Doctoral Thesis

Noise prediction of trains in curves

submitted in satisfaction of the requirements for the degree of
Doctor of Science
of the Vienna University of Technology, Faculty of Civil Engineering

---

Dissertation

Prognose der Schallemission von Zügen in der Bogenfahrt

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaften
eingereicht an der Technischen Universität Wien, Fakultät für Bauingenieurwesen
von

Dipl.-Ing. Michael Ostermann BSc

Betreuer: Univ.Prof.i.R. Dipl.-Ing. Dr.techn. Norbert Ostermann
Institut für Verkehrswissenschaften, TU Wien

Gutachter: Ao. Univ.Prof. Dipl.-Ing. Dr.techn. Christian Kirisits
Universitätsklinik für Strahlentherapie, MedUni Wien

Gutachterin: Prof. Dr.-Ing. Corinna Salander
Institut für Maschinenelemente, Universität Stuttgart

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material, which has been quoted either literally or by content from the used sources. The work complies with the graded one from the reviewers.

Wien, February 23, 2021 _____

Acknowledgement

First of all, I want to thank Univ.Prof.i.R. Dipl.-Ing. Dr.techn. Norbert Ostermann for supervising my thesis as well as the possibility to elaborate it during my employment as university assistant in the research centre of railway engineering, traffic economics and ropeways. In particular, he took always time for open-minded discussions, provided necessary contacts and gave helpful input out of his expert knowledge.

Thanks goes also to ao.Univ.Prof. Dipl.-Ing. Dr.techn. Christian Kirisits, who delivered important suggestions and ideas for the thesis as well as accepted to act as a reviewer of the thesis.

Moreover, I want to thank Prof. Dr.-Ing. Corinna Salander for her decision to review the thesis as well as for the kind and interesting discussions.

A big thanks goes to my colleagues, who supported my ongoing work and enabled to finish it during my employment by covering parts of my lecture duties especially in the last year.

Special thanks goes to Dipl.-Ing. Dr.techn. Thomas Maly, who introduced me to the topic of railway noise. I worked with him on several research projects and I especially acknowledge his ongoing motivation to examine every last detail of a topic, which partly transmitted to my working behaviour. He was always available for discussions and provided important input. Lastly, he proofread the thesis and gave valuable feedback for further improvements.

I want to thank the Austrian Federal Railways (ÖBB) and the Austrian Research Promotion Agency (FFG) for funding both research projects, in which the used measurement data were collected. Special thanks goes to Dr. Günter Dinhobl for his ongoing trust, interest and support with expertise and data from ÖBB. Furthermore, I want to thank Bernhard Kneidinger, Stefan Ojak and Florian Biebl for their helpful support.

I am thankful to my family and friends for supporting me throughout the elaboration.

Finally, the support of my love, Leonie, was overwhelming. She stood behind me from the beginning and was always understanding as well as gave me the strength to finish the thesis. She accepted my extended working hours especially at the end and helped me through lacks of motivation and difficult times. I thank her from all my heart for her unconditional love.

Abstract

People become more and more aware of noise and noise regulation thresholds decrease to satisfy the needs of a quiet environment, while also preventing health issues due to noise pollution. Avoidance of noise as a limiting factor in train operation results in a huge investment regarding noise mitigation measures, which raises building and maintenance costs significantly. Thus, noise prediction to plan mitigation measures adequately and cost-efficiently is going to become essential. Although standards for calculating the noise emission of trains were developed in some countries in Europe, the elevated noise emission in curves is either considered with a single additive factor dependent on the radius or not considered at all. Hence, neither higher rolling noise in curves nor curve squeal, which states one of the most disturbing noise sources emitted from railways, is sufficiently accounted in current noise emission models. The present thesis investigates influence factors on frequency of occurrence, peak levels and relative occurrence time of three distinguished curve squeal types as well as on general noise emission in curves. Examinations are carried out empirically and statistically using sound measurements of train pass bys collected in five different sections in the Austrian heavy rail network with radii between 226 m and 440 m and speed limits between 60 kmh^{-1} and 90 kmh^{-1}. After data filtering, 29097 train pass bys are considered in the final evaluation. An empirically elaborated algorithm detects time increments containing curve squeal (distinguished in all three stated types) out of short-time averaged (64 ms) third-octave band time spectra. Influence examination considers a total of 190 parameters, which are grouped in five categories – track, vehicle, dynamic, general and environmental. Main findings on frequency of occurrence are a high impact of relative rail humidity with peaking probabilities between 72 % and 75 % for all three types and decreasing rates in both directions in the remaining area, partly distinct train type dependencies in probability magnitude, positive correlation of radial steering index (RSI) and lateral acceleration and significantly decreased relative frequency of occurrence regarding rain, dew and frost presence. Building on formulated influence factors, classification models to predict occurrence of each curve squeal type with a general scope on practical usefulness (including a low quantity of parameters) and general applicability are developed. For the latter, model training is carried out on data points of two sections, while the remaining three are used to validate predictive behaviour on different sections with deviating properties, train type distribution and climatic conditions. Results are curve squeal detection rates of 68 % to 84 % in conjunction with an overall classification accuracy of 55 % to 68 % among the whole validation set. The final result is a regression model for noise emission prediction in curves. Since predictive performance can be significantly improved by inclusion of categorical predictors for occurrence of curve squeal, previously formulated classification models have to be executed prior to estimation of overall noise emission. With inclusion of predicted curve squeal occurrence, benchmarks of 4 dB RMSE and 57 % R^2 are achieved on the whole validation set. Application on the validation sets leads to deviations in energetic mean between prediction and originally measured levels from -1.1 dB to -0.5 dB – depending on the section. Building on that promising result, further validation on other datasets is recommended.

Kurzfassung

Um die zunehmende Lärmsensitivität der Bevölkerung zu berücksichtigen, werden Grenzwerte in Gesetzen und Verordnungen immer strenger. Zur Verhinderung, dass Lärm ein limitierender Faktor im Eisenbahnbetrieb wird, wird in hohem Ausmaß in Lärmschutzmaßnahmen investiert, die Bau- und Erhaltungskosten erhöhen. Zur Planung dieser besteht ein Bedarf an umfassenden Prognosemodellen. Trotz der Entwicklung solcher in der Vergangenheit, wird in diesen die Emission in der Bogenfahrt nicht ausreichend behandelt. Die Berücksichtigung erfolgt – wenn überhaupt – durch einen einzelnen additiven Zuschlag zur Gesamtemission. Die gegenständliche Arbeit beschäftigt sich mit einer umfassenden Untersuchung von Einflussfaktoren sowohl auf die Auftrittswahrscheinlichkeit, den Spitzenwert und die relative Auftrittszeit von drei unterschiedenen hochfrequenten Kurvengeräuschen, als auch auf die allgemeine Schallemission in der Bogenfahrt. Die Untersuchung umfasst empirische und statistische Auswertungen, welche auf Schallmessungen von fünf verschiedenen Messquerschnitten im österreichischen Vollbahnnetz mit Bogenradien zwischen 226 m und 440 m und zulässigen Geschwindigkeiten zwischen 60 kmh^{-1} und 90 kmh^{-1} basieren. Nach der Datenfilterung verbleiben insgesamt 29097 Zugsvorbeifahrten in der Auswertung. Ein automatisierter Algorithmus wertet Terzpegel-Zeitverläufe hinsichtlich dem Auftreten von erhöhten hochfrequenten Anteilen aus. Die durchgeführte Untersuchung beinhaltet insgesamt 190 Parameter. Tendenzen hinsichtlich Auftrittshäufigkeit zeigen relative Luftfeuchte der Schiene mit einem Maximum zwischen 72 % und 75 % und einem abnehmenden Verlauf in beide Richtungen, einzelne Zugstypen, positive Korrelationen mit dem Radialstellungsindex (RSI) und freier Seitenbeschleunigung, sowie stark verringerte Werte in Zusammenhang mit Regen, Tau und Frost. Im nächsten Schritt werden, aufbauend auf den formulierten Haupteinflussfaktoren, Klassifizierungsmodelle für die Auftrittsvorhersage von hochfrequenten Kurvengeräuschen entwickelt. Der Fokus liegt dabei auf allgemeiner Praxisanwendbarkeit (möglichst wenig Einflussfaktoren). Zur Einschätzung der Vorhersagequalität auf neuen Datensätzen wird das Modell auf Datenpunkten von zwei Messquerschnitten trainiert und in den verbleibenden drei validiert. Somit wird das Modell auf Datensätze mit abweichenden Streckeneigenschaften, Zugmix und klimatischen Bedingungen angewandt. Auf diesem Validierungsset wird eine Erkennungsrate von hochfrequenten Kurvengeräuschen von 68 % bis 84 % in Verbindung mit einer allgemeinen Klassifizierungsgenauigkeit von 55 % bis 68 % erreicht. Das finale Endergebnis stellt ein Regressionsmodell zur Vorhersage der Schallemission in der Bogenfahrt dar. Aufgrund der wesentlich verbesserten Prognosegenauigkeit, werden kategoriale Variablen, die das Auftreten von jeweils einem Typ hochfrequenter Kurvengeräusche beinhalten, in das Modell inkludiert. Somit ist die Anwendung der entwickelten Klassifizierungsmodelle vor der Schallemissionsprognose durchzuführen. Prognose des Validierungsset mit inkludiertem prognostizierten Auftritt von hochfrequenten Kurvengeräuschen resultiert in Kennwerten von 4 dB RMSE und 57 % R^2. Vergleich zwischen energetischen Mittelwerten gemessener und prognostizierter Emissionswerte in den einzelnen Validierungsquerschnitten über die gesamte Messdauer liefert Abweichungen zwischen -1.1 dB und -0.5 dB. Zur vertiefenden Validierung wird eine Anwendung auf weitere Datensätze empfohlen.

Contents

1 Introduction

In times of climate change and the aim of reducing greenhouse gas emissions, railways form the backbone of a future CO_2 neutral transport of both people and goods. In regulation (EU) No 1315/2013 [1], a trans-European transport network is set and billions of Euros are invested to build it. The mix of new high speed and conventional routes as well as improved existing tracks aim to increase the capacity and consequentially the number of train drives. In contrast to that, people become more and more aware of noise and noise regulation thresholds decrease to satisfy the needs of a quiet environment, while also preventing health issues due to noise pollution.

The World Health Organization (WHO) carried out a huge amount of studies in the last decades to research the impact of noise on human health and well-being. In the latest guideline from the WHO Regional Office for Europe [2] railway noise was associated mainly with annoyance and sleep disturbance. To limit effects on health, the Guideline Developer Group (GDG) recommends for the former to reduce the noise level below 54 dB L_{den} (day-evening-night-weighted sound pressure level) and for the latter to reduce the noise level below 44 dB L_{night} (equivalent continuous sound pressure level in the night period). However, these values are extremely low and difficult to achieve even with the number of train drives nowadays.

Considering all ongoing developments, noise can be a limiting factor for railway operation. In addition, demands for limits on capacity and resistance to expansion of the network are growing [3]. These trends are controversial to the objectives described in the first paragraph. The result is a huge investment in noise mitigation measures, which raises the building and maintenance costs significantly.

To harmonize the calculation methods of environmental noise in the European region, the European Parliament and the Council of the European Union elaborated the Commission Directive (EU) 2015/996 [4], which contains a full noise emission and propagation model for strategic noise mapping (called Common Noise Assessment Methods in Europe or short CNOSSOS-EU). The member states are obligated to establish strategic noise maps every five years starting in 2012. 2022 is the first time, where the harmonized model has to be used. However, it should be borne in mind that noise mapping is noise prediction on a higher scale with mostly approximated averaged values. In the future detailed noise prediction is going to become essential to plan mitigation measures adequately. Noise measurements are not sufficient to cover the whole rail network, however they have to deliver the input parameters for an universal and practice-oriented prediction model. This thesis deals with an approach of a noise emission model in curved sections with special focus on inclusion of curve squeal, which states one of the most disturbing noise sources emitted from railways [5].

1.1 Problem

Railway noise emission is dominated by rolling noise in a wide range of common velocities (roughly between 30 kmh^{-1} and 200 kmh^{-1} [6]). The origins of rolling noise are vertical force oscillations due to wheel and rail roughness (RR) leading to vibrations of wheel and rail. At

low speed, traction noise becomes relevant, whereas at higher speeds aerodynamic noise is prevalent [7]. In addition, many other noise sources can occur in train operation like warning signals, impact noise from shunting processes, switches, crossings or joints and so on. In curves, three effects are able to cause an elevated noise emission [8]: (i) the possibility of multiple contact points between rail and wheel; (ii) increased rail corrugation; (iii) occurrence of curve squeal noise. (i) and (ii) lead to higher rolling noise than on tangent track and (iii) introduces a phenomenon that radiates high frequency noise, which causes severe annoyance of line-side inhabitants.

Although standards for calculating the noise emission of trains are developed in some countries in Europe, the elevated noise emission in curves is either considered with a single additive factor dependent on the radius or not considered at all. If there is a factor, it is independent of train categories. A research project shows that the elevation of the emission in curves of trains with a basically high rolling noise (e.g. freight trains containing cast-iron block braked coaches) is negligibly small [9]. Thus, neither higher rolling noise in curves due to (i) and (ii), nor (iii) is sufficiently accounted in noise emission models.

1.2 Motivation and scope of the thesis

Building on the stated problem, the thesis focuses in a first step on examination of influence parameters regarding frequency of occurrence in terms of different curve squeal noise types, their peak amplitudes and their relative occurrence time. Moreover, dominant influence factors on general sound emission in curves are investigated. Finally, the scope lies in developing prediction models for curve squeal occurrence and sound emission in curves – in the standardized distance of 7.5 m from the track axis and 1.2 m above rail level according to ISO 3095 [10] – with the main objectives of practical usefulness (including a low quantity of parameters) and general applicability.

Thus, two main research goals are defined:

- Definition of dominant influence factors on frequency of occurrence, peak levels and relative time occurrence of curve squeal as well as on noise emission in curves
- Development of a practical prediction model for noise emission in curves with focus on general applicability

1.3 Methodology

Investigation is carried out empirically and statistically using sound measurement data (monitored in standardized distance according to ISO 3095 [10]) collected at five different curves in the Austrian heavy rail network between 2013 and 2017. Curve radii lie between 226 m and 440 m and speed limits are between 60 kmh^{-1} and 90 kmh^{-1}. In addition to the sound measurement, two axle counter sensors were used to monitor axle speeds and identify the train type by evaluating the axle pattern automatically. Moreover, a third sensor was used on the other track (always two-track lines) to detect presence of parallel train pass bys, which leads to exclusion of the data point in the evaluation. Track decay rates (TDR) and RR were

measured at least once in each section in the measurement period. Humidity, rail and air temperature, wind speed and direction, rainfall and air pressure was monitored by a mobile weather station at the measurement section during the measurements. A rail temperature sensor was located at the rail base between two sleepers. To complete and verify the dataset of environmental conditions, additional data was thankfully provided by the Central Institution for Meteorology and Geodynamics (ZAMG) [11] for the measurement periods. Data filtering is done on the one hand empirically with time line plots to detect obviously compromised data points and on the other hand by defining limits for braking/accelerating (≤ 0.2 ms^{-2}), speed gain/loss (≤ 10 kmh^{-1}) and wind speed (≤ 5 ms^{-1}). All data points above stated values are excluded. In addition, train pass bys against the standard driving direction are dropped. Data points, where no full measurement of the whole train length (buffer to buffer) as well as a friction control system is applied onto the rail head are also not considered. After the filtering step, 29097 train pass bys are left for further evaluation.

The term "curve squeal" may include several different phenomena, therefore a clear definition throughout the thesis is needed:

- flanging noise: some kind of shear noise, more broad-band, higher fundamental frequencies compared to squeal noise, lower levels in general (compared to squeal), more intermittent and associated with contact between wheel flange and rail gauge face [5]
- squeal noise: strongly tonal, mid to high frequency emission corresponding to natural frequencies of the wheel and caused by lateral stick-slip oscillations of the wheel on the rail tread [5]
- high frequency (HF) squeal noise: tonal, with higher fundamental frequencies and slower time behaviour compared to squeal noise, excitation mechanism not entirely clear [9]
- curve squeal: contains all three stated types – flanging, squeal and HF squeal noise

An empirically elaborated algorithm (developed in a research project [9]) detects time increments containing curve squeal (distinguished in all three stated types) out of short-time averaged (64 ms) third-octave band time spectra. It is used to detect time increments containing curve squeal, which allows a declaration on curve squeal occurrence, the corresponding peak magnitude (short-time averaged peak third-octave band level) and relative occurrence time of the event (further explained in Chapter 4). Influence examination considers a total of 190 parameters, which are grouped in five categories – track, vehicle, dynamic, general and environmental. They are calculated out of measurement data, contain directly measured values or include train type or section specific parameters. Analyses are carried out on ten different dependent variables – frequency of occurrence, peak levels and relative occurrence time of all three distinguished curve squeal types as well as on equivalent continuous noise level (energetic mean among all time increments from buffer to buffer summed up between 50 Hz and 10000 Hz third-octave bands). The latter range is chosen on behalf of considered third-octave bands in the CNOSSOS model [4]. Building on formulated parameters with the most impact on the particular variables, prediction models are built by application of several

statistical regression and classification algorithms. The final choice regarding performance is made with Root Mean Squared Error (RMSE) in case of regression and Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) for classification. Detailed methodology on carried out analyses and predictive modelling is provided in Subsection 6.2.

1.4 Limitations

Due to lack of measurement data, it is only possible to investigate trends for heavy rail. Light rail vehicles show different dynamic behaviour (e.g. lower normal load) and routes might include embedded tracks. To apply the model on light rail tasks, further research and at least model validation are needed. Psycho acoustic parameters and influences are excluded in the analyses. Other noise sources like traction, aerodynamic and impact noise are not separately considered due to lack of measurement data. However, they are included in monitored noise emission, respectively. Furthermore, noise propagation models, noise indicators (see [12]) and immission levels are not discussed in detail as the whole investigation is focused on emission. Further limitations for practical model application are stated in Subsection 6.6.

1.5 Outline

The thesis is divided into the following chapters.

Chapter 2 contains a literature review about curve noise in general, the mechanism behind curve squeal, past research on influence factors and prediction approaches. In addition, basics regarding meteorology and friction are explained.

In Chapter 3, properties of measurement sections and monitored train types are described as well as their comparability is discussed. Moreover, exclusion criteria for measurement data are defined and fundamental statistics among the dataset are provided.

Chapter 4 illustrates the functionality of the algorithm to detect all three distinguished types of curve squeal in short-time averaged third-octave band time increments.

Chapter 5 depicts the derivation of included input parameters and dependent variables as well as includes dew formation modelling on the rail head with a regression approach to predict rail temperature. Furthermore, trends of dependent variables among the dataset distinguished in sections and train types are illustrated and discussed.

Chapter 6 states the main part of the thesis – starting with an introduction on statistical modelling and detailed description of applied methodology, formulation of most important influence parameters on frequency of occurrence of all three curve squeal types and equivalent continuous noise level is carried out. Afterwards, prediction models for the stated tasks are developed and described in detail. The chapter is concluded by a summary of most important findings as well as limitations in practical application of developed models.

Chapter 7 provides an overall summary and Chapter 8 states future research needs.

Appendix A contains calculation steps of the applied dew formation approach, Appendix B illustrates basic statistic plots to shrunk the main part, Appendix C contains impact analyses on peak levels and relative occurrence time of all three curve squeal types and Appendix D

provides coefficients to derive needed principal components in the prediction model for equivalent continuous noise level.

2 Literature review

In the following subsections, basics regarding the coefficient of friction, curve squeal and meteorology are described. The scope is to give an introduction of fundamental mechanisms and an overview of past research work.

2.1 Coefficient of friction

The importance of friction between rail and wheel in regard to curve squeal is high and the main influence factors besides contaminants are environmental conditions. This chapter provides basic knowledge about adhesion, friction and past findings of environmental influence.

While adhesion states the force needed to separate two surfaces, which are in contact with each other, friction is accounted as resistance of one body moving over another. The friction coefficient is defined as the ratio between friction force and normal load in pure sliding motion. The adhesion coefficient is determined as the ratio of adhesion force and normal load in a rolling motion. Under pure sliding conditions, the adhesion and friction force are equal (see Figure 3 at the saturation point). [13]

The adhesion between wheel and rail depends on the so-called "third body" layer at the skin of the wheel and the rail. The thickness of the layer analysed from field tests is 15 µm, however can range from few micrometers to several dozen micrometers (another study finds e.g. 50 µm [14]). The layer characteristics depend on the rheological properties, which are complex due to a varying combination of particles from the wheel and rail as well as solid and fluid contaminants from climatic conditions – dead leaves, frost, water – and from operation – pieces of ballast, sand and oil. The fluid components are absorbed or adsorbed and can change the frictional conditions significantly. The third body layer is permanently in motion due to the high normal and shear stresses of the wheel-rail interaction and provides a protection of the rail and wheel bulk material from severe wear. [15]

Metals oxide in air and form an oxide layer (thickness between 1 nm and 10 nm) on the surface after a few minutes of exposure. Thus, corrosion products are also part of the third body layer. Dependent on the normal load, the oxide layer can prevent bare metal to metal contact and provide almost always a lower coefficient of friction. For a further tribological study, be referred to [16]. Iron oxides consist chemically of iron and oxygen. It can be distinguished between anhydrous and hydrated iron oxides (e.g. rust). The properties of iron oxides in terms of tribology are described in [17]. Another study points out that the thickness of the oxide layer is inversely proportional to the friction coefficient [18].

Beagley et al. [19–21] investigate the influence of humidity, oil and rail head debris on the friction coefficient. Small particles (mostly corrosion products) are formed from wear and during wet conditions on the rail head. The quantity of debris differs widely and can form substantial layers, which causes that no metal to metal contact takes place. On dry rails the debris quantity has no influence on the friction coefficient. The particles adsorb fluids (mostly oil and water are relevant in regular operation). While the oil quantity for

modifying the properties is rarely reached during regular operation, rail humidity[1], rain and dew can have a significant impact. The friction coefficient can be directly expressed as a function of rail humidity. With increasing rail humidity, the friction coefficient decreases. However, the impact of humidity is more noticeable if the debris quantity is high. Slight rain and dew reduce the friction coefficient significantly by building a viscose paste on the rail head, which is stable after wheel crossings, and also generate fresh debris. Heavy rain cleans the surface and the debris layer is washed away or removed by the wheel, which causes a higher friction coefficient compared to slight rain. A similar relation is found in laboratory tests with different methods by Buckley-Johnstone [23]. Application of low amounts of water (simulation of equivalent rainfall of about 2 mmh^{-1}) lead to a very low friction coefficient, whereas a considerably higher amount (simulation of heavy rain) or just a bulk of water cause a higher friction coefficient, however still lower than under dry conditions.

Lewis et al. [24] point out that a significant adhesion loss is associated with the dew point. It is hypothesized that the reason is condensed water on the rail head.

Chen et al. [25] evaluate on a twin-disc rolling contact machine an increase in the adhesion coefficient with raising water temperature under wet conditions.

Olofsson and Sundvall [26] carry out pin-on-disc experiments with a combination of two different humidities (40 % and 95 %) and temperatures (5 °C and 20 °C). The measured coefficient of friction is significantly higher at low humidity. The temperature has a negligible influence on the coefficient of friction.

Liew [27] reports a high and nearly constant coefficient of friction between 28 % and 50 % humidity and a lower and again constant one between 60 % and 80 %. The significant drop and lower friction at high humidity is explained with the formation of protecting layers on the surface consisting of iron hydroxides, ferri-oxide-hydrates, adsorption of water and normal atmospheric oxidation. A transition in the friction coefficient between 35 % and 63 % humidity is also observed by Hayashi et al. [28].

Zhu et al. [29] find in pin-on-disc tests a nearly linear decrease of the coefficient of friction until the humidity value reaches 65 %. At higher humidity the friction coefficient remains constant. Variations in ambient temperature result in a significantly higher friction value at middle humidity (45 %), followed by a steeper declination and similar values for higher humidity. Moreover, a correlation of friction coefficient and absolute humidity is shown – a decrease in friction takes place with raising absolute humidity. The slope is steeper at low temperatures indicating a sensible reaction of the coefficient of friction with slight changes in absolute humidity (and relative humidity respectively) at low temperatures. For the stabilized friction coefficient above 65 %, a water film development, which is capable of keeping iron oxides (predominately haematite α-Fe_2O_3 on the surface at high humidity levels, is made responsible. Haematite can increase friction and thus counteracts the declination of the friction coefficient – causing a constant value at humidity levels above the transition point. An overview of the impact of iron oxides on wheel-rail interaction can be found in [30]. Another study by Zhu et

[1]The term "rail humidity" states the ratio of partial water vapour pressure in the air and the saturated water vapour pressure calculated with the rail temperature. If the rail humidity reaches 100 %, water condenses on the rail surface [22].

al. [31] with the same method, shows the influence of temperature and humidity with friction maps using one clean and one pre-treated oxidized sample. In general, a declination of the friction coefficient with increasing humidity until 70 % humidity can be found. Above, the friction coefficient stabilizes and remains constant. A temperature dependent influence of the oxide layer is evident at low temperatures, where the clean sample shows a higher coefficient of friction at low humidity values, followed by a steeper declination. The tendency applies also if the absolute humidity is considered.

Apart from environmental influence, general tendencies of operational and vehicle dependent impacts on the friction coefficient can be summarized:

- with increasing speed, the friction coefficient decreases (e.g. [25, 32])
- with raising normal load, the friction coefficient decreases (e.g. [23]), reduced influence in wet conditions [32]
- with increasing surface roughness (wheel and rail), the friction coefficient raises (e.g. [25])

The most research in terms of wheel/rail friction is focused on traction problems (acceleration and braking) and hence discusses the longitudinal creepage. For squeal noise, the function of lateral creepage and adhesion coefficient is mainly of interest. Thus, it is unclear if the described impacts also apply on the relations in the lateral direction. In the following subsection, the basic mechanisms and influence factors in regard to curve squeal stated in the literature are explained. Moreover, the past findings of correlations with environmental conditions and exemplary consideration approaches in standards are discussed.

2.2 Curve squeal

A coned wheelset is able to traverse a curve with pure rolling motion, if it aligns itself radially and moves outwards to compensate the differences in travelling distance with the rolling radius difference. In case of a bogie (or a two-axle vehicle) two wheelsets are incorporated in a rigid frame leading to limitations in curves. The design of the yaw stiffness is always a compromise between stability and guidance. A high yaw stiffness leads to stability at high speeds, while the curving behaviour is worsened by preventing the wheelsets from fully radial alignment. [33] Four standard cases for a curving bogie can be distinguished: (a) both outer wheels are in flange contact, happening with a high free centrifugal force; (b) free curving occurs, only the outer wheel of the front wheelset is led by flange contact; (c) constraint curving takes place, the outer wheel of the leading wheelset and the inner wheel of the trailing wheelset are in flange contact indicating slow speed in a sharp curve; (d) both inner wheels are in flange contact, happening if a rail car is pulled slowly through a sharp curve. [34]

Due to a non-zero angle of attack (yaw angle between the wheel and ideally radial alignment), the lateral creepage value, which can be obtained by normalizing the lateral slip velocity by the rolling velocity, is also non-zero (see Figure 1). With the same approach, longitudinal and spin (rotational relative movement) creepage can be obtained, which also exists during curve negotiation. The angle of attack is usually highest at the leading wheelset and at sharp

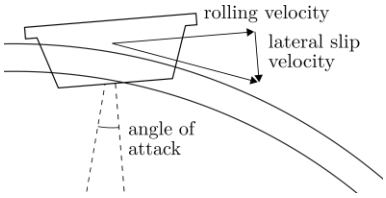

Figure 1: Generation of lateral creepage due to the present angle of attack [35][2]

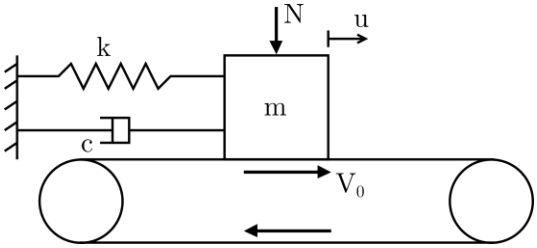

Figure 2: Simple mechanical model to illustrate stick-slip motions [35][2]

curves in combination with low speed. While flange contact may lead to flanging noise, the lateral creepage and the angle of attack are important factors in terms of squeal noise. [5]

Squeal noise originates from stick-slip motions between wheel and rail, which causes self-excited vibrations of a railway wheel. This can be illustrated by a simple mechanical model (see Figure 2). A mass m (wheel) restrained by a damper and a spring sticks on a moving belt (rail head) until the spring force ku exceeds the static friction force $\mu_{\mathrm{s}}N$ with μ_{s} as static friction coefficient. Afterwards, the mass slips on the belt, undergoes harmonic motion and the friction force changes to $\mu_{\mathrm{d}}N$ with μ_{d} as dynamic friction coefficient ($\mu_s > \mu_d$). The oscillation goes on until the dynamic friction force exceeds ku and the mass sticks again. Thus, stick and slip phases take place periodically. [35]

The contact patch between rail and wheel contains a slip and an adhesion region. As the creep force linearly increases at low creepage values, the slip region grows, while the adhesion region is reduced. When the creep force – creepage curve reaches a certain value, "saturation" takes place (refer to Figure 3). In this state, slip occurs over the whole contact area between wheel and rail. Depending on the materials in the third body region between wheel and rail (e.g. moisture or lubrication), the creep force can increase, decrease or remain constant in the full slip phase (schematically pictured in Figure 3) [36]. It has to be pointed out that both the amplitude of the creep force and the lateral creepage value, where saturation occurs, depend largely on friction conditions (e.g. [37]). The value of lateral creepage is approximately identical with the angle of attack in radians [5]. Creepage behaviour can be calculated for example with Kalker's model [38] using the approximate method FASTSIM or the more time-consuming CONTACT algorithm.

Rudd [39] describes three possible mechanisms, where squeal noise might be generated:

- longitudinal slip between inner and outer wheel

[2]Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Noise and Vibration Mitigation for Rail Transportation Systems by D. Anderson, P.-E. Gautier, M. Iida, J. T. Nelson, D. J. Thompson, T. Tielkes, D. A. Towers, P. de Vos, J. C. O Nielsen; © Springer International Publishing AG (2018)

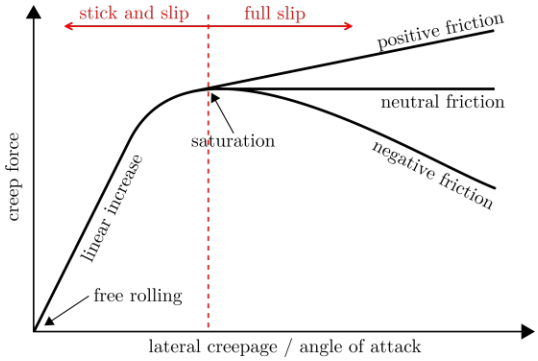

Figure 3: Schematic picture of creep force – creepage relationship

- wheel flange rubbing on the rail gauge face
- lateral slip between wheel tread and rail head

Longitudinal slip occurs if the conicity of the wheels is not able to compensate the difference in travel distance through the curve. Although it is identified as possible mechanism for enforced corrugation growth with oscillations in a frequency range of 50 Hz - 100 Hz, it is not thought to be responsible for squeal noise. Wheel flange contact with the rail gauge face is found to actually reduce the likelihood and the level of squeal noise. However, it may generate flanging noise. Lateral stick-slip motions are considered to be the main cause of squeal noise [5]. In a bogie or two-axle vehicle squeal noise is usually generated by the leading inner wheel, which is in most cases not in flange contact and has a large angle of attack. However, it is observed that squeal noise can also be radiated by the outer wheel, which is in flange contact [40–43].

Squeal noise is characterised as a pure tonal noise with extremely high sound pressure levels of up to 130 dB (at 0.9 m distance from the wheel [39]), which fundamental frequency corresponds to natural frequencies of axial wheel modes with zero nodal circles [44, 45]. In addition, noticeable frequencies of higher harmonics are often observed, which indicates the non-linear behaviour of squeal [41]. The wheel vibration dominates the sound radiation compared to the rail [44]. The frequency range is stated by Thompson [5] 250 Hz - 5000 Hz, whereas Vincent et al. [44] mention a much wider range of 400 Hz - 10000 Hz. Flanging noise is described as noise source with higher fundamental frequencies, lower levels, more intermittent and more broad-band compared to squeal noise [5].

In contrast to flanging noise, squeal noise was extensively researched in theory, test rigs and field measurements in the past decades. However, an overall conclusion of exact conditions leading to squeal and excitation mechanisms has not been achieved yet [7]. In the following subsections, possible excitation mechanisms and influence factors of curve squeal are discussed. As the thesis focuses on prediction and the impact of environmental conditions on squeal noise, past findings regarding these topics are outlined in more detail. The last subsection contains considerations of the elevation of the emission in curves in standards exemplarily.

2.2.1 Excitation mechanism

The two most discussed excitation mechanisms for squeal noise are the "falling friction" characteristic and "mode coupling". The former one is introduced by Rudd [39], which built the basis of many more detailed and improved squeal models. The latter phenomenon is explained e.g. by Hoffmann et al. [46]. Both are briefly discussed in the following paragraphs.

The "falling friction" mechanism is based on a declination of the creep force – creepage function with increasing creepage after the saturation region (outlined in Figure 3 as negative friction). In other words, the creep force decreases with increasing sliding velocity. This relation can be described as negative damping effect, which leads to permanent energy input and thus causes unstable self-excited vibration. The exponential growth of the amplitude is limited by non-linear effects in the creep force [47]. The presence of the declining friction is validated by several laboratory measurements [37, 48–52]. In contrast to that, Koch et al. [45] and Collette [53] observe no negative friction characteristic. Collette [53] states that vertical dynamics could be an important factor in squeal noise generation if neutral friction occurs.

Due to observations of squeal noise without negative friction, alternative mechanisms have to exist. Geometric coupling of different wheel modes rises as an important mechanism, which occurrence is possible with neutral friction. As responsible energy source, the wheel deformation pattern is identified [7]. In addition to lateral motions of the wheel on the rail head (see Figure 2), vertical variations of the contact force are taken into account (illustrated schematically in Figure 4). A mass m (wheel) is hold in place by two springs (k_1 and k_2). Another spring (k_H) represents the contact stiffness between mass and moving belt (rail head). Despite the displacement in x-direction on the moving belt, stick-slip oscillations occur [35]. The lateral oscillations are coupled with vertical oscillations (y-direction) due to variations in the normal load. Both vibrations depend on the friction force, which is alternated with normal load changes, and energy is constantly exchanged between them, which requires a phase shift. To illustrate the coupling further, both equations of motions are depicted [43]:

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = -2x - (1-\Delta)y \tag{1}$$

$$\frac{\mathrm{d}^2y}{\mathrm{d}t^2} = -x - 2y \tag{2}$$

In case of $\Delta < 1$, two coupled modes with different frequency and in perfect energy balance appear. At $\Delta = 1$, $y(t)$ becomes unstable because $x(t)$ feeds energy, but $x(t)$ is not coupled with $y(t)$ anymore. The amplitude of $y(t)$ is growing, while $x(t)$ remains constant. Values for $\Delta > 1$ cause unstable vibrations in both directions due to linear increases in amplitude. [43]

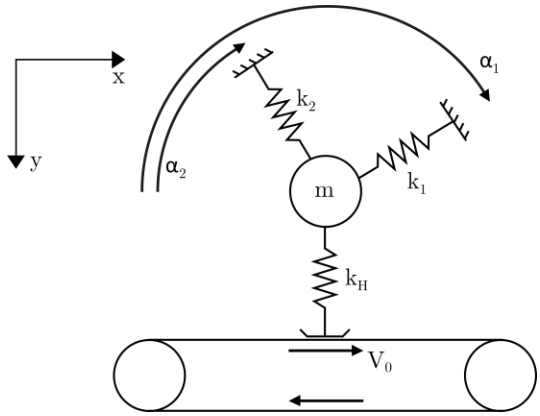

Figure 4: Simple mechanical model to illustrate the coupling of geometrical modes [35][3]

2.2.2 Influence factors

The range of influence parameters considered in relation to curve squeal is huge. An European research project categorizes parameters in track, vehicle and environment as follows [54]:

- vehicle:
  - bogie design (bogie stiffness, mass, axle flexibility)
  - wheel design/dynamics (wheel mobilities, axial vibration modes)
  - wheel damping (for axial vibration modes)
  - wheel base b (squeal for $R \cdot b^{-1} < 100$)
  - exact flange distance (can affect contact position)
  - wheel surface profile and wear (can affect contact position)
- speed (can affect contact position)
- vehicle/track:
  - vertical contact force and vertical dynamics (influenced by loading, unsprung mass, total mass, suspension properties, ...)
  - friction coefficient (depends on surface roughness, materials, surface moisture, dust, lubricants, ...)
- track:
  - curve radius R (strongest for $R < 200$ m)
  - rail surface profile and wear (can affect contact position)
  - track cant (can affect contact position in combination with speed)
  - track type, components and dynamics (including rail head mobilities)

[3]Reprinted by permission from Springer Nature Customer Service Centre GmbH: Springer Nature, Noise and Vibration Mitigation for Rail Transportation Systems by D. Anderson, P.-E. Gautier, M. Iida, J. T. Nelson, D. J. Thompson, T. Tielkes, D. A. Towers, P. de Vos, J. C. O Nielsen; © Springer International Publishing AG (2018)

  - exact rail gauge (can affect contact position)

- environment:
  - air humidity (can affect rail humidity)
  - rail humidity (linked to air humidity and rail temperature)
  - air temperature (can affect humidity)
  - wheel and rail temperature (can affect rail humidity)

Vincent et al. [44] identify three key parameters for squeal noise:

- local kinematic parameters at the wheel/rail contact
  - rolling speed
  - angle of attack
  - lateral position of the contact across the wheel tread
- contact friction law
- wheel modal parameters

Rolling speed determines the free centrifugal force and can influence the contact point as well as the curving behaviour. Although it can have an impact on squeal occurrence and amplitude [55], it is of secondary importance compared to lateral creepage [44].

Measurements for lateral creepage and angle of attack are mostly obtained by small scale test rigs. Remington [56] states that an angle of attack of at least 0.5 ° (8.7 mrad) is required for squeal to occur. Furthermore, the steady lateral creepage for the leading wheelset can be approximated as follows:

$$\frac{b}{R} > \gamma > \frac{b}{2R} \tag{3}$$

where b represents the bogie wheelbase, γ is the steady lateral creepage value and R states the curve radius. As the angle of attack in radians approximately complies with the lateral creepage value, the minimum for squeal occurrence, according to [56], is 0.87 %. De Beer et al. [57–59] evaluate on a 1:3 scale model a declination of the creepage – creep force relationship for angles of attack above 0.3 ° (lateral creepage 0.5 %) and squeal occurrence above 0.4 ° (lateral creepage 0.7 %). On the same rig, Monk-Steel et al. [49] measure the lateral friction curve with an implied longitudinal creepage of 2 % and find out that the slope changes sign above 1 ° (lateral creepage 1.7 %). Hsu et al. [50] find squeal occurrence at angle of attacks above 1.1 ° (lateral creepage 1.9 %). Jie et al. [60] observe squeal occurrence on a roller rig with a whole wheelset above 0.7 ° (lateral creepage 1.2 %). Matsumoto et al. [61] report a lateral creepage saturation point at 0.4 % and a slightly falling characteristic. As mentioned Koch et al. [45] and Collette [53] observe a squeal occurrence without negative friction on a 1:4 scale rig. Squeal is detected above 0.45 ° (lateral creepage of 0.8 %).

De Beer et al. [48] show on a 1:3 test rig that the lateral contact position has an influence on occurrence of squeal noise and the excited wheel mode. Squeal noise is only observed if the contact position of the leading inner wheel of a bogie is on the outside of the wheel tread (opposite to the flange).

Zenzerovic [7] carries out a parametric study to investigate the influence of variations in lateral creepage, friction coefficient and three lateral contact position between wheel and rail (on the outside of the wheel tread). The latter impacts the contact area size and shape as well as the coupling between vertical and lateral wheel dynamics, which determines which wheel mode is responsible for squeal noise and hence the radiated sound power level. A variation of 5 mm in lateral contact position causes a difference of up to 20 dB in sound power level. However, contact position does not affect the critical lateral creepage value, above which squeal occurs, significantly.

The maximum values of sound power levels are friction dependent. Friction can be interpreted as additional damping of the wheel/rail system. An increase in friction leads to a decrease in response frequency and hence to a shift in radiated squeal frequency. Furthermore, a strong dependency between friction and critical lateral creepage is observed. With increased friction, the critical creepage value also raises. It is concluded that wheel/rail kinematics and friction parameters have high importance in regard to occurrence and amplitude of squeal noise. [7]

With wheel modal parameters, the wheel response at the contact point can be fully evaluated due to the modal behaviour of the wheel. Since axial modes are thought to be mainly responsible for squeal noise, the damping conditions of those modes should be investigated firstly. [44]

2.2.3 Influence of environmental conditions

A probabilistic model to predict squeal noise only dependent on relative humidity is elaborated by Mehaan et al. [62]. The contact mechanics are similar to Rudd's model [39], where changes in the friction coefficient (due to relative humidity alterations) affect the system damping and hence the critical lateral creepage value (see Figure 3). As relationship between friction coefficient and relative humidity, the derived function of Beagley et al. [19] is linearly approximated. However, it has to be pointed out that the model uses the relative humidity of the air and not the rail humidity. Unlike most theoretical models, the outcome is validated with field measurements in Australia ($>$ 2 million wheel passes) and shows a very good correlation. In general, as the humidity increases, the probability of squeal also raises (about 2 % between 10 % and 100 % humidity). The model is further investigated by Liu and Mehaan [37] in a rolling contact two disc test rig, where the same trend is observed. The critical creepage value declines by 7 % with an increase in humidity from 50 % to 70 % and falls another 4 % with a raise from 70 % to 90 % humidity. The squeal probability is increased between 50 % and 90 % humidity by 6 %. However, no correlation between sound pressure level and humidity is derived. In contrast to that, investigations with field data show a probability increase for squeal noise until a peak value between 60 % and 70 % humidity followed by a

decrease at higher (air) humidity levels. Flanging noise shows the same relationship. Alteration in approximated sound power level in regard to humidity changes is low, indicating a slight increase with raising humidity [63]. Jiang et al. [43] report from field measurements that squeal occurrences are distributed equally between 20 % and 100 % humidity and thus no influence is observed. However, no squeal noise is occurred at very dry conditions (0 % - 20 % humidity). Dittrich et al. [54] elaborate a measurement procedure for type testing vehicles in terms of curve squeal and suggest dry conditions ($<$ 50 % humidity; no rain, frost, snow or dew) for frequent squeal occurrences. In a research project [64], long-term measurements in a light rail network are carried out and no correlation of relative humidity and occurrence of squeal noise is found.

In terms of absolute humidity, even less research is carried out. In the only paper that discusses the influence on curve squeal in detail, a decrease of A-weighted sound pressure level in the one-third octave bands between 2,5 kHz and 10 kHz with increasing absolute air humidity is observed. It is not distinguished between squeal and flanging noise in the evaluation [65]. Krüger [66] reports no correlation of absolute humidity and curve squeal.

Temperature influence in general is not discussed or not found to be correlated with curve squeal [66], however especially the dew point on the rail seems important. A significant decrease of probabilities of squeal noise and flanging noise is observed if the rail temperature drops below the dew point [63].

In wet conditions, the past research works accordingly report a significantly lower probability of squeal noise [5, 9, 35, 63, 66] and flanging noise [9, 63]. Moreover, in rainy conditions, a slight reduction in approximated sound power level is observed [63].

[66] and [65] observe an influence of daytime. While Krüger [66] reports no squeal noise occurrence in the morning hours (roughly between 6:00 and 9:30, presumably season dependent), Venghaus [65] points out the opposite. The environmental conditions during that time span are low temperatures and high relative humidity.

To conclude:

- findings on the influence of relative humidity and absolute humidity on curve squeal are not consistent
- wet and dew conditions significantly reduce the likelihood of curve squeal
- daytime effect is reported controversial
- other parameters (e.g. air pressure, sun radiation, cloudiness, ...) are not discussed

Hence, there is need for further research and detailed discussion of the influence of environmental conditions on elevated curve noise emission, which is carried out in this thesis.

2.2.4 Consideration in standards

As previously mentioned, the elevated sound emission of rail vehicles in curves is rarely considered in standards. A few examples are given for European noise calculation approaches. The list does not aim to be exhaustive.

In Switzerland, curve squeal is not considered in the noise emission calculation model SonRAIL [67] (for English sources refer to [68, 69]). However, an increase in noise emission in curves is taken into account by implementing an equivalent rail roughness, which is calculated with a roughness category of the vehicle (smooth/bad) and the curve radius. Wavelength spectra are given for both roughness categories.

In Germany, the Schall 03 model [70] contains correction factors for the calculation of noise emission of +8 dB for radii <300 m and +3 dB for radii 300 m $\leq R <$ 500 m. If permanently efficient measures against curve squeal are installed, the factor may be reduced by -3 dB in both categories.

In Austria, the RVE 04.01.02 [71] is elaborated in accordance to the Commission Directive (EU) 2015/996 [4], which contains the CNOSSOS-EU model. The noise emission model includes a correction factor for elevated curve noise of +8 dB for radii <300 m and +5 dB for radii 300 m $\leq R <$ 500 m. The factor is only valid for sections, where those radii exist over at least 50 m track distance. Based on a research project, it is recommended that for trains with a high rolling noise on tangent track (e.g. freight trains with cast-iron block brakes), the correction factor for middle radii should be neglected and for narrow radii should be reduced to +5 dB. Moreover, it is stated that climatic conditions can have a significant influence on the occurrence of curve squeal, thus if a decrease in occurrence can be proven, a proportional reduction of the factor is allowed.

The implemented correction factors stated in the previous paragraphs are questionable because every vehicle, regardless if it is a passenger coach, traction unit or freight wagon, gets the same consideration for elevated curve noise. As pointed out in RVE 04.01.02 [71], trains show different squeal likelihoods due to dependencies on wheel-rail-geometry and dynamics. Furthermore, climatic conditions and their influence on curve squeal behaviour are also not considered.

2.3 Basics in meteorology

The atmosphere contains a varying amount of water (from zero to four volume percent [72]) in all three aggregate conditions. For evaluations near ground level, as needed in this thesis, the water vapour concentration in those areas is of importance. In the following subsections, the basic parameters of meteorology – air pressure, temperature, humidity – are explained. Moreover, dew processes and the development of fog are described.

2.3.1 Air pressure

Air pressure can be illustrated as weight of the vertical air column over 1 m^2 earth surface with the same value in all directions. The ambient air pressure in the standardized atmosphere is defined as 1013.25 hPa (1 hPa = 100 Nm^{-2}) at 15 °C and an air density of 1.225 kgm^{-3} [72]. In Central Europe, fluctuations between 966 hPa and 1058 hPa (extreme values) are observed [73]. In weather forecasts, it is often spoken of areas with low pressure (indicating rain or storm) and high pressure (indicating clear sky), however this has not always to be the case. In stable weather conditions, daily fluctuations in the range of 2 hPa with two maxima

(around 10 and 22 o'clock local time) and two minima (roughly at 4 and 16 o'clock local time) over daytime can be observed [72]. In a weather change, the air pressure can drop or raise in the range of 20 hPa in a few hours [73].

2.3.2 Temperature

The air temperature dictates the value of water vapour saturation pressure and is thus of importance in terms of condensation. Vapour pressure is further discussed in Subsection 2.3.6. The air temperature shows a typical daily variation with a minimum in the morning hours at sunrise (mostly between 4 and 7 o'clock local time) and a maximum after the sun peak (roughly between 14 and 15 o'clock local time). Furthermore, a steep increase till midday followed by a lower one until the maximum is reached and a steep decrease in the evening hours followed by a lower one until the minimum value is observed. The development and especially the extreme values are influenced by the sun and the cloudiness of the sky. The surface temperature is of importance in terms of the development of dew drops as can be frequently observed on cars, roofs or grass. One reason is a cooling process induced by radiant heat loss (long-wave radiation) over night, where the surface temperature can drop 2 °C to 5 °C below the air temperature (further explained in Subsection 2.3.4) and thus dew processes can take place even though the relative air humidity is at 80 % or 90 %. [73]

2.3.3 Humidity

Humidity indicates the amount of water vapour in the air, which predominant generation mechanism is the evaporation above the sea [72]. From measurements, the most common parameter to express how humid the air is states the relative humidity (mostly indicated in %), which is the ratio of current partial water vapour pressure and saturated water vapour pressure. For an estimation of the amount of water vapour in the air, the absolute humidity, which depicts a density value (mass per volume), is needed. The computation and dependencies of the mentioned variables are described in Subsection 2.3.6. The relative humidity shows an inverse daily course to the temperature (again due to the dependency of the saturated water vapour pressure on temperature), whereas the absolute humidity indicates a different behaviour in autumn and winter compared to spring and summer. In the former seasons, the absolute humidity increases between sunrise and midday due to evaporation processes caused by the temperature increase. During the night, the value decreases due to dew processes on the surface. In the latter seasons, the same effects take place except in the hours around midday and in the afternoon convection takes place. Warmed air layers close to the surface rise in the air, including a high water vapour amount (from evaporation in the morning hours) and, as compensation, air with lower amounts of water vapour sinks onto the surface. Thus, the absolute humidity has a minimum in the afternoon and increases in the evening, as convection fades, till dew processes take place during the night. [73]

2.3.4 Dew formation

Since the wetness of the rail can alter the curve squeal behaviour significantly (refer to Subsection 2.2.3), the formation of dew on the rail is of interest as a wetting source in addition to fog and rain. In contrast to the two latter (liquid water droplets adhere on the surface), dew formation is a condensation process, where water vapour changes its state of aggregation to liquid on a surface, which causes a release of heat (latent heat of vaporization). If the surface temperature lies beneath the frost point (computable with (15)) of the ambient air, water vapour changes directly to the solid state and hoarfrost appears on the surface (causing release of the latent heat of sublimation) [74]. Another way for ice occurrence is possible if the surface is wet (e.g. from dew formation) and the surface temperature drops beneath the freezing point of water, which results in the release of the latent heat of fusion [75]. All three processes lead to an increase in surface temperature. To derive when the dew formation starts and at which rate the condensation takes place (leading to the accumulated mass e.g. per night), a general heat balance equation for a surface (e.g. the rail) can be formulated [76]:

$$Q^* = Q_\mathrm{H} + Q_\mathrm{E} + Q_\mathrm{G} \tag{4}$$

where Q^* is the net radiation, Q_H is the convective sensible, Q_E is the convective latent and Q_G is the conductive heat flux density all in units $\mathrm{Wm^{-2}}$. The radiation term contains the sum of heat flux densities received by solar (short-wave) and by terrestrial (long-wave) radiation from the atmosphere and emitted (i.e. lost) as upward heat flux density (long-wave) radiation to the atmosphere [74]. To stick to the chosen notation, the formula is as follows [76]:

$$Q^* = a \cdot K + \epsilon_\mathrm{sky} \cdot L - \epsilon_\mathrm{s} \cdot \sigma \cdot T_\mathrm{r}^4 \tag{5}$$

where a is the absorption coefficient of the surface for short-wave radiation (between 0 and 1, dimensionless), K is the incoming short-wave radiation, L states the incoming long-wave radiation (both with units $\mathrm{Wm^{-2}}$), ϵ_sky and ϵ_s are the emissivities of the sky and the surface for long-wave radiation (between 0 and 1, dimensionless), σ is the Stefan-Boltzmann constant ($5.67 \cdot 10^8$ $\mathrm{Wm^{-2}K^{-4}}$) and T_r is the temperature of the surface.

The right-side terms of (4) can be calculated as follows [76]:

$$Q_\mathrm{H} = h_\mathrm{c} \cdot (T_\mathrm{r} - T_\mathrm{a}) \tag{6}$$

$$Q_\mathrm{E} = h_\mathrm{v} \cdot (q_\mathrm{r} - q_\mathrm{a}) \tag{7}$$

$$Q_\mathrm{G} = h_\mathrm{k} \cdot (T_\mathrm{r} - T_\mathrm{u}) \tag{8}$$

where h_c, h_v, h_k are heat transfer coefficients for convective sensible heat, convective water vapour and conductive sensible heat (h_c, h_k with units $\mathrm{Wm^{-2}K^{-1}}$ and h_v in $\mathrm{Wm^{-2}}$), T_a, T_u are the temperatures of the ambient air and of the subsurface (e.g. in case of a rail the sleeper) both in K and q_r, q_a are the specific humidity quantities (expressed in (17)) at the surface

and in the ambient air (both with units $kgkg^{-1}$).

Introducing a convention that all fluxes away from the surface are positive, dew formation takes place if Q_E is negative and evaporation happens in case of Q_E stays positive [74]. Dividing Q_E by the latent heat of vaporization (in case of liquid dew formation) leads to the mass flux rate of water (with units $kgm^{-2}s^{-1}$), which states an equivalent depth of water (mm) per time unit [76]. Equation (4) can also be used to calculate the surface temperature (e.g. with iteration [77] or FE-modelling [78]).

Maximum dew formation rates of 0.06 mmh^{-1} - 0.07 mmh^{-1} and maximum quantities of dew formation of 0.2 mm per night - 0.4 mm per night are observed. In general, clear nights (no clouds) enhance dew formation (due to more radiant heat loss from the surface) and the probability of dew formation peaks at sunrise. Besides clouds, other environmental influence factors are wind speed, air temperature, relative humidity and surface temperature. [74]

2.3.5 Fog

In meteorology, fog is basically a cloud, which lies on the earth surface. Hence, if the relative humidity of the air is locally over 100 %, water vapour condensates to lower the humidity value to the saturation level again. The term "fog" is used if the sight is limited to below 1000 m. Four fog generating mechanism are distinguished [73]:

- radiation fog: air near ground level cools down due to radiation of the surface
- advection fog: warm and humid air is over a cold surface and cools down under the dew point (e.g. over lakes with cold water temperature)
- vaporization fog: raise of humidity level at constant temperatures due to vaporization (e.g. snow in warm ambient air)
- mixture fog: Decrease of temperature and increase of humidity simultaneously (e.g. in border areas between cold and warm air layers)

2.3.6 Computable values

Usually, data from weather stations contain at least absolute air pressure, relative humidity and temperature of the ambient air. With those three input values, lots of other parameters can be computed. The World Meteorological Organization (WMO) provides a guide for the measurement of meteorological variables [79], which also contains formulae for computation. In general, the guide aims for a standardization in measurement and calculation methods. In the following, the parameters relative air pressure, saturated water vapour pressure, partial water vapour pressure, dew point, absolute humidity, specific humidity and mixing ratio are covered.

Relative air pressure

For comparison of air pressure levels from different locations, the measured value of the barometer (called absolute air pressure – indicating the pressure of the ambient air) is reduced

to a standard level. For stations up to 750 m above mean sea level (MSL[4]), the absolute air pressure is usually reduced to MSL. The general reduction formula is as follows [79]:

$$\log_{10}\left(\frac{p_0}{p_s}\right) = \frac{K_p \cdot H_p}{T_s + \frac{a_l \cdot H_p}{2} + e_s \cdot C_h} \tag{9}$$

where the index s indicates the measured values at the station – p and e in hPa, T in K –, p_0 states the relative air pressure, K_p is the constant 0.0148275 Kgpm^{-1}, which can be computed by $\frac{g}{R} \cdot \log_{10}(e)$ with g as the standard acceleration of gravity (9.81 ms^{-2}), R as the gas constant of dry air (287.1 Jkg^{-1}K^{-1} [73]) and e the Euler number, H_p is the station elevation in gpm, a_l is the assumed lapse-rate between sea and station level in Kgpm^{-1} and C_h is a coefficient with value 0.12 KhPa^{-1} (for more information about a_l and C_h be referred to [81]).

As can be seen, the reduction formula depends not only on elevation, but also on temperature and relative humidity (in terms of the partial water vapour pressure). Furthermore, the elevation is expressed as geopotential altitude and not as geometrical altitude. An approximate equation to convert the scales is as follows [81]:

$$H_p \approx \frac{z}{1 + \frac{z}{r}} \tag{10}$$

where z states the geometric altitude in m and r is the radius of the earth in m ($6.37 \cdot 10^6$).

If the conversion of (10) is calculated with 750 m (the maximum where (9) is valid), the result is 749.912 gpm, thus, the deviation is 0.1 ‰. Hence, at low altitude levels, the difference is negligible.

Saturated water vapour pressure

The saturated water vapour pressure e_{sat} expresses the limit of partial water vapour pressure at a certain temperature in an air volume. It was shown experimentally that it depends only on air temperature and thus the temperature of the vapour. Several formulas for the calculation of e_{sat} were derived in the past decades. An overview and a plot of the divergences is given in [82]. According to [79], the following formulation is standardized (valid above water surface in an air temperature range of -45 °C $\leq t \leq$ 60 °C):

$$e_{sat} = 6.112 \cdot e^{\frac{17.62 \cdot t}{243.12 + t}} \tag{11}$$

where e_{sat} is the saturated water vapour pressure in hPa and t states the air temperature in °C.

Equation (11) calculates the value of saturated water vapour pressure in the pure phase (assuming pure water vapour without other air components). In reality, the air contains a few percent of water vapour, which causes that saturated water vapour pressure deviates from the

[4]The MSL is defined as "The average sea surface level for all stages of the tide over a 19-year period, usually determined from hourly heights observed above a fixed reference level." [80] p.376

one of the pure state. This is taken into account by an enhancement factor, which depends on temperature and pressure of the ambient air [83]:

$$e'_{\mathrm{sat}} = f(p,t) \cdot e_{\mathrm{sat}} \tag{12}$$

where e'_{sat} is the saturated water vapour pressure of moist air in hPa and $f(p,t)$ depicts the enhancement factor.

The enhancement factor can be extracted from Table 4.10 in [83] by interpolating the values of pressure and temperature linearly. According to [79], the temperature dependency is much lower ($\pm$ 0.1 %) than the dependency from pressure (0 % to +0.6 %) at pressure levels of around 1000 hPa, which allows neglecting the temperature dependency. If the enhancement factor is not considered at all, the error value is up to 0.5 % in the calculation [83]. The depicted equations are valid over water surfaces, which is a taken assumption in the thesis. Above ice surface, the equations are different (for further information be referred to [79]).

Partial water vapour pressure

The proportion of air pressure caused by water vapour is called partial water vapour pressure e' (for moist air) and can be calculated backwards from the relative humidity (if not measured explicitly) [79]:

$$e' = \frac{\Phi}{100} \cdot e'_{\mathrm{sat}} \tag{13}$$

where e' is the partial water vapour pressure for moist air in hPa and Φ is the relative humidity in %.

The calculated value of e' contains the enhancement factor, which is discussed in the previous paragraph.

Dew point and frost point

The temperature, where the current water vapour pressure is equal to the saturated water vapour pressure – or in other terms the relative humidity reaches 100 % – is called dew point. If a local temperature (e.g. the temperature of the surface) falls beneath the dew point of the ambient air, condensation takes place and dew drops appear. The standardized calculation is as follows (valid above water surfaces in an air temperature range of -45 °C $\leq t \leq$ 60 °C) [79]:

$$t_{\mathrm{d}} = \frac{243.12 \cdot \ln\left(\frac{e'}{6.112 \cdot f}\right)}{17.62 - \ln\left(\frac{e'}{6.112 \cdot f}\right)} \tag{14}$$

where t_{d} is the dew point temperature in °C.

If the surface temperature is beneath the frost point, hoarfrost appears instead [74]. The standardized calculation is as follows (valid above ice surfaces in an air temperature range of -65 °C $\leq t \leq$ 0 °C) [79]:

$$t_{\mathrm{f}} = \frac{272.62 \cdot \ln\left(\frac{e'}{6.112 \cdot f}\right)}{22.46 - \ln\left(\frac{e'}{6.112 \cdot f}\right)} \tag{15}$$

where t_{f} is the frost point temperature in °C.

As can be seen, by a combination of (11), (12) and (13), dew and frost point can be directly computed with the relative humidity, pressure and temperature of the ambient air.

Absolute humidity

The absolute humidity allows a direct estimation of the amount of water vapour in the air in terms of a density (mass per volume). Since water vapour states only a few percent of the air, the equation is depicted in gm^{-3} and not in the SI-unit for a better readability [72]:

$$a = 10^3 \cdot \frac{e' \cdot 10^2}{R_{\mathrm{w}} \cdot T} \tag{16}$$

where a is the absolute humidity in gm^{-3}, R_{w} is the gas constant of water vapour (461.525 $\mathrm{Jkg}^{-1}\mathrm{K}^{-1}$ [84]) and T is the temperature in K.

Specific humidity

The specific humidity is used to compare the humidity of air in different locations with alternating pressures and temperatures. It can be expressed as the ratio between the density of the water vapour and the density of the humid air or the ratio between the mass of water vapour and the total mass of the humid air, respectively. The value can be computed as follows [72]:

$$s = 10^3 \cdot \frac{M_{\mathrm{W}}}{M_{\mathrm{L}}} \cdot \frac{e'}{p - \left(1 - \frac{M_{\mathrm{W}}}{M_{\mathrm{L}}}\right) \cdot e'} \tag{17}$$

where s is the specific humidity in gkg^{-1}, M_{W} and M_{L} are the molar masses of water vapour (0.01801528 kgmol^{-1}) and dry air (0.0289645 kgmol^{-1}) and p states the absolute pressure in hPa [84].

Mixing ratio

The mixing ratio is similar to the specific humidity and again a method to compare the air humidity at different locations. It states the ratio of the density of the water vapour and the density of the dry part of the air or the ratio of the mass of water vapour and the mass of dry air. The formula is as follows [72]:

$$m = 10^3 \cdot \frac{M_{\mathrm{W}}}{M_{\mathrm{L}}} \cdot \frac{e'}{p - e'} \tag{18}$$

where m is the mixing ratio in gkg^{-1} [84].

Since the mass of water vapour in a kilogram of air is only a several grams, it is common to scale the values of the specific humidity and mixing ratio to gkg^{-1} [72].

3 Measurement data

The following subsections present an overview of the measurement data. Firstly, a description of measurement setup and sections, as well as the comparability of the latter, is given. Moreover, defined exclusion criteria as a first filtering step of train pass bys, which are not considered in further evaluations, are discussed. The last subsection provides fundamental statistics of the used dataset.

3.1 Measurement sections

All measurement campaigns were carried out with a mobile acoustic railway monitoring system (called acramos®), which allows automatically triggered sound pass by measurements. The general measurement setup contains at least one microphone in the standard position according to ISO 3095 [10] – 7.5 m from the track axis and 1.2 m above rail level –, two axle counter sensors on the measured track, one axle counter sensors on the other track (two-track lines in all sections), a weather station to monitor the environmental parameters and a rail temperature sensor located at the rail base between two sleepers. Axle counter sensors are used to trigger and stop the measurement, get a time reference to every axle in the sound signal, calculate axle speeds, identify the train type by axle pattern analysis and monitor if two trains are passing at the same time. Furthermore, at some measurement sections additional sensors are used, which are not used for further analysis and therefore only partly mentioned for completeness. Sections are termed C1 (described in Subsection 3.1.2), C2 (explained in Subsection 3.1.3), C3 (illustrated in Subsection 3.1.4), C4 (outlined in Subsection 3.1.5) and C5 (discussed in Subsection 3.1.6). The carried out measurements were part of two research projects – C1 - C3, final report in German [9] and C4/C5, final report in German [85]. In the following, a detailed description of the measurement setup and sections is given.

3.1.1 General measurement setup

As partly introduced, the acramos® system analyses train pass bys automatically. The output parameters are saved as ASCII-files and transferred to a database. The central operation unit, which is positioned near the section in a control box, consists of an industrial computer with two 8-channel A/D converters, an USV to ensure interruption-free power supply, evaluation modules for the axle counter sensors and a module for remote control. The computer is connected to the net by a GSM modem. For noise level measurements, class 1 microphones were used and calibrated with a CAL200 device before and after each measurement campaign. In the first three measurement campaigns (C1 - C3), the measurement data were recorded with a sampling rate of 32 kHz, whereas in C4 and C5, it was raised to 50 kHz. Key data about measurement system and sensors are depicted in Table 1 (C1 - C3) and Table 2 (C4/C5).

Rail roughness is monitored with an ODS TRM-05 device using a length of 1.2 m. Three sensors monitor the vertical profile of the rail head simultaneously. All three scan lines are averaged.

Track decay rate is measured either with the TNO method or the AEIF method. The

channel	name	type	serial no.	C1	C2	C3	used in further evaluation
1	M1	GRAS46AE	195314	✓	✓	✓	✓
2	M2	GRAS46AE	195315	✓			
3	M3	GRAS46AE	195316	✓			
4	M4	Gefell-ICP	1568	✓			
5	M5	GRAS46AE	77308	✓	✓	✓	
7	V1	IMI 352C03	144764	✓	✓	✓	
8	H1	IMI 352C04	139368	✓	✓	✓	
9	V2	IMI 352C05	139525	✓	✓	✓	
10	H2	IMI 352C06	139524	✓	✓	✓	
13	R2.1	Altpro ZK24	–	✓	✓	✓	
14	R1.1	Frauscher	–	✓	✓	✓	✓
15	R2.0	Altpro ZK24	–	✓	✓	✓	
16	R1.0	Frauscher	–	✓	✓	✓	✓
USB	–	Müller PT100	–	✓	✓	✓	✓
USB	–	Conrad Professional USB weather station	–	✓	✓	✓	✓

Table 1: Channel and sensor list for C1 - C3

channel	name	type	serial no. C4	serial no. C5	used in further evaluation
1	M1	GRAS46AE	195316	195314	✓
2	M2	GRAS46AE	58530	195315	✓
4	–	Müller PT100	–	–	✓
7	R1	Altpro ZK24	–	–	✓
8	R2	Altpro ZK24	–	–	✓

Table 2: Channel and sensor list for C4/C5

device	type	serial no.
impact hammer	086D05	31723
acceleration sensor vertical and lateral	393A03	–
calibrator	MF VC10	960202
system	MEDA (Wölfel Messsysteme Software GmbH & Co. KG) 8.102	08/695-2

Table 3: Device list for track decay rate measurements

TNO method is calculated with the software PBA arithmetically and uses the train itself as vibration source for track decay rate calculation (hence a pre-loaded track). In contrast, the AEIF method, which is applied in C2, C4 and C5, uses the vibration response of an impact hammer for calculation (track not pre-loaded). For further details and comparison be referred to [86]. The latter is standardized in EN 15641 [87], which is the reason why in C4 and C5 only that method was applied. A list of used devices for track decay rate measurement is provided in Table 3.

3.1.2 Kleinschwechat, S7 (C1)

The first measurement campaign – called C1 throughout the thesis – was carried out at a curve of the S7 line near Kleinschwechat (km 11.4) in Austria. The installation of the sensor components took place on 18th October 2013. For detailed information about the track parameters, refer to Table 4. In Figure 5, the measurement section with the sensor positions and the environment is depicted. The track is limited at the opposite side of the sensor components by a 3 m high wall with sound absorbing elements and on the other side by an earth wall. Four microphones were installed at the outer side of the curve and one between both tracks. The microphone positions were as follows (note that the distances refer only to the measured track 1):

- M1: 7.5 m from the axis, 1.2 m above rail level
- M2: 7.5 m from the axis, 13 m above rail level
- M3: 10.6 m from the axis, 10.6 m above rail level
- M4: 13 m from the axis, 7.5 m above rail level
- M5: 2.25 m from the axis, 0.5 m above rail level

For the analysis in this thesis, only measurement data from M1 are used. In addition to the sound measurements, acceleration sensors on the outer rail and on the inner rail were installed. The horizontal acceleration sensor was mounted on the outer rail head and the vertical one on the rail base – both magnetically. The measurement campaign lasted till 7th July 2014. On 27th February 2014, the rail roughness as well as the track decay rates were

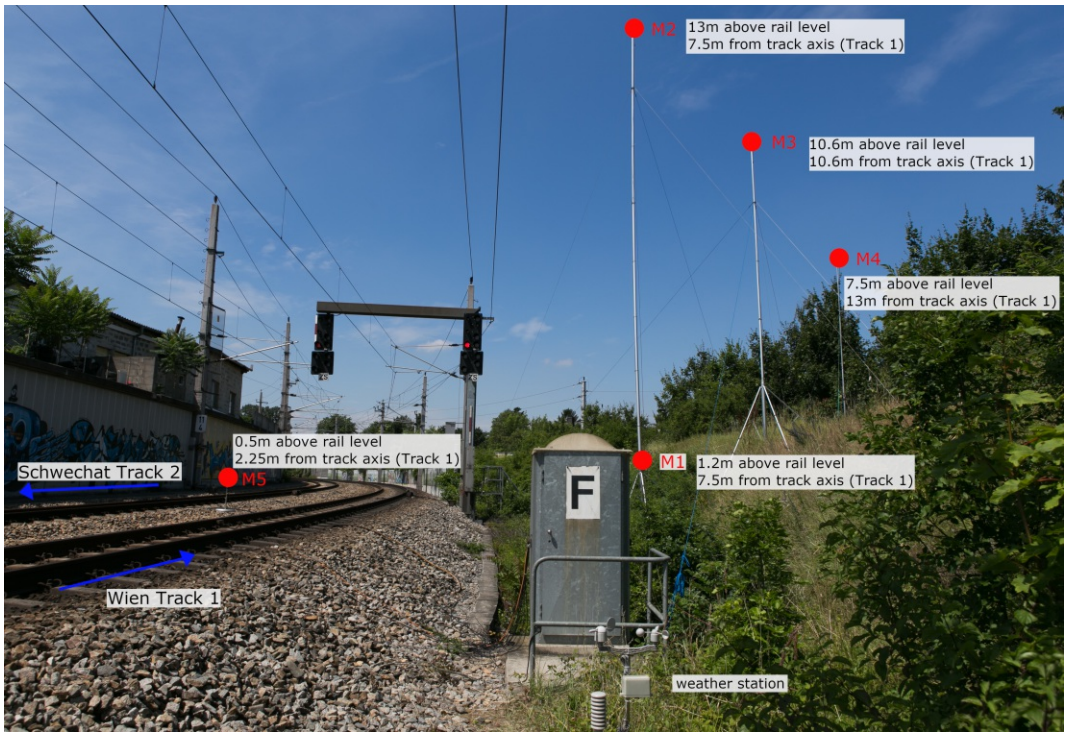

Figure 5: Overview of the measurement section of C1 (© T. Maly)

measured. The rail roughness of the inner rail is above the ISO 3095 [10] limit nearly over the whole spectrum, whereas the outer rail shows a rail roughness lower than the ISO 3095 limit over all considered wavelengths (depicted in Figure 6). The track decay rate was measured with the acceleration sensors and calculated with the TNO method. The lateral track decay rates of both rails satisfy the ISO 3095 [10] limit over the whole frequency range (see Figure 7). As can be seen in Figure 8, the vertical track decay rate drops under the limit partly on both rails between 400 Hz and 630 Hz. The traffic on that line is dominated by commuter trains.

track kilometre	11.4
curve radius	256 m
rail profile	UIC60
cant	70 mm
longitudinal gradient*	11.1 ‰
superstructure	gravel
sleepers	concrete
speed limit	60 kmh^{-1}
balanced speed (v_a)	39 kmh^{-1}

* in norm direction, see Figure 5

Table 4: Track parameters of C1

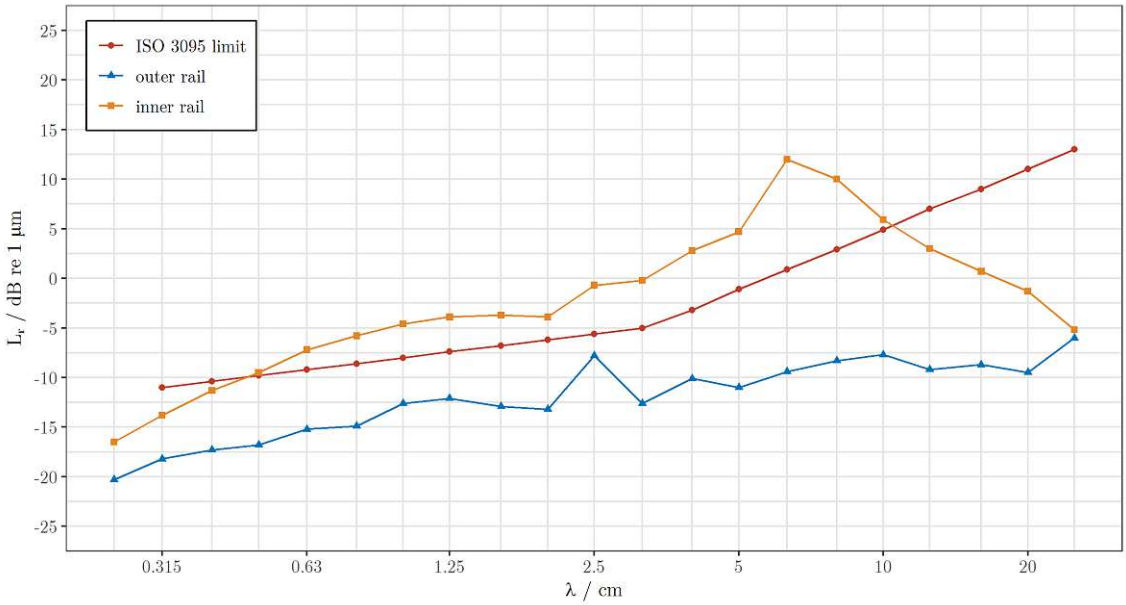

Figure 6: Rail roughness measurement in C1 on 7th February 2014

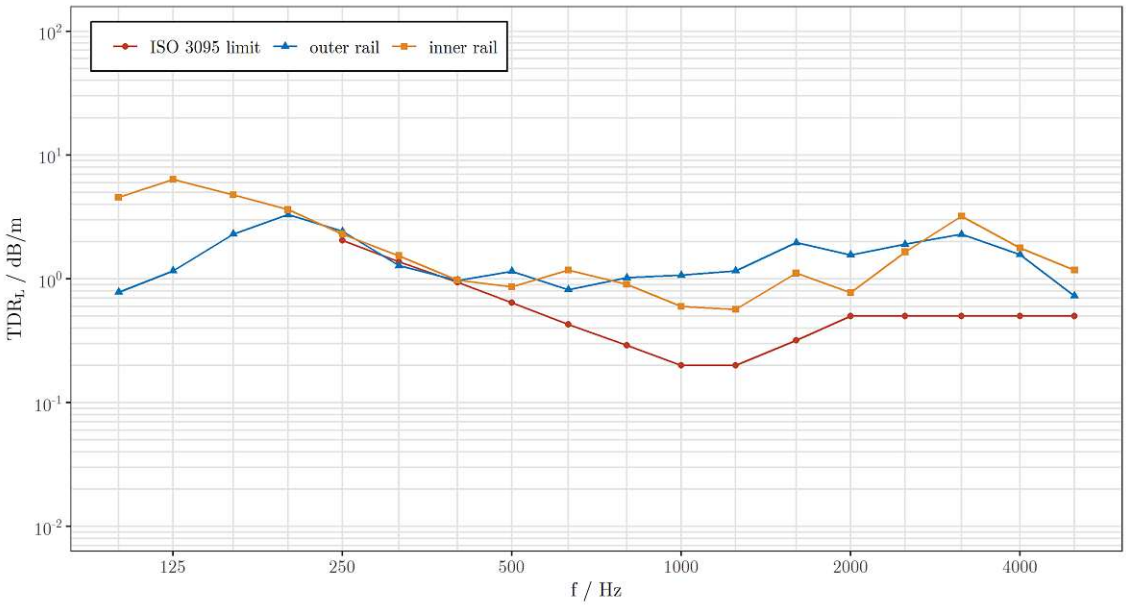

Figure 7: Lateral track decay rate measurement in C1 on 19th October 2013

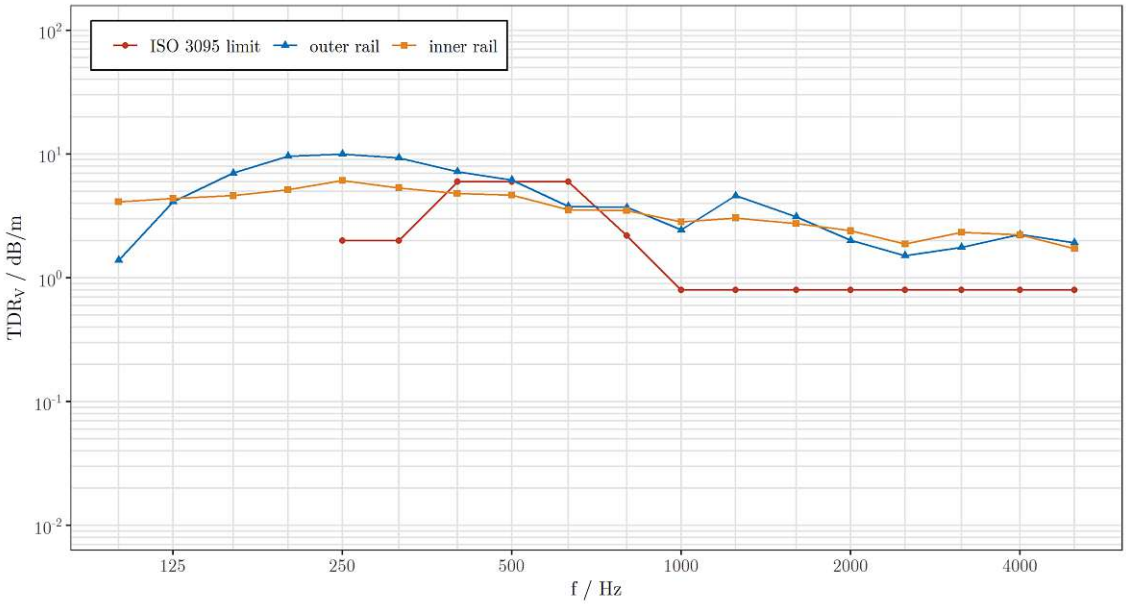

Figure 8: Vertical track decay rate measurement in C1 on 19th October 2013

3.1.3 Eichgraben, Westbahn (C2)

The second measurement campaign – called C2 throughout the thesis – was carried out at a curve of the Westbahn line near Eichgraben (km 27.2) in Austria. The measurement section is depicted in Figure 9. The outer side of the curve is limited by heavy vegetation, whereas the inner side allows free sound radiation over a field with slightly rising altitude. An additional microphone was placed in the middle of the two tracks and acceleration sensors were installed in the same way as in C1. The distances of the microphones were as follows (note that the distance only refers to the measured track 1):

- M1: 7.5 m from the axis, 1.2 m above rail level
- M5: 2.25 m from the axis, 0.5 m above rail level

track kilometre	27.2
curve radius	440 m
rail profile	54E2
cant	131 mm
longitudinal gradient*	-10.5 ‰
superstructure	gravel
sleepers	concrete
speed limit	90 kmh^{-1}
balanced speed (v_a)	69.9 kmh^{-1}

* in norm direction, see Figure 9

Table 5: Track parameters of C2, Eichgraben, Westbahn

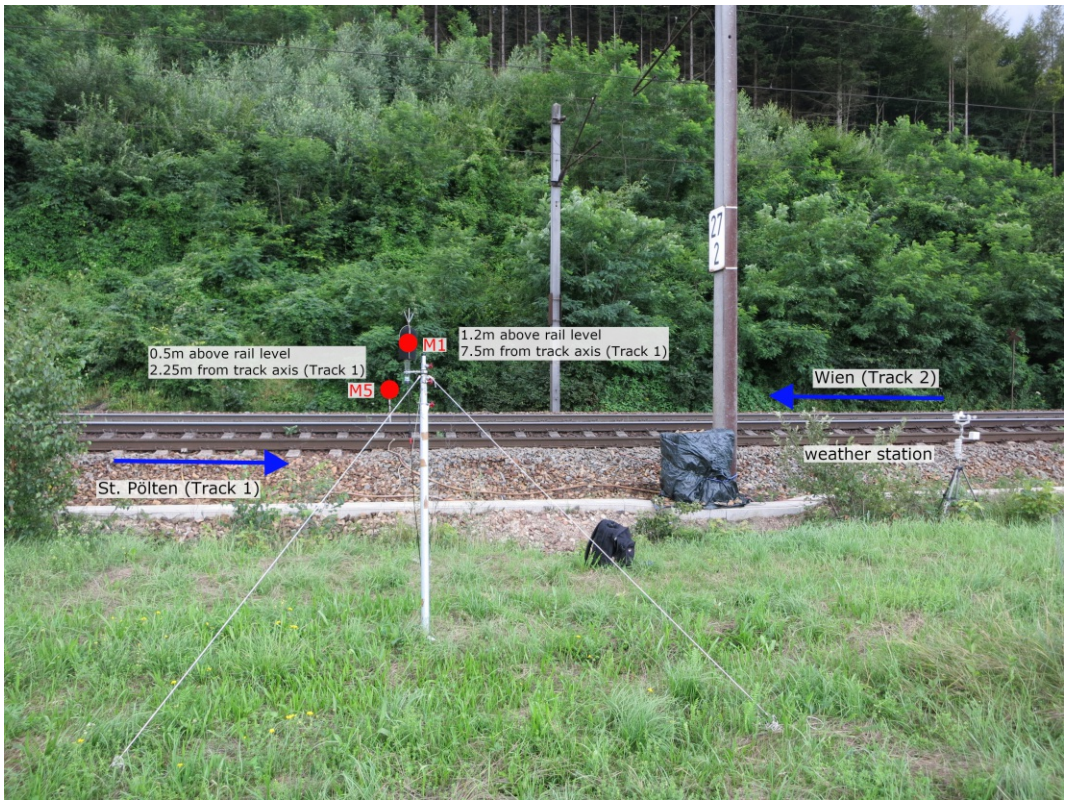

Figure 9: Overview of the measurement section of C2 (© psiacoustic Umweltforschung und Engineering GmbH)

For the analysis in this thesis, only measurement data from M1 are used. For detailed information about the track parameters, refer to Table 5. The measurement campaign was carried out between 8th July 2014 and 30th July 2014. On 8th July 2014, the rail roughness as well as the track decay rates were measured. The rail roughness is except at a wavelength of 2.5 cm beneath the ISO 3095 [10] limit at both rails (see Figure 10). To investigate differences in measurement methods, the calculation of the track decay rates was carried out with the TNO and the AEIF method. Although carried out only within four days difference, both methods show completely different results regarding lateral track decay rate (see Figure 11), while the results in vertical track decay rate are comparable (illustrated in Figure 12). The vertical track decay rate stays above the ISO 3095 limit on both rails nearly over the whole frequency range – only dropping between 1600 Hz and 3150 Hz beneath the limit with the AEIF method. The lateral one is with PBA calculation above the limit nearly over the whole frequency range, while dropping with the AEIF method between 250 Hz and 400 Hz as well as between 800 Hz and 2500 Hz partly under the limit on both rails. Since the AEIF method is standardized in EN 15641 [87], only those values are used in further evaluations. Train traffic on that line consists of evenly distributed freight and commuter trains.

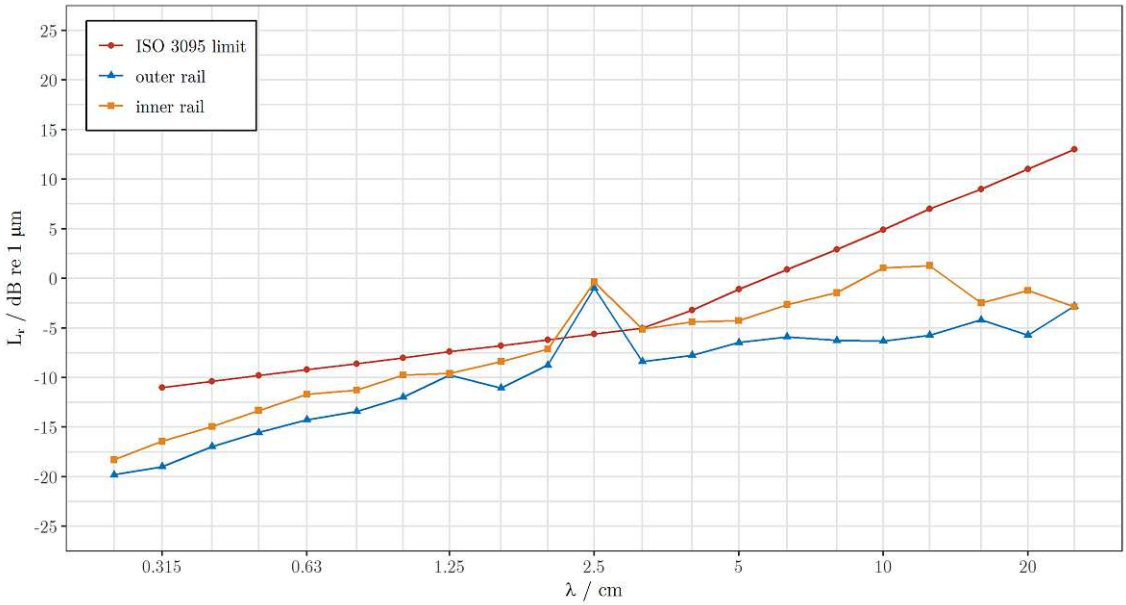

Figure 10: Rail roughness measurement in C2 on 8th July 2014

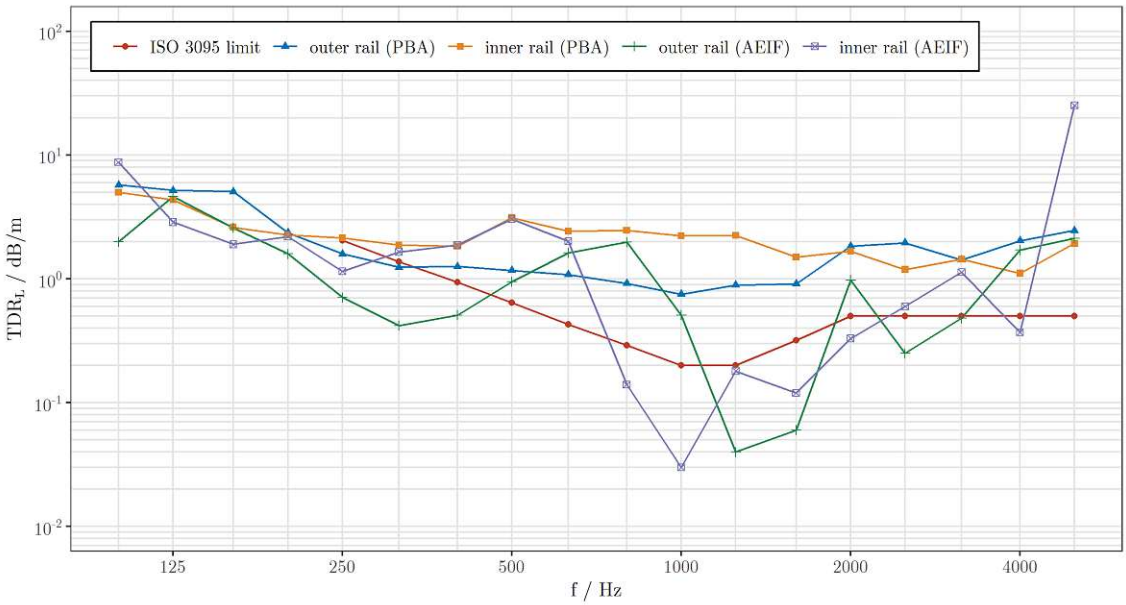

Figure 11: Lateral track decay rate measurement in C2 on 26th July 2014 (PBA) and on 30th July 2014 (AEIF)

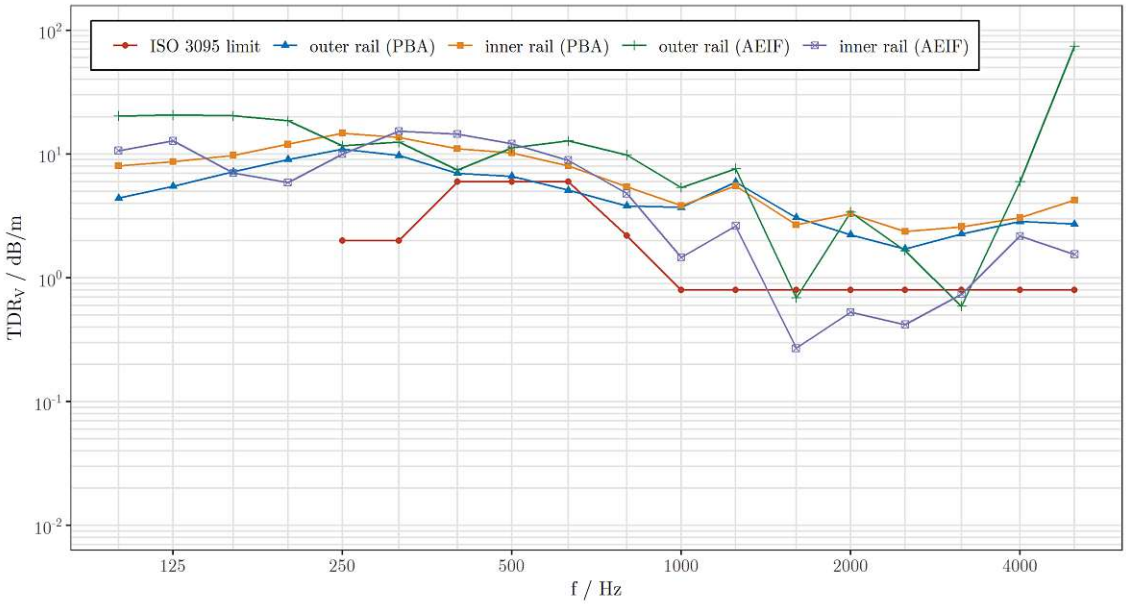

Figure 12: Vertical track decay rate measurement in C2 on 26th July 2014 (PBA) and on 30th July 2014 (AEIF)

3.1.4 Mürzzuschlag, Südbahn (C3)

The third measurement campaign – called C3 throughout the thesis – was carried out at a curve of the Südbahn line near Mürzzuschlag (km 112.85) in Austria. The measurement section can be seen in Figure 13. The track is located on a low dam with free sound radiation in every direction. An additional microphone was placed in the middle of the two tracks and acceleration sensors were installed in the same way as in C1. The distances of the microphones were as follows (note that the distance only refers to the measured track 1):

- M1: 7.5 m from the axis, 1.2 m above rail level
- M5: 2.25 m from the axis, 0.5 m above rail level

For the analysis in this thesis, only measurement data from M1 are used. For detailed information about the track parameters, be referred to Table 6. The measurement campaign lasted from 12th September 2014 till 3rd October 2014. The rail roughness (depicted in Figure 14) of the inner rail satisfies the ISO 3095 [10] limit. The roughness of the outer rail slightly exceeds the limit at a wavelength of 12.5 cm, but shows values under the limit in all other considered wavelengths. The track decay rates were calculated arithmetically with the TNO method. The lateral track decay rates (shown in Figure 15) of both rails satisfy the ISO 3095 [10] limit at all frequencies, except at 250 Hz. The vertical track decay rate (refer to Figure 16) does not satisfy the limit at 500 and 630 Hz (outer rail) and between 400 and 800 Hz (inner rail). The traffic on the line is a mix of long-distance passenger and freight trains.

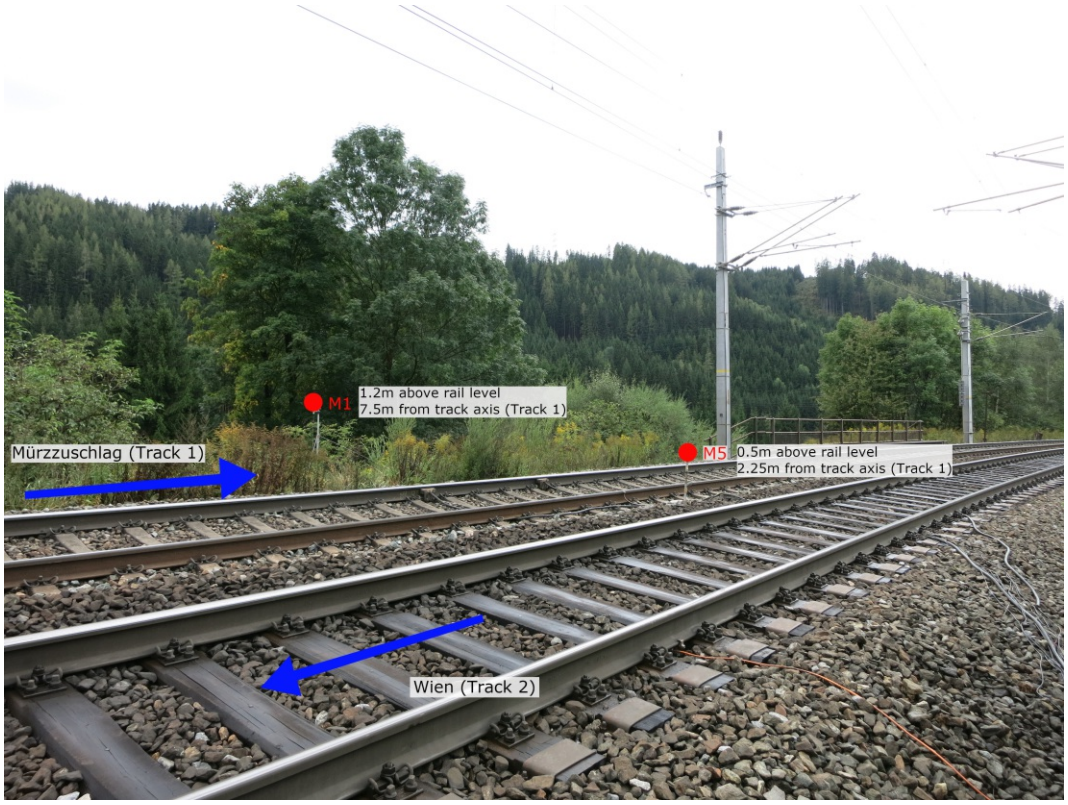

Figure 13: Overview of the measurement section of C3 (© psiacoustic Umweltforschung und Engineering GmbH)

track kilometre	112.85
curve radius	310 m
rail profile	UIC60
cant	150 mm
longitudinal gradient*	19 ‰
superstructure	gravel
sleepers	concrete
speed limit	80 kmh^{-1}
balanced speed (v_a)	62.8 kmh^{-1}

* in norm direction, see Figure 13

Table 6: Track parameters of C3, Mürzzuschlag, Südbahn

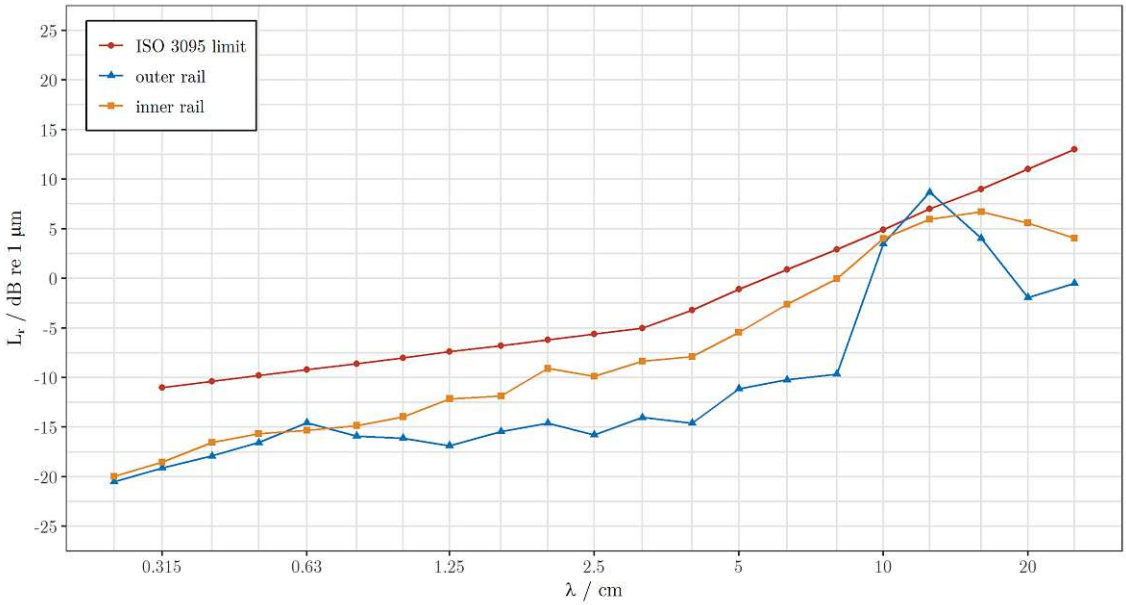

Figure 14: Rail roughness measurement in C3 on 3rd October 2014

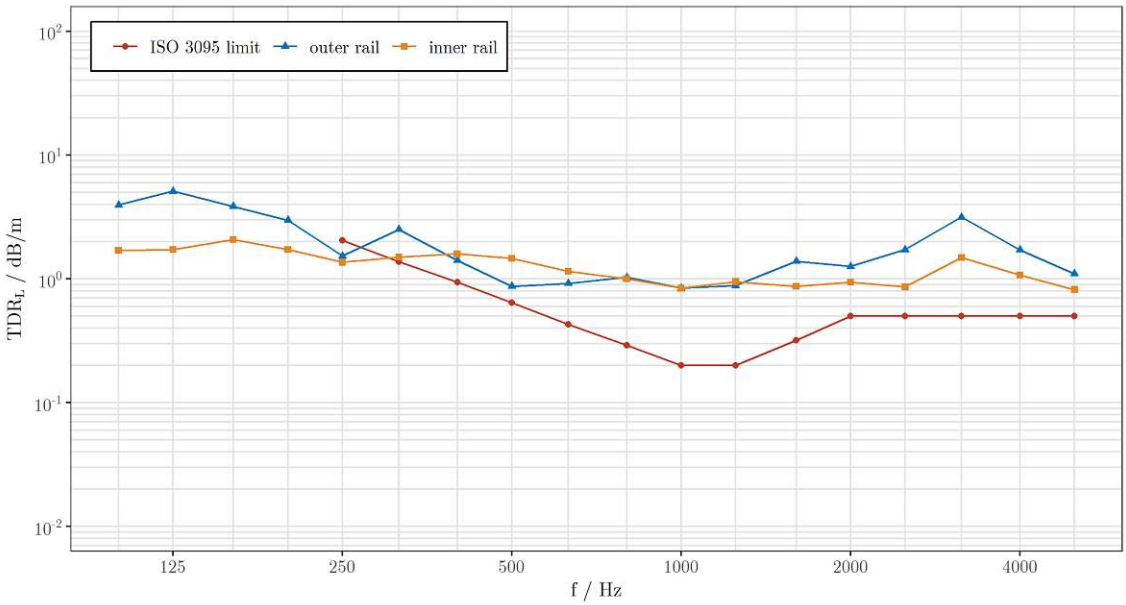

Figure 15: Lateral track decay rate measurement in C3 on 12th September 2014

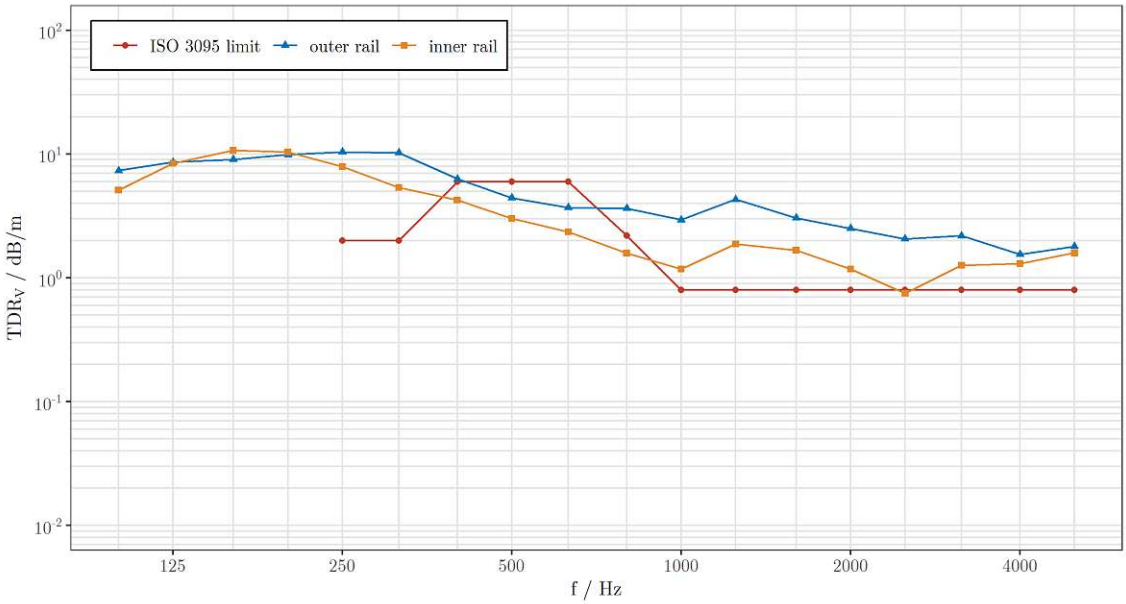

Figure 16: Vertical track decay rate measurement in C3 on 12th September 2014

3.1.5 Wien Gersthof, S45 (C4)

The fourth measurement campaign – called C4 throughout the thesis – was carried out at a curve of the S45 line (also called "Vorortelinie") in Vienna (km 5.6) in Austria. An impression of the measurement section is shown in Figure 17. The track is located in an urban environment with a decent population density. Near the track side with the microphones, a dead end street with low traffic is located. After the measurement section (direction Hütteldorf), the track leads over a bridge into a station, which causes that the trains tend to brake or at least do not accelerate anymore in the section and thus the speed limit is barely reached. Under the bridge, a street with heavy traffic and also a tram line is situated. The A-weighted equivalent sound pressure level of the environmental noise was investigated for 2 hours around midday and showed 57 dB, which is not considered influencing since train pass bys in 7.5 m distance from the axis show at least 10 dB higher noise levels. Furthermore, a rail defect was observed in a distance of 5 m to the measurement section. An investigation pointed out that the main influence on the emission lies between 160 Hz and 315 Hz with an elevation magnitude of 4 dB to 5 dB in third-octave band levels, which has to be borne in mind in the further evaluations. At the beginning of the curve (direction Hütteldorf), a wayside top-of-rail friction control system was located. In this campaign, two microphones in the same distance were used (note that the distance only refers to the measured track 2):

- M1: 7.5 m from the axis, 1.2 m above rail level
- M2: 7.5 m from the axis, 1.2 m above rail level

The measurement data for evaluation are only taken from M1, whereas M2 had only the function of validating the values of M1. For detailed information about the track parameters, refer to Table 7. The measurements were carried out between 4th February 2016 and 10th

Figure 17: Overview of the measurement section of C4 (© psiacoustic Umweltforschung und Engineering GmbH)

January 2017. From the beginning till 12th April 2016, the friction control system was disabled and afterwards enabled till the end of the campaign. Measurements of rail roughness and track decay rate (AEIF method) were carried out at the beginning and at the end of the measurement period. The rail roughness on the inner rail exceeds the ISO 3095 [10] limit in the wavelength spectra between 5 cm and 10 cm. The outer rail satisfies the ISO 3095 limit completely. The change in rail roughness is minor, except on the outer rail between 5 cm and 10 cm, where a significant increase can be seen – still under the limit. The lateral track decay rates of both rails are only between 1.25 kHz and 1.6 kHz below the ISO 3095 limit (see Figure 19). The alterations between beginning and end of the measurement period are minor, except a decline on the inner rail in the frequency range between 2000 Hz and 4000 Hz. The vertical track decay rate drops under the limit at 1600 Hz and 2500 Hz (only at the end) on the outer rail and at 400 Hz and 500 Hz (only at the end) on the inner rail (depicted in Figure 20). The traffic is dominated by commuter trains.

track kilometre	5.6
curve radius	226 m
rail profile	S49
cant	90 mm
longitudinal gradient*	0 ‰
superstructure	gravel
sleepers	wood
speed limit	60 kmh^{-1}
balanced speed (v_a)	41.5 kmh^{-1}

* in norm direction, see Figure 17

Table 7: Track parameters of C4, Wien Gersthof, S45

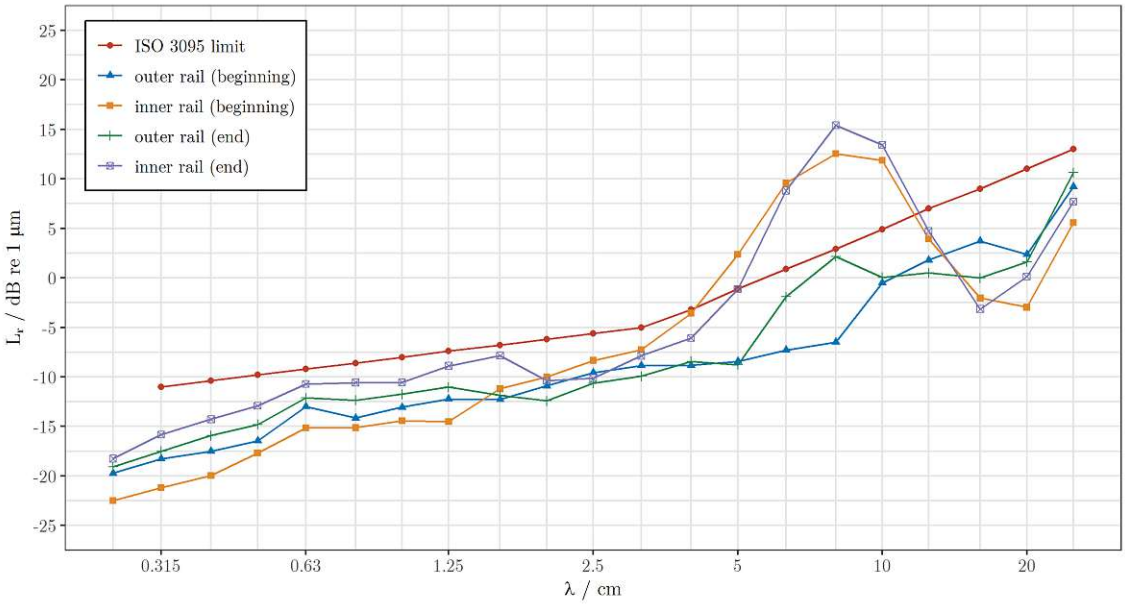

Figure 18: Rail roughness measurements in C4 on 2nd February 2016 and on 10th January 2017

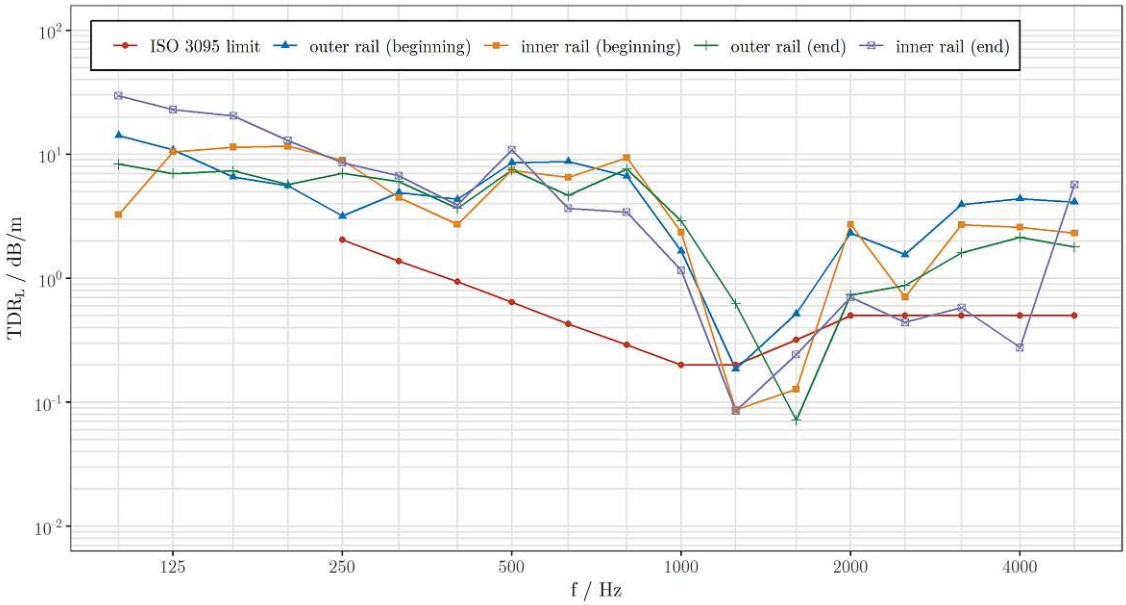

Figure 19: Lateral track decay rate measurements in C4 on 15th April 2016 and on 10th January 2017

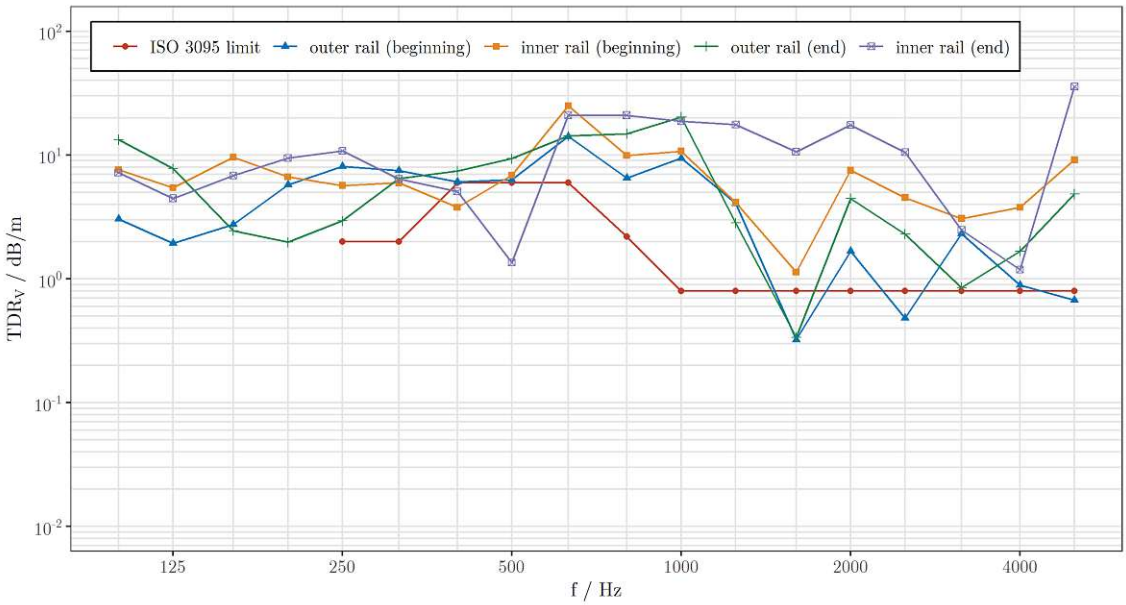

Figure 20: Vertical track decay rate measurements in C4 on 15th April 2016 and on 10th January 2017

3.1.6 Wien Heiligenstadt, S45 (C5)

The fifth measurement campaign – called C5 throughout the thesis – was carried out on the same line as C4 in Vienna (km 8.7) in Austria. An impression of the measurement section is shown in Figure 21. The track is situated on a high dam with a steep decrease on both sites. On the site of the microphones, private garden areas of housing blocks are located and on the opposite side a public park is next to the track. Two microphones in the same distance were used (note that the distance only refers to the measured track 2):

- M1: 7.5 m from the axis, 1.2 m above rail level
- M2: 7.5 m from the axis, 1.2 m above rail level

The measurement data for evaluation are only taken from M1, whereas M2 had only the function of validating the values of M1. For detailed information about the track parameters, refer to Table 8. The measurements lasted from 26th January 2016 till 28th January 2017. Rail roughness and track decay rate (AEIF method) were monitored at the beginning and at the end of the measurement period. The rail roughness of the inner rail exceeds the ISO 3095 [10] limit significantly in the wavelength spectrum between 5 cm and 16 cm (see Figure 22). The outer rail satisfies the ISO 3095 [10] limit nearly over the whole wavelength spectrum. Over the measurement period a slight decrease in rail roughness on the outer rail takes place, while the inner rail shows a decrease in the low wavelength area and an increase in the higher wavelengths. The lateral track decay rate drops only at 1.6 kHz under the ISO 3095 limit at the beginning of the measurement period on both rails (refer to Figure 23). One year later the ISO 3095 [10] limit is fully satisfied on both rails. The vertical track decay rate of the inner rail fulfils the ISO 3095 [10] limit in both measurements, whereas the measurements on the outer rail point out a drop under the limit between 500 Hz and 800 Hz at the beginning and slightly between 500 Hz and 630 Hz at the end of the measurement period (depicted in Figure 24). The change in track decay rate over the measurement period is minimal. As it is the same line like C4, the traffic consists of the same mixture.

track kilometre	8.7
curve radius	230 m
rail profile	S49
cant	90 mm - 95 mm
longitudinal gradient*	-10 ‰
superstructure	gravel
sleepers	wood
speed limit	60 kmh^{-1}
balanced speed (v_a)	42.5 kmh^{-1}

* in norm direction, see Figure 21

Table 8: Track parameters of C5, Wien Heiligenstadt, S45

Figure 21: Overview of the measurement section of C5 (© psiacoustic Umweltforschung und Engineering GmbH)

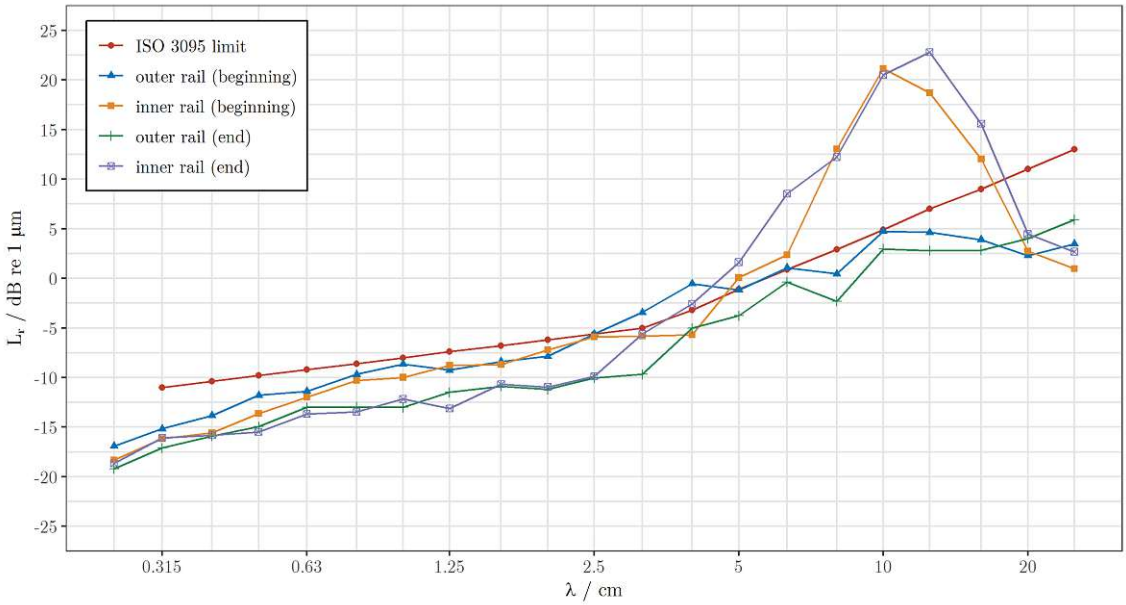

Figure 22: Rail roughness measurements in C5 on 27th January 2016 and on 12th January 2017

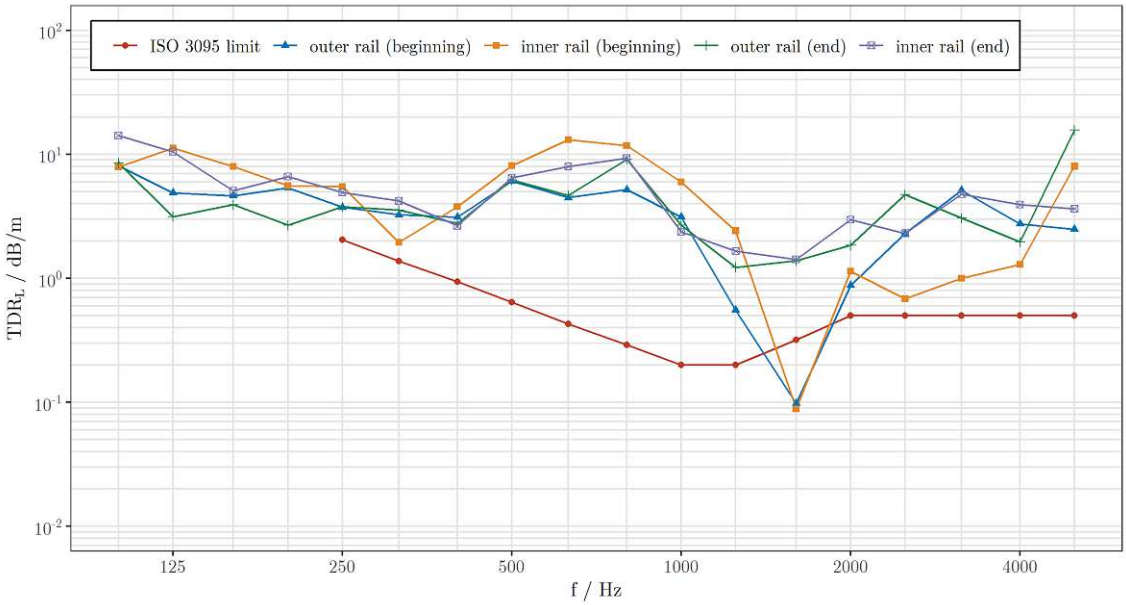

Figure 23: Lateral track decay rate measurements in C5 on 27th January 2016 and on 12th January 2017

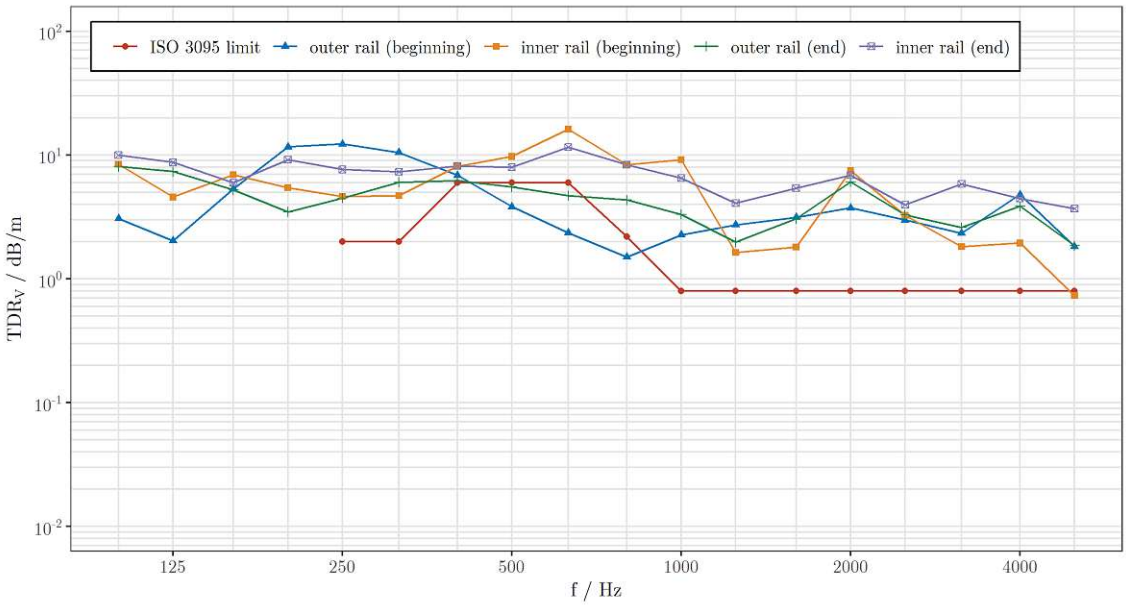

Figure 24: Vertical track decay rate measurements in C5 on 27th January 2016 and on 12th January 2017

3.2 Comparability

The measurement data were collected from five different sections with divergent topography and include various train types with variable vehicle dynamic and curving behaviour. In the following subsections, approaches to take those contrasts into account and allow evaluations with the whole dataset are described and discussed.

3.2.1 Time signal analysis

Third-octave spectra of the time signals are calculated using ArtemiS© SUITE [88] – performing a discrete Fourier transformation (DFT) with properties depicted in Figure 25. To avoid additional uncertainties, time weighting in form of the standardized method "fast" (125 ms, e.g. [10]) is not applied. The reason of standardized time weighting is usually to achieve comparable measurement data from different devices and/or setups, which is not the case in the present dataset. Thus, it is intentionally decided to diverge from the standard. Since C4 and C5 are recorded with 50 kHz sampling rate and no frequencies above the 12.5 kHz third-octave band are needed for the evaluations (see further justification in chapter 4), time signals are downsampled to 32 kHz with a resampling filter in the same software prior to performing the DFT. The outcome and basis of the most further evaluations are short-time averaged third-octave levels with a time increment length of 64 ms.

To allocate the emission to the train pass by, the recommended measurement duration from buffer to buffer in ISO 3095 [10] is chosen. The location of the buffers are set to 2.5 m before the first and 2.5 m after the last axle for every train pass by to define a standard duration and prevent train type influence. The exact cut time is calculated with the speed of the first and the last axle and in addition by adding the time difference to account the speed of sound between the train and the microphone. The latter is determined by the slope distance between axis and microphone (7.595 m) and the current speed of sound, which is calculated as follows [89]:

$$c = \sqrt{\kappa \cdot \frac{R}{M_{\text{mol}}} \cdot T} \tag{19}$$

where c is the speed of sound in ms^{-1}, κ is the ratio between specific heat at constant pressure and specific heat at constant volume (1.4), R states the universal gas constant (8.314 NmK^{-1}), M_{mol} is the molar mass of dry air (28.8 g) and T is the temperature in K [89].

To ensure that no divergences arise from time cutting, the third-octave band levels of the first and last time increment are cut energetically and added to the train pass by with a decreased time increment length.

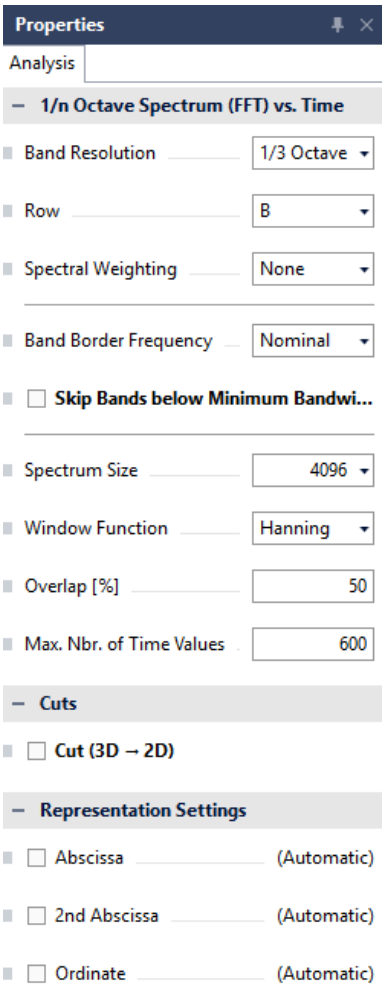

Figure 25: Chosen properties for the discrete Fourier transformation

3.2.2 Measurement sections

To allow a fast comparison of the measurement sections, key data are summarized in Table 9. Especially for taking differences of the sections into account, the following parameters are chosen and included in the further analysis:

- curve radius / m
- cant / mm
- rail profile
- sleeper material
- longitudinal gradient / ‰
- free curve breathing / mm
- wavelength spectrum of rail roughness / dB re. 10^{-6} m between 0.25 cm and 25 cm (both rails)
- frequency spectrum of lateral track decay rate / dBm^{-1} between 100 Hz and 5000 Hz
- frequency spectrum of vertical track decay rate / dBm^{-1} between 100 Hz and 5000 Hz

	C1	C2	C3	C4	C5
radius / m	256	440	310	226	230
length* / m	341	98	76	39	37
cant / mm	70	131	150	90	90-95
gradient** / ‰	11.1	-10.5	19	0	-10
rail	UIC60	54E2	UIC60	S49	S49
sleeper	concrete	concrete	concrete	wood	wood
speed limit / kmh^{-1}	60	90	80	60	60
balanced speed / kmh^{-1}	39	69.9	62.8	41.5	42.5
altitude / m a. MSL	165	376	748	217	173

* approximated value

** in norm direction

Table 9: Summarized properties of the different measurement sections

C1, C4 and C5 show similar curve radii, whereas curves in C3 and C4 have higher radii. An alteration of the curve radius may influence the curving behaviour and thus the noise emission. Moreover, as described in Subsection 2.2, an increasing curve radius is thought to decrease the probability of curve squeal significantly.

The cant is applied to compensate the centrifugal force during curving. In combination with operation speed and radius, it determines the amplitude and the direction of the resulting centrifugal force vector. Further impacts are discussed in the following subsections.

The longitudinal gradient may influence vehicle dynamic behaviour due to the need of more traction force in case of a positive gradient or vice versa.

The five sections contain three different rail types with a divergent mass per metre and thus an influence on vibration behaviour. Sleeper material has also an impact on radiated noise emission. [5]

Curve breathing means a lateral displacement of the rail due to stresses induced by fluctuations in the rail temperature around the neutral temperature (in the Austrian railway network between 20 °C and 25 °C [90]). Thus, the curve radius is narrowed (at low temperatures) or extended (at high temperatures) in the area of up to a few centimetres. Free curve breathing states the full expansion or shortening of the rail until full relaxation, which in reality is influenced and limited by superstructure properties, nevertheless allowing to take different thermal states of the rail into account. The value is calculated with the approach of Rangosch [91]:

$$\Delta f = R_0 \cdot \alpha \cdot \Delta T \cdot \left[1 - cos\left(\frac{\varphi_z}{2}\right)\right] \tag{20}$$

where Δf is the lateral displacement of the rail from the neutral state with full relaxation (free curve breathing) in mm, R_0 is the curve radius in mm, α states the thermal expansion coefficient (for steel $1.15 \cdot 10^{-5}$) K^{-1}, ΔT is the difference between rail temperature and neutral temperature (set to 25 °C) in K and φ_z is the central angle of the curve in rad.

Equation (20) follows the assumption that the rails are rigidly connected at the end of the curve, which is usually the case [91]. The central angle is calculated from curve length and radius.

Rail roughness and track decay rates are linearly interpolated and extrapolated in case of two measurements (C4 and C5) and otherwise held constant throughout the time of the measurements. Both approaches are not really realistic, since there may be complex relations causing the values to alter, however due to a lack of further data that assumption is the only choice. Regarding track decay rates, uncertainties in terms of comparability between values calculated by TNO and AEIF method have to be pointed out (see Subsection 3.1.3). As previously discussed in Subsection 3.1, single wavelengths and/or frequencies do not satisfy the normative limits of ISO 3095 [10] and may have a significant influence on the emission in certain frequencies. A research project [92] shows that differences between values can matter even though both satisfy the ISO 3095 [10] limit. Thus, inclusion of those spectra enables to consider differences in one or more third-octave band levels between the measurement sections.

3.2.3 Train types

The measurement data are dominated by passenger trains – especially commuter trains. Nevertheless, freight trains are also monitored. In Table 10, the most frequent train types are depicted. Train type A contains all sort of freight trains and thus operates with a huge variety of coaches and locomotives. Therefore, neither a number of axles, nor a standard car body length or bogie wheelbase can be stated. The same applies to train type B except that the

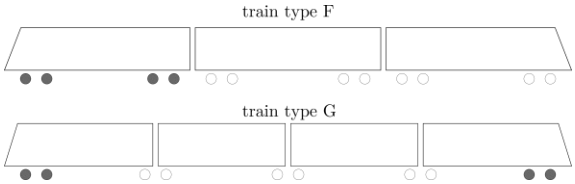

Figure 26: Schematic picture of trainsets F and G with powered bogies greyly filled and others not filled

trailers are similar in the whole train. Train type C consists of the same trailers, however the quantity – and thus the number of axles – differs. The other train types D to G operate – except a varying number of coaches in train type E – in a determined standard configuration. It has to be borne in mind, that train types D to G are also operated in double units frequently. All other train pass bys, which are not identified as A - G, are categorized as train type "other".

Train types F and G are trainsets with different locations of the powered bogies. As can be seen in Figure 26, train type F consists of a power coach and two coaches with no engine, whereas train type G is completely symmetrical powered. Thus, in the former case, it can be distinguished between pushed and pulled state, which may affect the curving behaviour. That issue refers to train types D and E as well. Influence is discussed in Subsection 6.3.

An approach to consider varying characteristics regarding curving behaviour is carried out with the inclusion of the following parameters:

- factor variable of different train types following the categorization in Table 10 A - G and "other"
- train orientation
- type of locomotive
- axles per length / m^{-1}
- bogie wheelbase / m
- angle of attack (AoA) / rad
- wheel diameter / mm
- radial steering index (RSI)

Train orientation takes different curving behaviour due to pushed or pulled state into account and can be extracted from the axle pattern of train types D, E and F. Considering the potential operation in double units, four different configurations are possible (see Figure 27). The train orientation is numbered as depicted. In case of single unit operation, the numbers 1 (pulled) and 2 (pushed) are used. Train types, where those states cannot be distinguished, are accounted with the number 0.

Train types A to E are conventionally powered with locomotives, thus the type can be evaluated from the axle pattern. The type is significant for noise emission characteristic and

train type	train category	number of axles	number of axles powered	bogie wheelbase / m	car body length buffer-buffer / m	traction	standard config-uration	
A	freight	–	4-8	powered: 2.9-3.0	–	loco	1-2 locos and trailers	[93]
B	freight (car transport)	–	4-8	powered: 2.9-3.0	–	loco	1-2 locos and trailers	[93]
C	long-distance	–	4	powered: 2.9-3.0, coaches: 2.5	26.4	loco	1 loco and trailers	[94]
D	long-distance / high-speed	32	4	powered: 3.0, coaches: 2.5	26.45-26.85	loco	1 loco, 6 trailers, 1 drive coach*	[95]
E	commuter	16-28	4	powered: 2.9-3.0, coaches: 2.5	26.80-27.13	loco	1 loco, 2-5 trailers, 1 drive coach*	[94]
F	commuter	12	4	powered: 2.5, other: 2.3	22.8-23.3	trainset	1 power coach, 1 inter-mediate coach, 1 drive coach*	[93]
G	commuter	10	4	powered: 2.3, other: 2.8 (ar-ticulated bogies)	14.75-17,54	trainset	2 power coaches, 2 inter-mediate coaches*	[96]

* also operated in double units

Table 10: Most common monitored train types

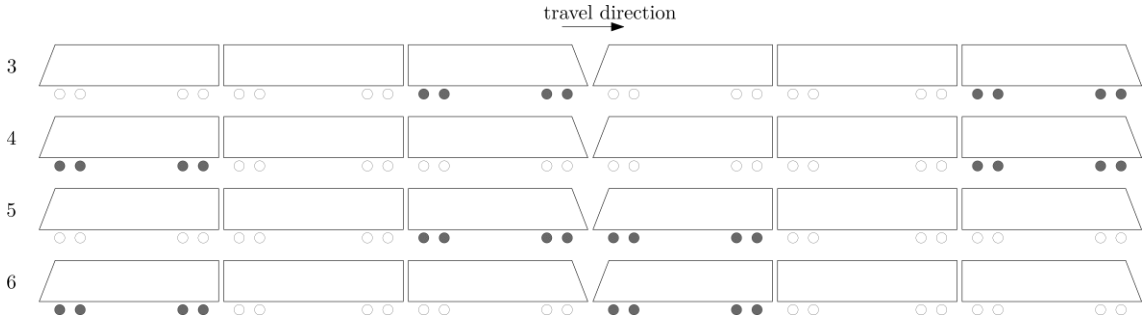

Figure 27: Schematic picture of the classification of train orientation in case of double unit operation with powered bogies greyly filled

	1116	1144	1142	2016	2043	
number in evaluation	1	2	3	4	5	
bogie wheelbase / m	3	2.9	3.4	2.7	2.5	[93]
axles	4	4	4	4	4	[93]
axles powered	4	4	4	4	4	[93]
wheel diameter / mm	1150	1300	1250	1100	950	[93]
traction	electric	electric	electric	diesel	diesel	

Table 11: Properties of the most frequent locomotive types

curving behaviour due to e.g. different bogie wheelbase, wheel diameter and engine noise. Table 11 gives an overview of the considered locomotives and important properties regarding curving behaviour and noise emission. The number of each power unit states the classification number determined by the Austrian Federal Railways. The number in the further evaluation is also depicted. In all other cases, the value is set to 0.

Since one of the main noise sources states the wheel-rail contact, a parameter to consider the number of axles dependent on the train length is meaningful – especially for freight trains due to their heterogeneous coaches. The value is derived by the ratio of the number of axles of a train and the train length (buffer to buffer).

The bogie wheelbase is an important value for curving behaviour since two wheelsets are rigidly connected in the bogie frame and prevented from orientating themselves fully radial in the curve. Without any other mechanisms, the curving behaviour gets worse with increasing bogie wheelbase [33]. Therefore, in train configurations with different bogies, the worst case is chosen for the further evaluation. In heterogeneous train sets like freight traffic, an extraction of the bogie wheelbase from the axle pattern is not possible because freight coaches can also have single axles. For those cases, the bogie wheelbase of the locomotive is chosen.

Using the described value for the bogie wheelbase, the worst case value of the AoA can be approximated by calculating the arc sine of the ratio between bogie wheelbase and curve radius, which, as discussed in Chapter 2, has a significant impact on curve squeal occurrence.

The wheel diameter influences the mass of the wheels and hence their natural frequencies. For determination, a similar approach to the boogie wheelbase is taken by always setting the value to the worst case of the whole train (mostly the locomotive). For trainsets F and G,

the wheel diameter is 950 mm and 760 mm [93].

The RSI depicts a value for deciding whether a single wheelset would be able to traverse a curve without flange contact by changing the direction only due to conicity and thus rolling radius difference or not. The calculation and a further description is provided in EN 14363 [97] and prEN 15302 [98]. For an exact calculation for every train, the wheel/contact position while curve traversing has to be measured or calculated with the worn rail and wheel profiles. Those data are not available, thus a simplified approach has to be taken. The RSI is calculated according to prEN 15302 [98] using contact geometry from ideal reference profiles wheel B (corresponding to the geometry of a common wheel in the Austrian heavy rail network) and rail A, with the following formula:

$$R_{\mathrm{E}} = r_0 \cdot \frac{2 \cdot b_{\mathrm{A}}}{\Delta r_{\mathrm{E}}} \tag{21}$$

$$q_{\mathrm{E}} = \frac{R_{\mathrm{E}}}{R} \tag{22}$$

where R_{E} states the narrowest curve radius, where free curving is possible, in m, r_0 is the wheel radius in m, b_{A} is the nominal contact point distance (for normal gauge 1.5 m) in m, Δr_{E} is the rolling radius difference at the point directly before flange contact is reached (for wheel B/rail A 0.00374 m) in m, q_{E} is the RSI and R is the curve radius in m.

In case of $q_{\mathrm{E}} \leq 1$, free curving is possible and with $q_{\mathrm{E}} > 1$ flange contact takes place. Again, it has to be pointed out that the RSI calculation is based on simplifying assumptions, nevertheless it allows to take different curving behaviour due to divergent wheel radii and curve radii into account.

3.2.4 Dynamic properties

Different vehicle dynamic behaviour while curving is considered by the following parameters:

- mean speed / kmh^{-1}
- maximum value of speed difference / kmh^{-1}
- train acceleration / ms^{-2}
- lateral acceleration / ms^{-2}

As described in the introduction of this chapter, the speed of every axle is calculated by the time difference between two axle counter sensors. The mean speed depicts the arithmetic mean of all axle speeds monitored for a train pass by. In case of braking and afterwards accelerating, the speed difference between the first and the last axle may conceal speed alterations during the train pass by. Therefore, the maximum value of the speed difference is considered. Train acceleration is calculated with the speed difference and the passing time between first and last axle.

If a vehicle traverses a curve, the centrifugal force causes a lateral acceleration, which is partly reduced by the inclination of the outer rail and depends on the driving speed. The relation for a normal gauge track is as follows [99]:

$$v_{\mathrm{b}} = \frac{\sqrt{\frac{D \cdot R}{11.8}}}{3.6} \tag{23}$$

$$a_{\mathrm{L}} = \frac{v^2}{R} - \frac{v_{\mathrm{b}}^2}{R} \tag{24}$$

where v_{b} is the balanced speed (lateral acceleration fully compensated with the inclination) in ms^{-1}, D is the inclination in mm, R is the curve radius in m, a_{L} is the lateral acceleration (if negative, the centrifugal force vector points to the centre of the curve – inclination is overcompensating, else if positive, it points in the outer direction) in ms^{-2} and v is the driving speed in ms^{-1}. The factor 11.8 is in units $\mathrm{mm} \cdot \mathrm{mh}^{-2}\mathrm{km}^{-2}$ and alters if it is not a normal gauge track.

For the calculation of lateral acceleration, the mean speed is used. The value can be interpreted, if highly positive, as forcing the wheel flange to the outer rail with flange contact, whereas highly negative values may cause the opposite – leading to flange contact on the inner rail.

3.2.5 Climate conditions

Regarding the altitude of the sections (refer to Table 9), C1, C4 and C5 show quite similar properties and also lie geographically in a distance of up to 14 km to each other. C2 and C3 are located in a valley structure – the former in the Austrian Wienerwald and the latter at the eastern beginning of the Austrian Alps with a much higher altitude than the other sections. Therefore, climate conditions are not equal, however also allowing to take divergent conditions into account. Apart from temperature and humidity behaviour due to different altitude, a highly influenced parameter is the absolute air pressure – decreasing with increasing altitude. Hence, it delivers no information in comparing, if all sections are evaluated together. Instead, the relative air pressure states a suitable value, since it depicts a reference pressure level on the same altitude (as described in Subsection 2.3).

Although the other measured parameters are also affected by climatic conditions, they are directly connected to the monitored noise emission and thus basically comparable. What should be discussed, are the properties of the used weather stations because they are different between C1 - C3 and C4/C5. For the former sections, a Professional USB wireless weather station manufactured by Conrad Electronics and for the latter ones, a Vaisala Weather Transmitter WXT520 was used. Table 12 provides an overview of differences in measurement range, accuracy and resolution – the values are taken from the corresponding manuals for C1 - C3 [100] and for C4/C5 [101].

The most significant divergences concern the rain measurement and the rain detection itself. The weather station of C1 - C3 counts rain with a resolution of 0.3 mm, which is bad, however the issue is not only the resolution, but also the rain detection, which is monitored only after an initial rainfall of 0.3 mm. Light, intermittent rain or foggy conditions are probably not monitored – therefore, the rain detection in C1 - C3 has to be critically seen. To account that uncertainty, rain data from C1 were validated and extended by measurements

		C1-C3			C4-C5		
parameter	unit	range	acc	res	range	acc	res
air temperature	°C	-40 - +65	± 1	0.1	-52 – +60	± 0.2 - ± 0.4	0.1
humidity	%	10 - 90	± 5	1	0 - 100	± 3 (0-90) - ± 5 (90-100)	0.1
absolute air pressure	hPa	300 - 1100	± 3 (700-1100)	0.1	600 - 1100	± 0.5 (0 °C - 30 °C) - ± 1 (-52 °C - 60 °C)	0.1
wind speed	ms^{-1}	0 - 44	± 0.1 (<10) - ± 10 % (>10)	0.1	0 - 60	± 3 % (at 10)	0.1
wind direction	°	0 - 360	–	22.5	0 - 360	± 3	1
rainfall	mm	0 - 9999	± 10 %	0.3	–	< ± 5 %	0.01

Table 12: Comparison of the two mobile weather stations in use

at the Vienna International Airport, which is located about 7 km southeast of the section. C2 and C3 was not validated due to the short measurement period of one month each.

Secondly, the bad accuracy of humidity values at both devices has to be borne in mind because that value is used to calculate all other wetness parameters and is observed in past research work to have a certain impact on curve squeal (see Chapter 2).

Beside different properties, data availability is also different. In C4/C5 data from the weather station were only saved in case of a train pass by, whereas in C1 - C3, a raw dataset from the weather station with 5 min time resolution was saved, in addition to the allocation to train pass bys. The former leads to gaps in the dataset, which occur mostly during the night. While for the statistical evaluation, values at the train pass by are predominantly of interest, the exact behaviour of climate conditions – especially during the night – cannot be fully depicted, which generates the need of alternative approaches regarding dew formation calculation. The latter is discussed further in Subsection 5.5.

3.3 Exclusion criteria

Measurement data have to be critically examined and validated. The further analyses with different statistical methods need another step of pre-processing, however this subsection summarizes criteria for data exclusion from the whole analyses. The outcome of the first filtering step builds the fundamental dataset, which is used in all further analyses. The scope is again to harmonize the data and improve comparability.

The following exclusion criteria are defined, as well as shortly justified:

- Operation on an other track due to not satisfying the standardized distance in ISO 3095 [10] of 7.5 m axis distance and 1.2 m above the rail level.

- Operation against the standard driving direction on the monitored track to prevent divergences due to vehicle dynamic behaviour or radiation conditions.
- During the measurement, another train is present on the other track in the section.
- The whole train length (buffer to buffer) is not fully monitored.
- A friction control system is applied to the rail head – to stay comparable with the other pass bys.
- Maximum acceleration or deceleration exceeds an absolute value of 0.2 ms^{-2}.
- Maximum absolute value of velocity difference lies above 10 kmh^{-1}
- Wind speed is over 5 ms^{-1} (according to ISO 3095 [10], an influence in the environmental noise level may take place)

Acceleration or deceleration during the train pass adds longitudinal dynamic effects, which may alter the curving behaviour as well as cause unwanted additional noise (e.g. brake squeal). Hence, a threshold value of 0.2 ms^{-2} is defined, which is thought to be a good compromise to limit additional dynamic impact, while on the other hand providing enough measurement data for the evaluation. Frequently, freight trains with large train lengths satisfy the determined acceleration limit, however the axle speeds are increasing or decreasing significantly over the whole train pass by. To prevent issues with different curving behaviour at divergent speeds in the same train pass by, a threshold value of 10 kmh^{-1} is set.

The second recommendation in ISO 3095 [10] regarding meteorological conditions, apart from wind speed, targets heavy rain as a potential influence in the environmental noise level. Measured pass bys in that condition are intentionally included in the dataset due to investigation of curve noise in rainy conditions. Nevertheless, that effect has to be borne in mind in the evaluations.

3.4 Fundamental statistics

To get an overview of the dataset distribution, basic statistical plots regarding train type counts and velocity – split into sections and/or train types – are provided.

3.4.1 Train type quantity

Table 13 contains the amount of train pass bys taken into account for further evaluation. In Figure 28, the distribution of train types in terms of the relative frequency is depicted. The quantities above the bars refer to the absolute amount of train pass bys related to that train type. Firstly, the amount of trains considered in the particular sections differs significantly. C5 contains nearly 60 % of all train pass bys, followed by C1 with over 30 %. C2, C3 and C4 share the remaining amount. While the measurement duration of C2 and C3 lasted only one month and thus the reason of the few train pass bys is obvious, C4 went on nearly a year (same as C5). However, two main reasons prevent the inclusion of more trains: (i) the

C1	C2	C3	C4	C5	sum
9314	864	479	1246	17194	29097

Table 13: Amount of train pass bys in C1 - C5

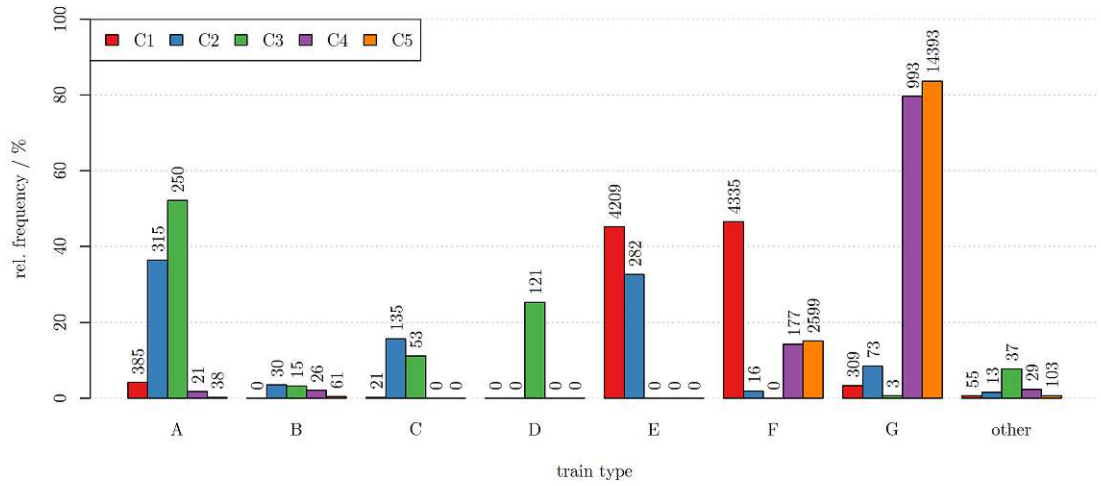

Figure 28: Relative frequency of train type occurrence split into C1 - C5

top-of-rail friction control system was operating over 80 % of the duration; (ii) due to the following station after the section, trains frequently start braking or at least not accelerating anymore and hence do not satisfy the exclusion criteria (defined in Subsection 3.3).

Secondly, train type distribution is not even. C4 and C5 are dominated by train type G (about 80 % relative frequency) and train type F – hardly any other train types are monitored. The former shows by far the most absolute quantity of all train types. Train types E and F are evenly distributed in C1, while the most other monitored types are A and G. Train type D is only present in C3. To conclude, analyses of the whole dataset is driven by the behaviour of train type E and F (C1) and train type G (mostly C5). To observe train type specific tendencies, evaluations are also carried out for each train type separately.

3.4.2 Velocity

As depicted in Subsection 3.2, the speed limit is not equal in all sections – in C2 90 kmh^{-1}, in C3 80 kmh^{-1} and in the other sections 60 kmh^{-1}. The distribution of mean velocity of train pass bys for each section is depicted in Figure 29 as an empirical cumulative distribution function (ecdf). It can be seen that in C4, trains rarely operate at the speed limit (again due to the following station, pointed out previously). The same issue, however not as severe, is present in C5. The latter is reasoned by an isolation joint in the section. Contrary, in C1, about 50 % of the trains drive at the speed limit. The velocity distribution for C3 indicates that the percentage is about 20 %. A closer look at the mean speed distribution per train type (refer to Figure 30) reveals that e.g. in C1, only train type A operates at significantly lower speed, whereas the other types drive at the speed limit frequently. C2, C3, and C5 show the

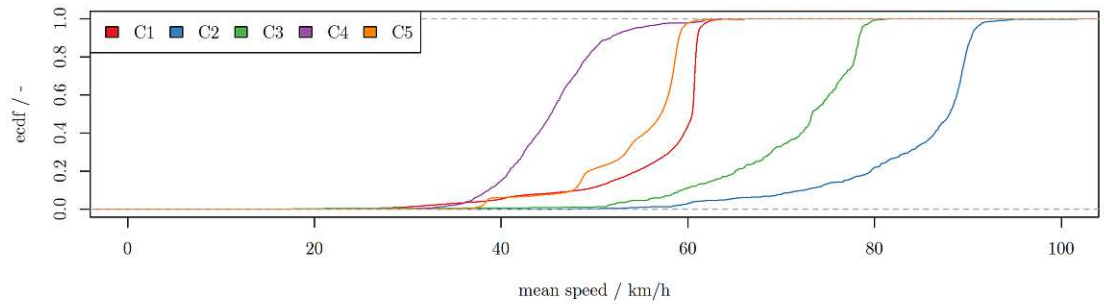

Figure 29: Mean speed distribution as a cumulative function

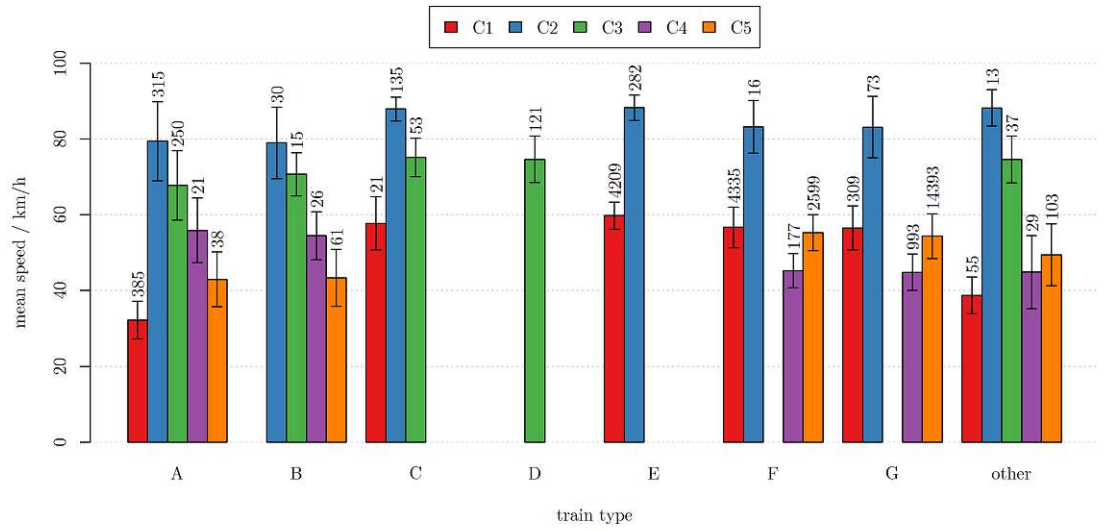

Figure 30: Mean of mean train speed with standard deviation intervals

same tendency that passenger trains operate at higher speeds than freight trains. C4 indicates the opposite, however that is again due to the following stop, which does not affect freight trains. Standard deviation intervals are generally higher for freight trains. Increasing train speed lead to higher rolling noise levels [5]. Moreover, the radiation frequency due to wheel and rail roughness changes due to the connection between wavelength spectrum and velocity – at steady roughness, the radiation frequency shifts upwards with increasing speed [5]. Thus, a comparison of noise level values without including speed as a parameter is not meaningful.

3.4.3 Timely distribution

Figure 31 shows the monthly distribution of train amount separated by C1 - C5. Measurements are always vulnerable to sensor defects or other issues that compromise the quality and therefore sorted out. Although C1 and C5 lasted over January, few train pass bys are available for evaluation. The same, however not as severe, applies to May. In August, the dataset consists almost only of pass bys from C5. The other months are well represented with at least 1500 train pass bys available. C4 contains few train pass bys after the friction

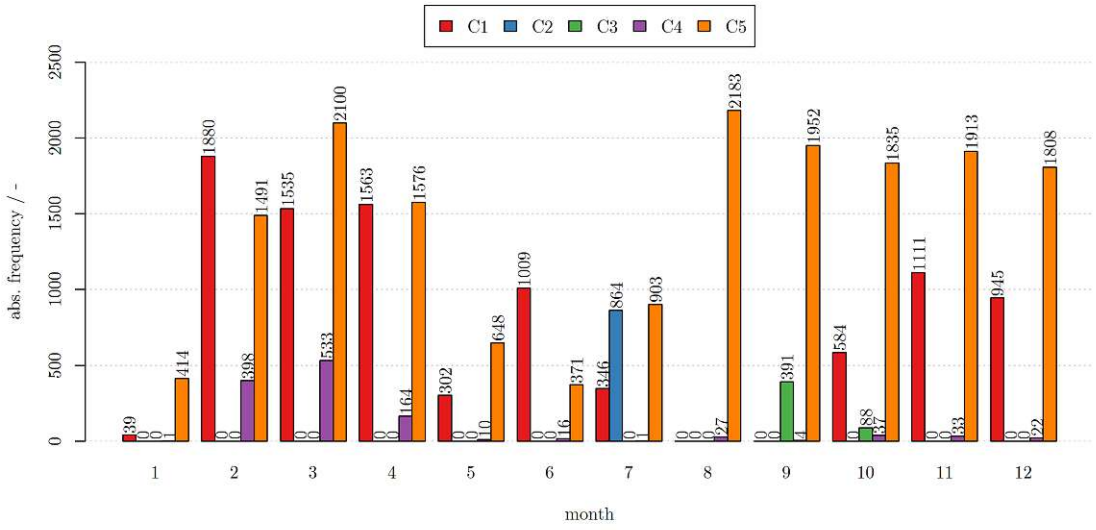

Figure 31: Amount of trains per month

control system was enabled – only in case of rain detection, no friction modifier was applied. Since the thesis focuses on investigation of dependencies regarding environmental conditions, which vary significantly over the year in all sections, monthly distribution of the dataset has to be taken into account. Moreover, since in C4 and C5, no continuous dataset of the weather station is available, data gaps also lead to no information about climate conditions over some days or weeks. The chosen solution is to look at data from nearby weather stations and interpolated grid data (further explained in Subsection 5.5).

4 Algorithm to detect curve squeal

To deal with such large numbers of measurement data, the detection of curve squeal has to be carried out automatically. Therefore, an algorithm to distinguish between train passes without or with little squeal or flanging noise (referred as "normal" pass bys) and conspicuous train passes, where squeal or flanging noise is radiated, is developed. Furthermore, the algorithm is able to differ between flanging, squeal and HF squeal noise automatically. The detection is applicable on monitored sound data in the standard point according to ISO 3095 [10] (7.5 m distance from the axis and 1.2 m above rail level). The first step of development consists of an empirical approach to find suitable evaluation parameters. Building on this, the algorithm is formulated, tested and validated. The database used for development, contains the measurement data of C1 - C3. Any depicted evaluations of train pass bys classified as radiating squeal or flange noise throughout the thesis are detected with that algorithm. An analyses of the applicability of the algorithm on measurement data from C4 and C5 was also carried out. The outcome is discussed in subsection 4.5. The algorithm was elaborated in an Austrian research project. Due to the fact that the final report of the project [9] is in German language and has not been published in English yet, the detection algorithm will be explained in detail in the following subsections.

4.1 Empirical elaboration

Because of the number of samples in C1 and the narrower curve radius compared to C2 and C3, the empirical elaboration was focused on measurement data of C1. The first evaluation was done with a random sample of 100 trains of the four most frequent passing train types (A, E, F and G), which were manually heard and categorized as normal or conspicuous by multiple project team members. Beside the audio file of the sound samples, the corresponding third-octave spectra and third-octave time functions (see Figure 32 for an example) built the foundation of the evaluation. The depicted spectrogram was not used in the manual classification, however it allows another view without the averaging process in the derivation of the third-octave bands. Due to a sampling frequency of 32 kHz, the unweighted third-octave spectra and the corresponding time gradients were determined between 10 Hz and 16000 Hz[5]. To validate the categorization, another 50 random samples from C1, C2 and C3 were investigated and categorized. In general, distinct squeal or flanging noise was clearly detected and separated from normal train pass bys. However, minimal squeal or flanging noise or overlapping squeal and flanging noise, was difficult or even impossible to distinguish. To get a better understanding of time behaviour, fluctuations of squeal and flanging noise over time were calculated and roughly assigned to positions of the train. The detected noise sources were categorized in three main categories:

- broadband high frequency noise (flanging noise)
- tonal noise with middle fundamental frequency (squeal noise)

[5]The 16000 Hz third-octave band is only calculated with components up to 16000 Hz.

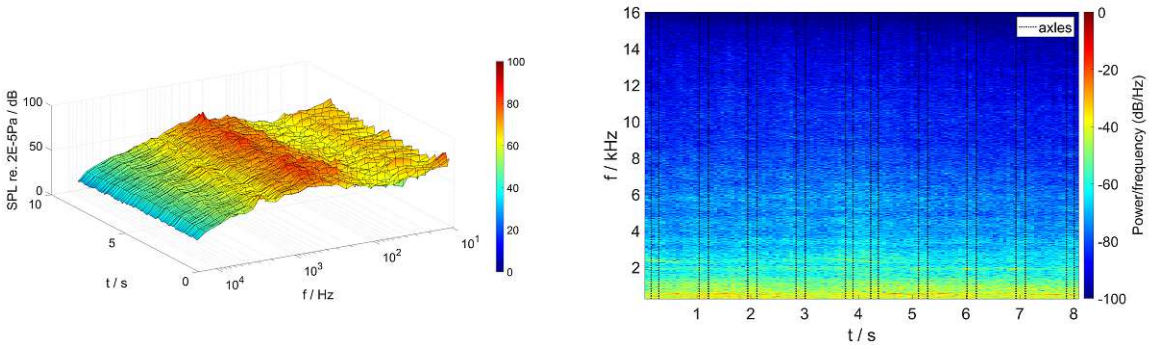

Figure 32: Third-octave time function (left) and spectrogram (right) of a train pass by classified as normal

- tonal noise with high fundamental frequency (HF squeal noise)

4.1.1 Flanging noise

The sound characteristic was described from hissing noise up to loud shearing of metals onto each other with a constant or intermittent time behaviour. As main difference to normal train pass bys, raises in third-octave levels from many kilohertz were observed – sometimes especially in the 12.5 kHz and 16 kHz third-octave bands. In addition, third-octave levels often remained in the same magnitude in the higher frequencies, whereas normal train passes showed a considerable decrease (see Figure 33 in comparison to Figure 32). The noise characteristic agrees with that described in the literature (e.g. [5]).

4.1.2 Squeal noise

The sound emission was characterized by the perceptibility of distinct tones, which were mostly not present in the whole train pass by and showed an audible increase in loudness over time until a maximum and afterwards a decrease. At the area of the maximum loudness, a slight frequency lowering was audible – indicating a Doppler effect of the moving sound source. In the third-octave time functions, elevations of one third-octave band or – if the

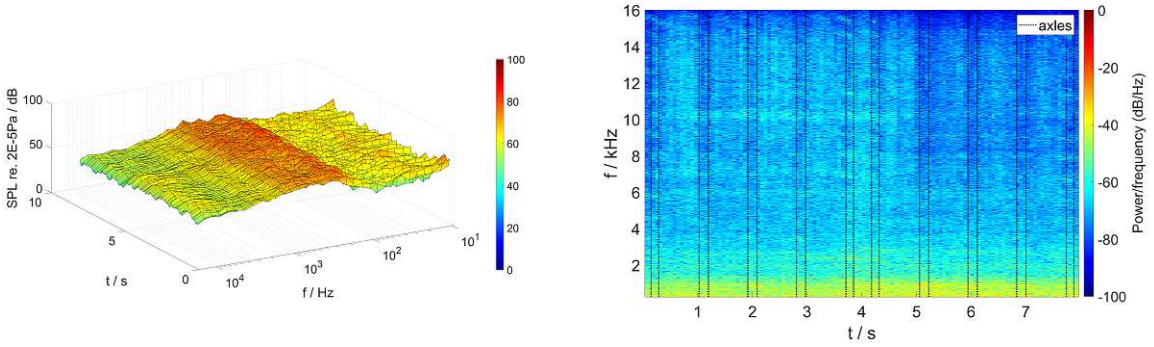

Figure 33: Third-octave time function (left) and spectrogram (right) of a train pass by containing flanging noise

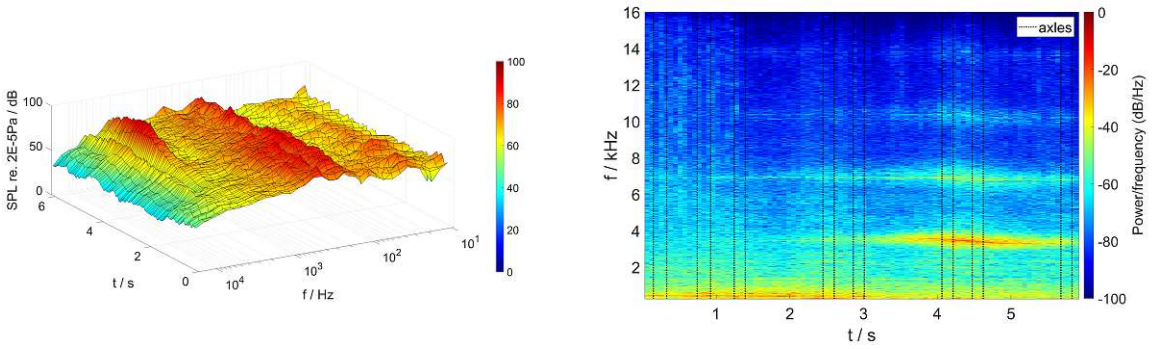

Figure 34: Third-octave time function (left) and spectrogram (right) of a train pass by containing squeal noise with clearly visible Doppler effect and harmonics

squealing frequency was near a band limit – two adjacent third-octave bands were visible (see Figure 34). The raises were normally located between the third-octave bands 1.25 kHz and 6.3 kHz with the most probabilities between 3 kHz and 5 kHz. In addition, raises with varying intensity due to harmonics were visible. The frequency range of squeal noise in the literature is wide. However, the observed frequency range lies in the area of previous publications (see Subsection 2.2).

4.1.3 High frequency squeal noise

In the third-octave bands between 8 kHz and 12.5 kHz, two forms of tonal noise were observed. Firstly, tonal components with a significantly slower time behaviour regarding level increase and decrease frequently occurred. Therefore, wheel squeal from a moving source as the reason is questionable (depicted in Figure 35 upper plots). Secondly, mostly in case of flanging noise, raises of one or two adjacent third-octave bands were visible in the third-octave time functions (see Figure 35 lower plots). Due to the overlapping, the audibility was bad and the identification of tonal sources was difficult. The third-octave view causes uncertainties with increasing frequency (due to the higher bandwidths), thus tonal components could not be identified clearly. The excitation mechanism for high frequency squeal noise remained unclear, however due to the time behaviour a connection to flange contact was assumed.

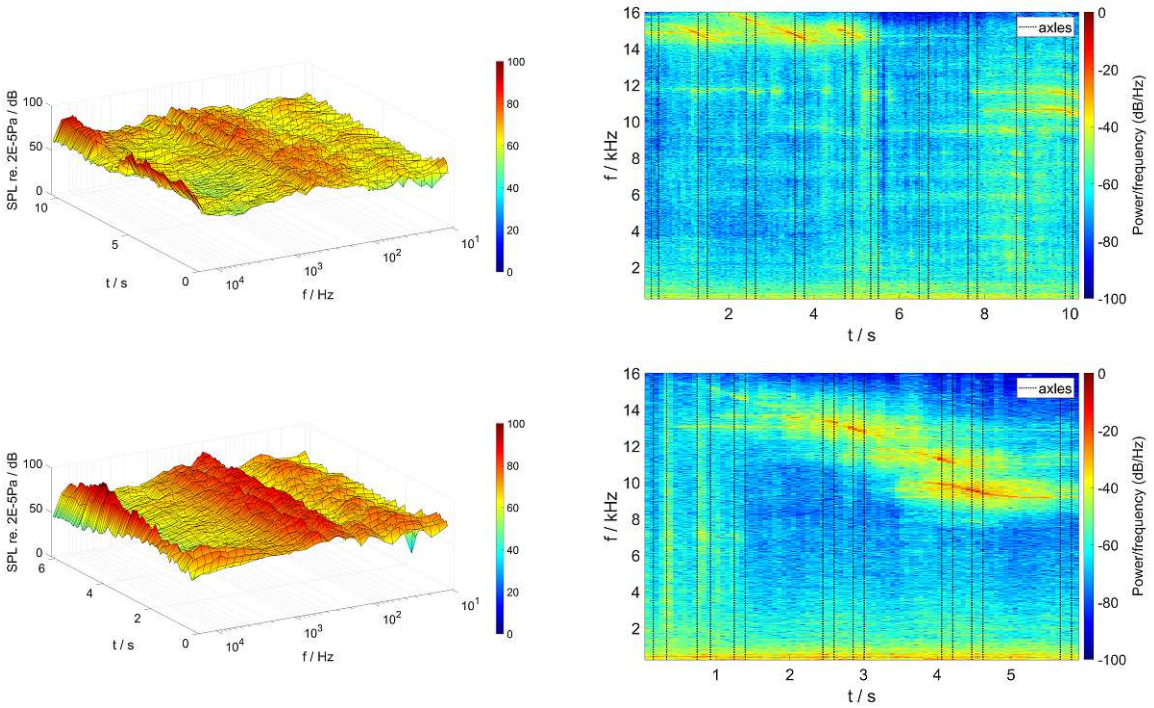

Figure 35: Third-octave time function (left) and spectrogram (right) of train pass bys containing high frequency squeal noise with constant and frequency shifting tonal components partly overlapped with broadband high frequency noise

4.2 Evaluation parameters

The principle of selecting appropriate evaluation parameters was the applicability to a random sample of train types and not to distinguish between train types and their specifications. As the main evaluation foundation, third-octave time functions were chosen. The advantages compared to third-octave spectra were the possibility to analyse slight and short-time raises clearly and distinguish between individual raises over time, which caused in addition an opportunity to assign raises to train specific positions (e.g. bogies). In the two following subsections, the selected parameters for flanging and squeal noise are described.

4.2.1 Flanging noise

Train passes classified as conspicuous regarding flanging noise showed a constant or even raising third-octave band value with increasing frequency from 2 kHz, whereas normal train pass bys indicated a significant decrease (see Figure 36). The 16 kHz third-octave level showed even for conspicuous train passes a decrease compared to the other values and was therefore excluded as a proper indicator. In addition to the growing level difference with higher frequency, the third-octave levels of the conspicuous train pass bys between 8 kHz and 12.5 kHz produced the highest arithmetic mean and median values. Building on those observations, two suitable evaluation parameters were derived:

- slope and/or third-octave levels of a linear regression line between 2 kHz or 4 kHz and 12.5 kHz

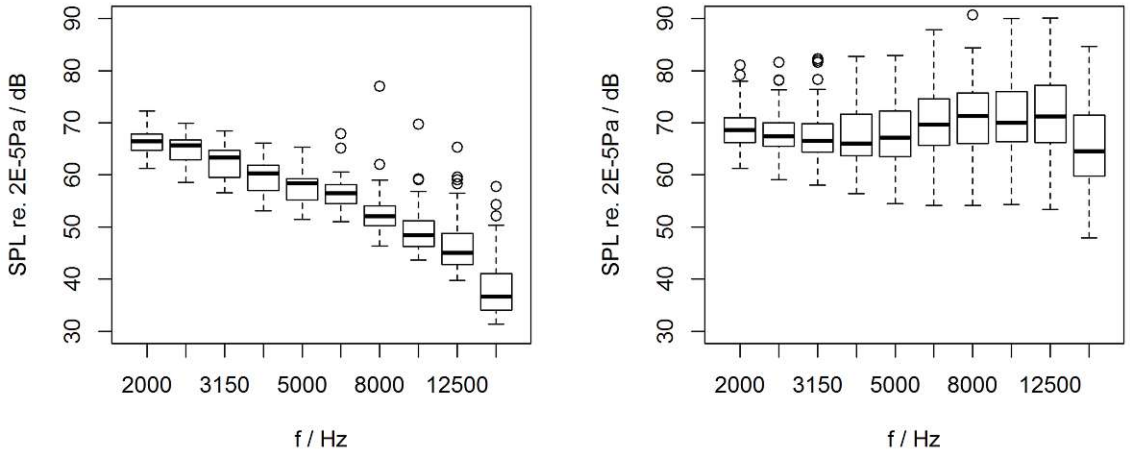

Figure 36: Box plots of the third-octave band levels from the random sample: 67 train classified as normal (left) and 124 trains categorized as containing flanging noise (right) [9]

- arithmetic mean of third-octave levels between 8 kHz and 12.5 kHz.

The slope of the regression line as only evaluation parameter was excluded because a formulation of a threshold level value is not possible. The former approach is nearly independent of outliers in the middle frequencies (e.g. from tonal components), whereas the latter is completely independent of outliers in those frequencies, however outliers between 8 kHz and 12.5 kHz can have a stronger influence. A comparison of the two approaches showed minor variances and thus the latter approach was chosen due to less computing time. As threshold values to differ between normal and flanging noise classification, 65 dB, 70 dB and 75 dB were considered. The middle value showed the best compromise between high detection rate and false detection. As an issue of the chosen approach, braking freight trains were identified. Due to the braking noise and thus the highly elevated levels in the middle frequency range, the decrease in the high frequency was visible, however the threshold of 70 dB was exceeded frequently. To solve that issue, the slope of the linear regression between 4 kHz and 12.5 kHz was implemented as additional condition to classify flanging noise for freight train pass bys. As threshold value for the slope, -15 dB per decade was defined empirically.

4.2.2 Squeal noise

Tonal components were characterised as an elevation of a single third-octave band or – if the squeal frequency is near a band limit – two adjacent third-octave bands. For the automated detection, an approach from Salz [102], which is capable of detecting raises of one and two adjacent third-octave bands, was used. In contrast to the thesis of Salz, an A-weighting was skipped to stay independent from the frequency dependent human audibility in the evaluation. The approach works as follows (refer to the illustration in Figure 37):

- An elevation of one third-octave band is present, if the value is 1.5 dB above both

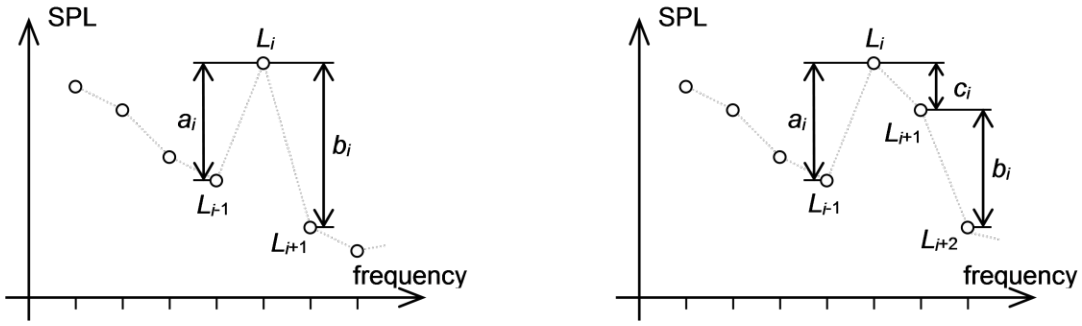

Figure 37: Approach to detect one (left) and two adjacent (right) elevated third-octave bands [9]

adjacent third-octave levels ($a_i > 1.5$ dB and $b_i > 1.5$ dB). The increase ΔL_i of the i-th third-octave band is calculated by the arithmetic mean of the difference values a_i and b_i:

$$\Delta L_i = L_i - \frac{L_{i-1} + L_{i+1}}{2} = \frac{a_i + b_i}{2} \tag{25}$$

- A raise of two adjacent third-octave bands exists, if the values are 1.5 dB above the respective neighbouring third-octave levels ($a_i > 1.5$ dB and $b_i > 1.5$ dB) and if the difference between the suspected third-octave bands c_i is smaller than the differences to the respective neighbouring third-octave levels ($c_i < a_i$ and $c_i < b_i$). The elevation ΔL_i of the i-th third-octave band is calculated by the difference between the energetic mean of the suspected third-octave levels and the arithmetic mean of the respective adjacent third-octave values:

$$\Delta L_i = 10 \cdot \log_{10}\left(\frac{10^{0.1 \cdot L_i} + 10^{0.1 \cdot L_{i+1}}}{2}\right) - \frac{L_{i-1} + L_{i+2}}{2} \tag{26}$$

In case of evaluating a third-octave band tonal elevated with both conditions, the condition resulting in a higher value of ΔL_i is determined as decisive. As mentioned previously, normal train pass bys show a significant third-octave level decrease from 2 kHz with increasing frequency and thus, the value of ΔL_i is higher with increasing frequency. To prevent an influence of that, third-octave bands with a level raise of greater or equal 10 dB are pre-selected ($\Delta L_i \geq 10$ dB) and only classified as squeal noise if their third-octave band level is greater or equal 80 dB ($L_i \geq 80$ dB). Both thresholds were elaborated empirically.

4.3 Formulation of the algorithm

The spatial resolution of the measurement data is even at higher speeds normally beneath 1.8 m, which is lower than the common wheelbase of a bogie (at least regarding passenger traffic). Thus, axle based evaluations are possible.

4.3.1 Flanging noise

Evaluation parameters for the identification of flanging noise are the arithmetic mean of the third-octave band levels between 8 kHz and 12.5 kHz and the slope of the linear regression line between 4 kHz and 12.5 kHz (for freight trains only) with threshold values of $\geq$ 70 dB and ≥ -15 dB per decade. Due to the possibly intermittent sound character, no low-pass filtering is used. To prevent the classification of short-term flanging noise, the time sum of detected time increments, in which flanging noise is detected, has to be at least 0.5 s. If that condition is fulfilled, the train pass by is classified as containing flanging noise.

4.3.2 Squeal noise

For the detection of squeal noise, the third-octave level elevations between 1.25 kHz and 12.5 kHz have to be calculated. Due to possible steep flanks at the beginning or end of squeal, no low-pass filtering is used. To prevent the classification of short-term squeal noise, the elevation for a time increment is only considered further if in the duration of $\pm$ 0.5 s, 20 % of level raises are above the determined thresholds ($\Delta L_i \geq 10$ dB and $L_i \geq 80$ dB). Afterwards, the evaluation is split in squeal noise (1.25 kHz - 6.3 kHz) and HF squeal noise (8 kHz - 12.5 kHz). In case that in one time increment various third-octave bands are classified as tonally elevated, the third-octave band with the highest level (L_i) is chosen and all other elevations are suppressed. That approach also causes to blank out harmonics, which normally have a lower level. Furthermore, if harmonics from tonal sources in the range of squeal noise are present in the range of high frequency squeal noise, they normally do not reach the threshold value of 80 dB. For the prevention of short-term squeal noise in certain third-octave bands, the time sum of the detected squeal noise in each third-octave band has to be at least 0.5 s to be classified as train pass by containing squeal noise in that certain third-octave band. In general, a train pass by is categorized as containing squeal noise, if at least one third-octave band fulfils the conditions.

4.4 Limitations of the algorithm

The formulated algorithm allows an objective evaluation of measurement data with equal parameters. However, it has to be borne in mind that a detection of all flanging and squeal noise is not possible. Furthermore, due to the evaluation of third-octave time functions, other noise sources like brake squeal or high frequency noise from the superstructure of freight vehicles, which radiate in similar frequencies or show similar behaviour, could be wrongly classified as flanging or squeal noise. Furthermore, due to the stationary sound measurement, wheels can squeal in other parts of the curve and thus may be beneath the classification threshold or not monitored at all. Building on exemplary verifications, the uncertainties of the evaluations are low, however the limitations have to be borne in mind especially regarding the interpretations (e.g. in case of few train passes).

4.5 Extension of the algorithm

An investigation showed that the algorithm is applicable on new measurement data (C4 and C5). In an extended frequency analysis of tonal noise components, it turns out, that the 800 Hz and 1 kHz third-octave bands may contain squeal noise as well. However, train type F showed tonal components in the 800 Hz third-octave band, which are velocity dependent, occur only at the power coach and show a rather different time behaviour than observed squeal noise [85]. In contrast, train type G appears to radiate squeal noise at that frequency. Thus, measurement data from C1 to C3 are also investigated regarding tonal noise in the 800 Hz and 1 kHz third-octave bands. It turns out, that train type F also shows the described behaviour in the 800 Hz third-octave band. Furthermore, train type A shows tonal components in the 800 Hz third-octave band mostly at the locomotive, which are identified as some sort of traction noise. Regarding the other train types, no tonal components in the 800 Hz third-octave band are detected. Tonal noise in the 1 kHz third-octave band shows characteristics of squeal noise and should therefore be included in squeal noise detection. Thus, the lower frequency threshold of the detection of squeal noise is lowered to 800 Hz – excluding identified tonal components in the 800 Hz third-octave bands at all in C1 to C3 and for train type F in C4 and C5 to minimize false detection of squeal noise.

5 Building the parameter set

The foundation of the parameter set consists of all monitored values in the time span of the five different measurement campaigns. Table 14 depicts an overview of directly measured parameters. Furthermore, data regarding cloudiness, sun radiation (short-wave) and snow depth are thankfully provided by the Central Institution for Meteorology and Geodynamics (ZAMG) [11] as grid data. The former is extracted from the INCA database [103, 104] with a spatial resolution of 1 km^2 and a time resolution of 15 min and the others are based on the STRAHLGRID database [105] with a spatial resolution of 0,01 km^2 and a time resolution of 15 min (sun radiation) and 1 h (snow depth). The values are from the cells, where the sections are located with no additional interpolation.

All values are linearly interpolated to the train passing times, which are determined at the start of the recording time (triggered by the first axle counter sensor), – if not already saved at the train pass by, respectively. Due to the worst time resolution of 1 h (snow depth), the impact of interpolation is thought to be low.

For validation of the measurement data, monthly time series plots from every measured parameter are examined (an example is depicted in Figure 38). Due to the limited quantity of train pass bys, an empirical examination seems more trustworthy rather than using some kind of validation algorithm. In case of single outliers, linear interpolation between the neighbour values is applied, if they lie within an hour. Otherwise or in case of consecutive compromised values, the whole data points are excluded from the further analysis. An imputation approach is skipped due to uncertainties regarding accuracy and thus maybe influencing the modelling process in a wrong direction.

In the following subsections, the derivation of all considered parameters is explained. Moreover, every parameter is mathematically described in terms of type, range of values and unit.

The variable types are described as follows [106]:

- Categorical parameters contain a finite quantity of groups. In contrast to discrete variables, they do not need to be numeric. An example would be considered train types

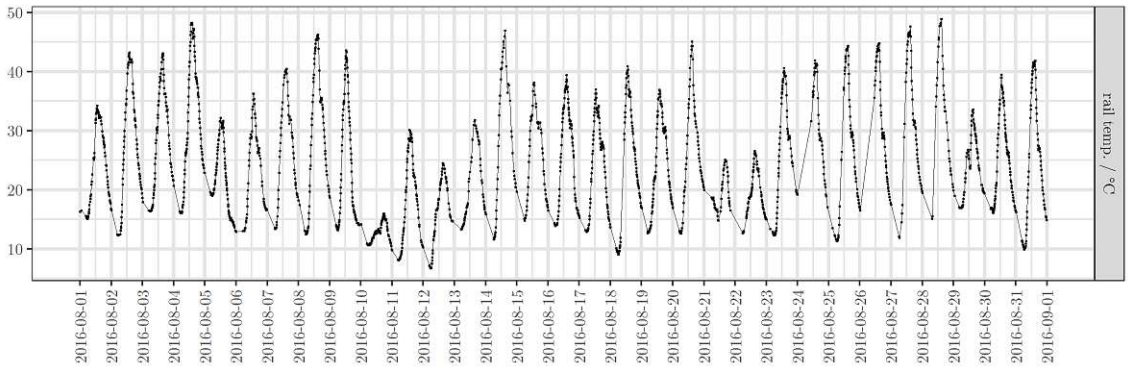

Figure 38: Example for a time series plot to validate and filter the measurement data empirically

parameter	sensor	location
axle pass by time	axle counter	mounted at the rail
sound pressure	microphone	7.5 m from track axis, 1.2 m above rail level
rail temperature	temperature sensor	rail base between two sleepers
air temperature	weather station	beside the track section
rel. humidity	weather station	beside the track section
absolute air pressure	weather station	beside the track section
wind speed	weather station	beside the track section
wind direction	weather station	beside the track section
rainfall	weather station	beside the track section
rainfall intensity	weather station	beside the track section
rainfall duration	weather station	beside the track section

Table 14: Monitored parameters during the measurement campaigns

(eight possible categories).

- Discrete variables are characterised as numerical and the amount of possible values between each other is finite – e.g. day of the year.
- Continuous variables consist of numeric values with a theoretically infinite amount of possible values between each other – e.g. train speed.

The parameters are grouped into five categories – track, vehicle, dynamic, environment and general. Derivation of those as well as for considered dependent variables is described in the following subsections. The last subsection provides a basic introduction of tendencies in the considered dataset with statistical plots, brief discussion and a detail on divergence among track sections.

5.1 Track parameters

Track parameters contain values, which are track and hence section dependent. A description is given in subsection 3.2.2. Table 15 provides the mathematical description of every parameter. Rail profiles are converted to numbers in the scale: (1) UIC60, (2) 54E2 and (3) S49. The same approach is taken for the sleeper material: (1) concrete and (2) wood. Rail roughness consists of 21 parameters per rail (wavelength third-octave spectra between 0.25 cm and 25 cm). Track decay rate contains 18 parameters for each rail and direction (third-octave band spectra between 100 Hz and 5000 Hz). Single bands are distinguished by the variable name after the second point.

5.2 Vehicle parameters

Parameters to take different train type properties into account are described in Subsection 3.2.3. The mathematical description is depicted in Table 16. An additional categorical

parameter	name in model	type	value range	unit
curve radius	R	categorical	[226;230;256;310;440]	m
cant	Cant	categorical	[70;90;92.5;131;150]	mm
rail profile	RailType	categorical	[1;2;3]	-
sleeper material	SleeperType	categorical	[1;2]	-
longitudinal gradient	LongSlope	categorical	[-10.5;-10;0;11.1;19]	‰
free curve breathing	CurveBreathing	continuous	]-∞,∞[	mm
rail roughness outer rail 0.25 cm - 25 cm wavelength	RR.OR.(0.25-25)	continuous	]-∞,∞[	dB re. 1 µm
rail roughness inner rail 0.25 cm - 25 cm wavelength	RR.IR.(0.25-25)	continuous	]-∞,∞[	dB re. 1 µm
lateral track decay rate 100 Hz - 5000 Hz outer rail	TDRh.OR.(100-5000)	continuous	[0,∞[	dBm^{-1}
lateral track decay rate 100 Hz - 5000 Hz inner rail	TDRh.IR.(100-5000)	continuous	[0,∞[	dBm^{-1}
vertical track decay rate 100 Hz - 5000 Hz outer rail	TDRv.OR.(100-5000)	continuous	[0,∞[	dBm^{-1}
vertical track decay rate 100 Hz - 5000 Hz inner rail	TDRv.IR.(100-5000)	continuous	[0,∞[	dBm^{-1}

Table 15: Mathematical description of track dependent parameters

parameter	name in model	type	value range	unit
train type	TrainType	categorical	[A;B;C;D;E; F;G;other]	-
train orientation	TrainOrientation	categorical	[0;1;2;3;4;5;6]	-
locomotive type	Loco	categorical	[0;1;2;3;4;5]	-
axle per length	Apl	continuous	[0,∞[	m^{-1}
bogie wheelbase	MaxBogieDist	continuous	[0,∞[	m
AoA	MaxAoA	continuous	[0,∞[	rad
wheel diameter	MaxWheelDm	categorical	[760;950;1100; 1150;1250;1300]	mm
RSI	RSI	continuous	[0,∞[	-
free curving possible	FreeCurving	categorical	[0;1]	-

Table 16: Mathematical description of vehicle properties

variable in conjunction with the RSI is added. In case of free curving (RSI ≤ 1), the value is set to 1, otherwise 0.

5.3 Dynamic parameters

Parameters influencing vehicle dynamics are mainly velocity and acceleration quantities as described in subsection 3.2.4. Table 17 provides the mathematical description.

5.4 General parameters

In this category, date and time variables are summarized. The information is important in terms of different seasons with alternating climate conditions. Therefore, the suitable detail needs to be investigated. The date and time information are split into numeric parameters representing meteorological seasons[6] (1 spring, 2 summer, 3 autumn, 4 winter), month (1 to 12), week (1 to 53), day of the year (1 to 365, in 2016 366) and hour of the day (0 to 23). For the latter, hour 0 for example states the time between 0:00 a.m. and 1:00 a.m. Due to application of daylight saving time in Austria, time zone between the last Sunday in March and the last Sunday in October is UTC+02 and in the remaining time span UTC+01. Since train operation also follows the change, no correction is applied to the data. That has

parameter	name in model	type	value range	unit
mean speed	Vmean	continuous	[0,∞[	kmh^{-1}
speed difference	VDiff	continuous	]-∞,∞[	kmh^{-1}
train acceleration	AccTrain	continuous	]-∞,∞[	ms^{-2}
lateral acceleration	AccSide	continuous	]-∞,∞[	ms^{-2}

Table 17: Mathematical description of dynamic quantities

[6] Meteorological seasons are defined as three month intervals of similar climatic conditions – spring (March, April, May), summer (June, July, August), autumn (September, October, November) and winter (December, January, February) [107]

parameter	name in model	type	value range	unit
quarter of the year	Quarter	discrete	[1,4]	-
month of the year	Month	discrete	[1,12]	-
week of the year	Week	discrete	[1,53]	-
day of the year	YearDay	discrete	[1,366]	-
hour of the day	Hour	discrete	[0,23]	-
hours since last train	HoursSLT	continuous	$[0,\infty[$	h

Table 18: Mathematical description of general parameters

to be borne in mind especially in conjunction with assumptions based on hours considered among the whole year. To examine an influence of longer train breaks, the hours since the last train pass by are also added. For pauses exceeding the common night break duration, hence not monitored trains in between likely, the value is set to missing. Table 18 contains the mathematical description.

5.5 Environment parameters

Many different values are derived from the measured quantities, to investigate, which parameters show significant correlations. In the next paragraphs, the calculation of each value is illustrated, followed by the mathematical description in Tables 22 and 23.

Pressure quantities

As previously stated, the absolute air pressure is not suitable for comparing sections with different altitudes. Nevertheless, the quantity is included into the evaluation because many derived parameters depend on that value. In C1 - C3, the barometer delivered no information, thus a backwards calculation approach is taken. As absolute air pressure input, the monitored value of the nearest stationary weather station in the ZAMG station network is used. Table 19 provides information in terms of distance and altitude. Equation (9) is used to derive the absolute air pressure at the section, which is equal to the calculated relative air pressure at the altitude of the section. The water vapour term is slightly influenced by absolute air pressure (by the enhancement factor, see Equation (12)), which is solved by iterating the enhancement factor until the difference in partial water vapour pressure is ≤ 0.001 hPa. The approach is validated by application to C4 and C5 and comparing the computed value with the measured one at the section. The uncertainty shows a magnitude of $\pm$ 3 hPa, which is on the one hand large because common daily fluctuations are in that range, however on the other hand the accuracy of the used barometer in C1 - C3 would have had the same uncertainty interval (see Table 12). Thus, the approach is chosen to derive absolute air pressure in C1 - C3.

The relative air pressure allows a comparison of air pressure quantities over all sections and is derived with Equation (9) projecting all air pressure values to MSL. The conversion between geometric and geopotential altitude is included with Equation (10). Since it is a computer-based calculation, the addition effort is negligible, however it would not be necessary in terms of accuracy at low altitudes.

section	name	air distance / km	altitude of barometer / m a. MSL
C1	Wien Unterlaa	4.53	200
C2	Buchberg	6.64	467
C3	Mürzzuschlag	4.04	709
C4	Wien Hohe Warte	2.74	209
C5	Wien Hohe Warte	0.54	209

Table 19: Key data of the nearest ZAMG weather stations of each section

Water vapour pressure quantities are saturated (derived with Equation (11)) and partial water vapour pressure (calculated with Equation (13)). Both values are valid for moist air in the whole occurring temperature range. Since the friction coefficient at the wheel-rail contact is of importance, all computed pressure values are calculated with air and rail temperature separately to take the conditions directly above the rail into account. In case of the relative air pressure, only different temperature and water vapour pressure are taken into account.

Humidity quantities

In addition to the measured value of the relative humidity, absolute and specific humidity as well as mixing ratio is calculated by using the Equations (16), (17) and (18). Due to the importance of conditions directly above the rail, the humidity quantities are also calculated with air and rail temperature separately.

Temperature quantities

Beside the measurements of air and rail temperature, dew and frost point (if air temperature is ≤ 0 °C, otherwise set to missing) are derived with Equations (14) and (15). Moreover, to investigate the influence of temperature and dew point further, the differences between air temperature and dew point as well as the same with rail temperature are added as two separate parameters – representing additional relative humidity indicators.

Rain quantities

As basic parameters, rain intensity during train pass bys and a categorical variable, if it is currently raining or not, are considered. Furthermore, rainfall and rain duration since the last train pass by are added to illustrate wetness conditions. If train pauses exceed the common night break duration, values are set to missing.

5.5.1 Dew formation

Dew formation has a significant impact on the friction coefficient and is common in terms of local climate conditions of all five sections. As introduced in Subsection 2.3.4, the convective latent heat flux density can be used to derive dew formation and evaporation rate (in general a mass flux rate of water). The fluctuations between air and rail temperature play a key role

in that matter. In all sections, rail temperature information is only available at the time of a monitored train pass by. While that issue is also the case in terms of air temperature in C4 and C5, C1 - C3 provide air temperature measurements beside the section in a time resolution of 5 min. Basically, the missing information is not a problem for other parameters and the model building process, however regarding dew formation, which occurs mainly between sunset and sunrise, where larger train pauses are common, information in a more detailed time resolution is necessary. There are four possible approaches to achieve that:

- linear interpolation of temperature quantities between train pass bys
- regression
- solving the heat balance equation (refer to Equation (4))
- FE-modelling of a short track section

For a complete solution of the heat balance equation, the conductive heat flux density between rail and sleeper has to be calculated, which requires the sleeper temperature near the contact point. Since no measurements of the sleeper temperature are available, this approach is not considered further. In general, the rail material has a high thermal conductivity (about 43 $Jm^{-1}s^{-1}°C^{-1}$ [78]), which leads to a homogeneous temperature distribution. The conductive heat transfer between rail and sleeper causes at hot temperatures that the rail base is cooler at the sleeper, however the rail head temperature distribution is nearly uniform [108]. Between the sleepers, the rail temperature distribution depends on the contact area with the ballast. Due to a homogeneous temperature distribution and an additional uncertainty added by the rail pad between the rail and the sleeper, which may have an insulation effect, a complex, numerical approach like FE seems not necessary and may lead to wrong assumptions.

The remaining approaches are investigated further. For regression models, predictor parameters have to be determined. Considering the heat balance equation, which can be formulated for every surface, the convective sensitive, convective latent and net radiation heat flux density can be calculated with air temperature, cloud cover, sun radiation (short-wave), wind speed and relative humidity. In the first modelling step with all five mentioned parameters, cloud cover and wind speed do not seem to have an impact on rail temperature and thus are sorted out. By far the most importance belongs to air temperature and depending on the model humidity or sun radiation shows a minor importance score. Necessary model dependent pre-processing steps are taken (refer to Table 25). The whole dataset is used for evaluation and is split in a training set (75 %) and a test set (25 %) prior to modelling. Performance metrics of the test set are depicted in Table 20 containing the best performing model, the performance of a neural network as best performing non-linear model and for comparison the benchmarks achieved by a linear regression model. As can be clearly seen, the relationship between predictors and outcome is not linear. The random forest model neglects sun radiation completely. Although the best performance is reached by that model, the estimated values are far from realistic (see Figure 39). Predicted values from the tree model show high fluctuations around the original values increasing and decreasing rapidly in the range of two to three

model	RMSE	R^2
Random Forest	2.30	0.96
Neural Network	2.54	0.95
Lin Reg	3.5	0.91

Table 20: Performance metrics of different regression models evaluated from the test set

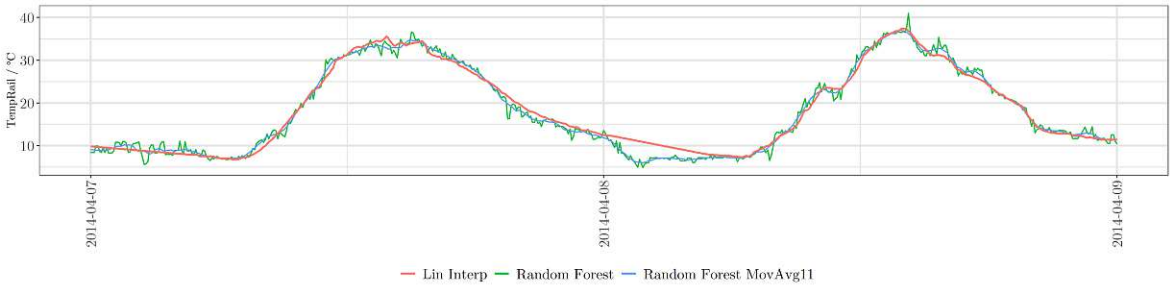

Figure 39: Example time series of 5 min linearly interpolated rail temperature (orange) and estimated values with a random forest regression model (with and without moving average)

degrees. That can be improved by applying a moving average. For the present task eleven averaged values are used to achieve a very good estimation. While that issue is observable in every considered tree or rule-based model, other models do not show that behaviour, hence they can be applied without a moving average. Further theoretical explanation about the whole model building process, applied steps and considered models is given in Subsection 6.1.

The second approach consists of linear interpolation between the train pass bys to get a time series with a time resolution of 5 min. Breaks above seven hours are filtered and set to missing due to uncertainties in the applicability of linear interpolation in case of longer time spans. Figure 39 depicts an example of a time series over two days and allows a comparison between linearly interpolated values and estimated values from the random forest model. It can be seen that during the night breaks the linearly interpolated function seems realistic. Since the train operation begins early in the morning, the frequent minimum value of the rail temperature during sunrise is monitored. A linear decrease of the rail temperature during the night is an assumption, however the tendency of heat loss due to emitted long-wave radiation is plausible and also quite corresponding to the regression path. Therefore, the simple approach of linear interpolation is chosen and used for dew formation calculation. Nevertheless, in terms of neglecting the necessity of rail temperature measurements for future applications, the regression approach is useful as well[7].

Extensive research is conducted in terms of dew formation for agricultural applications such as plant canopies or crop fields, where dew can play an important role as a water resource (e.g. [109]). Moreover, dew harvesting as water supply for villages in dry areas is investigated (e.g. [110]). Specific dew formation models for rails, especially the rail head, are missing. In contrast to approaches for agriculture, where soil is the basic surface, dew harvesting is investigated on artificial surfaces with low-density polyethylene (LDPE) or polymethylmethacrylate

[7]The final model object in R for the random forest regression is available at https://github.com/M-Ostermann/curve-noise-prediction

(PMMA) sheets [75], which, like steel, show no water storage potential, however completely different properties. Richards [76] adapts a leaf wetness model to estimate dew formation on a roof surface with asphalt shingles. The main difference in the derived models is the used convective sensible heat transfer coefficient, which is proportional to the mass transfer coefficient. Vuollekoski et al. [75] summarize different calculation approaches and show that the variance is quite large – growing with increasing wind speed. Monteith and Unsworth [74] provide a detailed physical explanation about the topic as well as a calculation approach. To elect a final model approach, three different models are applied for deriving the dew formation on rails:

- Model 1: exact calculation after Monteith and Unsworth [74]
- Model 2: dew collection model on a LDPE sheet [75]
- Model 3: dew formation model on a roof surface [76]

The selection is justified by the need of models for artificial surfaces, which are rare, as well as the criterion of model validation by field measurements. The calculation in all models is valid for a wet surface – in case of a dry surface, no evaporation can take place.

For the derivation of Model 1, the characteristic length of the surface is estimated with 10 cm, which is a good approximation for a wide range of wind directions on the rail head. The rail head is assumed as a horizontal flat plate for calculation of free and forced convection, which is a viable approximation for most surfaces [74]. The whole calculation is illustrated in Appendix A.

As input data, the measured quantities absolute air pressure, relative humidity and air temperature are also linearly interpolated to obtain a time series with 5 min time resolution. Afterwards, dew point and the values for specific humidity and water vapour saturation pressure are computed. Additionally, values for the specific heat capacity of air and the latent heat of vaporization are needed. The former is derived as follows [111]:

$$c_\mathrm{a} = 1002.5 + 275 \cdot 10^{-6} \cdot (T - 200)^2 \tag{27}$$

where c_a depicts the specific heat capacity of air in $\mathrm{Jkg^{-1}K^{-1}}$ and T states the temperature in K.

The latter is calculated with the following equation [109]:

$$L_\mathrm{v} = [2.501 - (2.361 \cdot 10^{-3}) \cdot t] \cdot 10^6 \tag{28}$$

where L_v states the latent heat of vaporization in $\mathrm{Jkg^{-1}}$ and t is the temperature in °C.

Model 2 and Model 3 use the same convective sensible heat transfer coefficient, which is derived as follows (valid at wind speeds beneath 5 $\mathrm{ms^{-1}}$) [76]:

$$h_c = 5.9 + 4.1 \cdot u \cdot \frac{511+294}{511+T_a} \quad (29)$$

where h_c states the convective sensible heat transfer coefficient in $Wm^{-2}K^{-1}$, u is the wind speed in ms^{-1} and T_a is the air temperature in K.

In Model 2, the convective water vapour coefficient is calculated with the following equation [76]:

$$h_v = 1.07 \cdot \frac{L_v}{c_a} \cdot h_c \quad (30)$$

where h_v is the convective water vapour coefficient in Wm^{-2} and 1.07 is a dimensionless ratio partly derived from molecular diffusion rates in the laminar boundary layer.

To compute a mass transfer rate in Model 2, the following formula is used [76]:

$$\frac{dm}{dt} = h_v \cdot \frac{s - s_r}{L_v} \cdot 3600 \quad (31)$$

where $\frac{dm}{dt}$ is the dew formation rate in mmh^{-1}, s_r is the specific humidity of the rail at saturation in $kgkg^{-1}$ and s depicts the specific humidity of the air $kgkg^{-1}$.

Model 3 uses the following equation to compute a mass transfer coefficient [76]:

$$k = \frac{0.622 \cdot h_c}{c_a \cdot p} \quad (32)$$

where k is the mass transfer coefficient in sm^{-1} and p states the absolute air pressure in Pa.

The mass transfer rate in Model 3 is derived as follows [76]:

$$\frac{dm}{dt} = k \cdot (e_{sat,d} - e_{sat,r}) \cdot 3600 \quad (33)$$

where $e_{sat,d}$ is the saturated water vapour pressure at the dew point temperature and $e_{sat,r}$ states the saturated water vapour pressure calculated with the rail temperature. For both values, the enhancement factor is included.

The dew formation rate is converted to the mass of dew formation in a time step by taking the arithmetic mean of two neighbour values and multiplying it with 0.083 h (corresponding to 5 min). In case of a negative sign, evaporation takes place. A cumulative sum of dew formation and evaporation allows an estimation, in which time increments dew is present on the rail. Wheel passes do not remove the viscose layer formed in dew conditions (refer to Subsection 2.1). However, further possible effects of wheel passes like displacement mechanisms (maybe leading to a thickness reduction), air flows during the pass by or impacts of water spray from the wheels add uncertainties, which are not researched yet. Figure 40 allows a comparison of the three derived model outcomes in terms of a violin plot[8]. Model 1 shows

[8]Violin plots combine conventional box plots with an underlying density function for estimating data point distribution. The filled black bar states the interquartile range with a white point representing the median.

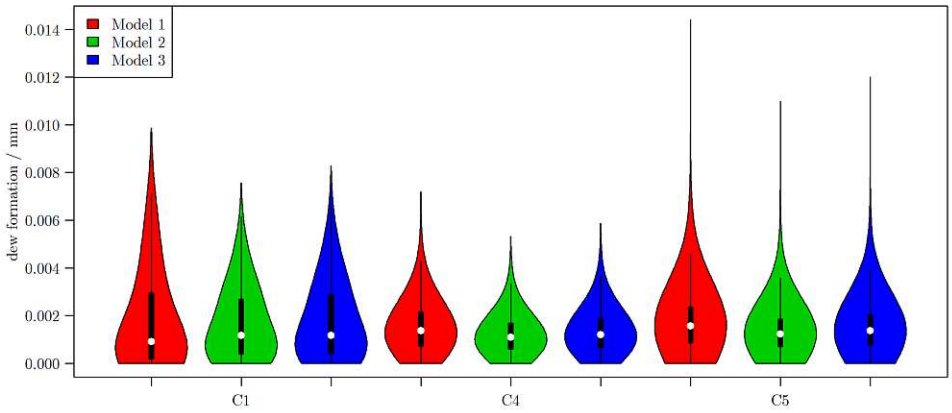

Figure 40: Violin plot for C1, C4 and C5 to compare the three models for dew formation

models	rel. frequency / %
1	33.0
2	34.2
3	35.0
simplified	25.9

Table 21: Relative frequency of dew presence on the rail for all sections

the largest dew formation maxima, followed by Model 3 and Model 2 the least. However, the distribution is quite similar, which points out that on the one hand, the assumptions for Model 1 are valid and on the other hand, the empirical approaches of Model 2 and Model 3 are applicable for approximating dew formation on rails reducing the calculation effort significantly. The comparison to the simplest approach of deriving dew presence on the rail in every condition, where the rail temperature is beneath the dew point, shows the importance to take evaporation into account (depicted in Table 21) – leading to a significant underestimation of dew presence. The calculated models obtain a similar relative frequency. However, for further calculation Model 1 is chosen on behalf of two reasons: (i) the model allows to exchange geometry and characteristic length freely; (ii) the calculation is valid for any wind speed.

In addition to the computed dew formation and evaporation in 5 min time steps, the cumulative sum of dew accumulated on the rail is calculated. If the value changes the sign to negative – giving the information that all dew is evaporated –, it is set to zero. Moreover, to investigate dew formation further, another variable, which expresses only dew formation is added (in case of evaporation the quantity is set to zero). All described parameters are linearly interpolated in the time domain to illustrate the conditions during the train pass by. In case of 5 min time steps, the interpolation uncertainty is low. A categorical variable to express, if dew is currently present (cumulative dew sum is above zero) on the rail or not is also considered. As described in subsection 2.3.4, hoarfrost can also accumulate on the rail as well as ice if dew or rain is present and the rail temperature drops beneath the freezing

parameter	name in model	type	value range	unit
absolute air pressure	Pressure	continuous	[0,∞[	hPa
air temperature	TempAir	continuous	]-∞,∞[	°C
rail temperature	TempRail	continuous	]-∞,∞[	°C
relative humidity	Hum	continuous	[0,100]	%
relative rail humidity	HumRail	continuous	[0,∞[	%
dew point of the ambient air	DewPoint	continuous	]-∞,∞[	°C
frost point of the ambient air	FrostPoint	continuous	]-∞,∞[	°C
wind speed	WindAbs	continuous	[0,∞[	ms^{-1}
wind direction	WindDir	continuous	[0,360[	°
rain intensity	RainMmH	continuous	[0,∞[	mmh^{-1}
rain duration since last train	RainDurSLT	continuous	[0,∞[	min
rainfall since last train	RainMmSLT	continuous	[0,∞[	mm
rainfall during train pass by	RainDetect	categorical	[0;1]	-
cloudiness of the sky	Clouds	continuous	[0,100]	%
sun radiation (short-wave)	SunRad	continuous	[0,∞[	Wm^{-2}
snow depth	SnowCm	continuous	[0,∞[	cm
relative air pressure	PressureRel	continuous	[0,∞[	hPa
relative air pressure at the rail	PressureRelRail	continuous	[0,∞[	hPa

Table 22: Mathematical description of environmental parameters part 1

point. Each mechanism is taken into account by a separate categorical variable. If hoarfrost or frost is likely on the rail, it is assumed that no liquid dew is present.

5.6 Dependent variables

The main scope is to investigate influences of the listed parameters on noise emission in general and especially their impact on curve squeal. Thus, suitable dependent variables have to be considered. The current guideline for noise emission calculation for a railway track in Austria (RVE 04.01.02 [71]) considers third-octave bands between 50 Hz and 10000 Hz. The calculated third-octave spectra from the monitored time signals (for methodology refer to subsection 3.2.1) range from 8 Hz to 12.5 kHz (neglecting the 16 kHz third-octave band due to the sampling rate of 32 kHz). Third-octave bands below 50 Hz are calculated with few frequency lines from the Fourier transformation and thus contain high uncertainties and fluctuations. The 12.5 kHz third-octave band usually contributes negligibly to the whole noise emission, however, as pointed out in Chapter 4, in case of flanging or HF squeal noise, it can be on the same level as other third-octave bands. Nevertheless, for comparability to the current noise emission calculation method, only third-octave bands between 50 Hz and 10000 Hz are used to calculate noise level quantities. An exception is made in case of variables, which are especially for the investigation of flanging or HF squeal noise dependencies.

The foundation of the computed values is built on 64 ms short-time averaged third-octave

parameter	name in model	type	value range	unit
saturated water vapour pressure	SatPressure	continuous	[0,∞[	hPa
saturated water vapour pressure at the rail	SatPressureRail	continuous	[0,∞[	hPa
partial water vapour pressure	PartPressure	continuous	[0,∞[	hPa
partial water vapour pressure at the rail	PartPressureRail	continuous	[0,∞[	hPa
absolute humidity	HumAbs	continuous	[0,∞[	gm^{-3}
absolute rail humidity	HumAbsRail	continuous	[0,∞[	gm^{-3}
specific humidity	HumSpec	continuous	[0,∞[	gkg^{-1}
specific rail humidity	HumSpecRail	continuous	[0,∞[	gkg^{-1}
mixing ratio	MixRatio	continuous	[0,∞[	gkg^{-1}
mixing ratio at the rail	MixRatioRail	continuous	[0,∞[	gkg^{-1}
dew formation and evaporation on the rail	DewEvapMm	continuous	]-∞,∞[	mm
dew formation on the rail	DewMm	continuous	[0,∞[	mm
cumulative sum of dew formation and evaporation	DewCumMm	continuous	[0,∞[	mm
dew present on the rail	DewPresent	categorical	[0;1]	-
hoarfrost present on the rail	HoarfrostPresent	categorical	[0;1]	-
ice present on the rail	FrostPresent	categorical	[0;1]	-
difference between air temperature and dew point	AT.DPDiff	continuous	[0,∞[	°C
difference between rail temperature and dew point	RT.DPDiff	continuous	]-∞,∞[	°C

Table 23: Mathematical description of environmental parameters part 2

parameter	name in model	type	value range	unit
equivalent continuous noise level	Level	continuous	[0,∞[	dB re. 20 µPa
peak flanging noise third-octave band level	BroadLevelMax	continuous	[0,∞[	dB re. 20 µPa
peak squeal noise third-octave band level	TonalLowLevelMax	continuous	[0,∞[	dB re. 20 µPa
peak HF squeal noise third-octave band level	TonalHighLevelMax	continuous	[0,∞[	dB re. 20 µPa
relative flanging noise time	BroadTimeRel	continuous	[0,100]	%
relative squeal noise time	TonalLowTimeRel	continuous	[0,100]	%
relative HF squeal noise time	TonalHighTimeRel	continuous	[0,100]	%
flanging noise occurrence	Broad	categorical	[0;1]	-
squeal noise occurrence	TonalLow	categorical	[0;1]	-
HF squeal noise occurrence	TonalHigh	categorical	[0;1]	-

Table 24: Mathematical description of considered dependent variables

band levels, which are considered between buffer and buffer for every train pass by. The algorithm in Chapter 4 identifies, if a time increment shows squeal and/or flanging noise. If the time sum among a train pass by is above 0.5 s, it is evaluated as containing squeal and/or flanging noise. The third observed mechanism – termed as HF squeal noise – is also investigated, although the excitation mechanism is not entirely clear. This information is considered as three categorical variables. Moreover, for examination of squeal and flanging noise behaviour, the time of squeal and flanging noise occurrence is included. To get rid of train speed and length dependency, the time is normalized by the passing time. The third chosen parameter for squeal and flanging noise investigation is the magnitude of the event in terms of a 64 ms short-time averaged peak third-octave level. For squeal noise, peak third-octave levels are considered between 800 Hz and 6300 Hz, for HF squeal and flanging noise peak third-octave levels are evaluated between 8000 Hz and 12500 Hz. For alterations in overall noise emission, the energetic sum of all considered third-octave levels and energetic mean over all time increments is added to the evaluation (referred as equivalent continuous noise level). Table 24 provides the mathematical description of the dependent variables.

5.6.1 Preparation of subsets

Although many track and vehicle parameters are taken into account to reach comparability in the whole dataset, the influence of including data from different sections and train types needs to be investigated. Thus, the following evaluation datasets are built:

- The whole dataset with all sections and train types together (referred as "all")
- The dataset is split into each train type, nevertheless considering all sections (called "TT")
- Separation of each section containing all operating train types (called "MC")
- Partitioning of each train type in each section (referred further as "MC_TT")

That approach leads in theory to 53 subsets, however not every train type operates in each section. Thus, the quantity is reduced to 41 (see Figure 28). To account statistical significance, subsets with less than ten train pass bys are excluded in advance. Categorization in terms of statistical significance is done by defining three levels:

- strong – over 500 train pass bys in the subset
- mediocre - between 100 and 500 train pass bys in the subset
- weak – under 100 train pass bys in the subset

5.7 Tendencies and anomalies in the dataset

To get a basic understanding of the dataset, especially of the behaviour of dependent variables in different measurement sections and in conjunction with distinct train types, some basic statistic plots are provided and discussed. Furthermore, one significant deviation is investigated in detail. Following the predefined limit for statistical significance, all categories with less than ten train pass bys are not depicted. In addition to illustration of sample quantity in each category (separated in measurement section and train type), median values with interquartile ranges are shown. The latter is chosen due to the decibel scale in the majority of discussed values as well as using mainly one statistic value throughout the analysis for simplicity. Moreover, median is generally more robust in terms of outliers compared to arithmetic mean. Stated deviations are always based on observed median differences. In terms of probability values, absolute frequency of occurrence and sample size are added to the plot for better clarity. For evaluation of equivalent continuous noise level, which contains trains with and without curve squeal, plots are split into one dataset without curve squeal and another dataset with all trains containing any type of curve squeal (including flanging noise, squeal noise and HF squeal noise). To avoid a large set of pictures in the main part of the thesis, only graphs of equivalent continuous noise level and frequency of occurrence of curve squeal types are depicted. Plots involving all other dependent variables are provided in Appendix B and are interpreted as well as referenced in the current subsection.

Equivalent continuous noise level (referred as Level) distribution of trains without curve squeal (illustrated in Figure 41 upper graph) indicates that train type G shows very low values (in C1 84 dB and in C4/C5 77 dB). Freight train category A shows similar values compared to train type B – both among the highest median levels with values around 90 dB. However, train types C, E and F radiate emissions with levels in the same range as train type A (in C1 and C2). In C4 and C5, train type F shows a median level of 83 dB. In terms of measurement sections, significant differences between C1 and C4/C5 are observable. Although radii are similar with 256 m compared to 226 m/230 m, median level differences of train types F, G and also the category "other" are in the range of 6 dB to 7 dB. That issue is further investigated in Subsection 5.7.1 below. Moreover, train speed seems to have little impact on median level values since regarding passenger trains between C1, C2 and C3 an even controversial trend of lower levels with increasing speed limit is observable. Only train type A shows a minor raise of 1 dB between C1 and C2/C3, where speed differences are 40 kmh^{-1} and 50 kmh^{-1} (freight trains in C1 drive at mean speeds of only 40 kmh^{-1}, refer to Figure 29). Following the approach of Thompson [5] the magnitude should be over 8 dB (A-weighted) higher – that approach is further discussed in Subsection 6.2. Detailed investigation shows that the rail roughness in C1 compared to C2/C3 is very high (on the inner rail), which might mask the speed influence as well as delivers an explanation of equal level values among train types A, C, E and F in C1. Nevertheless, regarding trains with very low rolling noise level (like train type G), a difference is present. Level magnitude of train pass bys containing curve squeal (refer to Figure 41 lower graph) are 1 dB to 3 dB higher than trains without curve squeal – in terms of freight traffic roughly equal. An exception is train type G, where the increase in C4 and C5 is 5 dB. That can be explained by its very low basic level, which is more increased in conjunction with curve squeal peaks. The magnitude of level difference between C1 and C4/C5 in case of curve squeal occurrence is slightly reduced to 3 dB - 6 dB, which indicates that median level values are impacted by track properties despite curve squeal appearance.

Flanging noise probability (referred as Broad – depicted in Figure 42) indicate a negative correlation to curve radius. However, statistical significance (at least weak) in all sections is only given regarding freight traffic and tendencies are partly controversial. The general pattern in freight traffic is a decrease in probability between curve radii $\leq$ 256 m in relation to values $\geq$ 310 m (reduction of 10% - 20%). In the former range, probability values between C1 and C4/C5 indicate a controversial tendency – increasing between 226 m and 256 m. In the latter area, similar values are achieved in C3 (310 m) and C4 (440 m). Apart from train type C in C1 (with weak statistical significance), the same behaviour is observable regarding passenger train categories. However, the pattern deviates among train types. Train type F shows an increase between C5 and C1 of 13 %, while train type G indicates the opposite behaviour with a decreasing probability of 10 % (that issue is further discussed in Subsection 6.3.2). Both show a significantly reduced magnitude in C2. Hence, a detailed statement of passenger train behaviour cannot be given, however, significant decrease in larger radii (440 m) is strongly indicated. Probabilities of flanging noise are

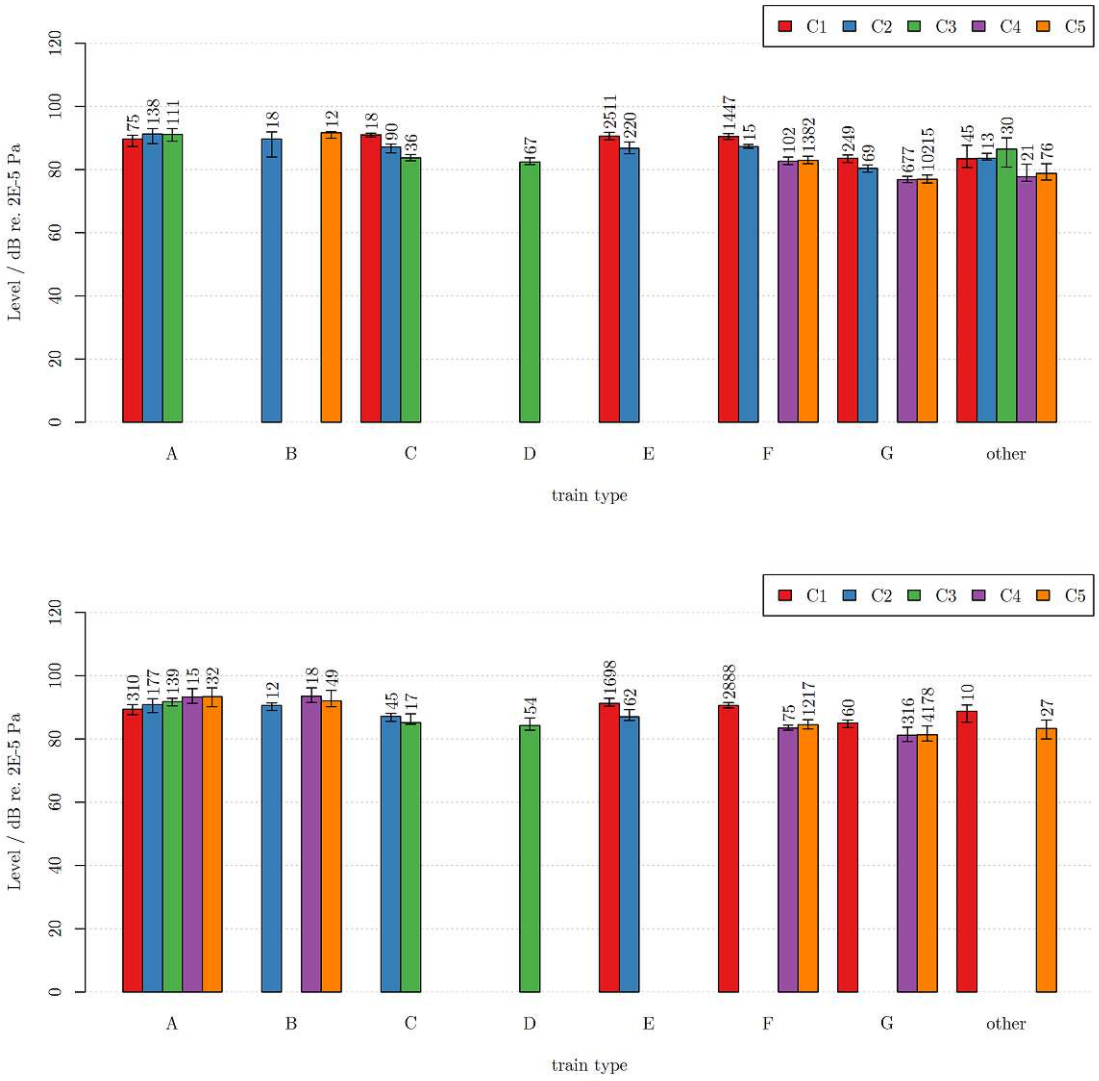

Figure 41: Median values and interquartile range for the dependent variable Level with datasets without any curve squeal (upper graph) and containing any type of curve squeal (lower plot)

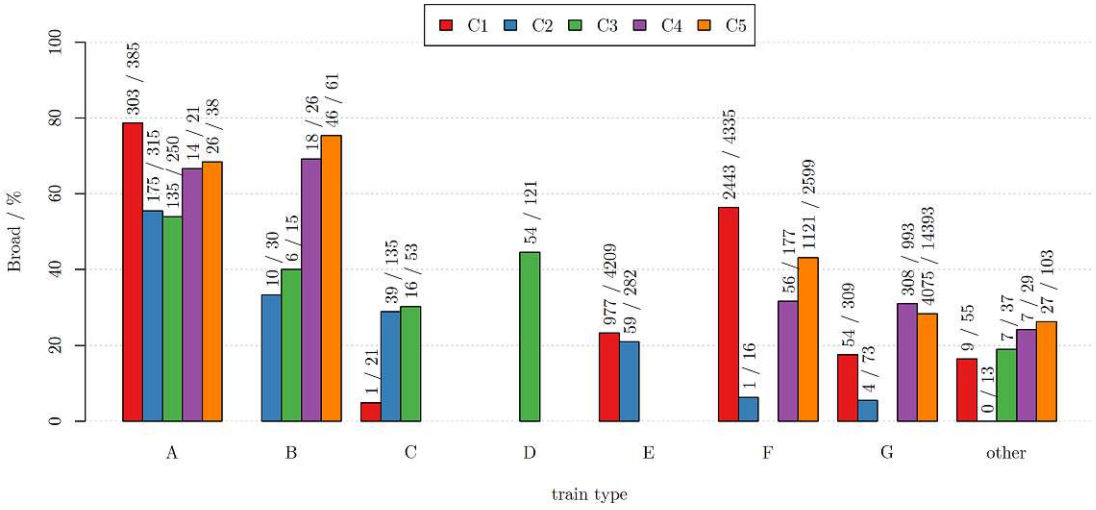

Figure 42: Relative and absolute frequency of flanging noise occurrence

clearly train type dependent – ranging from 55 % to 79 % for freight traffic, 32 % to 56 % for train type F and 18 % to 31 % for train type G. Other passenger trains show values between 21 % and 45 %, which depicts presumably a common probability range in passenger traffic regarding curve radii up to 440 m. Thus, train type tendencies (at least distinction between freight and passenger traffic) have to be taken into account in probability estimations.

Evaluation of short-time averaged peak third-octave band levels of flanging noise (BroadLevelMax – illustrated in Figure 92) reveals that magnitudes of flanging noise are usually in the same range as rolling noise peak third-octave band levels (evaluated among the whole considered range from 50 Hz to 10000 Hz – not depicted separately) or even lower for train types with mediocre and high rolling noise levels (train types A, B, E and F). The opposite tendency is present regarding train type G and train types C and D in C3, which all show low basic rolling noise levels. Freight trains show 5 dB - 10 dB higher peak third-octave band levels caused by rolling noise than flanging noise. In C1, third-octave band level peaks in the low and mid frequency range are higher regarding train type F (deviation 2 dB and 7 dB) and only about 3 dB lower for train type G, while levels of train type E are roughly equal. Train type F in C4/C5 shows no deviation between flanging and rolling noise peaks, while levels of train category G are raised about 6 dB if flanging noise occurs. Peak flanging noise levels for freight trains lie between 86 dB and 92 dB in narrow curve radii ($\leq$ 256 m), while larger radii show median peak levels of 83 dB (310 m) and 81 dB - 82 dB (440 m). Train type E shows the highest median among all categories with 95 dB in C1, however, levels are much lower in C2 (78 dB). Other passenger train categories show between 85 dB and 87 dB median values. Overall, lower flanging noise peak levels are present in conjunction with larger radii (440 m) – reduction compared to narrow ones ($\leq$ 256 m) ranges from 5 dB to 11 dB (freight traffic) and is up to 17 dB for train type E as only evaluable passenger train category in C2. In general, it has to be pointed out that intentionally no

A-weighting is applied. Moreover, in a presently considered frequency range of 8000 Hz to 12500 Hz (third-octave bands), significant weakening due to atmospheric absorption takes place (refer to ISO 9613-1 [112]). Thus, those peak levels quickly loose relevance in a sound propagation calculation with growing distance between source and immission point.

Investigation in duration of flanging noise (BroadTimeRel – depicted in Figure 93) allows an estimation of the contribution to overall radiated level. In all measurement campaigns, freight train pass bys with detected flanging noise show an occurrence over the whole passing time (100 %). However, in terms of freight traffic, flanging noise alone causes no significant raise in overall Level (median increase about 1 dB). Regarding passenger train types, median relative occurrence times lie between 27 % and 55 % (for radii $\leq$ 310 m), however the quite large interquartile range of about 30 % has to be pointed out. Significantly lower percentages are observed in C2 (440 m curve radius) with median values of 15 % and 18 % (interquartile range 8 % and 12 %) for train types C and E, which indicates a radius dependency at least for passenger trains.

Regarding squeal noise probability (TonalLow – illustrated in Figure 43), curve radius influence is observable. Train pass bys in C2 show significantly lower probability values – ranging from 2 % to 7 % for freight trains and from 0 % to 4 % for passenger traffic. Freight traffic indicates a distinct negative correlation to curve radius with values 67 %/79 % (226 m/230 m), 17 % (256 m), 9 % (310 m) and 2 % (440 m) for train type A and 62 %/79 % (226 m/230 m), 33 % (310 m) and 7 % (440 m) regarding train type B. Distribution among curve radii between 226 m and 310 m in other subsets indicates no clear pattern. Strong train type dependencies are present. In C4/C5, freight traffic shows extreme squeal noise probability (62 % - 79 %), while in terms of commuter trains values of 13 % and 16 % (type F) and about 7.5 % (type G) occur (further investigated in Subsection 6.3.3). Regarding C1, probabilities for passenger train categories E and F are much higher with 18 % and 31 %, while train type G shows a lower value compared to C4/C5 (about 3 %). For the latter observation, the significantly lower statistical significance has to be pointed out. Altogether, squeal noise shows a lower relative frequency of occurrence compared to flanging noise – ranging in narrower radii ($\leq$ 310 m) from 17 % to 33 % for freight traffic and from 3 % to 18 % for passenger trains (excluding outliers).

Short-time averaged peak third-octave band levels of squeal noise (referred as TonalLowLevelMax and pictured in Figure 94) depict the severity of those events. The increase in comparison to rolling noise peak third-octave band levels ranges from 4 dB to 7 dB for freight traffic and from 5 dB to 11 dB for passenger trains. An exception is present regarding train types E and F in C1, where equality to rolling noise peaks exist, which again indicates the magnitude of track contribution in that section. Due to hardly any squeal noise incidences of passenger trains in C2 and C3, sample points are too low to reach the defined limit for statistical significance. Thus, no general statement for larger radii ($>$ 256 m) is possible. In the remaining subsets, absolute values range from 96 dB to 105 dB regarding freight traffic –

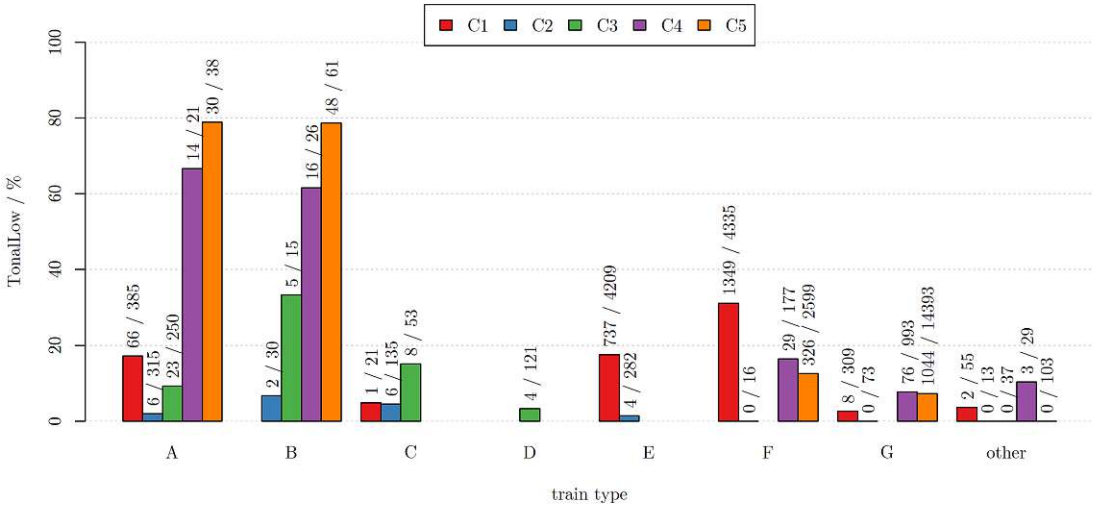

Figure 43: Relative and absolute frequency of squeal noise occurrence

showing a negative correlation to radius with values 104 dB/105 dB (226 m/230 m), 99 dB (256 m) and 96 dB (310 m). In conjunction with passenger trains, median peak levels of 89 dB to 93 dB are evaluated – showing no impact in conjunction with curve radius. An interesting discovery is that peak levels for freight traffic in C4/C5 exceed those of flanging noise by up to 16 dB. In over 60 % of freight train pass bys in C4/C5, both squeal and flanging noise is detected – hence, squeal noise peaks dominate the overall Level and completely mask the contribution of flanging noise. Regarding C1 to C3 as well as other categories in C4/C5, hardly any subset exceeds 20 % relative frequency of occurrence in terms of two curve squeal types in one train pass by. Figures 98 and 99 in Appendix B provide a complete overview of multiple detected curve squeal types in one train pass by.

Although peak levels of squeal noise incidences are severe, relative occurrence time (TonalLowTimeRel – refer to Figure 95) is significantly lower compared to flanging noise. Median values lie between 19 % and 29 % for passenger trains (sufficient data only available for train types E, F and G) with interquartile ranges between 9 % to 20 %. Freight traffic shows significant deviations among measurement sections. Relative occurrence time ranges from 20 % to 44 % (interquartile range 21% - 40%) in C4/C5, while values in C1 and C3 are with 8 % even lower compared to passenger trains in C1. Curve radius influence might be an explanation, however the difference in radius is low and would indicate a distinct non-linear decrease with growing radii. Moreover, passenger traffic shows no dependency on radius between 226 m and 256 m. Despite lower relative time occurrence compared to flanging noise, impact on overall Level can be up to 3 dB higher compared to flanging noise. However, median level increase of freight trains and all passenger train types in C1 hardly exceeds 1 dB and thus lies in the magnitude of flanging noise impact. Application of A-weighting as well as sound immission calculations would change the deviations to much higher importance of squeal noise since evaluated frequency range lies between 800 Hz and 6300 Hz (third-octave

bands).

HF squeal noise probability (referred as TonalHigh – depicted in Figure 44) shows the same pattern as squeal noise in terms of curve radius dependency regarding freight trains, however the differences between sections are more evenly distributed. Train types A and B show a negative correlation to curve radius with 52 %/47 % (C4/C5), 22 % (C1), 15 % (C3) and 9 % (C2) for the former and with 54 %/48 % (C4/C5), 7 % (C3) and 13 % (C2) for the latter. Moreover, passenger train pass bys indicate hardly any HF squeal noise incidences in C2 (0 % to 1 %). In general, passenger traffic shows no distinct curve radius dependencies for narrow curves ($\leq$ 310 m). Train type specific influence is present, however deviations are not as large as in terms of squeal noise. Train type G shows a probability value around 10 % in C1, C4 and C5, while regarding train type F 10 % (C4), 15 % (C5) and 23 % (C1) are evaluated. Other passenger train categories show magnitudes between 16 % and 21 % (data from C1 and C3), which represents presumably a good range for estimations.

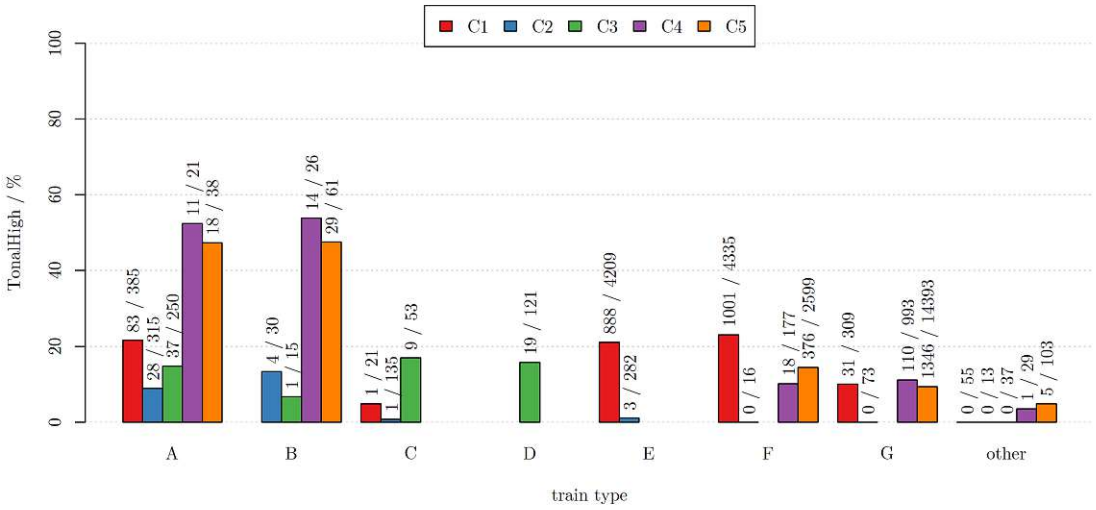

Figure 44: Relative and absolute frequency of HF squeal noise occurrence

Since for short-time averaged peak third-octave band levels of HF squeal noise (TonalHighLevelMax – illustrated in Figure 96) the same frequencies as flanging noise are evaluated, the difference between them is of interest. HF squeal noise peaks are 2 dB to 11 dB higher for all categories. Absolute values lie between 88 dB and 95 dB for freight traffic – indicating a slight radius dependency with a decrease between radii $\leq$ 310 m and 440 m of 4 dB to 7 dB. Peaks among passenger trains lie between 90 dB and 97 dB – thus slightly higher compared to freight traffic. Since hardly any HF squeal noise incidences are detected among the latter in higher curve radii (440 m), no general statement is possible. Regarding deviation to squeal noise, two different patterns are noticeable. Firstly, freight train pass bys show significantly lower peak levels compared to squeal noise in C4/C5 (reduction 10 dB to 13 dB) and slightly lower ones in C1 and C3 (reduction 4 dB). Passenger trains (only train types

E, F and G) indicate the opposite behaviour with partly higher HF squeal noise peaks compared to squeal noise in a range between 1 dB and 4 dB. The only exception states train type F in C1 with a decrease of 2 dB. Impact on Level due to HF squeal noise occurrence is roughly equal to squeal noise. However, despite peak levels can be as severe as squeal noise, application of A-weighting and significantly higher atmospheric absorption of those high frequencies (refer to [112]) in propagation weaken their impact on overall noise immission.

Duration of HF squeal noise (TonalHighTimeRel – outlined in Figure 97) lies roughly in the same area for passenger trains as the values of squeal noise with levels between 16 % and 28 % (interquartile range 14 % to 27 %). Regarding freight traffic, difference values in comparison to squeal noise are large in C4/C5 (between 7 % and 29 % lower), while C1 - C3 show minor deviations in the same range as for passenger traffic. Freight train pass bys indicate a negative correlation to curve radius with values of 13 %/14 % (226 m/230 m), 10 % (256 m), 6 % (310 m) and 5 % (440 m) for train type A (not sufficient data points for train type B). However, it has to be pointed out that statistical significance is weak. Interestingly, interquartile ranges are high in narrow radii ($\leq$ 256 m) with 13 % to 22 %, while in larger radii (310 m/440 m) they lie between 3 % and 4 %. Distribution among passenger train types and sections is quite different with no clear pattern. Train type G shows a homogeneous relative time duration of 21 % and 27 %, while the median value regarding train type F ranges between 16 % (C1) and 28 % (C5). Train type E is on the same scale as the upper limit of train type G, while train type D (only operating in C3) shows significantly lower relative occurrence time with a median value of 12 %. Thus, other influence factors than radius might be more prevalent at least in case of passenger traffic (further investigated in Appendix C.6).

5.7.1 Level difference C1 - C5

As previously described, significant level differences between C1 and C4/C5 are observed. This issue is investigated in detail in the present subsection. To avoid uncertainty of speed influence, only train pass bys between 55 kmh^{-1} and 65 kmh^{-1} are taken into account (speed limit in all three sections 60 kmh^{-1}). Moreover, train pass bys with detected curve squeal of any type are excluded. Since trains in C4 mainly drive at lower speeds (refer to Figure 29) and train quantity is also lower, only C1 and C5 are compared. Spectral properties are similar between C4 and C5. Figure 45 illustrates spectral distribution of train types F (upper graph) and G (lower plot) in C1 and C5. Median values, as well as interquartile range and one line for each train pass by in the background are depicted. As can be seen, difference spectra are similar regarding train type F and G, which strongly indicates influence of track section properties. The following characteristics are divergent:

- sleeper type (C1 concrete and C5 wood)
- rail type (mass difference C1 60 kgm^{-1} and C5 49 kgm^{-1})
- alterations in superstructure quantities – rail roughness and track decay rate

For investigation of sleeper type influence, a difference spectrum between concrete and wood sleepers is extracted from RVE 04.01.02 [71] (refer to Figure 46). The calculation is done by deriving sound power emission of rolling noise for vehicle and superstructure contribution and adding them together energetically. For the former, transfer functions for different wheel diameters – 680 mm, 860 mm, 920 mm and 1200 mm – are used and for the latter transfer functions for monoblock concrete sleepers with medium rail pad stiffness and wood sleepers are applied. All of that is provided in tables in the Appendix of RVE 04.01.02 [71]. Since difference spectra are calculated, the values are independent of total effective roughness, which depends on rail roughness, wheel roughness, train velocity and contact filter. In general, notably larger contribution (> 2 dB) of concrete sleepers in the low frequency range (160 Hz to 315 Hz) is observable. In the remaining range, sleeper type deviation is insignificant (mostly < 1 dB). Application of the difference spectrum on measurement data is illustrated in Figure 47 for train type G. For correction, 680 mm wheel diameter spectrum is used and an interpolation of the actual value (760 mm) is skipped due to negligible deviations. Changes for train type F are slightly lower and follow the same pattern. Altogether, sleeper type correction enhances the deviation of C5 and C1 in the third-octave bands 125 Hz and 160 Hz – the only range, where C5 third-octave band levels exceed the values of C1.

Variable mass of the rail alters oscillation properties and hence can influences the contribution in amplitude and frequency. However, the track decay rate monitors rail vibration behaviour directly. Thus, the deviation is covered by track decay rate comparison.

Microphone position lies on the outer side of the curve in both measurement sections. Thus, the outer rail is expected as main contributor. However, measurements of rail roughness as well as track decay rate on the outer rail deliver no sufficient explanation. A closer look reveals that deviations of properties monitored on the inner rail are the main cause, thus the contribution of the inner rail is also important and partly dominating in the present example. Rail roughness wavelength spectra of the inner rail (refer to Figure 48) for C1 (red) and both measurements in C5 (orange; the dotted line represents the measurement at the end of the campaign) show severe rail corrugation in C5 between wavelengths of 8 cm and 20 cm corresponding to frequencies at 60 kmh^{-1} of 100 Hz to 200 Hz. To estimate the impact of rail roughness deviation on total effective roughness, it is important to take the wheel roughness into account. Parameter variations in rail roughness and their effect on total effective roughness in conjunction with CNOSSOS model in Directive (EU) 2015/996 [4] are carried out in a research project [92]. The main finding is that a raise in rail roughness above the ISO 3095 limit [10] (black line in Figure 48) leads to nearly the same increase in total effective roughness over the whole considered wavelength spectrum between 25 cm and 0.25 cm in conjunction with disk braked vehicles, whereas deviations beneath the ISO 3095 limit [10] have a lower impact due to wheel roughness domination below a certain point. A change in total effective roughness directly influences rolling noise sound emission in the same magnitude. Thus, referring to Figure 48, deviations in rail corrugation in corresponding 125 Hz and 160 Hz third-octave band area are able to fully explain the higher noise emission in C5 compared to C1 in those frequencies. Moreover, in the remaining range between wavelengths 6.3 cm and 0.25 cm, rail roughness in C1 mostly exceeds ISO 3095 limit [10], while rail

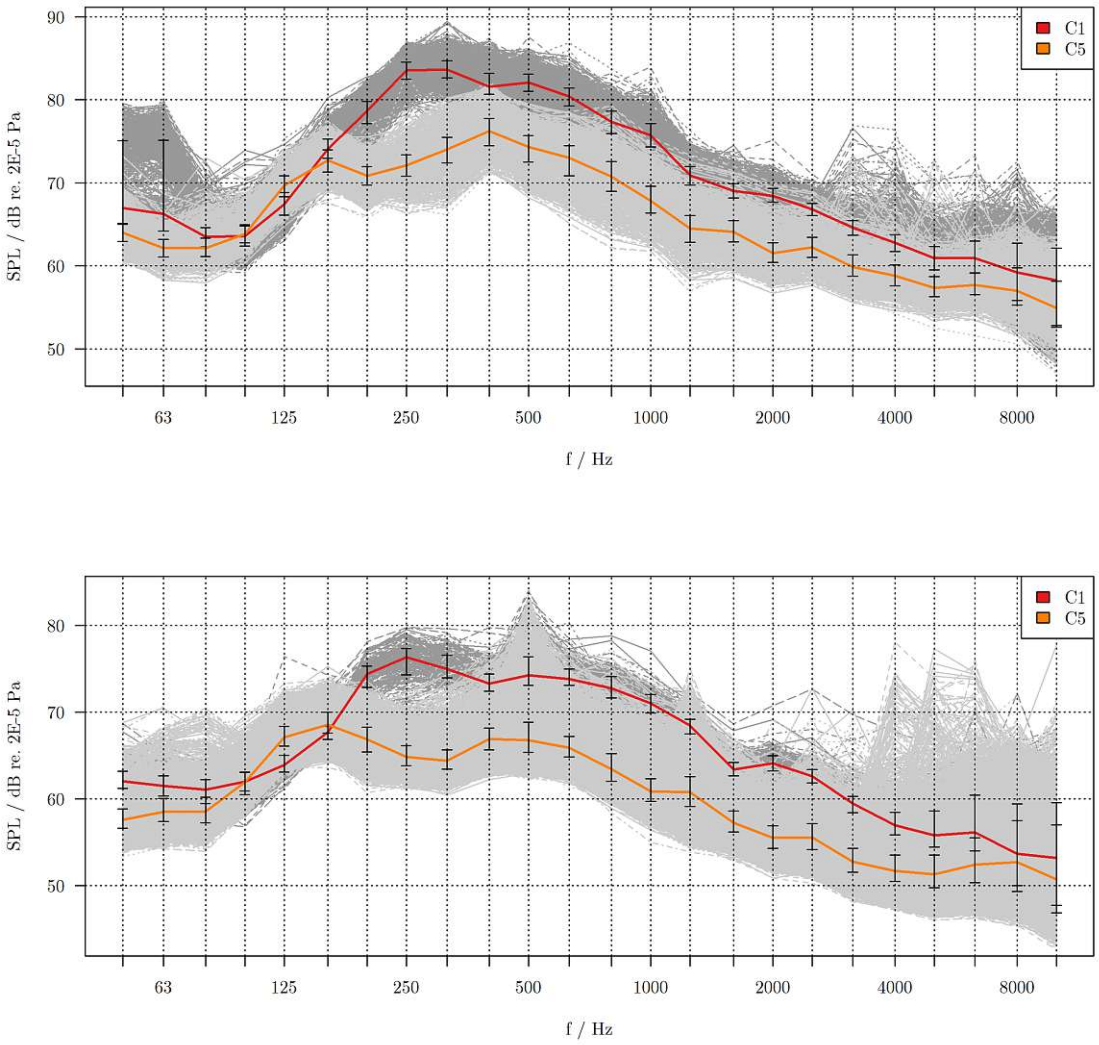

Figure 45: Third-octave band level distribution of train type F (upper graph) and train type G (lower plot) with colors for median values as indicators for C1 (red) and C5 (orange), interquartile range and grey lines in the background depicting every train pass by

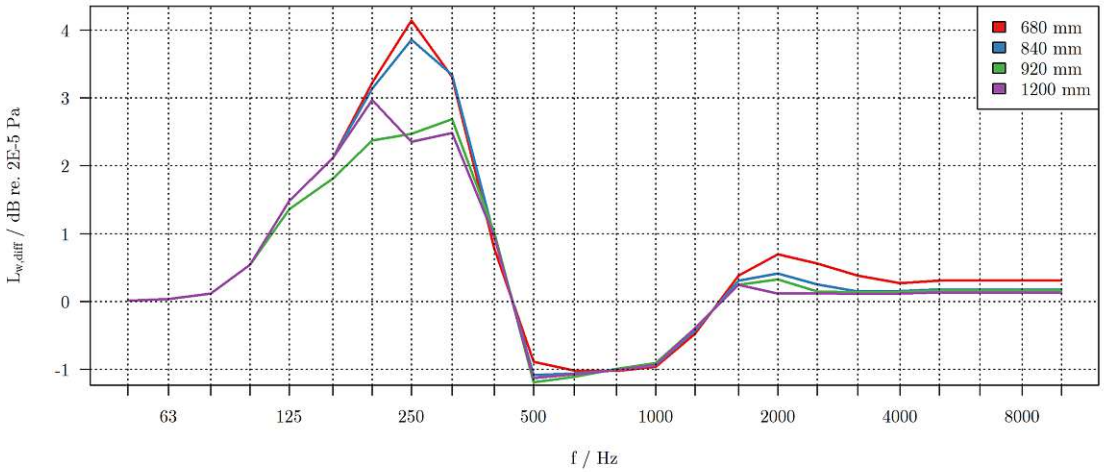

Figure 46: Third-octave band difference spectra of sound power level emission between monoblock concrete sleepers (medium rail pad stiffness) and wood sleepers for different wheel diameters – 680 mm to 1200 mm – according to RVE 04.01.02 [71]

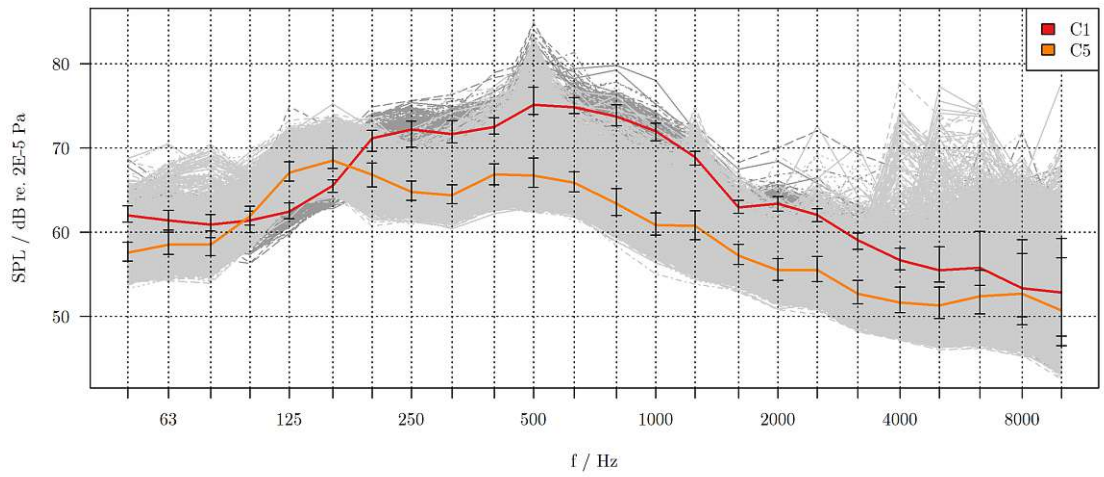

Figure 47: Third-octave band level spectrum of train type G with correction for differences in sleeper material, colors as indicators for median values, interquartile range and grey lines in the background depicting every train pass by

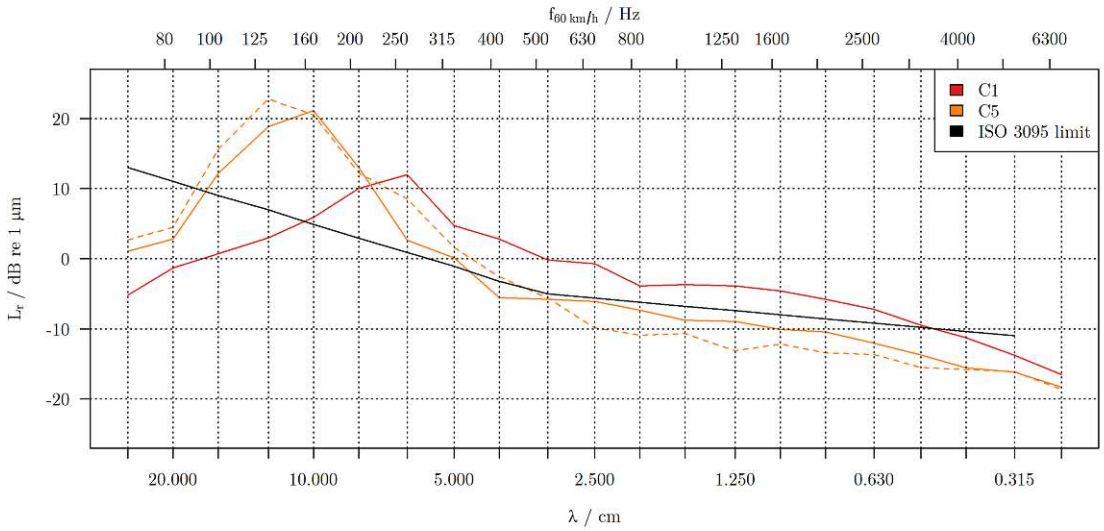

Figure 48: Rail roughness comparison between inner rails in C1 (red) and C5 (orange; dotted line for measurement after the campaign) – ISO 3095 [10] limit in black

roughness in C5 stays beneath the limit. Thus, a raise in sound emission in C1 of 5 dB to 10 dB between roughly 250 Hz and 6300 Hz is solely explainable by rail roughness difference.

Another influence factor states the track decay rate. Since it is a parameter of superstructure noise contribution, it is important to distinguish in frequency ranges, where either vehicle or superstructure emission dominates emitted rolling noise. Model calculations of CNOSSOS [4] show that between 125 Hz and 250 Hz third-octave bands, both parts are roughly equally contributing, from 315 Hz to 1600 Hz third-octave bands superstructure emission dominates and from 2000 Hz upwards vehicle sound emission states the main influence [92]. The thresholds of the mentioned ranges can slightly shift depending on wheel diameter, sleeper type and rail pad stiffness. Referring to Figure 47, it can be seen that especially between 500 Hz and 1250 Hz third-octave bands, median differences are around 10 dB and partly even exceeding. In addition to the stated influence of rail roughness, it is found that lateral track decay rate deviations are severe in that range (see Figure 49) – C5 indicates an up to 10 dBm^{-1} higher value compared to C1. Vertical track decay rates show no significant deviation. According to Thompson [5], regardless whether lateral or vertical track decay rate, a correction term of $10 \cdot \log_{10}(TDR_1/TDR_2)$ can be applied to sound emission level, which would result in correction terms in the present example of up to 10 dB. Due to dominating superstructure contribution between 500 Hz and 1250 Hz, the shown deviation in track decay rate is thought to have significant influence on overall sound emission and significant deviations in that range are explainable by differences in rail roughness and lateral track decay rate.

To conclude, rail roughness and track decay rate can have a significant influence on rolling noise, which is already well-known and further justified with the present example. Thus, those quantities have to be taken into account if data from different measurement sections are compared with each other. Moreover, sleeper material influence (only investigated between con-

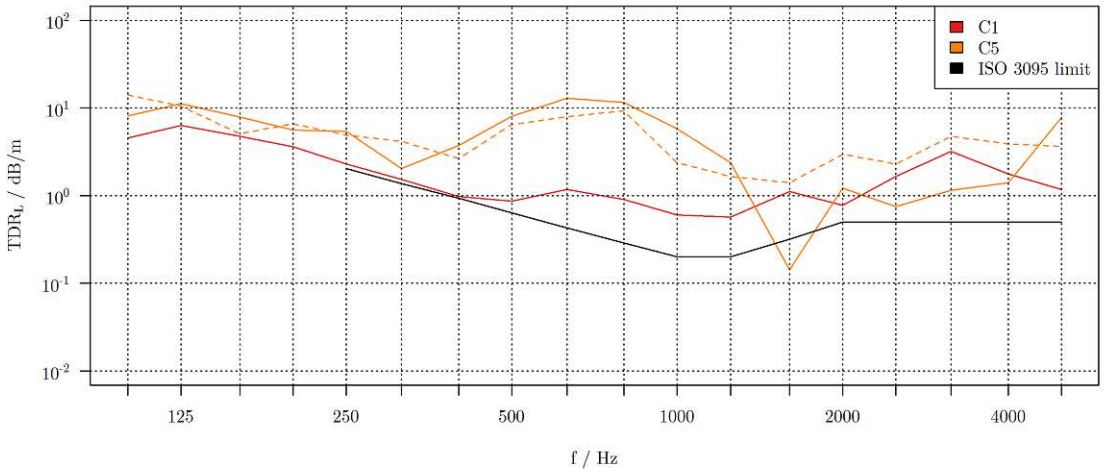

Figure 49: Lateral track decay rate comparison between inner rails in C1 (red) and C5 (orange; dotted line for measurement after the campaign) – ISO 3095 [10] limit in black

crete and wood) can impact primarily third-octave band contribution peaks between 160 Hz and 315 Hz, which are partly among the highest in emission spectra. However, in case of applying A-weighting, which is normally done for propagation and immission calculations, peaks in that range are weakened.

The stated significant deviation between C1 and C5 is statistically accounted in multivariate evaluations and thus an impact on those is not expected. Empirical examination as well as univariate statistical analysis is compromised since deviations cannot be directly pictured or taken into account in two dimensions. Thus, they are mostly focused on subsets, which do not contain data from more than one section. An impact estimation, which has to be thought of, lies in influence on curve squeal detection (algorithm described in Chapter 4). The algorithm is built on evaluation of short-time averaged third-octave levels – in case of squeal noise between 800 Hz and 6300 Hz and for HF squeal noise as well as flanging noise between 8000 Hz and 12500 Hz. Regarding squeal noise, detection relies on evaluating one or two neighboured increased third-octave bands by comparing their level with the outer neighbour values. Referring to Figure 47, it can be seen that a steep decline of median values in C1 starts at 1000 Hz. Thus, squeal noise detection between 800 Hz and 1250 Hz third-octave bands might be compromised by high rolling noise emission, whereas in higher frequency, no influence is expected. A closer look on median short-time averaged third-octave band levels of squeal noise (depicted in Appendix B Figure 96) reveals that they lie above 90 dB. Hence, impact on detection of squeal noise in the mid frequency range, even in conjunction with high rolling noise, is assumed to be minor. Detection of HF squeal noise and flanging noise is not influenced at all since deviations in high frequencies are low.

6 Feature selection and predictive modelling

After derivation of parameters and subsetting, the next step is to examine the dataset on the one hand empirically and on the other hand statistically. The former approach is predominately used to understand simple two dimensional correlations and for validation of the outcome of the statistical computation. The latter is applied to investigate multivariate dependencies in the dataset and to sort out parameters, which show no correlation with the outcome and hence may even compromise the following predictive modelling process depending on the applied model properties. The chapter is outlined by an introduction in statistical modelling, followed by an explanation of applied methodology. Afterwards two subsections dive deep into at first feature selection and afterwards predictive modelling. The chapter is concluded by a compact summary of key findings as well as limitations of the developed models in practical applications.

6.1 Introduction in statistical modelling

Statistical evaluations consist of a set of parameters (called predictors), which are connected to a dependent variable (outcome or response) by different modelling approaches. If the outcome is categorical (in the present case always two different categories), classification models are applied. Otherwise, regression models are used. Prior to applying any kind of model to the dataset, predictors have to be investigated in a pre-processing step. Evaluations of model performance and comparison of different models are carried out in the post-processing procedure by computing performance metrics. The following subsections give a brief overview of applied steps regarding pre-processing, considered performance benchmarks, different modelling approaches, feature selection techniques and post-processing strategy. Finally, computational details are provided.

6.1.1 Pre-processing

Categorical predictors are split into dummy variables, which consist of zero/one indicators to represent each category in an own vector to prevent influence of the magnitude of the assigned numeric number. If every category of the variable is represented by a separate vector, all dummy variables add up to one. Thus, some model algorithms, which include an intercept term (e.g. linear regression), are not able to solve the algorithm because the set of dummy variables contain identical information as the intercept. Therefore one category is left out and is represented if the values of all other dummy variables in that category are zero. [113]

Dummy variables are created for RailType (considering only 2 and 3 in the model), SleeperType (only using type 2), TrainType (numbering A to F with TrainType.(1-7)), TrainOrientation (leaving 0 out) and Loco (leaving 0 out). Other categorical variables are converted to numeric.

The pre-processing step is important to examine different properties of the predictor set. Statistical model algorithms show varying vulnerability to certain behaviour in the predictor

space (e.g. between predictor correlations), which can significantly worsen the performance if it is not accounted before. The following investigations can be carried out prior to the model building step in that order [114]:

1. Investigation of variance of each predictor to filter zero and near-zero variance predictors. Two criteria are used to identify near-zero variance: (i) percentage of unique values; (ii) ratio of the most frequent to the second most frequent value (frequency ratio). Standard thresholds are for the former less than 20% and for the latter greater than 20. If a predictor fulfils both conditions, it is evaluated as near-zero variance predictor. [114]

2. Evaluation of highly correlated predictors by applying a detection algorithm on the predictor set. The selection works as follows [114]:
   Steps repeated until no correlations are above the determined threshold (e.g. 90 %):

   (a) Evaluate the largest absolute correlation between two predictors;

   (b) For both predictors, compute the average correlation between each predictor and all of the other variables;

   (c) Mark the variable with the largest mean correlation for removal;

   (d) Remove this row and column from the correlation matrix

3. Examination of skewed predictor distribution – in the present case Yeo-Johnson transformation is applied because the method can deal with positive and negative predictor values (for detailed information be referred to [115])

4. Standardization[9] of predictors

5. Possibility of applying predictor reduction techniques like principal component analysis (PCA), which creates linear combinations of predictors to build an uncorrelated predictor set. For further information be referred to [116].

6.1.2 Performance metrics

For regression models, common quantitative performance measures are root mean squared error (RMSE) and the coefficient of determination (R^2 – value between zero and one). Generally speaking, RMSE is a measure for accuracy and should be minimised, whereas R^2 allows an interpretation of correlation between predicted and observed outcome and should be maximised. The foundation of RMSE states the sum of squared errors (SSE). The mean of SSE is called mean squared error (MSE). Equations are as follows [106]:

[9] Standardization consists of centring each predictor by subtracting the average value from all values leading to a mean of zero and scaling each predictor by dividing all values by the standard deviation leading to a standard deviation of one [113].

$$SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i) \tag{34}$$

$$MSE = \frac{SSE}{n} \tag{35}$$

$$RMSE = \sqrt{MSE} \tag{36}$$

where n represents the sample quantity, y_i states the observed outcome value and $\hat{y}_i$ is the estimated outcome.

The coefficient of determination depicts a quantity to estimate how much information of the data can be represented by the model. A R^2 value of 0.5 illustrates that the model is able to explain 50% of the variation in the outcome. Although several calculation methods for R^2 were derived, the standard way is to compute the correlation coefficient (R) between observed and predicted values and square it. For derivation of R, the Pearson method is applied (for further discussion be referred to [117]). [113]

In terms of modelling, there is always a trade-off between variance and bias in the model. Assuming statistical independence of data points and outcome with zero mean and constant variance σ^2, model bias and variance can be expressed as follows [113]:

$$E[MSE] = \sigma^2 + (Model\ Bias)^2 + Model\ Variance \tag{37}$$

where E is the expected value of MSE.

The first term states the alterations in variance that cannot be changed by modelling. In case of high bias, the functional form of the model is far from the true relation between predictors and outcome, whereas high variance illustrates that small changes in the data lead to significant alterations in the model fit. Thus, complex models tend to over-fitting (i.e. high variance), while simple models commonly under-fit the data (leading to high bias).

Regarding classification, predictions depend on a probability value. The simplest way of estimating accuracy is to depict a confusion matrix, where true and false positives (TP and FP), as well as false and true negatives (FN and TN) are illustrated in a 2x2-matrix. By dividing true predictions (TP+TN) by the total number of samples, an overall classification accuracy metric can be derived (further referred as Accuracy and abbreviated Acc). The problem is that the latter is influenced by class distribution. Thus, the metric is difficult to interpret in applications with high class imbalance[10] (as it is the case especially regarding squeal noise) and also not comparable between multiple classification tasks with deviating class distribution. To imagine, in a two-class classification problem with class 1 occurring in 90 % data points and for the rest (10 %) class 2 is assigned. If all samples would be predicted as having class 1, an overall accuracy of 90 % is achieved, however performance of the model is bad. In case of balanced class distribution among the samples to predict, a benchmark

[10] Class imbalance occurs if one class is overrepresented in the dataset. [113]

of 90 % would be pretty high and seen as good performance. To conclude, due to partly severe class imbalance in the present classification tasks, overall accuracy is only added in the predictive modelling step (Subsection 6.4) for additional information. [113]

As a benchmark, which is independent of class imbalance and thus comparable between all presented classification tasks, the Receiver Operating Characteristic (ROC) [118–120] is applied as the main metric of performance. It evaluates class probabilities rather than class values. ROC curves are drawn by plotting the value $1 - specificity$ on the x-axis and the *sensitivity* on the y-axis. They are derived as follows [113]:

$$sensitivity = \frac{TP}{n_{\mathrm{P}}} \tag{38}$$

$$specificity = \frac{TN}{n_{\mathrm{N}}} \tag{39}$$

where n_{P} represents the number of positive samples and n_{N} states the number of negative samples.

On the one hand, the ROC curve allows an estimation of how altering the cut-off probability threshold (default 50%) for decision, which class is predicted, influences sensitivity and specificity. On the other hand, the area under the ROC curve (AUC) represents a metric for model performance with an ideal model having a value of one and a completely uninformative model would show a value of 0.5. The AUC index is used for performance comparison and termed as ROC benchmark throughout the evaluations. [113]

6.1.3 Considered models

The evaluation contains continuous and categorical dependent variables – hence regression and classification models are applied. In general, linear, non-linear and tree/rule-based models can be distinguished. Some algorithms are able to deal with regression and classification tasks with minor alterations, others are only available for one of them. The scope of model selection is to consider a wide variety of different modelling approaches – from basic linear regression to highly complex random pattern search algorithms (e.g. artificial neural networks) – because no literature is available for the present or at least a similar modelling task. Table 25 provides an overview of chosen models. Furthermore, applicability to regression and classification tasks as well as vulnerabilities are stated. A short name for each model approach, which is mostly used in the following sections, is assigned. The last column describes whether a variable importance score can be evaluated from the model – all models with that possibility are used for the feature selection step in Subsection 6.3 as well. In the following, considered models are briefly described as well as further references for algorithmic detail provided.

Model	Short name	Recommended pre-processing	Reg.	Class.	Variable importance
Linear					
Linear regression	Lin Reg	1,2,4	✓		✓
Logistic regression	Log Reg	1,2,4		✓	✓
Linear discriminant analysis	LDA	1,4		✓	✓
Partial least squares	PLS	4	✓	✓	✓
Elastic net	E-Net	1,4	✓		✓
General logistic model net	GLM-Net	1,4		✓	✓
Nearest Shrunken Centroids	NSC	1,4		✓	✓
Non-linear					
Neural networks	Neural Networks	1,2,4	✓	✓	✓
Multivariate adaptive regression splines	MARS	–	✓		✓
Flexible discriminant analysis	FDA	–		✓	✓
K-nearest neighbour	KNN	1,4	✓	✓	
Naïve Bayes	NB	1,2		✓	
Tree/rule-based					
Basic Tree	CART	–	✓	✓	✓
Conditional Inference Tree	CTree	–	✓	✓	✓
Bagged tree	Bagged Tree	–	✓	✓	✓
Boosted tree	Boosted Tree	–	✓	✓	✓
Random forest	Random Forest	–	✓	✓	✓
Cubist	Cubist	–	✓		✓

Table 25: Considered models and their characteristics [113]

Linear regression

Linear regression can be directly described by the following equation [113]:

$$y_i = b_0 + b_1 \cdot x_{i,1} + b_2 \cdot x_{i,2} + ... + b_P \cdot x_{i,P} + e_i \tag{40}$$

where y_i is the predicted value for the i-th sample, b_0 states the intercept, b_j represents the coefficient for the j-th predictor, index P depicts the number of predictors and e_i is a random error value that the model cannot explain.

All regression models, which can be rewritten in the form of Equation (40), are categorized as linear models – that applies to PLS and E-Net as well. An important advantage of those is their high interpretability – especially in terms of the impact of every predictor and their relation among each other. The algorithm estimates the coefficients to minimize SSE. Linear regression models are limited if non-linear relations between predictors or between predictor and response are present. Due to their vulnerability to near-zero variance predictors, between predictor correlations and different predictor scales, pre-processing is essential for linear regression models. [113]

Logistic regression

Event probability calculation with logistic regression can be derived with the following equation [113]:

$$p = \frac{1}{1 + exp[-(\beta_0 + \beta_1 \cdot x_1 + ... + \beta_P \cdot x_P)]} \tag{41}$$

where p is the event probability (values between zero and one), β_0 represents the intercept, β_j states the estimated coefficient for the j-th predictor and index P depicts the number of predictors.

The algorithm again estimates the coefficients to minimize SSE, which leads to maximum likelihood estimates with the assumption of a normal distribution in the model outcome. Advantages and importance of pre-processing are identical to linear regression. [113]

Linear Discriminant Analysis

Discriminant analysis is focused on deriving a discriminant score, which in case of two classes (1,2) and two predictors (A,B) is computed for an unknown sample u as follows [113]:

$$D(u) = \mathbf{u}' \cdot S^{-1} \cdot (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \tag{42}$$

$$= u_{\mathrm{A}} \cdot \left(\frac{(\bar{x}_{1,\mathrm{A}} - \bar{x}_{2,\mathrm{A}} \cdot s^2_{\mathrm{B}})}{s^2_{\mathrm{A}} \cdot s^2_{\mathrm{B}} - s^2_{\mathrm{AB}}} - \frac{(\bar{x}_{1,\mathrm{B}} - \bar{x}_{2,\mathrm{B}} \cdot s_{\mathrm{AB}})}{s^2_{\mathrm{A}} \cdot s^2_{\mathrm{B}} - s^2_{\mathrm{AB}}} \right) + u_{\mathrm{B}} \cdot \left(\frac{(\bar{x}_{1,\mathrm{B}} - \bar{x}_{2,\mathrm{B}} \cdot s^2_{\mathrm{A}})}{s^2_{\mathrm{A}} \cdot s^2_{\mathrm{B}} - s^2_{\mathrm{AB}}} - \frac{(\bar{x}_{1,\mathrm{A}} - \bar{x}_{2,\mathrm{A}} \cdot s_{\mathrm{AB}})}{s^2_{\mathrm{A}} \cdot s^2_{\mathrm{B}} - s^2_{\mathrm{AB}}} \right) \tag{43}$$

where $\bar{x}$ represents the sample mean with indices for predictors A and B and calculated with data from class 1 or 2, s is the sample variance with indices for predictor A and B and s_{AB} states the sample covariance[11] between predictor A and B.

Following the depicted example, the classification is carried out by evaluating the difference between the discriminant score and the mean value of samples in each class and classifying the sample to the class with less difference. Thus, the algorithm minimizes the total probability of misclassification. As can be seen in Equation (43), a calculation with two parameters and two classes involves four mean values and three variance values. With growing predictor number, the computational effort grows large. Classification in two classes with P predictors needs $2P + P(P+1)/2$ parameters to be computed. However, due to evaluation of many mean and variance values, LDA is more robust to correlated predictors compared to logistic regression. Nonetheless, extreme between predictor correlations compromise the model as well as different predictor scales and near-zero variance predictors. Due to the need of an invertible covariance matrix (see Equation (42)), LDA is not applicable if more predictors than samples are present (this applies also to linear and logistic regression due to the same reason). Thus, filtering predictors with near-zero variance, between predictor correlations and predictor reduction methods (like PCA) are recommended. [113]

Partial Least Squares

PLS regression builds linear combinations between predictors, which on the one hand represent maximum variation of the predictors and on the other hand show the best correlation with the outcome. While the former is also the scope of PCA, the latter is an important feature. Thus, PLS, in contrast to linear regression, is able to deal with high between predictor correlations as well as with more predictors than samples due to the integrated generation of linear combinations. Furthermore, near-zero variance is not an issue because those predictors can be simply excluded in the built components. Mathematically, each iteration of the algorithm stores a vector of weights (called direction), which numerically summarizes the relationship between predictors and response. Predictors are afterwards projected orthogonally onto that direction and score vectors are evaluated. These are used to compute loadings, which represent a correlation value between score vector and original predictor. After each iteration the currently estimated predictors are subtracted from the response information. They are used to compute the next set of weights, scores and loadings (for more algorithmic details be referred to [121]). In the post-processing step, the optimal number of components can be evaluated by computing a performance metric (e.g. RMSE) to choose the final prediction model. Due to the computation of variance values of the predictors, different scales have a negative influence on the model building process. Thus, a standardized predictor set is favourable. [113]

Regarding classification, PLS involves a discriminant analysis approach with the same advantages over LDA as in the regression case over linear regression. The computation process

[11] A large covariance between two values indicates that they have the same sign and are on the same scale, whereas a near zero value means that they have no (or at least not a linear) relationship [113].

shows one significant deviation: The response in the regression case is univariate (one variable with continuous values), while in classification the response consist of a matrix with dummy variables of each class. The mathematical solution is provided in [122]. [113]

Elastic net

The elastic net regression model is part of the family of penalised models and combines the lasso [123] and ridge [124] approach. In general, the idea of penalised models is to add penalty terms to the calculation of SSE. That leads to high SSE values if parameter estimates grow large, which indicates over-fitting or between predictor correlations and thus prevents the model to include those due to the scope of finding minimum SSE. In case of elastic net, the equation is as follows [113]:

$$SSE_{\mathrm{E-Net}} = \sum_{i=1}^{n}(y_i - \hat{y}_i) + \lambda_1 \cdot \sum_{j=1}^{P}\beta_j^2 + \lambda_2 \cdot \sum_{j=1}^{P}|\beta_j| \tag{44}$$

where λ_1 is a ridge-type penalty value, P represents the number of Predictors, β_j state the parameter estimates and λ_2 is the lasso penalty.

The best performing values of λ_1 and λ_2 are evaluated in the post-processing step. Due to direct evaluation of the predictor set (without building combinations as in PLS), sample quantity has to exceed the number of predictors. Furthermore, again variances are evaluated like in the basic linear regression approach. Thus, predictor scales have an impact and therefore standardization of the predictor set is recommended. For more information about the elastic net algorithm be referred to [125]. [113]

General logistic model net

A GLM-Net follows the same approach for classification as the elastic net in regression. By adding penalty terms to a likelihood function from logistic regression (refer to Equation (41)), correlated parameters and tendencies to over-fitting are detected and those parameters are excluded from the model. The likelihood function is extended as follows [113]:

$$\ln L(p) - \lambda \left[(1-\alpha) \cdot \frac{1}{2} \cdot \sum_{j=1}^{P}\beta_j^2 + \alpha \cdot \sum_{j=1}^{P}|\beta_j| \right] \tag{45}$$

where $L(p)$ is the likelihood function, λ states a general penalty parameter and α represents a mixing value of lasso and ridge penalty terms.

In the post-processing step, λ and α values are chosen by comparison of performance metrics. Pre-processing needs are the same as in the elastic net approach with the same reasons. Further information about GLM-net is provided in [126]. [113]

Nearest Shrunken Centroids

NSC models work well with high-dimensional tasks, especially if the number of predictors significantly exceeds the sample quantity due to embedded feature selection. In general, the average value of each predictor in each class is taken as centroid. The overall centroid is derived by using the whole dataset (with all classes). If a predictor centroid is close to the overall centroid, its information for a certain class is low. Feature selection is carried out by evaluating the distance between class-specific centroid and overall centroid in combination with a shrinkage value, which is kind of a distance threshold. Only predictors with more distance between class-specific centroid and overall centroid (hence containing more information) than the shrinkage value are included in the model. The magnitude is chosen in the post-processing step by evaluating performance metrics. Detailed information is described in [127]. Since average values are computed, predictor scales matter and near-zero variance predictors can be wrongly interpreted as highly important. Thus, standardization and filtering of near-zero variance predictors as pre-processing is recommended. [113]

Neural networks

The model approach is focused on building one or more layers of hidden units (comparable to PLS components), which consist of linear combinations of some or all predictors. In contrast to PLS, the hidden units are built in a random nature (not with the constraint of maximised variability) and the transformation between predictor space and hidden unit is carried out usually with a non-linear function (e.g. a sigmoidal one) [113]:

$$h_k(\mathbf{x}) = g\left(\beta_{0,k} + \sum_{j=1}^{P} x_j \beta_{j,k}\right) \tag{46}$$

$$g(u) = \frac{1}{1+e^{-u}} \tag{47}$$

where $h_k(\mathbf{x})$ represents the function of the hidden unit k, $g(u)$ states the transformation function between predictor space and hidden unit layer (in the depicted form sigmoidal), $\beta_{j,k}$ is similar to regression coefficients with indices of predictor j and hidden unit k.

The outcome is modelled by one or multiple hidden units in a linear relationship. The total number of coefficients in the calculation grow with $H(P+1)+H+1$ with H as the quantity of hidden units and the number of predictors P. Therefore, the computational cost grows large with increasing predictor numbers. The basic algorithm tends to over-fitting. Thus, weight decay values (like penalties in elastic net) can be used to prevent that issue. The calculation of minimised SSE is altered as follows [113]:

$$\sum_{i=1}^{n}(y_i - f_i(x))^2 + \lambda \cdot \sum_{k=1}^{H}\sum_{j=0}^{P}\beta_{j,k}^2 + \lambda \cdot \sum_{k=0}^{H}\gamma_k^2 \tag{48}$$

where $f_i(x)$ is the linear function between hidden units and outcome for the ith sample, λ represents the weight decay value and γ_k are computed coefficients in $f_i(x)$ for each considered

hidden unit k.

The number of hidden units generated and the weight decay value are selected in the post-processing step by performance metric comparison. Neural networks are vulnerable in terms of near-zero variance predictors, between predictor correlations and different magnitudes in the predictor set. Further details are provided in [128]. [113]

Neural Networks can also be used in classification tasks. Differences in the application in contrast to regression are multiple outcome nodes (equal to the number of classes present) and a non-linear transformation to alter the linear relationship between hidden units and each class with a sigmoidal function to get values between zero and one. However, they are not directly probability values since the sum does not equal one. Thus, both relations predictors-hidden units and hidden units-outcome are non-linear (i.e. sigmoidal). [113]

Multivariate adaptive regression splines

Like PLS and neural network techniques, MARS creates separate components from the predictor set and models the outcome with them. In contrast to the former, which are based on linear combinations of the predictors (see Equation (46)), each predictor is split into two contrasted versions. Additionally, the separated components are mostly functions of only one or two predictors. Predictor separation is carried out by using hinge functions, which means that a certain range of the predictor values are set to zero and the remaining part is approximated with a linear regression line. To evaluate the cut point of the hinge function, each data point is modelled as a possible cut point with linear regression and the combination of predictor and cut point with the least RMSE value is chosen as a component in the model. The search process for combinations, which show the most reduction in RMSE, is continued until a defined stopping point. After derivation of the components, pruning can be used to filter predictors, which show a negligible reduction in RMSE – thus simplifying the model. Advantages of MARS are the robustness in terms of between predictor correlations, near-zero variance predictors and different scales in the predictor set due to building no predictor combinations, but only splitting predictors into two separate predictors. Furthermore, the model is highly interpretable because every predictor contribution to the outcome can be directly estimated. Embedded feature selection is also conducted. How much considered terms should be included in the final model can be evaluated in the post-processing step. [113]

Flexible discriminant analysis

FDA is an approach, where C (number of classes) linear regression models are fit to binary class indicators and their regression coefficients can be used to derive discriminant coefficients in a post-processing step. Any kind of regression approach, which produces slopes and intercepts for predictors can be applied (e.g. lasso or MARS). Due to the robustness and high interpretability of the approach, MARS is chosen as regression method. Hinge functions are converted to discriminant functions, which represent non-linear combinations of the original predictors. Further algorithmic details can be found in [129]. Since the technique consists mainly of the described MARS approach in the previous paragraph, no pre-processing is

necessary. [113]

K-nearest neighbour

The KNN method predicts a new sample based on the K-nearest neighbour samples from the training dataset. In case of regression the prediction value is derived with the mean of the K considered neighbour values. The computed distance metric can be chosen – in the present case Euclidean distance is applied due to being the most common [113]:

$$\left(\sum_{j=1}^{P}(x_{\mathrm{a},j} - x_{\mathrm{b},j})^2\right)^{\frac{1}{2}} \tag{49}$$

where x_{a} and x_{b} represent different sample value.

The only dependence on a distance metric between sample values makes the approach extremely vulnerable to different predictor scales. Thus, standardization is of high importance. Furthermore, near-zero variance predictors can worsen the predictive performance. Computational cost can be high if lots of samples should be predicted because for every estimated sample value, the whole set of distance metrics from the training set has to be computed. The number of neighbours considered in the final prediction model can be evaluated in a post-processing step by comparison of performance metrics. [113]

For application in classification, the derivation is also dependent on the computed distance metric (refer to Equation (49)). Instead of the mean value of neighbours, class probabilities are calculated by the proportion of neighbours in each class. The prediction classifies the sample into the class with the highest probability. If the proportion is tied, the neighbour value of $K + 1$ is decisive. [113]

Naïve Bayes

The foundation of the model is Bayes' Rule, which allows a direct estimation of class probabilities [113]:

$$Pr[Y = C_l|X] = \frac{Pr[Y] \cdot Pr[X|Y = C_l]}{Pr[X]} \tag{50}$$

where $Pr[Y = C_l|X]$ represents the posterior probability of the class l, $Pr[Y]$ states the prior probability (e.g. in the present case overall probability of squeal noise occurrence), $Pr[X|Y = C_l]$ is called conditional probability, which means the probability of observing the predictor values in class l, and $Pr[X]$ represents the probability of the predictors.

The NB approach assumes that all predictors are independent from each other, which is not realistic in most cases, however significantly simplifies the calculation of Bayes' Rule. The conditional probability can be rewritten as product of probability densities, which are

evaluated for each predictor separately:

$$Pr[X|Y = C_l] = \prod_{j=1}^{P} Pr[X_j|Y = C_l] \tag{51}$$

Estimation of each predictor individually leads to a simple computation of predictor probabilities – in case of continuous variables with an assumed normal distribution or kernel density functions and in case of categorical ones with the observed frequency in the training set. Despite that strong assumption, NB model performance can be competitive, especially due to low computational cost even in conjunction with large datasets. Since probability values are multiplied, the model breaks down if one predictor has no samples in a class (hence class probability equals zero). Thus, near-zero variance predictors have to be excluded prior to modelling. Strong between predictor correlations make the taken assumption more unrealistic and can worsen the model behaviour. Therefore, correlation analysis in pre-processing is recommended. Due to evaluation of each predictor independently, different scales do not matter. [113]

Basic Tree

Tree models in general can be applied to regression and classification tasks and consist of one or more if-then conditions, which are used to predict new sample values. The terminology is to refer each if-then statement in the model as split or node and each possible outcome value as terminal node or leave of the tree. Regarding regression, the algorithm searches every predictor value for the optimal split value, which is defined as the best minimisation in overall SSE, to split the dataset into two groups [113]:

$$SSE = \sum_{i \epsilon S_1} (y_i - \hat{y_1})^2 + \sum_{i \epsilon S_2} (y_i - \hat{y_2})^2 \tag{52}$$

where $\hat{y_1}$ and $\hat{y_2}$ are the average values of the training set outcome with indices for groups S_1 and S_2.

In the next step, the dataset of S_1 and S_2 is separately searched again for the optimal split value within each group. Due to that recursive behaviour, the technique is also known as recursive partitioning. The estimated value in each terminal node is calculated by averaging the remaining sample values of the training set. If the tree is fully grown, it is possibly very large and the advantage of a high interpretability is lost as well as tendency to over-fitting is present. A remedy for that is to apply pruning algorithms (e.g. [130]) as a post-processing step to reduce complexity by removing splits and leaves with minimal increase in overall SSE. Advantages of basic tree models are simplicity, high interpretability (if the tree is not large) and integrated feature selection (if predictors are never used in a split, they have no impact on the model). If high between predictor correlations are present, the chosen predictors in splits are quite random and thus that should be avoided. Due to its simplicity, the downside is its limited predictive performance (i.e. capturing the whole connection between predictors and outcome). Furthermore, the selection of predictors for splits depend on their number

of distinct values and may thus be influenced if many uninformative predictors are present. They can be chosen in the first few splits as decisive, while expressing no connection to the outcome at all. [113]

In classification tasks optimal splitting values are found by minimizing misclassification using the Gini index [130]. For a two-class model in a single node this value is calculated as follows [113]:

$$p_1 \cdot (1 - p_1) + p_2 \cdot (1 - p_2) \tag{53}$$

where p is the class probability with indices for class 1 and 2.

If only two classes are present, Equation (53) can be rewritten as $2p_1p_2$, which can be easily interpreted as low value if one class probability is nearly zero (indicating a good separation) and high value if both class probabilities are high. Nearly all possible split points (mid point between each unique predictor value) are evaluated and the maximum Gini index value is chosen. In each split, the searching process is repeated until the full tree is grown. Pruning is also applicable and advantages and downsides are identical with regression tasks. [113]

Conditional Inference Tree

The model building process is identical with basic trees for regression and classification with one significant exception. The optimal split values are selected by applying statistical hypothesis tests for each possible split point. The test is carried out by calculating a p-value[12] and thus a measure of statistical significance between the average values of the two created groups. This leads to a significant reduction of selection bias if many uninformative predictors are present, which may be wrongly chosen by a basic tree approach. Furthermore, if a certain threshold of statistical significance cannot be fulfilled, the splitting process is stopped for that node. Thus, pruning is usually not applied to conditional inference trees. For further algorithmic details be referred to [131]. [113]

Bagged Tree

Bagging or in long form bootstrap aggregation is a general technique applicable to any regression or classification model. The idea is to train a certain amount of models on different bootstrap samples [132] and predict new samples by averaging the outcome over all built models. The algorithm is as simple as that [113]:
for (i in 1:m) {

(a) Select a bootstrap sample out of the training set

(b) Train a model on this sample (in the present case an unpruned tree model)

}

In case of basic tree models, bagging can be highly beneficial in terms of improving stability

[12] p-values are a measure of statistical significance in terms of indicating a deviation from the null hypothesis in the relation between a predictor and the outcome (further mathematic details are provided in [131]).

and predictive performance. The averaging of the estimated values of multiple tree models lowers the variance in the outcome. Furthermore, the model is trained on a subset of the training data (the bootstrap sample), which enables a computation of a performance metric of each generated model by predicting the samples, which are left out (called out of bag samples). Pruning is skipped because over-fitting is prevented by the averaging process over multiple models. The number of calculated bagging iterations m can be evaluated in the post-processing step. [113]

Boosted Tree

In the present approach, a gradient boosting machine (GBM) [133] is applied. The general idea is to use a model with weak learning properties (e.g. basic trees with limited depth) and extend that model in each iteration to improve the predictive ability. The model building starts with a basic value for the response (e.g. average of the dependent variable in the regression case) and evaluates a gradient between observed and currently predicted value (e.g. residual) in each iteration. The algorithm consists of the following steps [113]:

1. Input tree depth D and number of iterations K
2. Compute a starting value (e.g. average response)
3. **for** (k in 1:K) {
   (a) Compute the residual between observation and prediction for each sample
   (b) Build a basic tree model of depth D
   (c) Estimate each sample value by using the tree model from (b)
   (d) Add the whole or a fraction (if shrinkage is used) of the currently predicted value to the estimated sample value in the previous iteration step

   }

A GBM tends to over-fitting because in every iteration the best performing model for the training set is chosen, which might not be the best overall model for the prediction of new samples. To prevent that issue, a shrinkage value (between zero and one), which works similar to penalties in E-Net and GLM-Net models, can be included. In the last iteration step, only a fraction of the newly predicted sample value (called a learning rate) is added to the model. While improving predictive performance, the downside is an increased computational cost (inversely proportional to the shrinkage value) due to the need of computing more iterations. The optimal tree depth, number of iterations and shrinkage value can be evaluated in the post-processing step by comparison of performance metrics. [113]

Regarding classification the algorithm uses the approach of logistic regression as an initial probability value (refer to Equation (41)) and extends that function with the estimates from a basic classification tree in each iteration. The general idea is equal to regression. Tree depth, number of iteration and shrinkage value for the final model are also evaluated in the post-processing step. For a detailed insight into the algorithm for classification be referred to [134].

Random forest

In contrast to bagged and boosted trees, where the generated trees are connected to the trees of previous iterations (a discussion on tree correlation can be found in [135]), a random forest model produces independent trees in each iteration. That is achieved by randomly selecting only a subset of k predictors and choosing the best predictor among them for the current split. The algorithm can be applied to regression and classification tasks in the same way and contains the following steps [113]:

1. Input the quantity of models to calculate m
2. **for** (i in 1:m) {
   (a) Select a bootstrap sample out of the training set
   (b) Train a basic tree model on this sample
   (c) **for** each split {
      i. Select a random set of k predictors
      ii. Evaluate the best predictor from the subset and create the split
      }
   (d) stop if the tree is fully grown (no pruning applied)
   }

Due to randomly subsetting the predictor set, the model is extremely robust in terms of between predictor correlation and reaches an even lower variance in the outcome than bagging. The overall number of trees m in the forest have to be chosen prior to model building, while the best performing value of k can be evaluated in the post-processing step. Further algorithmic details are provided in [136]. [113]

Cubist

The Cubist algorithm is a rule-based model, which is solely developed for regression tasks. The approach in general constructs model trees (i.e. every split is a model). Each node (child) is compared to the node above (parent) by applying a smoothing procedure to assign a weighting value to each model in the tree [113]:

$$a = \frac{Var[e_\mathrm{p}] - Cov[e_\mathrm{c}, e_\mathrm{p}]}{Var[e_\mathrm{p} - e_\mathrm{c}]} \tag{54}$$

where a is the smoothing coefficient and e represents the error value of the estimated sample value (i.e. the residual) with indices c for the child model and p for the parent model.

If the variance of the error value of the parent model is larger than the covariance, the smoothing procedure assigns more weight to the child model compared to the parent model. Each extracted rule from the model tree is represented by single linear model, which selection is based on the calculated weights. The next step consists of a pruning or rule combination

process by using the overall adjusted error rate in the training set. Any rule, for which an exclusion from the model results in no increased error rate, is dropped. The whole model building process can be further extended by considering committees, which represent a correction value if certain data points in the training set are under- or over-predicted. The final rule-based model is used to predict new samples by evaluating K similar neighbour values from the training set. The final prediction is computed as follows [113]:

$$\frac{1}{K} \cdot \sum_{l=1}^{K} w_l \cdot [t_l + (\hat{y} - \hat{t}_l)] \tag{55}$$

$$w_l = \frac{1}{D_l + 0.5} \tag{56}$$

where w is a weighting value depending on the distance D between neighbour l and new sample, t represents the observed outcome, $\hat{t}$ states the predicted outcome and $\hat{y}$ is the estimated value of the new sample.

The Cubist algorithm uses the Manhattan distance metric[13]. The number of calculated committees and the quantity of considered neighbour values for prediction can be evaluated by comparing performance metrics in the post processing step. [113]

6.1.4 Feature selection

Feature selection methods can be categorized in unsupervised and supervised techniques. The former ones are completely independent from the outcome, whereas the latter methods are designed to increase accuracy of a model by using different subsets of predictors. As unsupervised techniques, the previously described pre-processing filters to exclude near-zero variance and highly correlated predictors as well as PCA are considered. Supervised methods are subdivided in filter, wrapper and embedded approaches. The first compares each predictor with the outcome separately by using a certain filter criterion (e.g. p-value threshold between predictor and outcome) prior to model building. The latter can be illustrated as a loop around the model in which in each iteration only a subset of predictors is used to train the model. Afterwards, the final model is chosen by comparison of a performance metric. The outer loop ("wrapped" around the model) causes an increase in computational cost, which can be partly significant dependent on the applied model. Apart from filter and wrapper approaches, some models conduct embedded feature selection in their algorithm (e.g. PLS). Furthermore, in most considered models, predictor importance can be computed after the model building process (refer to the last column in Table 25). A general discussion about feature selection can be found in [137]. Since neither applied filter nor wrapper approaches achieve better results than embedded methods, only the latter are used in evaluation as well as discussed below. In addition, an univariate filter approach is applied for comparison to empirical findings. In the following paragraphs, considered approaches are briefly explained. [113]

[13] The difference to Euclidean distance (see Equation (49)) is that instead of squaring the gradient and taking the square root of the sum, the absolute value of the gradient is used [113].

Unsupervised techniques

The processes of filtering near-zero variance and between predictor correlations are explained in Subsection 6.1.1. PCA is more a predictor reduction and less a feature selection method due to the scope of generating a new, uncorrelated predictor set. The process is driven by explaining the variance in the predictor set with as few components, which consist of linear combinations of the original predictors. The full set of generated components is reducible by setting a variance threshold (e.g. 90%). The remaining components might contain not all original predictors or at least the contribution of some predictors to the components is very low, which causes a conducted feature selection incidentally. For more algorithmic details be referred to [116]. [113]

Filter methods

As previously mentioned, filter methods are solely used for comparison of univariate variable importance scores and empirical correlation analysis. Derivation in case of regression is carried out on the one hand with linear variance analysis and on the other hand with a smoothing spline function fitted between predictor and outcome (known as LOESS [138]). Regarding classification, a ROC benchmark is computed between each predictor and the response.

Predictor importance and embedded feature selection

Embedded feature selection is carried out in the algorithms of PLS, E-Net, GLM-Net, NSC, MARS, FDA and in all tree and rule-based models allowing a direct estimation of the best performing predictor set. Apart from that, predictor importance scores can be computed depending on the model type. The derivation is summarised in the following paragraphs.

For linear and logistic regression as well as LDA, t-statistic[14] is used on each model parameter and a p-value is computed with the null hypothesis of no difference between the two groups. [113]

PLS conducts feature selection in terms of excluding predictors, which are not part of the generated components. Due to the existence of three layers – predictor space, PLS components and outcome –, the contribution of each predictor to the model is quite complex to derive. The normalized weight of each predictor in each component is scaled with the explained variation in the response by that component and proportional to the achieved reduction in SSE by including it into the model. The numerator of the importance is calculated as a weighted sum of normalised weights assigned to each predictor, while the denominator is the overall response variation explained by all components. Thus, increased normalised weight of a predictor leads to a raise in its importance score. [113]

E-Net and GLM-Net models carry out feature selection by setting parameters to zero at certain penalty values causing that the predictor with coefficient zero is not integrated in the

[14]The t-statistic is a method for comparing means between two groups by taking the ratio of the difference of the mean values and a variability function. The general assumption is a normal distribution [113]. A full mathematical description is provided in [106].

final prediction model. For the remaining model parameters, t-statistic is used in the same way as for linear regression or LDA.

Variable importance in a NSC model is expressed by the difference between class centroids and overall centroid. Larger distance values imply a better class separation and thus indicate a larger importance. [113]

Predictor importance scores in neural networks are derived by the approach of Gevrey et al. [139], which is based on computation of weighting values (similar to PLS).

MARS models (and hence also FDA models) include a backwards feature elimination, which targets a minimization in generalized cross validation (GCV) [140] estimates of RMSE. Predictor importance scores are evaluated by tracking changes in the GCV RMSE if a hinge function of that predictor is added to the model and accumulating them. [113]

In general, all tree or rule-based models conduct feature selection automatically if one or more predictors are not used in any split or rule. Variable importance can be quantified by the reduction of the chosen performance metric (e.g. in case of regression SSE) achieved in a certain split. Variable appearance in the beginning nodes of a tree as well as multiple usage in splits indicate a larger predictor importance. The computation is identical for CART and conditional inference trees. In case of Bagged Tree, the same approach is applied to all generated trees by the bootstrap algorithm and an overall importance score is calculated. [113]

In case of Boosted Tree, every predictor is evaluated in terms of reduction of the chosen performance metric for splitting in each constructed tree and afterwards averaged over all trees in the ensemble. Due to dependencies between trees, many of the same predictors are selected in each tree, which magnifies their importance in the averaging process. For computational details be referred to [141] and [142]. [113]

Evaluation of predictor importance in random forest models is carried out by randomly permuting values of each predictor and observe the difference between predictive performance of the original sample compared to the altered one. If alteration of one predictor leads to hardly any difference in predictive performance, the predictor is assumed to be less important. More details about the approach are provided in [143]. [113]

Regarding Cubist, variable importance scores are based on the quantity of usage of each predictor in rule conditions and generated models. Distance metrics are not included in the calculation. [113]

6.1.5 Post-processing

Regardless what models are applied, model performance has to be estimated in a post-processing step. In general, the standard approach is to train a model on a major part of the dataset and to evaluate its predictive performance on a test set of data, which is not used for model training. There is no determined threshold of data splitting. In the present approach, 75 % of the data are used for training and 25 % as test set. In case of classification, 10 % of data points are solely used for examination of alternating cut-off thresholds. Thus, training is only done on 65 % of the data. While that approach is taken for the predictive modelling step in Subsection 6.4, the scope of the first evaluation is to build a final set of

parameters. Therefore, no test set is used and all data are applied for training. However, resampling methods in the model building process allow an integration of data splitting and thus an evaluation of predictive performance in the model training step. Resampling techniques generally choose a subset, which is used to train the model and the remaining samples are used for performance estimation. That process is repeated several times and the results are compared to select the final model parameters. The main difference in resampling algorithms is the subsetting method. Repeated k-fold cross validation[15] is recommended as a technique, which is able to deal with small sample sizes, which are present in different subsets of the evaluation, and shows good bias and variance properties, while also causing an acceptable computational effort. Thus, the described method is applied to all further evaluations following the recommendation of applying ten folds and five repetitions. [113]

6.1.6 Computational details

For all statistical calculations as well as graphical output of those in this thesis, R version 3.6.3 [144] is used in combination with the following packages: AppliedPredictiveModeling 1.1-7 [145], bestNormalize 1.6.0 [146], caret 6.0-86 [147], corrplot 0.84 [148], doParallel 1.0.15 [149], dplyr 0.8.5 [150], devtools 2.3.2 [151], e1071 1.7-3 [152], earth 5.1.2 [153], elasticnet 1.3 [154], extrafont 0.17 [155], forcats 0.5.0 [156], foreach 1.5.0 [157], Formula 1.2-3 [158], gbm 2.1.5 [159], ggplot2 3.3.0 [160], ggpubr 0.2.5 [161], ggsci 2.9 [162], ggthemes 4.2.0 [163], glmnet 4.0-2 [164], iterators 1.0.12 [165], lars 1.2 [166], lattice 0.20-38 [167], lubridate 1.7.4 [168], magrittr 1.5 [169], Matrix 1.2-18 [170], matrixStats 0.56.0 [171], nnet 7.3-12 [172] with customised plot_nnet function [173], plotmo 3.5.7 [174], plotrix 3.7-8 [175], pls 2.7-2 [176], plyr 1.8.6 [177], pROC 1.16.2 [178], prodlim 2019.11.13 [179], purrr 0.3.3 [180], R.matlab 3.6.2 [181], randomForest 4.6-14 [182], randomForestExplainer 0.10.1 [183], RColorBrewer 1.1-2 [184], readr 1.3.1 [185], recipes 0.1.12 [186], rlang 0.4.5 [187], reshape2 1.4.3 [188], ROCR 1.0-11 [189], scales 1.1.0 [190], scatterplot3d 0.3-41 [191], sm 2.2-5.6 [192], stringr 1.4.0 [193], TeachingDemos 2.12 [194], tibble 2.1.3 [195], tidyr 1.0.2 [196], tidyverse 1.3.0 [197], utf8 1.1.4 [198], vioplot 0.3.5 [199], XLConnect 1.0.1 [200] and zoo 1.8-8 [201]. Data preparation as well as application of the curve squeal detection algorithm (described in Chapter 4) is carried out with MATLAB® version 2019b [202].

The computation is done on a PC with an Intel® Core™ i5-9500 CPU and 32 GB DDR4 RAM. Parallel processing is carried out with six workers.

6.2 Methodology

The basic methodology is illustrated in Figure 50. Derivation of parameters, empirical filtering and subsetting are previously described in Chapter 5. They state the fundamental steps of data preparation. Parallel empirical examination and statistical feature selection result in formulation of a final predictor set, which is focused on practical suitability and general

[15] K-fold cross validation creates k random subsets of the data, trains a model on k-1 subsets and uses the remaining subset as test set to measure predictive performance. Every subset is used once as a test set. In the repeated case, the whole process is carried out again with different subsets. [113]

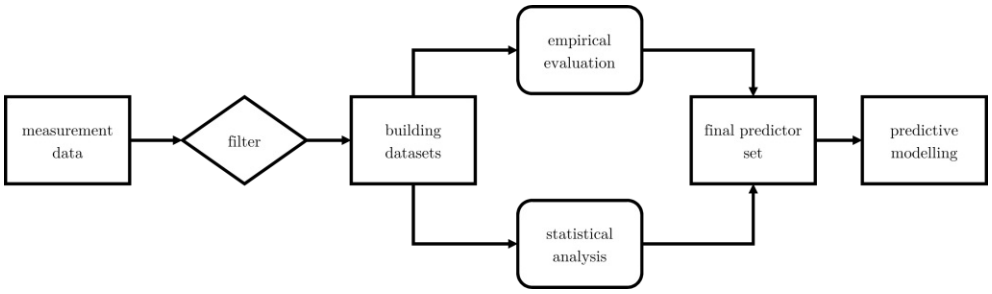

Figure 50: Chosen methodology approach for feature selection and predictive modelling

applicability on other datasets in the future (further explained in Subsection 6.2.1). Thus, the scope of the final set is a significantly reduced predictor amount and an exclusion of parameters, which are compromised by a special property in a certain section. Building on the selected parameters, predictive models are built and discussed. The latter process targets the development of practical and generally applicable predictive models. The latter is ensured by using only two sections for model training and the remaining three for model validation on data points, which consist of deviating track properties, train type distribution and climatic conditions. Final selection of correlated environmental parameters as well as general necessity of using environmental predictors at all are also added by excluding single predictors or groups out of the model and observing changes in accuracy (refer to Subsection 6.2.2).

6.2.1 Feature Selection

Feature Selection is carried out for each dependent variable (see Table 24) separately. The discussion is focused on subsets with strong statistical significance (refer to Subsection 5.6.1). This applies to "all", "MC_1", "MC_2", "MC_4", "MC_5", "TT_A", "TT_E", "TT_F", "TT_G", "MC_1_TT_E", "MC_1_TT_F", "MC_4_TT_G", "MC_5_TT_F" and "MC _5_TT_G". Dependent variables, which involve peak levels or relative occurrence time of curve squeal consist of reduced samples because only train pass bys, which are categorised as containing that type of curve squeal are available. In case of BroadLevelMax and BroadTimeRel, "MC_2", "MC_4" and "MC_4_TT_G" are downgraded to mediocre statistical significance. In terms of TonalLowLevelMax and TonalLowTimeRel "MC_2" and "MC_4_TT_G" are rated as weak and "MC_4" and "TT_A" are categorised as mediocre. Regarding TonalHighLevelMax and TonalHighTimeRel "MC_2" is downgraded to weak and "MC_4", "TT_A" and "MC_4_TT_G" are rated as mediocre. Furthermore, severe class imbalance is present – especially regarding TonalLow "MC_2" (only 18 true samples) and TonalHigh "MC_2" (only 36 true samples). That problem is generally present regarding TonalLow and TonalHigh and is accounted by using the ROC metric as performance benchmark (further explained in Subsection 6.1) and investigating alternative probability cut-offs in Subsection 6.4.

The discussion in the main part of the thesis (Subsection 6.3) is focused on predictor selection for curve squeal noise occurrence (Broad, TonalLow and TonalHigh) as well as equivalent

continuous noise level (Level). The reason is that investigated peak levels (BroadLevelMax, TonalLowLevelMax and TonalHighLevelMax) as well as relative occurrence time (BroadTimeRel, TonalLowTimeRel and TonalHighTimeRel) of curve squeal are found to be not satisfyingly predictable by a statistical model due to large RMSE values. Discussions on those are provided in Appendix C with either a heuristic model approach or simply a final list of most important predictors at the end. In the latter case, median values can be taken from figures in Appendix B for an approximation. It has to be pointed out that high variations in the dataset – especially regarding relative occurrence time, which reasons also the high occurring RMSE values –, add uncertainty to the described correlations and stated magnitudes. Building on mentioned issues, peak levels as well as relative time occurrence of curve squeal are not further considered in the predictive modelling step (Subsection 6.4).

Empirical examination

Each parameter of every subset is examined in conjunction with each dependent variable. The foundation is built, in case of categorical parameters, on violin plots (for continuous dependent variables, see Figure 52 left) and bar plots (for categorical dependent variables, depicted in Figure 53). For evaluation of continuous parameters, 2D-scatterplots with an underlying density function (in case of continuous dependent variables, see Figure 51) and violin plots (if a categorical dependent variable is evaluated, illustrated in Figure 52) are used. In addition, in case of 2D-scatterplots, parameters of a fitted linear regression line are also added for better estimation as well as an underlying density function. Evaluation of equivalent continuous noise level is separated in normal train pass bys (no curve squeal detected) and train pass bys categorised as containing curve squeal (regardless what type of curve squeal) due to significant level differences. The examination is focused on the former category. Rolling noise level increases with rising speed [6]. Thompson [5] states a usual proportionality to the logarithm of speed using the following equation:

$$L_{\mathrm{p}} = L_{\mathrm{p0}} + N \cdot \log_{10}\left(\frac{V}{V_0}\right) \tag{57}$$

where L_{p} is the A-weighted sound pressure level in dB, L_{p0} is the A-weighted sound pressure level at a reference speed V_0 in dB, V is the speed of the current measurement and N is a factor between 25 and 35 with a typical value of 30.

Using that equation with the stated value range for N on train pass bys without detected curve squeal shows bad accordance to the collected measurement data. Reasons can be on the one hand that intentionally no A-weighting is applied and on the other hand the relationship in a curve may be different. Thus, an approach for correction of the dataset is taken by setting the factor N in Equation (57) to 7.5 (shows best accordance – investigated empirically) and the reference speed to 60 kmh^{-1} (speed limit for majority of samples). The dependent variable Level (depicting equivalent continuous noise level) is corrected to the reference speed only for comparability reasons in the empirical examination. In all other analyses, no correction is applied due to the possible inclusion of speed as a parameter in multivariate approaches.

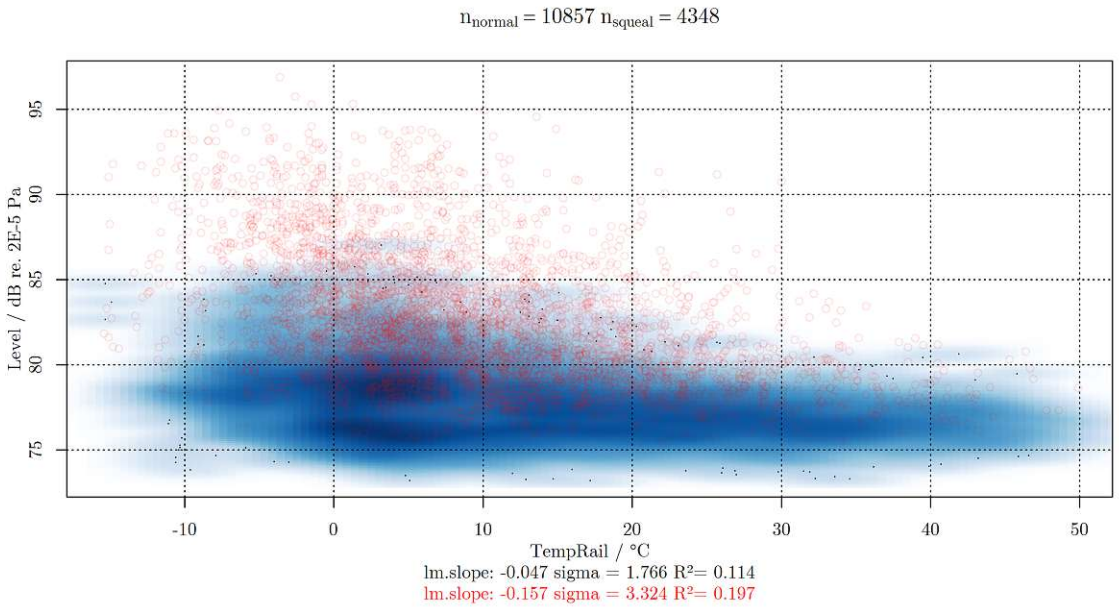

Figure 51: 2D-scatterplot with splitting of curve squeal (red) and normal measurement values (blue) with an underlying density function (outcome: Level, subset: "MC_5_TT_G")

Although a factor value of 7.5 shows best accordance in the present dataset, the applicability to other measurement data has to be critically examined.

To categorize visible correlations, a coloured matrix with rows for each parameter and columns for each subset is created. The colour scheme is separated in positive and negative correlation[16]. Every category is graded empirically into one of those levels:

- 3/-3: strong correlation visible
- 2/-2: correlation at least partly identified
- 1/-1: weak signs of a correlation assumed

In case of no correlations at all, the value zero is noted in the matrix. Figure 54 provides an example of the subset "MC_5_TT_G" with some parameters and from the dependent variable Level.

[16]Positive correlation means that the value of the dependent variable increases if the parameter value is raised, whereas negative correlation means the opposite.

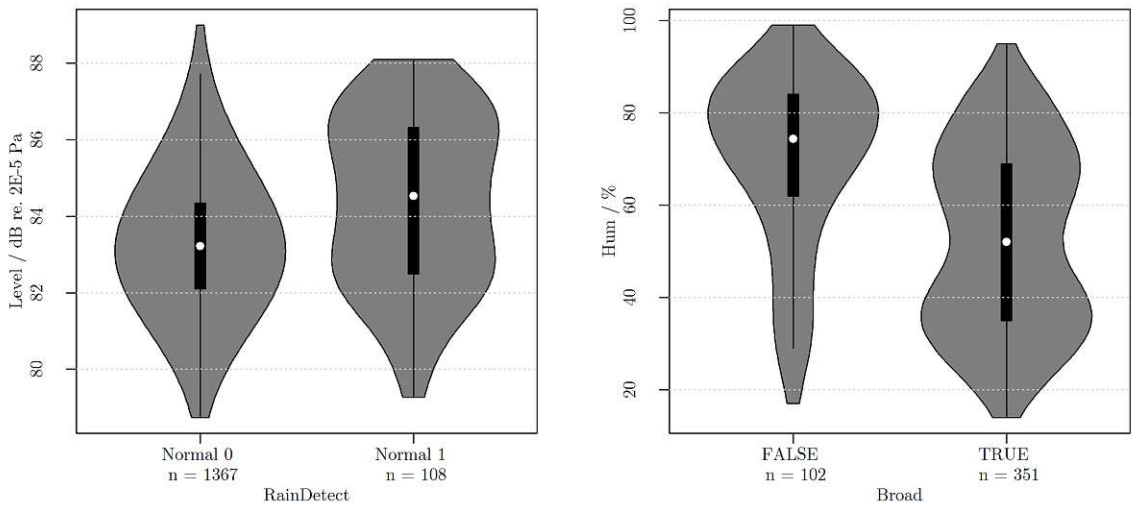

Figure 52: Left: violin plot of normal train pass bys and categorical parameter (outcome: Level, subset: "MC_5_TT_F"); Right: violin plot of categorical outcome and continuous parameter (outcome: Broad, subset: "MC_1_TT_A")

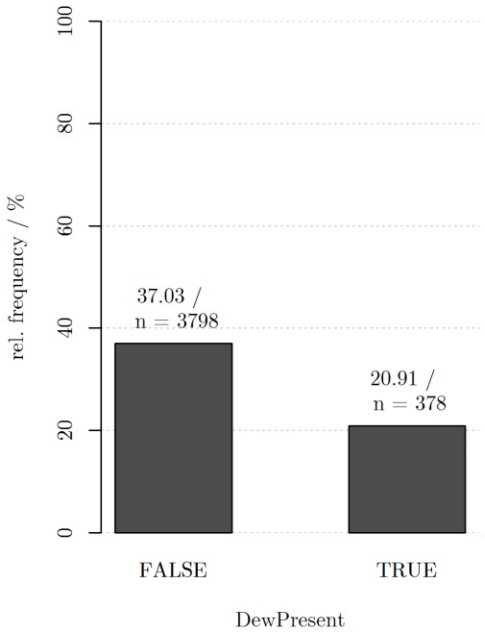

Figure 53: Bar plot with relative frequency of occurrence (outcome: Broad, subset: "MC_1")

Environment	MC_5_TT_G
Pressure	1
TempAir	-2
TempRail	-2
Hum	2
HumRail	2
DewPoint	0
FrostPoint	0
WindAbs	0
WindDir	0
RainMmH	0
RainDurSLT	0
RainMmSLT	0
RainDetect	1

Figure 54: Example of the matrix created for empirical examination

Statistical feature extraction

Statistical feature extraction is carried out with the approaches introduced in Subsection 6.1. Unsupervised techniques as well as models, which conduct embedded feature selection or at least allow an estimation of variable importance scores are applied. The latter state multivariate methods and build the foundation in selecting the final subset of predictors. Filter functions are only considered univariately for comparison to empirically observed correlations as they achieve worse results compared to other statistical methods in a multivariate analysis. As results of linear variance analysis are similar to the evaluation using the LOESS evaluation, only the former is depicted and discussed. In addition, a multivariate feature selection technique called recursive feature elimination (RFE) [203] is applied. However, it shows worse performance compared to other considered statistical algorithms, thus results are neither depicted nor discussed further. The application of completely randomized wrapper techniques like genetic algorithms or simulated annealing may be beneficial, however they had to be excluded due to their extreme computational cost. Although near-zero variance and between predictor correlation might compromise some models, they are intentionally not excluded prior to modelling in the feature selection step. The reason is that the identical dataset should be analysed by different approaches to select a final parameter set for the predictive modelling step in Subsection 6.4. Predictors showing zero variance are automatically excluded due to containing no information. Discussion is focused on the whole dataset, since the scope lies in selecting representative predictors for a generalized prediction model, which is applicable on different sections and divergent train types. Comparison is divided in univariate (empirical and statistical) and multivariate analyses. Afterwards, putting all findings together, a final predictor set is formulated for each dependent variable. Decision of the final model approach as well as best fitting tuning parameters of the model is carried out in case of regression with minimising RMSE (accuracy benchmark) and in case of classification with maximising the ROC benchmark. For theoretical reference about applied methods, be referred to Subsection 6.1.

Discussion outline

Firstly, unsupervised techniques are discussed for the whole dataset as well as all considered subsets as they are independent of the outcome. Afterwards each dependent variable is discussed separately following a consistent scheme. At first, empirical findings and univariate approaches are summarised as well as illustrated. The first figure states evaluated variable importance by a linear variance analysis (regression) or rankings by a achieved ROC benchmark (classification). The picture gives an overview of importance in conjunction with each parameter group – track, vehicle, dynamic, general and environmental. Afterwards, empirical interpretation in combination with univariate statistical predictor importance scores is carried out. The focus lies in examination of subsets "MC_TT", which are independent of different section and train type properties as they cannot be sufficiently accounted in a two dimensional evaluation. In the next step, multivariate statistical analyses are carried out with model algorithms, which conduct embedded feature selection or at least allow evaluation of predictor importance scores. The two best performing approaches among the whole dataset and subsets are depicted. Moreover, predictor importance scores of those are provided in a figure to finally interpret inclusion decisions of multivariate approaches. Each discussion on a dependent variable is concluded with a summary at the end of the subsection as well as a final predictor set, which, in case of Level, Broad, TonalLow and TonalHigh, is further applied and investigated in the predictive modelling step. Otherwise, a heuristic model approach or a recommendation for estimation from illustrated median values is stated.

6.2.2 Predictive Modelling

For predictive modelling, all described statistical algorithms are applied – regardless whether they allow predictor importance scores or provide embedded feature selection since that is already done in the previous step. Moreover, application of any filter or wrapper methods is skipped due to the same reason. Unsupervised techniques are partly applied in terms of zero variance filtering and PCA for rail roughness and track decay rate values – however, not separately discussed. Between predictor correlations are not present as well as near-zero variance. In contrast to the feature selection step, validation of predictive models is carried out by a separate test set (25 % data points). In case of classification, another small evaluation set is split prior to modelling (10 % data points). The latter is solely used to investigate alternating probability cut-offs to especially raise sensitivity (recommended by [113]). Thus, in case of regression, the model is trained on 75 % of data points and in case of classification on 65 % of data points. To improve predictive ability of the model and choose best performing tuning parameters, a resampling technique – again five times repeated 10-fold cross validation – is also applied. In general, resampled performance on the training set leads to similar benchmarks as on the test set, which justifies the choice of the resampling technique. Splitting is carried out by an algorithm, which selects groups (based on percentiles) and sampling in those subgroups. For classification, random sampling is done in the levels of the factor with the goal of balanced class distribution [147]. In general, model training is carried out on the whole dataset, on data points of C1 and C5, on C1 and on C5

as all other approaches have too few data points considering the applied splitting technique. Thus, a general validation of final model selection in terms of behaviour on each of the four considered sets is possible.

Firstly, best performing correlated environmental predictors, which are listed in the final predictor set, are evaluated by building models with only changing those predictors and compare performance benchmarks. In case of temperature, TempAir, TempRail, SatPressure and SatPressureRail are investigated. In case of relative humidity quantities, HumRail, Hum, RT.DPDiff and AT.DPDiff are examined. Finally, in case of absolute humidity quantities, HumSpec, HumSpecRail, MixRatio and MixRatioRail are considered. The latter choice is justified as they are suitable for comparing different location on different altitudes due to inclusion of absolute air pressure in the derivation. Regarding predictive modelling of equivalent continuous noise level, examinations show that especially squeal noise occurrence can have a significant impact. Thus, inclusion of predictors for curve squeal occurrence might be beneficial. The investigation is carried out by comparing model benchmarks with and without consideration of curve squeal occurrence predictors. For complete comparability in mentioned final predictor evaluations, the same data splits are used.

The final model building process (refer to Figure 55 left for regression and right for classification) is focused on data points in C1 and C5 as both are the only long-term measurement campaigns and thus allowing a wide variety of environmental conditions to include. Moreover, both campaigns show completely different track conditions, which is also beneficial in terms of model training. Thus, as described previously, models are trained on 65 % of data points in C1 and C5 in case of classification and 75 % of data points in the mentioned sections in case of regression. The strategy is justified by the main scope of building a model, which is applicable on new sections and new datasets. Thus, beside test and evaluation set (only in classification), the models can be applied to data points of C2, C3 and C4 (further referred as validation sets). Hence, a direct benchmark of application on a new section – unknown by the model – can be evaluated. Since there are completely different curve radii, climatic conditions and train type distributions involved, that depicts a value for general applicability on other sections and datasets. The final model is chosen on behalf of the latter application method in conjunction with reached benchmarks in ROC (classification) and RMSE (regression). Both are also used in model training to choose the best performing tuning parameters for the present task. Final models are described and interpreted in detail depending on their complexity.

For the classification tasks – predicting occurrence of flanging noise, squeal noise and HF squeal noise – the default probability cut-off (0.5 or 50 %), which decides whether a train pass by is predicted as positive (i.e. curve squeal occurrence) or negative (no curve squeal occurrence) is not ideal due to severe class imbalance especially regarding squeal noise. Alternative cut-off points are investigated by plotting the ROC curve of the model applied to the evaluation set (as described in the first paragraph of the current subsection). The ROC curve depicts *sensitivity* on the y-axis and $1 - specificity$ on the x-axis. It has to be borne in mind that defining an alternative cut-off value is not improving overall model performance – it is rather a shift between sensitivity and specificity. Thus, setting a lower cut-off probability

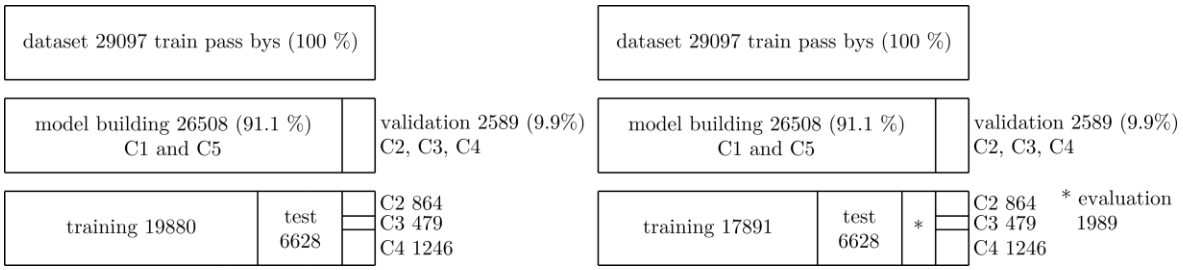

Figure 55: Overview of data quantity and distribution for the final modelling step focusing on data points in C1 and C5 for model building – left for regression and right for classification

leads directly to a higher sensitivity and a lower specificity. Since the scope in curve squeal detection is a high rate of detected events in conjunction with an acceptable rate of falsely positive predictions to stay on the safe side, the focus lies in achieving a high sensitivity. The examination is carried out on the one hand statistically by calculating the top left point on the ROC curve, which states a balanced compromise between sensitivity and specificity, and the Youden index[17] as well as on the other hand empirically by focusing on slope change points in the ROC curve. The slope of the ROC curve is an indicator of how much sensitivity is gained in comparison to the loss of specificity. A steep slope indicates a high gain of sensitivity with only minor losses in specificity, while the opposite illustrates a low sensitivity increase causing a high loss in specificity. Other approaches like up- or down-sampling of data points to achieve equal class frequencies or cost-sensitive training are not applied as they alternate the model or the dataset, which adds a reasonable uncertainty to the model building process as well as to the applicability on new datasets. Thus, they are seen as not appropriate for the present modelling task.

As the thesis focuses on impact of environmental conditions, the consequences of a complete exclusion of included environmental predictors is examined by training new models with the reduced predictor set and comparing them by achieved statistical benchmarks. The focus lies on models trained on data points of C1 and C5, while decision, which model performs best, is made building on achieved benchmarks in the validation sets. Thus, a justification of significance regarding environmental conditions for prediction is reached. The opposite case – using only environmental predictors – is examined as well with the same approach. For complete comparability, the same data splits as in the previously selected final model are used in those evaluations.

6.3 Feature selection

After derivation of parameters and subsetting (see Chapter 5), the next step is to examine the dataset on the one hand empirically and on the other hand statistically. The former approach is predominately used to understand simple two dimensional correlations and for validation of the outcome of the statistical computation. The latter is applied to investigate multivariate dependencies in the dataset and to sort out parameters, which show no correlation with the

[17]The Youden index is computed with the sum of sensitivity and specificity subtracted by 1. The value is maximised to evaluate the best performing cut-off probability. [113]

outcome and hence may even compromise the following predictive modelling process depending on the applied model properties. The subsection is outlined in application of unsupervised techniques, followed by discussion of each dependent variable, which is considered further in the predictive modelling step (Subsection 6.4) – Broad, TonalLow, TonalHigh and Level. Appendix C provides a discussion on the remaining dependent variables – peak level and relative occurrence time of each curve squeal type.

6.3.1 Unsupervised techniques

Unsupervised feature selection include dataset filtering in conjunction with near-zero or zero variance and between predictor correlations as well as PCA. As previously stated, those methods are independent of the outcome, which enables a discussion without including dependent variables.

Variance filter

The first step in pre-processing and parameter set investigation is filtering zero and near-zero variance. The basic methodology is described in Subsection 6.1. Standard thresholds are applied and conspicuous parameters are stated. For boolean or dummy variables, the criterion of unique values is meaningless. Considering the whole dataset (all sections and all train types), the following parameter are evaluated by applying standard threshold values:

- RainMmH: The frequency ratio is above 90, which is very high, however reasonable due to many values of zero (in case of no rain). Unique values are also sparse mainly due to the low resolution of the weather station in C1 - C3.
- RainDurSLT and RainMmSLT are influenced by the same issue as RainMmH with frequency ratios of 82 and 186.
- RainDetect: Frequency ratio value is 25, which indicates a strong class imbalance (low quantity of pass bys, where rain is present).
- SnowCm: Snow occurrence is very rare, which is also detected by that filter approach.
- HoarfrostPresent: High probability of hoarfrost occurrence is sparsely observed.
- TrainType.1: Freight train type A shows a frequency ratio of 28, which is primarily caused by very few train pass-bys in C4 and C5 (refer to Figure 28).
- TrainType.2, TrainType.3 and TrainType.4 show even more severe frequency ratios with 219, 138 and 239, which are all very high.
- TrainOrientation.3, TrainOrientation.4, TrainOrientation.5 and TrainOrientation.6 distinguish double unit operation configurations, which are only predominant in C1. Frequency ratio is with values between 31 and 36 evenly distributed.
- Loco.2: Locomotive type 2 is rarely observed in the train pass bys, which explains the high frequency ratio of 79.

- Loco.3 and Loco.4 show an even rarer occurrence with extreme frequency ratio values of 1118 and 14548.
- FreeCurving: A frequency ratio of 30 indicates class imbalance, which is explainable by the majority of data points with no curving restrictions (ideally calculated).
- RailType.2: This rail type is only present in C2, which is a section with few monitored train pass bys and reasons the frequency ratio of 33.

Subsetting between sections leads to zero variance values for all predictor, which are used to take different properties among sections into account. That includes R, Cant, LongSlope, RailType, SleeperType and all rail roughness and track decay rate (except in C4 and C5, where values are linearly interpolated between measurements at the beginning and at the end of the campaign). Additionally, the following predictors show zero variance:

- MC_1: TrainType.2, TrainType.4, Loco.3, Loco.4
- MC_2: SnowCm, HoarfrostPresent, FrostPresent, Quarter, Month, TrainType.4, FreeCurving
- MC_4: TrainType.3, Train.Type.4, TrainType.5, TrainOrientation.4, TrainOrientation.5, TrainOrientation.6, Loco.1, Loco.2, Loco.3, Loco.4
- MC_5: TrainType.3, TrainType.4, TrainType.5, TrainOrientation.4, TrainOrientation.5, TrainOrientation.6, Loco.1, Loco.2, Loco.3, Loco.4

Detection of near-zero variance is separated for further illustration only with two categories (for better clarity): (I) frequency ratio under 40, which might be acceptable for inclusion, if strong tendencies seem to be present; (II) frequency ratio of 40 or above, in which case consideration can be excluded at all. The following near-zero variance predictors are identified for section subsets:

- MC_1 I: RainMmH, RainDetect, TrainType.1, TrainType.7
- MC_1 II: RainDurSLT, RainMmSLT, DewMm, DewCumMm, HoarfrostPresent, FrostPresent, TrainType.3, Loco.2
- MC_2 I: RainMmH, RainDetect, TrainType.2, Loco.3
- MC_2 II: RainDurSLT, RainMmSLT, TrainType.6, TrainOrientation.2, TrainOrientation.3, TrainOrientation.4, TrainOrientation.5, TrainOrientation.6, Loco.4
- MC_4 I: RainMmH
- MC_4 II: SnowCm, HoarfrostPresent, TrainType.1, TrainType.2, TrainOrientation.3
- MC_5 I: RainDetect

- MC_5 II: RainMmH, RainDurSLT, RainMmSLT, SnowCm, TrainType.1, TrainType.2, TrainOrientation.3, FreeCuving

Train type separation over all sections leads to near-zero variance of train type predictors. Moreover, dependent on the subset, the following parameters show zero variance:

- TT_A: HoarfrostPresent, TrainOrientation.1, TrainOrientation.2, TrainOrientation.3, TrainOrientation.4, TrainOrientation.5, TrainOrientation.6, Loco.1, Loco.2, Loco.3, Loco.4, MaxWheelDm
- TT_E: TrainOrientation.3, TrainOrientation.4, TrainOrientation.5, TrainOrientation.6, Loco.4, RailType.3, SleeperType.2
- TT_F: Loco.1, Loco.2, Loco.3, Loco.4, MaxBogieDist, MaxWheelDm, FreeCurving
- TT_G: TrainOrientation.1, TrainOrientation.2, TrainOrientation.3, TrainOrientation.4, TrainOrientation.5, TrainOrientation.6, Loco.1, Loco.2, Loco.3, Loco.4, MaxBogieDist, MaxWheelDm, FreeCurving

From the remaining predictors, near-zero variance is present as follows:

- TT_A I: RainMmH, SnowCm
- TT_A II: RainDurSLT, RainMmSLT, FrostPresent
- TT_E I: RainMmH, RainDetect, MaxAoA
- TT_E II: RainDurSLT, RainMmSLT, DewMm, HoarfrostPresent, FrostPresent, TrainOrientation.1, TrainOrientation.2, Loco.3
- TT_F I: RainDetect
- TT_F II: RainMmH, RainDurSLT, RainMmSLT, HoarfrostPresent, RailType.2
- TT_G I: RainDetect, RailType.3, SleeperType.2
- TT_G II: RainMmH, RainDurSLT, RainMmSLT, SnowCm, RailType.2

Further data splitting causes zero variance regarding TrainType, R, Cant, RailType, SleeperType, LongSlope, rail roughness and track decay rate (for the two latter C4 and C5 is an exception). Subset dependent zero variance predictors are:

- MC_1_TT_E: TrainOrientation.3, TrainOrientation.4, TrainOrientation.5, TrainOrientation.6, Loco.3, Loco.4
- MC_1_TT_F: Loco.1, Loco.2, Loco.3, Loco.4, MaxBogieDist, MaxAoA, MaxWheelDm, RSI, FreeCurving

- MC_4_TT_G: TrainOrientation.1, TrainOrientation.2, TrainOrientation.3, TrainOrientation.4, TrainOrientation.5, TrainOrientation.6, Loco.1, Loco.2, Loco.3, Loco.4, MaxBogieDist, MaxAoA, MaxWheelDm, RSI, FreeCurving
- MC_5_TT_F: TrainOrientation.4, TrainOrientation.5, TrainOrientation.6, Loco.1, Loco.2, Loco.3, Loco.4, MaxBogieDist, MaxAoA, MaxWheelDm, RSI, FreeCurving
- MC_5_TT_G: TrainOrientation.1, TrainOrientation.2, TrainOrientation.3, TrainOrientation.4, TrainOrientation.5, TrainOrientation.6, Loco.1, Loco.2, Loco.3, Loco.4, MaxBogieDist, MaxAoA, MaxWheelDm, RSI, FreeCurving

In the remaining predictor set, near-zero variance is detected as follows:

- MC_1_TT_E I: RainMmH, RainDetect
- MC_1_TT_E II: RainDurSLT, RainMmSLT, DewMm, DewCumMm, HoarfrostPresent, FrostPresent, TrainOrientation.1, TrainOrientation.2, Loco.2, MaxBogieDist, MaxAoA
- MC_1_TT_F I: RainMmH, RainDetect
- MC_1_TT_F II: RainDurSLT, RainMmSLT, DewMm, HoarfrostPresent, FrostPresent
- MC_4_TT_G I: RainMmH
- MC_4_TT_G II: SnowCm, HoarfrostPresent
- MC_5_TT_F I: RainDetect, HoarfrostPresent
- MC_5_TT_F II: RainMmH, RainDurSLT, RainMmSLT, SnowCm, TrainOrientation.3
- MC_5_TT_G I: RainDetect
- MC_5_TT_G II: RainMmH, RainDurSLT, RainMmSLT, SnowCm

To conclude, rain, dew, snow and frost quantities are partly sparse. Quantities of train types B, C and D are low compared to the other categories – especially E, F and G. That issue also applies to locomotive types 3 and 4.

Between predictor correlations

Between predictor correlations can have a significant negative impact on statistic model algorithms – even robust tree and rule-based approaches are influenced due to making a somewhat random choice among correlated predictors in a split. Thus, although all predictors are included, that aspect has to be borne in mind in the interpretation of results. The illustration is separated in categories environment, general, vehicle, dynamic and track for better clarity. Pair-wise negative or positive correlation values above an absolute threshold of 0.9 (90 %) are depicted. Discussion on the whole dataset as well as alterations in subsets

are provided. Predictors, which show no pair-wise correlations, are not mentioned separately. Parameters included after applying the correlation search algorithm (refer to Subsection 6.1) with threshold values of 0.75 and 0.9, are mentioned separately. Deviation between threshold values is sparse, thus distinction is only noted if differences are present. It has to be pointed out that the algorithm is only able to deal with numeric predictors. Hence, categorical ones are converted to dummy variables with values zero (false) and one (true). However, the conversion has no effect on the result.

Predictor correlations regarding environmental parameters are provided in Table 26. Strong correlations are mainly notable between humidity quantities and also dew point. Since relative pressure is used to compare sections on different altitudes, subsetting by sections causes a high pair-wise correlation between Pressure and relative pressure quantities (PressureRel and PressureRelRail). Furthermore, CurveBreathing changes to high correlation with TempAir, TempRail, SatPressure and SatPressureRail, if only section subsets are considered. The search algorithm only retains RainMmH, RainMmSLT, RainDetect, DewEvapMm, DewMm and HoarfrostPresent as predictors in the whole dataset. Subsetting causes that partly RainDurSLT, SnowCm, DewCumMm, DewPresent and FrostPresent are also kept.

General parameters show high correlations between the predictors Month, Week and YearDay. Changes in subsets are not present. However, the search algorithm excludes every predictor in that category.

Regarding vehicle predictors, only TrainType.2 with MaxBogieDist and Loco.1 with TrainType.5 show high between predictor correlations in the whole dataset. Subsetting in sections lead to equality in information of MaxBogieDist and MaxAoA as well as MaxWheelDm and RSI, since only one curve radius value is present. Additionally, subset specific correlations occur (see Table 27 and Table 28). The search algorithm retains TrainType.(1-4), TrainOrientation.4, Loco.3, Loco.4 and MaxAoA in the whole dataset. Regarding different subsets, the following predictors are kept:

- MC_1: TrainType.1, TrainType.3, TrainType.7, TrainOrientation.4, Apl, MaxAoA
- MC_2: TrainType.2, TrainType.6, TrainType.7, TrainOrientation.(2-6), Loco.3, Loco.4, Apl, MaxAoA, RSI
- MC_4: TrainType.1, TrainType.2, TrainType.6, TrainOrientation.1, TrainOrientation.2, TrainOrientation.3, Apl, MaxAoA, RSI (only with threshold 0.9)
- MC_5: TrainType.1, TrainType.2, TrainOrientation.2, TrainOrientation.3, Apl, MaxAoA
- TT_A: Apl, MaxAoA, FreeCurving
- TT_E: TrainOrientation.2, Loco.2, Loco.3, Apl, MaxAoA
- TT_F: TrainOrientation.2, Apl, MaxAoA, RSI
- TT_G: Apl, MaxAoA, RSI

Predictor	Pair-wise correlation
TempAir	TempRail, SatPressure
TempRail	TempAir, SatPressure, SatPressureRail
Hum	AT.DPDiff
HumRail	RT.DPDiff
DewPoint	PartPressure, PartPressureRail, HumAbs, HumAbsRail, HumSpec, HumSpecRail, MixRatio, MixRatioRail
HumRail	RT.DPDiff
PressureRel	PressureRelRail
PressureRelRail	PressureRel
SatPressure	TempAir, TempRail, SatPressureRail
SatPressureRail	TempRail, SatPressure, DewEvapMm
PartPressure	DewPoint, PartPressureRail, HumAbs, HumAbsRail, HumSpec, HumSpecRail, MixRatio, MixRatioRail
PartPressureRail	DewPoint, PartPressure, HumAbs, HumAbsRail, HumSpec, HumSpecRail, MixRatio, MixRatioRail
HumAbs	DewPoint, PartPressure, PartPressureRail, HumAbsRail, HumSpec, HumSpecRail, MixRatio, MixRatioRail
HumAbsRail	DewPoint, PartPressure, PartPressureRail, HumAbs, HumSpec, HumSpecRail, MixRatio, MixRatioRail
HumSpec	DewPoint, PartPressure, PartPressureRail, HumAbs, HumAbsRail, HumSpecRail, MixRatio, MixRatioRail
HumSpecRail	DewPoint, PartPressure, PartPressureRail, HumAbs, HumAbsRail, HumSpec, MixRatio, MixRatioRail
MixRatio	DewPoint, PartPressure, PartPressureRail, HumAbs, HumAbsRail, HumSpec, HumSpecRail, MixRatioRail
MixRatioRail	DewPoint, PartPressure, PartPressureRail, HumAbs, HumAbsRail, HumSpec, HumSpecRail, MixRatio
DewEvapMm	SatPressureRail
AT.DPDiff	Hum, RT.DPDiff
RT.DPDiff	HumRail, AT.DPDiff

Table 26: Between predictor correlations for environmental parameters in the whole dataset

- MC_1_TT_E: TrainOrientation.2, Loco.2, Apl, MaxAoA
- MC_1_TT_F: TrainOrientation.2, Apl
- MC_4_TT_G: Apl
- MC_5_TT_F: TrainOrientation.3, Apl
- MC_5_TT_G: Apl

Dynamic properties show no correlations between predictors in the whole dataset. Depending on the subset, high correlations between Vmean and AccSide (all except TT_A and TT_E) as well as between VDiff and AccTrain (in MC_1, MC_4, TT_E, TT_F, MC_1_TT_E, MC_1_TT_F and MC_4_TT_G) appear. The search algorithm retains AccTrain and AccSide in the whole dataset and in most subsets.

Correlations between track parameters in the whole dataset are illustrated partly in Table 29. Rail roughness and track decay rate predictors show strong pair-wise correlations between each other, which has to be borne in mind in further evaluations. Subsetting in different sections removes most predictors due to zero variance. RR and TDR parameters in C4/C5 are perfectly correlated due to linear interpolation, which makes an exclusion preferable. The search algorithm keeps RailType.2, RailType.3 and SleeperType.2. Regarding subsets, the following predictors are retained:

- MC_4: RR.OR.10, TDRv.OR.1600, TDRv.OR.2500, TDRv.OR.5000, TDRh.OR.1250, TDRh.OR.1600, TDRh.IR.1250, TDRh.IR.1600, TDRh.IR.2500 (only with threshold 0.9)
- MC_5: RR.IR.5, TDRh.OR.1250, TDRh.OR.1600, TDRh.IR.1600, TDRh.IR.2500 (only with threshold 0.9)
- TT_A: RailType.2, RailType.3, SleeperType.2, TDRh.IR.1000, TDRh.IR.1250
- TT_E: RailType.2, RR.IR.3.15, RR.IR.2.5, TDRh.IR.1000, TDRh.IR.1250; only with threshold 0.9: RR.IR.16, TDRh.OR.100, TDRh.OR.5000, TDRh.IR.2000
- TT_F: RailType.2, RailType.3, SleeperType.2, TDRh.IR.2000 (only with threshold 0.9)
- TT_G: RailType.2
- MC_4_TT_G: RR.OR.10, TDRv.OR.1600, TDRv.OR.2500, TDRv.OR.5000, TDRh.OR.1250, TDRh.OR.1600, TDRh.IR.1250, TDRh.IR.1600; only with threshold 0.9: TDRv.OR.4000, TDRh.IR.2500
- MC_5_TT_F: RR.IR.5, TDRh.OR.1250, TDRh.OR.1600, TDRh.IR.1600, TDRh.IR.2500 (only with threshold 0.9)
- MC_5_TT_G: RR.IR.5, TDRh.OR.1250, TDRh.OR.1600, TDRh.IR.1600, TDRh.IR.2500 (only with threshold 0.9)

Predictor	Pair-wise correlation
MC_1	
TrainType.5	Loco.1, MaxBogieDist, MaxAoA
TrainType.6	MaxBogieDist, MaxAoA
Loco.1	TrainType.5
MaxBogieDist	TrainType.5, TrainType.6, MaxAoA
MaxAoA	TrainType.5, TrainType.6, MaxBogieDist
MC_2	
TrainType.2	MaxBogieDist, MaxAoA
TrainType.7	MaxWheelDm, RSI
MaxBogieDist	TrainType.2, MaxAoA
MaxAoA	TrainType.2, MaxBogieDist
MaxWheelDm	TrainType.7, RSI
RSI	TrainType.7, MaxWheelDm
MC_4	
TrainType.2	MaxBogieDist, MaxAoA
MaxBogieDist	TrainType.2, MaxAoA
MaxAoA	TrainType.2, MaxBogieDist
MaxWheelDm	RSI
RSI	MaxWheelDm
MC_5	
TrainType.2	MaxBogieDist, MaxAoA
TrainType.6	TrainType.7, Apl
TrainType.7	TrainType.6, Apl, MaxWheelDm, RSI
Apl	TrainType.6, TrainType.7
MaxBogieDist	TrainType.2, MaxAoA
MaxAoA	TrainType.2, MaxBogieDist
MaxWheelDm	TrainType.7, RSI
RSI	TrainType.7, MaxWheelDm

Table 27: Between predictor correlations for vehicle parameters in subsets "MC"

Predictor	Pair-wise correlation
TT_A	
MaxAoA	RSI
RSI	MaxAoA
TT_E	
Loco.1	MaxWheelDm
MaxAoA	RSI
MaxWheelDm	Loco.1
RSI	MaxAoA
TT_F	
Apl	MaxAoA, RSI
MaxAoA	Apl, RSI
RSI	Apl, MaxAoA
TT_G	
MaxAoA	RSI
RSI	MaxAoA
MC_1_TT_E	
TrainOrientation.1	TrainOrientation.2, Loco.2
TrainOrientation.2	TrainOrientation.1, Loco.2
Loco.1	MaxWheelDm, RSI, FreeCurving
Loco.2	TrainOrientation.1, TrainOrientation.2
MaxBogieDist	MaxAoA
MaxAoA	MaxBogieDist
MaxWheelDm	Loco.1, RSI, FreeCurving
RSI	Loco.1, MaxWheelDm, FreeCurving
FreeCurving	Loco.1, MaxWheelDm, RSI
MC_5_TT_F	
TrainOrientation.1	TrainOrientation.2
TrainOrientation.2	TrainOrientation.1

Table 28: Between predictor correlations for vehicle parameters in subsets "TT" and "MC_TT"

Predictor	Pair-wise correlation
R	RailType.2, RR.IR.8, TDRv.OR.160, TDRv.OR.5000
Cant	RR.IR.6.3, RR.IR.5
RailType.2	R, TDRv.OR.160, TDRv.OR.5000, TDRv.IR.315
RailType.3	SleeperType.2, LongSlope, RR.OR.25, RR.OR.20, RR.OR.16, RR.OR.12.5, RR.OR.10, RR.OR.8, RR.OR.6.3, RR.OR.5, RR.OR.0.63, RR.IR.25, RR.IR.12.5, RR.IR.10, RR.IR.2.5, TDRv.IR.800, TDRv.IR.1000, TDRv.IR.2000, TDRh.OR.100, TDRh.OR.315, TDRh.OR.400, TDRh.OR.500, TDRh.OR.630, TDRh.OR.800, TDRh.OR.1000, TDRh.IR.125, TDRh.IR.400, TDRh.IR.500, TDRh.IR.630, TDRh.IR.800
SleeperType.2	RailType.3, LongSlope, RR.OR.25, RR.OR.20, RR.OR.16, RR.OR.12.5, RR.OR.10, RR.OR.8, RR.OR.6.3, RR.OR.5, RR.OR.0.63, RR.IR.25, RR.IR.12.5, RR.IR.10, RR.IR.2.5, TDRv.IR.800, TDRv.IR.1000, TDRv.IR.2000, TDRh.OR.100, TDRh.OR.315, TDRh.OR.400, TDRh.OR.500, TDRh.OR.630, TDRh.OR.800, TDRh.OR.1000, TDRh.IR.125, TDRh.IR.400, TDRh.IR.500, TDRh.IR.630, TDRh.IR.800
LongSlope	SleeperType.2, RailType.3, RR.OR.20, RR.OR.8, RR.OR.6.3, RR.OR.5, RR.OR.4, RR.OR.2, RR.OR.1.6, RR.OR.0.63, TDRv.IR.100, TDRv.IR.500, TDRh.IR.400, TDRh.IR.500, TDRh.IR.630

Table 29: Between predictor correlations for track parameters in the whole dataset

Between predictor correlations are present especially in terms of environmental and track predictors. The applied detection algorithm sorts out important variables and does not seem to provide a satisfying predictor selection. The lack of connection to a dependent variable is clearly visible, however that issue refers to all unsupervised methods. Thus, reduction of between predictor correlations has to be carried out empirically building on the pair-wise correlation incidents, which are pointed out in the current subsection.

Principal component analysis

PCA is a suitable technique for high between predictor correlations and for the scope of parameter reduction by building a new, uncorrelated set of predictors. The application in the present case targets an estimation of how much components are needed to catch the majority of variance in the dataset and how well predictors can be represented by those. For the former, a recommended cut point is to dismiss principal components, which show an eigenvalue under one – meaning the principal components accounts for less variance in the dataset than a single predictor (in a standardized input set) [204]. To investigate quality of representation by a principal component, square cosine values are calculated. High quantities indicate good representation, while low values account for the opposite [205]. Prior to the calculation, predictors are pre-processed with steps 3 and 4 (see Subsection 6.1.1), which is necessary due to comparison of variance.

Table 30 provides an overview of how much principal components are needed to represent the bulk of variance in the predictor set. Rows state subsets and columns illustrate predictor sets. Separation is done in parameter groups environment (Env), general (Gen), vehicle (Veh), dynamic (Dyn) and track (Tr). The latter is further subcategorized in sets with only RR and/or TDR values as well as track quantities without TDR and RR predictors (referred as "Tr o"). As can be seen, although predictor quantity is large regarding TDR and RR, few principal components are needed to describe those values. Strong correlations in the parameter set of environment can also be avoided by application of PCA.

Quality of representation is depicted as a list of variables in each subset, which cannot be sufficiently described by principal components. No subset regarding predictors are pictured because bad description of parameters in different parameter groups are mostly the same. The threshold value of squared cosine values is set to 0.8 – hence only excluding very good representation values. The following predictors are not satisfactorily described (list sorted in decreasing squared cosine value):

- All: TrainOrientation.3, AccTrain, RainMmSLT, Loco.3, AccSide, SunRad, FreeCurving, CurveBreathing, TrainType.3, DewEvapMm, HoursSLT, RainMmH, VDiff, DewCumMm, RainDurSLT, TrainOrientation.2, Loco.2, DewMm, SnowCm, WindAbs, TrainType.4, Quarter, Clouds, Loco.4, WindDir
- MC_1: FreeCurving, SunRad, Loco.2, DewMm, DewCumMm, DewEvapMm, TrainOrientation.2, RainMmSLT, Quarter, WindAbs, HoursSLT, SnowCm, Clouds, Wind-

dataset	All	Env	Gen	Veh	Dyn	Tr	RR	TDR v	TDR h	TDR	TDR RR	Tr o
All	22	6	2	11	2	6	4	5	4	5	6	2
MC_1	16	6	2	8	2	–	–	–	–	–	–	–
MC_2	17	6	2	11	2	–	–	–	–	–	–	–
MC_4	14	7	2	5	2	2	1	2	1	2	2	–
MC_5	14	6	2	5	1	2	1	1	1	1	1	–
TT_A	12	5	2	2	2	4	3	4	4	4	4	3
TT_E	12	6	2	4	2	1	1	1	1	1	1	1
TT_F	19	6	2	5	2	5	3	4	3	5	5	2
TT_G	13	6	2	1	2	5	4	5	4	5	5	2
MC_1_ TT_E	12	6	2	2	2	–	–	–	–	–	–	–
MC_1_ TT_F	15	6	2	5	2	–	–	–	–	–	–	–
MC_4_ TT_G	11	6	2	–	2	2	1	2	1	2	2	–
MC_5_ TT_F	11	6	2	2	2	2	1	1	1	1	1	–
MC_5_ TT_G	10	6	2	–	2	2	1	1	1	1	1	–

Table 30: Quantity of principal components needed to describe the majority of variance divided in subsets of data points and predictors

Dir, TrainType.3

- MC_2: YearDay, DewEvapMm, RainMmH, SunRad, Apl, TrainOrientation.5, Loco.1, WindAbs, DewMm, DewCumMm, TrainOrientation.4, Hour, WindDir, Loco.3, RainDurSLT, TrainOrientation.2, Clouds, Loco.4, HoursSLT
- MC_4: TrainType.1, DewEvapMm, SunRad, Hour, DewMm, RainDurSLT, HoursSLT, WindAbs, DewCumMm, WindDir, SnowCm, TrainOrientation.3, Clouds
- MC_5: VDiff, RainMmSLT, FreeCurving, SunRad, WindDir, SnowCm, Hour, DewEvapMm, AccTrain, RainDurSLT, RainMmH, DewMm, WindAbs, Quarter, HoursSLT, DewCumMm, Clouds
- TT_A: DewEvapMm, RainDurSLT, RainMmH, SunRad, RainMmSLT, DewCumMm, SnowCm, WindAbs, DewMm, Quarter, Hour, HoursSLT, AccSide, Clouds, Apl, MaxBogieDist, WindDir
- TT_E: Hour, TrainOrientation.2, DewMm, DewCumMm, DewEvapMm, TrainOrientation.1, Loco.3, Apl, RainDurSLT, Quarter, WindAbs, SnowCm, Clouds, WindDir, HoursSLT
- TT_F: RainDurSLT, HoursSLT, SunRad, RainMmH, DewEvapMm, DewMm, Quarter, SnowCm, DewCumMm, Clouds, WindAbs, WindDir
- TT_G: RainMmSLT, CurveBreathing, SunRad, DewEvapMm, Hour, RainMmH, RainDurSLT, SnowCm, DewMm, WindAbs, DewCumMm, Quarter, WindDir, Clouds, HoursSLT
- MC_1_TT_E: RainMmSLT, DewCumMm, DewMm, DewEvapMm, WindDir, Quarter, WindAbs, SnowCm, Apl, Clouds, HoursSLT
- MC_1_TT_F: WindDir, Hour, Apl, RainMmH, DewMm, SunRad, RainMmSLT, DewCumMm, DewEvapMm, Quarter, WindAbs, SnowCm, Clouds, HoursSLT
- MC_4_TT_G: DewEvapMm, Hour, SunRad, SnowCm, Apl, WindDir, DewMm, RainDurSLT, WindAbs, HoursSLT, Clouds, DewCumMm
- MC_5_TT_F: AccTrain, RainMmH, SunRad, Hour, RainDurSLT, DewEvapMm, DewMm, WindDir, SnowCm, Quarter, WindAbs, DewCumMm, Clouds, HoursSLT
- MC_5_TT_G: RainMmSLT, Apl, AccTrain, DewEvapMm, SunRad, SnowCm, Hour, DewMm, RainMmH, RainDurSLT, WindAbs, Quarter, DewCumMm, HoursSLT, Clouds

The strength of PCA lies in describing high correlations between predictors as uncorrelated principal components, which are linear combinations of the originals. Weakness occurs if predictors are completely uncorrelated with other ones. In that case the solution of PCA is to generate a component for each of them, which misses the scope of predictor reduction.

Thus, for all mentioned predictors, which cannot be sufficiently described by PCA, a low correlation can be assumed and they could be included separately probably with some principal components generated from highly correlated predictor groups.

6.3.2 Flanging noise occurrence

Univariate statistical analysis is not capable of extracting highly important predictors in terms of flanging noise occurrence (referred as Broad) in the whole dataset (see Figure 56) as well as in any subset. Four vehicle predictors (MaxBogieDist, MaxAoA, TrainType.6 and Apl) are ranked top, followed by track parameters. Environmental predictors are listed in midfield. However, differences among parameters are low. Thus, empirical findings are discussed without comparison to univariate statistical outcome. The second part discusses performance and findings of applied multivariate statistical algorithms. The subsection is concluded by a brief summary of most important trends and building on that final influence parameters for predictive modelling are defined. General tendencies in sections and in conjunction with different train types are discussed in Subsection 5.7 as well as depicted there in Figure 42. To avoid influence from train type distribution and section dependencies overall and in subsets "TT" and "MC", the empirical evaluation is focused on "MC_TT" splits.

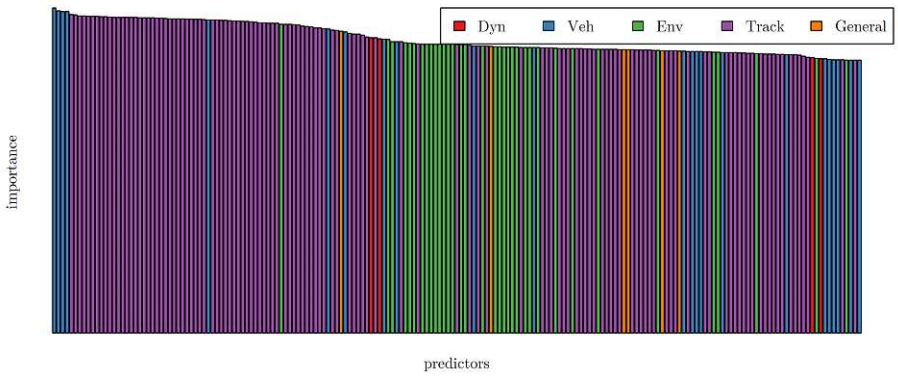

Figure 56: Predictor importance scores regarding flanging noise occurrence for the whole dataset evaluated with a univariate ROC curve evaluation (achieved benchmarks between 0.6 and 0.5) and coloured in groups dynamic (red), vehicle (blue), environmental (green), track (purple) and general (orange)

In general, negative correlation to curve radius is indicated by freight and passenger trains – especially between narrow radii ($\leq$ 256 m) and higher ones (440 m) (see Table 31). However, weak statistical significance (foremost regarding C2) for that observation has to be pointed out. Train type D only operates in C3 and is therefore skipped in the illustration. Among narrow radii ($\leq$ 256 m), freight trains show a generally high frequency of occurrence (67 % to 79 %), whereas train types E, G and other indicate low values (16 % to 31 %). The remaining

categories show mediocre values. Other track parameters show no correlations.

While loco type indicates partly controversial tendencies in subsets – all with weak statistical significance –, train orientation shows visible correlations regarding train type F in C1 (high statistical significance) with an increase of 6 % in the pulled stated compared to the pushed one. However, that trend is only hardly given (difference 1 %) in MC_5_TT_F. Furthermore, significant differences between single unit and double unit operation are observed. The former causes up to 20 % less frequency of occurrence (maximum between TrainOrientation.5 and TrainOrientation.1). Position of power coaches in double units is also relevant with lowest probability values in TrainOrientation.3 (power unit in front and middle position, refer to Figure 27) and highest in TrainOrientation.5 (both power units in the middle) – difference 5 %. The other two configurations lie in between. Thus, frequency of occurrence for single unit operation is equal to C5, however the overall probability in C1 is significantly higher because about 80 % of monitored trains are operated in double units. The reason behind is quite simple – a train pass bys of the same train type with more axles have a higher probability that one wheel is radiating flanging noise. Moreover, vehicle dynamics are not ideal in a pushed state because the traction force vector points straight forward while the train head already has to traverse the curve.

Distinct seasonal/monthly/weekly/daily alterations in frequency of occurrence among the year are visible (refer to Table 32). Beside their generally high deviation magnitudes, all depicted subsets show extreme values in January and partly lower levels in summer months. In C1, raised levels for both categories occur in October and November, while train type G in C5 (significantly higher statistical significance compared to train type F in the same section) indicate generally raised levels between September and December. Variations among months are large with up to 18 % and 31 % in C1 (train types E and F) as well as up to 64 % and 22 % in C5 (train types F and G) – excluding extremes in January. Detailed empirical examination reveals that those can be almost solely explained by relative humidity. Other environmental predictors – like temperature and absolute humidity – show correlations as well, however it seems that they represent the same trend as relative humidity. Their magnitude changes over the year significantly. Furthermore, the values are also dependent on general climatic conditions at certain sites. Both disadvantages can be countered by consideration of a relative humidity quantity. Figure 57 illustrates a function of absolute frequency of occurrence[18] regarding train pass bys with flanging noise occurrence (red) and without curve squeal in general (blue) in conjunction with relative rail humidity. From that, a relative frequency of occurrence among the range of relative rail humidity can be extracted (see Figure 58). As can be seen, considering the whole dataset, the probability of flanging noise occurrence exceeds 40 % between 46 % and 93 % and lies above 50 % in a range from 62 % to 77 % (peaking with 52 % at 72 %). That tendency is also visible in most subsets. Even in C2 and C3, where rather different climatic conditions are present and the campaign lasted only one month, a weak trend is visible, however sufficient validation over the year in those

[18]Since different sample quantities are depicted, a density function with a Gaussian kernel distribution using 512 points is calculated for both datasets and afterwards scaled by each sample quantity to achieve direct comparability.

datasets is not possible. Thus, the tendency seems to be section and train type independent. However, it has to be pointed out that such detailed examinations need a sufficient quantity of data points, which is only reached regarding train types E, F and G as well as in C1 and C5. Dry conditions ($<$ 40 % relative rail humidity) as well as humid ones ($>$ 100 % relative rail humidity) cause a significantly lower frequency of occurrence regarding flanging noise. The highly correlated predictor RT.DPDiff shows the same tendency in a value range between 1 °C and 12 °C (exceeding 40 %) and peaks above 50 % between 4 °C and 7 °C. Quantities derived with air temperature – Hum and AT.DPDiff – indicate roughly the same trend, however not as distinct as the former quantities. Therefore, consideration of rail temperature instead of air temperature seems necessary. The dependency on a narrow range of relative humidity leads to completely different daytime behaviour among the year. Hourly observations among months are provided in Table 33 for train type G in C5 as it is by far the most occurring, a very homogeneous one and independent to train orientation influence. Moreover, depiction ensures independence of deviating train type or section influence. However, it has to be pointed out that values depend largely on climatic conditions at different locations, where relative humidity distribution over the year can be completely different. Thus, monthly and daytime behaviour might deviate significantly from the illustration in other locations. Statistical significance is reached with usually monitored train pass bys in a range of 50 to 100 per month and hour. The only exceptions refer to January and May, where train counts drop to 12 to 40 per month and hour. The depicted hour value represents always the starting time – e.g. hour 5 covers a range between 5 a.m. and 5:59 a.m. Hour values, which are not added, lack of sufficient train pass by quantity. In months, where the mentioned humidity range is often present (e.g. January in C1 and C5 – in both, similar climatic conditions and altitudes are present), extreme frequency of occurrence distributed over the whole day is observable. The opposite is indicated in summer, where those conditions are usually only present in a narrow time in the morning and late in the evening. As can be seen, starting in September and peaking in January, overall frequency of occurrence raises and daytime influence begins to change from early morning/evening to morning/midday. Strong negative correlations are observed in conjunction with RainDetect (hardly any flanging noise during rain) in all considered subsets, which complies with past research. In the whole dataset, frequency of occurrence is reduced from overall 32.0 % (without any additional filtering of conditions) to 1.4 % in case of rain. Moreover, frost and hoarfrost conditions result in a large reduction of frequency of occurrence in all splits with decline magnitudes of 15 % to 20 % (train type E and G) and 35 % to 40 % (train type F). Interestingly, dew presence leads to a reduction, too, however not as much as rain or frost with about 8 % (train type E), 10 % to 20 % (train type F) and 15 % (train type G). Moreover, RainMmSLT in combination with RainDurSLT can also greatly reduce probability. The relationship behind depends on how fast the rail is drying, which is difficult to model and furthermore nearly impossible to include in a predictive model.

All other parameters show either a weak, a controversial or no trend at all in empirical examination.

train type	226 m (C4)	230 m (C5)	256 m (C1)	310 m (C3)	440 m (C2)	unit
A	66.7	68.4	78.7	54.0	55.6	%
B	69.2	75.4	–	40.0	33.3	%
C	–	–	4.8	28.9	30.2	%
E	–	–	23.2	–	20.9	%
F	31.6	43.1	56.4	–	6.2	%
G	31.0	28.3	17.5	–	5.5	%
other	24.1	26.2	16.4	18.9	0.0	%

Table 31: Frequency of occurrence regarding flanging noise distinguished by train type and curve radius with colour theme for statistical significance – red (weak), orange (mediocre), green (strong)

subset	month											
	1	2	3	4	5	6	7	8	9	10	11	12
MC_1_TT_E	79	14	27	32	15	16	22	–	–	32	31	17
MC_1_TT_F	100	44	58	59	68	49	68	–	–	75	70	50
MC_5_TT_F	69	13	33	77	56	69	62	45	66	45	19	45
MC_5_TT_G	67	27	19	23	23	29	16	18	36	38	30	36

Table 32: Monthly relative frequency of occurrence in units % regarding Broad over the year with colour theme for statistical significance – red (weak), orange (mediocre), green (strong)

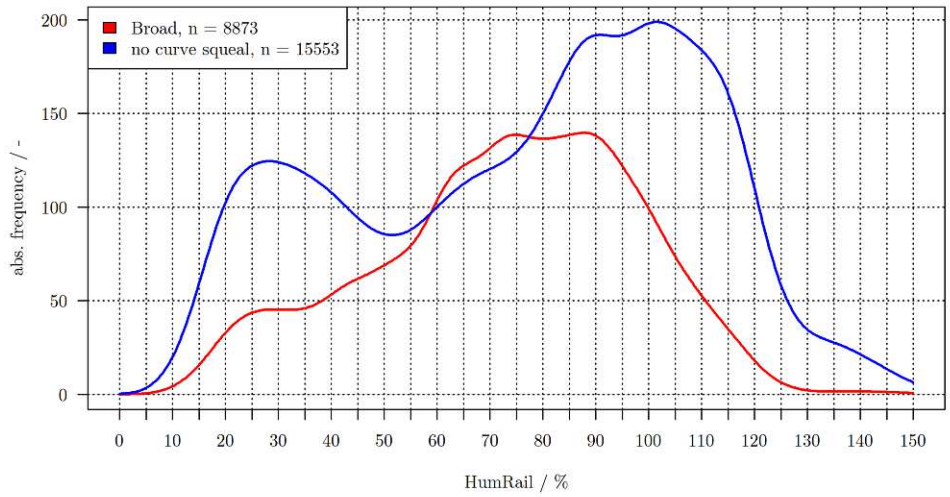

Figure 57: Absolute frequency function of relative rail humidity train pass by distribution with flanging noise occurrence (red) and without any type of curve squeal (blue) – data from the whole dataset

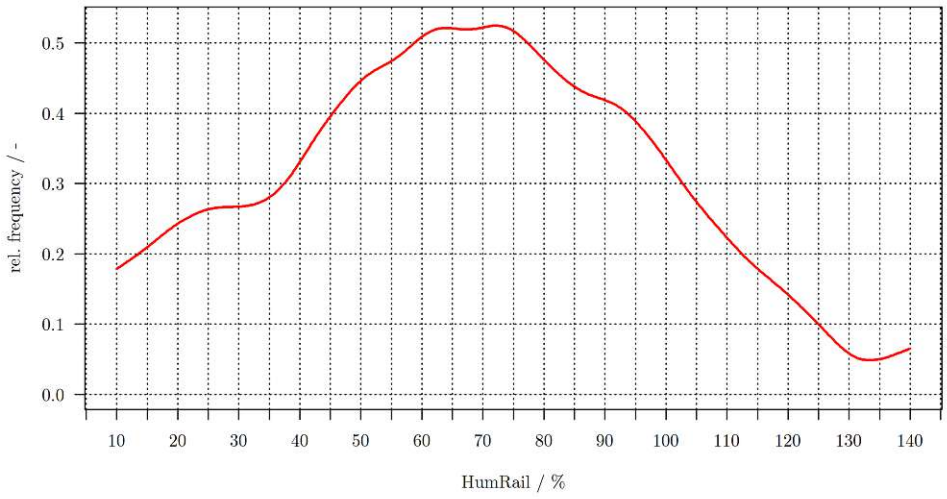

Figure 58: Relative frequency of flanging noise occurrence among the range of relative rail humidity – data from the whole dataset

hour	**month**											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
5												
6	+	+			+							
7	+			+			+	+				
8		+		+			+	+	+			
9	+	+		+					+			
10	+	+		+						+	+	
11		+									+	+
12	+	+									+	+
13											+	+
14	+									+		+
15	+						+					+
16	+									+	+	+
17	+		+							+		
18	+		+						+	+		
19	+		+						+	+		
20	+		+		+	+			+			
21	+				+	+		+	+			
22	+			+	+	+		+	+			
23					+	+	+		+			
range [%]	68-80	30-40	25-35	30-38	30-55	40-52	20-23	25-30	43-54	44-52	37-43	50-63

Table 33: Daytime influence over the year on flanging noise occurrence and additionally probability range in the marked area of the day (last row) – watch out, depicted time zone is UTC +1 and daylight saving time between March and October (UTC +2)

From all applied statistical models, complex tree algorithms (Boosted Tree and Random Forest) achieve the best performance with mostly distinct difference compared to linear or non-linear approaches. Both methods carry out embedded feature selection. Overall, Random Forest classification outperforms the Boosted Tree approach slightly (illustrated in Table 34), however computational effort is huge. Computation time with the whole dataset is 36 hours for Random Forest compared to 4.5 hours for the Boosted Tree algorithm. If computational cost is weighted high, the Bagged Tree algorithm (0.3 hours) is recommendable since it results only in slightly worse performance. In general, achieved benchmarks are good and due to applied five times repeated 10-fold cross validation, good predictive ability on the same boundary conditions is indicated. Moreover, the achieved benchmark of the whole dataset is similar compared to subsets, which indicates that the algorithms are capable of dealing with different sections and train types in one model. Referring to Table 34, the main difference to point out is the partly large reduction in quantity of used predictors in the Boosted Tree model.

subset	**Boosted Tree**		**Random Forest**	
	ROC	**number of predictors**	**ROC**	**number of predictors**
All	0.86	165	0.87	188
MC_1	0.85	54	0.86	64
MC_2	0.80	43	0.79	56
MC_4	0.85	46	0.85	171
MC_5	0.87	50	0.88	171
TT_A	0.73	74	0.74	169
TT_E	0.81	50	0.82	173
TT_F	0.85	126	0.87	174
TT_G	0.86	131	0.87	168
MC_1_TT_E	0.82	45	0.83	54
MC_1_TT_F	0.81	45	0.83	51
MC_4_TT_G	0.85	39	0.84	160
MC_5_TT_F	0.90	44	0.92	162
MC_5_TT_G	0.86	41	0.87	160

Table 34: Results of best performing regression models achieved with Boosted Tree and Random Forest

Since the present step is to select few competitive predictors for the final predictive modelling step, variable importance scores are compared in Figure 59 – depiction is capped at 25 % relative predictor importance. In addition to taking more predictors into account, the predictor importance scores in the Random Forest model decline not as steep as in the Boosted Tree. The cause is that in the latter every tree in each iteration is dependent on the tree in the previous iteration, thus, importance of high ranked predictors is magnified in each step of the loop (more about tree correlation is explained in Subsection 6.1).

Track parameters R, Cant, LongSlope, RailType and SleeperType are either not included

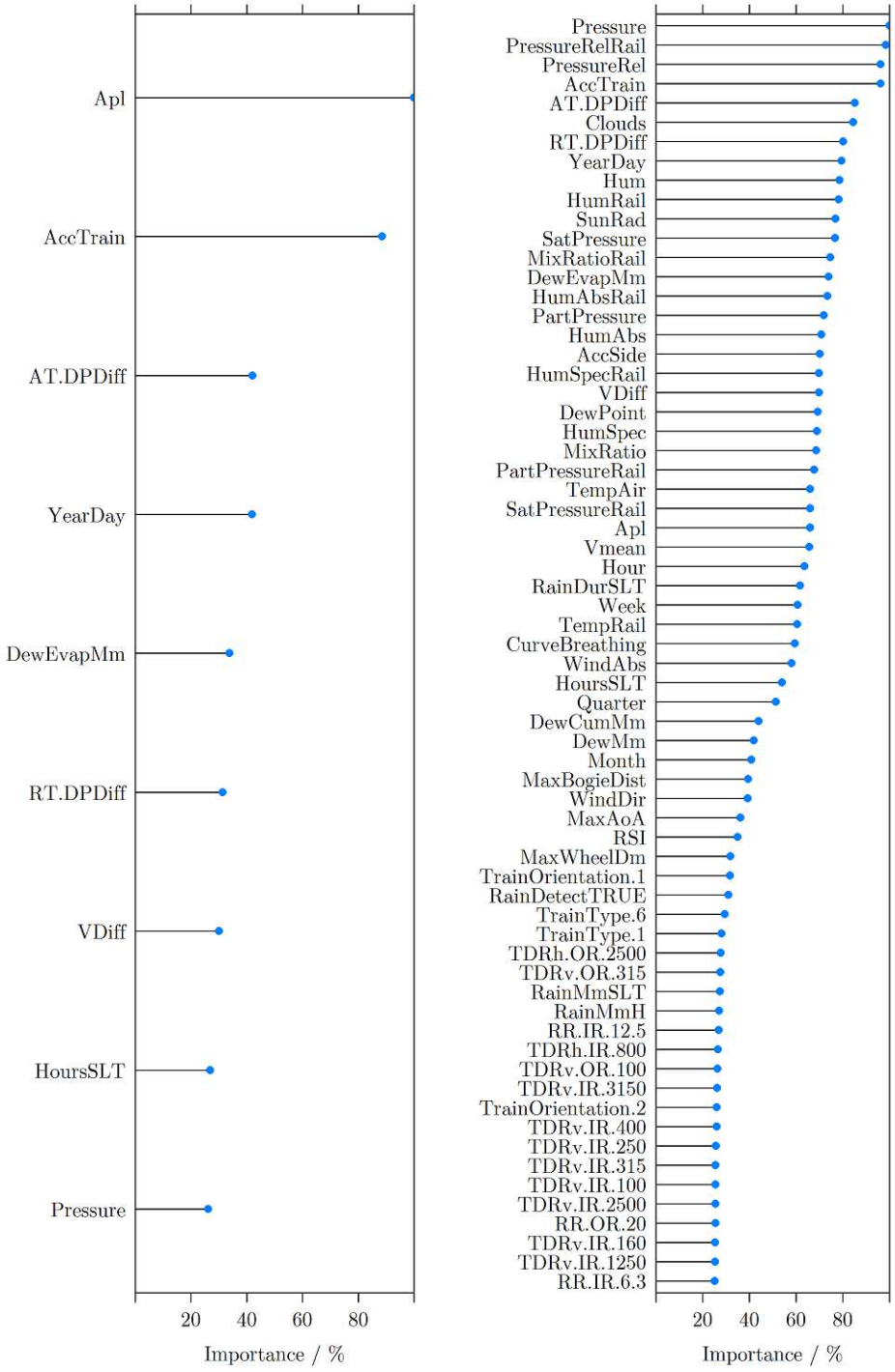

Figure 59: Relative predictor importance in the Boosted Tree (left) and Random Forest (right) model fitted to the whole dataset

at all (in Boosted Tree) or rated with a low importance score (in Random Forest). Few TDR and RR values are listed slightly above 25 %, however all others are ranked low or are excluded. Thus, significant section dependent influence on frequency of occurrence regarding flanging noise is not evaluated statistically, although empirical investigation indicates

differences partly (see Table 31). Therefore, observed deviations are taken into account in the shown statistical model by inclusion of other predictors. In terms of vehicle dependent predictors, Apl is rated top considering the whole dataset as well as ranked high in other subsets. An influence is also visible in empirical evaluation. By inclusion of Apl, train type distinction in general between freight and passenger trains as well as between single and double unit operation is possible. The latter is also separately included in Random Forest and thus a distinct influence is also detected statistically. However, axle count as well as train length is impossible to know in a prediction task. Hence, it is not suitable to consider Apl in a predictive model. The predictor group for bogie distance and wheel diameter – MaxBogieDist, MaxAoA, RSI and MaxWheelDm – is listed with relatively low importance. RSI and also the connected FreeCurving variable show the best correlation empirically and deliver a reasonable explanation – with growing wheel diameter and decreasing curve radius, the probability of traversing a curve without flange contact declines. TrainType.1 (type A) and TrainType.6 (type F) are also listed, presumably because both show a higher frequency of occurrence compared to other types. Dynamic predictors – especially AccTrain – are ranked high to mediocre. Empirical investigation shows hardly any influence of AccTrain, which indicates a misleading statistical correlation. The same applies to VDiff. Both predictors are compromised by the isolation joint in C5, where all trains loose speed due to a short section without traction. Overall dependency on train speed is not observed empirically. Regarding AccSide, a proportional relationship to probability is observable at least in tight curves and distinct regarding freight traffic in C4 and C5. The influence is reasonable since the outer wheels tend to flange contact with growing lateral acceleration. YearDay is ranked high, which indicates over-fitting of the model approaches. The Random Forest model lists also the other general predictors – Quarter, Month and Week – with mediocre to low importance. Influence is already discussed in empirical examination and justified by the statistical ranking. Interestingly, HoursSLT is also listed in both with low assigned importance score. No visible correlations are observed empirically, however influence might be compromised by freight trains operating in the night hours, where the predictor indicates a higher value. A bulk of environmental predictors is listed with high importance at least in the Random Forest model, which justifies their high impact further. Perfectly complying to the empirical examination, relative humidity indicators are top ranked – AT.DPDiff leads among them in both models. Pressure quantities are also top ranked, which is on the one hand for absolute pressure compromised by different section altitudes. On the other hand, impact of relative pressure quantities lacks of a physical explanation and is also only weakly shown in empirical investigation. Thus, it is assumed that they represent a misleading trend or at least can be sufficiently expressed by other environmental parameters. Temperature and absolute humidity quantities are also listed with high importance scores, however empirical examination indicates that they only depict the same trend as relative humidity with the downside of high magnitude changes over the year and also no comparability to other locations with different climatic conditions. Rain predictors are ranked with low to mediocre importance. Their impact is weakened statistically due to few data points, where rain is present at all. The same affects frost and hoarfrost, which causes that they are not listed at all.

To summarise, regarding track parameters, negative correlation to curve radius is indicated by freight and passenger trains – especially between narrow radii ($\leq$ 256 m) and higher ones (440 m). In terms of vehicle predictors, strong train type dependencies are present – freight trains indicate a generally high probability, train types C, D and F a mediocre one and the rest a lower probability value. Moreover, curve radius and wheel diameter influence can be represented by RSI, which shows a reasonable correlation to the outcome.

Separate vehicle influence is observed in terms of double unit operation and train orientation – distinguishing pushed and pulled state as well as different configuration in double units. The former is simply caused by a doubled axle quantity and raises frequency of occurrence by up to 20 %. The latter is foremost observable regarding single unit trainsets with a power coach (1 % to 6 %). Furthermore, train configuration of trainsets with power coaches in double unit operation shows deviations in a range of 5 % – with a worst case if both power units are in the middle and the best case if both are pulling their trainsets. Although train configuration shows a partly distinct influence, it is excluded from the final predictive model due to unknown alterations in operation.

Dynamic dependency is observed in terms of lateral acceleration. Speed difference, acceleration and deceleration as well as train speed show no general correlations.

Distinct alterations in probability magnitude among the year and additionally regarding daytime are observed. Detailed empirical investigation shows an almost exclusive connection to relative humidity quantities. Past research found an inverse relationship of relative surface humidity and friction coefficient [19]. However, the peak frequency of occurrence (exceeding 50 %) observed over the whole dataset lies between 62 % and 77 % relative rail humidity (peaking with 52 % at 72 %). In the remaining area, steep decreases below 45 % and above 90 % are present. That trend is also observed in most subsets and thus seems to be section and train type independent. The latter may be caused by beginning dew formation or rain occurrence. The tendency in dry air is not entirely explainable, nevertheless accords with the trend of low probability in conjunction with higher temperatures. Notably, relative humidity quantities derived with rail temperature show better and more distinct correlations to the outcome. Hence, consideration of rail temperature seems necessary. The correlation of relative humidity perfectly explains the alterations over the year and in daytime. In autumn and winter (between September and January), overall frequency of occurrence is higher due to more hours of relative rail humidity in the critical range (peaking in January). Hot months (July and August) show the least observed probability. Daytime influence is also connectable to relative rail humidity depending on typical hours, where critical values are present. In spring, summer and autumn, conditions favour occurrence early in the morning and in the afternoon (in summer only late in the evening), whereas in winter (especially November and December), daylight hours (late morning/midday) are predestined. In the latter months, frequent frost conditions are present, which additionally suppress flanging noise. It has to be pointed out that presented tendencies regarding seasons and daytime may alternate greatly in different local climatic conditions. Both measurement sections (C1 and C5), where those trends are observed, have similar climatic conditions as well as nearly the

parameter	name in model	type	value range	unit
train type dummy variables (1 mediocre, 2 low, rest zero)	TrainType.(1-2)	categorical	[0;1]	-
double unit operation	DoubleUnit	categorical	[0;1]	-
RSI	RSI	continuous	[0,∞[	-
lateral acceleration	AccSide	continuous	]-∞,∞[	ms^{-2}
difference between rail temperature and dew point	RT.DPDiff	continuous	]-∞,∞[	°C
rain occurrence	RainDetect	categorical	[0;1]	-
dew presence	DewPresent	categorical	[0;1]	-
frost conditions	FrostPresent	categorical	[0;1]	-

Table 35: Mathematical description of the final parameter set applied to predict Broad

same altitude. In general, daily variations in frequency of occurrence are well connectable to humid and/or frosty days, which leads to a significant decrease. In general, rain shows the most reduction in probability (decline in the whole dataset from 32.0 % without rain to 1.4 % train pass bys with detected flanging noise while rain), followed by hoarfrost (from 30.6 % to 2.5 %), frost (from 30.8 % to 8.7 %) and dew presence (from 29.9 % to 18.2 %).

Building on the results of empirical and statistical investigation and pursuing the scope of constructing a practical model, the following predictors are chosen for predictive modelling:

- Train type distinguished in three categories depending on their basic tendency to cause flanging noise – high (type A and B), mediocre (type C, D, F), low (type E, G, other)
- double unit operation
- RSI (represents curve radius and wheel diameter)
- lateral acceleration (leads to inclusion of radius and cant as well)
- relative humidity indicator (best performing predictor RT.DPDiff evaluated in Subsection 6.4)
- rain occurrence
- dew presence
- frost conditions (contains frost and hoarfrost)

Thus, predictor quantity is shrunk to nine. Table 35 provides a mathematical description of each predictor.

6.3.3 Squeal noise occurrence

Univariate statistical analysis in conjunction with squeal noise occurrence (referred as TonalLow) is not able to identify distinct influences (refer to Figure 60) among the whole dataset as well as in subsets. Track parameters (in first few places only TDR and RR values) are ranked top in combination with some high ranks of vehicle predictors – TrainType.7 (type G), RSI and Apl –, which indicates large differences between sections and train types. However, distinct train type differences lead to high rankings of track parameters since significant deviations in train type distribution among sections are present. Environmental, general and dynamic predictors are ranked in the lower half. Differences in scores among parameters are low. Thus, empirical findings are discussed without comparison to univariate statistical outcome. The second part discusses performance and findings of applied multivariate statistical methods. The subsection is concluded by a brief summary of most important trends and building on that final influence parameters for predictive modelling are defined. General tendencies in sections and in conjunction with different train types are discussed in Subsection 5.7 as well as depicted there in Figure 43. To avoid influence from train type distribution and section dependencies overall and in subsets "TT" and "MC", the empirical evaluation is focused on "MC_TT" splits.

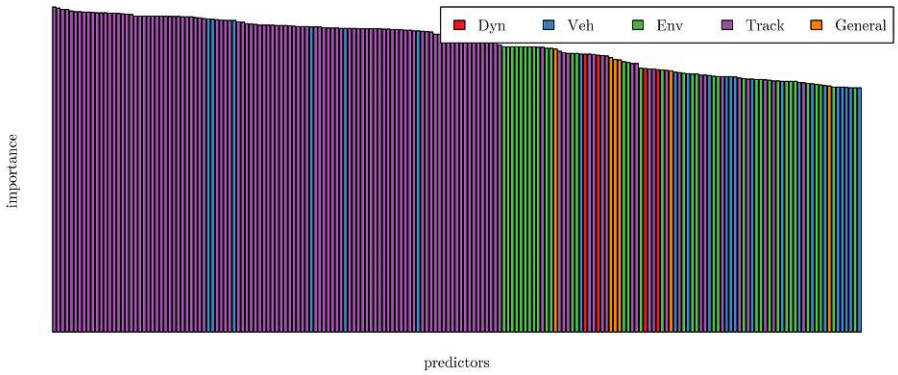

Figure 60: Predictor importance regarding squeal noise occurrence evaluated with a univariate ROC-curve evaluation (scores between 0.67 and 0.50) and coloured in groups dynamic (red), vehicle (blue), environmental (green), track (purple) and general (orange)

Detailed empirical analysis reveals a significant dependency on curve radius – in general less frequency of occurrence with increasing radius (illustrated in Table 36). It has to be pointed out that some values are derived with low statistical significance (red). Train type D only operates in C3 and is therefore skipped in the illustration. The significant increase of train type F between C4/C5 and C1 is explainable by an increased frequency of occurrence in double unit operation (further discussed below), while train type G and freight trains indicate a declining trend. A significant drop in frequency of occurrence is observable between all narrower radii ($\leq$ 310 m) and 440 m, while in the former range no steep decreases are shown

– apart from freight trains in C4 and C5, where probability values are extreme. Notably, the latter correlates with very low Apl values of 0.14 m^{-1} compared to monitored freight trains in other sections, which indicates a large distance between axles or bogies worsening the curving performance especially in conjunction with low curve radii. Regarding train type differences in narrow radii ($\leq$ 256 m), freight trains (excluding the abnormally high values in C4/C5 and evaluated double unit impact) show similar probability rates as passenger train categories C, E and F (13 % to 18 %), while in terms of train type G and other, a lower frequency of occurrence is observable (0 % to 10 %).

In addition to the illustrated deviations among train types, train orientation has a significant impact with differences of 5 % (train type E in C1), 8 % (train type F in C1, single units only) and 3 % (train type F in C5, single units only) – pushed operation always higher than pulled. Double unit operation leads to an increase of up to 22 % (the assumed reason is simply more axle quantity). Distinction of configuration in double units results in differences between TrainOrientation.3 (best case – both power units pulling their trainsets) and TrainOrientation.4 (worst case – power coaches in first and last position) of 6 %. The remaining configurations lie in between. While differences between pushed and pulled state comply with observations of flanging noise, train orientation in double units seems to cause more problems with a pulling power coach at the beginning and a pushing one at the end, which may lead to increased lateral creepage. RSI allows to consider curve radius and wheel diameter in one predictor and indicates an influence over the whole dataset.

In terms of dynamic predictors, AccSide shows a proportional relationship to frequency of occurrence, which leads to the assumption that lateral acceleration may also influence lateral creepage and not only flange contact. Other dynamic predictors show no general influence.

Deviations among the year and in daytime are observable in the dataset. Among months of the year (illustrated in Table 37) three out of four considered subsets indicate extreme values in January. Beside partly controversial trends, foremost train type G in C5, which states by far the most statistical significance, indicates lower values in summer months. Both tendencies comply with findings in terms of frequency of occurrence regarding flanging noise. Variations among months are large with up to 37 % and 34 % in C1 (train types E and F) as well as up to 25 % and 6 % in C5 (train types F and G) – excluding extremes in January. Detailed investigation reveals a distinct connection to relative humidity – absolute frequency function of train pass bys without any type of curve squeal (blue) and with detected squeal noise (red) among the range of relative rail humidity depicted in Figure 61 for the whole dataset[19]. From that, a relative frequency of occurrence over the whole range of relative rail humidity can be computed (refer to Figure 62). As can be seen, values between 60 % and 90 % lead to a significantly higher probability (exceeding 25 % and peaking with 30 % at 75 %), while in the remaining range probability values are steeply decreasing. That tendency is also visible in most subsets. Even in C2 and C3, where rather different climatic conditions are present and the campaign lasted only one month, a weak trend is visible, however

[19]Since different sample quantities are depicted, a density function with a Gaussian kernel distribution using 512 points is calculated for both datasets and afterwards scaled by each sample quantity to achieve direct comparability.

sufficient validation over the year in those datasets is not possible. Thus, the tendency seems to be section and train type independent. In very humid conditions ($\geq$ 100 %), the trend is reasonable due to common dew formation as well as probably rain, whereas the decline in dry conditions ($\leq$ 45 %) frequency of occurrence complies with past research [37, 62]. The correlated predictor RT.DPDiff shows the same tendency with exceeding 25 % relative frequency of occurrence between 2 °C and 7 °C and peaking at 4.5 °C with 29 %. Relative air humidity as well as the correlated predictor AT.DPDiff indicate roughly the same trend, however not as distinct. Thus, consideration of rail temperature seems necessary. The dependency on relative humidity leads to monthly deviations in frequency of occurrence with a generally higher value in autumn and winter (beginning in September, peaking in January and extending in C1 to February). Moreover, a partly distinct negative correlation to temperature and absolute humidity quantities is observable leading to very low probability rates in summer months (especially in July). However, it is thought that this effect can be sufficiently depicted by relative humidity as well – with advantages of no significant magnitude change among the year as well as comparability between different climatic conditions. Daily alterations are well connectable to rainy and/or frosty days, which is also visible directly by evaluating the predictors for wet and frost conditions. Considering the whole dataset, rain during a train pass by leads to a reduction of 15 % (to a total frequency of occurrence of 1.5 %), hoarfrost to a decrease of 16 % (to under 1 %), frost occurrence to a declination of 12 % (total 4 %) and dew presence to a reduction of 7 % (remaining probability 9 %). Moreover, RainMmSLT in combination with RainDurSLT can also greatly reduce probability. The relationship behind depends on how fast the rail is drying, which is difficult to model and furthermore nearly impossible to include in a predictive model. The most relying evaluation in terms of daytime influence represents the behaviour of train type G in C5 as train quantity exceeds other subsets by far, it states a very homogeneous train type and it is independent to train orientation influence. Daytime pattern among every month is provided in Table 38. Statistical significance is reached with usually monitored train pass bys in a range of 50 to 100 per month and hour except January and May, where train counts drop to 12 to 40 per month and hour. The depicted hour value represents always the starting time – e.g. hour 5 covers a range between 5 a.m. and 5:59 a.m. Hour values, which are not added, lack of sufficient train pass by quantity. It can be clearly seen that peak values of frequency of occurrence over daytime differs significantly – changing from summer (early morning, late in the evening) to daylight hours (late morning/midday). Thus, the observed pattern is similar to flanging noise occurrence. Summer months July and August show nearly no squeal noise incident in the evening hours, which indicates dry and warm conditions. Among the hottest hours around midday and early afternoon, hardly any squeal events are detected between May and August. Usually, the lowest relative humidity values are reached in those hours. It has to be pointed out that values depend largely on climatic conditions at different locations, where relative humidity distribution over the year can be completely different. Thus, monthly and daytime behaviour might deviate significantly from the illustration in other locations.

train type	226 m (C4)	230 m (C5)	256 m (C1)	310 m (C3)	440 m (C2)	unit
A	66.7	78.9	17.1	9.2	1.9	%
B	61.5	78.7	–	33.3	6.7	%
C	–	–	4.8	15.1	4.4	%
E	–	–	17.5	-	1.4	%
F	16.4	12.5	31.1	–	0.0	%
G	7.7	7.3	2.6	–	0.0	%
other	10.3	0.0	3.6	0.0	0.0	%

Table 36: Frequency of occurrence regarding squeal noise distinguished by train type and curve radius with colour theme for statistical significance – red (weak), orange (mediocre), green (strong)

subset	month											
	1	2	3	4	5	6	7	8	9	10	11	12
MC_1_TT_E	16	41	19	7	7	19	22	–	–	5	12	4
MC_1_TT_F	63	37	28	24	14	20	13	–	–	47	45	38
MC_5_TT_F	31	7	2	18	4	13	6	9	27	8	6	20
MC_5_TT_G	25	8	6	8	5	4	3	3	9	9	4	9

Table 37: Monthly relative frequency of occurrence in units % regarding TonalLow over the year with colour theme for statistical significance – red (weak), orange (mediocre), green (strong)

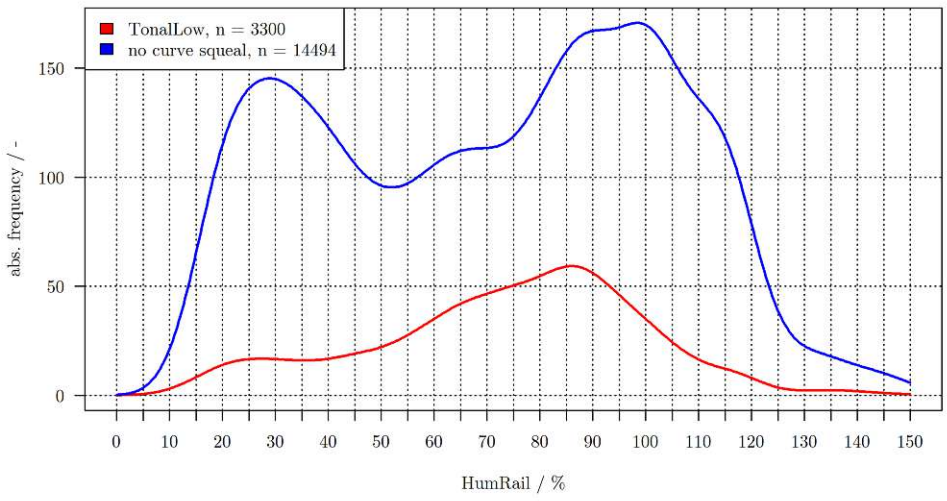

Figure 61: Absolute frequency function of relative rail humidity train pass by distribution with squeal noise occurrence (red) and without any type of curve squeal (blue) – data from the whole dataset

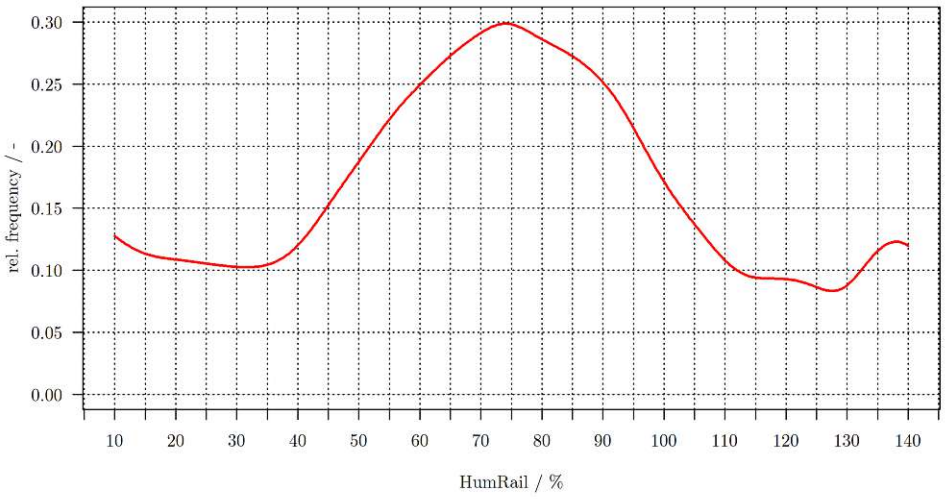

Figure 62: Relative frequency of squeal noise occurrence among the range of relative rail humidity – data from the whole dataset

hour	**month**											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
5						+	+					
6	+					+						
7				+	+	+	+	+	+			
8				+				+	+			
9		+	+	+					+	+	+	+
10			+							+	+	
11												
12												+
13	+											+
14												+
15	+									+		+
16	+											
17	+									+		
18										+		
19	+		+								+	
20		+	+							+	+	
21		+	+		+	+			+			
22	+				+	+			+	+		
23					+	+			+	+		
range [%]	30-58	14-17	9-12	15-20	14-25	7-8	8-9	15-16	12-18	10-13	6-9	13-19

Table 38: Daytime influence over the year on squeal noise occurrence and additionally probability range in the marked area of the day (last row) – watch out, depicted time zone is UTC +1 and daylight saving time between March and October (UTC +2)

Regarding applied statistical model approaches, Boosted Tree and Random Forest outperform the other algorithms regarding the whole dataset as well as all subsets significantly. However, for subset MC_2 even those complex and robust methods lack to find a prediction solution (only few squeal noise incidents at all). Thus, depiction is skipped. Table 39 provides an overview of achieved benchmarks. Good predictive performance in detecting squeal noise events on the same boundary conditions is indicated due to application of five times repeated 10-fold cross validation. Benchmark differences between the whole dataset and subsets are low leading to the conclusion that statistical algorithms are capable of dealing with different sections and train types. Again, the computational cost of Random Forest exceeds the Boosted Tree approach significantly. Moreover, included predictor quantity, is substantially lower in most sets, which makes Boosted Tree a reasonable choice for the task of feature selection. If computational effort is weighted high, the Bagged Tree algorithm is recommended leading to a minor performance loss.

subset	**Boosted Tree**		**Random Forest**	
	ROC	**number of predictors**	**ROC**	**number of predictors**
All	0.84	156	0.86	188
MC_1	0.79	55	0.80	64
MC_4	0.87	50	0.87	171
MC_5	0.84	49	0.86	171
TT_A	0.82	90	0.80	169
TT_E	0.87	48	0.88	173
TT_F	0.76	117	0.77	174
TT_G	0.85	126	0.86	168
MC_1_TT_E	0.86	46	0.87	54
MC_1_TT_F	0.69	47	0.70	51
MC_4_TT_G	0.90	37	0.89	160
MC_5_TT_F	0.82	44	0.85	162
MC_5_TT_G	0.84	42	0.85	160

Table 39: Results of best performing regression models achieved with Boosted Tree and Random Forest

Top predictor importance scores (refer to Figure 63) in both models are achieved by Apl, YearDay and VDiff/AccTrain. The former is compromised by the abnormally high frequency of occurrence of freight trains in C4 and C5, which show a notably low Apl, into account. Moreover, train length and axle quantity are impossible to know in a prediction task. YearDay depicts daily alterations, which are previously discussed in empirical examination. Although distinct variations occur, the predictor is not suitable for predictive modelling because conditions are different every year as well as location dependent. Thus, inclusion of other variables (like rain or frost presence) are preferable. Furthermore, the high ranking indicates over-fitting of the model. VDiff and AccTrain predictors are compromised by data

from C5, where every train pass by looses speed. Therefore, the evaluated correlation is misleading. A promising predictor states RSI. Although it is an indicator of flange contact probability, it depicts a general parameter of curving behaviour and thus might be useful in conjunction with squeal noise as well. Moreover, it also includes curve radius, which shows a distinct influence. The latter is the reason of partly high ranks regarding predictors, which deviate among sections (like Vmean and listed track parameters). Dynamically, AccSide is ranked in mid-range, which complies with empirical investigation. A mediocre ranking of HoursSLT is noticeable, which also shows a weak influence in empirical examination – among subsets, an inverse proportional relationship is present. Reasons may be dew formation during night breaks in operation as well as other influences altering the third-body layer. The correlation is also supported by daytime analysis as significantly lower frequency of occurrence is observed for the first operated trains in the morning. However, a separate inclusion is questionable since if the cause is dew formation, it is covered by the corresponding categorical variable. Train orientation and environmental influence is discussed in empirical examination and complies well with statistical rankings. Regarding humidity, relative air humidity or the correlated predictor AT.DPDiff is ranked higher than relative rail humidity and RT.DPDiff, which is contrary to empirical findings. In terms of temperature and absolute humidity predictors, SatPressure and HumAbsRail are listed with highest importance. SunRad depicts only influence in daylight hours and correlates well with the outcome. A trend is also visible in empirical examination. Despite found correlations, it is believed that influence of temperature, absolute humidity and sun radiation can be sufficiently and independently from local climatic conditions represented by a relative humidity indicator. All statistical evaluations tend to underrate categorical predictors like rain or frost presence. The assumed reason behind is that few data points contain those conditions and thus their general statistical impact is weakened. However, strong correlations are observed empirically.

To conclude, considering the whole dataset, distinct influence of curve radius on frequency of occurrence – decreasing with growing radius – is observed. That perfectly complies with past research. Possibilities of model integration are implicit inclusion by AccSide and RSI and/or explicitly with a separate predictor. The implicit approach is taken due to the scope of predictor reduction since two predictors include multiple trends together. Other track predictors show no general correlation with the outcome – neither empirically nor statistically.

Train type influence is distinct, allowing to summarise two groups – mediocre frequency of occurrence (train types A, B, C, E and F) and low one (train type G and other). Train type D is only monitored in C3 and operates with similar coaches as train type C, thus, it is added to the former category. Exceptional high probability rates are observed in terms of freight trains in C4/C5, which show notably low axle per length values and thus indicate severe curving behaviour in conjunction with narrow radii ($\leq$ 230 m). Curving behaviour in general is best represented by RSI. Distinct influence is observed in terms of double unit operation (up to 22 % more frequency of occurrence in double units), which is simply reasoned by doubled axle quantity. Moreover, pushed operation (in trainsets with power coach as well as in conjunction with locos) causes an up to 8 % higher probability. In case of double operation of trainsets

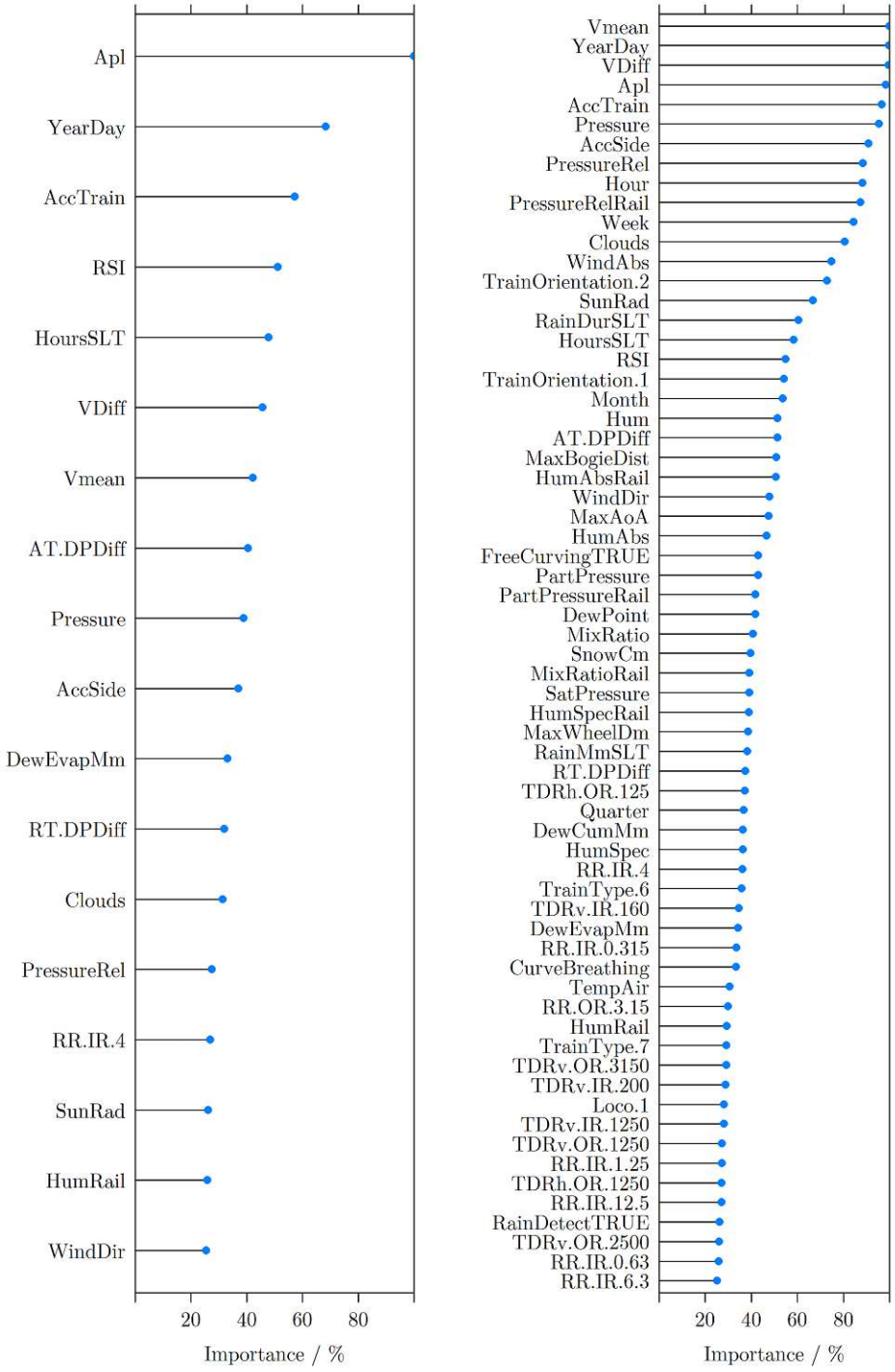

Figure 63: Relative predictor importance in the Boosted Tree (left) and Random Forest (right) model fitted to the whole dataset

with power coaches, a variation of 6 % depending on train configuration is evaluated (best case: both power coaches pulling their trainsets; wort case: one power coach pulling in front and one pushing at the end). Although train configuration shows a strong influence, an inclusion in a predictive model is not possible due to lack of knowledge in real operation.

Dynamic influence is present regarding lateral acceleration, which has presumably an effect on lateral creepage. Mean speed as well as deceleration or acceleration show no direct influence on frequency of occurrence.

Distinct deviations among the year and in daytime are present. Daily observations are perfectly connectable to rain or frost occurrence leading to days with few squeal noise incidences. Considering the whole dataset, rain during a train pass by leads to a reduction of 15 % (to a total frequency of occurrence of 1.5 %), hoarfrost to a decrease of 16 % (to under 1 %), frost occurrence to a declination of 12 % (total 4 %) and dew presence to a reduction of 7 % (remaining probability 9 %). Overall influence over the year is almost solely justifiable by relative humidity alterations. A critical range between 60 % and 90 % relative rail humidity leads to probability values of above 25 % (peaking at 75 % with 30 %), whereas in the remaining area relative frequency declines significantly. That trend is observed in the whole dataset as well as in most subsets and thus seems to be section and train type independent. In months, where the critical range is frequently reached, frequency of occurrence is higher. Extreme values are present in January, while probability in summer months (especially June and July) is very low. Daytime behaviour is also explainable in terms of in which hours the mentioned range in relative rail humidity is commonly reached. While in winter months, mainly daylight hours indicate a higher frequency of occurrence, in the remaining seasons, most probability peaks are connectable to dawn and/or sunset. In addition, in late spring and summer months (May to August), an even lower to zero frequency of occurrence is observed in the hours around midday and afternoon, where relative humidity is mostly lowest.

Building on the results of empirical and statistical investigation and pursuing the scope of constructing a practical model, the following predictors are chosen for predictive modelling:

- Train type distinguished in two categories depending on their basic tendency to cause squeal noise – mediocre (type A, B, C, D, E, F), low (type G, other)
- double unit operation
- RSI (represents curve radius and wheel diameter)
- lateral acceleration (leads to inclusion of radius and cant as well)
- relative humidity (evaluated as best performing predictor in Subsection 6.4)
- rain occurrence
- dew presence
- frost conditions (frost or hoarfrost likely)

Thus, predictor quantity is shrunk to eight. Table 40 provides a mathematical description of each predictor.

parameter	name in model	type	value range	unit
train type dummy variable (1 mediocre, rest zero)	TrainType.1	categorical	[0;1]	-
double unit operation	DoubleUnit	categorical	[0;1]	-
RSI	RSI	continuous	[0,∞[	-
lateral acceleration	AccSide	continuous	]-∞,∞[	ms^{-2}
relative humidity	Hum	continuous	[0,100]	%
rain occurrence	RainDetect	categorical	[0;1]	-
dew presence	DewPresent	categorical	[0;1]	-
frost conditions	FrostPresent	categorical	[0;1]	-

Table 40: Mathematical description of the final parameter set applied to predict TonalLow

6.3.4 HF squeal noise occurrence

Univariate statistical evaluation of HF squeal noise occurrence (termed as TonalHigh) is generally not able to extract a group of highly important predictors among the whole dataset as well as in subsets (depicted in Figure 64). The only exception states the first place – CurveBreathing – with a ROC benchmark of 0.69, which is decent in terms of a univariate predictor relation to the outcome. Curve breathing is derived by rail temperature, curve length and radius. A closer look reveals that magnitudes of curve breathing are very high in C2 (due to large curve length) and the opposite in C4 and C5 (due to a short curve length). The other sections lie in between. Since frequency of occurrence decreases with growing curve radii (further discussed below), the detected correlation of curve breathing is compromised if the whole dataset is evaluated. In general, the upper half is dominated by track predictors. Temperature and absolute humidity predictors are also ranked high. First vehicle predictors follow in midfield. Most environmental predictors as well as general and dynamic ones are ranked in the lower half with poor benchmarks. However, differences among predictors are low. Thus, empirical findings are discussed without comparison to univariate statistical outcome. The second part discusses performance and findings of applied multivariate statistical methods. The subsection is concluded by a brief summary of most important trends and building on that final influence parameters for predictive modelling are defined. General tendencies in sections and in conjunction with different train types are discussed in Subsection 5.7 as well as depicted there in Figure 44. To avoid influences from train type distribution and section dependencies overall and in subsets "TT" and "MC", the empirical evaluation is focused on "MC_TT" splits.

Empirical investigation reveals a distinct dependency on curve radius – declining frequency of occurrence with growing radius. Table 41 provides an overview of probability values separated by section and train type. Statistical significance is accounted by a colour theme. It has to be pointed out that some values are derived with low statistical significance (red). Train

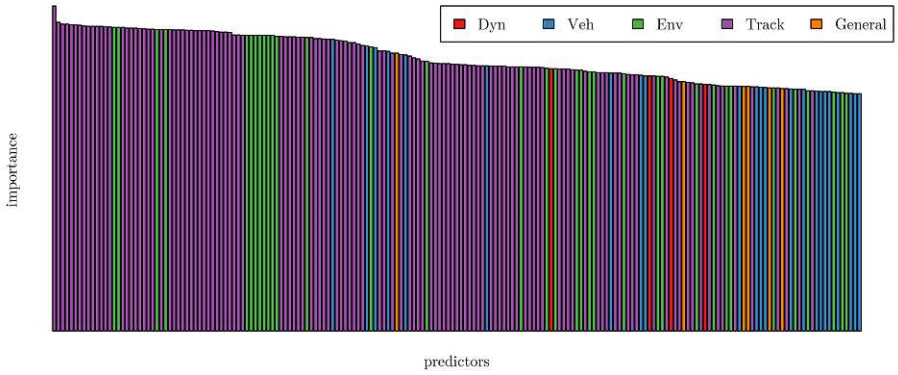

Figure 64: Predictor importance regarding HF squeal noise occurrence evaluated with a univariate ROC-curve evaluation (scores between 0.69 and 0.50) and coloured in groups dynamic (red), vehicle (blue), environmental (green), track (purple) and general (orange)

type D only operates in C3 and is therefore skipped in the illustration. With a frequency of occurrence of 15.7 %, the category indicates slightly lower values as train type C. Rates among the majority of other train types have one in common – a partly large decrease of probability between narrower radii ($\leq$ 310 m) and 440 m. Moreover, values between 226 m and 256 m show deviating trends depending on train types. In case of train type F, an increasing tendency between 230 m and 256 m is present, which is caused by higher frequency of occurrence in terms of double unit operation in C1 (further discussed below). Freight trains show a continuous declination among radii – only train type B shows an increase between 310 m and 440 m, however with weak statistical significance. Altogether, among narrow radii ($\leq$ 256 m), three similar train type groups can be extracted (outliers as well as double unit impact excluded in the stated ranges) – train types A and B with a high (22 % to 54 %), train types C, D, E and F with a mediocre (10 % to 21 %) and train types G and other with low (0 % to 11 %) frequency of occurrence. In general, probability ranges are mostly a bit higher compared to squeal noise, however much lower than those of flanging noise. Thus, the excitation mechanism of HF squeal noise might be similar to squeal noise, although time behaviour indicates rather a connection to flanging noise (refer to Chapter 4). Other track parameters do not seem to be directly correlated to the outcome.

Mediocre differences among train types occur (depicted in Table 41). Probability values of freight trains are high in C4 and C5 compared to other sections and train types, however, magnitude is not as severe as in terms of squeal noise. They show a notably low Apl value of 0.14, which indicates large distance between axles or bogies worsening their curving performance especially in conjunction with narrow radii (in the present case 226 m and 230 m). Interestingly, that seems to have significant impact on squeal noise and HF squeal noise occurrence, whereas no effect is observed regarding flanging noise occurrence. While in C1 and C3, passenger and freight trains show a similar frequency of occurrence (except train type G in C1), C2 indicates a significantly higher value in conjunction with freight trains. Double

unit operation increases frequency of occurrence by up to 12 %. General train orientation shows a controversial behaviour, leading to raises in the pushed state of 2 % (train type E in C1) and 9 % (train type F in C1, single units only) and causing a declination of 2 % in terms of train type F in C5. Train configuration in double units shows a 3 % lower value in TrainOrienation.4 (one power coach pulling in front and one pushing at the end) compared to all other configurations. Thus, alterations in vehicle dynamics depending on power unit position show little and partly controversial influence. Considering other vehicle parameters, only RSI shows a direct relationship to the outcome in empirical examination.

Regarding dynamic predictors, weak positive correlation is observed in conjunction with AccSide. VDiff, AccTrain and Vmean show no general correlations.

Distinct alterations over the year are observable among the whole dataset as well as all subsets. Monthly distribution (refer to Table 42) shows extreme values among all considered subsets in January and partly low ones in summer months. Both trends are identically shown in terms of flanging and squeal noise. Thus, that tendency refers to all distinguished curve squeal types. Beside partly controversial trends in other months, most subsets indicate a raise between October and December. Variations among months are large with up to 23 % and 29 % in C1 (train types E and F) as well as up to 22 % and 16 % in C5 (train types F and G) – excluding extremes in January. As main reason, relative humidity is identified. Figure 65 depicts a function of absolute frequency of occurrence[20] regarding train pass bys with HF squeal noise occurrence (red) and without any type of curve squeal (blue) in conjunction with relative rail humidity. From that, a relative frequency of occurrence among the range of relative rail humidity can be extracted (see Figure 66). Critical values lie between 59 % and 83 % exceeding a relative frequency of occurrence of 30 % and peaking to 34 % at 74 %. In the remaining range, steep decreases in both directions are present. That tendency is also visible in most subsets. Even in C2 and C3, where rather different climatic conditions are present and the campaign lasted only one month, a weak trend is visible, however sufficient validation over the year in those datasets is not possible. Thus, the tendency seems to be section and train type independent. Relative frequency of occurrence drops beneath 20 % in rather dry conditions ($\leq$ 54 % relative rail humidity) and in humid conditions ($\geq$ 100 % relative rail humidity). While the latter is explainable with beginning dew formation and/or probably rain, the reason for the former trend might be explainable with past research [37, 62] on squeal noise, however, it is unsure, since the excitation mechanism of HF squeal noise is not entirely clear. The correlated predictor RT.DPDiff shows the same tendency with exceeding 30 % relative frequency of occurrence between 2.5 °C and 6.5 °C and peaking at 4.5 °C with 34 %. Relative air humidity as well as the correlated predictor AT.DPDiff indicate roughly the same trend, however not as distinct. Thus, consideration of rail temperature seems necessary. Negative correlations to the outcome are observed regarding temperature and absolute humidity. However, it seems that they represent an identical trend as relative humidity and thus separate consideration is not needed. Moreover, advantages

[20]Since different sample quantities are depicted, a density function with a Gaussian kernel distribution using 512 points is calculated for both datasets and afterwards scaled by each sample quantity to achieve direct comparability.

train type	226 m (C4)	230 m (C5)	256 m (C1)	310 m (C3)	440 m (C2)	unit
A	52.4	47.4	21.6	14.8	8.9	%
B	53.8	47.5	–	6.7	13.3	%
C	–	–	4.8	17.0	0.7	%
E	–	–	21.1	–	1.1	%
F	10.2	14.5	23.1	–	0.0	%
G	11.1	9.4	10.0	–	0.0	%
other	3.4	4.9	0.0	0.0	0.0	%

Table 41: Frequency of occurrence regarding HF squeal noise distinguished by train type and curve radius with colour theme for statistical significance – red (weak), orange (mediocre), green (strong)

arise by inclusion of a relative humidity indicator in terms of no significant magnitude change among the year as well as comparability between different climatic conditions. Daytime behaviour deviates also strongly among the year, which complies with frequent occurrence of critical relative humidity values and the suppression effect in dry air conditions. Daytime tendencies and probability ranges in every month are provided in Table 43 for subset MC_5_TT_G. The choice is justified with by far most monitored train pass bys as well as independence to train orientation influence. Furthermore, it states a very homogeneous train type. Statistical significance is reached with usually monitored train pass bys in a range of 50 to 100 per month and hour except January and May, where train counts drop to 12 to 40 per month and hour. The depicted hour value represents always the starting time – e.g. hour 5 covers a range between 5 a.m. and 5:59 a.m. While frequency of occurrence in summer months (especially in July and August) is very low and peaks are in a narrow time range in the morning and evening hours, distribution in autumn and winter months change to daylight hours as well as absolute probability is also increased (extreme in January). It has to be pointed out that values depend largely on climatic conditions at different locations, where relative humidity distribution over the year can be completely different. Thus, monthly and daytime behaviour might deviate significantly from the illustration in other locations. In general, distribution changes are connectable to variations in dawn and sunset times among the year and can be fully explained with common relative humidity ranges depending on typical climatic conditions among months and daytime. Daily changes in frequency of occurrence correlate with rain and/or frost conditions. Considering the whole dataset, rain during the train pass by reduces probability by 16 % (to a total of 1.1 %), hoarfrost causes a decrease of 14 % (remaining 2.1 %), frost presence shows a declination of 10 % (total 6.3 %) and dew on the rail indicates 5 % reduction (to 10.3 %). Thus, distinct influence is especially present in terms of rain and hoarfrost, while dew seems to have a lower impact. Moreover, RainMmSLT in combination with RainDurSLT can also greatly reduce probability. The relationship behind depends on how fast the rail is drying, which is difficult to model and furthermore nearly impossible to include in a predictive model.

subset	month											
	1	2	3	4	5	6	7	8	9	10	11	12
MC_1_TT_E	74	18	27	27	15	7	4	–	–	21	26	25
MC_1_TT_F	50	25	19	18	13	8	15	–	–	37	37	26
MC_5_TT_F	61	5	7	17	10	15	16	8	14	16	9	27
MC_5_TT_G	37	11	5	6	6	8	3	3	8	10	13	19

Table 42: Monthly relative frequency of occurrence in units % regarding TonalHigh over the year with colour theme for statistical significance – red (weak), orange (mediocre), green (strong)

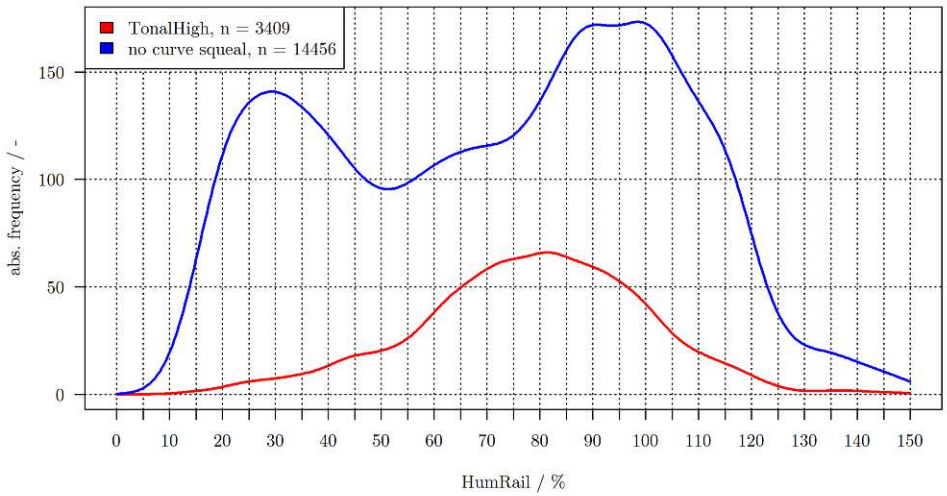

Figure 65: Absolute frequency function of relative rail humidity train pass by distribution with HF squeal noise occurrence (red) and without any type of curve squeal (blue) – data from the whole dataset

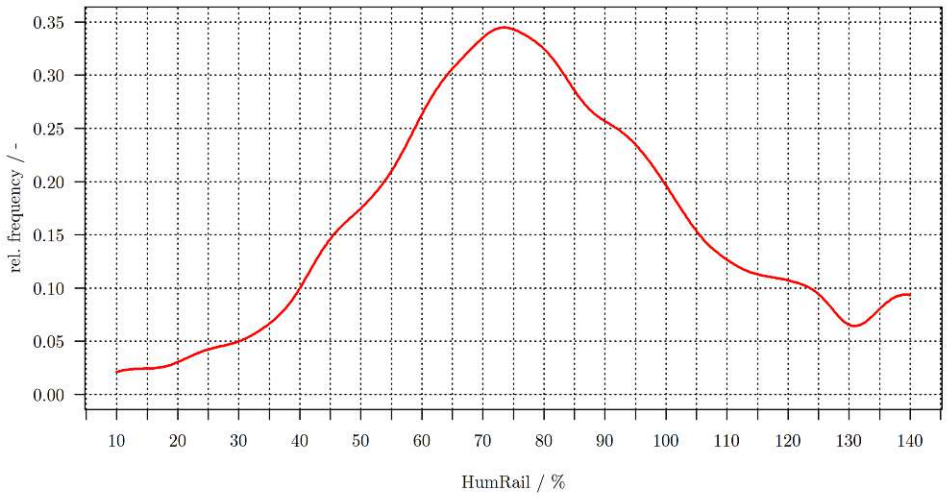

Figure 66: Relative frequency of HF squeal noise occurrence among the range of relative rail humidity – data from the whole dataset

hour	**month**											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
5						+		+				
6	+					+	+					
7	+			+	+	+	+	+				
8	+	+		+			+	+	+	+		
9	+	+		+					+	+		
10		+								+	+	+
11		+										+
12											+	+
13												+
14										+		+
15	+									+	+	+
16	+									+	+	+
17										+	+	+
18	+		+							+	+	+
19	+		+							+		+
20	+		+		+				+	+	+	+
21			+		+	+	+		+	+	+	
22	+		+		+		+		+	+		
23					+	+	+		+			
range [%]	36-55	16-24	8-12	13-22	10-25	10-15	5-7	5-8	13-18	10-16	15-23	18-33

Table 43: Daytime influence over the year on HF squeal noise occurrence and additionally probability range in the marked area of the day (last row) – watch out, depicted time zone is UTC +1 and daylight saving time between March and October (UTC +2)

Application of statistical algorithms reveals that the best performance in multivariate evaluations is reached by complex tree models – Boosted Tree and Random Forest (see Table 44). Good performance in most subsets is indicated by achieved benchmarks – even regarding C2 (low quantity of HF squeal noise incidents at all), an acceptable result is achieved. Altogether, due to application of five times repeated 10-fold cross validation, predictive ability in the same boundary conditions seems promising. Benchmark differences between the whole dataset and subsets are low – leading to the conclusion that statistical algorithms are capable of dealing with different sections and train types. Computational cost of Random Forest exceeds the Boosted Tree approach significantly. Moreover, included predictor quantity, is substantially lower in most sets, which makes Boosted Tree a reasonable choice for the task of feature selection. If computational effort is weighted high, the Bagged Tree algorithm is recommended leading to a minor performance loss.

Predictor importance scores are provided in Figure 67 using a cut point of 25 % relative importance for depiction. In general, few track predictors are listed, which indicates that deviations between sections – especially in terms of curve radius – are statistically represented by other variables (e.g. Pressure, Vmean, AccTrain and VDiff). The only track parameter ranked in midfield states CurveBreathing, which is compromised as already discussed in the first paragraph of the current subsection.

Regarding vehicle predictors, TrainType.1, TrainType.6 and TrainType.7 are directly listed train type dummy variables. The former shows high frequency of occurrence, while the latter two indicate a low probability. RSI represents a promising parameter, which connects curve radius and wheel diameter. The correlation also complies with empirical examination. TrainOrientation.1 and TrainOrientation.2 are listed in midfield (in Random Forest). However, empirical investigation evaluates a controversial trend among subsets between pushed and pulled operation. The ranking is presumably reasoned to account deviations regarding single and double unit operation, which are also observed empirically. Apl is ranked high as it allows the model to distinguish between train types as well as double unit operation and to include the notably high frequency of occurrence regarding freight trains in C4 and C5. However, train type dependencies are strong, thus it has to be connected to different train types and cannot be solely taken into account. Moreover, train length and axle quantity are impossible to know in a prediction task. Other vehicle dependent properties – MaxBogieDist, MaxAoA and FreeCurving – are listed in Random Forest with mediocre to low importance scores, however show no general trend in empirical examination.

All predictors in terms of dynamic influence are ranked high. AccTrain and VDiff importance scores are misleading due to speed loss in C5. Although the inclusion of Vmean is statistically reasonable (direct connection to curve radius in the dataset), velocity limit is also influenced by inclination and thus it has not to be always the case that increasing radii result in a raise of the speed limit. AccSide is directly connectable to dynamic forces while curving and thus flange contact as well as maybe lateral creepage is influenced. Therefore, inclusion seems reasonable.

YearDay is ranked highest regarding the group of general predictors, which indicates on

the one hand over-fitting of the models and on the other hand distinct daily alterations, which are also observed empirically. Nevertheless, it is not meaningful to include YearDay into a predictive model due to variable conditions every year and strong location dependency. The same refers to the predictor Hour. To reach independence of local climatic conditions, it is reasonable to represent trends among the year by environmental parameters. Empirical findings of a distinct relative humidity influence are supported by a top rank in the Boosted Tree model (AT.DPDiff) and high to mediocre ranks for all other relative humidity quantities in the Random Forest model (leading with Hum). Temperature and absolute humidity predictors as well as SunRad are also ranked high to mediocre. The latter depicts only influence in daylight hours and shows also a visible negative correlation to the outcome in empirical investigation. However, the found trends are thought to be covered by relative humidity consideration. Interestingly, HoursSLT is ranked high. It is also partly visible empirically that frequency of occurrence in conjunction with trains after the night break is very low. Reasons may be alterations in the third body layer over the night (e.g dew formation), which may impact friction coefficient. However, it is impossible to know time spans between operating trains in a prediction task. Thus, an inclusion in predictive modelling is not possible. Air pressure quantities seem to perfectly account section dependent deviations in the present dataset. However, it would be misleading to include that trend in a general predictive model due to dependency on altitude. Importance of relative pressure quantities is high, however it lacks of a physical explanation and only a weak correlation is shown in empirical investigation. Thus, it is assumed that it represents a misleading trend or at least can be sufficiently expressed by other environmental parameters.

To summarise, direct influence of track predictors is only observed regarding curve radius – increasing radii lead to declining frequency of occurrence especially between 310 m and 440 m. Implicitly, that trend can be well represented by free lateral acceleration as well as RSI in combination with dynamic and vehicle dependencies, respectively.

Impact of vehicle properties on frequency of occurrence is strong – allowing a categorisation in three groups: high (train types A and B), mediocre (train types C, D, E and F) and low probability (train type G and other). Double unit operation can increase frequency of occurrence by up to 12 % – with the simple reason of more axles in the train pass by. Deviations between pushed and pulled operation are controversial among the dataset – leading to an increase of 2 % (train type E in C1) and 9 % (train type F in C1) as well as a declination of 2 % regarding train type F in C5. Thus, a direct influence might not be present. In case of double unit operation, train configuration with one power coach pulling and the other one pushing at the end leads to 3 % lower frequency of occurrence compared to the other three possible combinations. In terms of other vehicle predictors, RSI shows a distinct positive correlation to the outcome. Other vehicle parameters seem to have no general connection to the response.

Regarding dynamic influence, only lateral acceleration is evaluated to have a direct link to the outcome as it influences flange contact probability and may alter lateral creepage as well.

subset	Boosted Tree		Random Forest	
	ROC	number of predictors	ROC	number of predictors
All	0.83	157	0.84	188
MC_1	0.78	53	0.80	64
MC_2	0.80	37	0.78	56
MC_4	0.84	47	0.83	171
MC_5	0.85	46	0.86	171
TT_A	0.72	95	0.72	169
TT_E	0.81	46	0.81	173
TT_F	0.80	89	0.82	174
TT_G	0.85	118	0.85	168
MC_1_TT_E	0.79	43	0.80	54
MC_1_TT_F	0.78	47	0.79	51
MC_4_TT_G	0.84	35	0.83	160
MC_5_TT_F	0.84	43	0.86	162
MC_5_TT_G	0.85	43	0.86	160

Table 44: Results of best performing regression models achieved with Boosted Tree and Random Forest

Distinct variations in frequency of occurrence among the year and in daytime are observed. In terms of relative rail humidity critical values lie between 59 % and 83 % exceeding a frequency of occurrence of 30 % and peaking to 34 % at 74 %. In the remaining range, steep decreases in both directions are present. That trend is observed in the whole dataset as well as in most subsets and thus seems to be section and train type independent. Relative frequency of occurrence drops beneath 20 % in rather dry conditions ($\leq$ 54 % relative rail humidity) and in humid conditions ($\geq$ 100 % relative rail humidity). Months as well as daytime hours (changing among the year), where relative humidity lies in the critical range frequently, show significantly higher probability values (extreme in January). In terms of daytime variations, in winter, frequency of occurrence is usually higher in daylight hours, whereas in other seasons, morning and evening hours (depending on dawn and sunset times) show raised incidences. In summer (especially July and August), overall frequency of occurrence is very low (in peak hours 5 % to 8 %). Daily variations of probability can be perfectly connected to rainy or frosty days (leading to a large declination of incidences). Considering the whole dataset, rain during the train pass by reduces probability by 16 % (to a total of 1.1 %), hoarfrost causes a decrease of 14 % (remaining 2.1 %), frost presence shows a declination of 10 % (total 6.3 %) and dew on the rail indicates 5 % reduction (to 10.3 %). Thus, distinct influence is especially present in terms of rain and hoarfrost, while dew seems to have a lower impact. In general, it has to be pointed out that distribution of environmental conditions can alter significantly in different locations. Thus, an inclusion of predictors representing numeric quarter/month/week/day of the year is not useful for a predictive model, which should be applicable to other datasets involving other sections and climatic conditions.

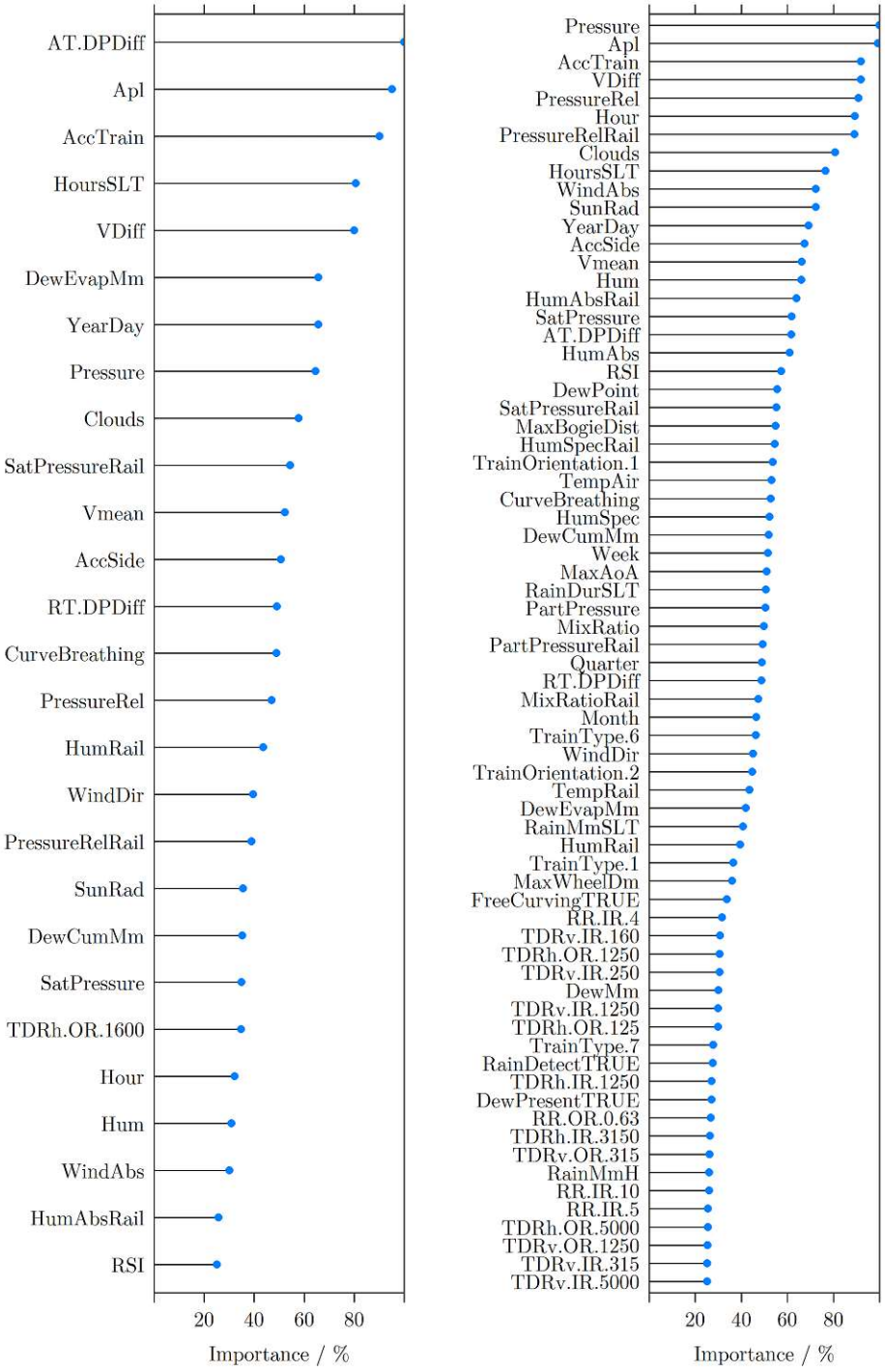

Figure 67: Relative predictor importance in the Boosted Tree (left) and Random Forest (right) model fitted to the whole dataset

Building on the results of empirical and statistical investigation and pursuing the scope of constructing a practical model, the following predictors are chosen for predictive modelling:

- Train type distinguished in three categories depending on their basic tendency to cause HF squeal noise – high (type A, B), mediocre (type C, D, E, F), low (type G, other)

parameter	name in model	type	value range	unit
train type dummy variable (1 mediocre, 2 low, rest zero)	TrainType.(1-2)	categorical	[0;1]	-
double unit operation	DoubleUnit	categorical	[0;1]	-
RSI	RSI	continuous	[0,∞[	-
lateral acceleration	AccSide	continuous	]-∞,∞[	ms^{-2}
relative rail humidity	HumRail	continuous	[0,∞[	%
rain occurrence	RainDetect	categorical	[0;1]	-
dew presence	DewPresent	categorical	[0;1]	-
frost conditions	FrostPresent	categorical	[0;1]	-

Table 45: Mathematical description of the final parameter set applied to predict TonalHigh

- double unit operation
- RSI (represents curve radius and wheel diameter)
- lateral acceleration (leads to inclusion of radius and cant as well)
- relative rail humidity (evaluated as best performing predictor in Subsection 6.4)
- rain occurrence
- dew presence
- frost conditions (frost or hoarfrost likely)

Thus, predictor quantity is shrunk to nine. Table 45 provides a mathematical description of each predictor.

6.3.5 Equivalent continuous noise level

In general, it has to be pointed out that statistical analyses regarding equivalent continuous noise level (termed as Level) is carried out without any distinction between train pass bys with and without detected curve squeal of any type. Thus, predictive accuracy as well as observed trends may be compromised by curve squeal influence. Empirically, evaluation is separated in train pass bys with and without any type of curve squeal. However, the discussion is exclusively focused on the latter category, since curve squeal behaviour is separately analysed in the previous subsections. Interquartile ranges among passenger trains (train types C to G) are in a range of around ± 2 dB in case of no detected curve squeal occurrence (refer to the upper graph of Figure 41). Thus, they can be summarised as quite homogeneous in their emission among all monitored train pass bys. In contrast, freight trains show roughly doubled interquartile ranges, which points out their heterogeneity – reasoned by their alternating configuration and various different car types with completely deviating acoustic characteristics. However, the same interquartile ranges are shown regarding train

type B, which is separated due to operation with homogeneous coaches. Thus, distinction of freight trains in two categories seems not necessary for the present evaluation. To avoid influence from train type distribution and section dependencies overall and in subsets "TT" and "MC", the empirical evaluation is focused on "MC_TT" splits.

Predictor importance scores of univariate statistical analysis (depicted in Figure 68) reveal domination of track influence – predictors, which are able to take the large differences between C1 and C4/C5 into account (e.g. TDRh.IR.(400-800), RailType.3, SleeperType.2), are top ranked. Additionally, two vehicle predictors are listed in the upper half – TrainType.7 (train type G) and MaxWheelDm. The former considers the generally lowest levels of train type G among all train categories and the latter allows general distinction between train types. One dynamic predictor – AccTrain – is ranked with mediocre importance. Detailed investigation reveals that the observable trend is compromised due to speed loss of all trains in C5 (isolation joint). Thus, a general impact on the outcome is not present. General and environmental as well as other dynamic predictors are listed in the lower half with mostly poor importance scores.

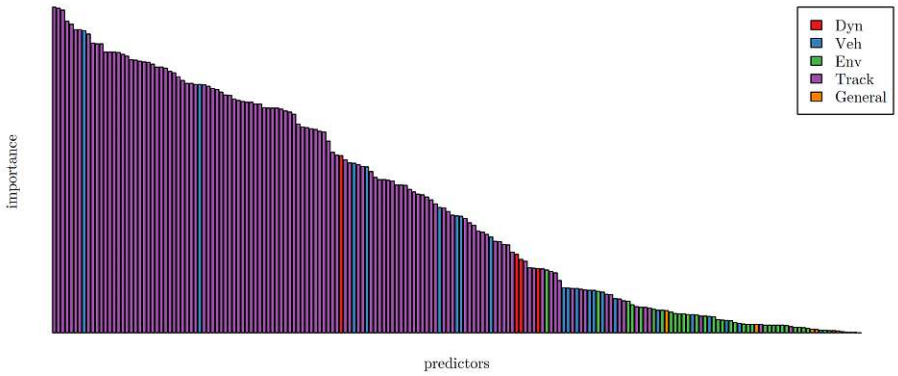

Figure 68: Predictor importance regarding equivalent continuous noise level evaluated with linear variance analysis and coloured in groups dynamic (red), vehicle (blue), environmental (green), track (purple) and general (orange)

Empirical examination regarding track influence complies with statistical findings. Strong correlations are observable if predictors are deviating between C1 and C4/C5. Directly connected issues are discussed in Subsection 5.7. To conclude, especially sleeper type, lateral TDR of the inner rail between 400 Hz and 1250 Hz and RR of the inner rail between 10 cm and 16 cm show significant differences. Apart from TDR and RR predictors as well as other mentioned quantities, track quantities show no direct connection to the outcome. In general, it seems that values are foremost dependent on RR and TDR (only considering track parameters), which agrees with behaviour on tangent track, and not influenced by curve radius alterations at all.

Distinct differences among train types are present. Overall, train type G shows about 7 dB lower median levels compared to other passenger train categories in the corresponding sections. Train type other indicates highly varying levels, however, median values are lower than other passenger train types (except train type G) in narrow radii ($\leq$ 256 m). However, the random category indicates higher levels in C2 and C3. In C1, track contribution is fairly high leading to almost equal median levels among train types A, C, E and F. In all other sections, freight trains show 4 dB to 8 dB higher median levels compared to the highest ones among passenger train types (decreasing with growing radius). It has to be borne in mind that measurement campaigns are carried out between 2013 and 2017, where most freight coaches were braked with cast-iron blocks. Nowadays, legislative of the EU forbids operation with cast-iron block brakes on certain routes, which have to be defined by each member state, by the end of 2024 [206]. In Germany, operation is forbidden on all normal gauge tracks by the end of 2020 (few exceptions are also stated) [207]. Hence, deviation between freight and passenger traffic levels will decrease significantly in the future. However, an integration of that trend in the model building process is not possible. Impact of train orientation as well as single and double unit operation is negligible (median differences in a range of 1 dB). Other vehicle predictors show no correlations to the outcome.

Influence of Vmean and hence also AccSide cannot be investigated, since levels are corrected for empirical examination to prevent a velocity dependent influence (see Subsection 6.2.1). As stated in derivation of the correction formula, median levels are increasing with raising mean speed, however not as much as found in the literature. VDiff and AccTrain indicate no impact, which justifies the selected thresholds for exclusion in the evaluation (refer to Subsection 3.3).

Influences among the year are partly observable. Presented trends are mainly extracted from train type G in C5 due to by far most quantity of samples and homogeneous sound emission. Among days of the year up to 8 dB median difference is evaluated, however, neighboured days show only up to 2 dB deviation. Thus, the tendency seems not directly dependent on daily alterations and is better represented by monthly distribution among the year (provided in Table 46). While train type F shows a homogeneous monthly distribution over the year, train type E in C1 and train type G in C5 indicate raised levels from October to December. Especially for the former category, October shows very high values compared to other months. Daytime alterations among months of the year are summarised in Table 47 for train type G in C5. Missing values mean that there are not sufficient data points. Statistical significance is reached with usually monitored train pass bys in a range of 30 to 60 per month and hour except May, where train counts drop to 15 to 30 per month and hour. In hour 0 and 4, where operation ends and begins, train counts are lower with 10 to 20 train pass bys. Hours are defined by their beginning time (e.g. hour 5 lasts from 5 a.m. to 5:59 a.m.). As can be seen, between May and September slightly lower levels are observed in the afternoon, while higher values are primarily present in the morning. However, overall daytime fluctuations among months are minor ($\pm$ 1 dB). It has to be pointed out that presented tendencies among the year depend presumably on climatic conditions. Thus, monthly and daytime behaviour might deviate significantly from the illustration in other

subset	month											
	1	2	3	4	5	6	7	8	9	10	11	12
MC_1_TT_E	–	89	90	90	91	91	91	–	–	95	93	92
MC_1_TT_F	–	90	90	91	91	91	92	–	–	91	91	90
MC_5_TT_F	84	84	83	84	84	83	83	84	83	84	83	83
MC_5_TT_G	78	76	76	76	76	76	77	77	77	79	79	79

Table 46: Monthly median values of Level over the year only considering trains without any type of curve squeal and with colour theme for statistical significance – red (weak), orange (mediocre), green (strong)

locations. Visible negative correlations are observed regarding temperature and difference between temperature and dew point (relative humidity indicator). An increasing trend of 1 dB to 2 dB in median level in case of dew presence is supported by all considered MC_TT subsets. Statistical significance is too weak to give a detailed statement on rain impact. Frost presence (regardless whether frost or hoarfrost is likely) leads to raised median levels of 1 dB to 2 dB. To investigate environmental impact further, spectral analyses are carried out and discussed in the following paragraph.

Spectral analyses are only focused on train pass bys without any detected curve squeal. Furthermore, in addition to generally formulated exclusion criteria (see Subsection 3.3), train pass bys are filtered to consider only data points in an area of 5 kmh^{-1} around the speed limit (60 kmh^{-1}) in C1 and C5. Hence, no correction of levels due to train speed impact is needed. The evaluation is done by comparing median values. Regarding temperature impact estimation, data points with dew and/or rain occurrence are excluded to prevent compromising impact. The focus stays on rail temperature as trends are more distinct compared to air temperature. Considered subsets are foremost train type G in C5 and additionally train type F in that section as well as train types E and F in C1 due to high statistical significance. Thus, the evaluation is only carried out in conjunction with passenger train categories. Investigation on freight traffic cannot be done due to lack of sufficient data points. For estimation of impact regarding dew presence, data points with rain or frost occurrence are excluded. Other boundary conditions (like speed filtering) are equal to the evaluation of rail temperature impact. Stated frequencies refer always to the whole third-octave band with that mid frequency.

The examination reveals that frost occurrence causes raised levels at 160 Hz and 500 Hz in a magnitude of 1.5 dB to 2 dB compared to the same rail temperature range ($\leq 0°C$) without frost occurrence (refer to blue lines in Figure 69). Both third-octave bands show the highest levels and thus influence the overall level significantly. The trend is also visible regarding train type F in the same section with raises at 125 Hz and 160 Hz of 1.5 dB to 3 dB and at 400 Hz of 1 dB. In C1, statistical significance regarding train pass bys with frost occurrence is low – especially in conjunction with train type F. Thus, evaluation of the latter category is skipped. Nevertheless, train type E indicates the same tendency with increases at 160 Hz and 200 Hz of 2 dB to 3 dB and at 400 Hz and 500 Hz of 1 dB to 2 dB. Thus, impact of frost

hour	month											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
0	–	0	0	0	–	–	0	0	+1	+1	0	+1
4	–	0	0	0	–	+1	+1	+1	+1	0	0	0
5	–	0	0	0	0	+1	+1	+1	+1	0	0	0
6	–	0	0	0	0	+1	+1	+1	+1	0	0	0
7	–	0	0	0	0	+1	+1	+1	+1	0	0	0
8	–	0	0	0	+1	0	0	+1	+1	0	0	0
9	–	0	0	+1	0	0	0	0	0	0	0	0
10	–	0	0	0	0	0	0	0	0	0	0	0
11	–	0	0	0	0	0	0	0	0	0	0	-1
12	–	0	0	0	0	0	0	0	0	0	0	0
13	–	0	0	0	0	0	0	0	-1	0	0	0
14	–	0	0	0	0	-1	0	0	0	0	0	0
15	–	0	0	0	0	-1	0	-1	0	0	0	+1
16	–	0	0	0	-1	0	0	0	0	0	0	+1
17	–	0	0	0	0	0	0	-1	-1	0	0	0
18	–	0	0	0	0	0	0	0	0	0	0	0
19	–	0	0	0	0	0	0	-1	-1	0	0	0
20	–	0	0	0	0	0	0	0	0	0	0	0
21	–	0	0	0	0	0	0	0	0	0	0	0
22	–	0	0	0	0	0	0	0	-1	0	0	0
23	–	0	0	0	+1	+1	0	0	0	0	0	0

Table 47: Daytime influence over the year on Level (without any type of curve squeal) with corresponding positive or negative median deviations among daytime in units dB compared to monthly median in Table 46 – watch out, depicted time zone is UTC +1 and daylight saving time between March and October (UTC +2)

occurrence lies around 160 Hz and 400 Hz to 500 Hz and is detected on three train types and among two sections with highly deviating properties.

Moreover, rail temperatures ≤ 0 °C without frost occurrence cause an increase of 1.5 dB to 2 dB at 125 Hz, 160 Hz and from 500 Hz to 1000 Hz compared to rail temperatures > 0 °C (illustrated in Figure 69). Thus, adding up both effects leads to up to 4 dB higher levels in certain third-octave bands compared to rail temperatures above the freezing point regarding train type G in C5. However, the mentioned impact of rail temperatures ≤ 0 °C without frost occurrence can only be partly validated by other subsets. Train type F in C5 shows only a slight increase (< 1 dB) at 160 Hz, the same type in C1 indicates a 2 dB raised level at 400 Hz and train type E in C1 shows slight increases (< 1 dB) at 125 Hz and 500 Hz. Thus, the trend in other subsets is not as distinct. Moreover, raises between 630 Hz and 1000 Hz are not indicated by other categories than train type G in C5. However, it has to be pointed out that statistical significance in the latter category exceeds all other MC_TT splits by far and is therefore most trustworthy. It is possible that train types, which show a higher basic noise emission (train types E and F compared to train type G) are not as sensible to changes in environmental conditions.

In general, application of A-weighting causes a significant reduction of the observed peak around 160 Hz, which is then about 10 dB lower than the peak at 400 Hz or 500 Hz. Thus, increases in the latter third-octave bands state the main issue in terms of impact on noise immission.

A third effect is observable (see again Figure 69) between 2000 Hz and 5000 Hz. While in the low and mid frequency area, median values among rail temperatures above the freezing point are roughly equal, deviations in the mentioned frequency range reach up to 4 dB between mid values of 45 °C and ≤ 0 °C with significantly lower levels in conjunction with high rail temperatures and a distinct separation between categories 25 °C and 35 °C. However, that particular trend is not indicated by other subsets. In general, third-octave band levels show a partly high difference among distinguished rail temperature categories – mostly a roughly linear positive or negative correlation. Considering only distribution till 5000 Hz (upper area discussed below), train type F in C5 shows deviations of up to 3 dB in third-octave band levels between 250 Hz and 800 Hz, the same train type in C1 shows up to 4 dB deviation between 200 Hz and 800 Hz and train type E in C1 indicates scattering at 160 Hz and between 630 Hz and 1250 Hz of up to 3 dB. Hence, trends are not equal among subsets and no consistent statements can be evaluated. To conclude, rail temperature magnitude above the freezing point can also have significant impact on certain third-octave band levels, which are among the highest in terms of rolling noise and thus can impact overall noise emission as well.

Considering the highest depicted frequency bands (6300 Hz to 10000 Hz), controversial trends are indicated regarding train type G in C5. However, train type F in C1 and C5 as well as train type E in C1 indicate a clear tendency of significantly decreased levels in conjunction with low rail temperatures with deviation magnitudes of up to 4 dB (6300 Hz), 7 dB (8000 Hz) and 10 dB (10000 Hz) between rail temperature categories 45 °C and ≤ 0 °C. Those frequencies may be influenced by atmospheric absorption even in a close monitoring range. A closer look into attenuation coefficients (calculated with formulae in ISO 9613-

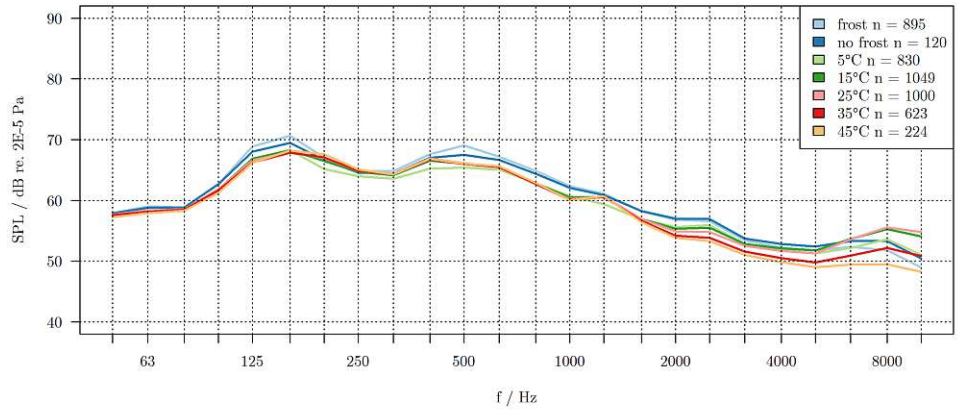

Figure 69: Median values of third-octave band levels from subset MC_5_TT_G with excluded rain and dew occurrence – blue lines for rail temperatures $\leq$ 0°C with distinction in frost occurrence (light blue) and no frost occurrence (blue); other values in the legend depict centre rail temperatures of the categories (e.g. category 5 °C ranges from 0 °C $<$ t $\leq$ 10 °C)

1 [112]) reveals that values in winter are more than doubled compared to summer (median values). Despite that trend complies with findings, the magnitude for 8000 Hz lies between 0.06 dBm^{-1} (minimum among the whole dataset) and 0.23 dBm^{-1} (maximum among the whole dataset) and thus in a distance of 7.5 m only up to 1.3 dB deviation is explainable. In terms of 10000 Hz up to 1.5 dB deviation occurs. Hence, other additional influences have to be responsible. However, emissions in that frequency range without any flanging or HF squeal noise occurrence are significantly lower than in mid frequencies and thus not much contributing to overall radiated sound. Moreover, due to high atmospheric absorption, those frequencies loose even more importance with growing distance to the immission point.

In addition to general influence of rail temperature as well as frost, dew occurrence leads to an increase of 1.5 dB to 2 dB at 125 Hz and 160 Hz (illustrated in Figure 70). Moreover, a decrease of 3 dB is visible regarding the high frequency bands 8000 Hz and 10000 Hz. Both trends comply with findings in all other considered MC_TT splits. Moreover, a slight raise in nearly all considered third-octave bands is observable – above 1 dB between 500 Hz and 1000 Hz. Although, the latter frequency bands are also increased in all other subsets, occurring magnitudes are lower ($<$ 1 dB). Again, it might be connected to the low noise emission of train type G and the resulting sensitivity to external influences. Altogether, impact of dew occurrence is also evaluated on three train types and among two sections with highly deviating properties.

Rain might have an influence, too, however statistical significance is weak in every subset. Thus, a generalised statement is not possible. Relative as well as absolute humidity show partly an influence, however it is not as distinct as observed in conjunction with rail temperature. Nevertheless, consideration in a multivariate approach might be beneficial, since impact of environmental conditions may be refined.

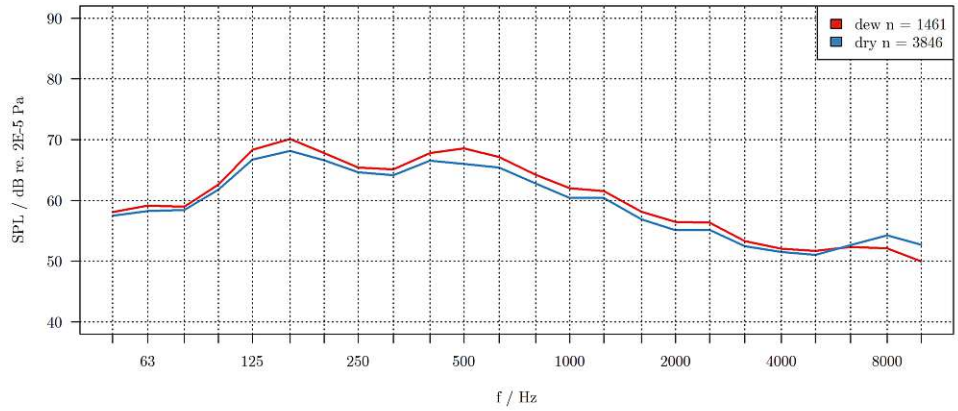

Figure 70: Median values of third-octave levels from subset MC_5_TT_G with excluded rain and frost occurrence – distinguished in dew occurrence (red) and dry conditions (blue)

To conclude, environmental conditions can have a significant influence on third-octave band levels – especially also on those, which are dominating in overall noise level and are not heavily impacted by A-weighting (400 Hz to 1000 Hz). Highest levels are reached on the one hand if frost occurrence (regardless whether frost or hoarfrost) is likely as well as partly additionally if rail temperature $\leq 0°C$ and on the other hand if dew is present on the rail, The latter complies with the observation of higher levels in the morning hours. Spectral distribution indicates raises in the area of 160 Hz and in mid frequencies (400 Hz/500 Hz). Generally, in the former range, track and vehicle contribution to rolling noise is roughly equal, while in the latter track emission is dominant [92]. Thus, a connection to superstructure emission is more plausible. Trends in both mentioned frequency areas are observed among three different train types and in two sections with largely deviating properties.

Assumptions for the behaviour in conjunction with frost occurrence are on the one hand, if ballast stones are frozen together, a bonding effect appears, which causes a loss in vibration absorption and hence results in elevated sound emission of the superstructure. On the other hand, in addition to the assumed mechanical impact, bonding effects between ballast stones may reduce porosity of the surface and change absorption among the transmission path. Moreover, if frozen earth is present in the transmission path between ballast bed and microphone, the ground changes in the worst case from partly absorbing to fully reflecting, which increases the monitored sound pressure in 7.5 m distance from track axis. Both ground effects can be summarised by a change in acoustic impedance of the surface along the transmission path. Another influence might be alterations in the third body layer on the rail head like changes in viscosity or different roughness pattern due to frost adherence. For the impact of dew occurrence, the only hypothesis is an alteration of the third body layer due to added humidity. All assumptions are based on general train type and section independence, which sorts out impact of different sleeper material, rail type, rail roughness, track decay rate, wheel diameter and eigenfrequencies of special vehicle components. General thoughts and questions,

which are partly countered by certain properties are:

- Examination if observed effects are also present on tangent track – cannot be investigated, since no data points are available. For a complete comparability, data from the same trains on tangent track would be needed, which cannot be provided.
- Colder temperatures have an impact on rail pad stiffness (increasing with decreasing temperature), which improves track decay rate and therefore indicates the opposite trend [9]. Thus, that influence is not present or masked by stronger tendencies in the other direction. However, it may be an explanation for mentioned trends of rail temperature scattering above the freezing point between 160 Hz and 1250 Hz if a positive correlation is indicated.
- Rail pad material might alter properties, if water is absorbed.
- Impact of longitudinal slip due to traction in conjunction with a lower friction coefficient in humid conditions – can be completely excluded since the isolation joint in C5 causes that every train looses speed and no traction is available in the section.
- Influence regarding general vehicle dynamics while curving – friction coefficient has an impact on lateral and longitudinal slip as well as spin.
- Longitudinal stick-slip effect – A frequency range between 50 Hz and 100 Hz is stated in past research [5], thus not according to observations
- Influence of the measurement setup (e.g. due to dew occurrence on the microphone) – Although the identical microphone is used in C1 and C5, in the latter section another microphone is applied for validation and control. Thus, measurement errors are not thought to account for the observed deviation.

Altogether, observed effects are not entirely explainable, which causes a need of further research (e.g by laboratory experiments in controlled climatic conditions). Moreover, data points on tangent track can also be investigated focusing on stated abnormalities. Due to the observed trends, measurements especially in conjunction with frost and dew conditions should be interpreted with caution. Deviations among temperature, humidity and relative humidity can have significant impact on certain third-octave band levels – especially in high frequencies (above 6300 Hz). However, also third-octave bands, which dominate overall sound emission and are not much weakened by A-weighting (400 Hz to 1000 Hz), are affected and can fluctuate significantly depending on environmental conditions. Thus, that has to be borne in mind especially in conjunction with short-time measurement campaigns and their comparability to each other.

Application of statistical models shows that tree and rule-based models outperform other approaches. Especially Cubist and Boosted Tree models achieve the first two ranks in the whole dataset as well as in all considered subsets (reached benchmarks depicted in Table 48). In most subsets as well as in the whole dataset, the Cubist model shows the best performance.

Moreover, the predictor importance ranking does not change much over subsets – excluding only zero variance predictors, respectively. In general, achieved benchmarks are very good with a RMSE value of 1.9 dB and 90 % correlation between predicted and original outcome over the whole dataset. It also indicates that the presented model approaches are capable of taking different section and train type properties into account. Due to application of five times repeated 10-fold cross validation, the listed benchmarks give also an approximation of predictive performance on a dataset with the same boundary conditions.

Figure 71 illustrates predictor ranking of the Cubist model among the whole dataset ordered regarding their relative importance score and depicted until 25 % is reached. Regarding track predictors, two RR predictors and one TDR predictor is listed (RR.OR.25, RR.OR.20 and TDRh.IR.500). The latter shows large differences between C1 and C4/C5. Rail roughness of the outer rail in listed wavelength bands is deviating between C1 and C5, however both satisfy ISO 3095 [10] limit. Furthermore, in conjunction with velocities of around 60 kmh^{-1}, frequencies of under 100 Hz are affected. Thus, impact of those might be overestimated by the model. Moreover, CurveBreathing is listed, however it is connected to rail temperature and also influenced in magnitude depending on the section. Hence, ranking of the latter is compromised and not suitable for inclusion in a predictive model. Among vehicle predictors, MaxWheelDm and Apl are top ranked. They are chosen to distinguish between train types (especially between freight traffic and train type G). The latter involves axle count and train length, which is impossible to know in a predictive task and thus not considered further. RSI is also listed with relatively low importance. All predictors for dynamic influence are listed. However, due to compromising influence of C5 on AccTrain and VDiff, both are not suitable for inclusion in a prediction model. AccSide as well as Vmean can have a reasonable impact since curving behaviour is affected by the former, while the latter shows a generally positive correlation to the outcome. The high ranking of YearDay indicates a tendency to over-fitting. Other listed seasonal predictors are Week and Month. Since the scope is a generally applicable prediction model, seasonal predictors are not considered due to dependency on local climatic conditions. Environmental predictors are ranked high and in large quantity, which justifies their impact further. Due to between predictor correlations, the set can be reduced to: HumAbsRail, SatPressureRail and RT.DPDiff (chosen on account of the highest ranking among correlated predictors). Due to lack of physical explanation, relative air pressure is not further considered and it is believed that its influence can be represented by other environmental predictors. Absolute air pressure depends on altitude and is thus not considered for predictive modelling.

To summarise, considering the whole dataset, influence of track properties (foremost TDR and RR) is dominating. Moreover, the sleeper material has also direct influence. Regarding vehicle properties, strong train type dependencies are observable. Train type G shows a very low basic rolling noise level and freight traffic (train types A and B) indicate high levels. The remaining categories show mediocre values. Distinguishing between train types can be carried out by inclusion of Apl and MaxWheelDm, which both show strong correlations in the dataset. The former is not suitable for predictive modelling, since axle count and train

subset	Boosted Tree			Cubist		
	RMSE	R^2	number of predictors	RMSE	R^2	number of predictors
All	1.98	0.90	166	1.94	0.90	96
MC_1	1.24	0.72	55	1.22	0.73	52
MC_2	2.46	0.63	41	2.45	0.64	43
MC_4	2.27	0.72	51	2.24	0.72	51
MC_5	2.23	0.67	51	2.14	0.70	63
TT_A	2.75	0.33	88	2.81	0.31	65
TT_E	1.26	0.72	47	1.26	0.73	45
TT_F	1.29	0.88	122	1.24	0.89	87
TT_G	2.25	0.57	143	2.23	0.57	100
MC_1_TT_E	1.21	0.69	42	1.21	0.69	40
MC_1_TT_F	0.98	0.53	47	0.98	0.52	45
MC_4_TT_G	2.28	0.48	39	2.26	0.49	45
MC_5_TT_F	1.73	0.46	44	1.58	0.54	54
MC_5_TT_G	2.28	0.55	43	2.22	0.57	54

Table 48: Results of best performing regression models achieved with Boosted Tree and Cubist

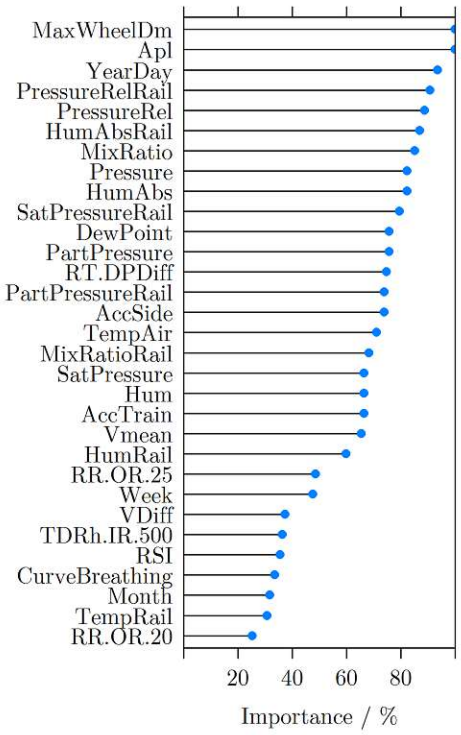

Figure 71: Relative predictor importance score of a Cubist model applied on the whole dataset

length are impossible to forecast in real operation. Despite the tendency of MaxWheelDm in the present dataset, an applicability on other datasets is questionable, since the correlation is also influenced by train type distribution among sections and different section properties. Thus, train type influence is solely included by categorising similar train types rather than including MaxWheelDm. Dynamic influence is visible in terms of Vmean and AccSide – both showing a positive correlation. Additionally, a seasonal impact is observed. Train pass bys in autumn and winter show increased levels (median deviation in a magnitude of up to 3 dB compared to spring and summer). Three trends are observed by detailed spectral analysis: (i) Dew occurrence leads to increased levels; (ii) Rail temperatures $\leq$ 0°C lead partly to raised levels compared to rail temperatures above the freezing point; (iii) Frost occurrence additionally raises levels. Assumptions for the reason behind (iii) are on the one hand a frozen superstructure, which shows worse damping behaviour and on the other hand possibly reduced porosity in ballast bed as well as frozen ground in the further transmission path, which causes reduced absorption of the surface. Other thoughts are alterations in the third body layer on the rail head in terms of viscosity and/or water absorption, which would fit foremost (i) and (iii). While (i) and (iii) are observed among three different train types in two deviating sections, (ii) only occurs mainly in conjunction with train type G in C5 with weak trends among other subsets. However, the latter category states by far the most statistical significance. It is assumed that due to the low basic rolling noise level, more sensibility on environmental conditions is present. It has to be pointed out that spectral analyses are only carried out on passenger train categories due to lack of data points regarding freight traffic. Since reasons are mainly unclear and stated assumptions lack of scientific proof, further research is recommended. Furthermore, influence of relative humidity and absolute humidity is observed, however not as distinct as rail temperature in empirical investigation. Nevertheless, consideration in a multivariate approach might be beneficial, since impact of environmental conditions may be refined. Especially inclusion of specific humidity (or mixing ratio) seems beneficial due to consideration of air pressure and thus comparability between different altitudes is improved. Statistical significance is too weak to give a detailed statement on rain impact. However, it is combined with dew occurrence as both state wetness sources.

Building on the results of empirical and statistical investigation and pursuing the scope of constructing a practical model, the following predictors are chosen for predictive modelling:

- TDR and RR predictors are combined to four principal components catching 95 % variance by using PCA on the whole dataset.
- sleeper type
- train type distinguished in three categories depending on their basic rolling noise level – high (type A and B), mediocre (type C, D, E, F, other), low (type G)
- mean speed
- lateral acceleration (leads to inclusion of radius and cant as well)

parameter	name in model	type	value range	unit
principal components of TDR and RR predictors	PC.(1-4)	continuous	]-∞,∞[	-
sleeper material (1 wood, 0 concrete)	SleeperTypeWood	categorical	[0;1]	-
train type dummy variables (1 mediocre, 2 low, rest zero)	TrainType.(1-2)	categorical	[0;1]	-
mean speed	Vmean	continuous	[0,∞[	kmh^{-1}
lateral acceleration	AccSide	continuous	]-∞,∞[	ms^{-2}
rail temperature	TempRail	continuous	]-∞,∞[	°C
specific humidity	HumSpec	continuous	[0,∞[	gkg^{-1}
relative rail humidity	HumRail	continuous	[0,∞[	%
wet conditions	Wet	categorical	[0;1]	-
frost conditions	FrostPresent	categorical	[0;1]	-

Table 49: Mathematical description of the final parameter set applied to predict Level

- rail temperature (evaluated as best performing predictor in Subsection 6.4.4)
- specific humidity (includes absolute air pressure in calculation) – evaluated as best performing predictor in Subsection 6.4.4
- relative rail humidity (evaluated as best performing predictor in Subsection 6.4.4)
- wet conditions (rain and/or dew occurrence)
- frost conditions (frost and/or hoarfrost)

Thus, predictor size is shrunk to 14. Mathematical description of those is provided in Table 49.

Since level values can be significantly influenced by curve squeal occurrence, inclusion of those predictors in the predictive model is additionally investigated in Subsection 6.4.4.

6.4 Predictive modelling

Building on the selection of the final parameter set for each considered dependent variable (refer to Subsection 6.3), prediction models are generated. As previously discussed, peak levels and relative occurrence time of curve squeal show high fluctuations and also high RMSE values in the models, which causes that a reliable prediction model cannot be formulated for those variables. They can be roughly estimated by heuristic models or median values (see Appendix C). The remaining dependent variables are three classification tasks – prediction of occurrence of flanging noise, squeal noise and HF squeal noise – and one regression task for predicting equivalent continuous noise level. The chosen methodology is described in Subsection 6.2.2 and is applied on all four modelling tasks separately. Thus, the present subsection

is outlined as discussion on all four dependent variables – beginning with examination of flanging noise and ending with equivalent continuous noise level investigation. Each subsection is sectioned in final model description, cut-off point investigation (only classification) and impact estimation of environmental predictors.

6.4.1 Flanging noise occurrence

Firstly, prediction models are trained with the final predictor set alternating each of the four considered predictors for relative humidity – Hum, AT.DPDiff, HumRail, RT.DPDiff. Models are trained on the whole dataset as well as on subsets C1, C5 and both together with the introduced splitting strategy. The focus lies on models trained on data points of C1 and C5 and their achieved benchmarks regarding application on validation sets C2, C3, C4 and altogether as an indicator for predictive ability on new datasets. In general, both predictors dependent on rail temperature show better performance compared to the other two derived with air temperature. Between the former, RT.DPDiff shows slightly better benchmarks and is therefore chosen as final predictor for relative humidity. It might not be as interpretable as HumRail, however, it behaves better statistically, which is the main scope for the present task. Figure 72 depicts the density function of absolute frequency in terms of flanging noise occurrence compared to train pass bys without any curve squeal occurrence. As can be seen, probability exceeds 50 % between roughly 4 °C and 7 °C difference. Lower values represent a high rail humidity, whereas high values indicate dry air.

Final prediction model

Benchmarks between models show that resampled performance on the training set as well as application on the test set is for most models hugely deviating from application on validation sets. Thus, the tendency especially of complex models like Random Forest or Boosted Tree indicates over-fitting, which worsens the predictive performance on new datasets. FDA significantly outperforms all other algorithms on the validation sets as well as suffers minor losses compared to test set benchmarks. Hence, the algorithm seems to be able to extract fundamental connections between predictors and the outcome. Table 50 depicts the ROC benchmark achieved by the final FDA model. The second best approach – CART – is also added for comparison and indication of magnitude in performance difference. As can be seen, benchmarks between test set and application on all validation sets together are roughly equal, which is fulfilling the modelling scope perfectly. In comparison to achieved benchmarks in feature selection (refer to Table 34), a significant decrease in model performance is observable with the reduced predictor set. However, that is rather caused by over-fitting in the former case and not depicting a benchmark of predictive ability on new datasets as well as not fulfilling the scope of a practical model. Hence, it can be concluded that if measurement data is available, complex tree-based model algorithms can be used to train a model, which is capable of predicting on the same sections with the same train mixture well. However, if a model should be trained on measurement data for general application to other datasets including different sections and train type distribution, a simpler

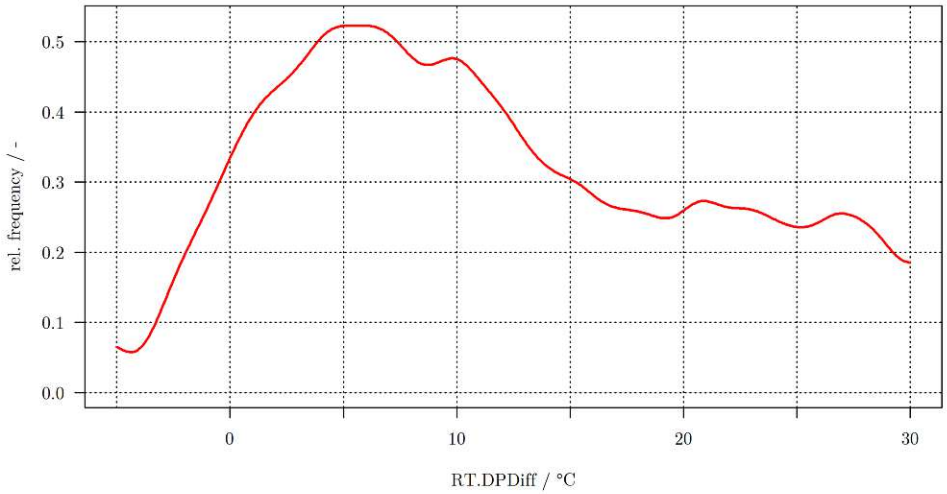

Figure 72: Relative frequency of flanging noise occurrence among the range of difference between rail temperature and dew point – data from the whole dataset

	FDA	**CART**
test set	0.725	0.720
C2	0.709	0.684
C3	0.672	0.636
C4	0.741	0.717
C2,C3,C4	0.722	0.693

Table 50: Achieved area under ROC curve by best performing model approaches FDA and CART trained on data points of C1 and C5

model without over-fitting is needed. For the latter task and in terms of practical applicability – achieved by inclusion of few predictors, worse performance benchmarks are acceptable.

The basic functionality of flexible discriminant analysis is briefly described in Subsection 6.1.3. It states a non-linear classification model, which is built with regression coefficients of the MARS algorithm. The final model consists of one degree (no square coefficients) and twelve additive terms in the discriminant function. Since it is only a two class problem, calculation of one discriminant score is needed for prediction. The function looks as follows:

$$
\begin{aligned}
D \;=\; & -5.0139 \\
& + 2.97275 \cdot TrainType.1 \\
& + 3.06280 \cdot TrainType.2 \\
& + 1.76713 \cdot RainDetect \\
& - 1.04218 \cdot DoubleUnit \\
& + 0.29722 \cdot h(3.95034 - RT.DPDiff) \\
& + 0.12448 \cdot h(RT.DPDiff - 3.95034) \\
& - 0.09363 \cdot h(RT.DPDiff - 17.8742) \\
& + 5.94799 \cdot h(0.86167 - RSI) \\
& + 54.11460 \cdot h(RSI - 0.82830) \\
& - 57.02336 \cdot h(RSI - 0.86167)
\end{aligned} \tag{58}
$$

Notably, the final model does not include lateral acceleration, dew and frost presence for prediction, which is surprising since especially frost presence decreases frequency of occurrence significantly in empirical examination. However, that simplifies the model further and modelling of dew formation can be skipped for flanging noise prediction. For regions with frequent frost or dew occurrence, it may be necessary to alternate the model if measurement data are available and worse model performance is suspected – e.g. by setting the rain predictor to one in case of frost. Categorical predictors are set to one (true) and zero (false). Five terms consist of hinge functions, which are introduced in the description of the MARS algorithm. If the value of the function $h(x)$ is above zero, it is x, else zero. The evaluated score of D for a set of predictors leads to the decision if the event is predicted positive or negative. The cut-off point is dependent on the chosen cut-off probability, which is investigated later. With the default value (50 %), a discriminant score of under -0.9166 causes a positive classification (i.e. flanging noise occurrence) and a score of greater or equal -0.9166 leads to negative classification. Since predictors are not standardized, the absolute difference between coefficients of each term do not represent the impact on the model directly. However, it is possible to analyse the impact of each term on the probability. Negative terms increase the probability, while positive ones cause a decline. Considering train type impact, it can be seen that train types summarised as category 1 (TrainType.1 – mediocre probability) and train types summarised as category 2 (TrainType.2 – low probability) both cause a decrease in probability, while the third category (freight trains), which is represented by trains, which are not category 1 and not category 2, contribute in terms of a significant raise in probability compared to the other train types. Coefficients between train category 1 and 2 indicate that probability of the former is higher. Thus, train type dependency is reasonably represented by the model. Rain occurrence decreases probability, while operation of double units causes an increase. Both mechanisms comply with observation and hence are well included in the model. Difference between rail temperature and dew point is present in

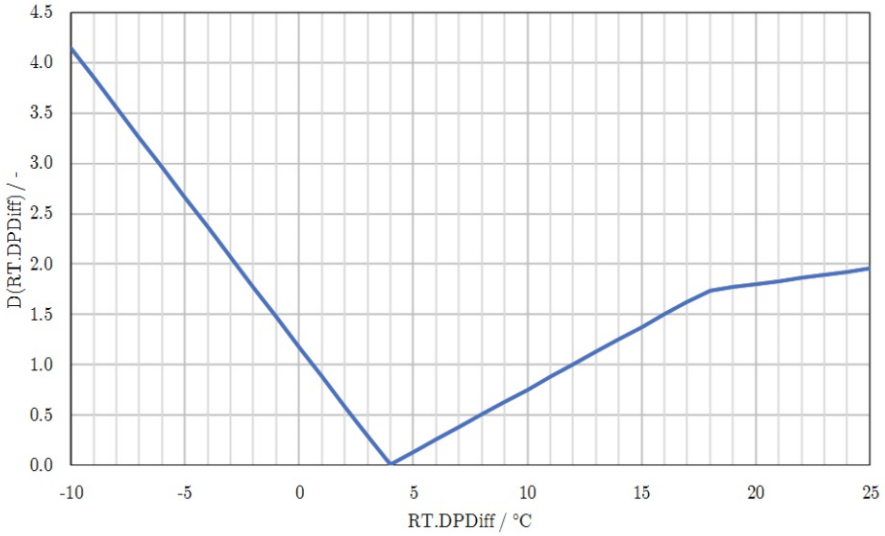

Figure 73: Contribution to the discriminant function of RT.DPDiff depicted between -10 °C and 25 °C

three different terms, which compromises a direct interpretation possibility. Thus, the whole impact is depicted in Figure 73 with a range between -10 °C and 25 °C and the corresponding discriminant score. As can be seen, the highest probability is set to the cut point of 3.95 °C in the model. In general, higher probability lies between roughly 2 °C and 8°C. Lower values steeply decrease probability, while higher ones also raise probability, however not as steep with an additional slope change at 17 °C. Hence, the model catches also the particular ranges in conjunction with difference between temperature and dew point, which are observed in the density function, very well. RSI contribution is also separated in three terms and illustrated in Figure 74 between 0.5 and 1.1. It can be seen that probability increases with growing RSI. In an area between 0.82 and 0.88, a steep probability decrease is modelled, followed again by a raise throughout higher values. In general, an increasing probability with growing RSI is reasonable and complies with examinations in the dataset. Compromising impact is caused by train type distribution. Thus, a slight uncertainty in RSI terms is present, however the impact on general applicability is thought to be low. Empirical correction can be done for example by deleting the second last term in Equation (58) and setting the coefficient of the last term to -10 to avoid a steep curve with higher RSI. That change would lead to the curve depicted in Figure 75, which seems reasonable in a qualitative manner. However, all coefficients and the intercept are connected together and are evaluated with a statistical optimisation criterion during model training. Thus, empirical changes add a great uncertainty to the model. Hence, it is recommended to leave the model formula unchanged and observe wrong classifications due to RSI in the mentioned range carefully. If they are present, model coefficient optimization might be beneficial[21].

[21] The model object from the final model in R is available at https://github.com/M-Ostermann/curve-noise-prediction

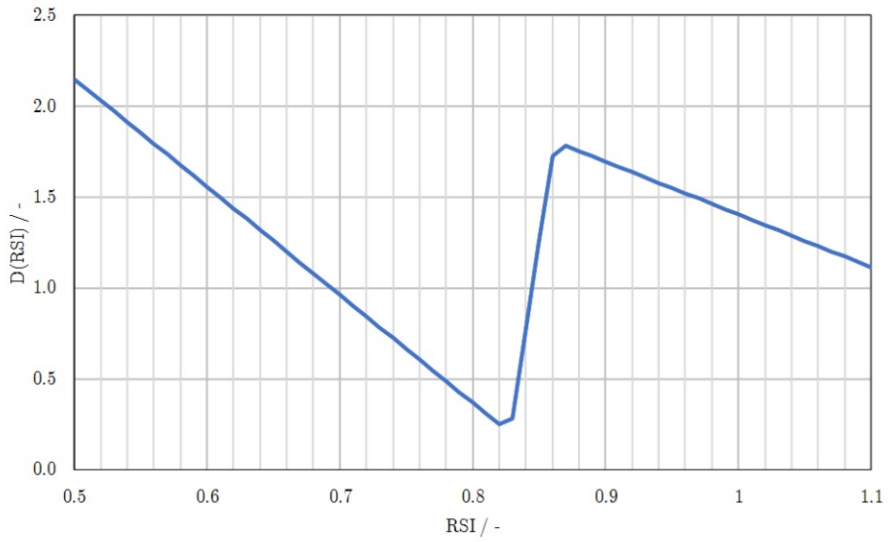

Figure 74: Contribution to the discriminant function of RSI depicted between 0.5 and 1.1

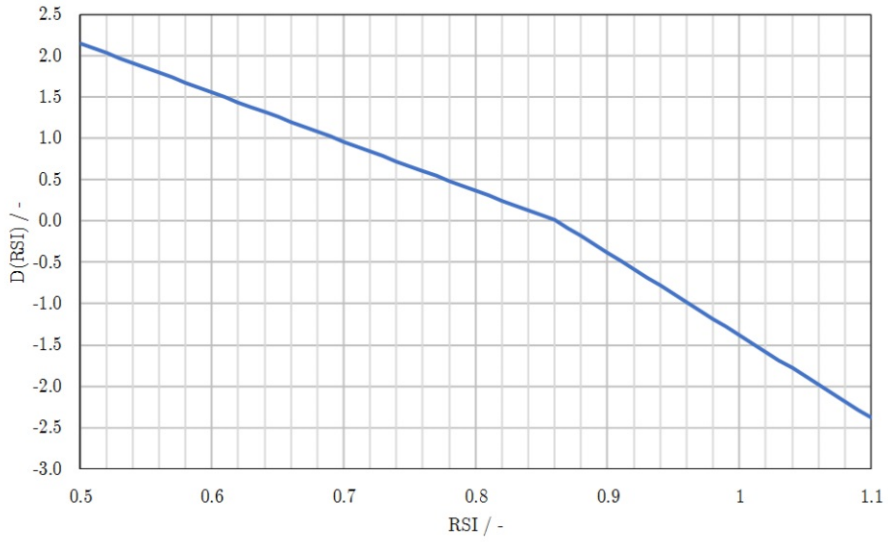

Figure 75: Contribution to the discriminant function of RSI with coefficient changes to avoid the model behaviour in the area around 0.85

Alternative cut-off probability

Alternative cut-off probabilities are investigated on a small dataset (10 %), which is neither used as training nor as test set. Application of the final FDA model delivers a ROC score of 0.732, which is slightly better compared to the score of the test set. Using the default probability cut-off value of 50 % results in a specificity of 0.89 (i.e 89 % of all train pass bys, which do not contain flanging noise, are detected correctly) and a sensitivity of 38 % (i.e 38 % of trains containing flanging noise are classified correctly by the model). Especially the latter score is not sufficient for flanging noise detection, hence a lower probability cut-off is needed. Statistically, the top left point as optimised balance between specificity and sensitivity is detected on the ROC curve at 37.1 %. Maximising the Youden index leads to a similar point and is thus not separately mentioned. Empirical investigation is focused on slope change points in the ROC curve. Roughly three slope changes are visible – slightly above 0.5, slightly below 0.371 and at 0.25. Between 0.371 and 0.25 the slope is similar to the 45 degree line, which indicates equality in sensitivity increase and specificity loss. Above 0.25, a further raise of sensitivity causes a high loss of specificity in comparison. Thus, lower probability thresholds seem not beneficial. The statistically evaluated point at 37.1 % leads to 71 % specificity and 63 % sensitivity, which would be a balanced option to prevent high false positive rates, however the sensitivity is still mediocre. The cut-off value of 0.25 leads to a low specificity of 44 %. The latter value states that more than half of trains without flanging noise are misclassified as containing flanging noise. However, reached detection of flanging noise is raised to 86 %, which is a reliable value.

To get an overview of changes in application on test set and validation data by setting different threshold values, impact is summarised in Table 51. As can be seen, reducing the threshold probability to 37.1 % causes no significant impact on overall classification accuracy combined with a distinct raise in sensitivity. Thus, the value of 37.1 % shows superior performance for the present task compared to 50 %. Further lowering of the cut-off probability causes a decrease in accuracy of 7 % to 9 %. Moreover, especially regarding C3, the specificity is lowered to 14 %, which results in a huge false positive classification. Hence, it can be concluded that the threshold of 37.1 % states a stable and balanced choice on all validation sets with especially well performance in C2 and C3 – both sections with higher radii. To increase sensitivity and thus prediction on the safer side accepting a lower accuracy, the threshold of 25 % is more suitable than 37.1 %. For application of the model, classification thresholds of the discriminant score function (Equation (58)) are provided in the last row of Table 51. Again, train pass bys are predicted positive (flanging noise occurrence) if the discriminant score is beneath the corresponding threshold and in case of equality or exceeding the threshold classified as negative (no flanging noise).

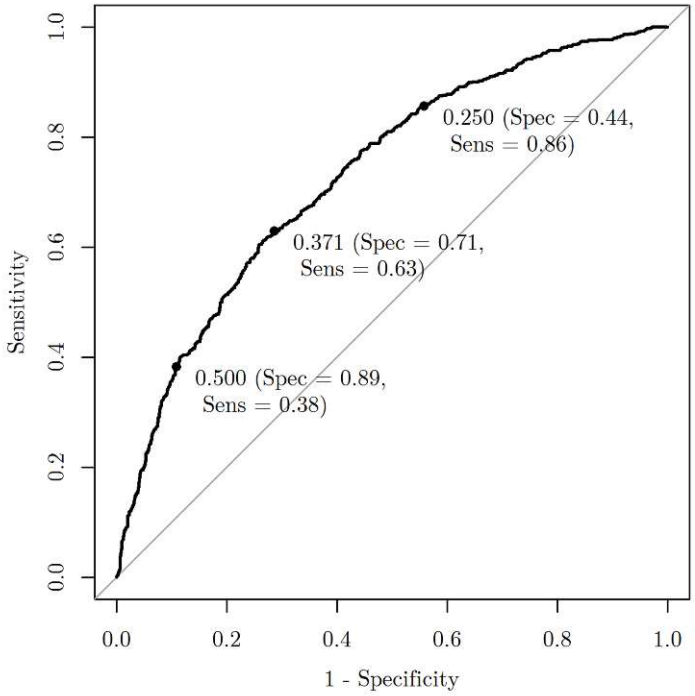

Figure 76: ROC curve for the final FDA model applied to the evaluation set with default cut-off (50 %), statistically optimised point (37.1 %) and even higher sensitivity (25 %)

cut-off probability		**0.500**	**0.371**	**0.250**
test set	Acc	0.71	0.68	0.59
	Sens	0.31	0.56	0.83
	Spec	0.92	0.74	0.47
C2	Acc	0.70	0.68	0.60
	Sens	0.64	0.67	0.75
	Spec	0.73	0.69	0.53
C3	Acc	0.58	0.59	0.52
	Sens	0.74	0.88	0.98
	Spec	0.44	0.35	0.14
C4	Acc	0.69	0.70	0.61
	Sens	0.10	0.53	0.83
	Spec	0.98	0.78	0.50
C2,C3,C4	Acc	0.67	0.67	0.59
	Sens	0.42	0.66	0.84
	Spec	0.81	0.68	0.45
discriminant score threshold		-0.9166	-0.2779	0.4073

Table 51: Achieved benchmarks on test set and validation sets with alternative cut-off probabilities for flanging noise prediction as well as corresponding threshold values for discriminant score function

Impact of environmental predictors

Predictor importance scores of the final FDA model for flanging noise predictions (refer to Figure 77) indicate that especially difference between rail temperature and dew point shows a significant impact on the model. Rain is rated in midfield, however in third place of six considered predictors. As previously mentioned, the final FDA model does not include frost or dew presence. Thus, altogether, environmental predictors seem important and necessary for prediction.

To estimate importance of considering environmental predictors, new models without those are built. Hence, predictors are only TrainType.1, TrainType.2, DoubleUnit, RSI and AccSide. Again, the focus lies on models trained on data points of C1 and C5, which are chosen on behalf of reached benchmarks in prediction of the validation sets C2, C3 and C4. In that case the Boosted Tree method outperforms all other approaches, followed by FDA (benchmarks summarised in Table 52). As can be seen, performance loss is heavily dependent on the validation set. While C2 shows a negligible worsening, application on C4 indicates a severe decline. Building on the pattern, it allows the estimation that the impact of environmental predictors is significantly higher in conjunction with lower curve radii ($\leq$ 310 m) than otherwise. Since the validation sets contain no even data point distribution among the year, an influence is also possible due to that. Data points in C2 are monitored in July with very hot temperatures, while data points in C4 are mainly from February, March and April. Altogether, it can be concluded that in the test set as well as in application on all data from validation sets, a performance loss due to exclusion of environmental predictors is present. However, the predictive ability is nevertheless given, since the benchmark of the test set is not severely deviating from performance on validation sets.

The opposite case – only considering environmental parameters for prediction – is also examined with the same boundary conditions. New models are trained on C1 and C5 and applied on validation sets for deciding which approach performs best. Included predictors are RT.DPDiff, RainDetect, FrostPresent and DewPresent. In that task, FDA and the Neural Network algorithm perform nearly equally. Due to simplicity and interpretability, the former

	Boosted Tree	**FDA**	**loss**
test set	0.660	0.647	0.065
C2	0.694	0.697	0.015
C3	0.610	0.601	0.062
C4	0.560	0.536	0.181
C2,C3,C4	0.648	0.621	0.074

Table 52: Achieved area under ROC curve by best performing model approaches without environmental conditions Boosted Tree and FDA trained on data points of C1 and C5 and loss in ROC compared between Boosted Tree and final FDA model with all selected predictors

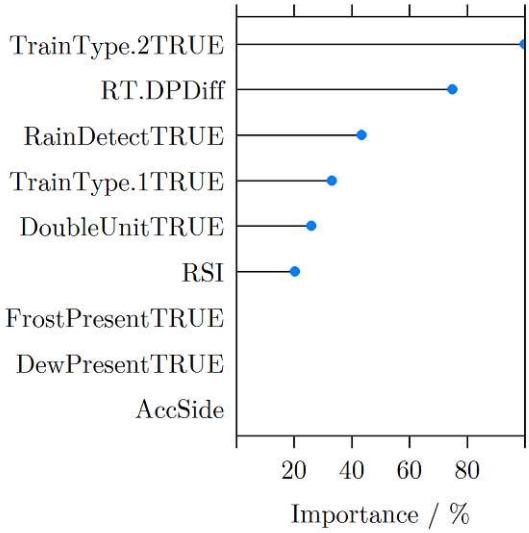

Figure 77: Predictor importance scores of the final FDA model

	FDA	**Neural Network**	**loss**
test set	0.646	0.649	0.079
C2	0.520	0.520	0.152
C3	0.631	0.624	0.041
C4	0.718	0.717	0.023
C2,C3,C4	0.637	0.637	0.085

Table 53: Achieved area under ROC curve by best performing model approaches considering only environmental conditions FDA and Neural Network trained on data points of C1 and C5 and loss in ROC compared between FDA and final FDA model with all selected predictors

method is chosen. Moreover, the FDA model only includes RT.DPDiff and RainDetect as predictors, thus modelling dew formation on the rail can be skipped. Compared to the other extreme, where no environmental conditions are taken into account, the opposite pattern is observable. The model is capable of predicting data points in C3 and C4 with minor losses, while it lacks to predict C2 sufficiently. Thus, it seems that environmental conditions have lower importance in terms of predicting flanging noise occurrence at lager curve radii (440 m) and/or in conjunction with hot temperatures. Overall performance loss is comparable between neglecting environmental predictors and considering only them, which means that both groups are roughly equally important. Hence, inclusion of environmental predictors is beneficial for flanging noise prediction.

6.4.2 Squeal noise occurrence

In a first step, models are built with the final predictor set only alternating the relative humidity indicators – Hum, AT.DPDiff, HumRail, RT.DPDiff – to evaluate the best performing

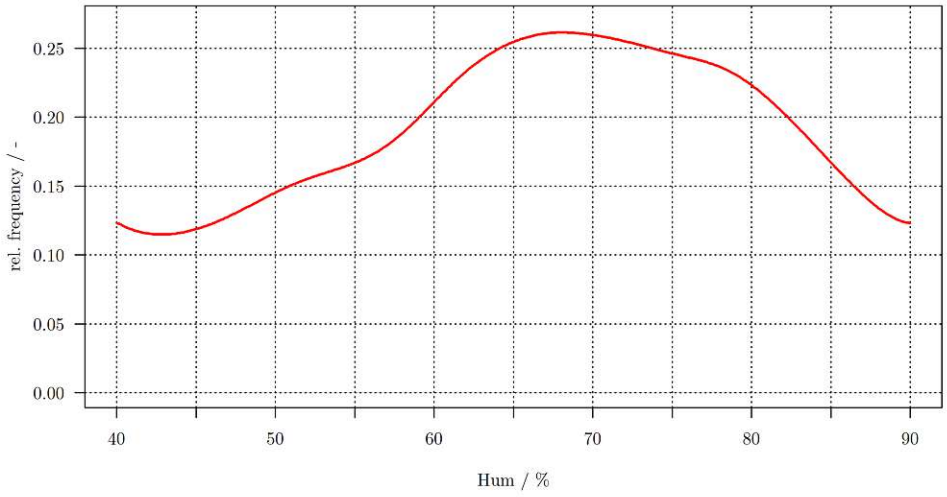

Figure 78: Relative frequency of squeal noise occurrence among the range of relative air humidity – data from the whole dataset

one for prediction of squeal noise occurrence. The focus lies on models trained on data points of C1 and C5 and applied to C2, C3 and C4 as validations sets. The best performing predictors are relative air humidity and relative rail humidity with slightly better performance of the former one. Thus, relative humidity of the ambient air (monitored beside the section) is chosen as final predictor. Relative frequency of occurrence distribution among that predictor (see Figure 78) indicates a high probability of squeal noise between 64 % and 74 % (exceeding 25 %), while low values cause a lower probability and higher values declines probability significantly. Not much thought should be given to the edge regions as data points (especially regarding dry conditions) are sparse and partly depict outlier points in certain sections. Thus, depiction is reduced to a range from 40 % to 90 %. Although empirical examination indicates a much better and smoother correlation regarding relative rail humidity, prediction benchmarks in C2 and C3 can be significantly improved by inclusion of relative air humidity. Thus, it seems that prediction of squeal noise occurrence becomes more robust in application on other datasets with different radii and climatic conditions.

Final prediction model

The difference between resampled training set performance as well as test set benchmarks compared to validation sets is even more severe than in case of flanging noise prediction due to higher class imbalance. Since probability decreases significantly with growing curve radius, prediction of data in C2 and C3 is very challenging for the model algorithms. The only approach, which is capable to achieve a slight predictive performance on C2 and C3, is a Neural Network. Although it also performs bad on C2 and C3 (see Table 54), it is able to nearly keep the performance from the test set in application on the whole validation set (C2, C3 and C4) as well as in C4. Just to see the difference, the second

	Neural Network	Naïve Bayes
test set	0.736	0.707
C2	0.562	0.471
C3	0.587	0.529
C4	0.790	0.725
C2,C3,C4	0.728	0.669

Table 54: Achieved area under ROC curve by best performing model approaches Neural Network and Naïve Bayes trained on data points of C1 and C5

best approach (Naïve Bayes), which completely fails to predict samples from C2 (area under ROC curve is below 0.5) as well as performing worse on data of C3, is also added. Again, complex tree-based models (like Random Forest) perform really well on the dataset, however predictive ability is limited to similar boundary conditions due to over-fitting. Thus, those approaches are suitable for predicting a particular section with a similar train type distribution, however they fail to predict new sections and their applicability on other datasets lacks for the present modelling task. Performance difference is not severe compared to feature selection (refer to Table 39), which justifies the predictor selection.

Functionality of neural network models is briefly described in Subsection 6.1.3. It is a non-linear modelling approach, which tries to catch correlations between predictors and outcome randomly. The final model consists of one hidden unit layer with nine neurons, which represent linear combinations of the predictors. Additionally, to prevent over-fitting, a weight decay value of 0.1 performs best in resampling. The whole model structure (refer to Figure 79) consists of eight neurons representing the input layer (illustrated as I1 to I8), nine neurons in the hidden unit layer (called H1 to H9), two bias neurons (B1 and B2) and one output neuron (O1). The score of the H1 to H9 is calculated as follows [113]:

$$H_i = g\left(B_{1,i} + \sum_{j=1}^{P} \alpha_{i,j} \cdot I_j\right) \tag{59}$$

$$g(x) = \frac{1}{1+e^{-x}} \tag{60}$$

where H_i represents the ith hidden unit score of H1 to H9, $B_{1,i}$ depicts the bias term, which is calculated as one multiplied with the assigned weight value to each neuron in the hidden unit layer, P is the predictor quantity, $\alpha_{i,j}$ are weight values for each predictor j corresponding to hidden unit i, I_j states the predictor input and $g(x)$ is the sigmoidal function applied to H_i to get scores between zero and one.

The outcome value is derived in a similar way and also transformed by Equation (60) [113]:

$$O1 = g\left(B_2 + \sum_{i=1}^{U} \beta_i \cdot H_i\right) \tag{61}$$

where $O1$ represents the outcome value (between zero and one), B_2 depicts the second bias term, which is also calculated as one times the assigned weight value, U is the neuron quantity in the hidden unit layer and β_i are weight coefficients for each hidden unit i.

Attention should be drawn to line width in Figure 79 as it depicts the magnitude of the weight value. It can be seen that high weight values are present for predictors RSI and DewPresent, while very thin lines are connected to AccSide. The colour of the lines indicates whether a positive (black) weight or a negative one (grey) is present. Since predictors are standardized and skewness removed (by Yeo-Johnson transformation), direct interpretation of the sign is not possible. Due to standardized input predictors, magnitude of weight coefficients can be compared and allow an impact estimation on the prediction model. They are used to calculate predictor importance scores (illustrated in Figure 80). The interpretation of line width is justified as dew presence and RSI have the most impact on the model. Relative air humidity, double unit operation, train type and frost presence are of mediocre importance (40 % to 50 %). Rain has few impact and lateral acceleration contributes negligibly to the model prediction. The former is surprising since rain suppresses squeal noise occurrence significantly. However, overall, rain is detected in very few pass bys at all, which explains that the correlation is not seen as important by the model due to lack of sufficient data points. Dew as a wetness source is far more common, thus the reduction in probability should be represented sufficiently. Table 55 provides the needed weight coefficient values for derivation of squeal noise probability. The first nine rows contain the weight values $B_{1,i}$ of the first bias neuron (B1) and $\alpha_{i,j}$ for calculation of each hidden unit neuron value by using Equations (59) and (60). The last two rows provide the value of the second bias neuron (B2) as well as weight coefficients β_i to derive the outcome probability $O1$ from the hidden unit scores by applying Equations (61) and (60). The outcome represents the probability value of a negative event (i.e no squeal noise occurrence). Thus, for computation of squeal noise probability, 1 minus the outcome has to be calculated. Again, it has to be pointed out that new predictor data cannot be fed directly to the model due to the need of pre-processing. Building on the derived probability, samples are classified as positive or negative depending on the chosen cut-off probability, which is investigated further in the following subsection[22].

[22]The model object from the final model in R is available at https://github.com/M-Ostermann/curve-noise-prediction

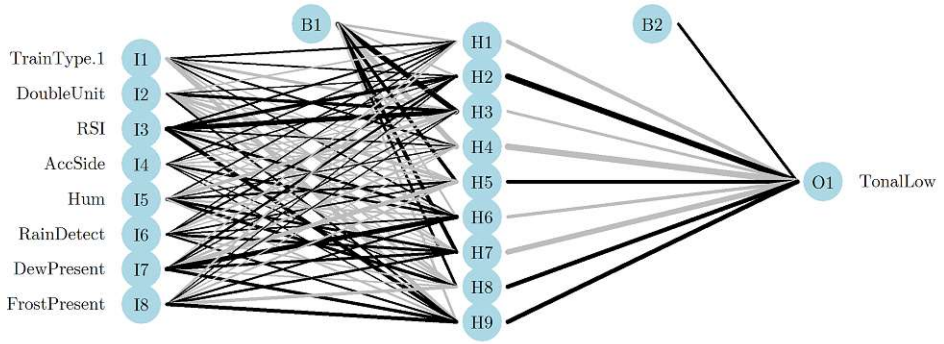

Figure 79: Final neural network model for squeal noise prediction with nine hidden units, two bias neurons and one outcome for class prediction – line width indicates the magnitude of the assigned weight value, whereas line colour shows if the sign of weight value is positive (black) or negative (grey)

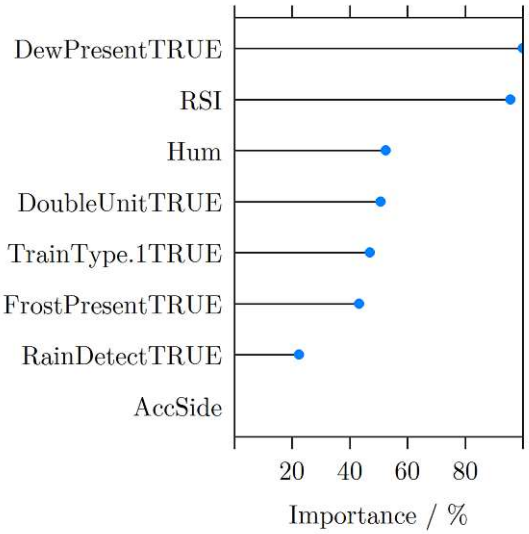

Figure 80: Predictor importance scores of the final neural network model calculated with assigned weight values for the neuron connections

	H1	H2	H3	H4	H5	H6	H7	H8	H9
B1	-1.6011	-1.0594	4.9150	-5.0952	2.5604	-1.2243	3.7536	1.9600	-1.8265
I1	1.0030	0.6102	-2.0090	1.3192	-0.3700	0.8061	-0.4430	-0.6646	-3.3272
I2	-0.6524	0.1368	2.3982	-0.4421	-2.9787	-0.5800	-1.7831	-1.9367	1.0906
I3	0.9203	3.0583	4.3213	0.8358	-1.3189	2.1036	-0.3300	-1.0595	3.6796
I4	0.6194	0.3523	-0.0242	-0.5894	-0.5167	1.4490	-0.1540	0.0819	0.0865
I5	0.7402	-0.2550	-2.8799	0.0226	1.6146	-1.2405	2.2779	2.4368	-0.9711
I6	0.1811	1.1695	-0.9406	-0.2602	0.6374	0.7746	1.9426	-0.5381	1.3541
I7	-0.6239	1.7616	0.0840	-1.7434	-3.8562	4.2855	2.2236	-0.5067	2.5537
I8	-1.3807	0.0578	0.3324	-0.0440	-1.0949	0.9296	1.5450	-2.4939	2.9403
B2	1.9913								
O1	-3.3387	5.3640	-2.2241	-5.9726	3.0204	-2.7273	-4.8879	4.1228	3.9468

Table 55: Weight values $\alpha_{i,j}$, β_i, B1 and B2 for derivation of squeal noise probability with the final neural network model

Alternative cut-off probability

Suitable cut-off probabilities are investigated on the evaluation set, which contains 10 % data points of C1 and C5. It is only used for the current examination task. Application of the final Neural Network model results in a ROC benchmark of 0.773, which is significantly better than the test set. However, it only indicates a scattering in the benchmarks, which could be a result of large possible alterations in the predictor data and/or of additional unknown impact factors. It states that the model can perform better than believed building on the test set scores. Application of the default probability cut-off value of 50 % or 0.5 leads to a specificity of 100 % and a sensitivity of 2 % (refer to Figure 81), which shows that the default value is not suitable for the present modelling task. Statistical investigation searching for the top left point of the ROC curve results in a threshold probability of 17.5 %. Maximising the Youden index leads to a similar value, hence depiction is skipped. Specificity at that point is 69 % and sensitivity is 75 %, which looks like a balanced choice as well as a reasonable value. Due to the severity of squeal noise occurrence and the possible impact on overall noise level, an even higher sensitivity might be aimed to stay on the safe side. On the other hand, low specificity leads to high false-positive rates. Thus, a balanced trade-off between sensitivity and specificity should be aimed. Empirical examination of the ROC curve focuses on slope changes. As can be seen, the first slope change lies at roughly 0.4 sensitivity, which is too low to consider. The second point is equal to the statistical optimisation. Afterwards, the slope magnitude drops slightly beneath the 45 degree line, which indicates that an increase in sensitivity, leads to a higher loss in specificity. Nevertheless, it might be beneficial to raise sensitivity further to stay on the safe side in prediction. The next slope change depicts a cut-off probability of 10 % – leading to a sensitivity of 90 % and a specificity of 45 %. Lower thresholds cause severe false positive rates in conjunction with a slight raise in sensitivity and are therefore not suitable.

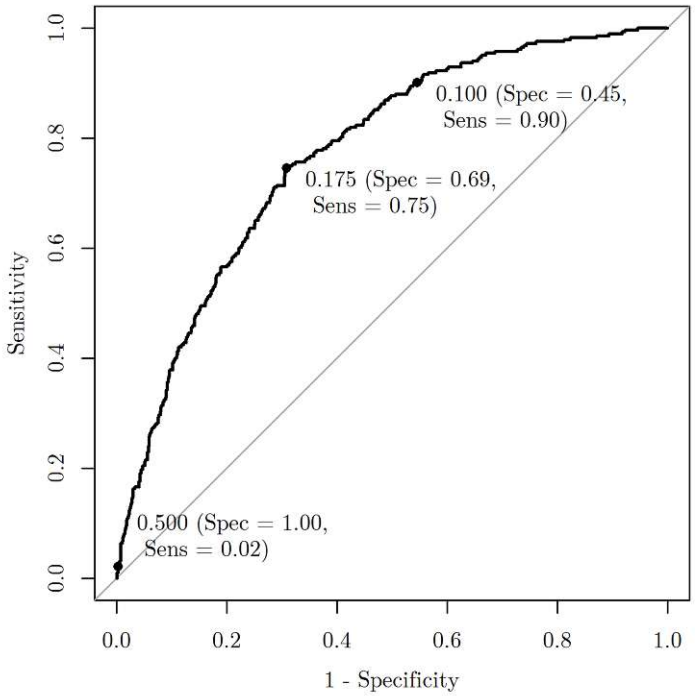

Figure 81: ROC curve for the final Neural Network model applied to the evaluation set with default cut-off (50 %), statistically optimised point (17.5 %) and even higher sensitivity (10 %)

Both threshold probabilities – 17.5 % and 10 % – seem to be a reasonable choice. Application on test set and validation sets (see Table 56) justifies that the default threshold of 0.5 is not suitable for squeal noise prediction. Moreover, a cut-off probability of 0.175 also shows a bad performance of 25 % to 28 % sensitivity on the validation sets. Thus, the probability threshold of 10 % is recommended for prediction as it performs well on test set (81 % sensitivity) and validation sets (44 % to 68 % sensitivity). Hence, for squeal noise prediction with the presented Neural Network model, an outcome score of greater than 0.9 leads to a negative prediction (i.e no squeal noise occurrence) and otherwise less or equal than 0.9 predicts the train pass by as positive (i.e squeal noise occurrence). If the scope is set to high overall classification accuracy, a cut-off probability of 0.175 can be applied.

cut-off probability		0.500	0.175	0.100
test set	Acc	0.86	0.72	0.53
	Sens	0.05	0.59	0.81
	Spec	1.00	0.74	0.48
C2	Acc	0.94	0.86	0.74
	Sens	0.06	0.28	0.44
	Spec	0.96	0.87	0.75
C3	Acc	0.90	0.88	0.74
	Sens	0.20	0.25	0.68
	Spec	0.99	0.96	0.74
C4	Acc	0.91	0.83	0.68
	Sens	0.14	0.27	0.68
	Spec	0.97	0.88	0.68
C2,C3,C4	Acc	0.91	0.83	0.68
	Sens	0.14	0.27	0.68
	Spec	0.97	0.88	0.68

Table 56: Achieved benchmarks on test set and validation sets with alternative cut-off probabilities for squeal noise prediction

Impact of environmental predictors

Variable importance scores in the fitted Neural Network for squeal noise prediction (see Figure 80) indicates that environmental predictors – especially dew presence – have a significant impact on the model. Relative humidity and frost presence are ranked in midfield and rain shows low importance. Hence, environmental predictors seem essential for squeal noise prediction.

Impact estimation of environmental parameters is carried out by building new models without them. Thus, only RSI, DoubleUnit, TrainType.1 and AccSide are included. The focus lies on models trained on data of C1 and C5 and performance estimation for deciding the final model on validation sets. In this scenario only linear models (logistic regression, PLS, GLM-Net) are able to reach an informative prediction (ROC benchmark above 0.5) in application on C2 and C3. They are also performing equally compared to other approaches on test set and other validation sets. Due to simplicity, logistic regression is chosen for depiction (benchmarks illustrated in Table 57). As can be seen, the loss in performance regarding C4 and the whole validation set is large, while test set and C2 indicate a mediocre decline. Interestingly, C3 is predicted with hardly any worsening, although the measurements took place in September, where dew occurrence is frequent and thus environmental conditions are assumed to have a significant impact. Altogether, performance of the fitted Neural Network model with all considered predictors is already bad in C2 and C3 and would be significantly worsened in the other sets by exclusion of environmental predictors. Thus, their inclusion is reasonable and result in a major improvement of the prediction model.

	Log Reg	loss
test set	0.689	0.047
C2	0.524	0.038
C3	0.581	0.006
C4	0.666	0.124
C2,C3,C4	0.624	0.104

Table 57: Achieved area under ROC curve by best performing model approach without environmental conditions Log Reg trained on data points of C1 and C5 and loss in ROC compared to the final Neural Network model with all selected predictors

	Neural Network	loss
test set	0.607	0.129
C2	0.412	0.150
C3	0.559	0.028
C4	0.710	0.080
C2,C3,C4	0.558	0.170

Table 58: Achieved area under ROC curve by best performing model approaches for squeal noise prediction considering only environmental conditions Neural Network trained on data points of C1 and C5 and loss in ROC compared to the final Neural Network model with all selected predictors

The opposite case – including only environmental predictors for squeal noise prediction – is examined using the same boundary conditions. New models are trained with predictors Hum, FrostPresent, RainDetect and DewPresent on data from C1 and C5, while decision of the best model approach is carried out by results on the validation sets. For that task, the Neural Network algorithm outperforms the other methods (benchmarks summarised in Table 58). The performance loss is severe compared to the case of considering all predictors and also worse than excluding all environmental parameters. The model is completely uninformative regarding prediction of data in C2 (ROC beneath 0.5) and also performs bad on test set as well as other validation sets except regarding C4, where an acceptable benchmark is reached. The fairly high result in C4 allows the assumption that environmental predictors are more important in months February, March and April as the majority of data points are from that duration. Dominating conditions are frequent dew occurrence and rather low temperatures. However, that connection is too weak to extract a generalised statement. To conclude, squeal noise prediction only depending on environmental conditions seems not sufficiently possible.

6.4.3 HF squeal noise occurrence

To evaluate which relative humidity predictor – Hum, AT.DPDiff, HumRail, RT.DPDiff – performs best regarding HF squeal noise prediction, models are trained with the final predictor set only alternating those four. Benchmark evaluation is focused on models trained on data

from C1 and C5 – the only two long-term campaigns –, which are applied to data of C2, C3, C4 and altogether as independent validation sets. The best performing predictor is relative rail humidity, which outperforms the other candidates significantly. Density distribution of frequency of occurrence regarding relative rail humidity is already depicted and discussed in Subsection 6.3 Figure 65 and thus not again illustrated.

Final prediction model

Performance – at least of the two best approaches – is not much worse between resampled training set, test set and validation sets (except C3), which indicates that the model is able to catch fundamental correlations between predictors and outcome. The best modelling approach is a Neural Network, followed by FDA (benchmarks illustrated in Table 59). The former significantly outperforms the latter in application on data points of C2, which justifies the model choice further. Regarding C3, performance is bad, however most models fail in prediction completely (ROC score beneath 0.5). Again, complex tree-based model algorithms (like Random Forest or Boosted Tree) seem to over-fit the dataset and are suitable for modelling a single measurement section from data and predicting similar data and train type distribution in the future. They are not applicable on new datasets. Performance loss due to the reduced predictor set is acceptable (beneath 0.1 ROC score compared to best performance in previous feature selection, see Table 44) for the scope of a practical model.

Detailed functionality of the Neural Network algorithm is described in Subsection 6.4.2 for squeal noise prediction. The final model to predict HF squeal noise occurrence consists of one hidden unit layer with nine neurons. To avoid over-fitting, a weight decay value of 0.1 performs best in resampling iterations. The whole model structure (refer to Figure 82) involves nine input neurons (depicting the predictors I1 to I9), two bias neurons (B1 and B2), nine neurons in the hidden unit layer (H1 to H9) and one outcome neuron (O1). Since predictors are standardized and skewness removed, weighting values can be directly compared to each other in terms of importance in the model. However, due to pre-processing measures, the sign of the weight coefficient cannot be interpreted. Considering the line width, which depicts magnitude of assigned weight value (black lines for positive sign and grey for negative sign), significant impacts come from RSI, HumRail and DewPresent, while AccSide indicates very thin lines and thus negligible impact. Observations are also justified by predictor importance

	Neural Network	**FDA**
test set	0.747	0.718
C2	0.739	0.588
C3	0.588	0.559
C4	0.710	0.691
C2,C3,C4	0.677	0.693

Table 59: Achieved area under ROC curve by best performing model approaches for HF squeal noise prediction Neural Network and FDA trained on data points of C1 and C5

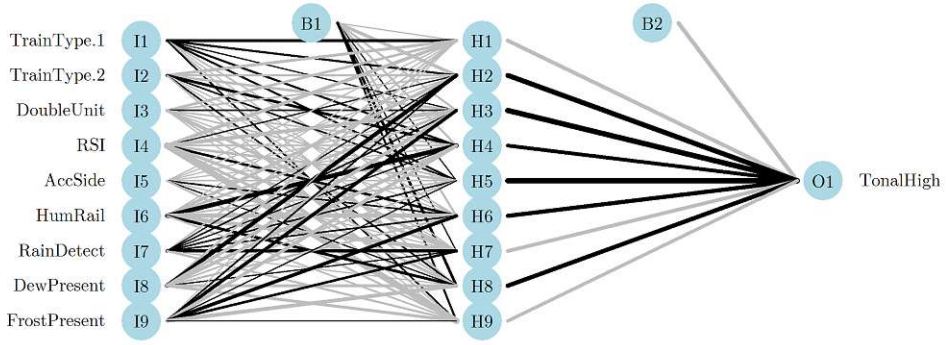

Figure 82: Final neural network model for HF squeal noise prediction with nine hidden units, two bias neurons and one outcome for class prediction – line width indicates the magnitude of the assigned weight value, whereas line colour shows if the sign of weight value is positive (black) or negative (grey)

scores (refer to Figure 83), which rely on assigned weight values. Top predictor scores are reached by RSI, relative rail humidity and dew presence. Frost presence lies in midfield, while train type as well as double unit operation and rain detect have low importance. The latter rankings are surprising, since vehicle impact is present in empirical examination as well as rain impact, which greatly reduces frequency of occurrence. However, data points with detected rain during a train pass by are rare, thus the model does not catch the impact completely. Nevertheless, influence of dew and frost seems to be well modelled. Complying with squeal noise, the value of lateral acceleration contribute negligibly to the model. Table 60 provides the needed weight coefficients for each connection. The first row represents the weight values between the first bias neuron (B1) and the hidden units (H1 to H9). The second to ninth row contain coefficients $\alpha_{i,j}$ for connections between input layer and hidden unit layer. The second last row depicts the weight value between the second bias neuron (B2) and the outcome (O1) and the last row provides coefficients β_i to connect the hidden unit layer to the outcome layer. Score of the hidden units can be derived with Equations (59) and (60). The probability value for classification is computed by using Equations (61) and (60). The calculated outcome represents the probability for a negative event (i.e no HF squeal noise occurrence). Thus, 1 minus the outcome has to be calculated to achieve the probability for HF squeal noise occurrence. It has to be pointed out that predictors need to be pre-processed prior to feeding them to the model. Depending on the outcome and the chosen threshold probability (default 0.5), train pass bys are classified as positive or negative. The cut-off probability is examined further in the following subsection[23].

[23]The model object from the final model in R is available at https://github.com/M-Ostermann/curve-noise-prediction

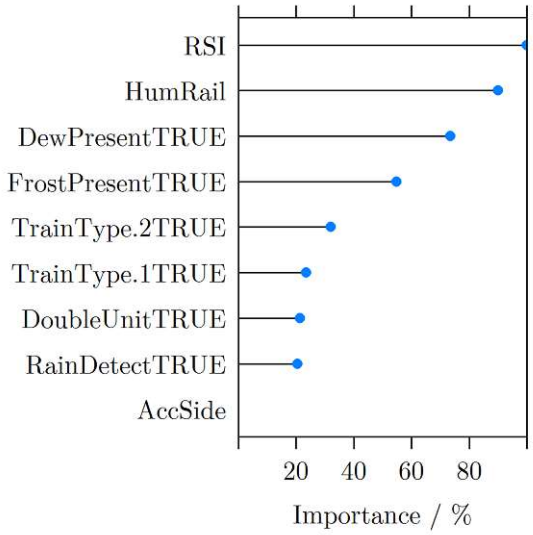

Figure 83: Predictor importance scores of the final neural network model calculated with assigned weight values for the neuron connections

	H1	**H2**	**H3**	**H4**	**H5**	**H6**	**H7**	**H8**	**H9**
B1	-0.1072	-3.2397	-0.9219	-1.2506	0.1875	-1.0961	0.1280	2.2175	1.1651
I1	2.0250	1.1442	0.0032	1.0861	1.0237	-0.5532	-0.1630	-0.5306	0.6735
I2	-2.2062	-1.0889	-1.6640	2.0364	0.5554	-0.3059	-0.3669	0.6917	-0.6115
I3	-1.5698	-0.7526	0.4639	-0.9005	-0.5749	-0.3442	-2.4189	-0.4696	-0.1516
I4	-3.1856	0.0725	-3.2890	-0.5510	-1.4485	1.8357	-2.8396	-1.9645	-6.1187
I5	-0.2508	-0.1924	-0.0131	-0.4151	-0.5256	0.9113	-0.1962	0.1258	-1.1019
I6	-0.0644	-3.1539	-2.2368	4.0832	-0.5237	-3.4321	-3.8272	2.1535	-0.7976
I7	-0.9006	0.1932	0.8117	0.3075	0.6903	0.7984	2.4606	-0.2296	-1.1134
I8	-1.8060	3.3942	-2.4855	-2.6270	-0.6449	-1.2058	0.1329	-2.0875	-2.1476
I9	-0.1416	0.5025	2.7009	-2.5285	0.0374	3.2375	1.5809	-3.2968	0.4198
B2	-3.3979								
O1	-3.7256	4.6184	4.9816	2.9120	4.8895	3.4005	-3.2407	4.0076	-3.1623

Table 60: Weight values $\alpha_{i,j}$, β_i, B1 and B2 for derivation of HF squeal noise probability with the final neural network model

Alternative cut-off probability

Examination of alternative cut-off probabilities is carried out on a small evaluation set (10 % data points in C1 and C5), which is neither used as test nor as training set. Application of the final neural network model on the evaluation set results in an area under ROC curve of 0.787, which indicates better performance as on the test set. Thus, scattering in predictions is present, which might be a result of large possible alterations in the predictor data and/or of additional unknown impact factors. However, it also shows that increased performance scores are possible. Plotting the ROC curve (see Figure 84) and marking the default probability threshold leads to a specificity of 100 % and a sensitivity of 4 %. Thus, it is not suitable for the present modelling task. Alternative cut-off points are found statistically at 19.3 % (top left point), which results in a balanced score of specificity 75 % and sensitivity 69 %. Maximizing the Youden index shows a slightly higher threshold (22.2 %), which leads to a lower sensitivity and thus is not considered further as the main scope is a high sensitivity. Empirical examination of the ROC curve focuses on the slope changes. Points beneath 19.3 % are not discussed further since achieved sensitivity is too low. The first slope change in the remaining range is located at a cut-off probability of 17 % with a specificity of 64 % and a sensitivity of 79 %, which is already an acceptable value. Looking for an even higher sensitivity leads to the second slope change point at a threshold value of 6 % probability – resulting in a specificity of 35 % and a sensitivity of 96 %. While the latter benchmark is extremely good, false positive rate is also fairly high. HF squeal noise can raise overall noise level significantly, however the importance of high emission frequencies between 8000 Hz and 12500 Hz is significantly weakened in propagation and hence in sound immission points due to atmospheric absorption. Another aspect is a reduction if A-weighting is applied, which is usually done in immission calculations. Thus, high sensitivity for HF squeal noise prediction might not be as important as regarding squeal noise.

Application of evaluated threshold probabilities on test set and validation sets (summarised in Table 61) reveals that sensitivity at cut-off points 0.193 and 0.170 is partly low, while at 0.06 false positive rate is huge. Thus, a cut-off value in between is iteratively derived searching for an acceptable score for both values. A threshold value of 10 % is chosen as a compromise. It achieves a stable sensitivity of 72 % to 78 % in the validation sets, while the specificity is held above 50 % – except regarding prediction on data points in C3, where a rather low specificity (27 %) is present. However, if overall classification accuracy should be improved on the cost of sensitivity, a threshold probability of 17 % is recommended. An illustration of the default value of 0.5 is skipped due to achieving hardly any sensitivity at all. Hence, for application of the presented Neural Network model with the scope of high sensitivity, an outcome score of greater than 0.9 classifies the train pass by as negative (i.e. no HF squeal noise occurrence), whereas values less or equal than 0.9 lead to a positive prediction (i.e HF squeal noise occurrence).

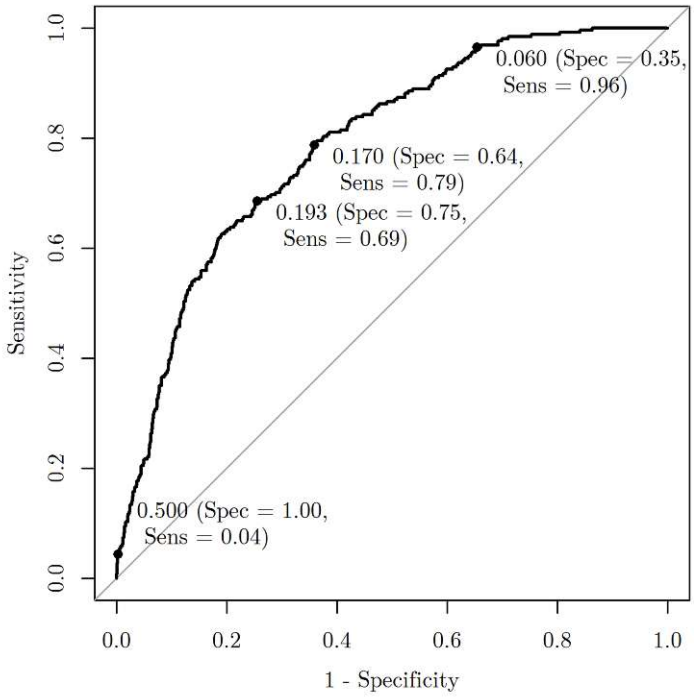

Figure 84: ROC curve for the final Neural Network model applied to the evaluation set with default cut-off (50 %), statistically optimised point (17.5 %) and even higher sensitivity (10 %)

cut-off probability		**0.193**	**0.170**	**0.100**	**0.060**
test set	Acc	0.73	0.66	0.53	0.43
	Sens	0.56	0.68	0.88	0.94
	Spec	0.76	0.66	0.47	0.35
C2	Acc	0.69	0.67	0.61	0.48
	Sens	0.61	0.64	0.72	0.92
	Spec	0.70	0.68	0.61	0.46
C3	Acc	0.56	0.49	0.35	0.30
	Sens	0.48	0.59	0.83	0.97
	Spec	0.57	0.48	0.27	0.20
C4	Acc	0.81	0.78	0.59	0.49
	Sens	0.25	0.32	0.77	0.88
	Spec	0.89	0.85	0.56	0.43
C2,C3,C4	Acc	0.73	0.69	0.55	0.45
	Sens	0.36	0.44	0.78	0.91
	Spec	0.77	0.72	0.53	0.40

Table 61: Achieved benchmarks on test set and validation sets with alternative cut-off probabilities for HF squeal noise prediction

Impact of environmental predictors

According to the achieved predictor importance scores in the final model for HF squeal noise prediction (illustrated in Figure 83), environmental parameters are ranked with high to mediocre importance. HumRail, DewPresent and FrostPresent show scores above 50 % relative importance and rain a score of 20 %. Thus, impact of environmental predictors on the prediction model is generally high and inclusion is therefore of importance.

Influence of environmental predictors is examined by training new models without considering them. Hence, only RSI, TrainType.1, TrainType.2, DoubleUnit and AccSide are used. The scope of model selection lies on achieved performance in prediction of the validation sets. Model training is carried out on data points of C1 and C5. The best performance is achieved by the Neural Network approach (benchmarks provided in Table 62). As can be seen, performance loss is high in nearly all applied sets. Predictive ability of the model is also bad, since achieved test set result is higher than in validation sets. Thus, exclusion of environmental predictors for HF squeal noise prediction worsens the predictive performance significantly.

The other extreme – HF squeal noise prediction only on behalf of environmental predictors – is investigated using the same strategy. Model training is carried out on data points of C1 and C5 and best performing model approach is evaluated from achieved benchmarks in validation sets. Considered predictors are HumRail, DewPresent, RainDetect and FrostPresent. FDA and Neural Network perform roughly equal (performance of both provided in Table 63). Due to simplicity, the former method is chosen for comparison. It includes only HumRail and RainDetect. Thus, modelling of dew formation can be skipped. Severe performance loss is observable in C2. On the test set as well as on other validation sets (except C3), a mediocre decline is present. Interestingly, in C3, no worsening of model performance compared to the final model with all predictors is shown. Thus, in that case, usage of only environmental predictors is possible without disadvantages. However, it states an exception and would comply with high impact from environmental conditions in September. Altogether, the performance loss is large. No complete failure of predictive ability among the validation set is observed –

	Neural Network	**loss**
test set	0.641	0.106
C2	0.607	0.132
C3	0.530	0.058
C4	0.614	0.096
C2,C3,C4	0.564	0.113

Table 62: Achieved area under ROC curve by best performing model approach without environmental conditions Neural Network trained on data points of C1 and C5 and loss in ROC compared to the final Neural Network model with all selected predictors

	FDA	Neural Network	loss
test set	0.675	0.683	0.072
C2	0.522	0.507	0.217
C3	0.589	0.591	-0.001
C4	0.660	0.656	0.050
C2,C3,C4	0.619	0.623	0.058

Table 63: Achieved area under ROC curve by best performing model approaches for squeal noise prediction considering only environmental conditions FDA and Neural Network trained on data points of C1 and C5 and loss in ROC between FDA and the final Neural Network model with all selected predictors

in contrast to squeal noise prediction. Thus, although bad performance is reached, modelling seems possible. However, it has to be pointed out that general applicability is questionable since performance deviation between test set and validation sets is fairly high. Hence, it is not recommended to carry out predictions considering only environmental predictors.

6.4.4 Equivalent continuous noise level

For evaluation of best performing environmental predictors, models are trained only altering those and comparing achieved benchmarks. The selection considers temperature predictors – TempAir, SatPressure, TempRail and SatPressureRail –, absolute humidity indicators – MixRatio, MixRatioRail, HumSpec and HumSpecRail – and relative humidity parameters – Hum, AT.DPDiff, HumRail and RT.DPDiff. Focus is given to models trained on data points of C1 and C5 and predictive performance on validation sets C2, C3 and C4. Building on that, rail temperature, specific humidity and relative rail humidity are chosen for the final prediction model. Using the formulated final predictor set, performance gain or loss by inclusion of categorical predictors, which cover curve squeal occurrence (Broad, TonalLow, TonalHigh), is investigated by training models adding one of them, afterwards only a combination of TonalLow and TonalHigh and finally all three together as predictors. In all carried out evaluations, the Random Forest approach significantly outperforms all other applied methods. Thus, performance change is depicted using the built Random Forest models in each step. RMSE benchmarks are summarised in Table 64. Performance increase is generally achieved by inclusion of each curve squeal indicator – the most regarding flanging noise occurrence (Broad). That seems surprising since flanging noise amplitudes are not as severe as those of squeal noise, however it can be explained by data point distribution. Train type G is the most common train type in the set and its equivalent continuous noise level is highly influenced by flanging noise due to its low basic rolling noise level. Moreover, flanging noise occurrence is more frequent compared to squeal or HF squeal noise. Thus, altogether, inclusion of flanging noise occurrence is more beneficial regarding overall prediction error. Performance can be further improved by inclusion of both squeal noise types (TonalLow and TonalHigh) compared to solely considering Broad. At last, including all three curve squeal types separately leads to an additional improvement especially regarding prediction in C4. Therefore, it is reasonable

	basic	Broad	TonalLow	TonalHigh	TonalLow and TonalHigh	all
test set	2.14	1.83	1.88	1.91	1.72	1.62
C2	4.14	4.10	4.15	4.17	4.08	4.04
C3	5.12	5.09	5.17	5.22	5.05	5.00
C4	4.11	3.36	3.57	3.64	3.24	2.92
C2,C3,C4	4.32	3.98	4.10	4.15	3.91	3.76

Table 64: Achieved RMSE (in dB) by a Random Forest model trained on data points of C1 and C5 with the final predictor set and alteration by including different categorical predictors for curve squeal occurrence

	basic	Broad	TonalLow	TonalHigh	TonalLow and TonalHigh	all
test set	0.89	0.92	0.91	0.91	0.93	0.93
C2	0.06	0.07	0.06	0.05	0.06	0.07
C3	0.25	0.13	0.22	0.25	0.16	0.11
C4	0.35	0.59	0.55	0.54	0.63	0.70
C2,C3,C4	0.52	0.58	0.57	0.55	0.58	0.61

Table 65: Achieved R^2 by a Random Forest model trained on data points of C1 and C5 with the final predictor set and alteration by including different categorical predictors for curve squeal occurrence

to consider all three distinguished types of curve squeal (Broad, TonalLow and TonalHigh) as additional categorical predictors in the final model to predict equivalent continuous noise level. The improvement in R^2 benchmarks (refer to Table 65) justifies the decision further. Although a Random Forest model is a black-box, it is described and analysed especially regarding predictor contribution to the predicted outcome in the following subsection. However, an explicit calculation form like for the previous classification tasks cannot be provided. Cut-off probabilities need not be discussed since it is a regression task and the prediction states the final outcome value rather than a probability.

Final prediction model

Predictive performance regarding the final Random Forest model shows extremely good benchmarks in application on test set – 1.6 dB RMSE and 93 % correlation between prediction and outcome. By inclusion of curve squeal predictors, results can be improved compared to the model built on the whole dataset with all considered predictors (depicted in Table 48). Thus, final predictor selection is further justified. Performance on validation sets – representing benchmarks on new section with different train types and climatic conditions – is significantly worse compared to the test set results. In C4, similar boundary conditions occur compared to C5, where the model is trained on, which leads to mediocre benchmarks with RMSE of

2.9 dB. Regarding C2 and C3, accuracy is improvable with RMSE values of 4 dB and 5 dB, which allows the conclusion that the model may show a weakness in taking deviating section properties from those in C1 and C5 into account. Therefore, more data points in different sections would be needed to improve the model.

In general, the construction algorithm for a Random Forest model is described in Subsection 6.1. The final model uses for each split decision five random predictors and consists of 500 unpruned regression trees. Each of them contains over 10000 nodes. Thus, depiction of every single tree is not possible. The basic functionality in terms of prediction is illustrated in Figure 85 with a simplified tree ensemble, where each tree has only three nodes. Despite the black-box behaviour due to the vast amount of large trees in the model, analysis of the model can be done by investigating the predictor distribution in each tree and thus their importance in the model. Some basic tendencies of the model can be extracted from Figure 86. To every predictor, a score of minimal depth[24] in each of the 500 trees in the ensemble is marked with the colour theme. Additionally, the minimal depth is averaged to achieve the stated score for each predictor, which dictates also the order of depiction. Environmental predictors are ranked among the last places as well as show no appearance in the first or second level of splits. Thus, their importance in the model seems lower compared to other predictors. All three curve squeal predictors are ranked high, which justifies their inclusion further. However, they are hardly ever used as the first splitting node. Other top ranks are achieved by train type variables and PC1, which states a principal component for taking TDR and RR into account. The latter is used most in the first splitting node, which indicates a high influence of RR and TDR values – 60 % variance among them is described by PC1. Sleeper material (distinction between wood and concrete) shows an interesting behaviour – on the one hand, the second most appearances in the first splitting node among all predictors, while on the other hand, it states the only parameter, which is not included in over 100 trees at all. Thus, its influence in some trees of the ensemble is high, while no impact is present in other trees at all. Hence, the dependency of sleeper material seems well represented, since it should not have too much decisive impact. Altogether, predictor representation in the model seems reasonable and complies with empirical findings discussed in Subsection 6.3.5[25].

[24] Minimal depth in conjunction with each predictor means the depth of the first appearance in a splitting criterion. Generally, a low minimal depth indicates a higher importance.

[25] The model object from the final model in R is available at https://github.com/M-Ostermann/curve-noise-prediction

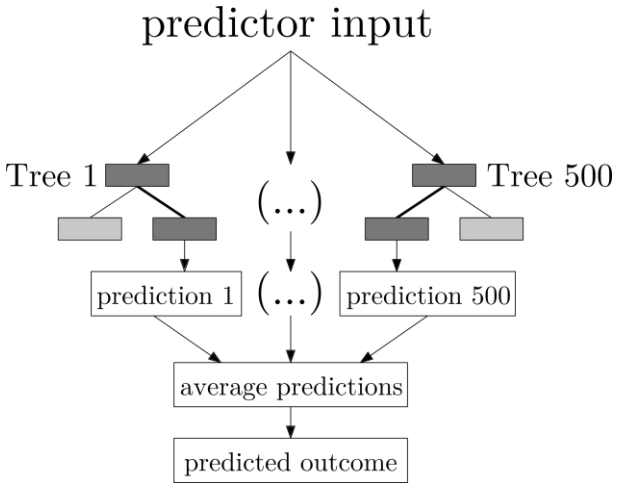

Figure 85: Simplified structure of a random forest model with 500 regression tress and each tree consisting of three nodes

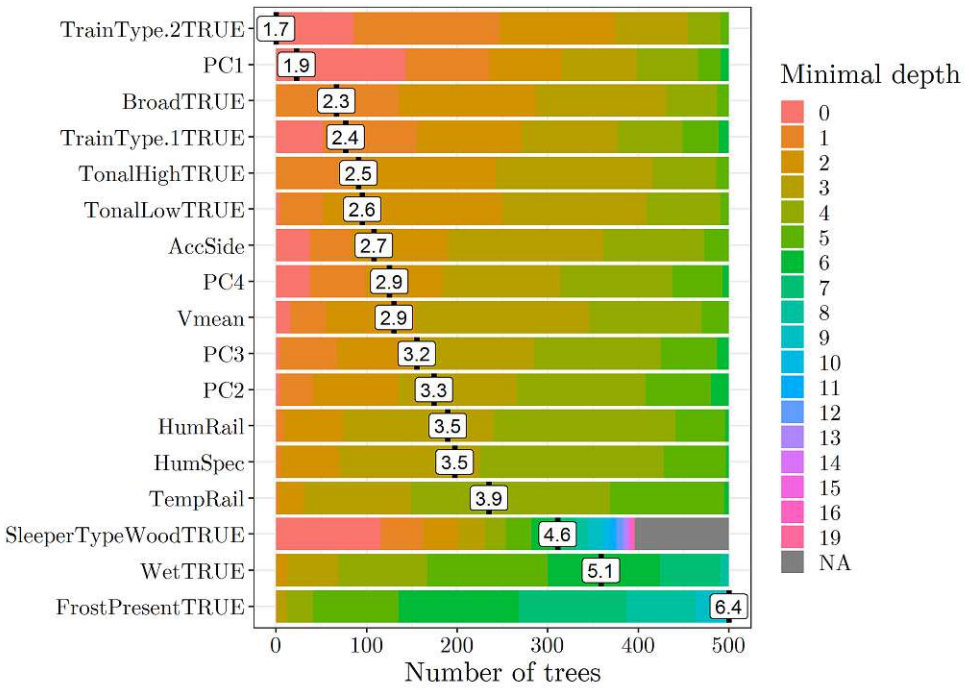

Figure 86: Plot of all considered predictors with their minimal depth among the 500 trees and their mean value of minimal depth

Impact of environmental predictors

Impact estimation of environmental predictors is carried out by building models without them. Thus, only PC.(1-4), SleeperTypeWood, TrainType.(1-2), Vmean, AccSide, Broad, TonalLow and TonalHigh are considered. The focus lies again on models trained on data points of C1 and C5 and their achieved benchmarks on validation sets C2, C3 and C4 (see Table 66). Again, a Random Forest model achieves top results. Performance regarding RMSE as well as R^2 seems weakly effected by exclusion of environmental predictors. All sets apart from C2 show a performance loss of up to 0.18 dB in RMSE as well as up to 0.04 in R^2, while C2 even indicates a slightly improved performance. That tendency also complies with analysed findings in terms of predictor impact on the final model – discussed in the previous paragraph. Thus, an exclusion of environmental predictors is possible causing a weak performance loss, however simplifies data acquisition for practical use significantly[26].

The other extreme – prediction of equivalent continuous noise level using only environmental predictors as well as curve squeal indicators – TempRail, HumRail, HumSpec, Wet, FrostPresent, Broad, TonalLow and TonalHigh – is examined applying the identical strategy. Model training is performed on data points of C1 and C5, while model selection is based on achieved benchmarks on validation sets C2, C3 and C4. A Random Forest model outperforms all other considered methods significantly (results provided in Table 67). Clearly, performance loss is severe in the test set as well as in all validation sets with a worsening in RMSE of up to 3.54 dB and in R^2 of up to 0.59. Thus, prediction of equivalent continuous noise level by using only environmental predictors – even in combination with curve squeal detection – is not recommendable. Moreover, it can also be concluded that considering environmental predictors as well as curve squeal indicators separately for predictive modelling in the present task leads to even worse results. Altogether, section, dynamic and vehicle properties are essential for prediction of equivalent continuous noise level, while environmental predictors contribute negligibly to predictive performance.

	RMSE	**loss**	**R^2**	**loss**
test set	1.71	0.09	0.93	0.00
C2	3.96	-0.08	0.11	-0.04
C3	5.03	0.03	0.09	0.02
C4	3.10	0.18	0.66	0.04
C2,C3,C4	3.81	0.05	0.61	0.00

Table 66: Achieved RMSE and R^2 by best performing model approach without environmental conditions Random Forest trained on data points of C1 and C5 and loss compared to the final Random Forest model with all selected predictors

[26]The model object from the final model in R without consideration of environmental conditions is also available at https://github.com/M-Ostermann/curve-noise-prediction

	RMSE	loss	R^2	loss
test set	5.16	3.54	0.34	0.59
C2	7.58	3.54	0.02	0.05
C3	6.06	1.06	0.05	0.06
C4	4.81	1.89	0.21	0.49
C2,C3,C4	6.09	2.33	0.08	0.53

Table 67: Achieved RMSE and R^2 by best performing model approach with only environmental predictors Random Forest trained on data points of C1 and C5 and loss compared to the final Random Forest model with all selected predictors

Application scenarios

In practical application of the model on other datasets involving new sections and train type distributions, no sound measurements for validation are available. Thus, curve squeal occurrence cannot be detected by the presented algorithm (refer to Chapter 4). It rather has to be predicted in the first place by applying the presented prediction models in the previous subsections (procedure illustrated in Figure 87). To estimate performance change by curve squeal prediction rather than evaluation of available sound measurements, all three formulated prediction models for curve squeal detection are applied to the whole dataset. In a next step, the results are implemented in the stated final Random Forest model for equivalent continuous noise level. Test set as well as validation set data points remain unchanged and especially the latter indicate performance change in application on new data points involving deviating section properties and train type distribution. Results are displayed in Table 68. The columns two to five contain RMSE and R^2 values achieved by using predicted curve squeal occurrence applying the recommended cut-off probability thresholds with the scope of high sensitivity – 0.25 for flanging noise and 0.10 for both squeal noise and HF squeal noise. Moreover, the loss or gain (in case of a negative value) compared to the final model with all predictors and curve squeal detection from measurement data (last columns in Table 64 and Table 65) is provided. As can be seen, performance loss is severe regarding test set and C4 with worsening of 2 dB in RMSE and R^2 decline of 0.20. In terms of other validations sets, changes are negligible – performance on data points of C2 even indicates an improvement. The partly high loss is not satisfying, thus prediction of curve squeal is additionally carried out with statistically recommended cut-off probabilities – 0.371 for flanging noise, 0.175 for squeal noise and 0.17 for HF squeal noise –, which results in an improvement of overall classification accuracy. The downside states a lower rate of correct curve squeal detection. Achieved benchmarks (depicted in the columns six to nine in Table 68) can be partly significantly improved – foremost on test set and validation set C4, however in all other sets also slightly. By comparison of Table 64 and Table 65 with Table 68, it can be seen that achieved benchmarks without inclusion of curve squeal predictors are worse among all validation sets in terms of aiming a high accuracy and partly taking a high sensitivity approach compared to prediction of curve squeal occurrence separately. Thus, for prediction of equivalent continuous noise level, inclusion of curve squeal predictors –

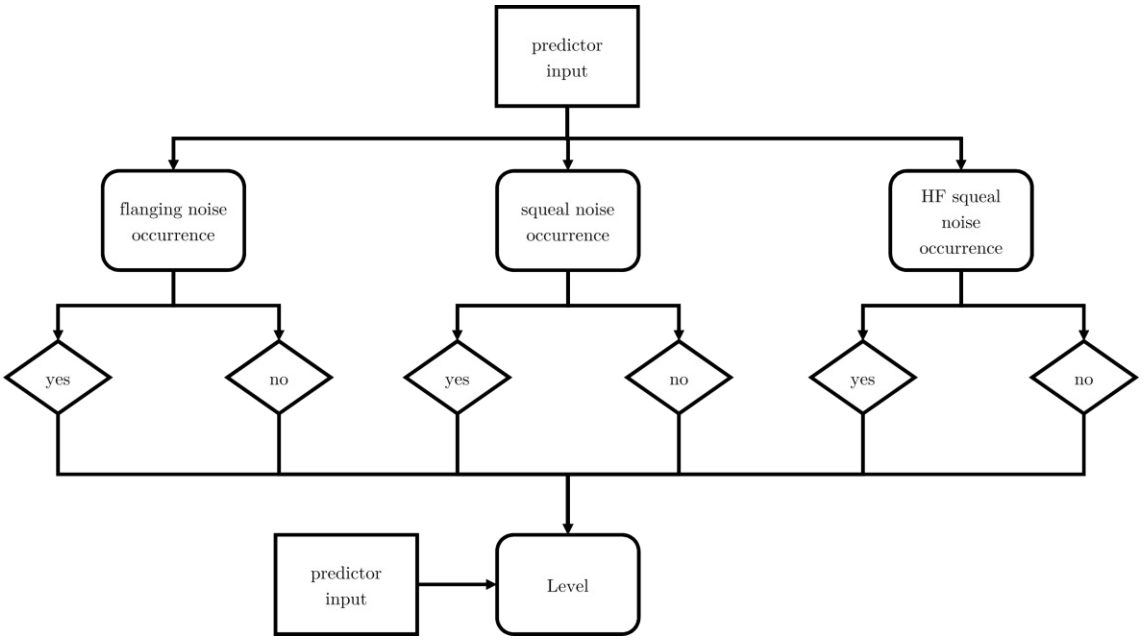

Figure 87: Structure of predicting equivalent continuous noise level on new datasets including prior curve squeal detection

	high sensitivity				high accuracy			
	RMSE	**loss**	**R^2**	**loss**	**RMSE**	**loss**	**R^2**	**loss**
test set	3.60	1.98	0.74	0.19	2.48	0.86	0.85	0.08
C2	3.85	-0.19	0.15	-0.08	3.83	-0.21	0.15	-0.08
C3	4.94	-0.06	0.02	0.09	4.95	-0.05	0.08	0.03
C4	4.98	2.06	0.32	0.38	3.64	0.72	0.42	0.28
C2,C3,C4	4.63	0.87	0.51	0.10	3.98	0.22	0.57	0.04

Table 68: Achieved RMSE and R^2 by best performing model approach with only environmental predictors Random Forest trained on data points of C1 and C5 and loss compared to the final Random Forest model with all selected predictors (refer to the last columns in Table 64 and Table 65)

even by using the separately formulated models for that task – results in an improvement in predictive performance compared to the case of excluding them out of the prediction completely. Regarding prediction of equivalent continuous noise level in conjunction with prior curve squeal occurrence prediction, cut-off probabilities of 0.371 for flanging noise, 0.175 for squeal noise and 0.17 for HF squeal noise are recommended.

Another thought has to be given to the behaviour of included TDR and RR values (sum of 114 predictors). They are represented by four principal components, which catch over 95 % variance among the whole dataset. In contrast to Table 30, where six components are stated – selected by the criterion of including all principal components with an eigenvalue of greater or equal one –, an overall variance cut-off at 95 % is defined. The decision is justified by the scope of general predictor reduction in the final prediction model. Each component

consists of linear combinations of all considered predictors. Thus, the score can be calculated as follows [113]:

$$PC_i = a_{i,1} \cdot P_1 + a_{i,2} \cdot P_2 + ... + a_{i,j} \cdot P_j \tag{62}$$

where index i represents the ith principal component, index j states the jth predictor, $a_{i,j}$ are assigned coefficient for each component and predictor and P_j is the value of each predictor.

In the present case, i counts from one to four and j from one to 114. Assigned coefficients to transform TDR and RR values from new sections are provided in Appendix D Tables 96, 97, 98 and 99[27]. In case of missing TDR and RR information, a general approach could be to apply the threshold spectra according to ISO 3095 [10]. To achieve values in the edge regions – in terms of rail roughness the 0.25 cm wavelength band is missing and regarding track decay rate, values of 100 Hz to 200 Hz third-octave bands are missing – rail roughness and lateral track decay rate are linearly extrapolated and vertical track decay rate is continued constant (values provided in Table 69). It has to be pointed out that this is an illustrative example, which should provide an estimation of performance change by application of a standardized spectrum rather than measured values. The extrapolation strategy can be adapted and builds on no scientific source. Moreover, if threshold spectra are used, they have to be applied to the outer and inner rail values, respectively. Corresponding scores for PC.(1-4) derived by application of ISO 3095 [10] threshold spectra and provided coefficients in Appendix D are listed in the last four rows of Table 69. Benchmarks and loss as well as gain (indicated by a negative sign) compared to the corresponding predictions with original TDR and RR values are provided in Table 70. In general, a rather high worsening of test set predictions in either case can be observed – especially in terms of RMSE (1.6 dB and 1 dB), while R^2 decline is not as severe. Among validation sets, performance is worse regarding C2 and C4 in case of detected curve squeal application, while results in conjunction with implemented curve squeal prediction – using the recommended approach of high accuracy – show only a slight performance decline in C2. Interestingly, in the latter case, predictions can be even improved in all other validation sets – at least in RMSE. R^2 shows a decline in C3. Thus, predictive performance – especially in case of considering predicted curve squeal occurrence – is not severely affected by using threshold spectra, which leads to the assumption that RR and TDR measurements are generally recommended to implement if they are available, however prediction – at least on all considered sections – could be carried out without them by using threshold spectra stated in ISO 3095 [10]. It has to be pointed out that applicability of threshold spectra for RR and TDR might compromise predictive performance at other sections presumably foremost in case of large deviation from the limits.

[27] The model object from PCA in R is available at https://github.com/M-Ostermann/curve-noise-prediction

λ / cm	RR / dB re. 1 μm	f / Hz	TDRv / dBm^{-1}	TDRh / dBm^{-1}
25.000	13.00	100	2.00	9.71
20.000	11.00	125	2.00	6.57
16.000	9.00	160	2.00	4.45
12.500	7.00	200	2.00	3.01
10.000	4.90	250	2.00	2.04
8.000	2.90	315	2.00	1.38
6.300	0.90	400	6.00	0.94
5.000	-1.10	500	6.00	0.64
4.000	-3.20	630	6.00	0.43
3.150	-5.00	800	2.19	0.29
2.500	-5.60	1000	0.80	0.20
2.000	-6.20	1250	0.80	0.20
1.600	-6.80	1600	0.80	0.32
1.250	-7.40	2000	0.80	0.50
1.000	-8.00	2500	0.80	0.50
0.800	-8.60	3150	0.80	0.50
0.630	-9.20	4000	0.80	0.50
0.500	-9.80	5000	0.80	0.50
0.400	-10.40			
0.315	-11.00			
0.250	-11.60			
PC.1	19.12142			
PC.2	19.70928			
PC.3	-15.73047			
PC.4	-11.31255			

Table 69: Threshold spectra of rail roughness and track decay rate according to ISO 3095 [10] – extrapolated values coloured grey and corresponding scores of PC.(1-4) provided in the last four rows

	detected curve squeal occurrence				predicted curve squeal occurrence			
	RMSE	loss	R^2	loss	RMSE	loss	R^2	loss
test set	3.19	1.57	0.86	0.07	3.46	0.98	0.74	0.11
C2	4.71	0.67	0.06	0.01	4.07	0.24	0.18	-0.03
C3	5.24	0.24	0.01	0.10	4.92	-0.03	0.00	0.08
C4	2.31	0.61	0.74	-0.04	3.29	-0.35	0.44	-0.02
C2,C3,C4	3.88	0.12	0.61	0.00	3.90	-0.08	0.57	0.00

Table 70: Achieved RMSE and R^2 by including threshold spectra for RR and TDR from ISO 3095 [10] – left applied with detected curve squeal and right by using predicted curve squeal occurrence with the scope of high accuracy (refer to Table 68)

At last, it is of interest if predicted values of equivalent continuous noise level stay generally above or beneath the original ones to get an estimation of deviation in averaged values calculated over a certain time span. Since that cannot be interpreted by RMSE or R^2, predicted values by several previously discussed approaches are applied and compared as well as illustrated exemplarily. Different methods are termed and numbered as follows:

1. Level_Original: equivalent continuous noise level of pass by measurements
2. Level_pred_CS_no: predicted levels with no curve squeal occurrence (i.e. setting all three curve squeal predictors to 0 among all data points)
3. Level_pred_CS_detected: predicted levels in conjunction with detected curve squeal occurrence
4. Level_pred_CS_pred_high_sens: predicted levels using priorly predicted curve squeal occurrence by application of cut-off probabilities favouring a high sensitivity
5. Level_pred_CS_pred_high_acc: predicted levels using priorly predicted curve squeal occurrence by application of cut-off probabilities favouring a high overall classification accuracy
6. Level_pred_ISO3095_CS_detected: predicted levels using threshold spectra for TDR and RR of ISO 3095 [10] with detected curve squeal occurrence
7. Level_pred_ISO3095_CS_pred_high_sens: predicted levels using threshold spectra for TDR and RR of ISO 3095 [10] and priorly predicted curve squeal occurrence by application of cut-off probabilities favouring a high sensitivity
8. Level_pred_ISO3095_CS_pred_high_acc: predicted levels using threshold spectra for TDR and RR of ISO 3095 [10] and priorly predicted curve squeal occurrence by application of cut-off probabilities favouring a high overall classification accuracy

Exemplary time series of all listed prediction approaches as well as the originally measured continuous noise levels (red) over two days are depicted in Figure 88 (C2), Figure 89 (C3) and Figure 90 (C4). The top plot contains a comparison of approaches 1, 2, 3 and 5, the middle one illustrates methods 1, 3 and 6 and the lower figure pictures the approaches 1, 4 and 7. Deviations of the listed methods from the originally measured values are provided for each validation set with median and energetic mean among the whole measurement periods in Table 71. Since energetic mean is commonly applied in noise calculation methods and regulations, discussion is focused on that quantity.

Notably, especially in C2 and C3, very high and very low occurring levels cannot be predicted with high accuracy (refer to Figures 88 and 89). Thus, that trend causes bad results regarding RMSE and R^2 in those sections. Detailed investigation identifies train type distribution as main reason. In C2, 40 % of considered train pass bys are freight trains (train types A or B), 59 % are passenger trains and the remaining ones are part of the category "other". In general, the model predictions underestimate the level of freight trains by 3.2 dB

to 6.8 dB in energetic mean (depending on the approach), while the most common passenger train category in C2 (train type E – 56 % of all passenger train pass bys) is predicted with an energetic mean deviation between original measurement and prediction approaches in a range from -1.2 dB to 2.1 dB (refer to Table 72). In C3, 55 % of data points are freight traffic (train types A or B), 37 % of train pass bys are passenger trains and the remaining ones are part of the category "other". Model predictions underestimate freight train pass by levels by 2.6 dB to 5.9 dB in energetic mean (depending on the approach), while the energetic mean of the most frequent passenger train category in C3 (train type D – 68 % of all passenger train pass bys) is overestimated by 2.4 dB to 5.8 dB (refer to Table 73). Thus, the partly high overall accuracy of the predictions regarding averaged values among the whole measurement period (see Table 71) is reached by a balanced underestimation of freight traffic and overestimation of passenger traffic. Moreover, in terms of energetic mean, the underestimation of freight trains leads to generally lower prediction values among the whole measurement period compared to the monitored ones. In C4, predicted levels show good accordance to the original measurement – even in conjunction with occurring variations (illustrated in Figure 90). Level predictions of the most occurring train category (train type G – 80 % of all train pass bys) deviate from the original measurement in a range from -1.0 dB to 3.0 dB in energetic mean (depending on the chosen approach), whereas the second most train category (train type F – 14 % of all train pass bys) shows overestimations in energetic mean in range from 0.9 dB to 3.5 dB (refer to Table 74). Freight traffic has only a share of 4 % in C4 and shows similar trends as observed in C2 and C3 – partly high underestimations in level predictions (not separately depicted). Despite low freight train pass by percentage, their level magnitudes exceed those of passenger trains significantly (refer to Figure 41), which causes an underestimation in energetic mean among the whole measurement period in some approaches, while regarding median values no method shows a lower value compared to the original measurements.

Between prediction approaches, it can be concluded that deviation between usage of predicted curve squeal occurrence with the scope of high overall classification accuracy (approach 5) and detected curve squeal occurrence by the algorithm (approach 3) is minor ($\leq$ 0.5 dB in energetic mean without splitting into individual train types). Energetic mean differences of approach 5 compared to the original measurement are in a range from -0.5 dB to -1.1 dB. Thus, the recommended cut-off probability thresholds in conjunction with equivalent continuous noise level prediction are further justified. Moreover, the applicability of the elaborated models for predicting curve squeal occurrence in Subsections 6.4.1, 6.4.2 and 6.4.3 is further confirmed. Comparison of approaches 2 and 3 shows that exclusion of curve squeal causes a decrease in energetic mean values among predictions in a range from 0.3 dB to 1.6 dB. Especially in C2 and C3, decrease magnitudes are $\leq$ 0.5 dB, which indicates that curve squeal consideration looses importance in those sections. Main reasons behind are presumably the train traffic mixture with significantly more freight traffic percentage and larger curve radii compared to C4, which leads to decreased probability rates of curve squeal in general. Generally speaking, the former is caused by higher rolling noise of freight trains, which leads to a low raise in overall sound emission compared to passenger train categories despite curve squeal occurrence. Using cut-off probabilities with the scope of high sensitivity (approach 4)

	median / dB			energetic mean / dB		
outcome	**C2**	**C3**	**C4**	**C2**	**C3**	**C4**
1	87.6	88.9	78.4	89.4	90.0	83.3
2	+0.5	-1.5	+0.9	-1.4	-2.1	-2.2
3	+0.9	-0.8	+1.6	-1.1	-1.6	-0.6
4	+1.6	+0.8	+5.2	-0.6	-0.6	+1.1
5	+1.2	-0.2	+1.4	-0.8	-1.1	-0.5
6	-1.7	-3.3	+0.2	-3.9	-3.9	-1.4
7	-1.2	-0.1	+4.6	-2.5	-1.3	+0.9
8	-1.5	-2.1	0.0	-3.1	-2.7	-1.3

Table 71: Deviation between different prediction approaches and originally measured values (outcome 1) in terms of median and energetic mean among the whole measurement period

leads to a slight overestimation in energetic mean in C4 (1.1 dB), while C2 and C3 can be predicted with -0.6 dB energetic mean deviation. Therefore, approach 4 – aiming for higher sensitivity in curve squeal prediction – also seems to represent a reasonable choice in terms of energetic mean estimation on the safe side. However, the probably high false positive rates in curve squeal detection and their impact on possible overestimations of level predictions have to be borne in mind. In terms of using threshold spectra in accordance with ISO 3095 [10], predicted levels show a general underestimation in energetic mean (except approach 7 in C4). That is reasonable, since all sections show an elevated rail roughness in some wavelength bands and/or a partly lower track decay rate compared to the threshold spectra.

To conclude, considering prediction accuracy (RMSE – refer to Table 68) as well as median and energetic mean prediction results among the whole measurement period, approach 5 – curve squeal prediction with the scope of high accuracy – shows the best benchmarks. Regarding approach 8 – same method, but using ISO 3095 [10] threshold spectra –, averaged levels are underestimated in both median and energetic mean. Thus, in curves, where rail roughness is in general commonly worse compared to ISO 3095 [10] limits, a measurement of TDR and RR is recommended. Otherwise, the prediction result could be raised by an additive factor to account track properties deviating from the ISO 3095 [10] limits. The magnitude has to be determined by further evaluations on appropriate sections.

Altogether, it has to be pointed out that train type distribution – especially ratio between freight and passenger trains – seems to have a distinct impact on predictive accuracy of averaged values. Using the recommended approach with the scope of high accuracy in curve squeal prediction (approach 5), averaged levels of the most frequent passenger train category are overestimated by 2.0 dB (in C2) and 5.4 dB (in C3) in energetic mean, whereas freight train levels are underestimated by 3.5 dB (in C2) and 3.2 dB (in C3) in energetic mean. Due to the balanced ratio of freight and passenger trains in both sections, the overall prediction results have a high accuracy in energetic mean of -1.1 dB (C3) and -0.8 dB (C2) deviation between prediction and original measurement. Thus, assuming the same boundary conditions as in C2 and C3, the model would significantly underestimate the energetic mean level in case of 100 %

Figure 88: Exemplary time series plots depicting two days of originally measured equivalent continuous noise levels versus different prediction outcomes in C2

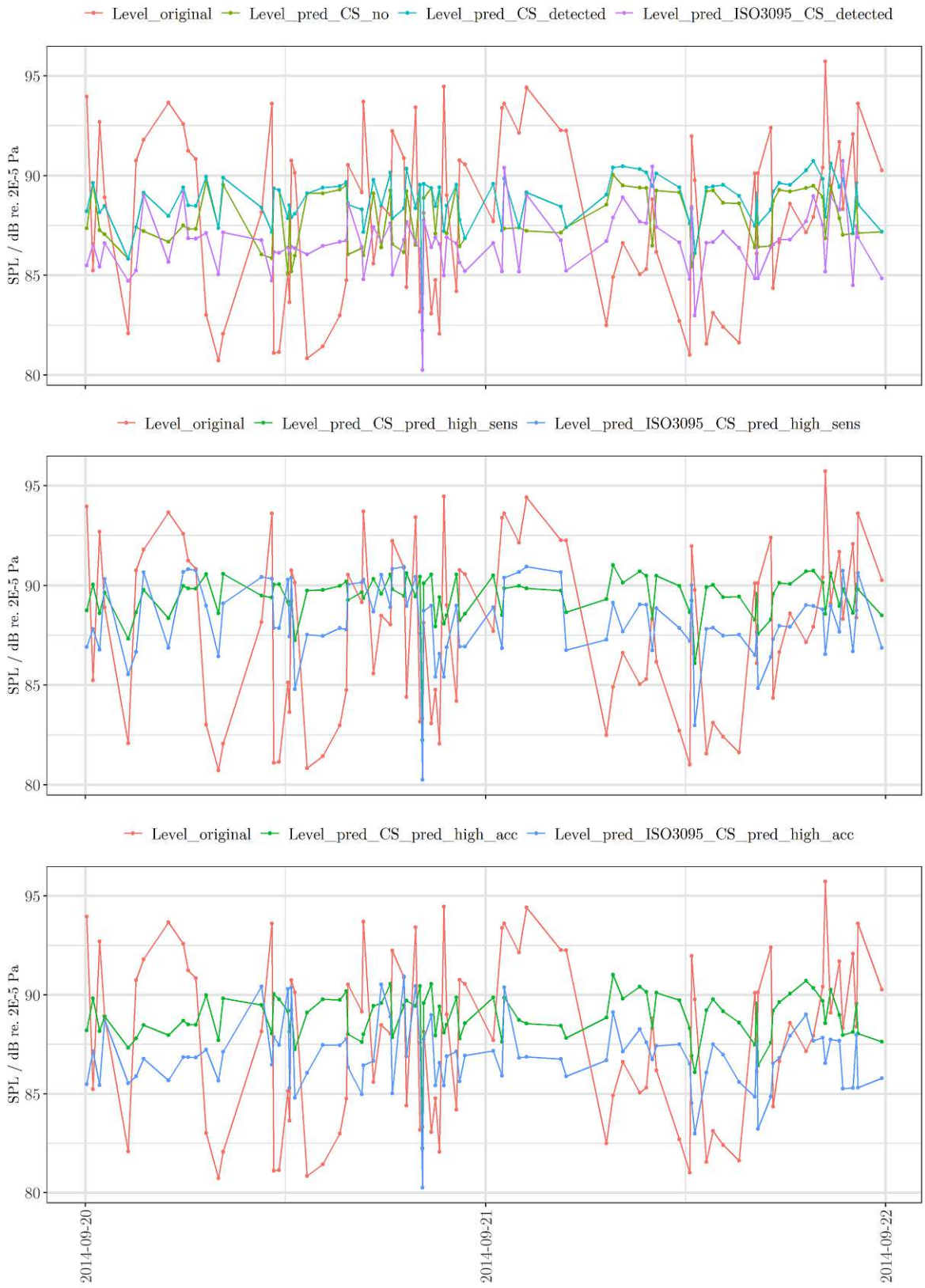

Figure 89: Exemplary time series plots depicting two days of originally measured equivalent continuous noise levels versus different prediction outcomes in C3

Figure 90: Exemplary time series plots depicting two days of originally measured equivalent continuous noise levels versus different prediction outcomes in C4

C2	median / dB		energetic mean / dB	
outcome	**A,B**	**E**	**A,B**	**E**
1	90.9	86.9	91.6	87.6
2	-4.3	+2.5	-5.0	+1.8
3	-3.8	+2.6	-4.4	+1.9
4	-2.7	+2.7	-3.2	+2.1
5	-2.8	+2.6	-3.5	+2.0
6	-6.2	-0.6	-6.8	-1.2
7	-4.4	-0.2	-4.1	-0.5
8	-4.6	-0.7	-5.0	-0.9

Table 72: Deviation between different prediction approaches and originally measured values (outcome 1) in terms of median and energetic mean among the whole measurement period in C2 separated in freight trains (category A and B) and the most common passenger train type E

C3	median / dB		energetic mean / dB	
outcome	**A,B**	**D**	**A,B**	**D**
1	91.6	83.2	91.8	84.3
2	-4.5	+6.1	-4.9	+4.9
3	-3.9	+6.2	-4.1	+5.1
4	-2.2	+7.0	-2.6	+5.8
5	-3.1	+6.6	-3.2	+5.4
6	-6.3	+3.4	-5.9	+2.4
7	-1.3	+5.6	-2.6	+4.0
8	-4.8	+4.3	-4.3	+3.2

Table 73: Deviation between different prediction approaches and originally measured values (outcome 1) in terms of median and energetic mean among the whole measurement period in C3 separated in freight trains (category A and B) and the most common passenger train type D

C4	median / dB		energetic mean / dB	
outcome	**F**	**G**	**F**	**G**
1	83.1	77.7	83.4	80.1
2	+1.9	+1.4	+1.6	-1.0
3	+2.6	+1.8	+2.3	+0.9
4	+4.0	+4.1	+3.5	+3.0
5	+3.4	+1.8	+2.8	+0.3
6	+0.8	+0.2	+0.9	+0.1
7	+2.6	+3.4	+2.8	+2.8
8	+1.5	+0.2	+1.6	-0.8

Table 74: Deviation between different prediction approaches and originally measured values (outcome 1) in terms of median and energetic mean among the whole measurement period in C4 separated in the two most common train types F and G

freight train operation and otherwise, in case of 100 % passenger traffic, overestimations would result. The bulk of monitored freight coaches in all sections are operated with cast-iron block brakes. Legislative of the EU forbids the operation of coaches with that braking system on certain routes, which have to be defined by each member state, by the end of 2024 [206]. Due to the underestimation of those by the model, it is possible that predictions are getting more accurate for freight trains consisting only of coaches with disc brakes or composite brake pads. Thoughts about other limitations of the elaborated model are provided in Subsection 6.6. Energetic mean estimations of each validation set among the whole measurement period with curve squeal prediction using cut-off probabilities with the scope of high accuracy lead to differences to originally measured values in a range from -1.1 dB to -0.5 dB. Thus, results are promising and further validation on other datasets is recommended.

6.5 Summarised findings

In the following, the main findings of the previous subsections are briefly summarised. The subsection is outlined in two main parts – discussing findings about curve squeal and equivalent continuous noise level. The former is separated in discussion on frequency of occurrence, peak level and relative occurrence time. Depiction is focused on general trends, which are presumably also indicated in other datasets involving different sections, train type distributions and climatic conditions.

6.5.1 Curve squeal

Summary on curve squeal findings is outlined in comparing all three distinguished curve squeal types – flanging, squeal and HF squeal noise – in conjunction with their frequency of occurrence, peak levels and relative occurrence times. For the first, discussion on most important influence factors and their magnitude is followed by a brief overview of derived prediction models.

Frequency of occurrence

Main findings on frequency of occurrence are depicted in Table 75. All depicted increase or decrease values are additive factors and not proportional. In the first three data rows, train types indicating a similar behaviour in terms of particular curve squeal types are stated. For flanging and HF squeal noise three different categories are defined – representing a high (TT 1), mediocre (TT 2) and low frequency of occurrence (TT 3). In general, freight trains show higher values compared to passenger trains for both mentioned noise sources, thus they are part of the first group. For the third type – squeal noise –, only two categories (TT 2 and TT 3) are used for grouping as freight traffic shows not much deviation to certain passenger train types. Interestingly, train type "other", which states a random category of all other operated trains (e.g. single locos or operation with different coaches than standardized for a certain train type) shows very low frequency of occurrence regarding all curve squeal types. The second train type, which is always categorised in TT 3, states train type G. Remaining categories show mediocre levels (except train type E regarding flanging noise). The following

rows allow an estimation of probability magnitude among the whole dataset as well as in terms of defined train type groups. The latter depiction is split in narrow curve radii ($\leq$ 256 m) and larger ones ($\geq$ 310 m), since deviation is partly significant. This complies with past research (e.g. [5]). The depicted range covers relative frequency values of subsets – distinguished in sections and train types – in the particularly separated groups. Partly large overlapping between ranges is caused by deviating behaviour in different radius categories. However, the separation is only chosen for depiction and is not included elsewhere. Notably, considering the whole dataset, flanging noise shows a significantly higher frequency of occurrence (34 %) compared to squeal noise (13 %) and HF squeal noise indicates a slightly lower probability (12 %) than squeal noise. However, it has to be pointed out that those magnitudes are highly influenced by distribution of train types and curve radii in the dataset and can be completely different on sections with deviating boundary conditions. Deviation magnitudes between groups TT 1 and TT 2 decrease with increasing radii, while TT 3 shows partly zero incidences in the larger radius category. Especially in the latter, partly weak statistical significance has to be pointed out. Double unit operation increases frequency of occurrence significantly for all curve squeal types in magnitudes of 12 % to 22 % with the simple cause that the same train type operates with doubled axle quantity. Nevertheless, that issue has to be accounted in curve squeal prediction tasks. Train orientation in single units (distinguishing pushed or pulled operation) as well as double units can significantly impact curve squeal occurrence with an observed trend to cause higher frequency of occurrence in the pushed state. Regarding double units, that finding is built on evaluation of train type F, which is a trainset with a power coach and two unpowered wagons. If two trainsets are operated together, the position of the power unit can impact the frequency of occurrence – in general, the least probability in terms of flanging and squeal noise is reached in case of both power units pulling their trainsets. In terms of flanging noise, the worst case states if both power units are in the middle and regarding squeal noise, if one power coach is at the end and one in front. HF squeal noise shows no impact of train orientation in conjunction with single units and a weaker maximum increase for double units compared to the other two curve squeal types. One of the most important findings in the thesis refers to the connection between curve squeal occurrence and relative humidity – especially relative rail humidity. Contrary to past research [37, 62], where an increase of squeal noise probability is evaluated among the whole range of relative humidity, all three distinguished curve squeal types show a peak frequency of occurrence in a narrow range to each other between 72 % and 75 % relative rail humidity and decreasing amplitudes in conjunction with lower as well as higher values (illustrated in Figure 91). While the former trend is complying to past research – with increasing relative humidity critical creepage value is lowered [37] –, the latter might be caused by beginning dew formation on the rail head. However, further research in terms of laboratory investigation under controlled climatic conditions are suggested. Moreover, past research is focused on squeal noise, which states only one out of three distinguished curve squeal types in the present thesis. Despite underlying mechanisms might deviate significantly, the connection to relative rail humidity is strongly indicated in all three separate evaluations. It also explains alternating frequency of occurrence among months of the year and even in

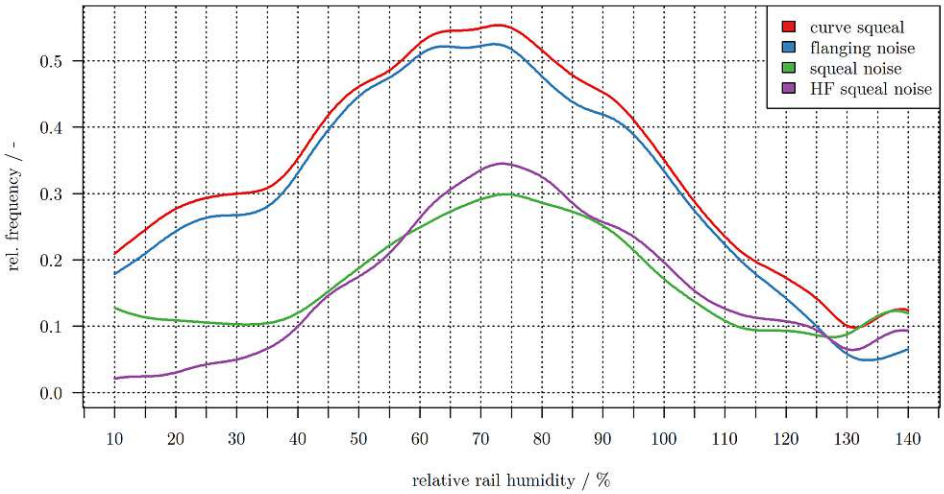

Figure 91: Relative frequency of all three distinguished curve squeal type as well as overall curve squeal among the range of relative rail humidity – data from the whole dataset

daytime. Regarding other impact of environmental conditions, rain nearly suppresses curve squeal completely (remaining relative frequency of occurrence 1 % to 2 % in the whole dataset), which complies with past research. Moreover, which is not discussed broadly in past investigations, frost occurrence – especially conditions, where hoarfrost is likely – shows similar impact compared to rain with remaining probabilities of 1 % to 3 % (hoarfrost) and 4 % to 9 % regarding frost conditions. The latter are assumed if rail temperature drops beneath the freezing point and liquid water due to rain or dew is present on the rail head. Dew formation is modelled separately (refer to Subsection 5.5.1). Dew itself shows also a declining impact on frequency of occurrence, however reduction in relative frequency of occurrence is not as distinct compared to rain or frost.

To conclude, in three separately carried out evaluations for frequency of occurrence regarding distinguished curve squeal types, the same parameters show the best correlations to the outcome. Altogether, the following indicators are evaluated as having the majority of influence on relative frequency of occurrence:

- Train type – grouped in two (squeal noise) or three groups of similar probability (flanging and HF squeal noise)
- operation in double units compared to single ones
- RSI (represents curve radius and wheel diameter)
- lateral acceleration (leads to inclusion of radius and cant as well)
- relative rail humidity
- rain

parameter	comment	unit	flanging noise	squeal noise	HF squeal noise
similar train types	high prob. (TT 1)	–	A,B	–	A,B
	mediocre prob. (TT 2)	–	C,D,F	A,B,C,D,E,F	C,D,E,F
	low prob. (TT 3)	–	E,G,other	G,other	G,other
rel. frequency of occurrence	whole dataset	%	34	13	12
	TT 1 ($\leq$ 256 m)	%	67 - 79	–	22 - 54
	TT 1 ($\geq$ 310 m)	%	33 - 56	–	7 - 15
	TT 2 ($\leq$ 256 m)	%	32 - 56	13 - 31	10 - 23
	TT 2 ($\geq$ 310 m)	%	6 - 45	0 - 33	0 - 17
	TT 3 ($\leq$ 256 m)	%	16 - 28	3 - 10	0 - 11
	TT 3 ($\geq$ 310 m)	%	0 - 21	0	0
double unit operation	max. increase	%	20	22	12
train orientation	max. increase (single unit)	%	6	8	–
	max. increase (double unit)	%	5	6	3
relative rail humidity	critical range	%	62 - 77	60 - 90	59 - 83
	prob. in critical range	%	$>$ 50	$>$ 25	$>$ 30
	peak	%	52 at 72	30 at 75	34 at 74
rain	whole dataset decrease (remaining prob.)	%	31 (1)	15 (2)	16 (1)
dew occurrence		%	12 (18)	7 (9)	5 (10)
hoarfrost		%	28 (3)	16 (1)	14 (2)
frost		%	22 (9)	12 (4)	16 (6)

Table 75: Observed main tendencies and impacts regarding relative frequency of occurrence of all three distinguished curve squeal types

- dew presence
- frost conditions (frost or hoarfrost likely)

RSI and lateral acceleration show a positive correlation to relative frequency of occurrence and allow an implicit consideration of wheel diameter as well as curve radius and cant. Both show general trends among considered sections, which seem to be sparsely influenced by data point distribution. Thus, they are chosen for inclusion in development of a generally applicable prediction model.

Building on the formulated main influence parameters, classification models to predict flanging, squeal and HF squeal noise occurrence are developed. Model training is done on 65 % data points of C1 and C5. 10 % of the remaining data points in those sections are held out as independent evaluation set to investigate alternative cut-off probability thresholds with the scope of improving sensitivity. Other data points in C1 and C5 are used as test set to validate model performance in similar boundary conditions. The remaining sections – C2, C3, C4 – are used as validation sets and represent an indicator, how the model performs on new datasets involving other sections, deviation train type distribution and completely different climatic conditions. Thus, focus lies on achieved benchmarks in the latter sections to estimate general applicability of the model to other datasets. For flanging noise prediction, flexible discriminant analysis (FDA) performs best. Regarding both squeal noise prediction tasks, the Neural Network method indicates top achieved benchmarks. Basic properties about the models are provided in Table 76. Cut-off probability examination reveals that the default value of 50 % to decide whether a train pass by is predicted as containing a particular curve squeal type is not suitable – especially in case of squeal and HF squeal noise prediction. Evaluation results in recommendation of two cut-off probability thresholds depending on the modelling scope. If overall classification accuracy is more important than high curve squeal detection rate (i.e. sensitivity), threshold values of 37.1 % for flanging noise, 17.5 % for squeal noise and 17 % are suggested. If scope lies in the opposite and lower overall classification accuracy as well as higher false positive rates are acceptable, cut-off probabilities of 25 % for flanging noise and 10 % for both squeal and HF squeal noise are recommended. Since the result of the model is a probability value for each train pass by, thresholds could be changed in application. However, it has to be borne in mind that they are chosen on behalf of an independent evaluation set, which is highly recommended statistically [113]. Thus, a change is generally not advised and should only be thought of in case of very implausible outcome rates for both stated cut-off probability sets. Despite the significantly reduced predictor set and the presumably complex underlying mechanism in terms of curve squeal occurrence, the developed models are capable of reaching curve squeal detection rates among the whole validation set (C2, C3 and C4) of 68 % to 84 % in conjunction with an overall classification accuracy of 55 % to 68 %. If the latter benchmark is the main modelling scope, values are improvable to 67 % to 83 % on the whole validation set, however causing a drop of sensitivity to values between 27 % and 66 %. Evaluation of predictor importance of environmental parameters compared to other included variables leads to the conclusion that both predictor groups – environmental and other – are

parameter		flanging noise			squeal noise			HF squeal noise		
statistical method		FDA			Neural Network			Neural Network		
standardization required		no			yes			yes		
cut-off prob. high acc. / %		37.1			17.5			17.0		
cut-off prob. high sens. / %		25.0			10.0			10.0		
benchmarks / %	dataset	Acc	Sens	Spec	Acc	Sens	Spec	Acc	Sens	Spec
high accuracy	test	68	56	74	72	59	74	66	68	66
	C2	68	67	69	86	28	87	67	64	68
	C3	59	88	35	66	30	69	49	59	48
	C4	70	53	78	88	25	96	78	32	85
	C2,C3,C4	67	66	68	83	27	88	69	44	72
high sensitivity	test	59	83	47	53	81	48	53	88	47
	C2	60	75	53	74	44	75	61	72	61
	C3	52	98	14	42	80	39	35	83	27
	C4	61	83	50	74	68	74	59	77	56
	C2,C3,C4	59	84	45	68	68	68	55	78	53

Table 76: Information on the built final classification models to predict flanging, squeal and HF squeal occurrence

necessary to include in a prediction model.

Peak level

Regarding peak levels, which state a maximum short-time averaged (64 ms) third-octave band level of all three distinguished curve squeal types, observed trends are provided in Table 77. The first row contains median levels among the whole dataset. Notably, squeal and HF squeal noise peak levels exceed those of flanging noise by 6 dB and 8 dB depending on squeal noise type. That complies with past research regarding higher amplitudes of squeal noise (e.g. [5]). However, it has to be pointed out that those magnitudes are highly influenced by distribution of train types and curve radii in the dataset and can be completely different on sections with deviating boundary conditions. Since distinction in multiple train type groups with similar behaviour is only evaluated as reasonable regarding squeal noise, depiction is split into passenger and freight traffic. Moreover, ranges in narrow radii ($\leq$ 256 m) and larger ones ($\geq$ 310 m) are illustrated. The depicted range covers median values of subsets – distinguished in sections and train types. In terms of flanging noise, magnitudes of peak levels regarding freight and passenger traffic are in the same range – both showing a decreasing trend in conjunction with larger radii. HF squeal noise indicates the same except it is unsure if passenger trains show decreasing levels in larger radii, too. Regarding squeal noise, freight trains show significantly higher peaks in lower radii compared to passenger trains and a decreasing trend in larger radii – however, even there, they are higher compared to the other two curve squeal types. A statement of passenger trains in larger radii is not possible due to lack of sufficient data points. Train orientation as well as double unit

parameter	comment	unit	flanging noise	squeal noise	HF squeal noise
median levels	whole dataset	dB	86	92	94
	freight ($\leq$ 256 m)	dB	86 - 92	99 - 105	92 - 95
	freight ($\geq$ 310 m)	dB	81 - 83	96	88 - 92
	passenger ($\leq$ 256 m)	dB	85 - 95	89 - 93	90 - 97
	passenger ($\geq$ 310 m)	dB	78 - 92	–	95
dew	median increase (raised median)	dB	2 (88)	0	0
hoarfrost		dB	10 (96)	3 (95)	4 (98)
frost		dB	6 (92)	0	3 (97)

Table 77: Observed main tendencies and impacts regarding peak levels of all three distinguished curve squeal types

operation show a negligible impact (deviations mostly $<$ 1 dB) on peak level magnitudes. Thus, depiction is skipped. Regarding environmental conditions, only dew, hoarfrost and frost impact can be estimated since rain reduce frequency of occurrence greatly and thus no sufficient data points are available. Interestingly, while investigated frost or wetness sources reduce frequency of occurrence, partly raised peak levels are evaluated. Dew occurrence leads to 2 dB median increase of flanging noise peak level among the whole dataset. Squeal and HF squeal peak levels are not affected by the latter. Regarding hoarfrost and frost occurrence, raises are higher compared to dew impact with magnitudes between 2 dB and 10 dB. The only exception states squeal noise, where no effect regarding frost occurrence is shown. Again, evaluations among the whole dataset are influenced by distribution of data points. However, subsetting to exclude train type and section influence leads to the same general trends – magnitude depending on the train type. In general, it has to be pointed out that statistical significance for stated values is partly weak due to a significantly reduced frequency of occurrence in conjunction with larger radii and certain environmental conditions.

Results of the evaluation on most important influence factors in conjunction with peak levels are depicted in Table 78. Contrary to frequency of occurrence, findings regarding importance deviate between all three distinguished curve squeal types. Three parameters are listed in all evaluations – temperature, absolute humidity and frost conditions. For the former, mostly rail temperature shows the best correlation to the outcome. Among absolute humidity indicators, specific humidity or mixing ratio (not much deviation between those) are most important. They include absolute air pressure in derivation, which makes them comparable on different altitudes. Furthermore, it shows hardly any difference if they are calculated with air or rail temperature. Inclusion of a relative humidity indicator indicates distinct daytime influence, which is especially present regarding flanging noise peak levels.

parameter	flanging noise	squeal noise	HF squeal noise
train type groups		✓	
curve radius	✓		✓
RSI		✓	
rail temperature	✓	✓	✓
specific humidity	✓	✓	✓
relative (rail) humidity	✓	✓	
wet conditions	✓		
frost conditions	✓	✓	✓

Table 78: Most important predictors regarding impact on peak level of all three distinguished curve squeal types

In general, temperature and absolute humidity show a negative correlation to the outcome in all three evaluations, while relative humidity shows a positive one. Curve radius is also accounted in every evaluation. Regarding flanging and HF squeal noise, it is considered explicitly, while in terms of squeal noise it is included implicitly with RSI. Additionally, squeal noise evaluation is also the only one, where distinction of similar train type groups seems reasonable. Thus, flanging and HF squeal noise are weakly influenced by train type, which is stated in the previous paragraph. Wet conditions (either rain or dew) show only an impact in terms of flanging noise.

Due to poor benchmarks of multivariate statistical analysis and the scope of practical applicability, development of a prediction model for curve squeal peak levels is not seen as meaningful. For rough estimation of peak levels, median values from Figures 92, 94 and 96 for similar train types and radii can be applied. Alternatively, presented ranges for passenger and freight traffic in Table 77 can be used.

Relative occurrence time

Tendencies of relative occurrence time regarding all three distinguished curve squeal types are summarised in Table 79. All depicted increase or decrease values are additive factors and not proportional. Notably, median values are nearly halved regarding both squeal noise types compared to flanging noise. However, it has to be pointed out that those magnitudes are highly influenced by distribution of train types and curve radii in the dataset and can be completely different on sections with deviating boundary conditions. Train type grouping is done in two out of three evaluations by separating passenger and freight traffic and in the remaining one not carried out at all. Thus, depiction is generally separated in passenger and freight traffic as well as trends in narrow curve radii ($\leq$ 256 m) and larger ones ($\geq$ 310 m). The depicted ranges cover median values of subsets – distinguished in sections and train types. In terms of relative occurrence time regarding flanging noise, depiction of train type

category "other" is not included, since it shows significantly higher values compared to passenger trains and thus is seen as an outlier group. 100 % relative occurrence time is observable in conjunction with flanging noise and freight trains. That trend is unaffected by different curve radii. Passenger trains in the same evaluation show decreased values in larger radii and overall also significantly higher relative occurrence time compared to both squeal noise examinations. Deviations between passenger and freight traffic in terms of relative occurrence time regarding squeal noise are not as distinct – showing a higher range, however not always larger values. Behaviour in larger radii can only be stated for freight trains, which show very low relative occurrence time, due to lack of sufficient data points. Relative occurrence time of HF squeal noise indicates larger values for passenger traffic compared to freight trains (about doubled). That trend continues in larger radii, however both groups show declining magnitudes. Altogether, except freight trains in conjunction with flanging noise, all evaluations indicate a decline in relative occurrence time among larger radii ($\geq$ 310 m) compared to narrower ones ($\leq$ 256 m). Contrary to peak levels, double unit operation as well as train orientation partly indicate an impact. For the former, decreasing magnitudes of 14 % to 22 % are observed. Thus, taking the doubled train length into account, absolute occurrence time is roughly equal between single and double units. Train orientation influence of single units – distinguishing between pulled and pushed operation – is only detected in terms of relative occurrence time regarding flanging noise with significantly increased magnitudes in the pushed state. Regarding double units, findings are built on evaluation of train type F, which is a trainset with a power coach and two unpowered wagons. If two trainsets are operated together, the position of the power unit can impact the frequency of occurrence in magnitudes of 4 % to 8 % for flanging noise and HF squeal noise, while regarding relative occurrence time of the third type, no deviations are indicated. Complying to observation regarding peak levels, relative occurrence time is also increased in conjunction with hoarfrost and frost conditions (raises of up to 13 %). Dew occurrence shows a controversial trend among curve squeal types – indicating a huge raise for flanging noise as well as a slight increase for squeal noise and a decrease for HF squeal noise. Again, evaluations among the whole dataset are influenced by distribution of data points. Subsetting to exclude train type and section influence reveals that impact of dew on relative occurrence time of both squeal noise types is negligible (mostly $\pm$ 1 %). Moreover, the magnitude of increase regarding flanging noise is significantly lower in single subsets (up to 15 %). General trends of frost and hoarfrost comply with subset evaluations – magnitudes depending on train type. Overall, it has to be pointed out that statistical significance for stated values is partly weak due to a significantly reduced frequency of occurrence in conjunction with larger radii. Moreover, due to high interquartile ranges in the whole dataset as well as among subsets, a general uncertainty is present regarding the presented trends and their magnitude.

Results of most important influence parameters in conjunction with relative time occurrence of all three distinguished curve squeal types reveal that environmental conditions are important (provided in Table 80). The list includes temperature, absolute, relative humidity and frost conditions in every evaluation. The former two show generally negative correlations

parameter	comment	unit	flanging noise	squeal noise	HF squeal noise
median values	whole dataset	%	45	25	23
	freight ($\leq$ 256 m)	%	100	8 - 43	10 - 14
	freight ($\geq$ 310 m)	%	100	8	5 - 6
	passenger ($\leq$ 256 m)	%	28 - 55	19 - 29	16 - 28
	passenger ($\geq$ 310 m)	%	15 - 37	–	12
double unit operation	max. median decrease	%	20	22	14
train orientation	max. median increase (single unit)	%	12	–	–
	max. median increase (double unit)	%	8	–	4
dew	median increase (raised median)	%	55 (100)	2 (27)	-7 (16)
hoarfrost		%	8 (53)	12 (37)	13 (36)
frost		%	9 (54)	3 (28)	10 (33)

Table 79: Observed main tendencies and impacts regarding relative occurrence time of all three distinguished curve squeal types

parameter	flanging noise	squeal noise	HF squeal noise
train type groups	✓		✓
train orientation	✓		
double unit	✓	✓	✓
curve radius		✓	✓
rail temperature	✓	✓	✓
specific humidity	✓	✓	✓
relative (rail) humidity	✓	✓	✓
wet conditions	✓		
frost conditions	✓	✓	✓

Table 80: Most important predictors regarding impact on relative occurrence time of all three distinguished curve squeal types

to the outcome, whereas relative humidity indicates a positive one. Frost impact is already discussed in the previous paragraph. Among temperature values, rail temperature shows mostly the best performance. Absolute humidity indicators are listed as specific humidity or mixing ratio – although not much deviation between them is indicated. They include absolute air pressure in their derivation, which leads to comparability at different altitudes. Moreover, hardly any difference in performance between using air or rail temperature in their calculation is observed. Among relative humidity indicators, difference between rail temperature and dew point is listed in two of three evaluation, which is highly correlated with relative rail humidity, however its values show a negative correlation to the outcome. Train type group distinction is seen reasonable regarding flanging and HF squeal noise due to deviations between freight and passenger traffic. Other listed impacts are already discussed in the previous paragraph.

Due to high fluctuations in the dataset (resulting in high RMSE values in multivariate statistical approaches) and pursuing the scope of constructing a practical model, heuristic approaches to estimate relative time occurrence are developed – only including curve radius and partly passenger and freight train distinction (refer to Table 81). It has to be pointed out that those state tries to model a simplified relationship empirically, which needs further radii for comparison and validation.

flanging noise		
radii / m	**other / %**	**passenger / %**
≤ 310	100	45
$310 < R < 440$	100	$109.385 - 0.2077 \cdot R$
≥ 440	100	18

squeal noise	
radii / m	**all train types / %**
≤ 256	30
$256 < R < 310$	$124.815 - 0.3704 \cdot R$
≥ 310	10

HF squeal noise			
radii / m	**freight / %**	**radii / m**	**other / %**
≤ 230	15	≤ 256	28
$230 < R < 440$	$25.952 - 0.0476 \cdot R$	$256 < R < 310$	$103.852 - 0.2963 \cdot R$
≥ 440	5	≥ 310	12

Table 81: Most important predictors regarding impact on relative occurrence time of all three distinguished curve squeal types

6.5.2 Equivalent continuous noise level

Main findings in the evaluation are summarised in Table 82. The whole examination is focused on train pass bys without any detected curve squeal type. Stated quantities are always median values or median difference magnitudes. The first row states three groups of train types, for which similar behaviour in terms of equivalent continuous noise level are observed. The first group (TT 1) consists of freight trains (high levels), the second one of passenger trains as well as other trains (e.g. single locos) – mediocre levels – and the third category includes one train type, which shows a very low basic rolling noise level. Depiction of the random train category "other" is skipped in presented ranges in the table due to indicating low levels in narrow radii (≤ 256 m) and higher ones in larger radii (≥ 310 m). This would compromise the stated ranges. The further rows contain median levels distinguished in narrow radii and larger ones. Moreover, worst case (highest among subsets) median alterations in case of curve squeal occurrence are provided in brackets – indicating from left to right flanging/squeal/HF squeal noise. Since peak magnitudes of flanging noise are significantly lower compared to both types of squeal noise (refer to Table 77), evaluation of flanging noise impact is done by excluding train pass bys, which contain in addition to flanging noise also any type of squeal noise. In the other direction, for squeal and HF squeal noise evaluation, flanging noise occurrence is tolerated because otherwise too few data points would be left for evaluation. However, in each squeal noise evaluation, parallel occurrence of the other squeal noise type is excluded. Since magnitudes are highly influenced by distribution of train types and curve radii in the dataset, depiction of an overall median value is uninformative. The stated range covers median values of subsets – distinguished in sections and train types – in the particularly separated groups. An important finding is that

regarding freight traffic, hardly any raise in median levels due to curve squeal occurrence is shown. In terms of flanging and HF squeal occurrence, a slight decrease in median values is indicated. However, that is only reasoned by general fluctuations in monitored freight train pass by levels in the dataset. Thus, train type distinction regarding inclusion of curve squeal occurrence in noise prediction models is essential (also pointed out in [9]). Secondly, level seems not influenced by curve radius in that category. In terms of TT 2, slight raises are shown due to curve squeal occurrence (up to 2 dB in narrow radii and up to 4 dB in larger radii) and no influence of higher radii is visible. In the third group (TT 3), median values are significantly raised by curve squeal occurrence (up to 6 dB), however radius influence is also not indicated. In general, deviating track parameters can have significant influence on equivalent continuous noise level. Especially in curves, partly significant rail corrugation is common. The stated values in narrow radii ($\leq$ 256 m) cover sections with partly large deviations in RR and TDR compared to each other. Thus, the range as well as worst case curve squeal impact covers a wider range of track properties. Among higher radii, better track properties (mostly ISO 3095 [10] limit satisfied) are present. Thus, results in worse conditions cannot be covered. Regarding environmental impact, an increase in terms of dew as well as frost/hoarfrost occurrence of up to 2 dB is observed. The latter value is derived from the carried out empirical examination, which includes a speed normalisation to 60 kmh^{-1} (refer to Subsection 6.2). The trend is indicated by all statistical significant subsets – involving three different passenger train types and two sections with deviating properties. For freight traffic, no general statement can be given due to lack of sufficient data points. Detailed spectral analysis identifies section and train type independent raises around 160 Hz and 400 Hz/500 Hz third-octave bands in magnitudes of up to 2 dB for both dew and frost occurrence separately. Both frequencies are among the highest in rolling noise emission of the investigated train types. Hence, their impact on overall equivalent continuous noise level is reasonable. It has to be pointed out that intentionally no A-weighting is applied. Thus, if that measure is taken in propagation and immission calculation, the increase around 160 Hz is significantly weakened ($>$ 10 dB reduction). However, the latter raise (at 400 Hz/500 Hz) is not heavily impacted. Moreover, third-octave band level can alter up to 4 dB depending on temperature, which is partly connectable to variations in track decay rate if a positive correlation is indicated. However, the opposite case is also present in some third-octave bands.

Building on carried out evaluation on equivalent continuous noise level, the following parameters are evaluated as having a major influence:

- track decay rate
- rail roughness
- sleeper material (only evaluated wood and concrete)
- train type distinguished in three categories depending on their basic rolling noise level – high (type A and B), mediocre (type C, D, E, F, other), low (type G)
- mean speed

parameter	comment	unit	equivalent continous noise level		
similar train types		–	A,B (TT 1)	C,D,E,F (TT 2)	G (TT 3)
median levels*	≤ 256 m	dB	90 - 92 (-1/0/-1)	83 - 91 (+1/+2/+2)	78 - 83 (+3/+6/+6)
	≥ 310 m	dB	91 (+1/+1/0)	82 - 87 (+1/–/+4)	80 (–/–/–)
dew occurrence	max. median increase	dB	–	2	2
hoarfrost		dB	–	2	2
frost		dB	–	2	2

* worst case alteration with flanging/squeal/HF squeal noise occurrence in brackets

Table 82: Observed main tendencies and impacts regarding equivalent continuous noise level generally without any type of squeal noise – raises due to squeal noise occurrence stated in brackets

- lateral acceleration (leads to inclusion of radius and cant as well)
- rail temperature
- specific humidity
- relative rail humidity
- wet conditions (rain and/or dew occurrence)
- frost conditions (frost and/or hoarfrost likely)
- curve squeal occurrence (distinguished in flanging/squeal/HF squeal noise)

In addition to temperature, dependencies on relative rail humidity and specific humidity are also evaluated, however not as distinct as in terms of temperature. In general, temperature and specific humidity show a negative correlation, while relative rail humidity indicates a positive one. Positive correlation to train velocity accords with past research (e.g. [5]). However, the stated range of the factor N in formula 2.2 in [5] of 25 to 35, which is presumably evaluated from tangent track, shows bad accordance in the present dataset. Instead, an empirically derived value of 7.5 works best for speed influence description (without A-weighting). Thus, level increase with speed might be slower in curves. However, it is unsure if the stated factor is applicable to other datasets. Since speed can be included as separate parameter in multivariate approaches, no correction is applied in statistical evaluations.

Building on the formulated main influence parameters, a regression model to predict equivalent continuous noise level in curves is developed. TDR and RR are included by using principal component analysis and extracting the first four components, which catch over 95 %

variance. Coefficients for building the components out of RR and TDR measurements are provided in Appendix D. Model training is done on 75 % data points of C1 and C5. Other data points in C1 and C5 are used as test set to validate model performance in similar boundary conditions. The remaining sections – C2, C3, C4 – are used as validation sets and represent an indicator, how the model performs on new datasets involving other sections, deviating train type distribution and completely different climatic conditions. Thus, focus lies on achieved benchmarks in the latter sections to estimate general applicability of the model to other datasets. The best performing statistical algorithm is Random Forest. It achieves a RMSE of 1.6 dB and a R^2 of 0.93 in the test set and a RMSE of 3.8 dB and a R^2 of 0.61 on the whole validation set (C2, C3 and C4 data points). Investigation of importance regarding environmental predictors reveals that prediction with only them (and curve squeal occurrence) leads to severe performance loss (decrease in RMSE on test set of 3.5 dB and on the whole validation set of 2.3 dB). Thus, separately modelling with only environmental predictors or only curve squeal indicators results in even worse benchmarks. Otherwise, if environmental conditions are excluded, marginal loss in predictive performance is observed (decrease in RMSE of 0.1 dB in test and validation set). Hence, prediction without considering environmental predictors is possible if a minor performance loss is accepted. Especially inclusion of all three curve squeal types is evaluated as beneficial (improvement in RMSE of 0.5 dB in test set and 0.6 dB in the whole validation set as well as in R^2 of 0.04 in test set and 0.09 in the whole validation set), which generates the need of applying all three formulated classification models for prediction of curve squeal occurrence prior to application of equivalent continuous noise level prediction. That scenario is carried out on the dataset – predicting curve squeal occurrence in the whole dataset on the one hand with the scope of overall classification accuracy and on the other hand with focus on improving sensitivity (refer to Subsection 6.5.1). The former application leads to significantly better RMSE values in test set and the whole validation set (improvement of 1.1 dB and 0.7 dB). Thus, for that application usage of stated cut-off probabilities with the scope of high classification accuracy is recommended to improve prediction accuracy in the present regression task. Moreover, considering the whole validation set, a 0.3 dB lower RMSE value is achieved by implementing curve squeal predictions compared to neglecting curve squeal predictors in the regression model, which justifies their importance and the stated way of application.

Investigation on deviation between several application scenarios of the prediction model – including curve squeal detection with the detection algorithm (Chapter 4), curve squeal prediction either using cut-off probabilities with the scope of improving sensitivity or overall classification accuracy and usage of ISO 3095 [10] threshold spectra instead of TDR and RR measurements – and the original measurements is carried out by comparison of averaged values (median and energetic mean) of each validation set among the whole measurement period. In general, considering both median and energetic mean results, equivalent continuous noise level predictions using cut-off probabilities with the scope of improving overall classification accuracy in prior curve squeal occurrence estimation deliver promising benchmarks. Median deviations are in a range from -0.2 dB to 1.4 dB and energetic mean differences lie in a range from -0.5 dB to -1.1 dB between predictions and original measurements – depending on the

validation set. Deviations between using the detection algorithm and the previously stated approach are $\leq$ 0.6 dB in median and $\leq$ 0.5 dB in energetic mean. Thus, the recommended cut-off probability thresholds for curve squeal occurrence estimation in conjunction with equivalent continuous noise level prediction stated in the previous paragraph are further justified. Moreover, the applicability of the elaborated models for predicting curve squeal occurrence is further confirmed. Exclusion of curve squeal (i.e. setting all three curve squeal predictors to 0 among all data points) causes a decrease in energetic mean values among predictions in a range from 0.3 dB to 1.6 dB compared to the recommended approach. Especially in C2 and C3, decrease magnitudes are $\leq$ 0.5 dB, which indicates that curve squeal consideration looses importance in those sections. Main reasons are presumably significantly higher freight train percentage compared to C4 – pass by levels of those are nearly unaffected by curve squeal occurrence (refer to Table 82) – as well as larger curve radii compared to C4, which generally lowers curve squeal probability. Using cut-off probabilities with the scope of high sensitivity in curve squeal prediction leads to median deviations in a range from 0.8 dB to 5.2 dB and energetic mean differences in a range from -0.6 dB to 1.1 dB between prediction and original measurement. Despite the promising result in terms of energetic mean, the probably high false positive rates in curve squeal detection and their impact on possible overestimations of level predictions has to be borne in mind. Using the recommended approach with application of threshold spectra according to ISO 3095 [10] instead of TDR and RR measurements leads to underestimations in median in a range from 2.1 dB to 0.0 dB and in energetic mean in a range from 1.3 dB to 3.1 dB. That is reasonable, since all sections show an elevated RR in some wavelength bands and/or a partly lower TDR compared to the threshold spectra. Thus, in curves, where RR is in general commonly worse compared to ISO 3095 [10] limits, a measurement of TDR and RR is recommended. Otherwise, the prediction result could be raised by an additive factor to account track properties deviating from the ISO 3095 [10] limits. The magnitude has to be determined by further evaluations on appropriate sections. It has to be pointed out that the accuracy in averaged values among the whole measurement periods seem significantly dependent on train type distribution – especially ratio between freight and passenger trains. Considering the results of the recommended approach, averaged levels of the most frequent passenger train category are overestimated by 2.0 dB (in C2) and 5.4 dB (in C3) in energetic mean, whereas freight train levels are underestimated by 3.5 dB (in C2) and 3.2 dB (in C3) in energetic mean. Due to the balanced ratio of freight and passenger trains in both sections, the overall prediction results have a high accuracy in energetic mean of -1.1 dB (C3) and -0.8 dB (C2) deviation between prediction and original measurement. Thus, assuming the same boundary conditions as in C2 and C3, the model would significantly underestimate the energetic mean level in case of 100 % freight train operation and otherwise, in case of 100 % passenger traffic, overestimations would result. The bulk of monitored freight coaches in all sections are operated with cast-iron block brakes. Legislative of the EU forbids the operation of coaches with that braking system on certain routes, which have to be defined by each member state, by the end of 2024 [206]. Due to the underestimation of those by the model, it is possible that predictions are getting more accurate for freight trains consisting only of coaches with disc brakes or composite brake pads. Thoughts about other limitations

of the elaborated model are provided in Subsection 6.6. Altogether, results are promising and further validation on other datasets is recommended.

6.6 Limitations

section	freight / %	passenger / %
C1	4	95
C2	40	59
C3	55	37
C4	4	94
C5	1	99

Table 83: Ratio of freight and passenger traffic among sections

Since no data points from tangent track are available, it is unsure if stated findings regarding equivalent continuous noise level without curve squeal occurrence are curve specific. The carried out evaluations are built on measurement data collected in five different sections. In general, all data points are monitored on the heavy rail network. Hence, applicability on light rail networks cannot be proven. Two long-term measurement campaigns (C1 and C5) with rather narrow curve radii (230 m and 256 m) deliver the bulk of data points (91 %). Thus, the whole examination is impacted by trends in those – especially detailed investigations, where further data splitting is applied. Table 83 provides the ratio of passenger and freight traffic among all sections. The remaining percentage covers train pass bys categorised in the random category "other". Hence, the bulk of data points are passenger trains – especially in the two main sections (C1 and C5), which causes that statistical significance in evaluations regarding freight traffic is mostly weaker compared to those of passenger trains. Nevertheless, developed prediction models for curve squeal occurrence and equivalent continuous noise level are validated on the three short-time measurement campaigns, where partly different section properties, train type distributions (especially freight to passenger train ratio) and climatic conditions are present. Despite that, predictive performance might be compromised if tested and validated boundary conditions are left. The latter refers especially to applicability on higher or narrower curve radii (< 226 m and > 440 m), significantly deviating track properties (rail roughness and track decay rate), higher curving speeds (> 100 kmh^{-1}) and extreme climatic conditions – e.g. curves in mountain routes or generally significantly deviating from those in the east of Austria. Another aspect to point out is that measurement data is collected between 2013 and 2017, where the majority of freight wagons are braked with cast-iron blocks. Nowadays, legislative of the EU forbids operation with cast-iron block brakes on certain routes, which have to be defined by each member state, by the end of 2024 [206]. Hence, deviation between freight and passenger traffic levels will decrease significantly in the future, which is impossible to include in the developed models. However, since grouping of train type groups with similar outcome values is carried out in each prediction model, freight trains on routes, where operation with cast-iron brakes is forbidden in the future, can be categorised in the mediocre category of passenger traffic. Validation of that approach cannot be pro-

vided, respectively. If newly developed train coaches and/or trainsets are used in operation, behaviour of those train types have to be investigated by collected measurement data from real operation and afterwards evaluation of similar properties to a train type in the present evaluation. Otherwise, categorisation into the mediocre passenger train category is recommended since it is not thought that newly developed trainsets or coaches show higher curve squeal probability rates or rolling noise than those. The developed models are not capable of including modifications due to application of lubrication systems on the rail head or gauge face. It might be an approach to include those conditions as rain or dew occurrence into the model. However, validation for that assumption cannot be provided.

7 Summary

The present thesis investigates influence factors on frequency of occurrence, peak level and relative occurrence time of three distinguished curve squeal types – flanging, squeal and high-frequency (HF) squeal noise. The excitation mechanism of the former is rubbing of the wheel flange on the rail gauge face (emission more broad-band and high fundamental frequencies – evaluated between 8000 Hz and 12500 Hz), while the second is caused by lateral stick-slip motion of the wheel on the rail tread (lower fundamental frequency and tonal character – detection range set to 800 Hz - 6300 Hz). Underlying mechanisms of the third are not entirely clear (tonal character with high fundamental frequencies – evaluated between 8000 Hz and 12500 Hz). Moreover, influence factors on general noise emission in curves are investigated. Examinations are carried out empirically and statistically using sound measurements of train pass bys (standardized distance according to ISO 3095 [10]) collected in five different sections with radii between 226 m and 440 m and speed limits between 60 kmh^{-1} and 90 kmh^{-1} in the Austrian heavy rail network. After data filtering, 29097 train pass bys are considered in the final evaluation. In addition to sound measurement data, environmental conditions are monitored with a mobile weather station beneath the sections and a rail temperature sensor located at the rail base between two sleepers, train speed as well as axle pattern are collected by two axle counter sensors. A third one on the other track (all sections on two-track lines) detects presence of parallel train pass bys, which are excluded, respectively. Different evaluations from measurement data combined with spectra of measured rail roughness (RR) and track decay rates (TDR) result in a set of 190 considered influence parameters. Dew formation modelling on the rail head as well as a developed regression approach for rail temperature prediction are also included. An empirically elaborated algorithm detects time increments containing curve squeal (distinguished in all three stated types) out of short-time averaged (64 ms) third-octave band spectra. It is used to categorise train pass bys as containing curve squeal as well as to extract the corresponding peak magnitude (short-time averaged peak third-octave band level) and relative occurrence time of the event. In addition, behaviour of equivalent continuous noise level (energetic mean among all time increments from buffer to buffer summed up between 50 Hz and 10000 Hz third-octave bands) is also evaluated. Thus, the formulated parameter set is investigated in conjunction with ten dependent variables. In a first step, main influence factors are extracted on the one hand with empirical and univariate statistical methods and on the other hand by applying multivariate statistical approaches.

Main findings on frequency of occurrence are a high impact of relative rail humidity with peaking probabilities between 72 % and 75 % for all three types and decreasing rates in both directions in the remaining area, partly distinct train type dependencies in probability magnitude, positive correlation of radial steering index (RSI) and lateral acceleration and significantly decreased relative frequency of occurrence regarding rain, dew and frost presence. The latter includes frozen water (from previous dew formation or rain) on the rail head as well as hoarfrost. Influence factors generally agree between all three curve squeal types, which indicates that their probability rates are raised in the same conditions, despite their deviating excitation mechanisms. Peak level evaluation reveals that amplitudes of both squeal noise

types are significantly higher compared to flanging noise. Moreover, negative correlations of curve radius, absolute humidity and temperature to the outcome as well as higher levels in frost conditions are observed in all three evaluations. Examinations of relative occurrence time indicate a higher value for flanging noise compared to roughly equal levels for both squeal noise types. Trends among all curve squeal types comply with stated ones regarding peak levels with the addition of a generally positive correlation of relative humidity. Equivalent continuous noise level tendencies are primarily evaluated without any curve squeal occurrence. Beside known impacts on rolling noise – TDR, RR, sleeper material, train type, train speed –, which are clearly indicated in the evaluation, raised levels in conjunction with dew occurrence and frost are observed. Each of them cause a raise in median values of up to 2 dB – behaviour extracted from three different train types among two sections separately. Spectral investigation reveals a complying raise around 160 Hz and 400 Hz/500 Hz third-octave band levels of up to 2 dB each and shown in terms of dew and frost similarly. Both areas are among the highest spectral contributors in considered train types (A-weighting intentionally not applied). Thus, impacts are shown in multiple train types and sections in the same frequency range. Temperature as well as absolute humidity show negative correlations to the outcome, while relative humidity shows a positive one.

Building on formulated influence factors on frequency of occurrence, classification models to predict occurrence of each curve squeal type are developed. For flanging noise, flexible discriminant analysis is used, while for both squeal noise types, separate neural networks are built. They are trained on two sections, which contain the bulk of data points, and validated on the other three as well as on an independent test set split prior to modelling from the training sections. Thus, applicability is proven in three different sections, which are not used in model training involving different section properties, train type distributions and climatic conditions. Alternative cut-off probability thresholds are investigated with a ROC curve in a separate evaluation set, which is solely used for that purpose. Developed models are capable of reaching curve squeal detection rates (i.e. sensitivity) among the whole validation set of 68 % to 84 % in conjunction with an overall classification accuracy of 55 % to 68 %. If the latter benchmark is the main modelling scope, values are improvable to 67 % to 83 % on the whole validation set, however causing a drop of sensitivity to values between 27 % and 66 % by application of different cut-off probability thresholds. Due to poor benchmarks of multivariate statistical analysis and the scope of practical applicability, development of a prediction model for curve squeal peak levels is not seen as meaningful. High fluctuations in the dataset (resulting in high RMSE values in multivariate statistical approaches) prevent also a predictive model building process regarding relative occurrence time. For rough estimation of the former, stated median values for similar train types and radii can be applied. For the latter, heuristic models are formulated, which use only curve radius as parameter as well as partly distinguish between passenger and freight traffic.

At last, a regression model for prediction of equivalent continuous noise level of single train pass bys in curves in the standardized measurement point according to ISO 3095 [10] – 7.5 m distance from track axis and 1.2 m above rail level – is developed. As statistical method, a random forest approach performs best. For model training and validation, the

same approach as for the presented classification tasks is used. Since it improves predictive performance, categorical predictors for occurrence of each distinguished curve squeal type are included. Thus, in practical application, occurrence of curve squeal has to be predicted using the previously described classification models prior to application of the model for equivalent continuous noise level prediction. For that application, usage of cut-off probability thresholds with the scope of improving overall classification accuracy results in better statistical benchmarks. Performance quantities of the model in application on the whole validation set with inclusion of predicted curve squeal occurrence are 4 dB RMSE and 57 % R^2. Averaged values among the whole measurement period achieve median deviations in a range from -0.2 dB to 1.4 dB and energetic mean differences in a range from -0.5 dB to -1.1 dB between predictions and original measurements – depending on the validation set. Using the same approach but with application of threshold spectra according to ISO 3095 [10] instead of TDR and RR measurements leads to underestimations in median in a range from 0.0 dB to 2.1 dB and in energetic mean in a range from 1.3 dB to 3.1 dB. That is reasonable, since ISO 3095 [10] limits are partly exceeded in all sections. It has to be pointed out that the stated accuracy in averaged values seem significantly dependent on train type distribution – especially ratio between freight and passenger trains due to general underestimations of the former and tendency to overestimations regarding the latter. Nevertheless, results are promising and further validation on other datasets is recommended.

8 Future work

Future research need is strongly indicated by carried out evaluations. Firstly, the underlying mechanism of HF squeal noise is not entirely clear as it shows significant time behaviour deviations to squeal noise, however amplitudes are in the same range of the latter with higher fundamental frequencies. Considering frequency of occurrence, peak levels and relative occurrence time, stated influence factors and their trends partly lack of explanations. Theoretical modelling as well as validation in laboratory measurement under controlled conditions is recommended. Moreover, behaviour of passenger trains in conjunction with higher curve radii ($\geq$ 310 m) cannot be sufficiently evaluated due to lack of data points. Despite lower frequency of occurrence and thus need of more data points in evaluations, additional long-term measurement campaigns are beneficial. In general, fine tuning and validation of the developed prediction models on new datasets is recommended. Especially applicability on higher or narrower curve radii ($<$ 226 m and $>$ 440 m), significantly deviating track properties (rail roughness and track decay rate), extreme climatic conditions (e.g. curves in mountain routes) and higher curving speeds ($>$ 100 kmh^{-1}) needs to be proven. Moreover, changes in freight traffic behaviour compared to evaluated trends on tracks, where cast-iron block braked wagons are going to be forbidden in future operation, is interesting.

Regarding equivalent continuous noise level, two raising mechanisms – dew and frost occurrence – are observed, which reasons are unknown. Despite some stated assumptions and thoughts, there is a complete lack of scientific proof. Moreover, the general trend of negative correlations to temperature, absolute humidity and the positive one to relative humidity is not entirely explainable with past research findings. It is foremost of interest if the stated deviations also occur on tangent track or if it states a phenomenon in curves.

All finally applied model objects in R are available online[28]. They can be taken and applied to new sections as well as extended by application to new data points with deep learning algorithms to improve their accuracy and general applicability. Moreover, mitigation measures (e.g. friction control systems) can be included. An idea for that would be to account pass bys, where the friction control system is enabled, equal to rain or dew – setting the categorical variable for rain to 1.

At last, an investigation in the same scale for light rail traffic considering both an embedded track and a conventional one is recommended. Operation of light rail is mostly carried out in densely populated areas, thus the need of a detailed prediction approach regarding noise emission in curves is high.

[28] https://github.com/M-Ostermann/curve-noise-prediction

Bibliography

[1] European Parliament and Council of the European Union. *Regulation (EU) No 1315/2013 of the European Parliament and of the Council of 11 December 2013 on Union Guidelines for the Development of the Trans-European Transport Network and Repealing Decision No 661/2010/EU.*

[2] WHO Regional Office for Europe. *Environmental Noise Guidelines for the European Region*. 2018. ISBN: 978-92-890-5356-3.

[3] International Union of Railways (UIC). *Railway Noise Technical Measures Catalogue*. UIC003-01-04fe. 2013.

[4] European Commission. *Comission Directive (EU) 2015/996 of 19 May 2015 Establishing Common Noise Assessment Methods According to Directive 2002/49/EC of the European Parliament and of the Council*.

[5] D. Thompson. *Railway Noise and Vibration: Mechanisms, Modelling and Means of Control*. 1st ed. Oxford, UK: Elsevier, 2009. ISBN: 978-0-08-045147-3.

[6] N. Ostermann, ed. *Anwenderhandbuch - Systematische Bahnlärmbekämpfung*. Hamburg, Germany: DVV Media Group GmbH | Eurailpress, 2013. ISBN: 978-3-7771-0464-5.

[7] I. Zenzerovic. "Engineering Model for Curve Squeal Formulated in the Time Domain". Doctoral thesis. Gothenburg, Sweden: Chalmers University of Technology, 2014.

[8] B. Stallaert and P. Vanhonacker. "Rolling Noise and Corrugation in Curves: Modelling and Solutions". In: *Proceedings IWRN12*. 12th International Workshop on Railway Noise. Terrigal, Australia, Sept. 12–16, 2016, p. 8.

[9] T. Maly et al. *Bewertung des akustischen Einflusses von Gleisbögen für die Erstellung von Lärmkarten BEGEL*. 2015. URL: `https://www2.ffg.at/verkehr/studien.php?id=1101&lang=de&browse=programm` (visited on 02/11/2021).

[10] International Organization for Standardization. *ISO 3095:2013 Acoustics - Railway Applications - Measurement of Noise Emitted by Railbound Vehicles*.

[11] Zentralanstalt für Meteorologie und Geodynamik Kundenservice Österreich, Wien, Niederösterreich und Burgenland. *Measurement Data for Scientific Research*. Oct. 24, 2019.

[12] European Parliament and Council of the European Union. *Directive 2002/49/EC of the European Parliament and of the Council of 25 June 2002 Relating to the Assessment and Management of Environmental Noise*.

[13] Y. Zhu. "Adhesion in the Wheel-Rail Contact". Doctoral thesis. Stockholm, Sweden: KTH Royal Institute of Technology, 2013.

[14] A. Meierhofer et al. "Third Body Layer – Experimental Results and a Model Describing Its Influence on the Traction Coefficient". In: *Wear* 314.1-2 (2014), pp. 148–154. DOI: `10.1016/j.wear.2013.11.040`.

[15] Y. Berthier et al. “The Role and Effects of the Third Body in the Wheel-Rail Interaction”. In: *Fatigue and Fracture of Engineering Materials and Structures* 27 (2004), pp. 423–436. DOI: 10.1111/j.1460-2695.2004.00764.x.

[16] I. Hutchings and P. Shipway. *Tribology: Friction and Wear of Engineering Materials*. 2nd ed. Oxford, UK: Elsevier, 2017. ISBN: 978-0-08-100910-9.

[17] D. Godfrey. “Iron Oxides and Rust (Hydrated Iron Oxides) in Tribology”. In: *Lubrication Engineering* 55.2 (1999), pp. 33–37.

[18] Y. Zhu, U. Olofsson, and R. Nilsson. “A Field Test Study of Leaf Contamination on Railhead Surfaces”. In: *Proceedings of the Institution of Mechanical Engineers Part F: Journal of Rail and Rapid Transit* 228.1 (2012), pp. 71–84. DOI: 10.1177/0954409712464860.

[19] T. M. Beagley, I. J. McEwen, and C. Pritchard. “Wheel/Rail Adhesion – the Influence of Railhead Debris”. In: *Wear* 33.1 (1975), pp. 141–152. DOI: 10.1016/0043-1648(75)90230-6.

[20] T. M. Beagley, I. J. McEwen, and C. Pritchard. “Wheel/Rail Adhesion – Boundary Lubrication by Oily Fluids”. In: *Wear* 31.1 (1975), pp. 77–88. DOI: 10.1016/0043-1648(75)90123-4.

[21] T.M. Beagley and C. Pritchard. “Wheel/Rail Adhesion – the Overriding Influence of Water”. In: *Wear* 35.2 (1975), pp. 299–313. DOI: 10.1016/0043-1648(75)90078-2.

[22] J. M. Lujnov and S. I. Kossikov. “Paper 3: Friction on Railway Rails”. In: *Proceedings of the Institution of Mechanical Engineers* 178.5 (1963), pp. 16–23. DOI: 10.1243/PIME_CONF_1963_178_148_02.

[23] L. E. Buckley-Johnstone. “Wheel/Rail Contact Tribology: Characterising Low Adhesion Mechanisms and Friction Management Products”. Doctoral thesis. Sheffield, UK: The University Of Sheffield, 2017.

[24] S. R. Lewis et al. “Investigation of the Isolation and Frictional Properties of Hydrophobic Products on the Rail Head, When Used to Combat Low Adhesion”. In: *Wear* 314.1-2 (2014), pp. 213–219. DOI: 10.1016/j.wear.2013.11.024.

[25] H. Chen et al. “Experimental Investigation of Influential Factors on Adhesion between Wheel and Rail under Wet Conditions”. In: *Wear* 265.9-10 (2008), pp. 1504–1511. DOI: 10.1016/j.wear.2008.02.034.

[26] U. Olofsson and K. Sundvall. “Influence of Leaf, Humidity and Applied Lubrication on Friction in the Wheel-Rail Contact: Pin-on-Disc Experiments”. In: *Proceedings of the Institution of Mechanical Engineers Part F: Journal of Rail and Rapid Transit* 218.3 (2004), pp. 235–242. DOI: 10.1243/0954409042389364.

[27] W.Y.H. Liew. “Effect of Relative Humidity on the Unlubricated Wear of Metals”. In: *Wear* 260.7-8 (2006), pp. 720–727. DOI: 10.1016/j.wear.2005.04.011.

[28] O. Hayashi, T. Nomura, and K. Nagase. "Influence of Atmospheric Conditions upon Adhesion between Rails and Running Wheels". In: *Transactions of the Japan Society of Mechanical Engineers Series C* 63.606 (1997), pp. 566–571. DOI: 10.1299/kikaic.63.566.

[29] Y. Zhu, U. Olofsson, and H. Chen. "Friction Between Wheel and Rail: A Pin-On-Disc Study of Environmental Conditions and Iron Oxides". In: *Tribol Lett* 52 (2013), pp. 327–339. DOI: 10.1007/s11249-013-0220-0.

[30] Yi Zhu. "The Influence of Iron Oxides on Wheel-Rail Contact: A Literature Review". In: *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 232.3 (2018), pp. 734–743. DOI: 10.1177/0954409716689187.

[31] Y. Zhu, Y. Lyu, and U. Olofsson. "Mapping the Friction between Railway Wheels and Rails Focusing on Environmental Conditions". In: *Wear* 324-325 (2015), pp. 122–128. DOI: 10.1016/j.wear.2014.12.028.

[32] W. J. Wang et al. "Experimental Study on Adhesion Behavior of Wheel/Rail under Dry and Water Conditions". In: *Wear* 271.9-10 (2011), pp. 2699–2705. DOI: 10.1016/j.wear.2011.01.070.

[33] A. H. Wickens. *Fundamentals of Rail Vehicle Dynamics: Guidance and Stability*. Lisse, Netherlands: Swets & Zeitlinger Publishers, 2003. ISBN: 0-203-97099-3.

[34] J. Ihme. *Schienenfahrzeugtechnik*. 2nd ed. Wiesbaden, Germany: Springer Vieweg, 2019. ISBN: 978-3-658-24923-6.

[35] D. J. Thompson, G. Squicciarini, and B. Ding. "A State-of-the-Art Review of Curve Squeal Noise: Phenomena, Mechanism, Modelling and Mitigation". In: *Noise and Vibration Mitigation for Rail Transportation Systems*. Vol. 139. Notes on Numerical Fluid Mechanics and Multidisciplinary Design. Cham, Germany: Springer, 2018, pp. 3–41. DOI: 10.1007/978-3-319-73411-8_1.

[36] D. J. Fourie. "Mechanisms Influencing Railway Wheel Squeal Excitation in Large Radius Curves". Doctoral thesis. Johannesburg, South Africa: University of Johannesburg, 2011.

[37] X. Liu and P. A. Meehan. "Investigation of the Effect of Relative Humidity on Lateral Force in Rolling Contact and Curve Squeal". In: *Wear* 310.1-2 (2014), pp. 12–19. DOI: 10.1016/j.wear.2013.11.045.

[38] J. J. Kalker. *Three-Dimensional Elastic Bodies in Rolling Contact*. Solid Mechanics and Its Applications. Dordrecht, Netherlands: Springer Science+Business Media, 1990. ISBN: 978-94-015-7889-9.

[39] M. J. Rudd. "Wheel/Rail Noise – Part II: Wheel Squeal". In: *Journal of Sound and Vibration* 46.3 (1976), pp. 381–394. DOI: 10.1016/0022-460X(76)90862-2.

[40] D. Curley et al. “Field Trials of Gauge Face Lubrication and Top-of-Rail Friction Modification for Curve Noise Mitigation”. In: *Noise and Vibration Mitigation for Rail Transportation Systems*. Ed. by J. Nielsen et al. Vol. 126. Notes on Numerical Fluid Mechanics and Multidisciplinary Design. Berlin Heidelberg, Germany: Springer, 2015, pp. 449–456. DOI: 10.1007/978-3-662-44832-8_54.

[41] Ch. Glocker, E. Cataldi-Spinola, and R.I. Leine. “Curve Squealing of Trains: Measurement, Modelling and Simulation”. In: *Journal of Sound and Vibration* 324.1-2 (2009), pp. 365–386. DOI: 10.1016/j.jsv.2009.01.048.

[42] G. Squicciarini et al. “Curve Squeal in the Presence of Two Wheel/Rail Contact Points”. In: *Noise and Vibration Mitigation for Rail Transportation Systems*. Ed. by J. Nielsen et al. Vol. 126. Notes on Numerical Fluid Mechanics and Multidisciplinary Design. Berlin Heidelberg, Germany: Springer, 2015, pp. 603–610. DOI: 10.1007/978-3-662-44832-8_71.

[43] J. Jiang, R. Dwight, and D. Anderson. “Field Verification of Curving Noise Mechanisms”. In: *Noise and Vibration Mitigation for Rail Transportation Systems*. Ed. by T. Maeda et al. Vol. 118. Notes on Numerical Fluid Mechanics and Multidisciplinary Design. Tokyo, Japan: Springer, 2012, pp. 349–356. DOI: 10.1007/978-4-431-53927-8_41.

[44] N. Vincent et al. “Curve Squeal of Urban Rolling Stock – Part 1: State of the Art and Field Measurements”. In: *Journal of Sound and Vibration* 293.3 (2006), pp. 691–700. DOI: 10.1016/j.jsv.2005.12.008.

[45] J. R. Koch et al. “Curve Squeal of Urban Rolling Stock – Part 2: Parametric Study on a 1/4 Scale Test Rig”. In: *Journal of Sound and Vibration* 293.3-5 (2006), pp. 701–709. DOI: 10.1016/j.jsv.2005.12.009.

[46] N. Hoffmann et al. “A Minimal Model for Studying Properties of the Mode-Coupling Type Instability in Friction Induced Oscillations”. In: *Mechanics Research Communications* 29.4 (2002), pp. 197–205. DOI: 10.1016/S0093-6413(02)00254-9.

[47] B. Ding et al. “An Assessment of Mode-Coupling and Falling-Friction Mechanisms in Railway Curve Squeal through a Simplified Approach”. In: *Journal of Sound and Vibration* 423 (2018), pp. 126–140. DOI: 10.1016/j.jsv.2018.02.048.

[48] F.G. de Beer, M.H.A. Janssens, and P.P. Kooijman. “Squeal Noise of Rail-Bound Vehicles Influenced by Lateral Contact Position”. In: *Journal of Sound and Vibration* 267.3 (2003), pp. 497–507. DOI: 10.1016/S0022-460X(03)00710-7.

[49] A. D. Monk-Steel et al. “An Investigation into the Influence of Longitudinal Creepage on Railway Squeal Noise Due to Lateral Creepage”. In: *Journal of Sound and Vibration* 293.3-5 (2006), pp. 766–776. DOI: 10.1016/j.jsv.2005.12.004.

[50] S. S. Hsu et al. “Experimental and Theoretical Investigation of Railway Wheel Squeal”. In: *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 221.1 (2007), pp. 59–73. DOI: 10.1243/0954409JRRT85.

[51] X. Liu and P. A. Meehan. "Investigation of the Effect of Lateral Adhesion and Rolling Speed on Wheel Squeal Noise". In: *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 227.5 (2013), pp. 469–480. DOI: 10.1177/0954409713501653.

[52] L. Xiaogang and P. A. Meehan. "Investigation of Squeal Noise under Positive Friction Characteristics Condition Provided by Friction Modifiers". In: *Journal of Sound and Vibration* 371 (2016), pp. 393–405. DOI: 10.1016/j.jsv.2016.02.028.

[53] C. Collette. "Importance of the Wheel Vertical Dynamics in the Squeal Noise Mechanism on a Scaled Test Bench". In: *Shock and Vibration* 19.2 (2012), pp. 145–153. DOI: 10.3233/SAV-2011-0620.

[54] M. Dittrich et al. *ACOUTRAIN: Virtual Certification of Acoustic Performance for Freight and Passenger Trains*. Deliverable D1.10 Braking noise and Curving noise: corresponding indicators and measurement procedures (final, update of D1.3) EC Contract No. FP7 - 284877. 2014.

[55] L.G. Kurzweil. "Wheel/Rail Noise – Means for Control". In: *Journal of Sound and Vibration* 87.2 (1983), pp. 197–220. DOI: 10.1016/0022-460X(83)90555-2.

[56] P.J. Remington. "Wheel/Rail Squeal and Impact Noise: What Do We Know? What Don't We Know? Where Do We Go from Here?" In: *Journal of Sound and Vibration* 116.2 (1987), pp. 339–353. DOI: 10.1016/S0022-460X(87)81306-8.

[57] F. De Beer et al. "Curve Squeal of Railbound Vehicles (Part 1): Frequency Domain Calculation Model". In: *Proceedings of the 29th International Congress on Noise Control Engineering*. InterNoise. Vol. 3. Nice, France, Aug. 27–30, 2000, pp. 1560–1563.

[58] P.-P. Kooijman et al. "Curve Squeal of Railbound Vehicles (Part 2): Set-up for Measurement of Creepage Dependent Friction Coeffiecient". In: *Proceedings of the 29th International Congress on Noise Control Engineering*. InterNoise. Vol. 3. Nice, France, Aug. 27–30, 2000, pp. 1564–1567.

[59] M. Janssens et al. "Curve Squeal of Railbound Vehicles (Part 3): Measurement Techniques for Wheel/Rail Contact Velocities and Forces at Squeal Noise Frequencies". In: *Proceedings of the 29th International Congress on Noise Control Engineering*. InterNoise. Vol. 3. Nice, France, Aug. 27–30, 2000, pp. 1568–1571.

[60] E. Jie et al. "An Experimental Study of Squeal Noise Characteristic for Railways Using a Scale Model Test Rig". In: *Proceedings of the Third International Conference on Railway Technology*. Third International Conference on Railway Technology. Ed. by J. Pombo. Vol. 110. Civil-Comp Proceedings. Cagliari, Sardinia, Italy: Civil-Comp Press, Apr. 5–8, 2016. DOI: 10.5050/KSNVE.2015.25.5.352.

[61] A. Matsumoto et al. "Creep Force Characteristics between Rail and Wheel on Scaled Model". In: *Wear* 253.1-2 (2002), pp. 199–203. DOI: 10.1016/S0043-1648(02)00100-X.

[62] P. A. Meehan et al. “Probabilistic Prediction of Wheel Squeal under Field Humidity Variation”. In: *Proceedings of 20th International Congress on Acoustics*. ICA 2010. Ed. by M. Burgess et al. Sydney, Australia: Curran Associates, Inc., Aug. 23–27, 2010, pp. 2198–2203. ISBN: 978-1-61782-745-7.

[63] T. Maly, F. Biebl, and M. Ostermann. “The Effects of Weather Conditions and Wheel Wear on Curve Squeal”. In: *Proceedings of the 23rd International Congress on Acoustics*. ICA 2019. Ed. by M. Ochmann, M. Vorländer, and J. Fels. Aachen, Germany: German Acoustical Society, DEGA e.V., Sept. 9–13, 2019, pp. 1559–1566. ISBN: 978-3-939296-15-7.

[64] Forschungsverbund Leiser Verkehr. *Geräusche im Schienennahverkehr: Forschung Kurvengeräusche Zusammenfassung*. 2010.

[65] H. Venghaus. “The Impact of Weather Conditions on the Noise Radiation Level of Curve Squeal”. In: *Proceedings IWRN13*. 13th International Workshop on Railway Noise. Ghent, Belgium, Sept. 16–20, 2019, p. 7.

[66] F. Krüger. *Verbundprojekt Leiser Verkehr: Kurvengeräusche - Entwicklung von anwendungsreifen und wirksamen Maßnahmen zur Reduzierung von Kurvenquietschen an Rad und Schiene*. Final report. Köln, Germany: STUVA - Studiengesellschaft für unterirdische Verkehrsanlagen e.V., 2009.

[67] D. Sehu et al. *sonRail - Projektdokumentation*. 2010. URL: `https://www.empa.ch/documents/56129/160826/sonRail-projektdokumentation.pdf/f036553f-80c8-4b87-bc74-ead535ea29ab` (visited on 02/11/2021).

[68] J. M. Wunderli. “sonRAIL - From the Scientific Model to an Application in Practice”. In: *Proceedings of Ninth European Conference on Noise Control*. 9th European Conference on Noise Control. Prague, Czech Republic: Czech Acoustical Society, June 10–13, 2012, pp. 475–480. ISBN: 978-80-01-05013-2.

[69] M. Hecht et al. “sonRAIL – The New Swiss Calculation Model for Railway Noise”. In: *Noise and Vibration Mitigation for Rail Transportation Systems - Proceedings of the 10th International Workshop on Railway Noise*. 10th International Workshop on Railway Noise. Ed. by T. Maeda et al. Vol. 118. Nagahama, Japan: Springer Japan, Oct. 18–22, 2010, pp. 515–522. DOI: `10.1007/978-4-431-53927-8_61`.

[70] Bundesministerium der Justiz und für Verbraucherschutz. *Sechzehnte Verordnung zur Durchführung des Bundes-Immissionsschutzgesetzes, Verkehrslärmschutzverordnung, Anlage 2 (zu §4), Berechnung des Beurteilungspegels für Schienenwege (Schall 03)*. 2014.

[71] Austrian Research Association for Roads, Railways and Transport. *RVE 04.01.02 - Calculation of Railway Noise Emission*. 2019.

[72] B. Klose and H. Klose. *Meteorologie - Eine Interdisziplinäre Einführung in Die Physik Der Atmosphäre*. 3. Auflage. Springer-Lehrbuch. Berlin Heidelberg, Germany: Springer Spektrum, 2016. ISBN: 978-3-662-43622-6.

[73] H. Malberg. *Meteorologie Und Klimatologie*. 5th ed. Berlin Heidelberg, Germany: Springer-Verlag, 2007. ISBN: 978-3-540-37222-6.

[74] J. L. Monteith and M. H. Unsworth. *Principles of Environmental Physics*. 4th ed. Oxford, UK: Academic Press, 2013. ISBN: 978-0-12-386993-7.

[75] H. Vuollekoski et al. "Estimates of Global Dew Collection Potential on Artificial Surfaces". In: *Hydrology and Earth System Sciences* 19 (2015), pp. 601–613. DOI: 10.5194/hess-19-601-2015.

[76] K. Richards. "Adaptation of a Leaf Wetness Model to Estimate Dewfall Amount on a Roof Surface". In: *Agricultural and Forest Meteorology* 149.8 (2009), pp. 1377–1383. DOI: 10.1016/j.agrformet.2009.02.014.

[77] H. Xiao et al. "Analysis of the Effect of Meteorological Factors on Dewfall". In: *Science of The Total Environment* 452-453 (2013), pp. 384–393. DOI: 10.1016/j.scitotenv.2013.03.007.

[78] H. Wang and P. Balaguru. *Low Solar Absorption Coating for Reducing Rail Temperature and Preventing Buckling*. Final report DOT/FRA/ORD-18/28. Washington, DC, USA: The State University of New Jersey, 2018.

[79] World Meteorological Organization. *Volume 1 - Measurement of Meteorological Variables*. WMO No. 8. Geneva, Switzerland, 2018. ISBN: 978-92-63-10008-5.

[80] World Meteorological Organization. *International Meteorological Vocabulary*. 2nd ed. WMO No. 182. Geneva, Switzerland, 1992. ISBN: 978-92-63-02182-3.

[81] World Meteorological Organization. *Technical Note No. 61 - Note on the Stadardization of Pressure Reduction Methods in the International Network of Synoptic Stations*. WMO-No. 154.TP.74. 1964.

[82] H. Vömel. *Saturation Vapor Pressure Formulations*. Boulder, CO, USA: National Center for Atmospheric Research, 2016. URL: http://cires1.colorado.edu/~voemel/vp.html (visited on 02/11/2021).

[83] World Meteorological Organization. *International Metereological Tables*. Ed. by S. Letestu. WMO-No. 188.TP.94. 1966.

[84] T. Foken. *Angewandte Meteorologie: Mikrometeorologische Methoden*. 3rd ed. Berlin Heidelberg, Germany: Springer Spektrum, 2016. ISBN: 978-3-642-25525-0.

[85] T. Maly et al. *Einflüsse auf Schallemissionen in Bögen ESB*. 2019. URL: https://projekte.ffg.at/projekt/1413863 (visited on 02/11/2021).

[86] M. T. Kalivoda. "Track Decay Rate of Different Railway Noise Test Sites". In: *Proceedings of Forum Acusticum 2005*. Forum Acusticum. Ed. by F. Augusztinovicz, A. B. Nagy, and Z. Hunyadi. Budapest, Hungary, Aug. 29–Sept. 2, 2005, p. 3. ISBN: 978-963-8241-68-9.

[87] DIN Deutsches Institut für Normung e. V. *DIN EN 15461:2011 Railway applications - Noise emission - Characterisation of the dynamic properties of track sections for pass by noise measurements*. 2011.

[88] HEAD acoustics GmbH. *ArtemiS SUITE 10.7*. Version 10.7.19182.3. 2020.

[89] M. Möser. *Technische Akustik*. 10th ed. VDI-Buch. Berlin Heidelberg, Germany: Springer Vieweg, 2015. ISBN: 978-3-662-47704-5.

[90] F. Pospischil. *Längsverschweißtes Gleis im engen Bogen: eine Betrachtung der Gleislagestabilität*. Ed. by M. Mailer. Vol. 1. Schriftenreihe des Arbeitsbereichs Intelligente Verkehrssysteme an der Universität Innsbruck. Innsbruck, Austria: Studia Universitätsverlag, 2015. ISBN: 978-3-903030-20-6.

[91] S. Rangosch. "Lagestabilität lückenloser Meterspurgleise in kleinen Bogenradien". Doctoral thesis. Zurich, Switzerland: ETH Zürich, 1995. DOI: 10.3929/ethz-a-001517517.

[92] T. Maly et al. *Ermittlung von längenbezogenen Schallleistungspegeln und Eingangsparameter für CNOSSOS-EU ELSEC*. 2019. URL: https://projekte.ffg.at/projekt/1695860 (visited on 02/11/2021).

[93] R. Rotter and H. Petrovitsch. *Triebfahrzeuge österreichischer Eisenbahnen: Elektrische Lokomotiven und Triebwagen*. 2nd ed. Düsseldorf, Germany: Alba Fachverlag, 1999. ISBN: 978-3-87094-740-8.

[94] P. Reinthaler and H. Heless. *Reisezugwagen österreichischer Eisenbahnen: Vierachsige Reisezugwagen in Ganzstahlbauart der ÖBB*. Düsseldorf, Germany: Alba Fachverlag, 2006. ISBN: 978-3-87094-194-9.

[95] ÖBB-CI & M Werbeagentur GmbH. *railjet - technische Kurzbeschreibung*. 2008. URL: http://www.bahnzauber-europa.at/Railjet%20Folder.pdf (visited on 02/11/2021).

[96] R. Beier. *Loks Der ÖBB: Österreichische Bundesbahnen Seit 1947*. 2nd ed. Typenkompass. Stuttgart, Germany: Transpress-Verlag, 2014. ISBN: 978-3-613-71413-7.

[97] DIN Deutsches Institut für Normung e. V. *DIN EN 14363:2019 - Railway Applications - Testing and Simulation for the Acceptance of Running Characteristics of Railway Vehicles - Running Behaviour and Stationary Tests*. 2019.

[98] DIN Deutsches Institut für Normung e. V. *E DIN EN 15302:2018 - Railway Apllications - Wheel-Rail Contact Geometry Parameters - Definitions and Methods for Evaluation*. 2018.

[99] DIN Deutsches Institut für Normung e. V. *DIN EN 13803:2017 - Railway applications - Track - Track alignment design parameters - Track gauges 1435 mm and wider*. 2017.

[100] Conrad Electronic SE. *Bedienungsanleitung: Funk-Wetterstation Professional USB*. 2014.

[101] Vaisala Oyj. *User's Guide - Vaisala Weather Transmitter WXT520*. 2012.

[102] D. Salz. "Terzspektrenbasierte Bewertung der Tonhaltigkeit von Schienenfahrzeuggeräuschen". Doctoral thesis. Berlin, Germany: Technische Universität Berlin, 2006.

[103] T. Haiden et al. "The Integrated Nowcasting through Comprehensive Analysis (INCA) System and Its Validation over the Eastern Alpine Region". In: *Weather and Forecasting* 26.2 (2011), pp. 166–183. DOI: 10.1175/2010WAF2222451.1.

[104] A. Kann and T. Haiden. "INCA – an Operational Nowcasting System for Hydrology and Other Applications". In: *Berichte Geol. B.-A.* NÖ GEOTAGE. Vol. 88. Haindorf bei Langenlois, Austria, Sept. 29–30, 2011.

[105] M. Olefs and W. Schoener. "A New Solar Radiation Model for Research and Applications in Austria". In: *EGU General Assembly 2012*. Vienna, Austria, Apr. 22–27, 2012.

[106] J. Hedderich and L. Sachs. *Angewandte Statistik*. 16th ed. Berlin, Germany: Springer Spektrum, 2018. ISBN: 978-3-662-56657-2.

[107] C. D. Schoenwiese. *Klimatologie*. 3rd ed. Stuttgart, Germany: Eugen Ulmer, 2008. ISBN: 978-3-8252-1793-8.

[108] H. Wang et al. "Thermal Benefits of Low Solar Absorption Coating for Preventing Rail Buckling". In: *Proceedings of the 2015 Joint Rail Conference*. 2015 Joint Rail Conference. San Jose, California, USA: American Society of Mechanical Engineers, Mar. 23–26, 2015. DOI: 10.1115/JRC2015-5669.

[109] R. G. Allen et al. *Crop Evapotranspiration - Guidelines for Computing Crop Water Requirements*. Rome, Italy: FAO - Food and Agriculture Organization of the United Nations, 1998. ISBN: 92-5-104219-5. URL: http://www.fao.org/3/X0490E/X0490E00.htm (visited on 02/11/2021).

[110] V. S. Nikolayev et al. "Water Recovery from Dew". In: *Journal of Hydrology* 182.1-4 (1996), pp. 19–35. DOI: 10.1016/0022-1694(95)02939-7.

[111] J. C. Dixon. *The Shock Absorber Handbook*. 2nd ed. Wiley-Professional Engineering Publishing Series. Chichester, UK: John Wiley & Sons, 2007. ISBN: 978-0-470-51020-9.

[112] International Organization of Standardization. *ISO 9613-1:1993 - Acoustics - Attenuation of Sound during Propagation Outdoors*. Geneva, Switzerland, 1993.

[113] M. Kuhn and K. Johnson. *Applied Predictive Modeling*. New York, USA: Springer-Verlag, 2013. ISBN: 978-1-4614-6848-6.

[114] M. Kuhn. "Building Predictive Models in R Using the Caret Package". In: *Journal of Statistical Software* 28.5 (2008). DOI: 10.18637/jss.v028.i05.

[115] I.-K. Yeo and R. A. Johnson. "A New Family of Power Transformations to Improve Normality or Symmetry". In: *Biometrika* 87.4 (2000), pp. 954–959. JSTOR: 2673623.

[116] H. Abdi and L. J. Williams. "Principal Component Analysis". In: *Wiley Interdisciplinary Reviews: Computational Statistics* 2.4 (2010), pp. 433–459. DOI: 10.1002/wics.101.

[117] T. O. Kvalseth. "Cautionary Note about R^2". In: *The American Statistician* 39.4 (1985), pp. 279–285. DOI: 10.1080/00031305.1985.10479448.

[118] D. G. Altman and J. M. Bland. "Statistics Notes: Diagnostic Tests 3: Receiver Operating Characteristic Plots". In: *BMJ* 309.6948 (1994). DOI: 10.1136/bmj.309.6948.188.

[119] C. D. Brown and H. T. Davis. "Receiver Operating Characteristics Curves and Related Decision Measures: A Tutorial". In: *Chemometrics and Intelligent Laboratory Systems* 80.1 (2006), pp. 24–38. DOI: 10.1016/j.chemolab.2005.05.004.

[120] T. Fawcett. "An Introduction to ROC Analysis". In: *Pattern Recognition Letters* 27.8 (2006), pp. 861–874. DOI: 10.1016/j.patrec.2005.10.010.

[121] P. Geladi and B. R. Kowalski. "Partial Least-Squares Regression: A Tutorial". In: *Analytica Chimica Acta* 185 (1986), pp. 1–17. DOI: 10.1016/0003-2670(86)80028-9.

[122] M. Barker and W. Rayens. "Partial Least Squares for Discrimination". In: *Journal of Chemometrics* 17.3 (2003), pp. 166–173. DOI: 10.1002/cem.785.

[123] R. Tibshirani. "Regression Shrinkage and Selection Via the Lasso". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (1996), pp. 267–288. DOI: 10.1111/j.2517-6161.1996.tb02080.x.

[124] A. E. Hoerl and R. W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems". In: *Technometrics* 12.1 (1970), pp. 55–67. DOI: 10.1080/00401706.1970.10488634.

[125] H. Zou and T. Hastie. "Regularization and Variable Selection via the Elastic Net". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2 (2005), pp. 301–320. DOI: 10.1111/j.1467-9868.2005.00503.x.

[126] J. Friedman, T. Hastie, and R. Tibshirani. "Regularization Paths for Generalized Linear Models via Coordinate Descent". In: *Journal of Statistical Software* 33.1 (2010), pp. 1–22. DOI: 10.18637/jss.v033.i01.

[127] R. Tibshirani et al. "Diagnosis of Multiple Cancer Types by Shrunken Centroids of Gene Expression". In: *Proceedings of the National Academy of Sciences* 99.10 (2002), pp. 6567–6572. DOI: 10.1073/pnas.082099299.

[128] B. D. Ripley. *Pattern Recognition and Neural Networks*. Cambridge, Massachusetts, USA: Cambridge University Press, 2008. ISBN: 978-0-521-71770-0.

[129] T. Hastie, R. Tibshirani, and A. Buja. "Flexible Discriminant Analysis by Optimal Scoring". In: *Journal of the American Statistical Association* 89.428 (1994), pp. 1255–1270. DOI: 10.2307/2290989.

[130] L. Breiman et al. *Classification and Regression Trees*. Reprint. The Wadsworth Statistics Probability Series. Boca Raton, Florida, USA: Chapman & Hall/CRC, 1998. ISBN: 978-0-412-04841-8.

[131] T. Hothorn, K. Hornik, and A. Zeileis. "Unbiased Recursive Partitioning: A Conditional Inference Framework". In: *Journal of Computational and Graphical Statistics* 15.3 (2006), pp. 651–674. DOI: 10.1198/106186006X133933.

[132] B. Efron and R. Tibshirani. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy". In: *Statistical Science* 1.1 (1986), pp. 54–75. DOI: 10.1214/ss/1177013815.

[133] J. Friedman. "Greedy Function Approximation: A Gradient Boosting Machine". In: *The Annals of Statistics* 29.5 (2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451.

[134] J. Friedman, T. Hastie, and R. Tibshirani. "Additive Logistic Regression: A Statistical View of Boosting". In: *The Annals of Statistics* 28.2 (2000), pp. 337–407. DOI: 10.1214/aos/1016218223.

[135] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. 2nd ed. Springer Series in Statistics. New York, USA: Springer-Verlag, 2009. ISBN: 978-0-387-84858-7.

[136] L. Breiman. "Random Forests". In: *Machine Learning* 45 (2001), pp. 5–32. DOI: 10.1023/A:1010933404324.

[137] I. Guyon and A. Elisseeff. "An Introduction to Variable and Feature Selection". In: *Journal of Machine Learning Research* 3 (2003), pp. 1157–1182. DOI: 10.5555/944919.944968.

[138] W. S. Cleveland and S. J. Devlin. "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting". In: *Journal of the American Statistical Association* 83.403 (1988), pp. 596–610. DOI: 10.2307/2289282.

[139] M. Gevrey, I. Dimopoulos, and S. Lek. "Review and Comparison of Methods to Study the Contribution of Variables in Artificial Neural Network Models". In: *Ecological Modelling* 160.3 (2003), pp. 249–264. DOI: 10.1016/S0304-3800(02)00257-0.

[140] G. H. Golub, M. Heath, and G. Wahba. "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter". In: *Technometrics* 21.2 (1979), pp. 215–223. DOI: 10.2307/1268518.

[141] J. Friedman. "Stochastic Gradient Boosting". In: *Computational Statistics & Data Analysis* 38.4 (2002), pp. 367–378. DOI: 10.1016/S0167-9473(01)00065-2.

[142] G. Ridgeway. "Generalized Boosted Models: A Guide to the Gbm Package". In: (2020), p. 15. URL: https://cran.r-project.org/web/packages/gbm/vignettes/gbm.pdf (visited on 02/11/2021).

[143] L. Breiman. "Randomizing Outputs to Increase Prediction Accuracy". In: *Machine Learning* 40 (2000), pp. 229–242. DOI: 10.1023/A:1007682208299.

[144] R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2020. URL: https://www.R-project.org/ (visited on 02/11/2021).

[145] M. Kuhn and K. Johnson. *R Package "AppliedPredictiveModeling": Functions and Data Sets for Applied Predictive Modeling*. Version 1.1-7. 2018. URL: https://CRAN.R-project.org/package=AppliedPredictiveModeling (visited on 02/11/2021).

[146] R. A. Peterson and J. E. Cavanaugh. "Ordered Quantile Normalization: A Semiparametric Transformation Built for the Cross-Validation Era". In: *Journal of Applied Statistics* (2019), pp. 1–16. DOI: 10.1080/02664763.2019.1630372.

[147] M. Kuhn. *R Package "Caret": Classification and Regression Training*. Version 6.0-86. 2020. URL: https://CRAN.R-project.org/package=caret (visited on 02/11/2021).

[148] T. Wei and V. Simko. *R Package "Corrplot": Visualization of a Correlation Matrix*. Version 0.84. 2017. URL: https://github.com/taiyun/corrplot (visited on 02/11/2021).

[149] Microsoft Corporation and S. Weston. *R Package "doParallel": Foreach Parallel Adaptor for the 'parallel' Package*. Version 1.0.15. 2019. URL: https://CRAN.R-project.org/package=doParallel (visited on 02/11/2021).

[150] H. Wickham et al. *R Package "Dplyr": A Grammar of Data Manipulation*. Version 0.8.5. 2020. URL: https://CRAN.R-project.org/package=dplyr (visited on 02/11/2021).

[151] H. Wickham, J. Hester, and W. Chang. *R Package "Devtools": Tools to Make Developing R Packages Easier*. Version 2.3.2. 2020. URL: https://CRAN.R-project.org/package=devtools (visited on 02/11/2021).

[152] D. Meyer et al. *R Package "E1071": Misc Functions of the Department of Statistics, Probability Theory Group*. Version 1.7-3. 2019. URL: https://CRAN.R-project.org/package=e1071 (visited on 02/11/2021).

[153] S. Milborrow. *R Package "Earth": Multivariate Adaptive Regression Splines*. Version 5.1.2. 2019. URL: https://CRAN.R-project.org/package=earth (visited on 02/11/2021).

[154] H. Zou and T. Hastie. *R Package "Elasticnet": Elastic-Net for Sparse Estimation and Sparse PCA*. Version 1.3. 2020. URL: https://CRAN.R-project.org/package=elasticnet (visited on 02/11/2021).

[155] W. Chang. *R Package "Extrafont": Tools for Using Fonts*. Version 0.17. 2014. URL: https://CRAN.R-project.org/package=extrafont (visited on 02/11/2021).

[156] H. Wickham. *R Package "Forcats": Tools for Working with Categorical Variables (Factors)*. Version 0.5.0. 2020. URL: https://CRAN.R-project.org/package=forcats (visited on 02/11/2021).

[157] Microsoft and S. Weston. *R Package "Foreach": Provides Foreach Looping Construct*. Version 1.5.0. 2020. URL: https://CRAN.R-project.org/package=foreach (visited on 02/11/2021).

[158] A. Zeileis and Y. Croissant. "Extended Model Formulas in R: Multiple Parts and Multiple Responses". In: *Journal of Statistical Software* 34.1 (2010), pp. 1–13. DOI: 10.18637/jss.v034.i01.

[159] B. Greenwell et al. *R Package "Gbm": Generalized Boosted Regression Models*. Version 2.1.5. 2019. URL: https://CRAN.R-project.org/package=gbm (visited on 02/11/2021).

[160] H. Wickham. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer, 2016. ISBN: 978-3-319-24277-4.

[161] A. Kassambara. *R Package "Ggpubr": 'ggplot2' Based Publication Ready Plots*. Version 0.2.5. 2020. URL: https://CRAN.R-project.org/package=ggpubr (visited on 02/11/2021).

[162] N. Xiao. *R Package "Ggsci": Scientific Journal and Sci-Fi Themed Color Palettes for Ggplot2*. Version 2.9. 2018. URL: https://CRAN.R-project.org/package=ggsci (visited on 02/11/2021).

[163] J. B. Arnold. *R Package "Ggthemes": Extra Themes, Scales and Geoms for Ggplot2*. Version 4.2.0. 2019. URL: https://CRAN.R-project.org/package=ggthemes (visited on 02/11/2021).

[164] J. Friedman, T. Hastie, and R. Tibshirani. "Regularization Paths for Generalized Linear Models via Coordinate Descent". In: *Journal of Statistical Software* 33.1 (), pp. 1–22. DOI: 10.18637/jss.v033.i01.

[165] Revolution Analytics and S. Weston. *R Package "Iterators": Provides Iterator Construct*. Version 1.0.12. 2019. URL: https://CRAN.R-project.org/package=iterators (visited on 02/11/2021).

[166] T. Hastie and B. Efron. *R Package "Lars": Least Angle Regression, Lasso and Forward Stagewise*. Version 1.2. 2013. URL: https://CRAN.R-project.org/package=lars (visited on 02/11/2021).

[167] D. Sarkar. *Lattice: Multivariate Data Visualization with R*. New York: Springer, 2008. ISBN: 978-0-387-75968-5.

[168] G. Grolemund and H. Wickham. "Dates and Times Made Easy with Lubridate". In: *Journal of Statistical Software* 40.3 (2011), pp. 1–25. DOI: 10.18637/jss.v040.i03.

[169] S. M. Bache and H. Wickham. *R Package "Magrittr": A Forward-Pipe Operator for R*. Version 1.5. 2014. URL: https://CRAN.R-project.org/package=magrittr (visited on 02/11/2021).

[170] D. Bates and M. Maechler. *R Package "Matrix": Sparse and Dense Matrix Classes and Methods*. Version 1.2-18. 2019. URL: https://CRAN.R-project.org/package=Matrix (visited on 02/11/2021).

[171] H. Bengtsson. *R Package "matrixStats": Functions That Apply to Rows and Columns of Matrices (and to Vectors)*. Version 0.56.0. 2020. URL: https://CRAN.R-project.org/package=matrixStats (visited on 02/11/2021).

[172] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. 4th ed. Statistics and Computing. New York, USA: Springer-Verlag, 2002. ISBN: 978-0-387-21706-2.

[173] M. W. Beck. *Visualizing Neural Networks in R – Update*. 2013. URL: https://beckmw.wordpress.com/2013/11/14/visualizing-neural-networks-in-r-update/ (visited on 02/11/2021).

[174] S. Milborrow. *R Package "Plotmo": Plot a Model's Residuals, Response, and Partial Dependence Plots*. Version 3.5.7. 2020. URL: https://CRAN.R-project.org/package=plotmo (visited on 02/11/2021).

[175] J. Lemon. "Plotrix: A Package in the Red Light District of R". In: *R-News* 6.4 (2006), pp. 8–12. ISSN: 1609-3631.

[176] B.-H. Mevik. *R Package "Pls": Partial Least Squares and Principal Component Regression*. Version 2.7-2. 2019. URL: https://CRAN.R-project.org/package=pls (visited on 02/11/2021).

[177] H. Wickham. "The Split-Apply-Combine Strategy for Data Analysis". In: *Journal of Statistical Software* 40.1 (2011), pp. 1–29. DOI: 10.18637/jss.v040.i01.

[178] X. Robin et al. "pROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves". In: *BMC Bioinformatics* 12 (2011), p. 77. DOI: 10.1186/1471-2105-12-77.

[179] T. A. Gerds. *R Package "Prodlim": Product-Limit Estimation for Censored Event History Analysis*. Version 2019.11.13. 2019. URL: https://CRAN.R-project.org/package=prodlim (visited on 02/11/2021).

[180] L. Henry and H. Wickham. *R Package "Purrr": Functional Programming Tools*. Version 0.3.3. 2019. URL: https://CRAN.R-project.org/package=purrr (visited on 02/11/2021).

[181] H. Bengtsson. *R Package "R.Matlab": Read and Write MAT Files and Call MATLAB from Within R*. Version 3.6.2. 2018. URL: https://CRAN.R-project.org/package=R.matlab (visited on 02/11/2021).

[182] A. Liaw and M. Wiener. "Classification and Regression by randomForest". In: *R News* 2.3 (2002), pp. 18–22. ISSN: 1609-3631.

[183] A. Paluszynska, P. Biecek, and Y. Jiang. *R Package "randomForestExplainer": Explaining and Visualizing Random Forests in Terms of Variable Importance*. Version 0.10.1. 2020. URL: https://CRAN.R-project.org/package=randomForestExplainer (visited on 02/11/2021).

[184] E. Neuwirth. *R Package "RColorBrewer": ColorBrewer Palettes*. Version 1.1-2. 2014. URL: https://CRAN.R-project.org/package=RColorBrewer (visited on 02/11/2021).

[185] H. Wickham, J. Hester, and R. Francois. *R Package "Readr": Read Rectangular Text Data*. Version 1.3.1. 2018. URL: https://CRAN.R-project.org/package=readr (visited on 02/11/2021).

[186] M. Kuhn and H. Wickham. *R Package "Recipes": Preprocessing Tools to Create Design Matrices*. Version 0.1.12. 2020. URL: https://CRAN.R-project.org/package=recipes (visited on 02/11/2021).

[187] L. Henry and H. Wickham. *R Package "Rlang": Functions for Base Types and Core R and 'Tidyverse' Features*. Version 0.4.5. 2020. URL: https://CRAN.R-project.org/package=rlang (visited on 02/11/2021).

[188] H. Wickham. "Reshaping Data with the Reshape Package". In: *Journal of Statistical Software* 21.12 (2007), pp. 1–20. DOI: 10.18637/jss.v021.i12.

[189] T. Sing et al. "ROCR: Visualizing Classifier Performance in R". In: *Bioinformatics* 21.20 (2005), pp. 3940–3941. URL: http://rocr.bioinf.mpi-sb.mpg.de (visited on 02/11/2021).

[190] H. Wickham and D. Seidel. *R Package "Scales": Scale Functions for Visualization*. Version 1.1.0. 2019. URL: https://CRAN.R-project.org/package=scales (visited on 02/11/2021).

[191] U. Ligges and M. Maechler. "Scatterplot3d - an R Package for Visualizing Multivariate Data". In: *Journal of Statistical Software* 8.11 (2003), pp. 1–20. DOI: 10.18637/jss.v008.i11.

[192] A. W. Bowman and A. Azzalini. *R Package "Sm": Nonparametric Smoothing Methods*. Version 2.2-5.6. 2018. URL: http://www.stats.gla.ac.uk/~adrian/sm/ (visited on 02/11/2021).

[193] H. Wickham. *R Package "Stringr": Simple, Consistent Wrappers for Common String Operations*. Version 1.4.0. 2019. URL: https://CRAN.R-project.org/package=stringr (visited on 02/11/2021).

[194] G. Snow. *R Package "TeachingDemos": Demonstrations for Teaching and Learning*. Version 2.12. 2020. URL: https://CRAN.R-project.org/package=TeachingDemos (visited on 02/11/2021).

[195] K. Müller and H. Wickham. *R Package "Tibble": Simple Data Frames*. Version 2.1.3. 2019. URL: https://CRAN.R-project.org/package=tibble (visited on 02/11/2021).

[196] H. Wickham and L. Henry. *R Package "Tidyr": Tidy Messy Data*. Version 1.0.2. 2020. URL: https://CRAN.R-project.org/package=tidyr (visited on 02/11/2021).

[197] H. Wickham et al. "Welcome to the Tidyverse". In: *Journal of Open Source Software* 4.43 (2019), pp. 1–6. DOI: 10.21105/joss.01686.

[198] P. O. Perry. *R Package "Utf8": Unicode Text Processing*. Version 1.1.4. URL: https://CRAN.R-project.org/package=utf8 (visited on 02/11/2021).

[199] D. Adler and S. T. Kelly. *R Package "Vioplot": Violin Plot*. Version 0.3.5. 2020. URL: https://github.com/TomKellyGenetics/vioplot (visited on 02/11/2021).

[200] Mirai Solutions GmbH. *R Package "XLConnect": Excel Connector for R*. Version 1.0.1. 2020. URL: https://CRAN.R-project.org/package=XLConnect (visited on 02/11/2021).

[201] A. Zeileis and G. Grothendieck. "Zoo: S3 Infrastructure for Regular and Irregular Time Series". In: *Journal of Statistical Software* 14.6 (2005), pp. 1–27. DOI: 10.18637/jss.v014.i06.

[202] The MathWorks Inc. *MATLAB*. Version R2019b Update 4 (9.7.0.1296695). Natick, Massachusetts, USA.

[203] I. Guyon, J. Weston, and S. Barnhill. "Gene Selection for Cancer Classification Using Support Vector Machines". In: *Machine Learning* 46 (2002), pp. 389–422. DOI: 10.1023/A:1012487302797.

[204] H. F. Kaiser. "A Note on Guttman's Lower Bound for the Number of Common Factors". In: *British Journal of Statistical Psychology* 14.1 (1961), pp. 1–2. DOI: 10.1111/j.2044-8317.1961.tb00061.x.

[205] A. Kassambara. *Practical Guide To Principal Component Methods in R*. 1st ed. Multivariate Analysis II. Montpellier, France: STHDA, 2017. ISBN: 978-1-975721-13-8.

[206] European Commission. *Comission Implementing Regulation (EU) 2019/774 of 16 May 2019 Amending Regulation (EU) No 1304/2014 as Regards Application of the Technical Specification for Interoperability Relating to the Subsystem 'Rolling Stock — Noise' to the Existing Freight Wagons*.

[207] Bundestag. *Gesetz zum Verbot des Betriebs lauter Güterwagen und zur Änderung des Allgeminen Eisenbahngesetzes - Schienenlärmschutzgesetz*. 2017.

A Dew formation – calculation steps of Model 1

From a physical perspective, dew formation is a mass transfer of water vapour from humid air onto a surface. For calculation, the convective sensible and latent heat flux density are of importance. Convective sensible heat exchange is distinguished in free and forced convection depending mainly on the current wind speed. In both, laminar and turbulent flow conditions can occur. [74]

The following equations provide a step-by-step guide through the calculation. Firstly, basic physical quantities have to be computed. Most of them are nearly constant at the considered temperature and pressure ranges, many can be extracted by interpolation from tables, however the focus lies on an expression by equations due to much lower programming effort.

Kinematic viscosity of air [111]:

$$\nu = \frac{\mu}{\rho} \tag{63}$$

$$\mu = \frac{1.458 \cdot 10^{-6} \cdot T^{1.5}}{T + 110.4} \tag{64}$$

$$\rho = \frac{p}{T_\mathrm{v} \cdot R} \tag{65}$$

$$T_\mathrm{v} = T \cdot (1 + 0.6078 \cdot s) \tag{66}$$

where ν depicts the kinematic viscosity of air in $\mathrm{m^2 s^{-1}}$, μ states the dynamic viscosity of air in Pas, T is the air temperature in K, ρ is the density of moist air in $\mathrm{kgm^{-3}}$, p is the air pressure of the ambient air in Pa, T_v depicts the virtual air temperature (correction between moist and dry air) in K, R is the gas constant of dry air (287.0586 $\mathrm{Jkg^{-1}K^{-1}}$) and s states the specific humidity in $\mathrm{kgkg^{-1}}$.

Thermal expansion coefficient of air in units $\mathrm{K^{-1}}$ [111]:

$$\alpha = \frac{1}{T} \tag{67}$$

Thermal conductivity of air in units $\mathrm{Wm^{-1}K^{-1}}$ [111]:

$$k = 0.02624 \cdot \left(\frac{T}{300}\right)^{0.8646} \tag{68}$$

Specific thermal capacity of air at constant pressure in units $\mathrm{Jkg^{-1}K^{-1}}$ [111]:

$$c_\mathrm{a} = 1002.5 + 275 \cdot 10^{-6} \cdot (T - 200)^2 \tag{69}$$

Thermal diffusivity of air in units $\mathrm{m^2 s^{-1}}$ [74]:

$$\kappa = \frac{k}{\rho \cdot c_\mathrm{a}} \tag{70}$$

The magnitude of free and forced convection can be described by the Nusselt number. In the former case, the derivation is based on the Reynolds number and in the latter on the

Grashof number. [74]

The Reynolds number is defined as follows [74]:

$$Re = u \cdot \frac{d}{\nu} \tag{71}$$

where Re depicts the Reynolds number (dimensionless), u states the wind speed in ms^{-1} and d is a characteristic length of the surface in flow direction with units m.

The Nusselt number for forced convection is derived for a flat plate in laminar conditions (Re $< 2{\cdot}10^4$) and turbulent conditions (Re $\geq 2{\cdot}10^4$) with the following equations [74]:

$$Nu_{\text{forced,lam}} = 0.6 \cdot Re^{0.5} \tag{72}$$

$$Nu_{\text{forced,turb}} = 0.032 \cdot Re^{0.8} \tag{73}$$

where $Nu_{\text{forced,lam}}$ depicts the Nusselt number for forced convection with laminar flow conditions and $Nu_{\text{forced,turb}}$ states the same in the turbulent case – both dimensionless.

The Grashof number is calculated with formula (74) [74]. In case of simultaneous heat and water vapour transfer, the quantity is calculated with the virtual temperature difference between surface and ambient air [74]. To compute the virtual temperature of the rail, the air temperature in Equation (66) is replaced by the rail temperature.

$$Gr = \frac{\alpha \cdot g \cdot d^3 \cdot |T_{\text{r,v}} - T_{\text{v}}|}{\nu^2} \tag{74}$$

where Gr is the Grashof number (dimensionless), g depicts the gravitational acceleration (9.80665 ms^{-2}) and $T_{\text{r,v}}$ states the virtual temperature of the rail in K.

With the Grashof number, the Nusselt number for free convection can be computed – again with the assumption of a flat plate. Laminar flow conditions are present for Gr $< 10^5$ and turbulent ones are the case with Gr $\geq 10^5$. The equations are as follows [74]:

$$Nu_{\text{free,lam}} = 0.5 \cdot Gr^{0.25} \tag{75}$$

$$Nu_{\text{free,turb}} = 0.13 \cdot Gr^{0.33} \tag{76}$$

where $Nu_{\text{free,lam}}$ states the Nusselt number for free convection in case of laminar flow conditions and $Nu_{\text{free,turb}}$ depicts the Nusselt number for free convection in turbulent ones – both dimensionless.

The threshold between free and forced convection cannot be set to one value because both mechanisms usually coexist. However, ranges, where one mechanism is the major contributor can be distinguished. One approach to determine limit values is to compute the ratio of the Grashof number and the Reynolds number with the following equation [74]:

$$C = \frac{Gr}{Re^2} \tag{77}$$

In general, if C is very low, forced convection dominates and if C is high, heat transfer is governed by free convection. The range in between is referred as mixed convection, where both mechanisms contribute significantly. The threshold values are set to [74]:

- C $\leq$ 0.1: only forced convection is included
- 0.1 $<$ C $<$ 10: mixed convection occurs and both Nusselt numbers are added up
- C $\geq$ 10: free convection dominates heat transfer and is therefore solely taken

However, it has to be pointed out that the threshold values are derived from a study with leaf boundary layers. Although the characteristic length of a leaf is on the same scale, the flow conditions may differ between a leaf and the rail head.

With the Nusselt number, the computation of the convective sensible heat flux density is possible, but for deriving dew formation not necessary and therefore skipped. To convert the sensible convective heat variables in latent heat variables, the calculated Nusselt number and the Lewis number can be used to derive the Sherwood number. The latter allows a computation of the latent heat flux density and hence the dew formation and evaporation rate for a wet surface. Again, the equations are split in free and forced convection and afterwards categorized in the same way as explained above. [74].

The Lewis number and the Sherwood number for free and forced convection are calculated with the following equations [74]:

$$Le = \frac{\kappa}{D_\mathrm{v}} \tag{78}$$

$$Sh_\mathrm{forced,lam} = Nu_\mathrm{forced,lam} \cdot Le^{\frac{1}{3}} \tag{79}$$

$$Sh_\mathrm{forced,turb} = Nu_\mathrm{forced,turb} \cdot Le^{\frac{1}{3}} \tag{80}$$

$$Sh_\mathrm{free,lam} = Nu_\mathrm{free,lam} \cdot Le^{0.25} \tag{81}$$

$$Sh_\mathrm{free,turb} = Nu_\mathrm{free,turb} \cdot Le^{\frac{1}{3}} \tag{82}$$

where Le depicts the Lewis number (dimensionless), D_v states the diffusivity of water vapour in air with units $\mathrm{m^2 s^{-1}}$ (can be calculated by linear interpolation over the temperature range in Table 4.17 in [83]) and Sh is the Sherwood number with indices for forced and free convection as well as each state in laminar and turbulent flow conditions (dimensionless).

The Sherwood number is used to compute a resistance to latent heat transfer (in other terms to water vapour transfer), which allows to estimate the dew formation rate depending on the difference of absolute humidity. The equations are as follows [74]:

$$r_{\mathrm{v}} = \frac{d}{D_{\mathrm{v}} \cdot Sh} \tag{83}$$

$$\frac{\mathrm{d}m}{\mathrm{d}t} = \frac{a - a_{\mathrm{r}}}{r_{\mathrm{v}}} \cdot 3600 \tag{84}$$

$$\tag{85}$$

where r_{v} depicts the resistance to water vapour transfer in sm^{-1}, $\frac{\mathrm{d}m}{\mathrm{d}t}$ states the dew formation rate in mmh^{-1}, a is the absolute humidity of the ambient air in kgm^{-3} and a_{r} is the absolute humidity at the rail in kgm^{-3}.

If the dew formation rate is positive, dew is formed on the surface – in case of a negative sign evaporation takes place.

B Basic statistic plots

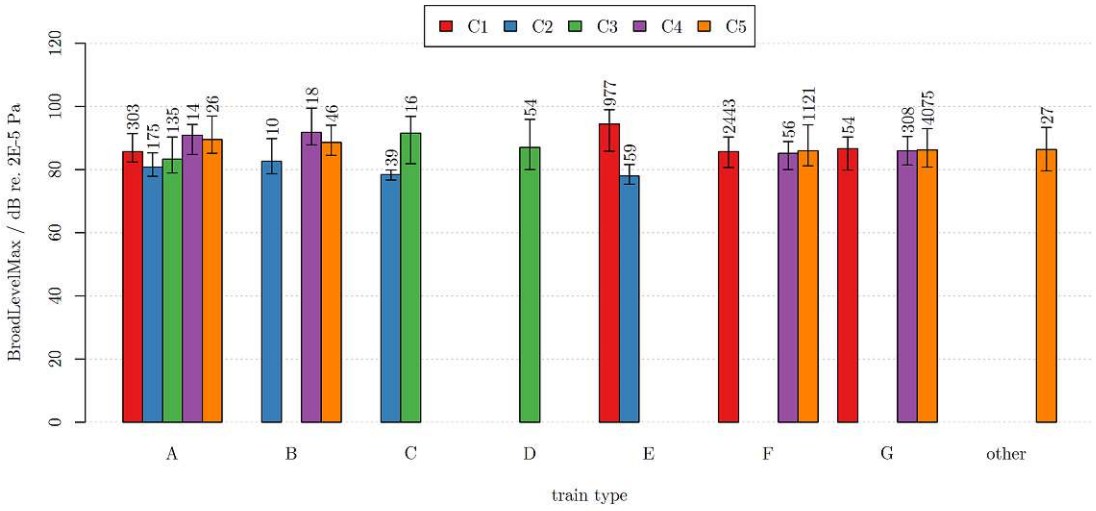

Figure 92: Median values and interquartile ranges for the dependent variable BroadLevelMax

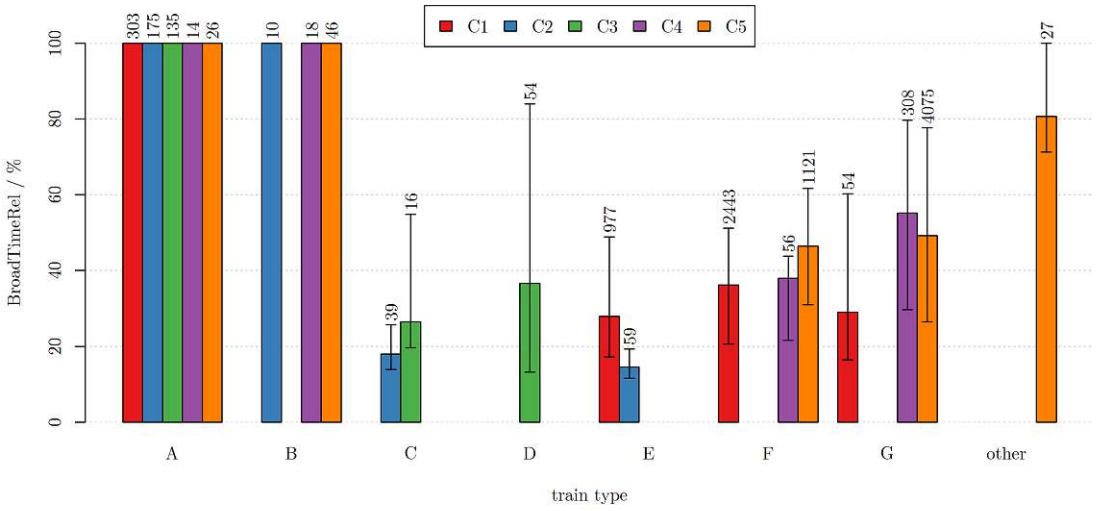

Figure 93: Median values and interquartile ranges for the dependent variable BroadTimeRel

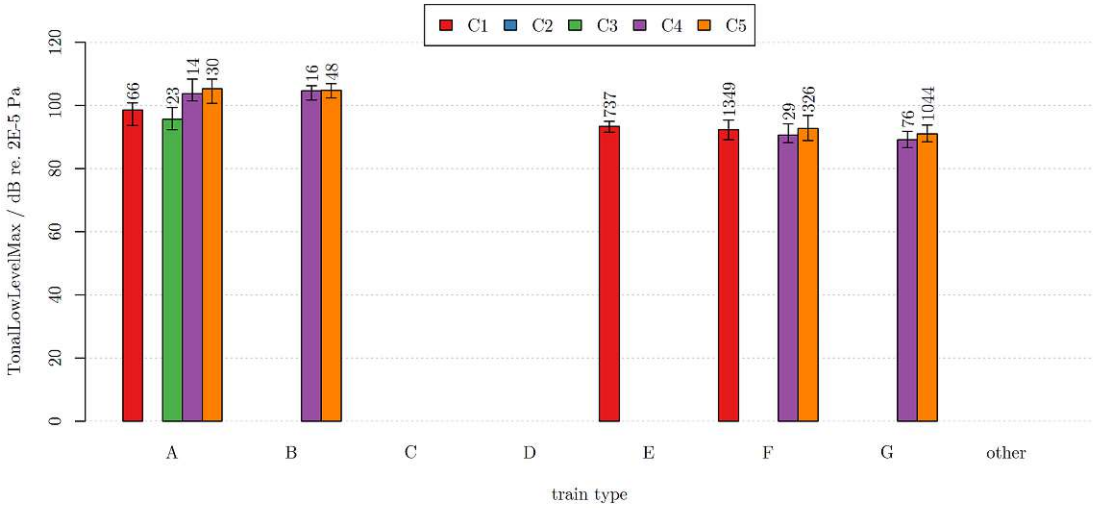

Figure 94: Median values and interquartile ranges for the dependent variable TonalLowLevelMax

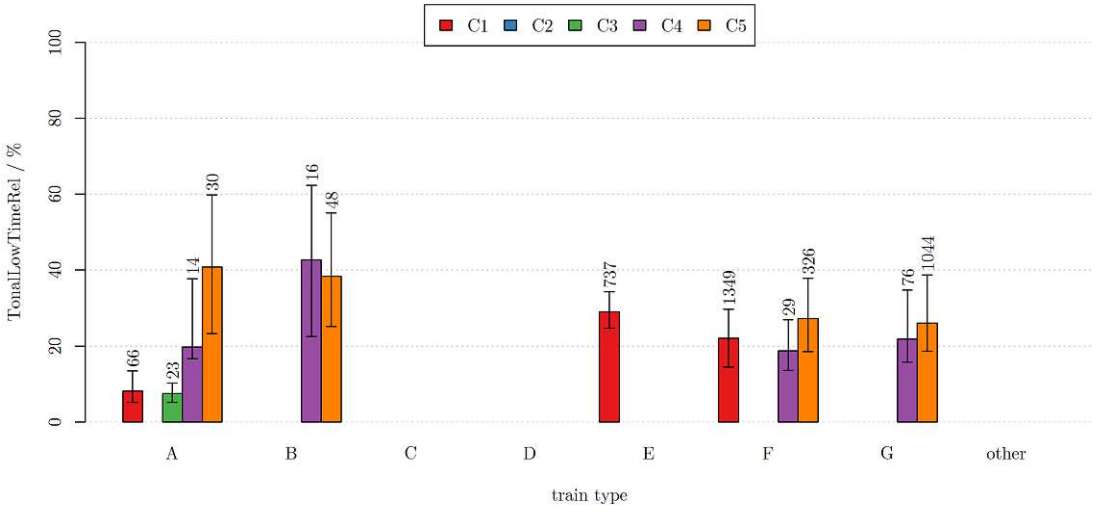

Figure 95: Median values and interquartile ranges for the dependent variable TonalLowTimeRel

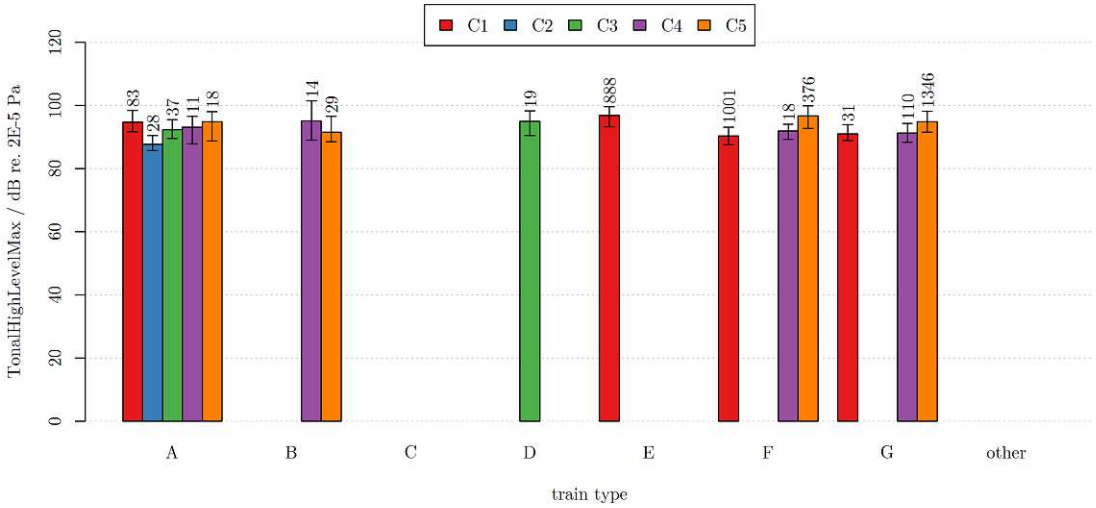

Figure 96: Median values and interquartile ranges for the dependent variable TonalHighLevelMax

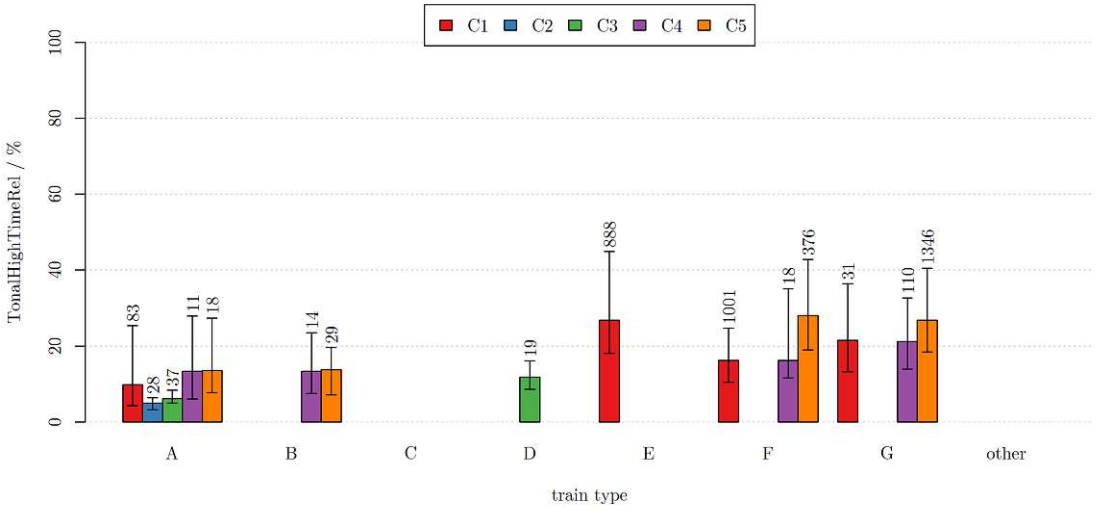

Figure 97: Median values and interquartile ranges for the dependent variable TonalHighTimeRel

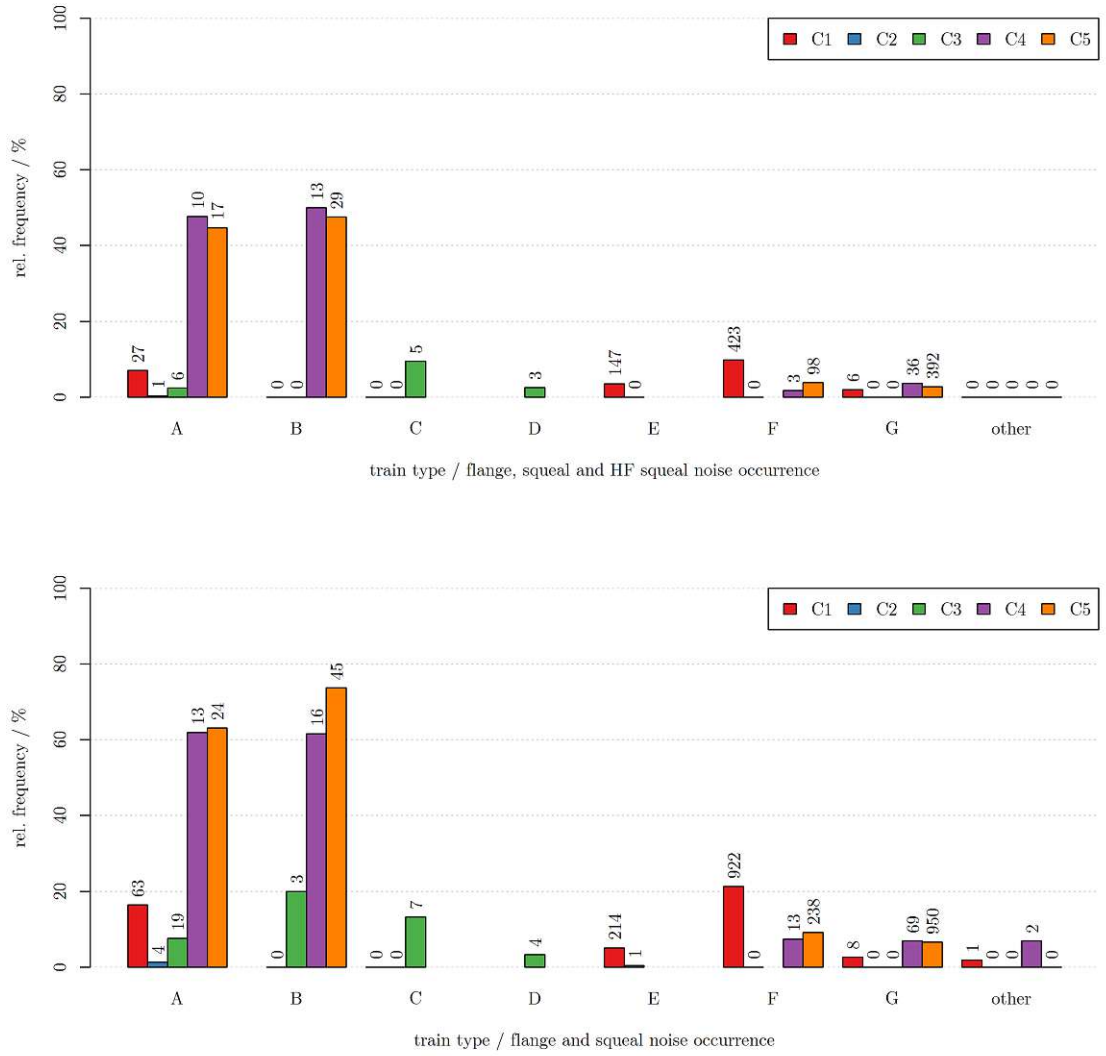

Figure 98: Relative frequency of occurrence regarding train pass bys, where multiple curve squeal types are detected in the same train pass by – top plot: all three types are observed; bottom graph: flanging and squeal noise are evaluated

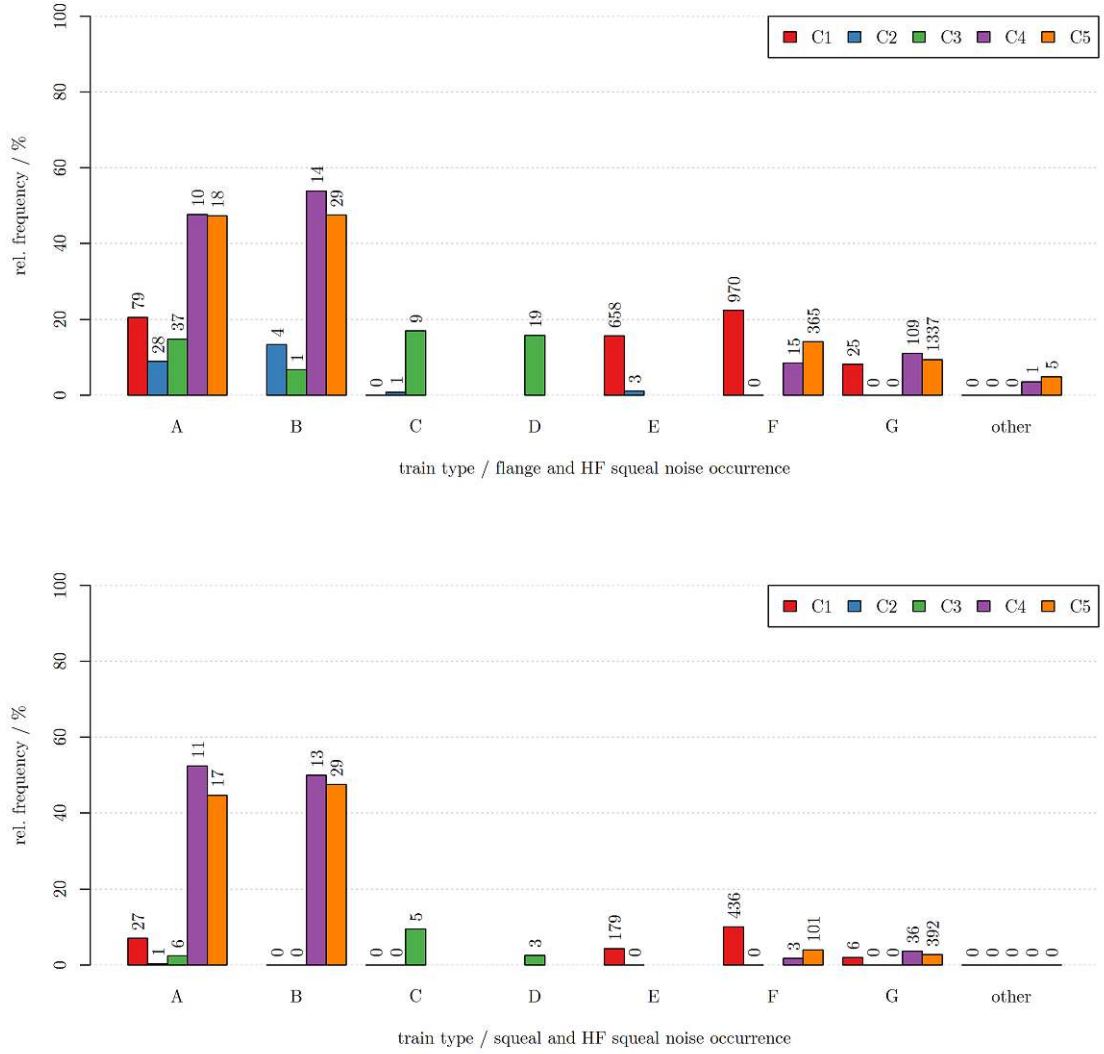

Figure 99: Relative frequency of occurrence regarding train pass bys, where multiple curve squeal types are detected in the same train pass by – top plot: flanging noise and HF squeal noise are assessed; bottom graph: squeal and HF squeal noise are identified

C Discussion on curve squeal peak level and relative occurrence time

C.1 Flanging noise peak level

Univariate statistical analyses in conjunction with flanging noise peak level (termed as BroadLevelMax) show a strong dependency on environmental conditions in the whole dataset (see Figure 100) as well as in most subsets. Top predictors are TempRail (and the correlated other temperature parameters), followed by relative humidity (top ranked RT.DPDiff) and absolute humidity variables. SunRad is also ranked with high importance. The listed vehicle predictor with the most importance is Loco.1. Track predictors (mainly RR and TDR values) and vehicle parameters, which are used to distinguish between sections and train types, are ranked in midfield. Dynamic predictors are listed with very low importance scores – regarding the whole dataset as well as subsets.

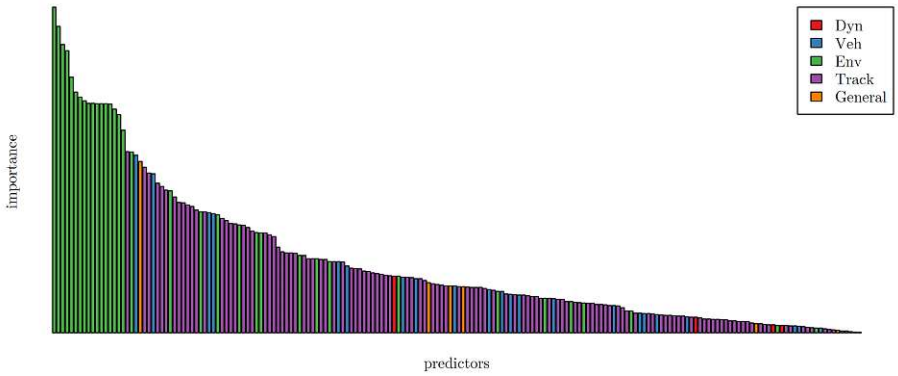

Figure 100: Predictor importance evaluated with linear variance analysis and coloured in groups dynamic (red), vehicle (blue), environmental (green), track (purple) and general (orange)

Due to partly distinct section and train type influence (refer to Figure 92), empirical investigations are focused on MC_TT subsets. As previously discussed in Subsection 5.7, negative correlation to curve radius is visible – foremost between C2 (440 m) and narrower radii ($\leq$ 310 m). Considering other track predictors, CurveBreathing shows a strong negative correlation to the outcome (empirically ranked as category 2 and 3) – caused by its dependence rail temperature (discussed below). However, curve breathing itself is not thought to have a direct connection to the outcome. Remaining track features show no general influence.

Regarding vehicle predictors, train type influence is especially strong in terms of train type E in C1, which shows the largest peak levels overall – even exceeding freight trains in C4/C5. Detailed investigation reveals no alternating external conditions compared to other categories, which leads to the assumptions that it is caused by an unknown train type specific property. Due to comparable levels with other categories in C2 (low statistical significance),

the cause may only refer to narrow radii. Train orientation shows a negligible (1 dB median difference) influence with controversial tendencies among train type F in C1 and C5. Double unit operation indicates hardly any influence with the similar peak levels as single units as well as no significant deviation among double unit configuration (median deviations up to 1 dB). Other vehicle predictors show no correlation to the outcome.

Dynamic predictors show no influence at all empirically, which complies with very low importance scores in univariate statistical analyses.

Deviations among the year are distinct, indicating a general tendency of significantly lower levels in summer months and high ones in winter (refer to Table 84). In C1, median values deviate in a magnitude of up to 17 dB (train type E) and up to 9 dB (train type F) between summer and winter months. The former category is generally raised in spring and autumn as well. Train types F and G in C5 support the trend with deviations between summer and winter months of up to 18 dB for the former and up to 12 dB for the latter. Observations among months result in a strong negative correlation (category 3) in conjunction with temperature and AT.DPDiff/RT.DPDiff (relative humidity indicators). Furthermore, category 2 negative correlations are classified in terms of absolute humidity predictors. Alterations over daytime are partly significant with median deviations compared to stated monthly values among hours of up to 9 dB in both directions. Generally, the majority of level increases are shown in morning and evening hours, whereas around midday and afternoon lower magnitudes are evaluated – only tendencies from train type G with low statistical significance. Thus, daytime variations also accord with indicated correlations regarding environmental predictors. It has to be pointed out that values may depend largely on climatic conditions at different locations. Thus, monthly and daytime behaviour might deviate significantly from the illustration in other locations. Distinct influence is also observed in conjunction with frost and hoarfrost presence – increasing median peak levels in C5 in a magnitude of 10 dB (frost) and 12 dB (hoarfrost) regarding train type F and 8 dB (frost) and 9 dB (hoarfrost) in terms of train type G. Additionally, dew occurrence causes raises of 5 dB (train type F) and 2 dB (train type G) in the same section. Trends of dew and frost impact are also supported by train types E and F in C1. In general, it has to be pointed out that flanging noise occurrence is reduced in dew, frost and rain conditions (refer to Subsection 6.3.2). Thus, statistical significance for those correlations is mediocre (dew) to weak (frost). An estimation of rain influence among the mentioned subsets is not possible due to a greatly reduced frequency of occurrence and thus hardly any level values available. In spectral evaluations of equivalent continuous noise level (refer to Subsection 6.3.5), low temperatures indicate lower levels at high frequencies (deviation up to 7 dB in 8000 Hz and up to 10 dB in 10000 Hz third-octave bands). Moreover, dew occurrence lowers the radiated emission in those frequency bands. However, flanging noise peak levels show the opposite behaviour in both observations. Thus, the excitation mechanism seems not connected. Moreover, higher attenuation coefficients of atmospheric absorption in winter (discussed in Subsection 6.3.5) indicate a contrary impact to the stated findings. Hence, the latter effect seems to be masked by a much more dominant impact mechanism, which causes the opposite trend. Previously stated assumptions of influence

subset	month											
	1	2	3	4	5	6	7	8	9	10	11	12
MC_1_TT_E	97	94	96	96	93	80	82	–	–	93	95	92
MC_1_TT_F	89	87	85	84	82	80	83	–	–	88	88	87
MC_5_TT_F	100	90	83	85	84	82	85	83	84	89	88	95
MC_5_TT_G	95	88	85	85	84	86	84	83	83	85	88	93

Table 84: Monthly median values of BroadLevelMax over the year with colour theme for statistical significance – red (weak), orange (mediocre), green (strong)

regarding a frozen superstructure and frozen earth along the transmission path may be an explanation of raised levels in conjunction with frost occurrence (refer to Subsection 6.3.5). However, loss in mechanical damping property is not thought to impact high frequencies. Despite low to mediocre statistical significance, correlations are indicated by three different train types in two sections with deviating properties. Altogether, the functionality of the underlying mechanism regarding the distinct influence of environmental predictors temperature, relative and absolute humidity as well as frost and dew occurrence remains unknown.

Application of statistical algorithms reveals that complex rule-based and tree models outperform other approaches significantly – results of Cubist and Boosted Tree are provided in Table 85. Achieved benchmarks considering the whole dataset are 6.3 dB RMSE and 31 % correlation between prediction and original. Thus, the task cannot be modelled well even with complex algorithms. Not much deviations between benchmarks in the whole dataset and among subsets indicate that the model is capable of including different section and train type properties. Due to application of five times repeated 10-fold cross validation, illustrated benchmarks are a good indicator of predictive behaviour on the same boundary conditions. Predictor importance scores are illustrated in Figure 101 with a cut point of 25 % relative importance. Beside environmental and general predictors, only Loco.1 (in Boosted Tree), Apl and RSI are ranked. Thus, track and dynamic influence seems negligible or included into the model by other predictors. Loco.1 accounts foremost the high levels of train type E in C1. Therefore, inclusion is not preferable in a predictive model. Apl is compromised by train type E in C1 as well and shows no correlation at all empirically. Moreover, axle count and train length is impossible to know in a predictive task. Empirically, no direct influence of RSI is visible, however the ranking might be connected to the observed correlation regarding curve radius. However, inclusion of curve radius alone might be more beneficial, since train type dependencies in the dataset could be misleadingly interpreted statistically. Environmental influences dominate the built model in terms of direct predictors as well as connected general parameters. Rail temperature is top ranked in both models. The Cubist model also ranks absolute humidity quantities (highest HumSpec) and pressure predictors in midfield as well as relative humidity indicators (best Hum) with low importance. YearDay and Week are also listed, however due to dependencies on local climatic conditions, they are not useful in a predictive model. Pressure quantities are not thought to have a direct physical relation to the

subset	Boosted Tree			Cubist		
	RMSE	R^2	number of predictors	RMSE	R^2	number of predictors
All	6.28	0.30	127	6.27	0.31	97
MC_1	6.04	0.34	51	6.12	0.32	42
MC_2	4.66	0.24	35	4.80	0.21	22
MC_4	5.78	0.16	38	6.05	0.13	18
MC_5	6.44	0.28	49	6.35	0.31	53
TT_A	6.15	0.22	83	6.26	0.20	25
TT_E	7.31	0.30	47	7.53	0.27	35
TT_F	5.56	0.35	117	5.54	0.35	77
TT_G	6.37	0.25	123	6.46	0.24	73
MC_1_TT_E	7.43	0.21	41	7.69	0.18	32
MC_1_TT_F	5.32	0.26	44	5.40	0.25	34
MC_4_TT_G	5.38	0.18	35	5.52	0.14	19
MC_5_TT_F	6.02	0.44	43	5.73	0.49	43
MC_5_TT_G	6.46	0.25	42	6.53	0.24	46

Table 85: Results of best performing regression models achieved with Boosted Tree and Cubist

outcome. Moreover, absolute air pressure is solely selected by the algorithm for distinction between sections and not useful in a general prediction model. The indicated correlation regarding relative air pressure is thought to be representable by temperature and humidity values.

To conclude, track influence is only visible in terms of curve radius – showing a negative correlation to the outcome, especially between 310 m and 440 m, while the range between 226 m and 310 m indicates only marginal deviations. Regarding vehicle predictors, general differences among train types are partly observable, however their magnitude is only relevant in terms of train type E in C1 and the occurring contrast between freight and passenger traffic in C4/C5. Thus, a predictive model might not need train type distinction predictors. Other vehicle properties as well as dynamic influence seem to have negligible impact. Alterations among the year are distinct and connectable to alterations in environmental conditions. A general tendency of significantly lower levels in summer months and high ones in winter is indicated (deviation magnitude of up to 9 dB to 18 dB depending on train type). Moreover, the majority of level increases are shown in morning and evening hours, whereas around midday and in the afternoon lower magnitudes are evaluated. Both tendencies result in negative correlations to rail temperature, absolute and relative humidity quantities. Detailed empirical investigation reveals a distinct influence of dew presence (raise in a magnitude of 2 dB to 5 dB) as well as frost occurrence (increases of 8 dB to 12 dB). Assumptions for the latter case are bonding effects between gravel stones, which cause on the one hand a loss in mechanical damping property of the superstructure, however this is not thought to impact high frequencies. On the other hand, a reduced porosity in

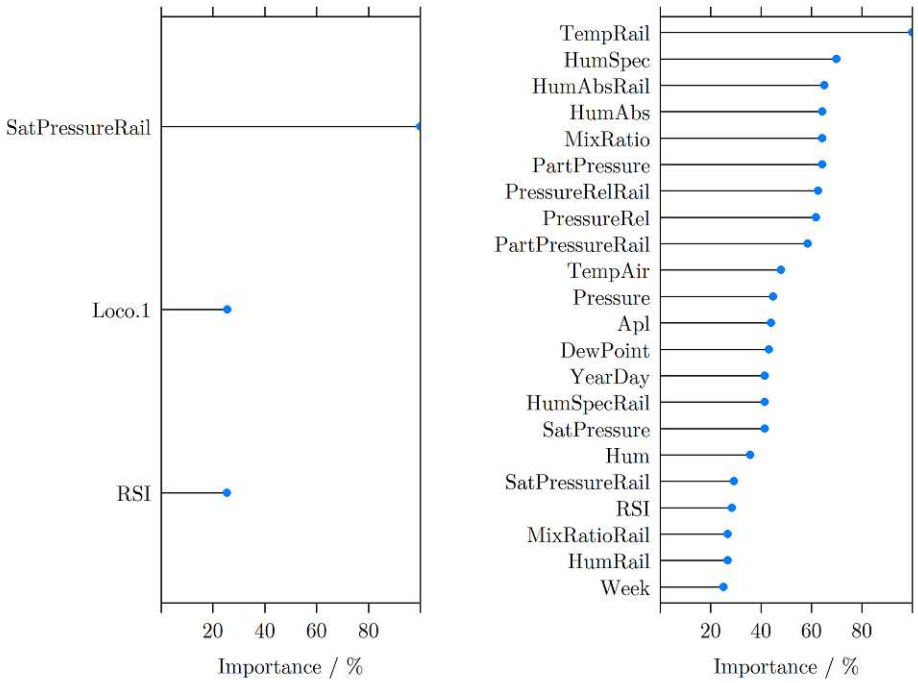

Figure 101: Relative predictor importance in the Boosted Tree (left) and Cubist (right) model fitted to the whole dataset

the ballast bed might occur in combination with frozen ground between ballast bed and microphone, which both cause a reduced surface absorption along the transmission path. The general trend is contrary to observed sound emission behaviour among the temperature range as well as in conjunction with dew occurrence in high frequencies evaluated in spectral analyses of equivalent continuous noise level (refer to Subsection 6.3.5). Observed deviations of attenuation coefficients between winter and summer indicate also an opposite trend. Thus, the latter impact is masked by another more dominant influence. Hence, excitation mechanisms between equivalent continuous noise level and observed alterations in the present evaluation seem not connected as well as remain unclear.

Building on the results of empirical and statistical investigation, the following predictors have the most impact on the outcome:

- curve radius
- rail temperature
- specific humidity (also takes absolute air pressure into account)
- relative humidity
- wet conditions (rain and/or dew present)
- frost conditions (frost or hoarfrost likely)

Due to poor benchmarks of multivariate statistical analysis and building on the scope of practical applicability, development of a prediction model for flanging noise peak levels is not seen as meaningful. However, due to unknown reasons behind observed correlations as well as no scientific proof for stated assumptions, further research need is indicated by the present evaluation. For rough estimation of flanging noise peak levels, median values from Figure 92 for similar train types and radii can be applied.

C.2 Relative occurrence time of flanging noise

Univariate statistical analysis of relative occurrence time regarding flanging noise (referred as BroadTimeRel) evaluates train type dependencies as most important – more detailed predictor TrainType.1 (illustrated in Figure 102). The reason behind is that all freight trains show a relative occurrence time of 100 % (see Figure 93). Thus, if a freight train emits flanging noise, it continues throughout the whole train pass by. Beside that dominating correlation, predictor ranks in midfield represent mostly track influence (TDR and RR) as well as other variables (e.g. Pressure) to distinguish between sections. Exceptions are AccSide (ranked in fifth place overall) and TrainOrientation.1. The former is compromised due to lower operation speeds of freight trains in C1, while regarding the latter an influence is also detected empirically (discussed below). General and environmental predictors are ranked with low importance.

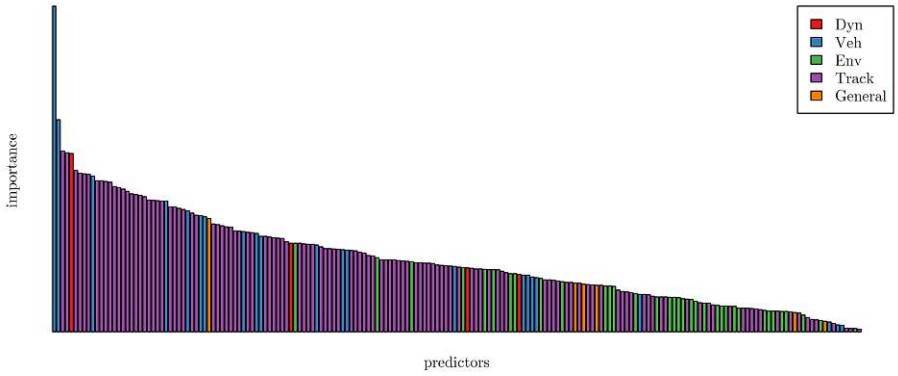

Figure 102: Predictor importance evaluated with linear variance analysis and coloured in groups dynamic (red), vehicle (blue), environmental (green), track (purple) and general (orange)

Due to alterations among train types and sections, empirical analysis is focused on subsets MC_TT. Evaluations of track predictors indicate a distinct decrease between 310 m and 440 m curve radius (only train type C evaluable with 9 % decline) and a partly visible overall negative correlation of radius – dependent on train type. Train types F and G show a decrease in median values between 230 m and 256 m of 10 % and 20 %, while passenger trains in C3 (310 m) show a controversial behaviour with partly higher values compared to 256 m radius.

Radial influence corresponds only to passenger trains as freight trains show 100 % occurrence time – even at 440 m radius. CurveBreathing shows a partly distinct negative correlation to the outcome, which indicates an influence of rail temperature (further stated below). The predictor itself is not thought to have a direct connection to the outcome.

Regarding train type deviations, median values among passenger trains in narrower curve radii ($\leq$ 310 m) range from 27 % to 55 %, while a larger radius (440 m) leads to median values of 15 % to 18 %. Train category other shows a significantly higher relative occurrence time with 81 %. Thus, its behaviour corresponds rather to the freight train category. Pushed train operation causes significantly higher values (difference in median to pulled 8 % to 12 %). The trend is supported by all statistically significant subsets (train type E in C1 and train type F in C1 and C5). Double units show up to 20 % lower values compared to single unit (pushed) operation. Between configurations in double units, up to 8 % deviation in median occurs (worst case TrainOrientation.6 – both power coaches pushing their trainsets – and lowest in TrainOrientation.5 – both power units in the middle). Thus, train orientation can have a significant effect in single and double unit operation. Other vehicle predictors show no correlations to the outcome.

Dynamic predictors indicate no influence. Overall, statistical tendency of AccSide is compromised by freight trains in C1, which operate at lower speed (hence also lower lateral acceleration). Thus, a direct connection is not present.

Alterations among the year are distinct. Deviations among months indicate extreme values in January and generally raised ones from October to December (see Table 86) – indicated by all considered subsets with deviating magnitudes. Excluding extremes in January (median 58 % to 98 %), variations among other months reach up to 18 % to 20 % in C1 and up to 28 % to 31 % in C5. A partly weak statistical significance for those values has to be borne in mind. Alterations in daytime are partly large (around 20 % in both directions for train type G in C5), however they do not seem to follow a certain pattern like peak level or frequency of occurrence. It has to be pointed out that values may depend largely on climatic conditions at different locations. Thus, behaviour among the year might deviate significantly from the illustration in other locations. Regarding environmental predictors, empirical investigation finds a visible to strong negative correlation in conjunction with rail temperature (category 2 and 3). Moreover, the relative humidity quantity RT.DPDiff also indicates a partly visible negative correlation. Dew as well as frost presence show a positive correlation, however magnitude largely depends on the subset. Impact of dew occurrence ranges from hardly any (train type G in C5) to up to 15 % (train type E in C1) increase, while frost conditions cause a raise of 4 % to 10 %. Due to the mediocre increase compared to deviations between winter and other seasons, frost might not be the only underlying mechanism. However, frost occurrence partly explains magnitude deviation between C1 and C5 in December due to vastly higher train pass bys in C5 (66 %) under that condition compared to C1 (2.8 %). Due to significant reduction in frequency of occurrence in conjunction with rain, evaluation is not possible due to a lack of data points. In general, high interquartile ranges are present, which causes an uncertainty in stated values.

subset	month											
	1	2	3	4	5	6	7	8	9	10	11	12
MC_1_TT_E	61	21	35	26	21	19	21	–	–	39	38	34
MC_1_TT_F	58	34	34	32	38	27	39	–	–	45	42	34
MC_5_TT_F	70	35	38	45	43	40	45	34	48	54	46	63
MC_5_TT_G	98	55	34	44	38	48	41	37	47	51	45	65

Table 86: Monthly median percentage values of BroadTimeRel over the year with colour theme for statistical significance – red (weak), orange (mediocre), green (strong)

Application of statistical model algorithms shows that complex tree and rule-based models – Boosted Tree and Cubist – significantly outperform other approaches (results depicted in Table 87). The only exception states C2, where a MARS model using only two predictors (TrainType.1 and TrainType.2) achieves the best overall performance. Thus, deviation between freight traffic (100 % relative occurrence time) and passenger trains (20 %) is huge enough to include only that issue. In the whole dataset, nearly 50 % correlation between prediction and original is achieved. However, a high fluctuation is previously indicated by high interquartile ranges, which also leads to a large RMSE value of 21 %. Not much deviations between benchmarks in the whole dataset and among subsets (except C2) indicate that the model is capable of including different section and train type properties. Due to application of five times repeated 10-fold cross validation, illustrated benchmarks are a good indicator of predictive behaviour on the same boundary conditions.

Predictor importance scores with a cut point of 25 % relative importance are depicted in Figure 103. While the Boosted Tree model is focused on the predominant influence of freight trains, the Cubist model shows a different approach. The former is caused by tree correlation, since every generated tree is dependent on the tree in the previous iteration. Thus, correlating predictors are magnified in each step (further discussed in Subsection 6.1). The Cubist model depends mostly on environmental influence. Beside CurveBreathing (connected to rail temperature), no track predictors are listed. Among different section, the latter is compromised due to deviating curve lengths, which are misleadingly interpreted statistically. Vehicle influence is accounted by Apl, MaxBogieDist and MaxAoA. All three allow a distinction between freight and passenger trains, however no general correlations are shown empirically. Regarding dynamic variables, three of them (Vmean, AccSide and AccTrain) are ranked with low importance. The former takes the difference between sections for passenger trains into account, while the latter two are compromised by distribution of data points. General predictors YearDay and Week are listed, which indicates a detailed, complex model fit – maybe a tendency to over-fitting. However, an inclusion of those in a predictive model is not meaningful due to location dependencies and yearly variations. Regarding environmental influence, relative air pressure is ranked top, followed by absolute humidity quantities, absolute air pressure, temperature predictors and relative humidity values. Due to lack of physical explanation, PressureRel is not thought to be directly connected and is rather represented by inclusion of other environmental predictors. Absolute

subset	Boosted Tree			Cubist		
	RMSE	R^2	number of predictors	RMSE	R^2	number of predictors
All	21.34	0.47	148	21.13	0.49	98
MC_1	18.10	0.57	52	18.02	0.57	45
MC_2	9.21	0.94	40	8.31	0.95	36
MC_4	23.40	0.39	41	23.65	0.38	48
MC_5	23.77	0.28	49	23.31	0.31	53
TT_E	21.03	0.21	43	21.08	0.23	31
TT_F	17.54	0.35	93	17.24	0.35	78
TT_G	25.28	0.23	121	25.20	0.25	94
MC_1_TT_E	21.43	0.19	37	21.56	0.20	31
MC_1_TT_F	17.22	0.37	38	17.23	0.37	40
MC_4_TT_G	25.33	0.23	35	26.55	0.17	23
MC_5_TT_F	17.43	0.39	44	17.05	0.42	43
MC_5_TT_G	25.23	0.23	42	24.87	0.26	49

Table 87: Results of best performing regression models achieved with Boosted Tree and Cubist

air pressure allows a distinction between sections, however it also indicates no general influence. Dependencies on environmental conditions comply with empirical findings.

To conclude, track influence is only observed in conjunction with curve radius. Passenger trains show a significant decrease between $\leq$ 310 m and 440 m (from 27 % - 55 % to 15 % - 18 %), while narrower radii ($\leq$ 310 m) indicate a decline between 230 m and 256 m of 10 % to 20 % and no clear pattern between 256 m and 310 m. Freight trains (train type A and B) show 100 % relative occurrence time in every section. Beside deviations among train types, pushed train operation causes increases of 8 % to 12 %. Double units show deviations among their configuration of 8 % (worst case both power coaches pushing their trainsets and best case both power units in the middle) and up to 20 % lower values compared to single unit (pushed) operation. Dynamic influence is not indicated. Distinct alterations among the year are observed. Especially January indicates extreme values (58 % to 98 %). It states the coldest month of the year with frequent frost occurrence and high relative humidity. Apart from that, raised levels from October to December are observed. Variations among months (excluding January) reach up to 18 % to 20 % in C1 and up to 28 % to 31 % in C5. A partly weak statistical significance for those values has to be borne in mind. However, unlike frequency of occurrence and peak level of flanging noise, relative occurrence time show no certain pattern in daytime behaviour. Regarding environmental predictors, temperature (especially rail temperature), absolute humidity and relative humidity quantities show correlations to the outcome – empirically and statistically. Impact of dew occurrence ranges from hardly any to up to 15 % increase depending on the subset and frost conditions cause raises of 4 % to 10 %. Trends of seasonal and environmental impact show similarities with observed findings regarding peak level of flanging noise. Thus,

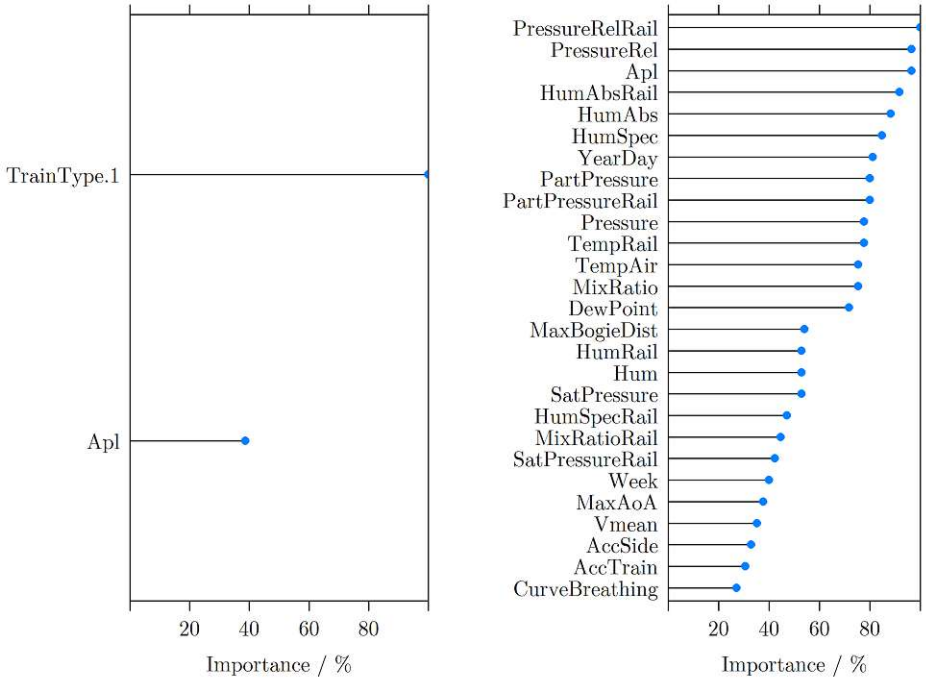

Figure 103: Relative predictor importance in the Boosted Tree (left) and Cubist (right) model fitted to the whole dataset

reasons might be related. It has to be pointed out that due to high interquartile ranges in subsets, a general uncertainty is present regarding the stated trends and their magnitude. Reasons behind observed findings are unknown, which indicates need of further research.

Building on the results of empirical and statistical investigation, the following predictors have the most impact on the outcome:

- train type based on observed relative occurrence time – high (train types A, B, other), mediocre (train types C, D, E, F, G)
- double unit operation
- train orientation (in both single and double unit operation)
- rail temperature
- specific humidity (also takes absolute air pressure into account)
- relative humidity
- wet conditions (rain and/or dew present)
- frost conditions (frost or hoarfrost likely)

Due to high fluctuations in the dataset (resulting in high RMSE values and a general uncertainty) and pursuing the scope of constructing a practical model, a heuristic approach

is chosen rather than constructing a predictive model. Freight trains and train category other are set to 100 % relative occurrence time and all passenger train types are modelled with 45 % for curve radii less or equal than 310 m and afterwards using a linear regression line of $109.385 - 0.2077 \cdot R$ for calculation (using the point at 440 m and 18 % for derivation). To prevent a value of zero for larger radii (would happen at 527 m), values for radii greater than 440 m are set to a constant of 18 %. It has to be pointed out that this is a try to model a simplified relationship empirically, which needs further radii for comparison and validation.

C.3 Squeal noise peak level

In general, it has to be pointed out that data points of squeal noise are sparse and in many subsets statistical threshold for inclusion is not met. The latter applies to train types C, D and other as well as to C2 (see Figure 94). Among remaining subsets, a clear trend of higher levels in conjunction with freight trains compared to passenger traffic is visible.

Univariate linear variance analysis of squeal noise peak level (termed TonalLowLevelMax) indicates a dominating vehicle influence with RSI in the top rank (depicted in Figure 104). The latter perfectly illustrates deviation between freight and passenger traffic (due to modelling freight trains with high wheel diameter) as well as slightly lower levels of train type G (lowest wheel diameter). Moreover, negative correlation to curve radius is also taken into account. Other vehicle predictors follow, however representing the same trends. Interestingly, RainDetect is listed in ninth place, followed by other environmental predictors in the upper segment. Dynamic influence (foremost AccTrain) is compromised by data point distribution. Beside track parameters in midfield (RR and TDR) for distinction among sections, Hour is ranked relatively high. Further discussion is carried out in the next paragraph in combination with empirical findings.

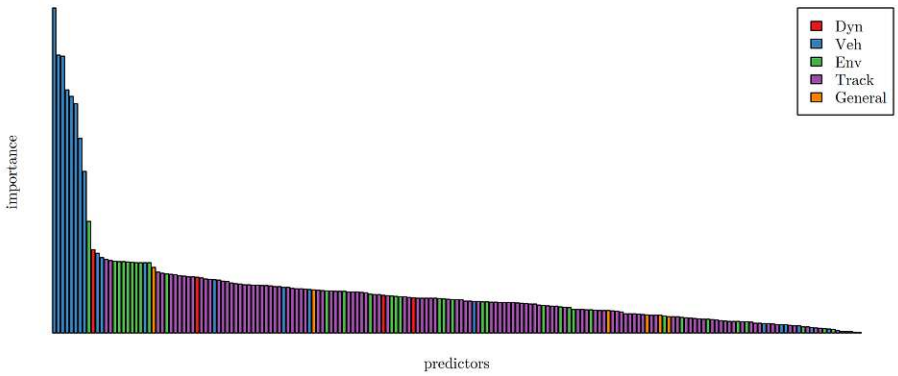

Figure 104: Predictor importance evaluated with linear variance analysis and coloured in groups dynamic (red), vehicle (blue), environmental (green), track (purple) and general (orange)

Empirically, track influence is only indicated in conjunction with a negative correlation of curve radius, however only regarding freight trains. It has to be pointed out that due to a lower frequency of occurrence of squeal noise in larger radii (> 256 m), no statement can be given for passenger trains. Additionally, CurveBreathing shows a negative correlation (category 2). However, it depends on rail temperature (further discussed below) and no general correlation to the outcome is evaluated. Train type influence is distinct. Beside the increased levels of freight trains in general and especially in C4/C5, slight differences among passenger train categories are observed with median peak levels of 93 dB (train type E), 91 dB to 93 dB (train type F) and 89 dB to 91 dB (train type G). Thus, especially train type G emits slightly lower amplitudes – maybe connected to the low wheel diameter. Freight trains in C4/C5 show a notably low Apl value of about 0.14 compared to monitored freight trains in other sections, which indicates a large distance between bogies or axles worsening the curving behaviour and presumably also increasing angle of attack. According to past research, the latter shows a positive correlation to the outcome [37]. Train orientation in single units shows a controversial trend among the dataset. While train type E in C1 and train type F in C5 indicate a raise of 1 dB and 2 dB due to pushed operation, regarding train type F in C1 a decrease of 2.5 dB is observable. Configuration in double units indicate deviations of up to 2 dB (worst case both power coaches in the middle and best case both power units pushing their trainsets). Absolute peak levels remain in the area of single units. Altogether, train configuration shows a partly controversial trend with mostly negligible differences. All vehicle predictors, which allow distinction between passenger and freight traffic (Apl, MaxAoA, MaxWheelDm, RSI, FreeCurving) show mediocre or strong correlation to the outcome in the whole dataset as well as section and train type (only TT_A) subsets. Influence of dynamic predictors is not visible in empirical examination.

Seasonal and monthly alterations are observable. Further distinction in weeks and days of the year show not much fluctuation, however threshold for statistical significance in the latter is not often reached. Monthly median level distribution among the year is illustrated in Table 88. As can be seen, trends are not fully agreeing between subsets – presumably due to influence of mostly weak statistical significance. However, following the pattern of train types F and G in C5, raised levels occur between October and January. A weak trend to lower levels in summer months is partly present. In C1, no clear tendencies are shown. Altogether, deviations among months are large with up to 9 dB and 3 dB in C1 (train types E and F) as well as up to 9 dB and 4 dB in C5 (train types F and G). Daytime behaviour cannot be evaluated due to a generally low frequency of occurrence in terms of squeal noise and hence not enough monthly data points among hours. It has to be pointed out that values may depend largely on climatic conditions at different locations. Thus, behaviour among the year might deviate significantly from the illustration in other locations. Regarding environmental predictors, TempRail, SunRad, absolute humidity quantities and RT.DPDiff all indicate a negative correlation (category 2 and 3), while HumRail shows a positive one. The mentioned tendencies are distinct in terms of train types E in C1 and G in C5, whereas train type F shows generally weak correlations in both sections. Dew presence seems to have a negligible effect on peak levels, while frost presence causes an increase in median

subset	month											
	1	2	3	4	5	6	7	8	9	10	11	12
MC_1_TT_E	–	93	94	99	–	90	91	–	–	93	98	93
MC_1_TT_F	92	91	94	93	94	92	93	–	–	93	93	92
MC_5_TT_F	98	90	–	90	–	94	94	89	92	94	96	93
MC_5_TT_G	94	92	91	90	91	89	88	89	89	92	92	93

Table 88: Monthly median levels of TonalLowLevelMax over the year with colour theme for statistical significance – red (weak), orange (mediocre)

level of 2 dB to 5 dB – indicated by all considered subsets. Assumed reasons behind the latter are equal to stated ones in Subsection 6.3.5 – generally lower surface absorption in the transmission path. Loss in mechanical damping property of the superstructure is not seen as important in conjunction with frequencies between mostly 1250 Hz and 6300 Hz, where vehicle contribution dominates rolling noise [92]. Due to hardly any occurrence of squeal noise in rain conditions, no statement is possible.

Application of statistical algorithms reveals that complex tree and rule-based models – Boosted Tree and Cubist – achieve the best overall results (see Table 89). Considering the whole dataset, Cubist performs best, however regarding subsets no clear winner among these two can be stated. RMSE values are mostly around 4 dB and correlation between prediction and original is 35 % in the whole dataset. While the former is mediocre, the latter indicates a rather bad accordance. Not much deviations between benchmarks in the whole dataset and among subsets (except train type F especially in C1) indicate that the model is capable of including different section and train type properties. Due to application of five times repeated 10-fold cross validation, illustrated benchmarks are a good indicator of predictive behaviour on the same boundary conditions.

Predictor importance scores (depicted in Figure 105 with 25 % relative importance cut point) show RSI as top ranked predictor, which perfectly complies with univariate statistical and empirical analysis. RSI allows a distinction between freight and passenger traffic as well as inclusion of negative correlation to curve radius. While in Boosted Tree MaxAoA and Apl follow (for the same purpose), Cubist ranks many environmental predictors as high to mediocre important – especially absolute humidity quantities. Interestingly, TempAir shows a significantly higher importance score compared to TempRail, which is contrary to empirical examination. No relative humidity quantity is ranked, instead, relative air pressure achieves a high rank. Due to lack of physical explanation, it is not thought to be directly connected and is rather represented by inclusion of other environmental predictors. The listing of YearDay and Week indicates over-fitting – especially due to lack of visible alterations empirically. Both are generally not meaningful to include in a predictive model due to location dependencies and yearly variations. Track influence as well as dynamic predictors are included with low importance and only used to distinguish between sections.

subset	Boosted Tree			Cubist		
	RMSE	R^2	number of predictors	RMSE	R^2	number of predictors
All	4.05	0.33	140	4.03	0.35	87
MC_1	4.00	0.20	52	3.99	0.22	35
MC_5	3.99	0.43	51	4.05	0.42	43
TT_E	3.57	0.38	42	3.49	0.41	37
TT_F	4.27	0.12	96	4.31	0.10	58
TT_G	3.45	0.24	86	3.56	0.21	35
MC_1_TT_E	3.60	0.37	39	3.52	0.40	37
MC_1_TT_F	4.05	0.08	40	4.09	0.06	28
MC_5_TT_F	5.00	0.24	38	5.02	0.26	14
MC_5_TT_G	3.47	0.23	39	3.54	0.22	33

Table 89: Results of best performing regression models achieved with Boosted Tree and Cubist

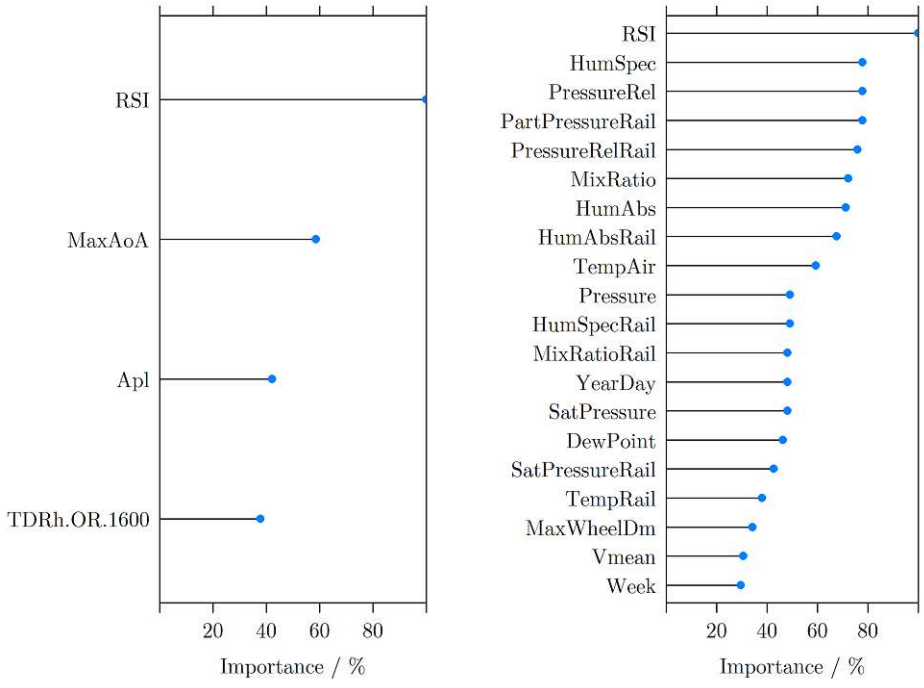

Figure 105: Relative predictor importance in the Boosted Tree (left) and Cubist (right) model fitted to the whole dataset

Track influence is only indicated by a negative correlation to curve radius. However, that trend is only shown in conjunction with freight trains and with low statistical significance. Peak level magnitude is dominated by vehicle dependencies. Freight trains show generally higher peak levels (extreme in C4 and C5 with differences in median values of up to 16 dB) compared to passenger train categories. Among the latter, train type G indicates lower levels (deviation 2 dB to 4 dB) than other passenger train types (only train types E and F evaluable). The best performing predictor is RSI, which allows distinction between freight and passenger traffic as well as takes radius dependency into account. Regarding dynamic influence, no direct connection to the outcome is observable. Seasonal and monthly variations are visible with no full agreement between subsets. Train types F and G in C5 indicate raised levels between October and January. A weak trend to lower values in summer months is partly present. In C1, no clear tendencies are shown. Altogether, deviations among months are large with up to 9 dB and 3 dB in C1 (train types E and F) as well as up to 9 dB and 4 dB in C5 (train types F and G). Daytime behaviour among months cannot be evaluated due to lack of statistical significant data (caused by a generally low frequency of occurrence of squeal noise). Altogether, seasonal dependencies are strongly correlated with temperature and absolute humidity, which is both statistically and empirically indicated. Empirical investigation finds an additional relation of relative rail humidity to the outcome. Furthermore, frost presence causes an increase in peak levels of 2 dB to 5 dB – indicated by all considered subsets, hence among three different train types in two sections with deviating properties. The assumed reason behind the latter is a generally lower surface absorption in the transmission path due to reduced porosity in the ballast bed and frozen earth between ballast bed and microphone (refer to Subsection 6.3.5). Dew occurrence shows a negligible effect and rain conditions cannot be evaluated due to hardly any squeal noise incidences.

Building on the results of empirical and statistical investigation, the following predictors have the most impact on the outcome:

- Train type distinguished in three categories depending on magnitude of observed peak levels – high (type A, B), mediocre (type C, D, E, F, other), low (type G)
- RSI
- rail temperature
- specific humidity (also takes absolute air pressure into account)
- relative rail humidity
- frost conditions (frost or hoarfrost likely)

Due to poor benchmarks of multivariate statistical analysis and building on the scope of practical applicability, development of a prediction model for squeal noise peak levels is not seen as meaningful. However, due to unknown reasons behind observed correlations as well as no scientific proof for stated assumptions, further research need is indicated by the present

evaluation. For rough estimation of squeal noise peak levels, median values from Figure 94 for similar train types and radii can be applied. Due to notably deviating freight train peak levels in C4 and C5 compared to other sections, a better estimation for narrow radii ($\leq$ 256 m) states presumably the value in C1 as freight traffic in the latter section shows a more heterogeneous distribution of Apl values.

C.4 Relative occurrence time of squeal noise

In general, it has to be pointed out that data points of squeal noise are sparse and in many subsets statistical threshold for inclusion is not met. The latter applies to train types C, D and other as well as to C2 (see Figure 95).

Univariate statistical analysis of relative occurrence time regarding squeal noise (referred as TonalLowTimeRel) ranks Apl with highest importance (see Figure 106). Detailed investigation shows that freight trains in C4 and C5 have significantly higher relative occurrence time compared to all other subsets and Apl values lie mostly around 0.14. Thus, statistical evaluations detects that trend, however a general correlation between Apl and the outcome is not found. The upper half is dominated by track predictors – an indication of distinct variations among sections. Other vehicle, dynamic, environmental and general parameters are ranked low.

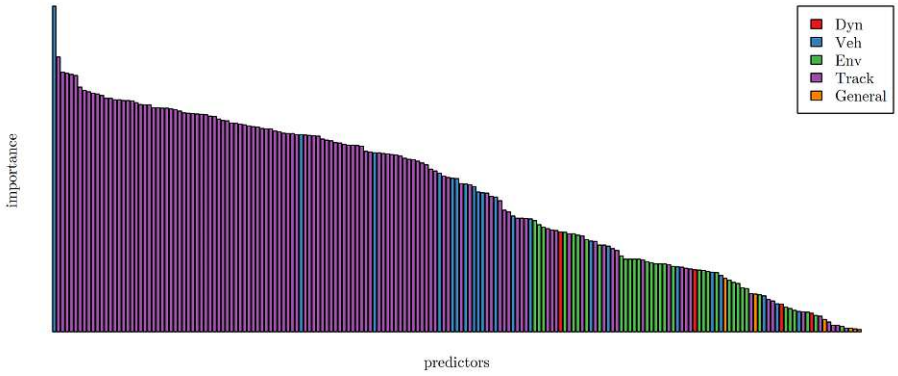

Figure 106: Predictor importance evaluated with linear variance analysis and coloured in groups dynamic (red), vehicle (blue), environmental (green), track (purple) and general (orange)

Empirical examination reveals regarding freight traffic a steep drop between 226 m and 256 m of up to 35 % (presumably due to Apl values), followed by a negligible decrease to 310 m ($<$ 1 %). Passenger train categories indicate no curve radius dependency with general values for narrow radii ($\leq$ 256 m) between 19 % and 27 %. Other track predictors show weak to no correlation to the outcome.

Apart from train type influence, train orientation in single units indicates a controversial

trend – in C1, pushed operation shows 3 % to 5 % lower values, while in C5, the pushed state causes an increase of 2 %. Double unit operation indicates significantly lower relative occurrence time with around 20 % compared to up to 42 % (pulled single units). Among train configurations in double units, no differences are observed. Altogether, on the one hand, single units of train type F in C1 show higher values (difference up to 14 %) than in C5 and on the other hand double units indicate approximately half of the occurrence time. Hence, if train length is taken into account, absolute occurrence time is identical between single and double units. Nevertheless, since relative time values are considered, the observed deviation has to be taken into account. Other vehicle predictors show hardly any correlation to the outcome.

Correlations of dynamic predictors Vmean and AccSide to the outcome are partly weak to not visible. Other general dynamic influence is not observed.

In contrast to statistical importance ranking, seasonal and monthly deviations are evaluated (illustrated in Table 90) empirically. Beside partly controversial trends – presumably caused by weak statistical significance as well as high interquartile ranges –, agreement is reached with overall low relative occurrence time in June. In C5, lower levels are indicated in summer months (especially regarding train type G). In general, deviations among months are large with up to 11 % and 7 % in C1 (train types E and F) as well as up to 18 % and 9 % in C5 (train types F and G). It has to be pointed out that values may depend largely on climatic conditions at different locations. Thus, behaviour among the year might deviate significantly from the illustration in other locations. Daytime behaviour cannot be evaluated due to a generally low frequency of occurrence in terms of squeal noise and hence not enough monthly data points among hours. Altogether, subsets of C1 indicate only weak correlations in conjunction with environmental predictors – thus, variations over the year seem somewhat random. In contrast, subsets of C5 show visible negative correlations (category 2) in terms of TempRail as well as regarding absolute humidity quantities and RT.DPDiff. Frost presence causes an increase of about 3 % (frost) and 5 % to 10 % in case of hoarfrost conditions – indicated by all considered subsets. Dew occurrence indicates a slight positive correlation, however in a negligible magnitude of 1 % to 2 %. Thus, especially hoarfrost seem to have a significant influence, however statistical significance is low and occurrences among the year are rare. Rain cannot be evaluated due to hardly any squeal noise incidences and hence lack of sufficient data points. In general, high interquartile ranges are present, which causes an uncertainty in stated values.

subset	month											
	1	2	3	4	5	6	7	8	9	10	11	12
MC_1_TT_E	–	29	33	32	–	22	29	–	–	22	32	23
MC_1_TT_F	19	20	24	23	26	21	21	–	–	25	21	22
MC_5_TT_F	38	20	–	21	–	25	30	23	30	25	33	27
MC_5_TT_G	25	26	26	28	23	21	19	20	24	30	26	28

Table 90: Monthly median values of TonalLowTimeRel over the year with colour theme for statistical significance – red (weak), orange (mediocre)

Application of statistical algorithms reveals that complex tree and rule-based models – Boosted Tree and Cubist – outperform other approaches especially in the whole dataset and in some subset, where complex connections are present. Partly, simpler tree models or MARS, which are more interpretable, perform equally. Benchmarks are provided in Table 91. A high fluctuation is indicated by large interquartile ranges (refer to Figure 95), which leads to a large RMSE value of 13 % in the whole dataset. Correlation between prediction and original is 20 %, which is also poor. Thus, even complex algorithms lack to extract a useful pattern among predictors. Due to application of five times repeated 10-fold cross validation, illustrated benchmarks are a good indicator of predictive behaviour on the same boundary conditions.

Predictor importance scores for the whole dataset (refer to Figure 107 – depiction with 25 % relative importance cut point) indicate hardly any direct dependency on track predictors. Instead, train type and section distinction is carried out by inclusion of Apl, AccTrain, HoursSLT, VDiff and Pressure. The issue about Apl is previously discussed. AccTrain and VDiff correlations are compromised by speed loss of every train pass by in C5. The inclusion of HoursSLT is reasoned by possibility to extract freight trains, which tend to higher values due to less frequent train operation in the night. However, a general trend is not observed empirically. Regarding Vmean and AccDiff, a weak positive correlation is observed empirically. However, they are presumably compromised in the model by lower velocities of freight trains in C1 compared to passenger trains. Listing of TrainOrientation complies with empirical examination and is used to distinguish between single and double units. General predictors Week and YearDay indicate over-fitting and are not suitable for representation of general trends due to location dependencies and yearly alterations. Among environmental influence, absolute humidity quantities are ranked top, followed by relative air pressure, air temperature and relative rail humidity (low importance). Generally, accordance to empirical observations (in C5) is reached. Relative air pressure is thought to be representable by other environmental predictors since it lacks of a physical explanation for direct correlation. The importance ranking of air temperature above rail temperature is contrary to empirical examination.

To summarise, track influence is observed in conjunction with curve radius. A significant decrease between radii 226 m - 256 m (median roughly 25 %) and radii between 310 m -

subset	Boosted Tree			Cubist		
	RMSE	R^2	number of predictors	RMSE	R^2	number of predictors
All	12.97	0.20	147	12.98	0.21	72
MC_1	11.14	0.24	51	11.17	0.24	47
MC_5	15.13	0.10	48	15.27	0.11	15
TT_E	8.58	0.27	41	8.85	0.24	36
TT_F	12.25	0.19	95	12.25	0.20	44
TT_G	15.27	0.07	72	15.77	0.06	30
MC_1_TT_E	8.56	0.26	39	8.84	0.23	34
MC_1_TT_F	11.91	0.20	42	12.00	0.20	27
MC_5_TT_F	12.86	0.16	38	12.95	0.17	16
MC_5_TT_G	15.33	0.07	37	15.65	0.07	27

Table 91: Results of best performing regression models achieved with Boosted Tree and Cubist

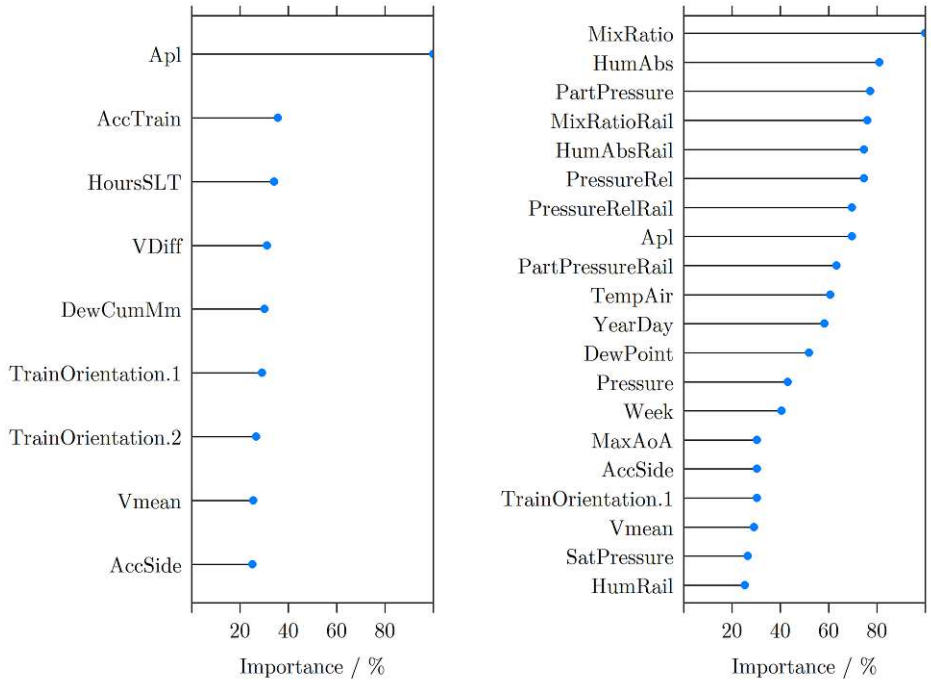

Figure 107: Relative predictor importance in the Boosted Tree (left) and Cubist (right) model fitted to the whole dataset

440 m (median about 10 %) is indicated in the whole dataset. Freight traffic shows a steep drop between 226 m and 256 m of up to 35 % (presumably due to Apl values), followed by a negligible decrease to 310 m (< 1 %). Passenger train categories indicate no curve radius dependency with general values for narrow radii (≤ 256 m) between 19 % and 27 %.

As previously described, distinct train type alterations are observed among subsets – freight trains in C4/C5 show extreme values, followed by passenger trains in midfield and freight traffic in C1 with low relative occurrence time. Thus, trends are controversial, which allows the assumption of no general train type influence, but rather high deviations among freight trains. However, it has to be pointed out that statistical significance is mostly weak due to generally low frequency of occurrence regarding squeal noise. Double unit operation halves the relative occurrence time (from roughly 40 % to 20 %). Thus, absolute squeal noise occurrence time remains unchanged whether if single or double units are operated.

Dynamic influence is partly observed regarding mean speed and lateral acceleration, however with weak correlations.

Seasonal and monthly variations are observed. Alterations among the year are partly controversial with the only agreement of generally low values in June. In C5, lower levels are indicated in summer months (especially regarding train type G). In general, deviations among months are large with up to 11 % and 7 % in C1 (train types E and F) as well as up to 18 % and 9 % in C5 (train types F and G). Negative correlations are shown in terms of temperature, absolute humidity quantities and difference between temperature and dew point (relative humidity indicator). While dew occurrence seems to have negligible impact and rain is not evaluable due to lack of sufficient data points, frost – especially hoarfrost – causes increases of up to 10 % relative occurrence time. It has to be pointed out that due to high interquartile ranges in subsets, a general uncertainty is present regarding the presented trends and their magnitude. Reasons behind stated findings are unknown, which indicates need of further research.

Building on the results of empirical and statistical investigation, the following predictors have the most impact on the outcome:

- curve radius
- double unit operation
- rail temperature
- mixing ratio (also takes absolute air pressure into account)
- difference between rail temperature and dew point (relative humidity indicator)
- frost conditions (frost or hoarfrost likely)

Due to high fluctuations in the dataset (resulting in high RMSE values and a general uncertainty) and pursuing the scope of constructing a practical model, a heuristic approach is chosen. Prediction is separated into two groups – curve radii ≤ 256 m and radii ≥ 310 m.

For the former group, relative time occurrence is set to 30 % for all train categories and for the latter range to 10 %. In between, linear interpolation is used. Thus, for radii greater than 256 m and less than 310 m, relative occurrence time can be derived by $124.815 - 0.3704 \cdot R$. It has to be pointed out that this is a try to model a simplified relationship empirically, which needs further radii for comparison and validation.

C.5 HF squeal noise peak level

It has to be borne in mind that frequency of occurrence regarding HF squeal noise lies in the same range as squeal noise. The only exception is present regarding freight trains in larger curve radii, where higher rates are observed. Thus, the issue of sparse data point distribution and many subsets missing the defined threshold for statistical significance also refers to the evaluation of peak levels in conjunction with HF squeal noise (see Figure 96).

Predictor importance scores of univariate statistical analysis (depicted in Figure 108) in conjunction with HF squeal noise peak levels (termed TonalHighLevelMax) indicate a dominant vehicle influence. In first and second place are TrainType.6 (train type F) and TrainOrientation.1, followed by TrainType.5 (train type E) in third place. While train type E shows overall the highest median peak levels, the ranking of train type F is caused by relatively low peak levels – especially in C1. Influence of train orientation is further discussed in the next paragraph. High importance scores are achieved by track predictors (mainly TDR and RR), which indicates deviations among sections. Absolute humidity quantities and rail temperature are ranked in the upper half as well as some other vehicle parameters. General variables are ranked in midfield, while dynamic predictors achieve very low importance scores.

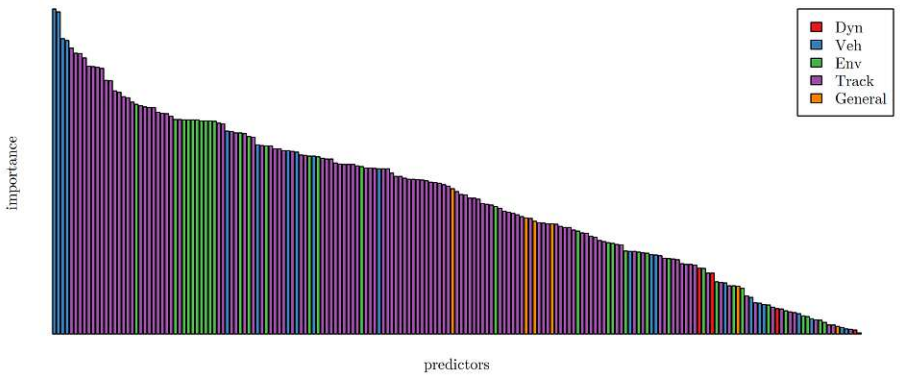

Figure 108: Predictor importance evaluated with linear variance analysis and coloured in groups dynamic (red), vehicle (blue), environmental (green), track (purple) and general (orange)

Empirically, negative correlation of curve radius to the outcome is observable for freight trains between 256 m (95 dB) and 440 m (88 dB). Narrower radii ($\leq$ 256 m) show median

levels of 92 dB to 95 dB. Due to lack of sufficient data points, trends of passenger trains in conjunction with larger radii ($>$ 256 m) cannot be evaluated. Regarding narrower radii ($\leq$ 256 m), a significant decrease of 4 dB to 7 dB between C5 (230 m) and C1 (256 m) occurs for both train type F and G. A solely cause of curve radius is thought to be unlikely, since in C4 (226 m) median peak levels lie in between. However, data points from C4 are not as trustworthy due to low statistical significance and uneven distribution among the year. Thus, additional underlying mechanisms might be present, which are unknown. Other track predictors show no direct connection to the outcome.

Distinct train type dependencies are observed. Interestingly, in contrast to squeal noise, freight traffic does not show the highest peak levels. Train type E in C1 shows largest overall median values (97 dB), while regarding train types F and G, 7 dB lower median levels are observed in the same section. As reason for the former, an unknown train type specific property is assumed. Freight traffic lies in between. In C5, median levels are more evenly distributed among all train types deviating between 94 dB and 96 dB. Thus, freight and passenger trains emit roughly equal peak levels. Train orientation influence regarding train type F is negligible – showing fluctuations of 1 dB to 2 dB (in controversial directions between C1 and C5) between pushed and pulled operation. Moreover, single and double unit operation indicate no difference in radiated peak levels. Train orientation among double units shows also insignificant differences ($<$ 1 dB). Thus, the high predictor importance score regarding train orientation in statistical evaluation is caused by generally high levels of train type E in C1 as the majority of train pass bys are operated in the pulled state (TrainOrientation.1). All vehicle predictors, which allow distinction between C1 and C5 (e.g. MaxAoA and RSI) show correlations to the outcome. However, no general connections are evaluated among them.

Regarding dynamic predictors, empirical examination reveals no correlations, which complies with low importance ranks statistically.

Variations among the year deviate significantly between C1 and C5 (refer to Table 92 for monthly median values among the year). While in the former, for both train type E and F, differences over the whole year are in a range of up to 3 dB, in the latter section, monthly distribution differs significantly (up to 11 dB for train type F and up to 6 dB for train type G). Raised values in autumn (especially October and November) and extreme values in winter (foremost in December and January) are shown by both train types in C5. Subsets partly indicate a trend of lower values in summer months. Due to generally most statistical significance of train type G in C5 and measurement over the whole year in that section, the indicated trend of that subset is most trustworthy. However, since not much overall deviation is shown in C1, the observed tendency in C5 with that magnitude might not apply to all sections. The presented deviations among the year can be described by alterations in environmental conditions – foremost temperature and absolute humidity quantities, which show visible negative correlations in subsets of C5 and weak ones in other considered subsets. Observed deviations of attenuation coefficients between winter and summer, which can impact high frequencies between 8000 Hz and 12500 Hz even in a close monitoring range, indicate a controversial trend (see Subsection 6.3.5). Thus, the latter influence seems to be masked by another more dominant effect. Hence, excitation mechanisms between equivalent

subset	month											
	1	2	3	4	5	6	7	8	9	10	11	12
MC_1_TT_E	97	96	96	98	96	97	–	–	–	98	98	96
MC_1_TT_F	–	90	90	90	89	89	90	–	–	91	91	92
MC_5_TT_F	100	–	91	98	–	89	94	94	94	97	97	99
MC_5_TT_G	98	94	94	93	93	93	92	92	93	94	94	98

Table 92: Monthly median peak levels of TonalHighLevelMax with colour theme for statistical significance – red (weak), orange (mediocre)

continuous noise level and observed alterations in the present evaluation seem not connected as well as remain unclear. Daytime influence cannot be evaluated due to lack of monthly data points among hours in a sufficient number. Moreover, relative humidity quantities show weak positive correlations and partly none at all. Thus, they might not influence the outcome directly. While dew presence does not show any impact except regarding train type E in C1 (raise in median of 2 dB), frost conditions indicate a significant influence. In C5, increases of 3 dB (train type G) and 4 dB (train type F) are observed regardless whether frost or hoarfrost is likely. Thus, median level raises in December and January compared to neighbour months seem solely explainable by frost presence. The assumed reason behind the tendency regarding frost occurrence is a generally lower surface absorption in the transmission path due to reduced porosity in the ballast bed and frozen earth between ballast bed and microphone (refer to Subsection 6.3.5). General percentage of frosty conditions of train pass bys in C1 is very low and in contrast slightly higher in November. Thus, differences in climatic conditions in winter can explain deviations between C1 and C5 partly, which enhances the exclusion of general predictors from a generally applicable predictive model. Impact of rain cannot be evaluated due to hardly any HF squeal noise incidences.

Application of statistical algorithms reveals that complex rule-based and tree models – Boosted Tree and Cubist – achieve best overall benchmarks (results illustrated in Table 93). As can be seen, the whole dataset can be best modelled by the Boosted Tree approach. Among subsets, no clear winner between Boosted Tree and Cubist can be stated. In general, achieved RMSE values are mediocre with around 4 dB and correlation between prediction and outcome is poor with 28 % agreement between prediction and original in the whole dataset. Thus, slightly worse performance compared to squeal noise peak level prediction is achieved. Not much deviations between benchmarks in the whole dataset and among subsets indicate that the model is capable of including different section and train type properties. Due to application of five times repeated 10-fold cross validation, illustrated benchmarks are a good indicator of predictive behaviour on the same boundary conditions.

Predictor importance scores (depicted in Figure 109 with a cut point of 25 % relative importance) of the Boosted Tree model list only two vehicle predictors – MaxAoA and Apl. Thus, the model focuses on distinguishing between train types and sections in general. While MaxAoA includes curve radius and bogie wheelbase, Apl is only responsible for train

subset	Boosted Tree			Cubist		
	RMSE	R^2	number of predictors	RMSE	R^2	number of predictors
All	4.25	0.28	142	4.29	0.27	79
MC_1	4.01	0.37	48	4.06	0.36	43
MC_5	4.46	0.15	47	4.49	0.14	23
TT_A	4.91	0.23	64	5.07	0.20	11
TT_E	4.15	0.12	39	4.23	0.10	27
TT_F	4.04	0.33	94	4.07	0.32	63
TT_G	4.24	0.17	96	4.30	0.16	34
MC_1_TT_E	4.14	0.12	35	4.22	0.10	25
MC_1_TT_F	3.82	0.07	39	3.85	0.05	23
MC_5_TT_F	4.56	0.22	37	4.50	0.26	27
MC_5_TT_G	4.29	0.14	34	4.35	0.12	17

Table 93: Results of best performing regression models achieved with Boosted Tree and Cubist

type distinction. In general, both are compromised by high levels of train type E in C1. Regarding the former, bogie wheelbase is difficult to estimate in a predictive modelling task. Instead, inclusion of RSI is possible. However, neither bogie wheelbase nor wheel diameter seems directly connected to the outcome in empirical examination. The same applies to Apl. Moreover, axle count and train length are impossible to know in a predictive task. Hence, using solely curve radius as predictor might be beneficial. The Cubist algorithm is heavily focused on environmental conditions – especially absolute humidity quantities (highest rank HumSpec) and temperature, which complies with empirical findings. Inclusion of YearDay and Week indicates over-fitting. Both as well as Pressure are not meaningful in a general predictive model due to location dependencies. Relative air pressure lacks of a physical explanation and is thought to be represented by other environmental predictors.

To conclude, negative correlations of curve radius to the outcome are shown especially for freight traffic. Regarding passenger trains, the trend can only be confirmed between 230 m and 256 m (lack of data points for larger radii) with a significant decrease in median levels of up to 7 dB, while values for 226 m radius lie in between. However, the latter tendency is weakened by low statistical significance. Other track predictors show no direct correlation to the outcome.

Train type influence is distinct. In contrast to squeal noise observations, freight trains do not show generally higher median peak levels compared to passenger traffic. Largest median peak values are observed for train type E in C1 (97 dB), while regarding other passenger train categories, up to 7 dB lower values are observed in the same section. Train types F and G show roughly equal levels (± 1 dB) in the same sections. Train orientation as well as operation of single or double units show no impact on the outcome. Every predictor, which takes the observed train type dependencies and radius influence into account (i.e. MaxAoA and RSI) indicates a visible correlation. However, general correlations are neither regarding

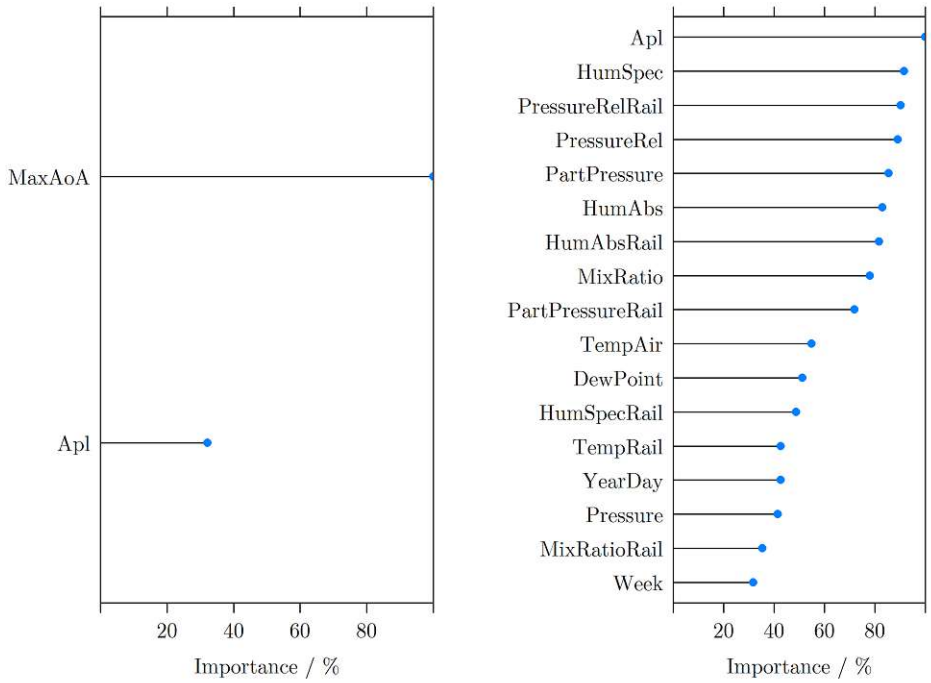

Figure 109: Relative predictor importance in the Boosted Tree (left) and Cubist (right) model fitted to the whole dataset

bogie wheelbase nor in conjunction with wheel diameter observable. Thus, due to controversial trends between passenger and freight traffic in C1, C4 and C5, train type deviations indicate no clear pattern and hence might generally be best represented by solely considering curve radius influence.

Dynamic predictors show no correlations in both statistical and empirical evaluation.

Seasonal influence is observed especially in C5, where median levels are significantly higher in winter (especially December and January) with 98 dB (for train type G) than in summer months (92 dB), while train types in C1 only show monthly deviations of up to 3 dB among the year. However, a slight trend of lower values in summer months complies with findings in C5. Detailed investigation shows that climatic conditions in winter between C1 (2013/2014) and C5 (2016/2017) differ significantly. While in C5 much more train pass bys are monitored in conditions where frost or hoarfrost is likely, C1 shows very few train pass bys in that conditions (e.g. C5 in December 66 % of train pass bys with frost compared to C1 in the same month with 2.8 %). Train pass bys, where frost and hoarfrost is likely, show 3 dB to 4 dB increased median levels, which explains the majority of observed deviation magnitude between C1 and C5 (train type F) in winter months. The assumed reason behind the latter is a generally lower surface absorption in the transmission path due to reduced porosity in the ballast bed and frozen earth between ballast bed and microphone (refer to Subsection 6.3.5). Moreover, raised levels in autumn (especially in October and November) and generally low levels in summer are observed. That trend is directly connectable to variations in temperature and absolute humidity (both negatively correlating). Observed

deviations of attenuation coefficients between winter and summer indicate an opposite trend. Thus, the latter impact seems to be masked by another more dominant influence. Hence, excitation mechanisms between equivalent continuous noise level and observed alterations in the present evaluation seem not connected as well as remain unclear. A statement regarding daytime alterations among months is not possible due to lack of sufficient data points. The same applies to rain. Remaining environmental predictors show weak to no correlations to the outcome (including relative humidity quantities).

Building on the results of empirical and statistical investigation, the following predictors have the most impact on the outcome:

- curve radius
- temperature (no clear winner between air and rail temperature)
- specific humidity (also takes absolute air pressure into account)
- frost conditions (frost or hoarfrost likely)

Due to poor benchmarks of multivariate statistical analysis and building on the scope of practical applicability, development of a prediction model for HF squeal noise peak level is not seen as meaningful. However, due to unknown reasons behind observed correlations as well as no scientific proof for stated assumptions, further research need is indicated by the present evaluation. For rough estimation of HF squeal noise peak levels, median values from Figure 96 for similar train types and radii can be applied.

C.6 Relative occurrence time of HF squeal noise

It has to be borne in mind that frequency of occurrence of HF squeal noise lies in the same range as squeal noise. The only exception is present regarding freight trains in larger curve radii, where higher rates are observable. Thus, the issue of sparse data point distribution and many subsets missing the defined threshold for statistical significance also refers to the evaluation of relative occurrence time in conjunction with HF squeal noise (see Figure 97).

Predictor importance ranking of univariate linear variance analysis regarding relative occurrence time of HF squeal noise (referred as TonalHighTimeRel) is dominated by track influence – mostly RR and TDR parameters (illustrated in Figure 110). Thus, distinct deviations among sections are indicated. In the upper half, some vehicle and environmental predictors are listed as well. In the former category, highest ranks are achieved by Apl (distinction between freight and passenger traffic), TrainType.6 (train type F) and TrainOrientation.1. Impact of the latter two is further discussed in the next paragraph. Regarding environmental influence, TempRail and absolute humidity quantities as well as Pressure (only useful for representation of different sections) are ranked high. Dynamic and general predictors are listed in the lower half.

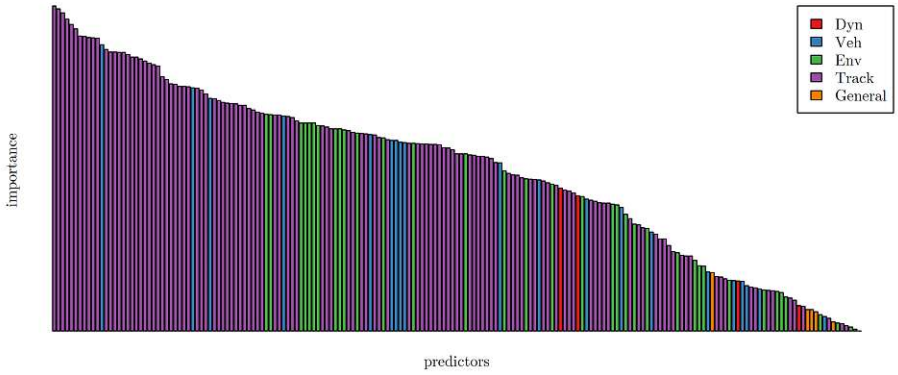

Figure 110: Predictor importance evaluated with linear variance analysis and coloured in groups dynamic (red), vehicle (blue), environmental (green), track (purple) and general (orange)

Empirical examination also indicates partly distinct deviations among sections, which is explainable with a negative correlation of curve radius to the outcome. In case of freight traffic, a decrease between 226 m/230 m and 256 m of up to 4 % and a decline to 310 m of 4 % is present – followed by a negligible one to 440 m (1 %). It has to be pointed out that statistical significance in all freight train subsets is low. Regarding passenger traffic, the trend is supported by train type F and G with reductions in median between 230 m and 256 m of 12 % and 5 %. Tendencies in larger radii cannot be evaluated due to lack of sufficient data points. Negative correlations are also observed regarding CurveBreathing (category 2 to 3), however it is foremost influenced by rail temperature (discussed below) and shows no general connection to the outcome. Other track predictors show no direct influence.

Train type dependencies are present – especially between freight and passenger traffic. In contrast to other dependent variables, freight trains show generally lower median values (< 15 %). Among passenger train types, higher medians of 16 % to 28 % are observable in narrower radii (≤ 256 m). In C1, significant differences between train type E (median 27 %) and train type F (16 %) are present. Detailed investigation reveals that deviations among sections (regarding train type F) as well as between train types E and F in C1 are caused by variations between single and double unit operation. In the former case, median values of train types E and F in C1 as well as between C1 and C5 are roughly equal. In contrast, double units show significantly lower median values (about halved). Considering the doubled train length, absolute time of occurrence is roughly equal as in single units. While pushed and pulled state show negligible differences regarding single unit operation, train configuration in double units indicate differences in a magnitude of up to 4 % (worst case both power coaches pushing their trainsets and best case both power units pulling). The mentioned deviation between single and double unit operation explains also the high statistical ranking of Apl. RSI and MaxAoA show partly visible correlation due to their ability to account decreasing values with increasing radii for freight trains. However, they are compromised by modelling

subset	month											
	1	2	3	4	5	6	7	8	9	10	11	12
MC_1_TT_E	31	26	40	30	25	24	–	–	–	33	24	23
MC_1_TT_F	–	15	15	16	17	12	18	–	–	19	17	16
MC_5_TT_F	37	–	18	23	–	16	25	26	22	24	30	37
MC_5_TT_G	32	28	24	25	19	26	26	25	23	24	26	35

Table 94: Monthly median of TonalHighTimeRel over the year with colour theme for statistical significance – red (weak), orange (mediocre)

the decrease between C5 and C1 for train type F, which is solely explainable by the impact of double unit operation and not a general trend. Other vehicle predictors show weak to no correlations to the outcome.

Regarding dynamic predictors, no direct influence on the outcome is observable.

Distinct fluctuations among the year are observable with variations among months of up to 16 dB and 7 dB in C1 (train types E and F) and of up to 11 dB and 16 dB in C5 (train types F and G) – summarised in Table 94. In C1, train type E indicates peaks in March and October. However, train type F in the same section only supports a slight raise in October and not much fluctuation among the year at all. Contrary, in C5, especially December and January show extreme values – consistent in conjunction with train types F and G. Moreover, agreement is reached in terms of generally lower values in June, which complies with findings on relative occurrence time of squeal noise. Frost or hoarfrost conditions cause increases in median of up to 14 % (train type F in C5) and up to 12 % (train type G in C5). Thus, explanation can be provided by the fact that in C5 those conditions are present in 66 % of all train pass bys in December, while in comparison hardly any frost conditions (2.8 %) are shown in C1 in the same month. Moreover, rail temperature and absolute humidity quantities show a visible to strong (category 2 to 3) negative correlation to the outcome in the whole dataset as well as in all subsets. A visible negative correlation is also observed regarding RT.DPDiff (representing relative humidity). Thus, warm and dry conditions seem to cause a lower relative occurrence time of HF squeal noise. Daytime influence among months as well as rain impact cannot be investigated due to lack of sufficient data points. Dew occurrence shows negligible influence. In general, high interquartile ranges are present, which causes an uncertainty in stated values.

Application of statistical algorithms reveals that complex tree and rule-based models – Boosted Tree and Cubist – achieve the best benchmarks (results are summarised in Table 95). As can be seen, the Boosted Tree algorithm outperforms the Cubist model in the whole dataset as well as in every subset partly significantly (considering RMSE). Overall, a high fluctuation in the dataset is indicated by large interquartile ranges (refer to Figure 97), which leads to a large RMSE value of 15.7 % in the whole dataset. Correlation between prediction and original is 24 %, which is also poor. Thus, even complex algorithms lack to extract a useful pattern among predictors. Due to application of five times repeated 10-fold cross validation,

subset	Boosted Tree			Cubist		
	RMSE	R^2	number of predictors	RMSE	R^2	number of predictors
All	15.69	0.24	146	16.06	0.23	81
MC_1	14.76	0.28	45	15.17	0.26	41
MC_5	16.72	0.14	46	16.89	0.14	35
TT_A	12.08	0.23	34	12.90	0.21	21
TT_E	18.31	0.14	41	18.81	0.12	28
TT_F	12.62	0.35	102	12.97	0.32	72
TT_G	16.71	0.09	96	17.15	0.08	21
MC_1_TT_E	18.28	0.15	36	18.80	0.13	30
MC_1_TT_F	10.16	0.17	41	10.39	0.16	26
MC_5_TT_F	17.50	0.25	35	17.56	0.27	32
MC_5_TT_G	16.52	0.09	40	16.85	0.09	14

Table 95: Results of best performing regression models achieved with Boosted Tree and Cubist

illustrated benchmarks are a good indicator of predictive behaviour on the same boundary conditions.

Relative predictor importance scores are depicted in Table 111 with a cut point of 25 % relative importance. The Boosted Tree model ranks MaxAoA as top predictor, which is reasonable since negative correlation to curve radius in terms of freight trains can be taken into account, closely followed by Apl. The latter is compromised by single unit and double unit operation and depicts no generally applicable trend. Moreover, two dynamic predictors – AccTrain and VDiff – are listed, which are compromised due to speed loss in C5 as well as show no correlations empirically. While HoursSLT is used to identify freight trains – indicating partly higher values due to operation during the night –, the reason behind inclusion of DewCumMm is unclear as no correlation is visible empirically. Ranking of YearDay is an indicator of over-fitting and DewEvapMm is directly correlated to SatPressureRail – hence, depended on temperature. To conclude, building on empirical and univariate statistical evaluation only MaxAoA and DewEvapMm are predictors with an assumed direct connection to the outcome. However, the former is compromised by taking the decrease between C5 and C1 for train type F into account, which is caused by deviations between single and double unit operation. Thus, although the Boosted Tree model achieves the best benchmarks, it is built on misleading trends without a general applicability. The importance ranking in the Cubist model seems more reasonable with also MaxAoA as top predictor, followed by absolute humidity quantities and temperature features. Although Quarter might be appropriate for the present modelling task, inclusion in a generally applicable predictive model is not preferable due to yearly alterations and locations dependencies. The high ranking of relative air pressure lacks of a physical explanation for direct impact, however is thought to be represented by other environmental predictors. In contrast to empirical examination, no relative humidity predictor is detected as important.

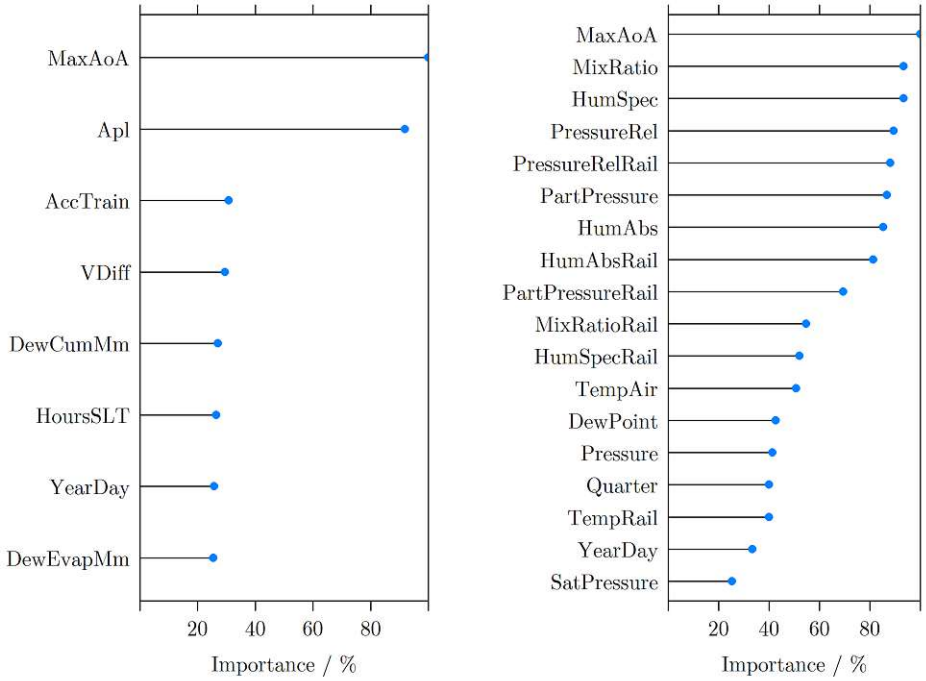

Figure 111: Relative predictor importance in the Boosted Tree (left) and Cubist (right) model fitted to the whole dataset

To conclude, among track predictors, only curve radius shows a direct correlation to the outcome – decreasing median values with increased radii. However, the influence can only be evaluated regarding freight trains due to lack of data points for passenger traffic in larger radii. Regarding freight traffic, a decrease between 226 m/230 m and 256 m of up to 4 % and a decline to 310 m of 4 % is present – followed by a negligible one to 440 m (1 %).

Train type influence is foremost visible between freight and passenger trains. The former category indicate about 12 % to 17 % less median values than the latter in conjunction with the same radii. Another distinct influence is observed regarding operation of single and double units. In the former case, median relative occurrence time complies with other passenger train categories (absolute values between 22 % and 28 %), while double units show significantly lower median levels (about halved). Considering the doubled train length, absolute occurrence time is roughly equal compared to single unit operation. Among train configurations, negligible differences between pushed and pulled state are observed in terms of single units. Double unit configurations are showing up to 4 % difference with a worst case if both power units are pushing their trainsets and a best case if both power coaches are pulling. Vehicle predictors, which allow inclusion of radius and train type dependencies (MaxAoA and RSI) show positive correlations to the outcome. However, it has to be borne in mind that due to the misleading trend of train type F in C1 (significantly lower overall median value due to mostly double unit operation), the mentioned predictors are compromised in statistical evaluations among multiple sections and train types. The same applies to Apl. Other vehicle predictors show no direct correlation to the outcome.

General influence of dynamic predictors is not observed.

Alterations among the year are observable with variations among months of up to 16 dB and 7 dB in C1 (train types E and F) and of up to 11 dB and 16 dB in C5 (train types F and G). In C5, median values are raised significantly in winter (especially in December and January). The trend is connected to frost or hoarfrost occurrence, which applies to 66 % of all train pass bys in December. Caused increases reach up to 14 % in median. Climatic conditions in C1 are much different with frost conditions in only 2.7 % train pass bys in December. Thus, missing raises in that section among winter months are explainable by that. Train type E in C1 shows high values in March and October, however, that trend is only slightly supported by train type F in the same section for October as well as not directly connectable to special environmental conditions. Agreement is reached in terms of slightly lower median values in summer months (especially June). In general, visible negative correlations of temperature, absolute humidity predictors and difference between temperature and dew point (representing a relative humidity quantity) to the outcome are observed – leading to the conclusion that warm and dry conditions cause less relative occurrence time. Dew presence shows a negligible impact. Influence of daytime among months as well as rain cannot be evaluated due to lack of sufficient data points. Other environmental predictors seem not directly connected to the outcome. It has to be pointed out that due to high interquartile ranges in subsets, a general uncertainty is present regarding the presented trends and their magnitude. Reasons behind stated findings are unknown and stated assumptions lack of scientific proof, which indicates need of further research.

Building on the results of empirical and statistical investigation, the following predictors have the most impact on the outcome:

- curve radius
- train type distinguished in two categories – freight (train types A, B) and other trains (train types C, D, E, F, G, other)
- double unit operation
- rail temperature
- mixing ratio (also takes absolute air pressure into account)
- difference between rail temperature and dew point (relative humidity indicator)
- frost conditions (frost or hoarfrost likely)

Due to high fluctuations in the dataset (resulting in high RMSE values and a general uncertainty) and pursuing the scope of constructing a practical model, a heuristic approach is chosen. Models are separated in freight trains and passenger trains. For the former, relative occurrence time is set to 15 % for curve radii $\leq$ 230 m, 5 % for curve radii $\geq$ 440 m and in between linearly interpolated with $25.952 - 0.0476 \cdot R$ for radii greater than 230 m and less than

440 m. It has to borne in mind that all values are indicated with weak statistical significance. Passenger trains are modelled with 28 % for radii $\leq$ 256 m, with 12 % for radii $\geq$ 310 m (only from train type D) and in between with a linear regression line of $103.852 - 0.2963 \cdot R$. Due to lack of sufficient data points, passenger train behaviour for radii $>$ 310 m is unclear and thus set to constant at a rather high value. Hence, future research can be used for fine tuning. Since it is unknown if the observed trend of significantly lower values in conjunction with double unit operation is generally shown, it is not included in the heuristic approach. It has to be pointed out that this is a try to model a simplified relationship empirically, which needs further radii for comparison and validation.

D Principal component coefficients for equivalent continuous noise level prediction

	coefficients for PC.(1-4)			
predictors	$a_{1,j}$	$a_{2,j}$	$a_{3,j}$	$a_{4,j}$
RR.OR.25	0.11307	0.05376	-0.07554	0.00938
RR.OR.20	0.11690	0.05099	0.03049	-0.00419
RR.OR.16	0.11667	-0.01664	-0.01960	0.00050
RR.OR.12.5	0.10614	-0.02434	-0.00051	-0.02160
RR.OR.10	0.11433	-0.03418	0.05397	0.00106
RR.OR.8	0.10821	-0.08340	0.05968	0.01980
RR.OR.6.3	0.10901	-0.07364	0.08549	0.01257
RR.OR.5	0.10866	-0.08037	0.07737	-0.00450
RR.OR.4	0.10380	-0.09849	0.07835	0.01699
RR.OR.3.15	0.10375	-0.11394	0.01780	-0.01267
RR.OR.2.5	-0.01062	-0.16024	0.04319	-0.09021
RR.OR.2	0.09868	-0.11568	0.01717	-0.06527
RR.OR.1.6	0.10286	-0.10129	0.07231	-0.00465
RR.OR.1.25	0.07531	-0.11680	0.07373	-0.02881
RR.OR.1	0.06781	-0.16352	0.06868	0.04652
RR.OR.0.8	0.10575	-0.09974	0.05910	0.00228
RR.OR.0.63	0.11733	-0.05385	0.00696	0.00355
RR.OR.0.5	0.10595	-0.09772	0.06759	-0.00437
RR.OR.0.4	0.10412	-0.09120	0.09254	0.01921
RR.OR.0.315	0.10145	-0.08850	0.09394	0.06950
RR.OR.0.25	0.10732	-0.09570	0.05524	0.01649

Table 96: Coefficients for derivation of principal components used in prediction of equivalent continuous noise level in conjunction with RR values for the outer rail

predictors	coefficients for PC.(1-4)			
	$a_{1,j}$	$a_{2,j}$	$a_{3,j}$	$a_{4,j}$
RR.IR.25	0.10245	0.06302	-0.12448	-0.00673
RR.IR.20	0.10012	0.03052	0.12298	-0.03785
RR.IR.16	0.10136	0.01884	0.14924	0.01703
RR.IR.12.5	0.11117	0.01808	0.11107	0.01224
RR.IR.10	0.11525	0.00043	0.07735	0.04503
RR.IR.8	0.10266	-0.00626	0.01111	0.14944
RR.IR.6.3	-0.08434	0.09808	0.00923	0.16432
RR.IR.5	-0.09016	0.04619	0.03763	0.17915
RR.IR.4	-0.10397	0.05961	0.05118	0.09142
RR.IR.3.15	-0.11026	-0.00426	0.07609	0.07856
RR.IR.2.5	-0.11229	-0.06883	0.01408	0.00114
RR.IR.2	-0.10759	-0.09112	0.02539	0.06128
RR.IR.1.6	-0.10847	-0.05581	0.05006	0.07552
RR.IR.1.25	-0.10608	-0.07883	0.06114	0.07352
RR.IR.1	-0.10613	-0.05529	0.08776	0.06788
RR.IR.0.8	-0.10269	-0.07174	0.07982	0.08475
RR.IR.0.63	-0.10579	-0.05389	0.07671	0.07150
RR.IR.0.5	-0.10639	-0.06227	0.09407	0.05404
RR.IR.0.4	-0.10568	-0.02192	0.13007	0.04602
RR.IR.0.315	-0.08685	-0.01898	0.19406	0.05512
RR.IR.0.25	-0.09187	-0.03559	0.17982	0.04202

Table 97: Coefficients for derivation of principal components used in prediction of equivalent continuous noise level in conjunction with RR values for the inner rail

predictors	coefficients for PC.(1-4)			
	$a_{1,j}$	$a_{2,j}$	$a_{3,j}$	$a_{4,j}$
TDRv.OR.100	0.08366	0.04365	0.05294	-0.21001
TDRv.OR.125	-0.00930	0.09420	0.13793	-0.24403
TDRv.OR.160	-0.07683	-0.05106	0.13346	-0.18051
TDRv.OR.200	-0.05927	-0.18840	-0.02755	-0.08683
TDRv.OR.250	-0.04784	-0.20693	-0.05125	-0.01041
TDRv.OR.315	-0.05772	-0.19154	-0.03426	-0.08113
TDRv.OR.400	-0.09900	-0.10513	0.07035	-0.02471
TDRv.OR.500	-0.08631	0.10648	-0.05476	-0.10817
TDRv.OR.630	-0.02543	0.16729	-0.13457	-0.12601
TDRv.OR.800	-0.04813	0.15276	-0.09772	-0.15218
TDRv.OR.1000	0.02197	0.06043	-0.22852	-0.06194
TDRv.OR.1250	-0.10601	-0.06274	-0.11486	-0.02603
TDRv.OR.1600	-0.00394	-0.01350	0.25615	0.12688
TDRv.OR.2000	0.09897	0.06078	0.13762	-0.06958
TDRv.OR.2500	0.10336	0.01388	0.15054	-0.02178
TDRv.OR.3150	0.09829	0.05848	0.02641	0.13904
TDRv.OR.4000	0.08062	-0.08176	0.11854	-0.15596
TDRv.OR.5000	-0.02477	-0.04924	0.16735	-0.22486
TDRv.IR.100	0.10506	0.02674	0.03314	-0.13676
TDRv.IR.125	0.07899	0.09413	0.01525	-0.19667
TDRv.IR.160	0.09581	-0.02219	-0.14527	-0.08507
TDRv.IR.200	0.09230	0.12187	0.02711	-0.08003
TDRv.IR.250	-0.01839	0.15224	0.07308	-0.21270
TDRv.IR.315	0.00553	0.03498	-0.02334	-0.29558
TDRv.IR.400	0.07890	-0.03811	0.11671	-0.18880
TDRv.IR.500	0.09500	-0.06933	0.01020	-0.13891
TDRv.IR.630	0.11086	-0.02043	-0.09873	0.00569
TDRv.IR.800	0.11410	0.01167	-0.01217	0.01205
TDRv.IR.1000	0.11141	-0.02587	-0.04475	0.08108
TDRv.IR.1250	-0.01546	0.18232	0.08645	-0.02091
TDRv.IR.1600	0.04358	0.14750	0.18047	0.03138
TDRv.IR.2000	0.11417	0.00385	0.01737	0.06348
TDRv.IR.2500	0.10309	0.05911	0.00439	0.09333
TDRv.IR.3150	0.07229	0.16188	0.08439	0.07109
TDRv.IR.4000	0.07565	0.15323	0.00486	0.01095
TDRv.IR.5000	0.02821	0.21116	-0.06654	-0.00941

Table 98: Coefficients for derivation of principal components used in prediction of equivalent continuous noise level in conjunction with vertical TDR values

predictors	coefficients for PC.(1-4)			
	$a_{1,j}$	$a_{2,j}$	$a_{3,j}$	$a_{4,j}$
TDRh.OR.100	0.11534	0.02405	-0.07987	0.01655
TDRh.OR.125	0.10110	-0.02489	-0.14894	-0.06187
TDRh.OR.160	0.11062	-0.00492	-0.11444	0.02712
TDRh.OR.200	0.05717	-0.11547	-0.10424	0.19826
TDRh.OR.250	0.10401	0.01681	0.08224	0.07938
TDRh.OR.315	0.11126	0.04338	-0.07983	0.06032
TDRh.OR.400	0.11273	0.00907	-0.08486	0.06656
TDRh.OR.500	0.11611	0.02193	-0.05360	0.04541
TDRh.OR.630	0.11259	0.02340	-0.09125	0.02202
TDRh.OR.800	0.11444	0.06733	0.02293	-0.00633
TDRh.OR.1000	0.11326	-0.02754	0.08930	0.04014
TDRh.OR.1250	-0.04491	0.11573	0.20160	0.11735
TDRh.OR.1600	-0.09382	0.11426	0.06788	0.09500
TDRh.OR.2000	-0.02229	0.19523	-0.10044	0.08971
TDRh.OR.2500	0.08524	0.11471	0.13912	0.05059
TDRh.OR.3150	0.10026	-0.08768	-0.01963	0.11605
TDRh.OR.4000	0.10863	-0.04238	-0.11072	0.03713
TDRh.OR.5000	0.10628	0.09215	0.05643	-0.03828
TDRh.IR.100	0.08684	0.06611	0.15277	-0.07540
TDRh.IR.125	0.10688	0.00027	0.04570	0.08220
TDRh.IR.160	0.07504	-0.04315	-0.09636	0.18953
TDRh.IR.200	0.09661	0.05830	-0.08912	0.11251
TDRh.IR.250	0.10669	0.01542	-0.09075	0.09808
TDRh.IR.315	0.10229	0.11592	-0.02778	0.00020
TDRh.IR.400	0.11615	-0.06081	0.01390	-0.00107
TDRh.IR.500	0.11959	-0.02172	-0.00847	0.00346
TDRh.IR.630	0.11694	-0.04602	0.03226	0.02378
TDRh.IR.800	0.11877	-0.01277	0.00635	0.04531
TDRh.IR.1000	0.11131	-0.07043	0.03174	0.06433
TDRh.IR.1250	0.10250	-0.05949	0.12922	0.03022
TDRh.IR.1600	-0.04669	0.17142	0.15485	0.04945
TDRh.IR.2000	0.09988	0.12079	-0.01651	0.04410
TDRh.IR.2500	-0.00794	0.18778	0.16198	0.03917
TDRh.IR.3150	-0.00320	0.21164	0.08179	0.05094
TDRh.IR.4000	0.06583	0.17375	0.04812	0.07866
TDRh.IR.5000	0.09234	-0.07726	0.03897	-0.15427

Table 99: Coefficients for derivation of principal components used in prediction of equivalent continuous noise level in conjunction with lateral TDR values