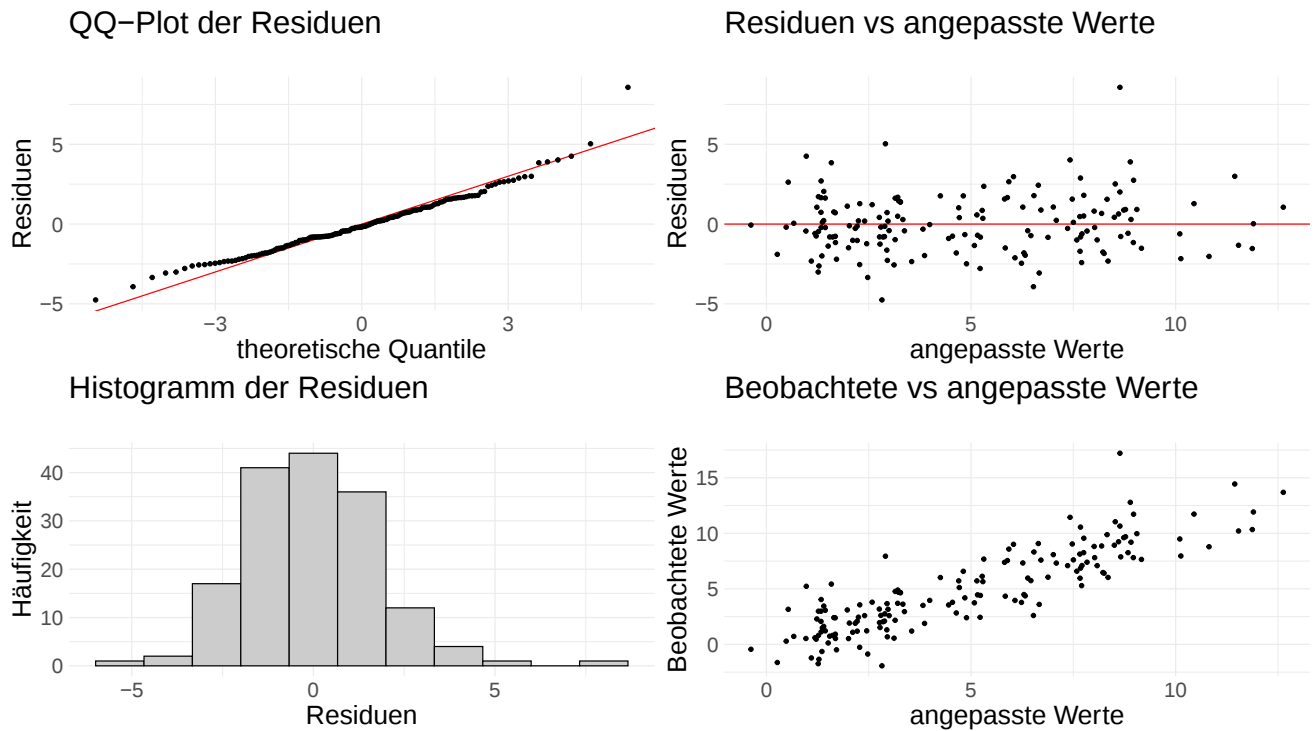


Abbildung A.30.: Diagnostische Plots für das Modell für die Stahlqualität R260.

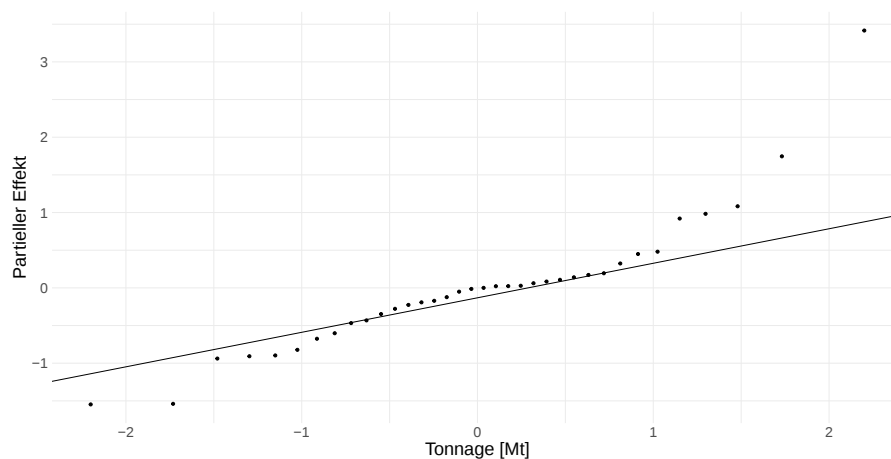


Abbildung A.31.: QQ-Plot für den zufälligen Effekt Ort für das Modell für die Stahlqualität R260.

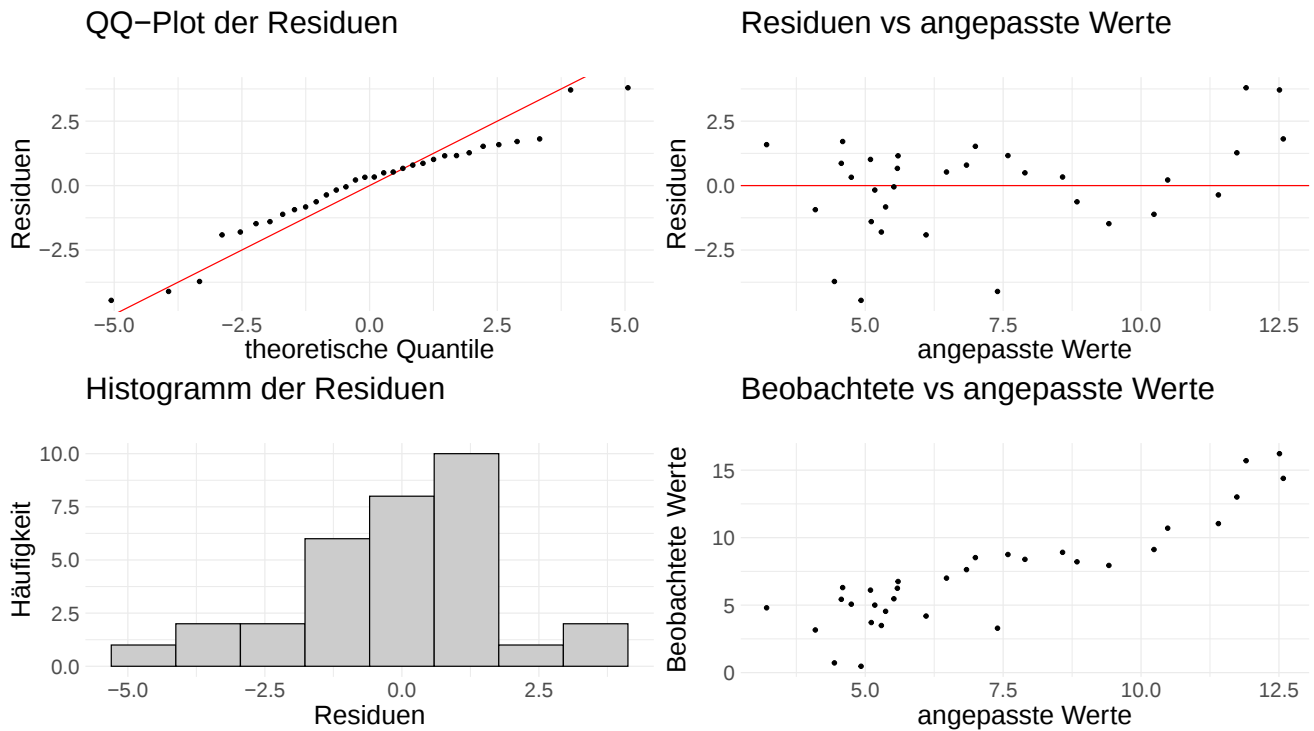


Abbildung A.32.: Diagnostische Plots für das Modell für die Stahlqualität R350. Die Plots bestätigen, dass man dieses Modell nicht verwenden sollte.

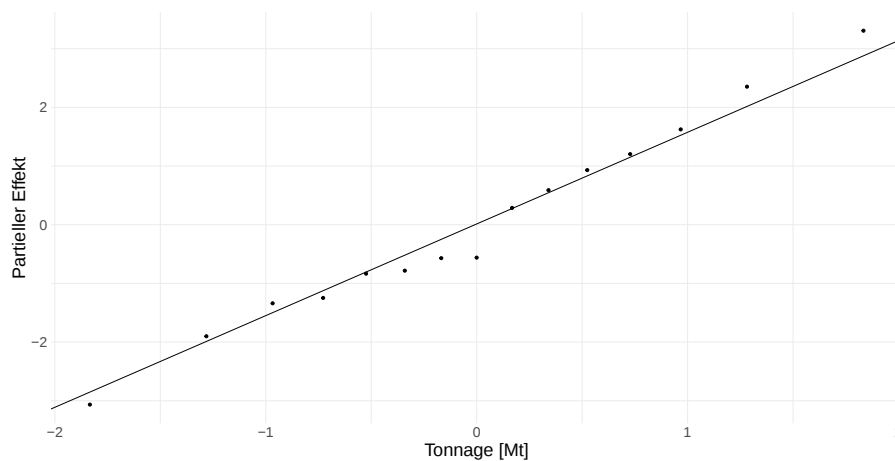


Abbildung A.33.: QQ-Plot für den zufälligen Effekt Ort für das Modell für die Stahlqualität R350.

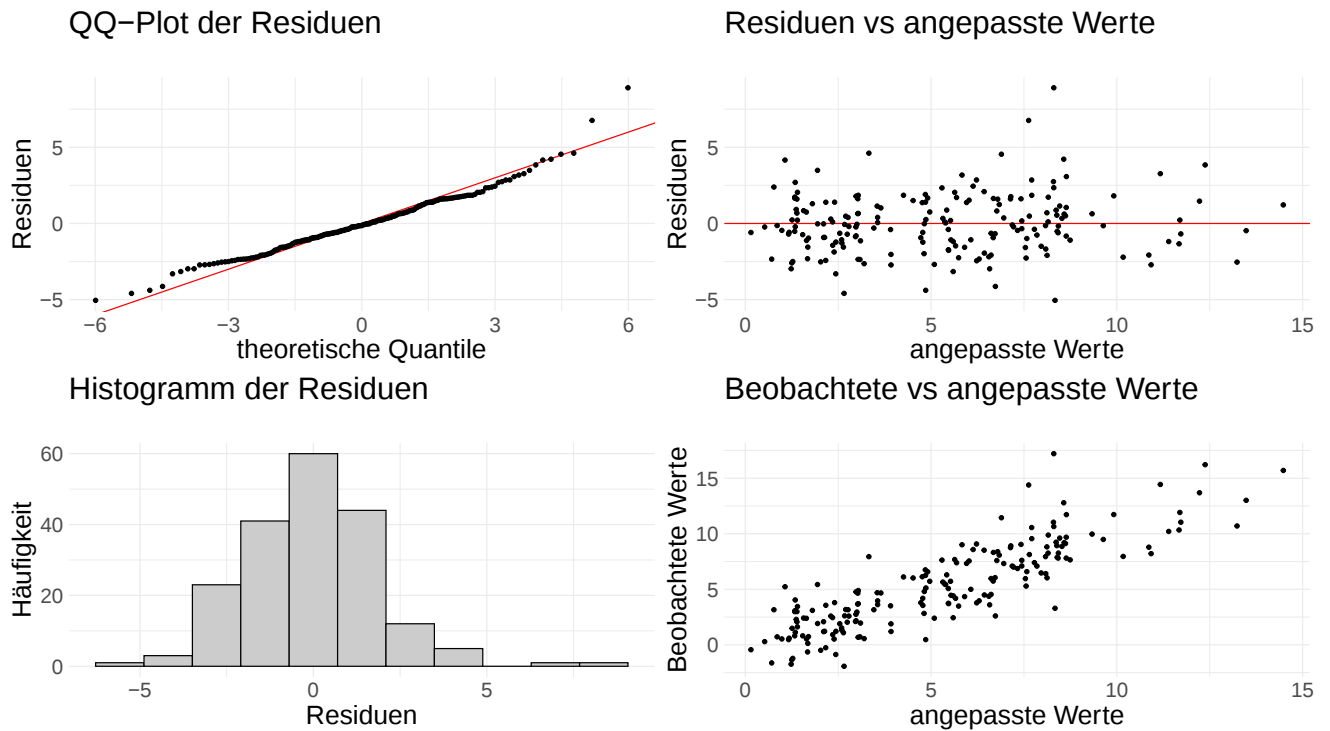


Abbildung A.34.: Diagnostische Plots für das Modell für beide Stahlqualitäten.

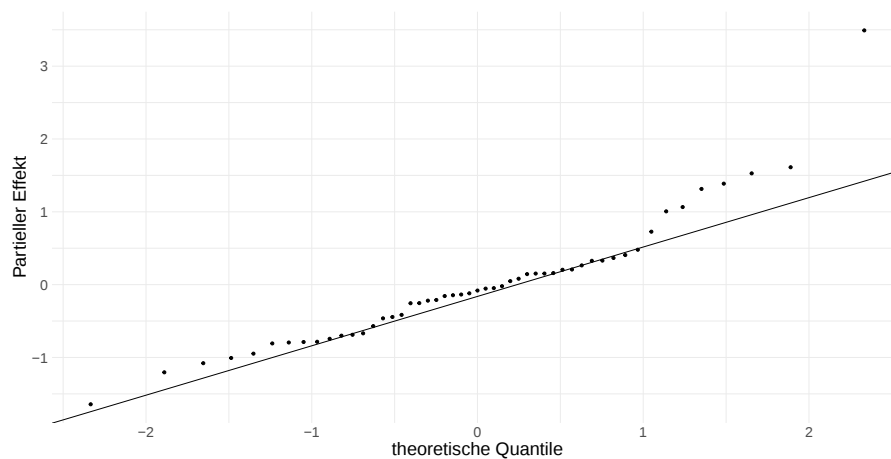


Abbildung A.35.: QQ-Plot für den zufälligen Effekt Ort für das Modell für beide Stahlqualitäten.

Literaturverzeichnis

- [1] T. Hastie, R. Tibshirani, 1986, *Generalized additive models (with discussion)*, Statistical Science 1, 297-318.
- [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] F. Mauz, R. Wigger, T. Wahl, M. Kuffa, K. Wegener, 2022 *Acoustic roughness measurement of railway tracks: implementation of an optical measurement approach & possible improvements to the standard*, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2022, ISSN 2041-3017, DOI <https://doi.org/10.1177%2F09544097221086487>
- [5] 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>
- [6] G. Simpson, 2022, *gratia: Graceful ggplot-Based Graphics and Other Functions for GAMs Fitted using mgcv*. R package version 0.7.3. <https://gavinsimpson.github.io/gratia/>.
- [7] S.N. Wood, 2004, *Stable and efficient multiple smoothing parameter estimation for generalized additive models.*, Journal of the American Statistical Association. 99:673-686.
- [8] S.N. Wood, 2011, *Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models.*, Journal of the Royal Statistical Society (B) 73(1):3-36
- [9] S.N. Wood, 2017, *Generalized Additive Models: An Introduction with R (2nd edition)*, Chapman and Hall/CRC, Boca Raton.
- [10] 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*.
- [11] Europäisches Komitee für Normung, EN ISO 15610:2009: *Bahnanwendungen – Geräuschemission – Messung der Schienenrauheit im Hinblick auf die Entstehung von Rollgeräusch*.
- [12] 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

- [13] R Core Team, 2022, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.