



UNIVERSIDADE ESTADUAL DE CAMPINAS
Faculdade de Engenharia Mecânica

Arthur Cancellieri Pires

**Measuring railway track irregularities from
instrumented railway vehicle data using
machine learning techniques**

**Medição de irregularidades da ferrovia
usando dados de um vagão instrumentado e
técnicas de aprendizado de máquina**

CAMPINAS
2023

Arthur Cancellieri Pires

Measuring railway track irregularities from instrumented railway vehicle data using machine learning techniques

Medição de irregularidades da ferrovia usando dados de um vagão instrumentado e técnicas de aprendizado de máquina

Dissertation presented to the School of Mechanical Engineering of the University of Campinas in partial fulfillment of the requirements for the degree of Master/Doctor in Mechanical Engineering, in the area of Solid Mechanics and Mechanical Design.

Dissertação apresentada à Faculdade de Engenharia Mecânica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Mecânica, na Área de Mecânica dos Sólidos e Projeto Mecânico.

Orientador: Prof. Dr. Auteliano Antunes dos Santos Júnior
Coorientador: Prof. Dr. Guilherme Fabiano Mendonça dos Santos

ESTE TRABALHO CORRESPONDE À VERSÃO FINAL DA DISSERTAÇÃO/TESE DEFENDIDA PELO ALUNO ARTHUR CANCELLIERI PIRES, E ORIENTADA PELO PROF. DR. AUTELIANO ANTUNES

CAMPINAS
2023

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Área de Engenharia e Arquitetura
Rose Meire da Silva - CRB 8/5974

P665m Pires, Arthur Cancellieri, 1996-
Measuring railway track irregularities from instrumented railway vehicle data using machine learning techniques / Arthur Cancellieri Pires. – Campinas, SP : [s.n.], 2023.

Orientador: Auteliano Antunes dos Santos Júnior.
Coorientador: Guilherme Fabiano Mendonça dos Santos.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Engenharia Mecânica.

1. Aprendizado de máquina. 2. Engenharia ferroviária. 3. Inteligência artificial. 4. Manutenção baseado na condição. 5. Ciência de dados. I. Santos Júnior, Auteliano Antunes dos, 1963-. II. Santos, Guilherme Fabiano Mendonça dos. III. Universidade Estadual de Campinas. Faculdade de Engenharia Mecânica. IV. Título.

Informações Complementares

Título em outro idioma: Medição de irregularidades da ferrovia usando dados de um vagão instrumentado e técnicas de aprendizado de máquina

Palavras-chave em inglês:

Machine learning

Railway engineering

Artificial intelligence

Condition based maintenance

Data science

Área de concentração: Mecânica dos Sólidos e Projeto Mecânico

Titulação: Mestre em Engenharia Mecânica

Banca examinadora:

Auteliano Antunes dos Santos Júnior [Orientador]

Romis Ribeiro de Faissol Attux

Luís Gustavo Giacon Villani

Data de defesa: 31-01-2023

Programa de Pós-Graduação: Engenharia Mecânica

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0003-0980-5000>

- Currículo Lattes do autor: <http://lattes.cnpq.br/5178519013420583>

**UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA**

DISSERTAÇÃO DE MESTRADO ACADÊMICO

**Measuring railway track irregularities from
instrumented railway vehicle data using
machine learning techniques**

**Medição de irregularidades da ferrovia
usando dados de um vagão instrumentado e
técnicas de aprendizado de máquina**

Autor: Arthur Cancellieri Pires

Orientador: Prof. Dr. Auteliano Antunes dos Santos Júnior

Coorientador: Prof. Dr. Guilherme Fabiano Mendonça dos Santos

A Banca Examinadora composta pelos membros abaixo aprovou esta Dissertação/Tese:

Prof. Dr. Romis Ribeiro de Faissol Attux (Titular)
Faculdade de Engenharia Elétrica e de Computação (FEEC)/UNICAMP

Prof. Dra. Rosângela dos Santos Motta (Titular)
Departamento de Engenharia Civil/Poli-USP

Prof. Dr. Luís Gustavo Giacon Villani (Titular)
Departamento de Engenharia Mecânica/UFES

A Ata de Defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 31 de janeiro de 2022.

Acknowledgments

I would like to thank my supervisors, professor Auteliano Antunes dos Santos Jr. and Guilherme Fabiano Mendonça dos Santos. Your advice, ideas, and direction were very important for this thesis. Working at Lafer and LabTDF in your research group has been a pleasure.

I would also like to thank the data science and machine learning team members at LabTDF for their valuable feedback and discussions. This includes both current and former members: William Ludoviko Homem, Maria Clara Assunção Viana, Lorenzo Miniguite Scaramussa, Leandro Furlam Turi and Antonio Tameirão.

A very special thanks to my parents for all the love and support during my master's degree at Unicamp.

This study was financed in part by VALE S.A, grant 03-P-18453/2019

Resumo

Esta dissertação propõe uma abordagem baseada em dados para estimar as irregularidades geométricas da ferrovia a partir de dados medidos pelo vagão instrumentado (VI). Aprendizado de máquina é utilizado para encontrar o mapeamento não linear entre os dados do VI e as irregularidades da via. Um modelo dinâmico do veículo ferroviário BRA1 foi utilizado para gerar um conjunto de dados artificial que contém variáveis medidas pelo VI verdadeiro e também de outros VIs encontrados na literatura. Foi realizada uma etapa de análise de dados extensiva para verificar se a instrumentação atual do VI da BRA1 é suficiente para obter irregularidades tanto laterais quanto verticais da via. Engenharia de variáveis baseado nos movimentos do vagão, integração de sinais e métricas estatísticas foram aplicadas para extrair variáveis e, em seguida, as melhores variáveis foram selecionadas usando um método de *wrapper*. Oito modelos diferentes de regressão foram treinados e otimizados usando Optuna. Os resultados mostram que, com a instrumentação atual do VI, obter irregularidades laterais é improvável devido à baixa correlação, no entanto, as irregularidades verticais podem ser obtidas com uma raiz do erro quadrático médio (RMSE) de 0,556 mm. Com o pós-processamento, o RMSE foi ainda mais reduzido para 0,410 mm.

Palavras Chave: Aprendizado de máquina; Vagão instrumentado; Análise exploratória de dados; Irregularidades geométricas; Inspeção de via permanente.

Abstract

This thesis proposes a data-driven approach to estimating geometric track irregularities from instrumented railway vehicle (IRV) data. Machine learning is used to find the nonlinear mapping between IRV data and track irregularities. A dynamic model of the BRA1 railway vehicle was used to generate an artificial dataset that contains variables that are measured by the real BRA1 IRV and other variables measured by IRVs found in the literature. An extensive data analysis step was done to verify if the current instrumentation of the BRA1 IRV is sufficient for obtaining both lateral and vertical track irregularities. Feature engineering based on wagon movements, signal integration and time domain statistical metrics were applied to extract features and then the best features were selected using a wrapper method. Eight different regression ML models were trained and optimized after the feature selection using Optuna. The results show that, with the current instrumentation of the BRA1 IRV, obtaining lateral track irregularities is unlikely due to low correlation, however, vertical irregularities can be obtained with a root mean squared error (RMSE) of 0.556 mm. With postprocessing, the RMSE was further reduced to 0.410 mm.

Key Words: Machine learning; Instrumented railway vehicle; Exploratory data analysis; Track irregularities; Track inspection.

List of figures

Figure 1: Layers forming a ballasted railway track - (BERAWI, 2013).	21
Figure 2: Main components of the rail track superstructure - (CARLVIK, 2020).	21
Figure 3: Types of track irregularities. a) Alignment (lateral); b) Gauge; c) Longitudinal (vertical); d) cross-level; e) torsion/twist; f) superimposed - (ZHAI, 2020).	22
Figure 4: Track geometry car EM 120 - (BERAWI, 2013).	24
Figure 5: Track measurement system installed on a track geometry car - (KHAJEHEI, 2021).	24
Figure 6: Half-car suspension model – modified from (ZHANG et al., 2021).	25
Figure 7: Vehicle response recording at 69 km/h on tangent segment due to combined cross-level, alignment and surface deviations - (LI et al., 2006).	27
Figure 8: Example of an instrumented railway vehicle (IRV) used in a Brazilian railway. The left figure shows some of the IRVs instrumentation such as solar panels along the carbody. The right figure highlights one of its sensors (load cell attached to spring) for measuring spring displacement – modified from (SANTOS; REICHL, 2014).	28
Figure 9: IRV instrumentation layout - (HARDIE et al., 2016).	28
Figure 10: BRA1 Instrumented railway vehicle - (SOBRINHO et al., 2021).	30
Figure 11: Spring deflection response before and after tamping - (MONAKALI; RAVITHARAN, 2016).	31
Figure 12: Linear regression of track geometry signal for two consecutive tamping actions (deterioration period) - (NEUHOLD; VIDOVIC; MARSCHNIG, 2020).	32
Figure 13: Positional error in two datasets from two different inspection runs. (a) A constant shift between the two datasets, (b) compression of the second dataset to the first - (KHOSRAVI et al., 2021).	33
Figure 14: Regression analysis with 5 deterioration periods. Outliers are highlighted in red - (NEUHOLD; VIDOVIC; MARSCHNIG, 2020).	35
Figure 15: Degradation model after detection and elimination of outliers - (NEUHOLD; VIDOVIC; MARSCHNIG, 2020).	35
Figure 16: Illustration of how different speed profiles affect the measured vibration. Letter (a) shows the speed profiles in the time domain; (b) speed profiles in the spatial domain; (c) the roughness interpolated in time; (d) the roughness in space; (e) the acceleration of the oscillator in time, the box highlights two oscillations from the same bump which occur at different points in time due to the oscillator's speed profile; and (f) the acceleration of the oscillator in space - (LEDERMAN et al., 2017a).	37
Figure 17: Algorithm schema of the BC-IVAM sub-model for minimizing false positives - (BAI; LIU; LI, 2020).	38
Figure 18: Distribution map of vibration defect data in the down direction of the Lanxin Railway between km 548 and km 985.5 - (BAI; LIU; LI, 2020).	39
Figure 19: Letter (a) shows the carbody vertical acceleration. Letter (b) shows the Hilbert spectra for frequencies between 0 – 30 Hz, highlighting a joint depression. Letter (c) shows the Hilbert spectra for the range of 0 – 5 Hz - (TSUNASHIMA; HIROSE, 2020).	42
Figure 20: Multi-resolution analysis (MRA) using the discrete wavelet transform - (TSUNASHIMA et al., 2012).	43
Figure 21: Multi-resolution analysis (MRA) of vertical carbody acceleration. Different types of detected track irregularities are highlighted - (TSUNASHIMA et al., 2012).	44
Figure 22: Measurement results of a curved section with corrugation. a) Axlebox sensor b) Carbody sensor - (TSUNASHIMA et al., 2012).	45

Figure 23: Setup for measuring acceleration. The left figure shows axlebox mounted accelerometers. The right figure shows bogie mounted accelerometers - (LEE et al., 2012).	46
Figure 24: The effect of low vehicle speed on calculated vertical irregularities - (YEO, 2017).	49
Figure 25: Wavelengths that affect ride quality - (ALCOCER, 2019).	53
Figure 26: Track curvature.	55
Figure 27: Generated track irregularities from SIMPACK. Letter a) shows the lateral track irregularities for the right rail. Letter b) shows the vertical track irregularities for the right rail.	56
Figure 28: Frameworks for three different monitoring approaches: physical models, conventional data-driven models, and deep learning models. Shaded boxes denote data-driven components - (ZHAO et al., 2019).	58
Figure 29: Diagnosis procedure for intelligent fault diagnosis (IFD) using traditional machine learning - (LEI et al., 2020).	58
Figure 30: Example of overfitting using polynomial regression - (GÉRON, 2019).	63
Figure 31: Example of a decision tree for univariate splits - (AGGARWAL, 2015).	65
Figure 32: Example of a decision tree for multivariate splits - (GÉRON, 2019).	66
Figure 33: Training diverse classifiers - (GÉRON, 2019).	67
Figure 34: Combining models with a hard voting classifier - (GÉRON, 2019).	68
Figure 35: Illustration of bagging - (GÉRON, 2019).	69
Figure 36: A single decision tree (left figure) versus a bagging ensemble of 500 trees (right figure) - (GÉRON, 2019).	69
Figure 37: Illustration of how boosting works, exemplified by an AdaBoost classifier - (MARSH, 2016).	70
Figure 38: Simple flowchart of how model stacking works - (PIRES et al., 2021).	72
Figure 39: Aggregating predictions using stacking - (GÉRON, 2019).	72
Figure 40: Simplified neural network architecture for obtaining wheel loads from track geometry data - (LI et al., 2006).	73
Figure 41: SIMPACK model – modified from (SILVA, 2022).	75
Figure 42: Methodology flowchart.	76
Figure 43: Before outlier removal.	80
Figure 44: After outlier removal.	81
Figure 45: Y component of the triaxial accelerometer.	82
Figure 46: Zoom of the curved section between 4100 m and 4800 m.	82
Figure 47: Boxplot of the sensor variables.	83
Figure 48: Distribution plot of the track irregularities.	84
Figure 49: Correlation heatmaps.	85
Figure 50: Multivariate correlation using a wrapper method for 6 different models.	87
Figure 51: Representation of a bogies' six degrees of freedom - (YEO, 2017).	89
Figure 52: Excitation of oscillatory modes in the wagon body - (BLEAKLEY, 2006).	90
Figure 53: Accelerometer locations and axis naming convention - (BLEAKLEY, 2006).	91
Figure 54: Relating corner accelerations to 5 degrees of freedom - (BLEAKLEY, 2006).	91
Figure 55: Example of rigid body motions that can be recreated using feature engineering on the IRV measurements.	92
Figure 56: Objective history plot.	97
Figure 57: The effect that the activation function has on the objective function.	98
Figure 58: The effect that the number of neurons of each layer has on the objective function.	98
Figure 59: Optimized neural network architecture.	100
Figure 60: Neural network learning curve.	101

Figure 61: PSD of the target variable and model output compared to the FRA norm specifications.	102
Figure 62: Comparison between the model prediction and filtered prediction for a section of the test set.	103
Figure 63: Flowchart of how to aggregate the machine learning with explainability techniques for decision making.	104
Figure 64: SHAP values for determining feature importance.	105
Figure 65: Shap dependence plot of the variable representing the wagon's roll movement. .	106

List of tables

Table 1: Track quality levels from the norm EN 13848-5 - (BERAWI, 2013).....	50
Table 2: Coefficients for the power spectral density functions - (PIRES et al., 2021).	52
Table 3: Commonly used time domain features - (KIMOTHO; SEXTRO, 2014)	60
Table 4: Preprocessed dataset.....	78
Table 5: Illustration of the rolling window process employed.....	94
Table 6: Dataset after the rolling window process	94
Table 7: Model results.	99

List of abbreviations and symbols

Latin letters

A_k	Gaussian random variable with mean 0 and standard deviation σ_k , independent of each other for $k=1,2,\dots,N$	
$D1$	Wavelength of irregularities inside the range $3 < \lambda \leq 25 \text{ m}$	
$D2$	Wavelength of irregularities inside the range $25 < \lambda \leq 70 \text{ m}$	
$D3$	Wavelength of irregularities inside the range $70 < \lambda \leq 200 \text{ m}$	
f	Frequency per time unit	[Hz]
F	Frequency per space unit	[1/m]
$J(\theta)$	Cost function	
K_z	Vertical track curvature	[1/m]
N	Total number of frequency increments (harmonics) in $[w_{max} + w_{min}]$	
Ns	Number of samples (user-defined)	
V	Vehicle speed	[m/s]
x	Length of the signal in increments of $1/Ns$	
r	Mixing ration for ElasticNet regularization	
w_{min}	Minimum angular frequency (wavenumber)	[rad/s]
w_{max}	Maximum angular frequency (wavenumber)	[rad/s]
\ddot{z}	Vertical acceleration	[m/s ²]
$z(t)$	Vertical displacement	[m]
\hat{z}	Vertical displacement obtained from K_z	[m]

Greek letters

α	Weight of the regularization term	
Δw	Frequency increment $\left(\frac{w_{max}+w_{min}}{N}\right)$;	[rad/s]
$\Delta \Omega$	Angular frequency increment $(2\pi\Delta F)$	[rad/s]
ΔF	Frequency increment	[1/m]
θ_k	Random phase angle uniformly distributed in the range $(0,2\pi]$ and independent of A_k and of each other for $k = 1, 2, \dots, N$	[rad]
θ	Regularization term	
σ_k	Standard deviation expressed as $\sigma_k = \sqrt{2S_x(w_k)\Delta w}$	
Ω	Spatial wavenumber $(2\pi F)$	[rad/m]
λ	Wavelength of the track excitation	[Hz]
λ_{min}	Lower limit of the analyzed wavelength band	[m]
λ_{max}	Upper limit of the analyzed wavelength band	[m]

Subscript

A_v, A_a	Roughness coefficient – table constant	[cm ² rad/m]
Ω_c, Ω_s	Critical wavenumber – table constant	[rad/m]
$S_{al}(\Omega)$	PSD of FRA lateral track irregularities	[cm ² rad/m]
$S_v(\Omega)$	PSD of FRA vertical track irregularities	[cm ² rad/m]
$S_{cl}(\Omega)$	PSD of FRA crosslevel track irregularities	[cm ² rad/m]
$S_v(\Omega)$	PSD of FRA vertical track irregularities	
$S_{al}(\Omega)$	PSD of FRA lateral track irregularities	
$S_{cl}(\Omega)$	PSD of FRA crosslevel track irregularities	

Abbreviations

<i>AAR</i>	Association of American Railroads
<i>CEN</i>	European Committee for Standardization
<i>EN</i>	European Standard
<i>FRA</i>	Federal Railroad Administration
<i>IRV</i>	Instrumented railway vehicle
<i>MSE</i>	Mean squared error
<i>RMSE</i>	Root mean squared error
<i>SHAP</i>	Shapley additive explanations

Variable list

Acc_max	Maximum acceleration of all uniaxial accelerometers
BackBounce_Acc	Bounce movement on the trailing bogie calculated using uniaxial accelerometers
BackBounce_Susp	Bounce movement on the trailing bogie calculated using suspension displacement
BackRoll_Acc	Roll movement on the trailing bogie calculated using uniaxial accelerometers
BackRoll_Susp	Roll movement on the trailing bogie calculated using suspension displacement
Bounce_Acc	Largest value between FrontBounce_Acc and BackBounce_Acc
Bounce_Susp	Largest value between FrontBounce_Susp and BackBounce_Susp
FrontBounce_Acc	Bounce movement of the leading bogie calculated using uniaxial accelerometers
FrontBounce_Susp	Bounce movement of the leading bogie calculated using suspension displacement
FrontRoll_Acc	Roll movement of the leading bogie calculated using uniaxial accelerometers
FrontRoll_Susp	Roll movement of the leading bogie calculated using suspension displacement
kurt	Kurtosis of the time window
mean	Mean of the time window
MA	Mean absolute value of the time window
peak_to_peak	Peak to peak value of the time window

Pitch_Acc_L	Pitch movement of the left side of the wagon calculated using uniaxial accelerometers
Pitch_Acc_R	Pitch movement of the right side of the wagon calculated using uniaxial accelerometers
Pitch_Acc	Largest value between Pitch_Acc_L and Pitch_Acc_R
Pitch_Susp_L	Pitch movement of the left side of the wagon calculated using suspension displacement
Pitch_Susp_R	Pitch movement of the right side of the wagon calculated using suspension displacement
Pitch_Susp	Largest value between Pitch_Susp_L and Pitch_Susp_R
RMA	Root mean absolute of the time window
RMS	Root mean square of the time window
Roll_Acc	Largest value between BackRoll_Acc and FrontRoll_Acc
Roll_Susp	Largest value between BackRoll_Susp and FrontRoll_Susp
skew	Skewness of the time window
std	Standard deviation of the time window
SS_Z_LR	Secondary suspension displacement on the right side of the leading bogie
SS_Z_LL	Secondary suspension displacement on the left side of the leading bogie
SS_Z_TR	Secondary suspension displacement on the right side of the trailing bogie
SS_Z_TL	Secondary suspension displacement on the right side of the trailing bogie
SuspTravel_L	Largest suspension displacement on the left side of the wagon
SuspTravel_R	Largest suspension displacement on the right side of the wagon
SuspTravel	Largest value between SuspTravel_L and SuspTravel_R
UA_Z_LR	Uniaxial vertical acceleration on the right side of the leading bogie
UA_Z_LL	Uniaxial vertical acceleration on the left side of the leading bogie
UA_Z_TR	Uniaxial vertical acceleration on the right side of the trailing bogie
UA_Z_TL	Uniaxial vertical acceleration on the right side of the trailing bogie
UA_Z_LR_Vel	Numerical integration of the UA_Z_LR variable, obtaining velocity
UA_Z_LL_Vel	Numerical integration of the UA_Z_LL variable, obtaining velocity
UA_Z_TR_Vel	Numerical integration of the UA_Z_TR variable, obtaining velocity
UA_Z_TL_Vel	Numerical integration of the UA_Z_TL variable, obtaining velocity
UA_Z_LR_Pos	Numerical integration of the UA_Z_LR_Vel variable, obtaining displacement
UA_Z_LL_Pos	Numerical integration of the UA_Z_LL_Vel variable, obtaining displacement
UA_Z_TR_Pos	Numerical integration of the UA_Z_TR_Vel variable, obtaining displacement
UA_Z_TL_Pos	Numerical integration of the UA_Z_TL_Vel variable, obtaining displacement
TA_X	X component of the triaxial accelerometer
TA_Y	Y component of the triaxial accelerometer
TA_Z	Z component of the triaxial accelerometer

Summary

1	INTRODUCTION	18
1.1	Motivation	18
1.2	Objective	19
1.2.1	Research goals	19
2	LITERATURE REVIEW	20
2.1	Track components and geometry parameters	20
2.2	Describing railway track quality	23
2.2.1	Track geometry monitoring	23
2.2.2	Dynamic response monitoring	25
2.3	Practical considerations for quantifying track quality	30
2.3.1	Using maintenance information	30
2.3.2	Data alignment and matching	32
2.3.3	Track segments with different characteristics	33
2.3.4	Outlier removal.....	34
2.3.5	Robustness issues	36
2.3.6	Correcting false positives when detecting track defects.....	37
2.3.7	Summary of practical considerations	40
2.4	Instrumented railway vehicle sensors	40
2.4.1	Carbody mounted sensors.....	40
2.4.2	Axlebox mounted sensors.....	44
2.4.3	Bogie mounted sensors	46
2.5	Obtaining track irregularities from accelerometer signals.....	47
2.6	Norms for evaluating track quality	49
2.6.1	European Regulation Standard (EN 13848)	50
2.6.2	Federal Railroad Administration (FRA).....	51
2.6.3	Defining the track irregularity wavelength range.....	52
2.6.4	Generating artificial track irregularities in SIMPACK.....	54
2.7	Track condition monitoring	56
2.7.1	Machine learning for intelligent fault diagnosis.....	58

2.7.2	Data collection.....	59
2.7.3	Feature extraction and engineering	59
2.7.4	Health state recognition	60
2.8	Machine learning algorithms	61
2.8.1	Linear regression with regularization	61
2.8.2	Decision trees	64
2.8.3	Model ensembling	66
2.8.4	Neural networks.....	73
3	METHODOLOGY	75
3.1	Dataset creation and preprocessing.....	77
3.2	Exploratory data analysis	79
3.2.1	Outlier removal.....	79
3.2.2	The effect of track severity on measured signals	81
3.2.3	Correlation analysis	84
3.2.4	Conclusions from EDA	88
3.3	Feature engineering	88
3.3.1	Domain knowledge of wagon movements	89
3.3.2	Integrating accelerometer signals	92
3.3.3	Statistical metrics.....	93
3.4	Feature selection.....	95
3.5	Machine learning models	95
3.6	Hyperparameter optimization	96
4	RESULTS.....	99
4.1	Data postprocessing	101
4.2	Model interpretability	103
5	CONCLUSION	107
6	FUTURE STUDIES.....	109
	Bibliography.....	110
	APPENDIX A – Optuna hyperparameter search space	118

APPENDIX B – Optimal model parameters.....	120
APPENDIX C – Feature engineering equations for wagon movements	121

1 INTRODUCTION

Railway companies routinely face the problem of deciding which track segments should be prioritized for inspection and maintenance (GERUM; ALTAY; BAYKAL-GÜRSOY, 2019). Like any other mechanical system, railway components deteriorate with usage, time, and environmental conditions, causing lower safety and possibly railway asset failure (KASRAEI; ZAKERI; BAKHTIARY, 2020). Track degradation is especially important due to its large effect on vehicle dynamics and safety. Thus, it is crucial to have an accurate way of measuring and monitoring track irregularities in both the short and long term (KHAJEHEI, 2021; URDA et al., 2021).

The recent development of sensors and information technology (IT) has led to the emergence of condition-based monitoring of track assets (TSUNASHIMA; HIROSE, 2020). By recognizing how track condition changes over time, it is possible to predict the residual lifetime of the asset and optimize maintenance with accurate planning. Furthermore, track deterioration is closely related to ride quality and railway safety, so it is also important to keep track irregularities in check and follow the railway's norms (TSUNASHIMA, 2019).

By installing an onboard sensing device in an in-service vehicle, track inspection can be done at lower costs and higher frequency when compared to track inspection cars or manual inspection (TSUNASHIMA; HIROSE, 2020). These vehicles are referred to as IRVs (instrumented railway vehicles). From IRVs measured data, different types of models and techniques can be applied for condition monitoring and degradation forecasting.

1.1 Motivation

Data is constantly being measured from an instrumented railway vehicle (IRV) and used for maintenance planning and decision-making by the railway engineers all around the world. This is particularly true for Brazilian Railway, like BRA1. The railway company currently compares the measured data from the IRV with preestablished maintenance limits. However, the results are not useful in their completeness. More sophisticated methods involving condition-based

maintenance (CBM) are being invested to improve upon the current system, where the IRV is the main source of data.

Despite the abundance of available data, the systems that will transform this data into useful information for maintenance teams are in their early stages. Therefore, there is a large margin for improvement and innovation in this area. This thesis represents a step towards transforming the current paradigm of track evaluation by offering novel contributions to the analysis methods pointing towards a more comprehensive and secure approach to the digital inspection system.

1.2 Objective

The objective of this thesis is to develop a methodology to indirectly measure geometric track irregularities from the dynamic response measured by an instrumented railway vehicle (IRV).

1.2.1 Research goals

- Analyze the effect of track irregularity class (FRA6 – FRA3) on the measured dataset;
- Analyze the effectiveness of the current IRVs sensor arrangement for obtaining track irregularities;
- Utilize data-driven methods to measure rail-specific parameters, creating virtual sensors;
- Apply model interpretability techniques to understand how the optimal model operates and observe if the learned characteristics correspond to the known domain knowledge;
- Propose a flowchart of how to integrate the optimal data-driven model with day-to-day operations.

2 LITERATURE REVIEW

This section will focus on the core concepts and theories necessary for application of machine learning in the analysis of railway systems. Specifically, a brief review of track geometry parameters and types of geometric track irregularities is presented, to contextualize the reader before going into how track quality is quantified and why instrumented railway vehicles could be an adequate choice for applications involving condition monitoring. Different IRV sensor arrangements are also discussed, highlighting their respective applications, advantages, and disadvantages.

Later on, the dataset generation process is explained. Track irregularities are generated using a power spectral density function while dynamic simulation provides the sensor values that correspond to the BRA1 IRV. From this information, a machine learning problem can be formulated to map the input (sensor variables) to the output (track irregularities). More details on how this is done in the context of fault diagnosis and health monitoring are provided, alongside with the more commonly used machine learning models.

2.1 Track components and geometry parameters

The railway track is composed of several different components that work together as the track's support. The main purpose of this structure is to provide an economical transportation system by guiding the vehicle and transmitting contact loads through the track components to the subgrade (ATTO-OKINE, 2017). The basic layers that compose a ballasted railway track are shown in Figure 1 and Figure 2. The superstructure is made of rails, ties (sleeper), fasteners, turnouts, and crossings, while the substructure consists of a ballast, sub-ballast, subgrade, and other drainage facilities. These two structures are separated by the sleeper-ballast interface. The research developed in this thesis focuses only on the ballasted railway tracks and the superstructure.

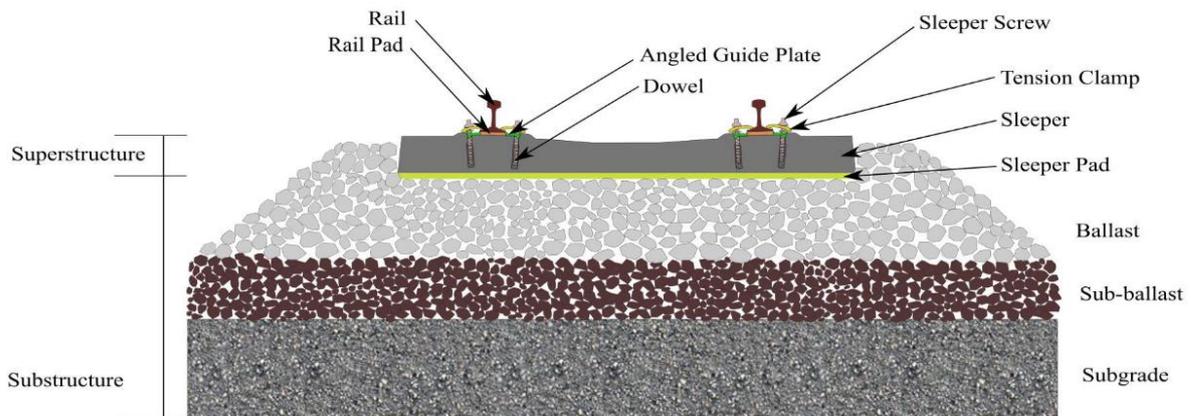


Figure 1: Layers forming a ballasted railway track - (BERAWI, 2013).



Figure 2: Main components of the rail track superstructure - (CARLVIK, 2020).

Track irregularities can be defined as geometrical deviations of the rail cross-section from an ideal track geometry. They can be divided into two main groups (URDA et al., 2021):

- **Distributed track irregularities:** exhibit regular patterns along the track for multiple wavelengths. These defects depend on the damage mechanism (wear or rolling contact fatigue)
- **Isolated track defects:** changes in the rail cross-sections at specific locations.

This thesis will focus on the distributed track irregularities category, which is shown in Figure 3. Distributed track irregularities are commonly represented by the following variables (KASRAEI; ZAKERI; BAKHTIARY, 2020; MORI et al., 2010):

- Left and right alignment (lateral irregularities);
- Left and right longitudinal level (vertical irregularities);
- Cant (superelevation);
- Gauge;
- Twist.

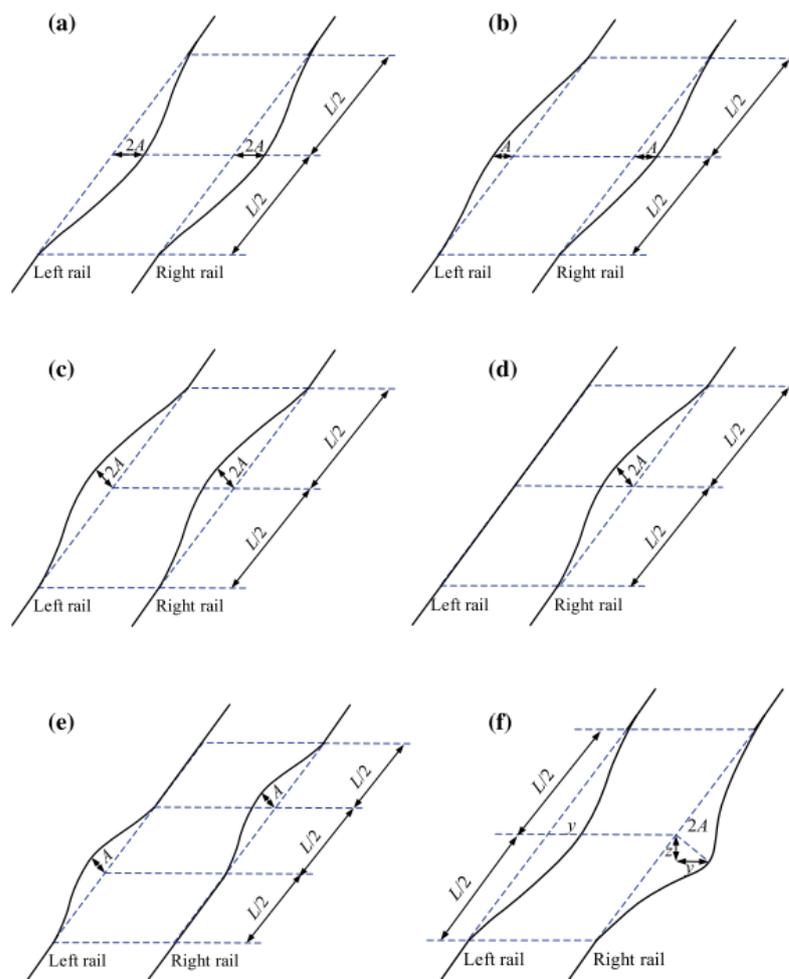


Figure 3: Types of track irregularities. a) Alignment (lateral); b) Gauge; c) Longitudinal (vertical); d) cross-level; e) torsion/twist; f) superimposed - (ZHAI, 2020).

2.2 Describing railway track quality

When planning maintenance actions, it is crucial to describe track quality and understand the current track condition to reduce derailment risks (KASRAEI; ZAKERI; BAKHTIARY, 2020; WEI et al., 2014; ZHU et al., 2013). The two main ways of quantifying track quality are track geometry measurement and monitoring the dynamic response of the vehicle.

2.2.1 Track geometry monitoring

Ideally, railway vehicles would be guided along a smooth path, following the intended design parameters. In practice, there are deviations from this ideal state which can lead to unwanted vehicle responses such as poor ride quality and derailment (WESTON et al., 2007a). Many factors influence the track's degradation, such as wagon speed, weather, topography, track characteristics, and components such as ballast, fasteners, sleepers, etc (BAI et al., 2016; BERAWI, 2013; SOLEIMANMEIGOUNI; AHMADI; KUMAR, 2016; YEO, 2017). Defects are not deterministically known a priori; they are stochastic on where and when they occur (GERUM; ALTAY; BAYKAL-GÜRISOY, 2019). Only after inspection and labeling the defect will be quantified; therefore, it is important to continuously measure track irregularities.

Track irregularities are typically measured monthly using a track geometry car. This interval is chosen due to costs, track access rights, and other maintenance issues, since operation is interrupted in the whole segment of the track being inspected (BAI et al., 2016; DE ROSA et al., 2020; MOHAMMADI et al., 2019; PIRES et al., 2021; TSUNASHIMA; NAGANUMA; KOBAYASHI, 2014). This approach is, therefore, not optimal for diagnostics. An example of a track recording car is shown in Figure 4. This vehicle can measure gauge, cross-level (superelevation), twist, alignment, and longitudinal profiles with a sampling rate of 0.25 m. It can use either a mechanical contact system or an optical system that measures rail positions using an inertial platform as a reference (WESTON et al., 2007a).



Figure 4: Track geometry car EM 120 - (BERAWI, 2013).

An example of a measurement system used by another track geometry car is shown in Figure 5. The instrumentation consists of vertical and lateral accelerometers (ACC), a gyroscope, linear variable differential transformers (LVDTs), and an optical system with lasers and cameras. Different track geometry cars can have different sensor arrangements for measuring track irregularities.

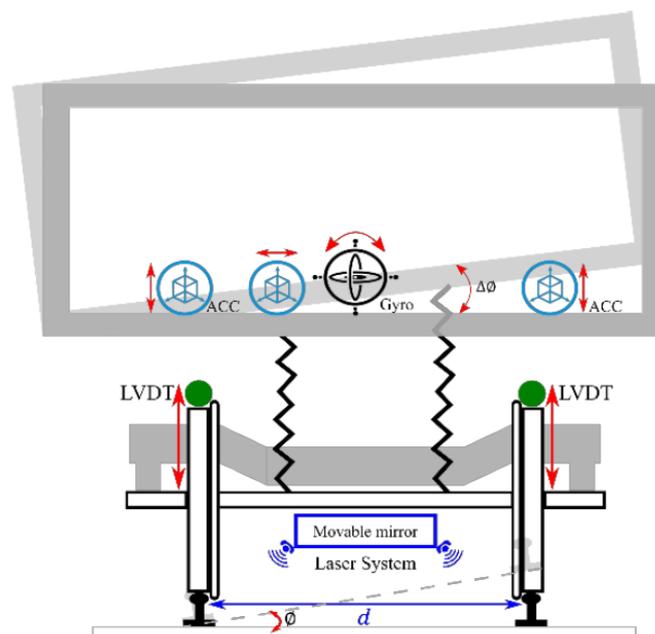


Figure 5: Track measurement system installed on a track geometry car - (KHAJEHEI, 2021).

2.2.2 Dynamic response monitoring

The alternative to track recording cars involves relying on inferring information from real freight vehicle motion (TAN et al., 2018a; THOMPSON et al., 2016; URDA et al., 2021; WESTON et al., 2007a, 2007b). This is done by monitoring the dynamic response (vehicle vibrations) of the vehicle with respect to track excitations by using sensors. Since vehicle vibrations are expected to correlate with track irregularities, the magnitude of the dynamic response is considered a possible means of assessing general track condition trends (BAI; LIU; LI, 2020; SANTOS; REICHL, 2014; TAN et al., 2018a; THOMPSON et al., 2016; TSUNASHIMA, 2019; URDA et al., 2021).

To better illustrate the inherent relationship between track excitation and the dynamic response, consider the half-car suspension model shown in Figure 6. A series of springs and dampers connect the bodies. The track excitations on the bottom will cause a dynamic response to the bogie mass m_{wr} and wagon mass m_b . Sensors are put on specific locations to measure these responses, such as an accelerometer s_1 placed on the bogie and the secondary spring displacement measurement s_2 . From these measurements, it is possible to infer the state of the track from the measured responses. Instrumented railway vehicles (IRVs) are the name given to the vehicles commonly used to measure these responses.

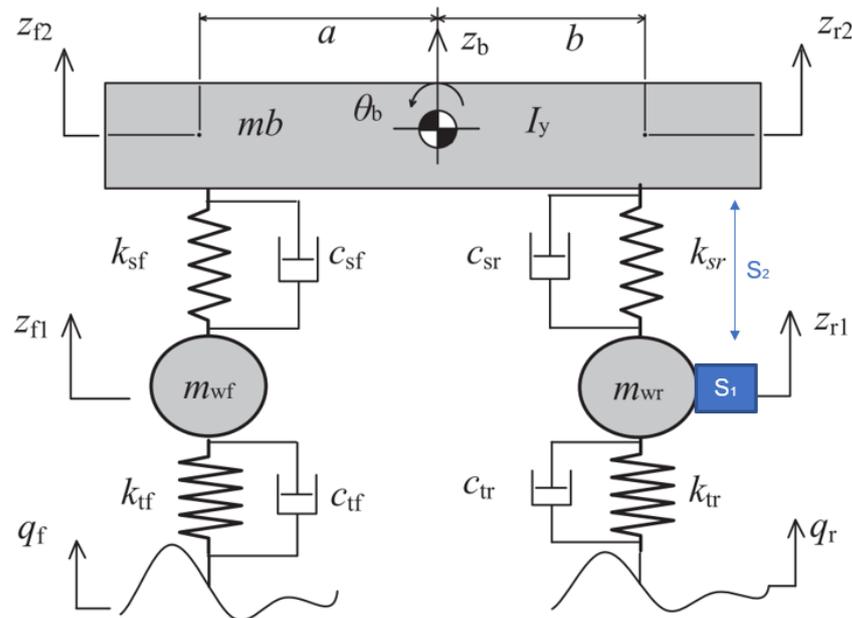


Figure 6: Half-car suspension model – modified from (ZHANG et al., 2021).

An instrumented railway vehicle is defined as a regular railway service vehicle that is equipped with sensors and instrumentation capable of measuring and communicating operation conditions. These vehicles act as probes, serving to detect and analyze real-time vibrations and sensor data during regular operation. They are also capable of informing essential maintenance areas using GPS information.

When the IRV monitored variables exceed a specified threshold, a track defect flag is generated. Depending on the severity of the flag, maintenance workers are assigned to visit the defective track segment to check and review these defects using other inspection tools, such as track geometry trolleys (BAI; LIU; LI, 2020). True and false positives are identified by these workers as well as their likely causes.

The main advantages of using IRVs over track geometry cars can be summarized as follows (CHONG et al., 2017; KOJIMA; TSUNASHIMA; MATSUMOTO, 2006; LEDERMAN et al., 2017a; LI et al., 2006; RAVITHARAN, 2019; SANTOS; REICHL, 2014):

- Operation is not interrupted since it is a regular wagon, but with added instrumentation;
- Measurements of the vehicle's response are performed in near real time and under normal operating conditions for any loading scenario;
- The higher frequency of trips compared to track geometry cars allows more passes over areas of interest, meaning greater statistical confidence about the state of the track and faster detection of defects;
- The use of vehicle responses allows the identification of critical defects which could not have been identified solely from geometry parameters, thus improving maintenance operation by complementing geometry-based track assessment.

The work of LI et al. (2006) helps illustrate the last topic. In his work, around 50% of the cases of poor vehicle performance cannot be detected by solely analyzing exceptions in the track geometry. Figure 7 exemplifies this, where bad dynamic performance is characterized by the ratio between the lateral and vertical forces (L/V) being above the standardized value of 1.0, as defined by the Association of American Railroads (AAR). Individually, none of the geometric parameters exceeded the Federal Railway Administration (FRA) standards, however, the combination of these deviations resulted in the vehicle exceptions shown.

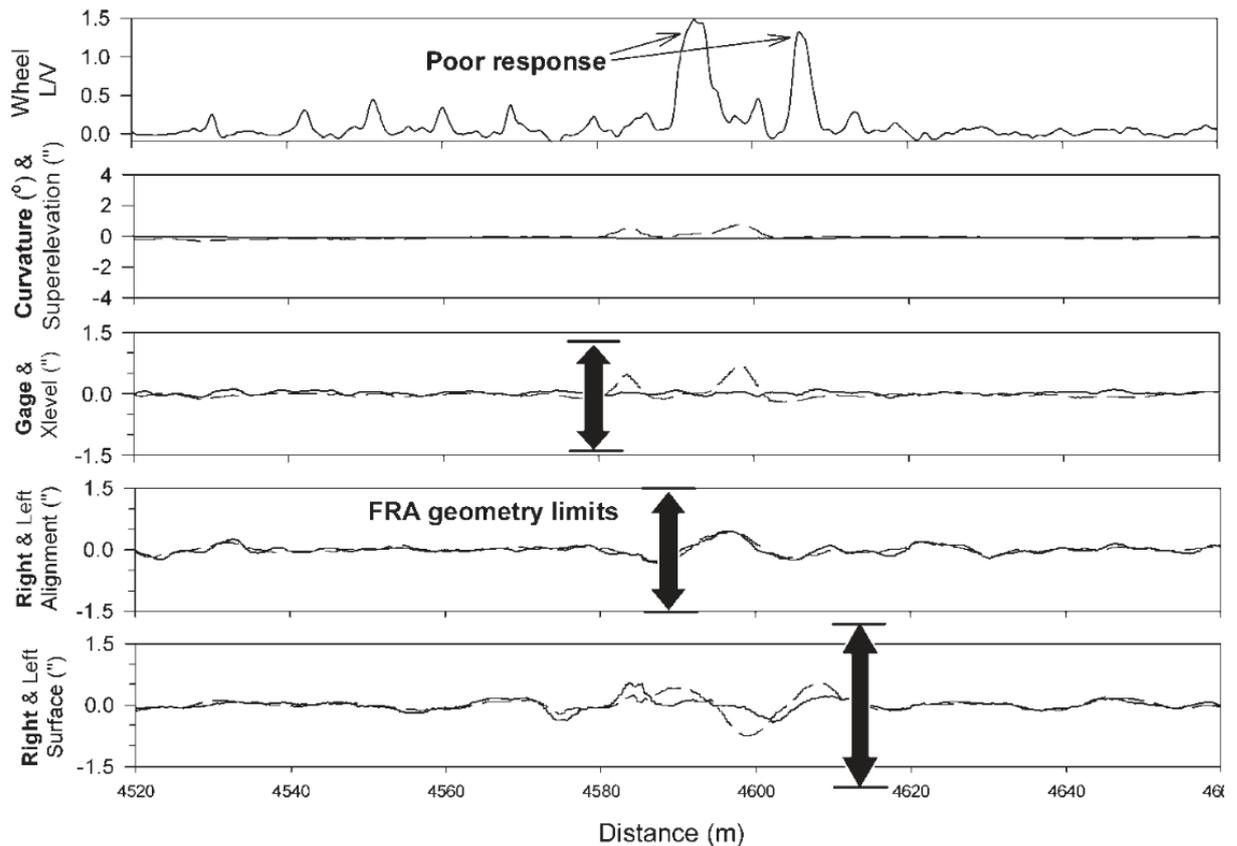


Figure 7: Vehicle response recording at 69 km/h on tangent segment due to combined cross-level, alignment and surface deviations - (LI et al., 2006).

There are downsides to consider, however. The first problem is that the sensors are in an aggressive environment subject to large vibrations, which can lead to early sensor failures and erroneous measurements. A readily available maintenance team is needed to avoid downtime and maximize availability and monitoring. Data transmission needs to be well thought out. For real-time applications, for example, 3G transmission is necessary, which will require extensive data storage capacity, organization (in both the database and data pipelines), and easy access within the organization for usage.

Figure 8 shows an IRV used in a Brazilian railway. In the left figure, there are solar panels installed on the wagon side, so the system is self-sustainable. The right figure shows a zoomed image of one of the sensors installed in the wagon. This sensor highlighted by the red circle is a spring attached to a load cell, which is responsible for measuring the vertical displacement of the secondary suspension system. Knowing the spring's elastic modulus and the applied force due to the load cell, it is possible to estimate the spring's displacement. Many researchers

employ this sensor in their IRVs (COWIE et al., 2015; HARDIE et al., 2016; THOMPSON et al., 2016). An example of another IRV is shown in Figure 9.



Figure 8: Example of an instrumented railway vehicle (IRV) used in a Brazilian railway. The left figure shows some of the IRVs instrumentation such as solar panels along the carbody. The right figure highlights one of its sensors (load cell attached to spring) for measuring spring displacement – modified from (SANTOS; REICHL, 2014).

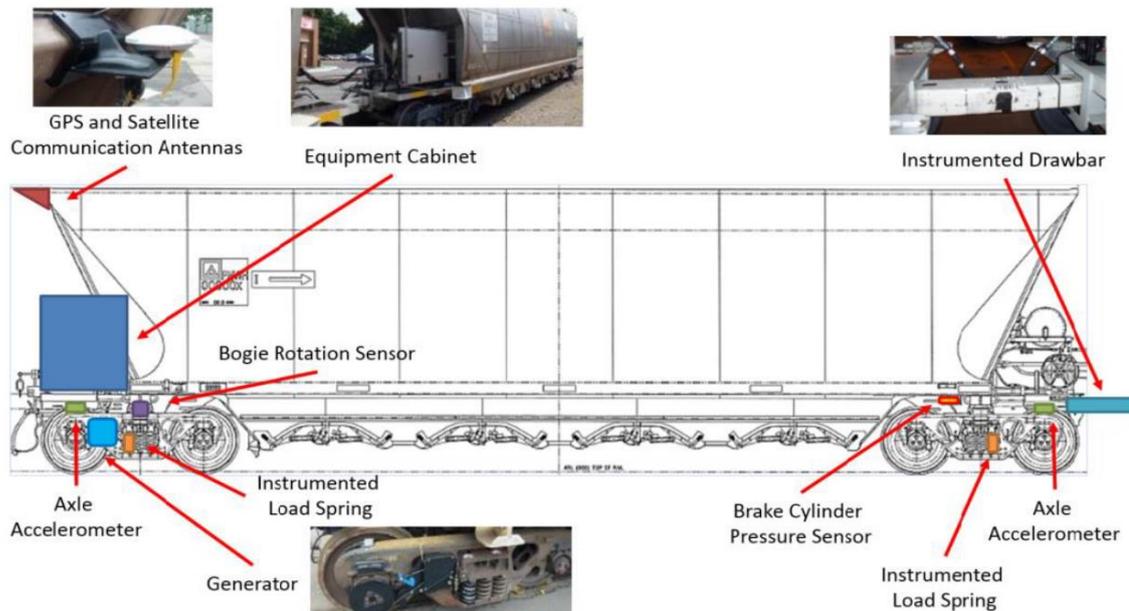


Figure 9: IRV instrumentation layout - (HARDIE et al., 2016).

There are many examples in the literature of researchers that used data measured by IRVs to either infer track quality, predict asset degradation, or detect local track defects. Thompson

et al. (2016) used an IRV for track degradation prediction based on different wagon classes. Hardie et al. (2016) used an IRV for track defect detection, peaked/dipped weld detection, track stiffness monitoring, and maintenance forecasting. This IRV also measures geometric parameters such as rail vertical alignment, crosslevel, and twist. Cowie et al. (2015) used an IRV for maintenance forecasting, detection of damaged welds, track defect detection, and coupler force monitoring for fatigue damage analysis. This IRV also measures superelevation and twist.

Although the BRA1 IRV doesn't measure roll and yaw to obtain superelevation and track twist like in Cowie et al. (2015) and Hardie et al. (2016), there are similar variables measured to those employed by the researchers stated above. The measured variables of the BRA1 are described in detail by Pires et al. (2021). They are:

- **Vertical acceleration of the bogie:** detects small impacts caused by track defects (uniaxial accelerometers). Two sensors are located on the left and right sideframe above the first wheelset of the leading bogie. Two other sensors are located on the left and right sideframe above the last wheelset of the trailing bogie;
- **Triaxial acceleration of the carbody:** monitors wagon stability due to Hunting (triaxial accelerometer). Located on the carbody above the trailing bogie;
- **Vertical displacement of suspension springs:** monitor dynamic wheel loading and it is used to establish velocity restrictions (DARBY et al., 2003). It is a load cell tied to a spring located on the left and right side of the secondary suspension on the leading and trailing bogie;
- **GPS:** identifies track location and measures velocity;
- **Coupler force (only for GDE wagons):** used to better establish velocity restrictions in locations with large forces between wagons. Basically, it uses strain gauges on the shaft connecting two wagons. It is also employed to monitor longitudinal wagon dynamics (BOWEY, 2018a; COWIE et al., 2015; YUEN et al., 2018);
- **Brake pipe pressure:** used for monitoring brake applications and estimating their resulting forces (BOWEY, 2018a).

Figure 10 shows the location of the sensors on the BRA1 IRV.

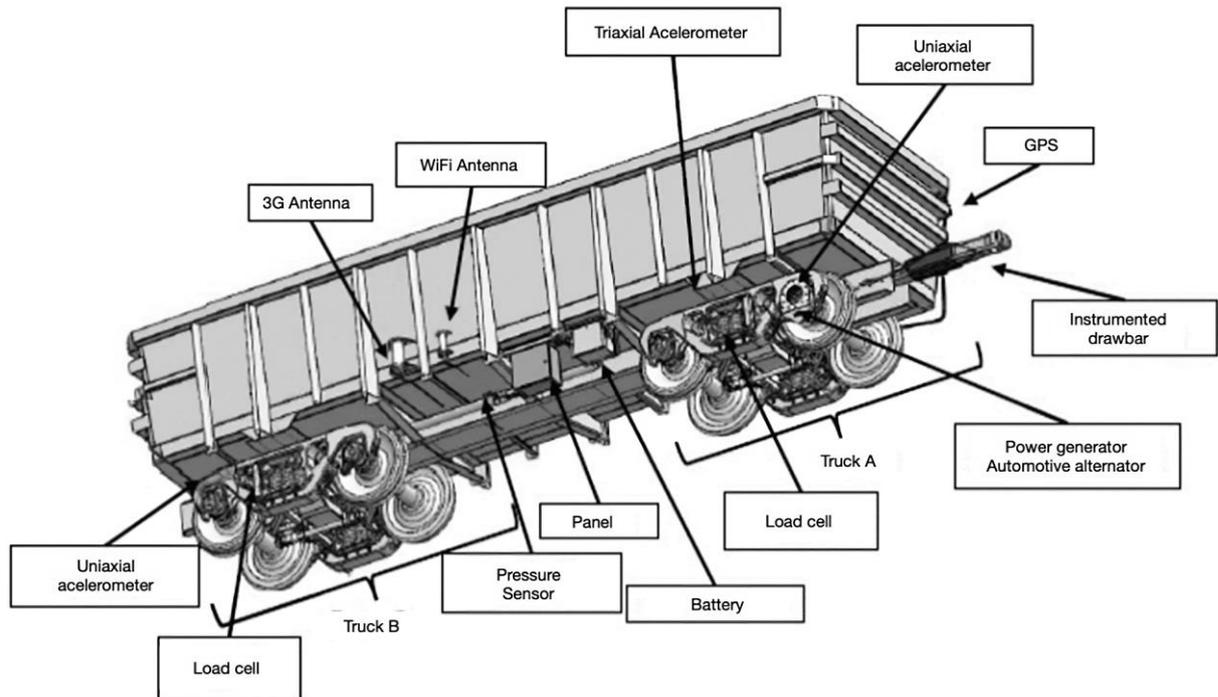


Figure 10: BRA1 Instrumented railway vehicle - (SOBRINHO et al., 2021).

2.3 Practical considerations for quantifying track quality

Through the bibliography review, many researchers highlighted problematic situations that occurred when validating their algorithms on real measured data. This section will talk about the most important factors to consider for a more practical application.

2.3.1 Using maintenance information

For maintenance activities such as tamping, the monitored signal is expected to improve after maintenance. Therefore, the behavior of the track segment before and after will likely differ, as shown in Figure 11. To model track improvement, the quality signals (time series) are split into segments between successive maintenances. Neuhold, Vidovic, and Marschnig (2020)

developed an algorithm that extracted tamping actions from quality signals after outlier removal and was confronted with the executed actions. In this way, recorded tamping actions are validated and the addition of missing tamping actions is also possible. Linear models are fitted for each segment to monitor signal degradation. Based on the model's prediction, the future state of the track can be estimated. With this information, a time frame can be established as to when maintenance must be done. This is shown in Figure 12, where the monitored variable is the standard deviation of a longitudinal track irregularity. IRV signals can also be used in this context, as shown in Figure 11 for a spring displacement variable.

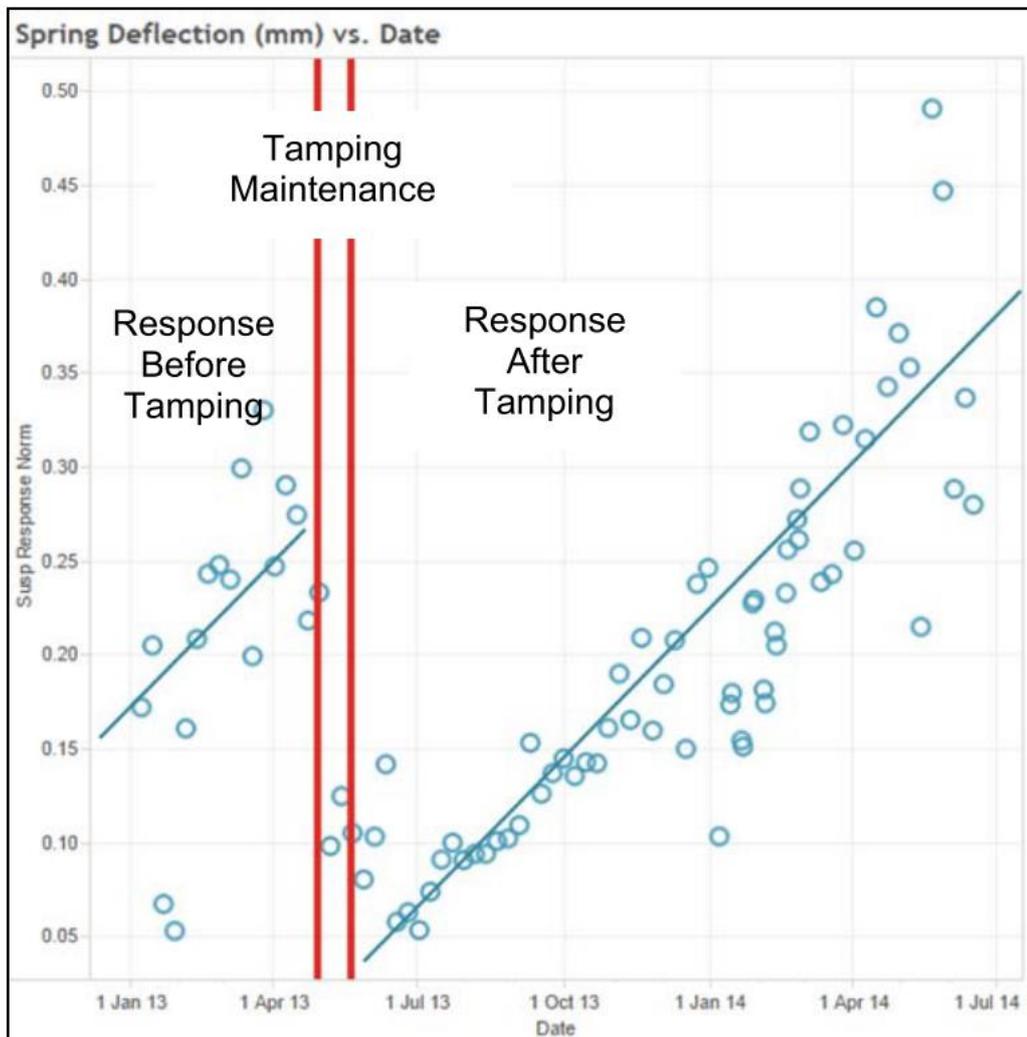


Figure 11: Spring deflection response before and after tamping - (MONAKALI; RAVITHARAN, 2016).

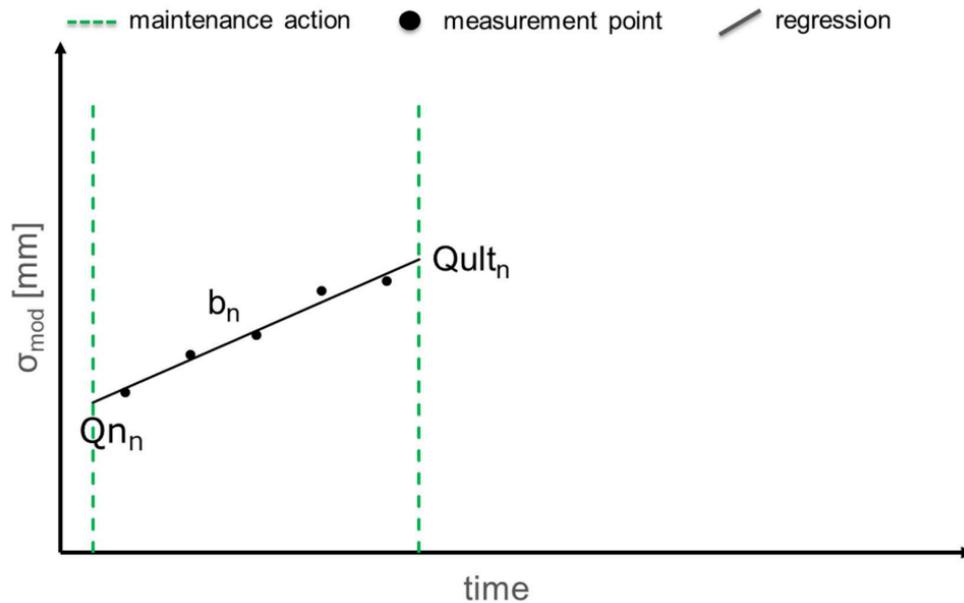


Figure 12: Linear regression of track geometry signal for two consecutive tamping actions (deterioration period) - (NEUHOLD; VIDOVIC; MARSCHNIG, 2020).

2.3.2 Data alignment and matching

Positioning accuracy between consecutive measurement trips of a specific track segment must have precise positional accuracy to avoid making incorrect conclusions (KHOSRAVI et al., 2021; LEDERMAN et al., 2017b, 2017a; NEUHOLD; VIDOVIC; MARSCHNIG, 2020). Figure 13 shows two types of positional error between two datasets: the first inspection dataset was conducted in February 2016 while the second dataset was conducted on two different dates in April and September 2016, respectively.

Figure 13a shows that the time series of the second inspection is shifted to the right, which makes analyzing the local defect difficult due to its positional uncertainty. The time series between successive measurement trips may differ in length due to different velocities during measurement, wheel sliding, geometry degradation, wheel wear, and environmental conditions (KHAJEHEI, 2021; KHOSRAVI et al., 2021). To the left of the 130.32 km mark, a shift was detected between successive inspections. So, considering the 130.31 km mark, it seems that a severe track condition appeared, increasing the longitudinal level from around -1 mm to around -4.5 mm, but that “defect” was due to a misalignment.

Figure 13b highlights that the time series can also be compressed or stretched, as seen by the different shift sizes. Wheel sliding causes compression since the traveled distance is more than what was calculated from wheel rotation, while wheel slippage causes stretching because the traveled distance is less than the calculation resulting from wheel rotation, supposing that an odometer is used (KHOSRAVI et al., 2021).

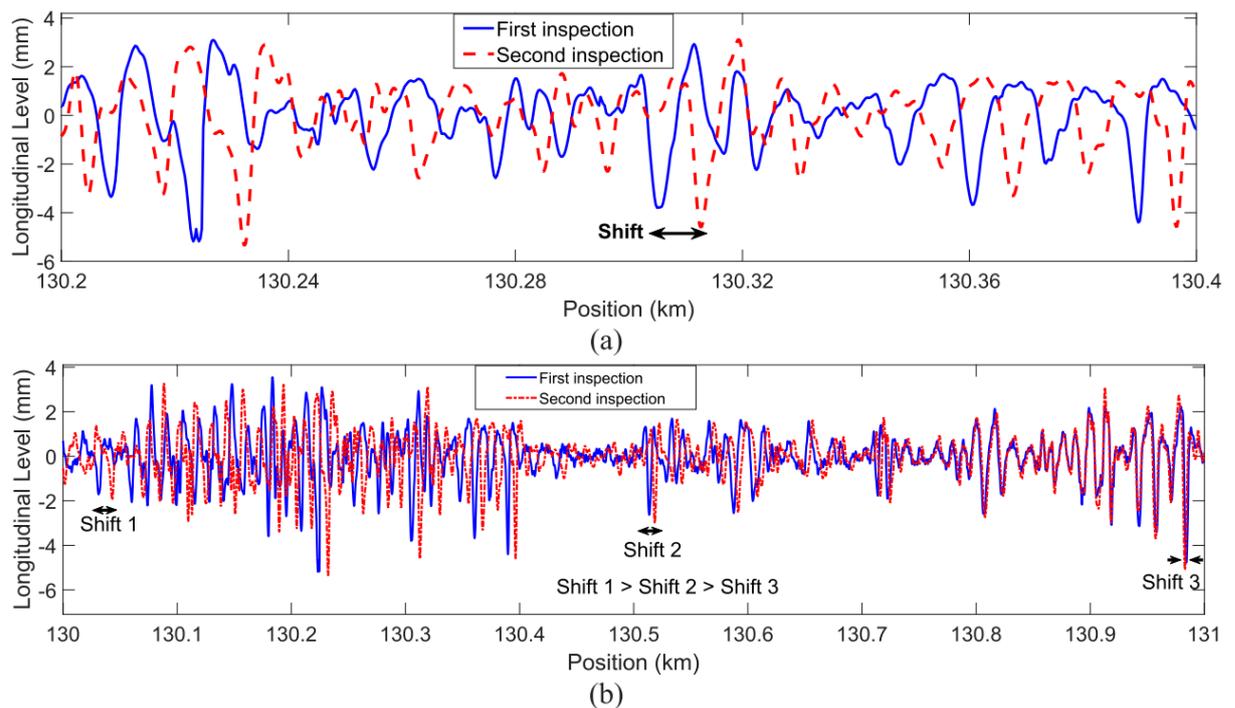


Figure 13: Positional error in two datasets from two different inspection runs. (a) A constant shift between the two datasets, (b) compression of the second dataset to the first - (KHOSRAVI et al., 2021).

2.3.3 Track segments with different characteristics

Railway networks have numerous rail segments that differ in terms of their defect proneness due to conditions such as loading, seasonal conditions, wear, fatigue, etc (GERUM; ALTAY; BAYKAL-GÜRSOY, 2019). Although a generalized model prediction method is desired, segments can lack data or have very different characteristics. Exploratory data analysis (EDA) is necessary to determine which is the best approach for the given data.

Gerum, Altay, and Baykal-Gürsoy (2019) used clustering to pool similar segments to obtain sufficient data for defect detection. The segments were grouped into 14 clusters using the K-means algorithm. Clusters 1 to 4 contain 24, 28, 47, and 41 rail segments with sufficient

data for prediction. Therefore, only clusters 1 to 4 were evaluated, representing 140 of the 170 rail segments (82.4%).

The dataset from Kasraei, Zakeri, and Bakhtiary (2020) contained 45 kilometers of track geometry data measured by a track recording car and stored each 0.25 m. The track was divided into 225 sections of 200 m for the left and right rails. The K-means algorithm was used to separate these 225 track sections into groups with the most similarity to decrease modeling uncertainty. The optimal number of clusters was 9, however, clusters 2 and 8 had fewer data. Therefore, these clusters were omitted and only 7 clusters were used for analysis (89% of original data).

2.3.4 Outlier removal

To exemplify the importance of outlier removal, consider both Figure 14 and Figure 15 from the paper done by Neuhold, Vidovic, and Marschnig (2020). From Figure 14, the second and last regression lines are steeper when compared to the other three, which means that the track deteriorated rapidly. By studying the measurements more closely and comparing them to executed maintenance data, the researchers discovered that two measurements are likely to be outliers (red measurement points 8 and 15), which lead to a false deterioration rate and period. The presence of these outliers would have led to 6 maintenance actions.

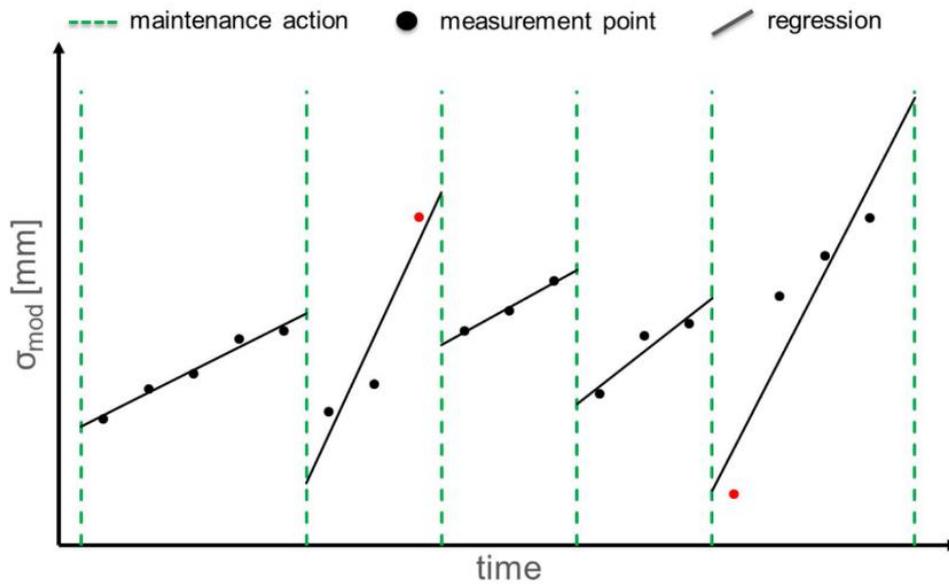


Figure 14: Regression analysis with 5 deterioration periods. Outliers are highlighted in red - (NEUHOLD; VIDOVIC; MARSCHNIG, 2020).

The regression model after removing these outliers is shown in Figure 15. Now only three deterioration periods are present with a similar degradation rate, leading to 3 maintenance actions instead of 6. Therefore, outlier detection and removal must be done when quantifying track quality.

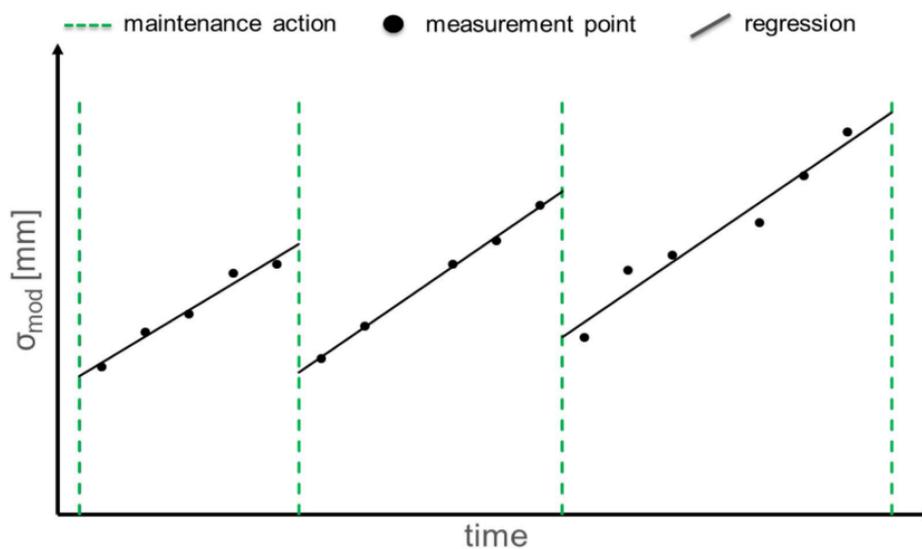


Figure 15: Degradation model after detection and elimination of outliers - (NEUHOLD; VIDOVIC; MARSCHNIG, 2020).

2.3.5 Robustness issues

Condition monitoring must be robust to data uncertainties such as vehicle speed and position uncertainty. Lederman et al. (2017a) highlighted that different speeds cause different vehicle dynamics and affect data alignment between different passes on the track or data. For sensors placed on the carbody, the filter effect of the suspension system also depends on the velocity, so these sensors will stress additional problems. They also state the larger position uncertainty, which makes detecting track changes harder. A unidimensional mass-spring model (toy model) was studied to better comprehend the effect that vehicle speed and position uncertainty had on the condition monitoring of trains.

Figure 16 shows two passes of different speeds for the toy model in the time domain (a) and spatial domain (b). If the velocity is zero, this means that the train stopped to drop passengers at a station. The defect in the spatial domain is shown in the letter (d). Although the defect is the same for both passes, the different speed profiles cause the mass-spring systems to experience a different excitation in the time domain, as shown in letter (c). The excitation at different points in time leads to a different dynamical response in the time domain, shown in letter (e). The response in the spatial domain (f) between both passes has better similarity, although there continues to be some effect due to speed variation seen at the end of the signal. Data alignment needs to be used to correct this problem. Alignment techniques such as Dynamic Time Warping (DTW) are commonly used to solve this problem (TAN et al., 2018b).

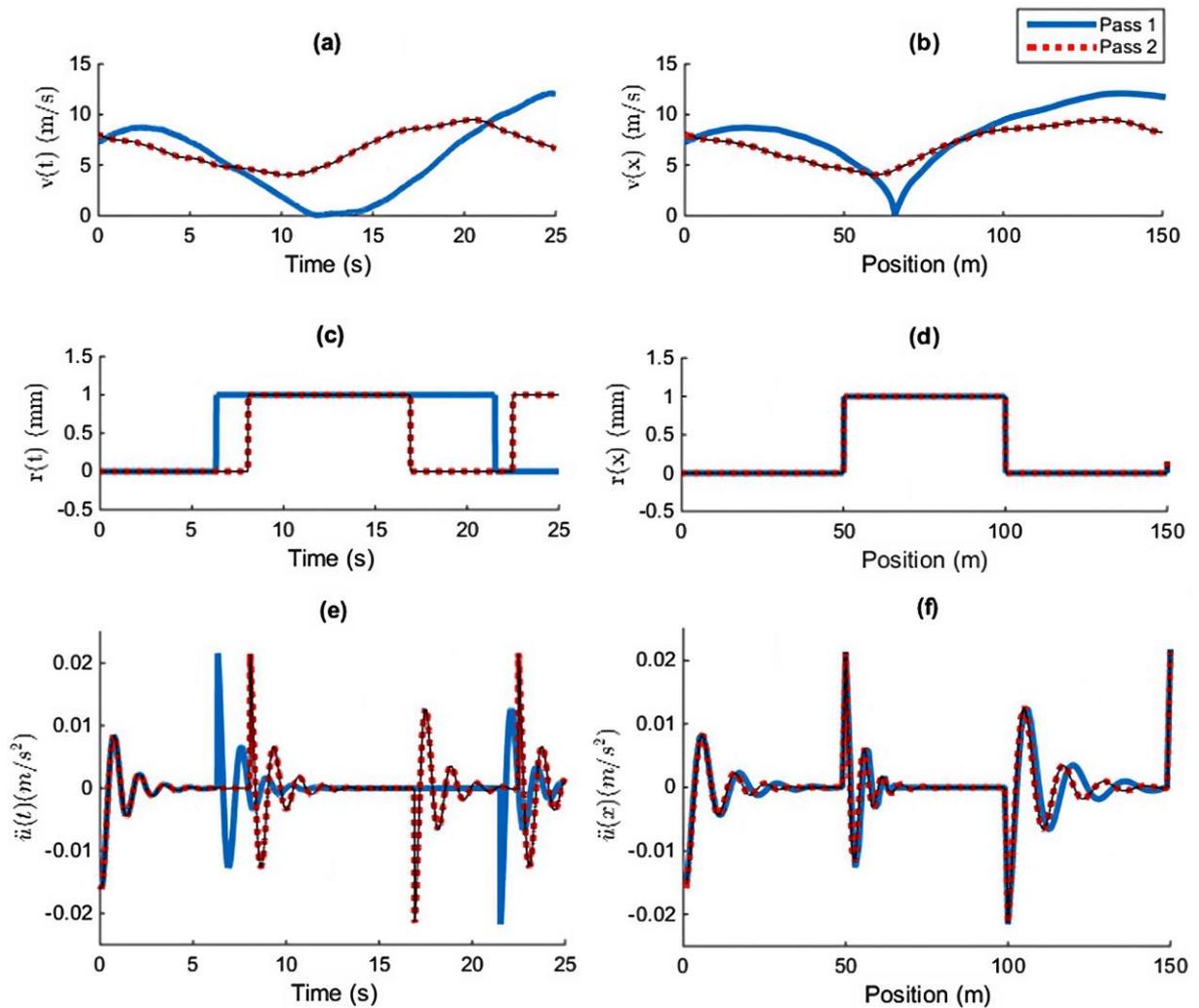


Figure 16: Illustration of how different speed profiles affect the measured vibration. Letter (a) shows the speed profiles in the time domain; (b) speed profiles in the spatial domain; (c) the roughness interpolated in time; (d) the roughness in space; (e) the acceleration of the oscillator in time, the box highlights two oscillations from the same bump which occur at different points in time due to the oscillator's speed profile; and (f) the acceleration of the oscillator in space - (LEDERMAN et al., 2017a).

2.3.6 Correcting false positives when detecting track defects

Bai et al. (2020) gave an example of how measurement errors can disrupt accurate maintenance planning. Suppose that an IRV runs over a specific track segment with the same track defect being detected and recorded repeatedly. Measurement bias or different vehicle speeds can easily cause the track defect to be detected in different locations, leading to the misinterpretation that there are multiple defects. These false positives must be minimized. Thus,

the authors proposed a data-driven bias correction and defect diagnosis model for in-service vehicle acceleration measurements (DBCDD-IVAM) to eliminate false positive detections of track vibration defects and to diagnose the causes of track vibration defects. The proposed model comprises two parts: a bias correction sub-model (BC-IVAM) and a defect diagnosis sub-model (DD-IVAM).

Figure 17 shows how the BC-IVAM sub-model works. The horizontal axis represents the location of track vibration defects along the railway line, and the vertical axis represents the inspection date for track vibration defects. The points represent different track vibration defects detected on different dates. Two constants C_{FT} and C_{FD} are used to describe a rectangular area that represents the tolerance for measurement bias. If a track vibration defect point u_i is within this rectangular area, these defect points are assumed to result from the same track geometry defect. The most recent point u_j in the rectangular area is selected as the representative track vibration defect. If the total number of track vibration defect points that fall within the rectangular area is greater than or equal to a limit C_{FN} , it is a true positive. If it is less than C_{FN} , the defect is due to the rolling stock itself or from the external environment instead of track irregularities. For this scenario, these points are classified as false positives.

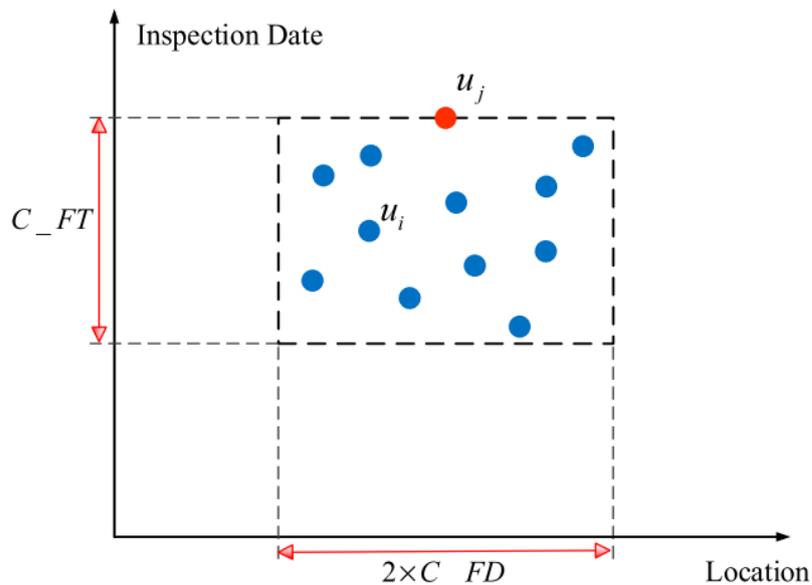


Figure 17: Algorithm schema of the BC-IVAM sub-model for minimizing false positives - (BAI; LIU; LI, 2020).

To illustrate this, consider $C_{FN} = 8$ points. From Figure 17, there are 11 blue points (variable u_i). This means that all 11 blue points are due to the same track defect, thus, are not

false positives. The most recent track defect point u_j is considered the representative defect of all previous points.

Although Bai, Liu, and Li (2020) used expert knowledge to define the constants C_{FT} , C_{FD} , and C_{FN} , an optimization problem to minimize the false positive rate with these constants as the optimization parameters can be defined. The main takeaway here is that a confidence range for successive inspections should be created to minimize the detection of false positives, which would decrease the workload of maintenance workers. The result of BC-IVAM is shown in Figure 18 with an 84.1% success rate. The DD-IVAM sub-model has a success rate of 75.8%.

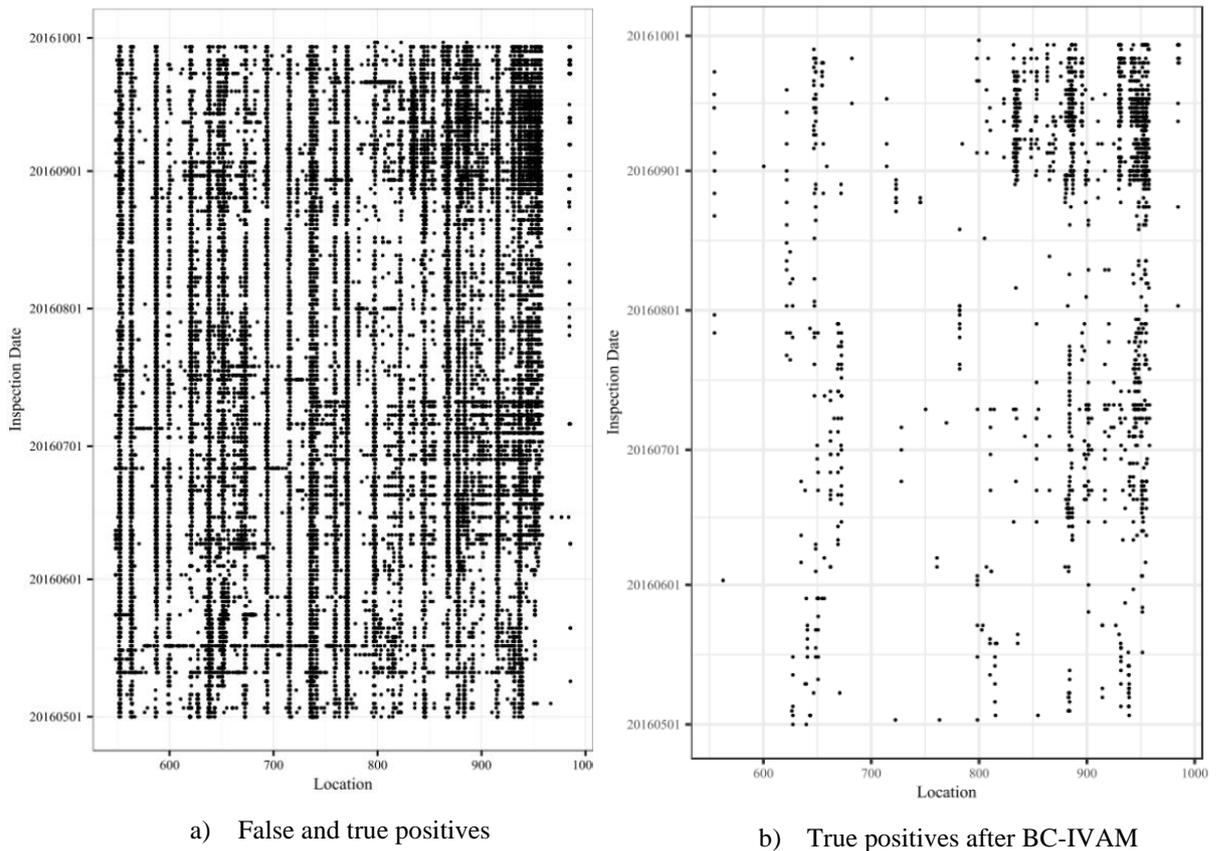


Figure 18: Distribution map of vibration defect data in the down direction of the Lanxin Railway between km 548 and km 985.5 - (BAI; LIU; LI, 2020).

2.3.7 Summary of practical considerations

- Track degradation and renewal must be validated and confronted with maintenance data;
- Data alignment must be done to avoid drawing incorrect conclusions;
- Using statistical metrics of the raw signal can help correct alignment errors depending on the specified length for computation;
- Outlier removal is a necessary step when quantifying track quality. Besides traditional statistical analysis, maintenance data will help determine which data point is, in fact, an outlier;
- Track quality monitoring techniques must be robust towards data uncertainties such as vehicle speeds and position uncertainty. Alignment algorithms can help with this issue.

2.4 Instrumented railway vehicle sensors

IRV sensors for measuring dynamic responses can be divided into 3 main groups based on their location: carbody, bogie, and axlebox sensors. This section will detail the sensors of each respective group, highlight their advantages and drawbacks, and show the IRV sensors used in this thesis.

2.4.1 Carbody mounted sensors

Carbody mounted sensors are typically accelerometers used to monitor wagon stability and the occurrence of hunting oscillation (BARBOSA, 2016; ODASHIMA et al., 2017; PIRES et al., 2021). Tsunashima et al. (2019) detected the presence of corrugation in the track by observing the vertical carbody acceleration. They noted that this signal is greatly influenced by low-frequency vibration, masking the effect that corrugation has on the signal. This causes corrugation detection to be a difficult task with traditional threshold boundary limits. Therefore,

signal processing techniques are required to be used in conjunction with the measurements to extract useful information. In this case, the root-mean-square (RMS) of the carbody acceleration was used as an index for monitoring track condition. A new work by the same authors stated that using the RMS metric results in a loss of frequency information and makes the analysis more difficult (TSUNASHIMA; HIROSE, 2020). To complement the RMS index, the Continuous Wavelet Transform (CWT) and the Hilbert-Huang Transform (HHT) were applied and compared to determine the best technique for detecting faulty tracks, which is shown in Figure 19. The color bars represent the magnitude of the signal.

Figure 19a shows a vertical carbody acceleration signal while letters (b) and (c) show the HHT of the signal at different frequency scales. Letter (b) reveals a large frequency component near 25.65 km detected at the same position where a joint depression exists. By zooming into a frequency scale between 0-5 Hz, Figure 19c shows a different defect, the degraded longitudinal level irregularity. Depending on the problem, different signal processing techniques and signal representations shall be tested to find the best representation, as was done by Lederman et al. (2017a) for detecting track defects from an instrumented train.

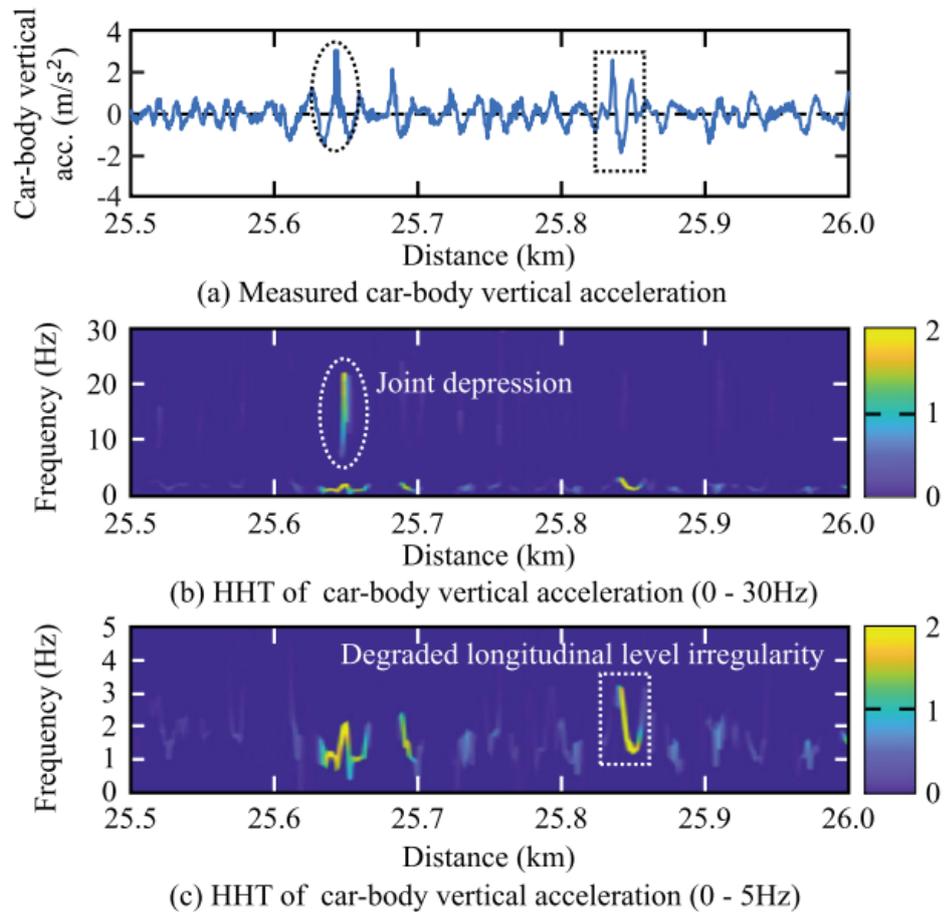


Figure 19: Letter (a) shows the carbody vertical acceleration. Letter (b) shows the Hilbert spectra for frequencies between 0 – 30 Hz, highlighting a joint depression. Letter (c) shows the Hilbert spectra for the range of 0 – 5 Hz - (TSUNASHIMA; HIROSE, 2020).

Track fault detection using instrumentation on the carbody or in-cabin does not have to deal with the harsh environment close to the wheel-rail contact interface like bogie or axlebox mounted sensors, so it is easier to maintain (MORI et al., 2010). One downside, however, is that the distinctive signal of track faults is hidden in the carbody vibration or is filtered out by the suspension system (MORI et al., 2010; TSUNASHIMA; HIROSE, 2020; WESTON et al., 2015). Signal processing techniques are necessary for in-cabin fault detection.

Tsunashima et al. (2012) used the wavelet transform to decompose the car body acceleration signal into components of different frequency bands using multi-resolution analyses (MRA). With each specific frequency resolution, different defects were detected from the original signal. The sum of the signal at each frequency band returned the original signal. This process is illustrated in Figure 20 and exemplified in Figure 21. In the case of Figure 20, the starting signal S was measured at 2 kHz, which means the signal has up to 1 kHz of

frequency content due to Nyquist's theorem. This 1 kHz signal is now split into two components: one from 0-500 Hz (a_1) and another from 500-1000 Hz (d_1). The component with the lowest frequency range is split again with the above methodology, but the starting signal now has frequency content between 0-500 Hz. At the end of the process, the signal of 1 kHz was split into the components d_1 , d_2 , d_3 , d_4 , and a_4 corresponding to 1000-500 Hz, 500-250 Hz, 250-125 Hz, 125-62.5 Hz, and frequencies not greater than 62.5 Hz, respectively. The user can specify the number of desired frequency bands, which was 10 in the case of Figure 21.

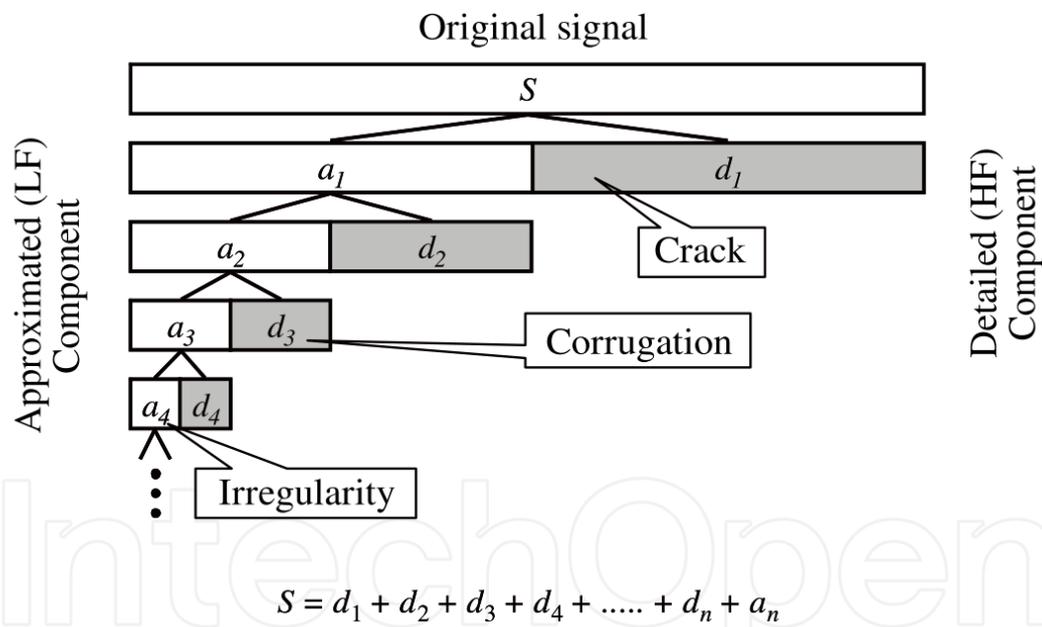


Figure 20: Multi-resolution analysis (MRA) using the discrete wavelet transform - (TSUNASHIMA et al., 2012).

Another issue is that the carbody acceleration waveform is considerably different from track geometry and the amplitude greatly depends on the vehicle speed (TSUNASHIMA; NAGANUMA; KOBAYASHI, 2014; YEO, 2017), highlighting the need for signal processing and possible robustness issues. In the case of Tsunashima et al. (2012), they measured their data on a curve with a radius of 202 m at a constant velocity of 38 km/h for detecting track faults from cabin vibration. Many researchers prefer to use a constant velocity and not deal with variations due to environmental or operational conditions.

Lee et al. (2012) stated that carbody acceleration is highly dependent on the primary and the secondary suspension, so the effect of the track irregularities is difficult to extract from such data. Therefore, measuring irregularity values from carbody mounted sensors is unlikely.

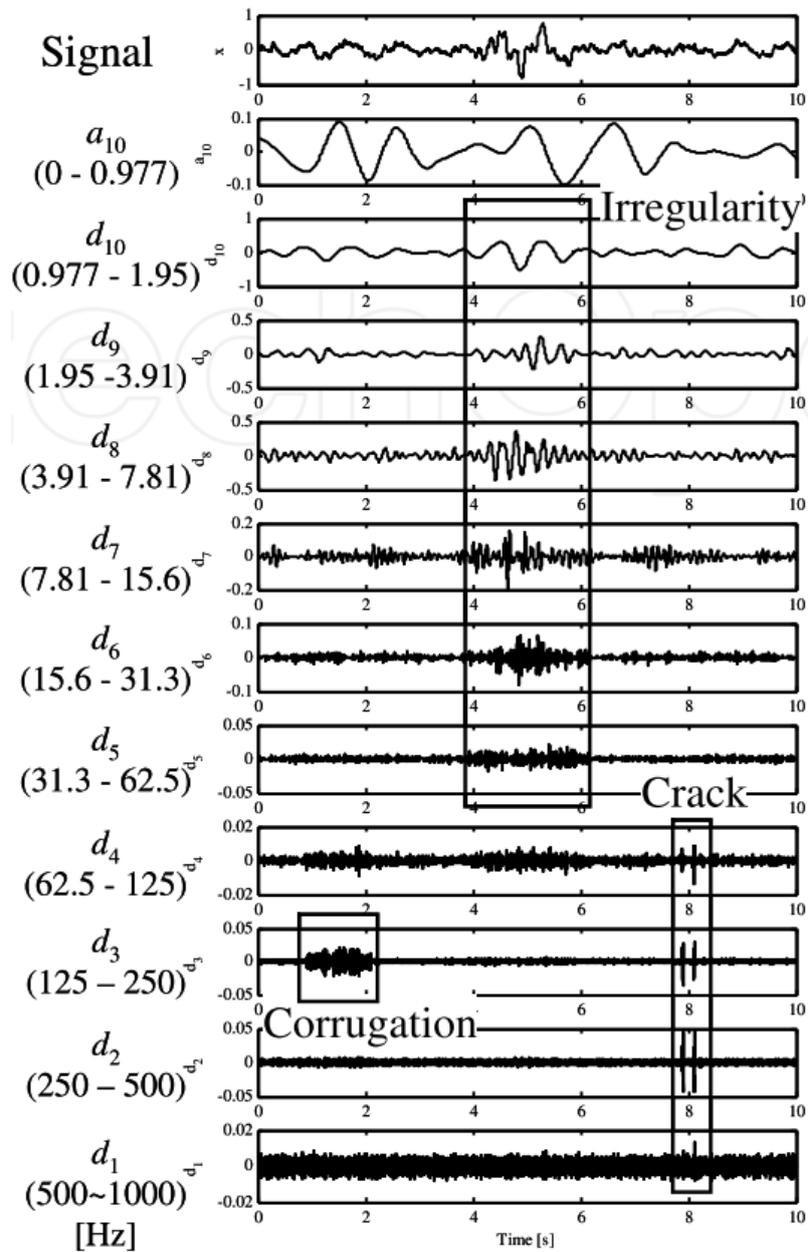


Figure 21: Multi-resolution analysis (MRA) of vertical carbody acceleration. Different types of detected track irregularities are highlighted - (TSUNASHIMA et al., 2012).

2.4.2 Axlebox mounted sensors

Axlebox mounted sensors consist mainly of vertical and lateral accelerometers. They are generally used for detecting track defects such as spalling, dipped welds, and insulated joints (DARBY et al., 2003).

Axlebox sensors tend to require higher frequency maintenance due to larger impact loads caused by the closer proximity to the wheel-rail contact interface (BAASCH et al., 2019; MORI et al., 2010; TAKIKAWA, 2012; TSUNASHIMA; NAGANUMA; KOBAYASHI, 2014). Due to the lack of filtering from the secondary suspension system, however, they can lead to a better representation of the dynamic response from track excitations. A comparison between carbody and axlebox accelerations was highlighted in the work of Tsunashima et al. (2012) and illustrated in Figure 22 for a curved track section with corrugation. The difference in signal amplitude and wave format is very clear; the vertical carbody acceleration stays relatively the same while the axlebox acceleration varies greatly.

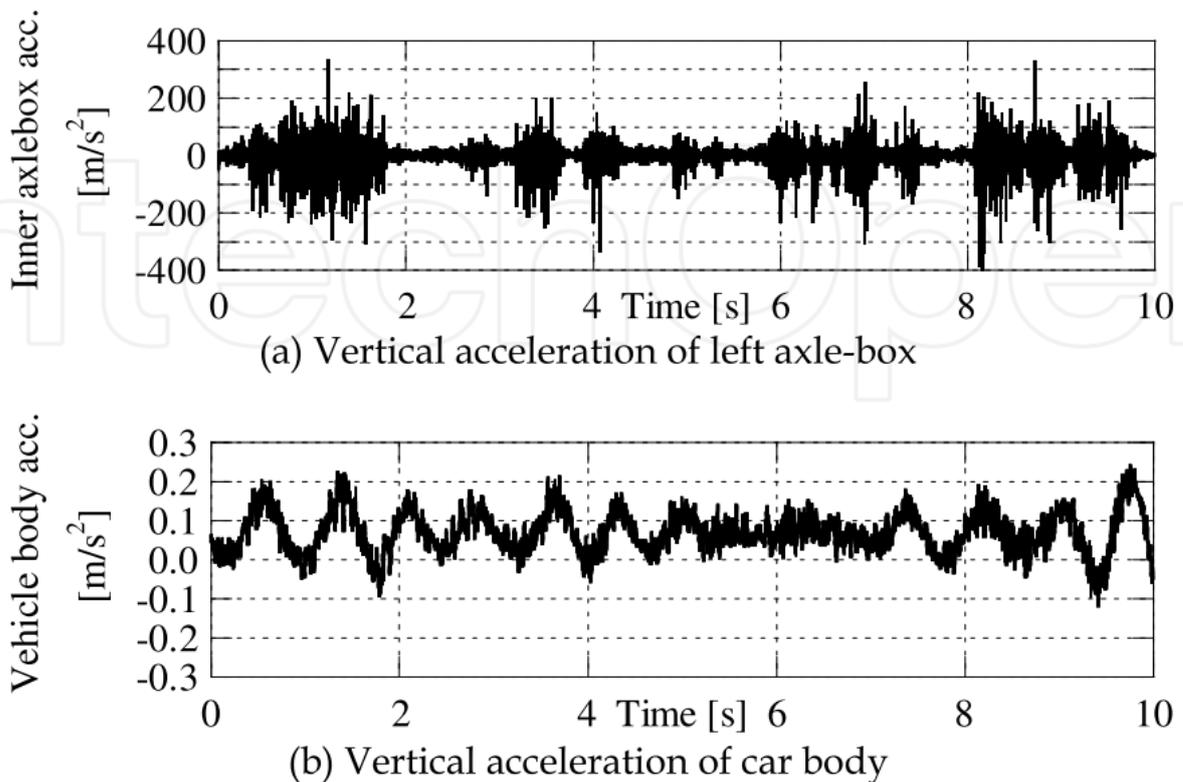


Figure 22: Measurement results of a curved section with corrugation. a) Axlebox sensor b) Carbody sensor - (TSUNASHIMA et al., 2012).

2.4.3 Bogie mounted sensors

There are many different types of bogie mounted sensors in the IRV setting. Typically, vertical and lateral accelerometers, gyroscopes, and displacement sensors are used. Lee et al. (2012) used lateral and vertical accelerometers mounted on the bogie and axlebox to estimate lateral and vertical track irregularities by using a Kalman filter alongside bandpass and compensation filters. Track irregularities obtained from both axlebox and bogie mounted sensors were compared to a track geometry measurement system. The placement of these sensors is shown in Figure 23.

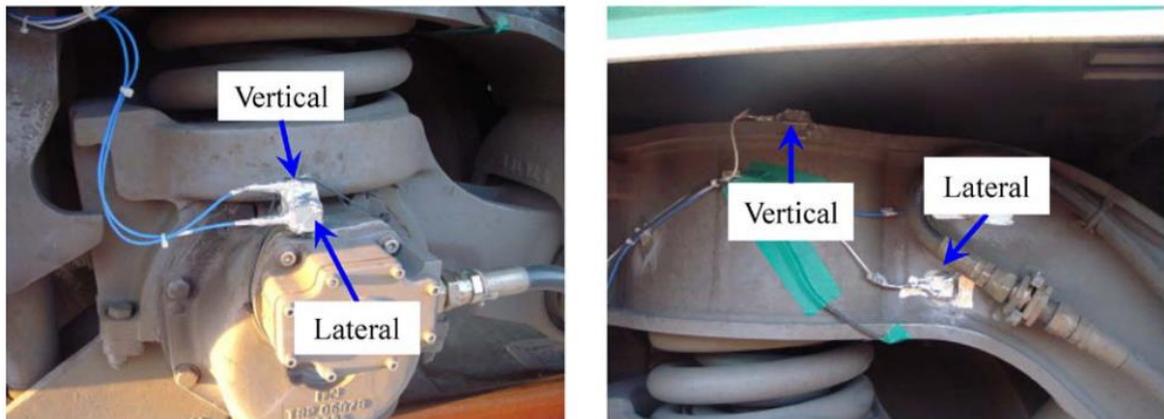


Figure 23: Setup for measuring acceleration. The left figure shows axlebox mounted accelerometers. The right figure shows bogie mounted accelerometers - (LEE et al., 2012).

Some authors use gyroscopes, roll, and yaw sensors to obtain geometric information about the track, such as orientation, curvature, twist, and crosslevel by using mathematical formulae (HARDIE et al., 2016; HESSER; ALTUN; MARKERT, 2022; YEO, 2017). Hardie et al. (2016) used a roll rate sensor for calculating track twist, while Lingamanaik et al. (2017) used both yaw and roll rate sensors to determine track curvature and superelevation. An example can be seen in Figure 9, where a roll rate gyroscope was secured to a crossbar on the bolster.

The main advantages that come with using bogie mounted sensors are their higher maintainability when compared to axlebox sensors, better signal representation compared to carbody sensors, and the possibility of using different types of sensors due to the many parameters to monitor such as the secondary spring's displacement. The larger number of locations to install sensors is also a factor. The main disadvantage is that some signal information can be lost due to the filtering effect of the suspension system when compared to

the axlebox mounted sensors. Overall, this is probably the best location to install instrumentation on a railway vehicle.

2.5 Obtaining track irregularities from accelerometer signals

Theoretically, vertical track irregularity values can be calculated by double integrating the vertical acceleration of the bogie or axlebox mounted sensors (BLEAKLEY, 2006; LEE et al., 2012; NADARAJAH et al., 2018; THOMPSON et al., 2016), as shown in Equation (2.1). The main difference is that bogie mounted sensors are filtered in the vertical direction by the primary suspension while axlebox mounted sensors aren't. For the lateral track irregularities, Lee et al. (LEE et al., 2012) stated that bogie-mounted accelerometers are very similar to those using the axlebox mounted accelerometers due to the lack of substantial lateral suspension when compared to the vertical direction. When considering more practical scenarios in obtaining both vertical and lateral track geometry from integrating acceleration, a few problems appear (BLEAKLEY, 2006; LEE et al., 2012; YEO, 2017):

- Large drift in either the positive or negative direction due to an offset (the accelerometer rarely gives a 'zero' output when acceleration is exactly zero);
- Double integration further exacerbates sensor drift. Yeo et al.(2017) applied a high pass filter to minimize low frequency offset and drift;
- Accelerometer output has some high frequency random noise due to sampling electronics.

$$z(t) = \int \int \ddot{z} dt dt \quad (2.1)$$

This conversion is necessary if comparisons are to be made between passes of the train over the same section of track at differing speeds, as comparisons will not be possible in the time domain (YEO, 2017). The conversion can be done by simply changing the time domain to the spatial domain from the sampling rate and vehicle speed.

Another option is to obtain the signal already in the spatial domain by finding the vertical curvature K_z of the track from the accelerometer signal \ddot{Z} and velocity V , as shown in Equation (2.2).

$$K_z = \frac{\ddot{Z}}{V^2} \quad (2.2)$$

Since vertical curvature is equal to the second derivative of z with respect to x , the vertical displacement \hat{z} can be obtained with Equation (2.3)

$$\hat{z} = \iint K_z dx dx = \iint \frac{d^2z}{dx^2} dx dx \quad (2.3)$$

Based on Equations (2.2) and (2.3), it can be deduced that, as the vehicle speed decreases, smaller accelerations are needed to obtain the same curvature and corresponding vertical displacement. At low speeds, the resulting vertical acceleration can be so small that it falls below the noise threshold of the accelerometer, causing drift (NADARAJAH et al., 2018; WESTON et al., 2007a, 2007b; YEO, 2017). This causes the calculated displacement to reflect accelerometer drift, as seen in Figure 24 by the large blue values at near zero vehicle speed (pink line), which surpass the scale of the figure. This issue is exacerbated by the presence of the squared speed term in Equation (2.3). In contrast, vertical irregularities obtained from the gyroscope are not as severely impacted due to the absence of the squared speed term.

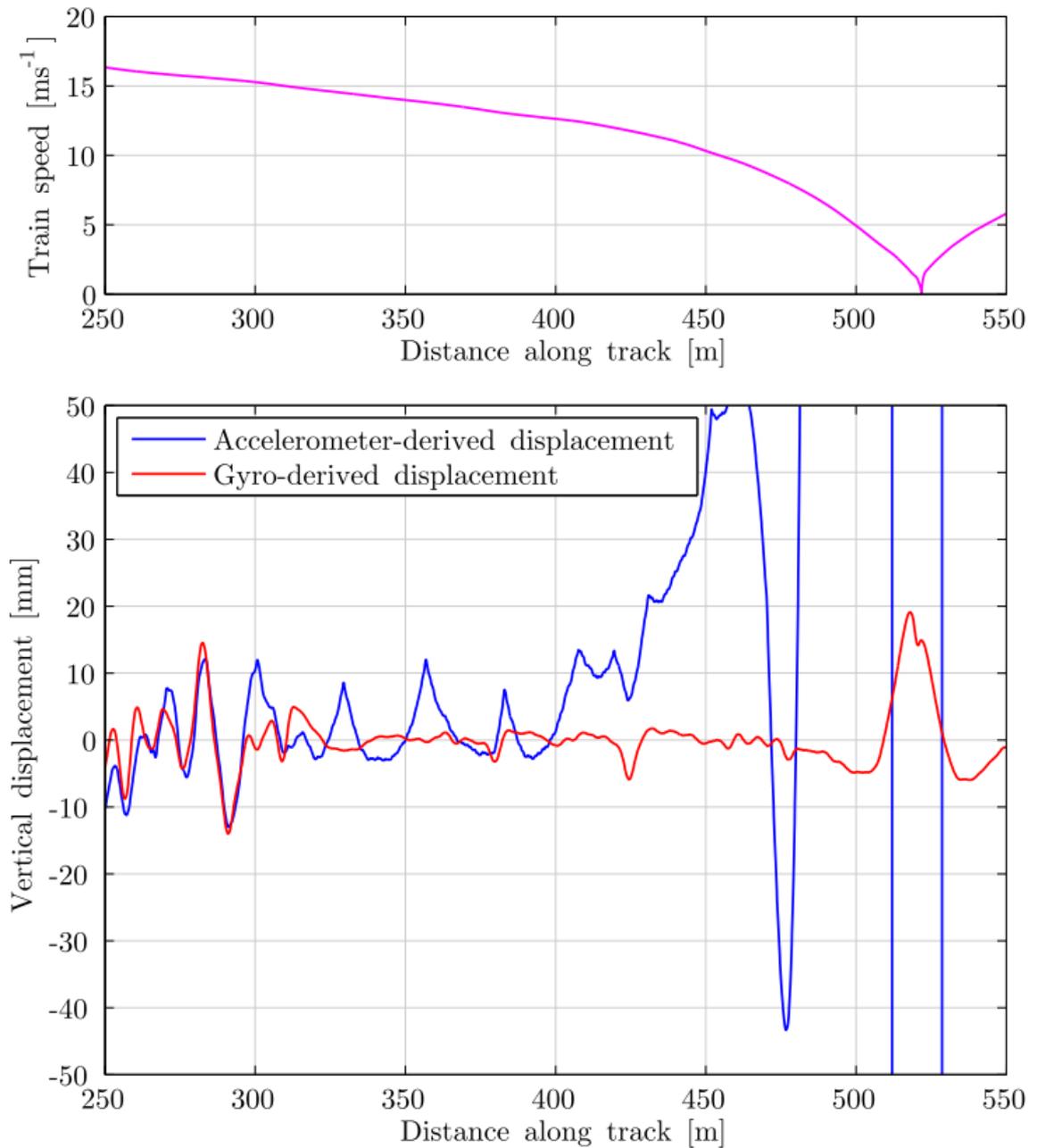


Figure 24: The effect of low vehicle speed on calculated vertical irregularities - (YEO, 2017).

2.6 Norms for evaluating track quality

This section will go over the two main norms used in this thesis to generate the artificial track irregularities that will be used in the multibody dynamic simulation model. By running

the simulation, the dataset for obtaining track irregularities from sensor variables will be generated.

2.6.1 European Regulation Standard (EN 13848)

The European Committee for Standardization (CEN) has created a group of standards EN 13848 for the characterization of track geometry and track quality. This standard is divided into five different parts:

- **Part 1:** Terminology and framework for specifying track geometry parameters such as track gauge, lateral and vertical irregularities, crosslevel, and twist;
- **Part 2:** Measuring system and track recording vehicle;
- **Part 3:** Track construction and maintenance machines;
- **Part 4:** Manual and lightweight devices;
- **Part 5:** Minimum requirements for track geometry quality and safety related limits for the geometry parameters defined in part 1.

Focusing on part 5 of EN 13848 (EN 13848-5), maintenance strategies are based on three different quality levels: safety limit, intervention limit, and alert limit. The formal definition for each level is shown in Table 1.

Alert limit (AL)	If a limit value is exceeded, an action to correct the error has to be considered in the regularly planned maintenance.
Intervention Limit (IL)	If a limit value is exceeded, an action to correct the error has to be done immediately before the next inspection.
Safety Limit (IAL)	If a limit value is exceeded, an action should be done to reduce the risk of derailment (closing the line, reducing speed, immediate tamping, etc.)

Table 1: Track quality levels from the norm EN 13848-5 - (BERAWI, 2013).

For each limit, the standard defines track quality based on the wavelength spans D1, D2 and D3 presented below (BERAWI, 2013).

- D1: wavelength of irregularities inside the range $3 < \lambda \leq 25 \text{ m}$ (related to safety);
- D2: wavelength of irregularities inside the range $25 < \lambda \leq 70 \text{ m}$ (related to ride comfort);
- D3: wavelength of irregularities inside the range $70 < \lambda \leq 200 \text{ m}$ (related to ride comfort).

To obtain a specific class, a bandpass filter is applied to the wavelength range of interest. Wavelengths below D1 ($0.03 < \lambda \leq 1$ for example) are related to corrugation, which results in an audible vibration (KARIS, 2018).

2.6.2 Federal Railroad Administration (FRA)

Track irregularities are normally modeled as being stochastic (or random). Thus, it is possible to describe their content in the frequency domain through its power spectral density (PSD) (PIRES et al., 2021). Countries like the United States, China, France, and Germany have modeled their respective spectrum of irregularities to represent their typical railway condition (BERAWI, 2013). These functions were obtained from measured data and are usually previously filtered and smoothed with statistical methods. Because of that, they represent the average track irregularity of their respective country.

The FRA categorized the quality of railway systems based on their one-sided PSD into nine different classes. The first six classes correspond to normal-speed railways while classes seven to nine are designated for high-speed tracks, accommodating passenger trains with a maximum speed of 145 km/h and freight trains with a maximum speed of 130 km/h. Note that these PSDs can only be applied for a wavelength range between 1.524 m and 304.8 m (BERAWI, 2013).

The mathematical expressions for the PSDs according to the geometric parameter are shown by Equations (2.4), (2.5), and (2.6):

For vertical alignment:

$$S_v(\Omega) = k \frac{A_v \Omega_c^2}{\Omega^2(\Omega^2 + \Omega_c^2)} \quad (2.4)$$

For lateral alignment:

$$S_{al}(\Omega) = k \frac{A_a \Omega_c^2}{\Omega^2(\Omega^2 + \Omega_c^2)} \quad (2.5)$$

For gauge and cross level (superelevation):

$$S_{gauge/cl}(\Omega) = k \frac{4 A_v \Omega_c^2}{(\Omega^2 + \Omega_c^2)(\Omega^2 + \Omega_s^2)} \quad (2.6)$$

Table 2 provides the coefficients for the PSDs according to their quality classification. Since the FRA PSD is very similar to the track conditions of the BRA railway, this standard and its PSDs are used for generating virtual track irregularities for multibody simulation and creating the dataset used in this thesis. The maximum permissible speed in this railway is 65 km/h for freight wagons and railway experts stated that they work with, at worse, FRA 3 irregularities. Therefore, line grades 6 to 3 were used in this study.

Class	Max velocity (km/h)		Parameters			
	Freight transportation	Passengers	A_v (cm ² rad/m)	A_a (cm ² rad/m)	Ω_c^2 (rad/m)	Ω_s^2 (rad/m)
1	16	24	1.2107	3.3634	0.6046	0.8245
2	40	48	1.0181	1.2107	0.9308	0.8245
3	64	97	0.6816	0.4128	0.8520	0.8245
4	97	129	0.5376	0.3027	1.1312	0.8245
5	129	145	0.2095	0.0762	0.8209	0.8245
6	177	177	0.0339	0.0339	0.4380	0.8245

Table 2: Coefficients for the power spectral density functions - (PIRES et al., 2021).

2.6.3 Defining the track irregularity wavelength range

The first step in generating artificial track irregularities is to correctly define the wavelength range. It is important to note that only a few specific wavelengths are responsible for affecting the dynamic response of the vehicle (BERAWI, 2013). This is due to the relationship that the excitation wavelength has with the vehicle speed and the vehicle's natural frequency, expressed by Equation (2.7).

$$\lambda = \frac{V}{f} \quad (2.7)$$

Where λ is the wavelength of the track excitation (m), V is the vehicle speed (m/s) and f is the natural frequency of the vehicle (Hz).

Equation (2.7) is represented graphically for different natural frequency values in Figure 25. Depending on the vehicle speed, different wavelength values would be responsible for exciting the railway vehicle. Since low velocity values do not typically cause large and dangerous dynamic responses in comparison to high velocities, the largest permissible speed value (65 km/h) is used to determine the wavelength range. For the BRA railway, this is the maximum permissible speed.

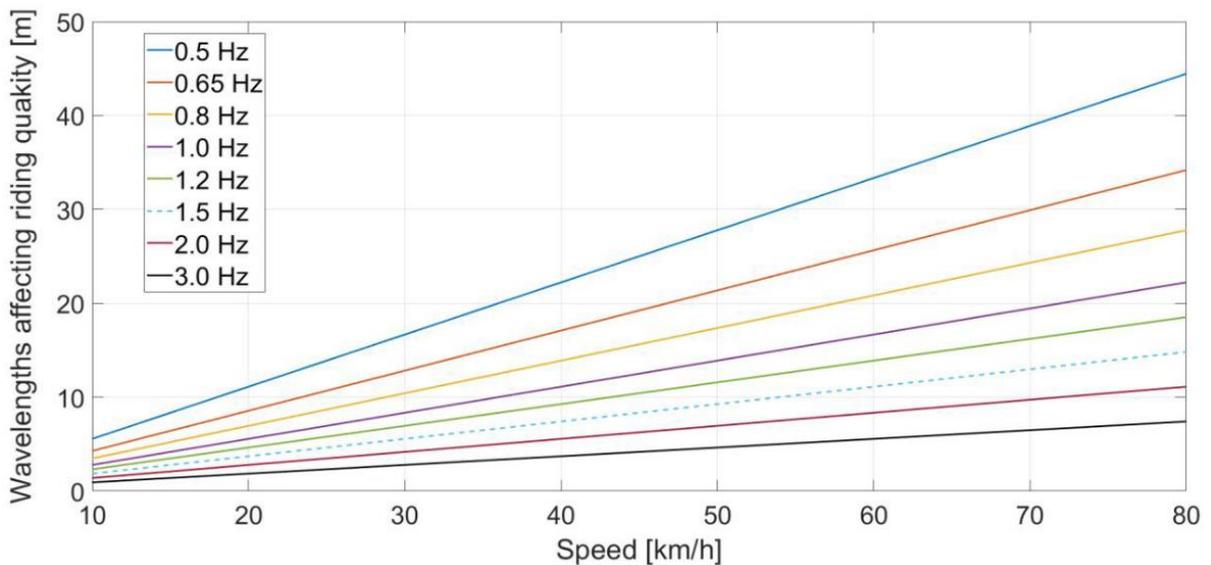


Figure 25: Wavelengths that affect ride quality - (ALCOCER, 2019).

Alcocer (2019) showed an example of applying Equation (2.7) to determine the wavelength range. For a natural frequency range between 1 and 2 Hz and a speed of 80 km/h, the wavelength that excites the vehicle would be between 11.111 m and 22.222 m. This was used to justify the usage of D1 class wavelengths for generating artificial track irregularities for their simulations.

For this thesis, the natural frequency of the studied railway vehicle was measured, and the lowest and largest frequencies were 1.10 Hz and 4.17 Hz respectively. Considering a maximum permissible speed of 65 km/h and applying Equation (2.7), the wavelength range that would excite the wagon would be between 4.33 m and 16.414 m. The D1 class from EN 13848-

5 contemplates this range, therefore, this class will be used for generating artificial track irregularities.

2.6.4 Generating artificial track irregularities in SIMPACK

SIMPACK uses Equation (2.8) for the amplitude of the signal. Given the frequency band F_1 and F_2 , the number of discrete frequencies and the PSD function $S(\Omega)$, the amplitude of the trigonometric Fourier series that describes the spatial domain track excitation can be calculated. This is done by summing the resulting amplitudes and phases of the harmonic functions created, shown in Equation (2.8). Note that the frequency resolution, range, and PSD formula affect the amplitude value. From Equation (2.9), both the frequency and amplitude of the signal change, depending on the wavelength range.

$$r(x) = \sum_{k=1}^N A_k \cos(w_k x + \theta_k) \quad (2.8)$$

$$A_k(\Omega) = \Delta\Omega \sqrt{\frac{S(\Omega)}{2\pi\Delta\Omega}} \quad (2.9)$$

The procedure described led to the creation of four irregularities: two vertical irregularities from Equation (2.4) (left and right rail) and two lateral irregularities from Equation (2.5) (left and right rail). It must be emphasized that the same formula was used for both left and right irregularities; however, a different value was set in the random number generator inside SIMPACK to guarantee that the excitations would be different and, therefore, better represent reality.

Using the parameters listed below as input for SIMPACK and the PSD formula constants of Table 2, artificial irregularities were created for both lateral and vertical directions. The method consists in using a polynomial quotient (1 in SIMPACK) to describe the PSD expressions and then importing them as a stochastic function from PSD (108 in SIMPACK). For SIMPACK automatically convert the units to S.I., the DIMLESS function was used to wrap the coefficients of the polynomial quotient and give the desired units.

- **Class:** FRA6, FRA5, FRA4, FRA3 and no irregularities;
- **Upper and lower spatial frequency (1/m):** 1/3 and 1/25 (D1);
- **Number of frequencies (harmonics):** 3000 (BERAWI, 2013; PODWÓRNA, 2015);
- **Representation of independent:** Angular frequency;
- **Free factor alpha:** 2π ;
- **Sides:** One-sided;
- **Constant k:** 0.25 (BERAWI, 2013; PODWÓRNA, 2015);
- **Track extension:** 7407 m (Figure 26).

Figure 26 shows the macrogeometry used. It is composed of nine curves and ten tangents. The horizontal lines where curvature equals zero represent tangent sections while those that are not zero represent curved sections. The transition from tangent to curve or vice-versa is shown by the inclined lines, where track curvature changes linearly over the distance. Figure 27 shows the generated track irregularities from SIMPACK according to their FRA class.

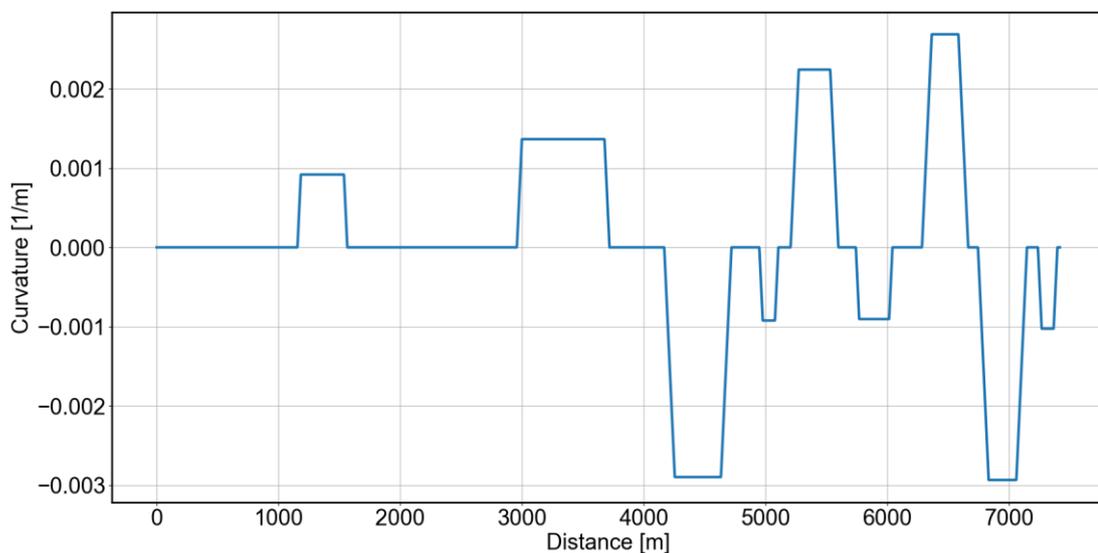


Figure 26: Track curvature.

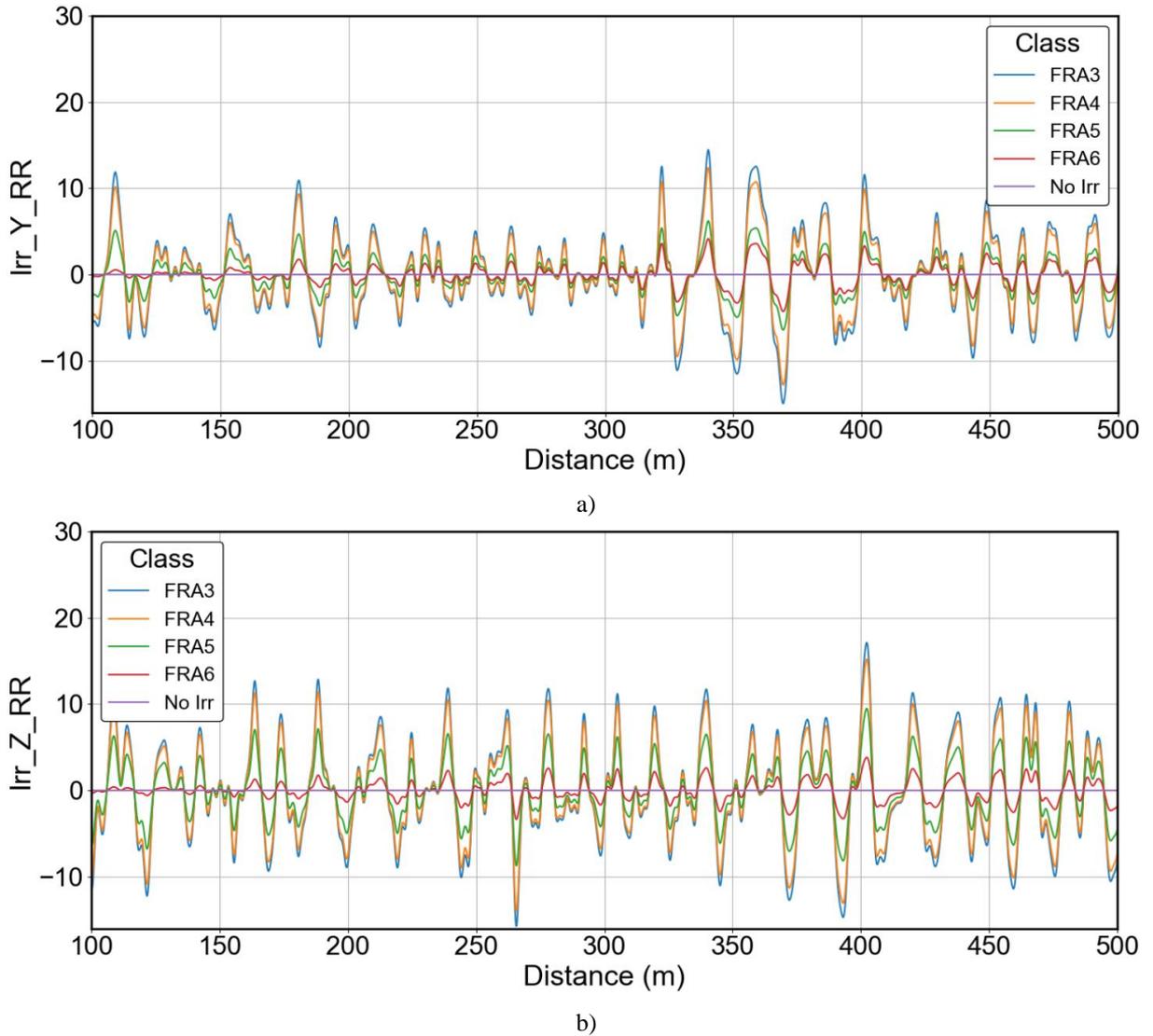


Figure 27: Generated track irregularities from SIMPACK. Letter a) shows the lateral track irregularities for the right rail. Letter b) shows the vertical track irregularities for the right rail.

2.7 Track condition monitoring

There are three main approaches to condition monitoring: physical models, conventional data-driven models, and deep learning models, exemplified in Figure 28. Traditionally, physics-based models were commonly used to monitor asset degradation, however, the Internet of Things (IoT) has shifted this paradigm towards data-driven methods. Most of the physics-based models are unable to be updated with online measured data and the model needs to represent the dynamic system through accurate parameter identification, which limits their effectiveness and flexibility (ESCALONA; URDA; MUÑOZ, 2021; ZHAO et al., 2019). On the other hand,

the development of sensor networks, computers, and the increasing amount of data collected from assets have led data-driven monitoring methods to become more and more attractive.

Traditional data-driven methods likely require the design of new features (feature engineering) to better find the mapping between the dataset and the target variable. The number of features created can be so vast that it is common to use feature selection or dimensionality reduction to reduce the number of variables to only the most relevant ones. This process of designing and choosing new features is heavily dependent on the domain knowledge and expertise of the data scientist, although there are algorithms that try to automate these steps such as `tsfresh` (CHRIST et al., 2018). After creating the dataset, the model is trained using traditional machine learning techniques like Support Vector Machines (SVM) or RandomForests.

To alleviate the problem of needing domain knowledge, deep learning extracts hierarchical representations from input data by building deep neural networks with multiple layers of non-linear transformations. Each layer can be regarded as a function that applies a transformation from input values to output values (REZAEIANJOUYBARI; SHANG, 2020; ZHAO et al., 2019). Therefore, the application of one layer can learn a new representation of the input data, and then, the stacking structure of multiple layers can enable the model to learn complex concepts from the raw input. This is one of its main selling points since it automatically learns internal representations from the raw input and target variable, therefore, not requiring extensive human labor and domain knowledge for handcrafting features and feature selection. For this thesis, traditional machine learning was used.

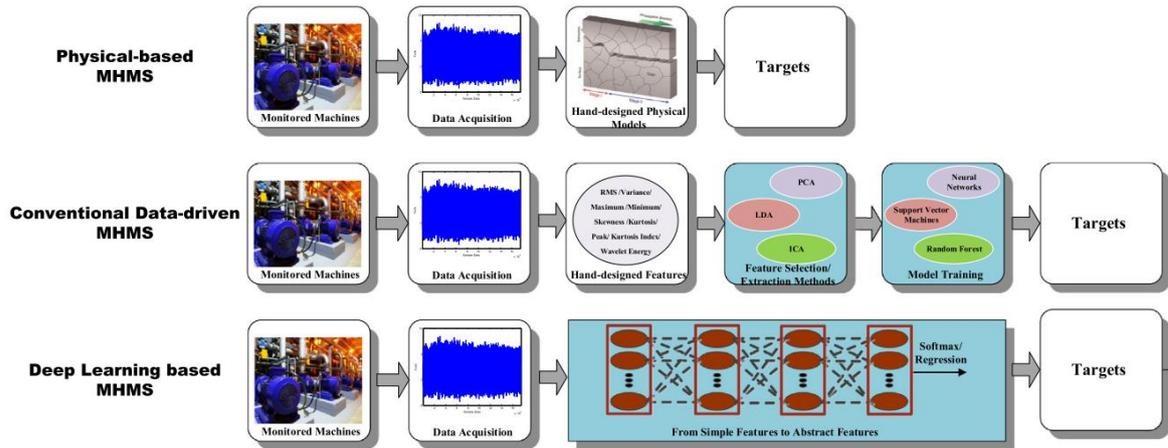


Figure 28: Frameworks for three different monitoring approaches: physical models, conventional data-driven models, and deep learning models. Shaded boxes denote data-driven components - (ZHAO et al., 2019).

2.7.1 Machine learning for intelligent fault diagnosis

Machine learning has helped automate intelligent fault diagnosis (IFD) to recognize the health state of machines without specialized knowledge of the engineering scenario (LEI et al., 2020). The diagnosis procedure can be summarized in 3 steps: data collection, artificial feature extraction, and health state recognition, as shown in Figure 29. Each step of Figure 29 will be described in more detail in the following sections highlighting how to take these concepts and apply them to the railway system.

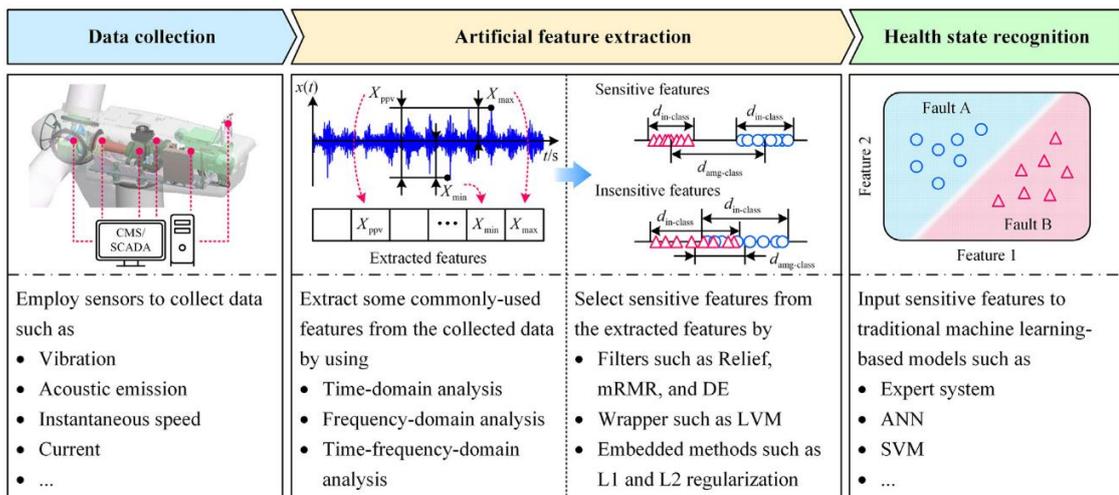


Figure 29: Diagnosis procedure for intelligent fault diagnosis (IFD) using traditional machine learning - (LEI et al., 2020).

2.7.2 Data collection

Sensor measurements are the main source of data for machine diagnosis. Different sensors can be used to measure vibration (accelerometers), temperature, and pressure, among other variables. For instrumented railway vehicles (IRVs), the dynamical response from track excitations is measured using accelerometers positioned on the carbody, bogie, and axlebox. Other sensors such as load cells and strain gauges are used to measure suspension displacement (secondary or primary) and wagon drawbar impact loads respectively (BOWEY, 2018b; COWIE et al., 2015; PIRES et al., 2021). More information on the measured variables is available in Section 2.2.2.

2.7.3 Feature extraction and engineering

Feature extraction is a common step in all diagnostic and prognostic approaches. Feature extraction is defined as the process of obtaining time, frequency, and time-frequency domain features from raw signal data (ANKRAH; KIMOTHU; MUVENGEI, 2020; LEI et al., 2007, 2020; PIRES et al., 2021; TSUI et al., 2015). When faults occur in machines, the time-domain signal may change in amplitude and distribution compared to the signal under normal conditions. The frequency spectrum and its distribution may also change, meaning that new frequency components may appear associated with degradation (LEI et al., 2007).

Examples of time, frequency, and time-frequency domain features are shown in Table 3. Since raw sensor data is a time series signal, a rolling window can be applied and generate new columns of data from these metrics. Tsunashima et al. (2019) used a rolling window of 4 samples to calculate the root mean square of the carbody vertical acceleration, although the exact size of the window is a parameter to be optimized.

Time domain features	Shape factor $SF = \frac{RMS}{\frac{1}{n} \sum_{i=1}^n x_i }$	Power spectral density of FFT $PSD = \sum_{k=-\infty}^{\infty} r_k e^{-i\omega k}$
RMS $RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$	Line integral $LI = \sum_{i=0}^n x_{i+1} - x_i $	Time –frequency domain
Variance $var = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$	Peak to peak value $PP = \max(x_i) - \min(x_i)$	Energy of WPD detail coefficient one $D_1 = \sum_{i=1}^{n1} cD1_i$
Peak value $PvT = \max(x_i)$	Shannon entropy $Ent = - \sum_{i=1}^n x_i^2 \log(x_i^2)$	Energy of WPD detail coefficient two $D_2 = \sum_{i=1}^{n2} cD2_i$
Crest factor $CF = \frac{PvT}{RMS}$	Skewness $Sk = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}}$	Energy of WPD detail coefficient three $D_3 = \sum_{i=1}^{n3} cD3_i$
Kurtosis $Kurt = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n \times var^2} - 3$	Frequency domain	Energy of WPD approximate coefficient three $A_3 = \sum_{i=1}^{n3} cA3_i$
Clearance factor $Clf = \frac{PvT}{\left(\frac{1}{n} \sum_{i=1}^n \sqrt{ x_i } \right)^2}$	Peak value of FFT $PvF = \max(r_k),$ $r_k = \sum_{i=-\infty}^{\infty} x(t) e^{-i\omega t}$	
Impulse factor $IF = \frac{PvT}{\frac{1}{n} \sum_{i=1}^n x_i }$	Energy of FFT $En = \sum_{f=1}^N r_k$	

Table 3: Commonly used time domain features - (KIMOTHO; SEXTRO, 2014)

Extracted features must be sensitive to the fault condition, as shown by the 3rd block of Figure 29. By accurately separating normal and faulty conditions, the machine learning algorithms will have an easier time training the model, leading to better performance.

In the case of the problem of obtaining track geometric irregularities from measured IRV data, it is expected that large geometric peaks lead to large dynamic excitations. Therefore, feature engineering and extraction are responsible for creating variables that are sensitive to these conditions.

2.7.4 Health state recognition

Health state recognition can be summarized as a classification problem, where machine learning models are used to establish the relationship between the selected features and the health state of the asset (LEI et al., 2020). Given a dataset that has already undergone feature

extraction, the diagnosis model is trained with labeled samples of the health state. Once trained, the models can recognize the health states when fed with unlabeled data.

Unfortunately, IRV data does not have a label that characterizes track health. The limits utilized by maintenance teams are not optimal and depend on the engineer's expertise during measurement and analysis. That said, the BRA railway does have norms that help differentiate a good track from a bad one in terms of its geometry. Therefore, knowing that the limits of the track geometry must satisfy international norms, it might be possible to find what sensor values correspond to these limits. However, this is not the scope of the thesis.

2.8 Machine learning algorithms

This section will go over the machine learning algorithms used to map the input dataset to the target variable (geometric track irregularities). The algorithms used were: Linear regression with regularization (Lasso, Ridge, ElasticNet, SGDRegressor), RandomForestRegressor, ExtraTreesRegressor, AdaBoostRegressor, XGBRegressor (Extreme Gradient Boosting Regressor), LightGBM (Light Gradient Boosting Machine) and neural networks.

2.8.1 Linear regression with regularization

Linear regression is one of the simplest models to use and is given by Equation (2.10), where \hat{y} is the predicted value, n is the number of features, x_i is the i^{th} feature value and θ_j is the j^{th} model parameter. There is also a vector representation given by the transpose of the parameter vector θ and the feature vector x . This parameter vector contains the weights that each variable has on the model output.

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n = \theta^T x \quad (2.10)$$

An error metric is minimized to find the best fitting line to the data. The mean squared error (MSE) is commonly used as the cost function, which is shown in Equation (2.11), where y_i is a vector of containing the values of the variable and m is the number of data points. Note that the *MSE* is only parameterized by θ .

$$MSE(\theta) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \theta^T x_i)^2 \quad (2.11)$$

The generalization performance of the linear regression model may be affected by the size or complexity of the parameter vector θ , leading to overfitting (AGGARWAL, 2015; GÉRON, 2019). Overfitting occurs when a model fits the training data too closely, thereby losing its ability to generalize to unseen data. This concept can be illustrated through polynomial regression, as shown in Figure 30. When a polynomial of degree 1 is used, a linear model is obtained (red line), however, this model may not accurately represent the dataset due to its simplicity, known as underfitting. On the other hand, using a polynomial of degree 300 results in excessive oscillations to fit the data points. There are too many parameters when compared to the number of observations, which leads to the model representing the noise in the data rather than the underlying relationship, causing overfitting. The best fit was obtained using a 2nd-degree polynomial, highlighting the need to modify the linear regression cost function to mitigate overfitting.

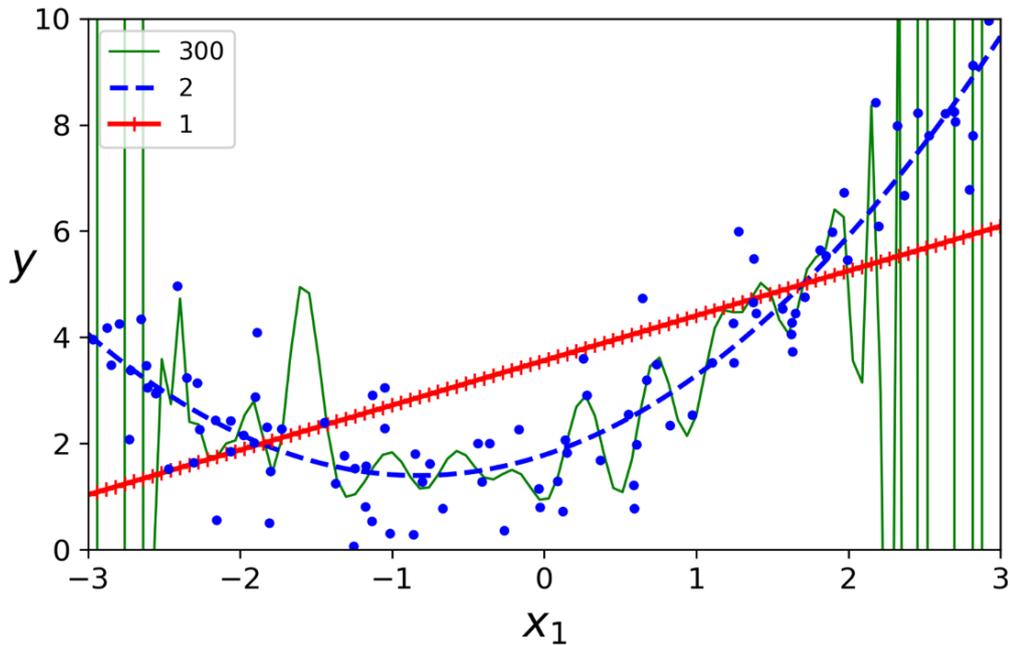


Figure 30: Example of overfitting using polynomial regression - (GÉRON, 2019).

Regularization is a technique that addresses overfitting by adding a penalty term to the cost function. This penalty term discourages very large coefficients and reduced the weight of less important variables, thereby forcing the model to have a simpler and more generalizable representation of the data (AGGARWAL, 2015; GÉRON, 2019). In essence, regularization trades a small increase in error on the training data for improved generalization performance on unseen data. There are three types of regularization: Lasso, Ridge, and Elastic Net, each with a different representation of the regularization term.

Ridge regularization adds a regularization term equal to $\frac{\alpha}{2} \sum_{i=1}^n \theta_i^2$ to the cost function, where α is a user-defined hyperparameter representing how much influence the regularization term will have on the cost function. If α is very small, regularization does not affect the algorithm. However, if α is very large, then all weights end up very close to zero and the result will be a flat line. The Ridge regression cost function $J(\theta)$ is the MSE of the model's weights added with Ridge's regularization term, as shown in Equation (2.12).

$$J(\theta) = MSE(\theta) + \frac{\alpha}{2} \sum_{i=1}^n \theta_i^2 \quad (2.12)$$

The regularization term above has an equivalent representation using a squared L2-norm, also known as Euclidian norm or L2 regularization. Using this representation alongside the cost function $J(\theta)$ gives:

$$J(\theta) = MSE(\theta) + \frac{\alpha}{2} \|\theta\|_2^2 \quad (2.13)$$

Lasso regularization has a different regularization term that is the L1 norm of the weight vector instead of half the square of the L2 norm. Lasso regularization tends to eliminate the weights of the least important features by setting them to zero, offering a sparse model (GÉRON, 2019). The cost function alongside Lasso regularization is shown in Equation (2.14).

$$J(\theta) = MSE(\theta) + \alpha \sum_{i=1}^n |\theta_i| \quad (2.14)$$

Elastic Net regularization is a mix of both Ridge and Lasso regularization, where a mixing ratio r is defined to combine both regularization functions, as shown in Equation (2.15). If $r = 0$, Elastic Net becomes Ridge regularization and when $r = 1$, it becomes Lasso regularization.

$$J(\theta) = MSE(\theta) + \frac{1-r}{2} \alpha \sum_{i=1}^n \theta_i^2 + r\alpha \sum_{i=1}^n |\theta_i| \quad (2.15)$$

2.8.2 Decision trees

Decision trees utilize hierarchical decisions on the feature variables to obtain the output, resulting in a tree-like structure, as shown in Figure 31. The first decision is done using the “Age” variable, leading to two branches that used the “Salary” variable to continue making decisions and splitting the data until they are well separated. Although Figure 31 uses a single attribute for the splits, it is possible to consider multiple variables simultaneously, as shown in Figure 32. Algorithms such as RandomForests utilize multiple decision trees trained on different subsets of data to produce an output and aggregate these outputs into a final value.

This leads to more robust and generalizable solutions compared to using a single decision tree (AGGARWAL, 2015; GÉRON, 2019; YI et al., 2019). This technique is a type of model ensemble, which will be discussed in the following sections.

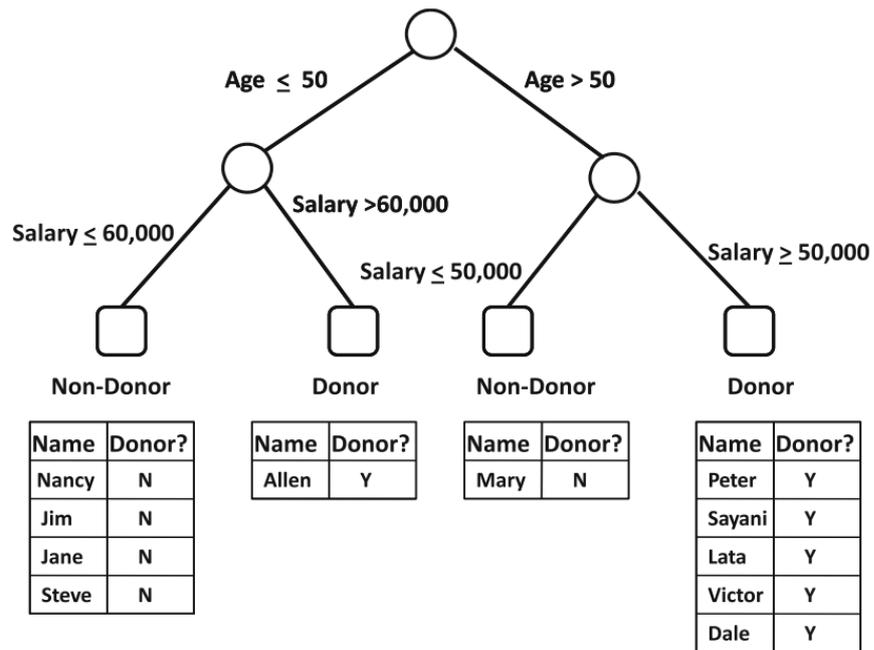


Figure 31: Example of a decision tree for univariate splits - (AGGARWAL, 2015).

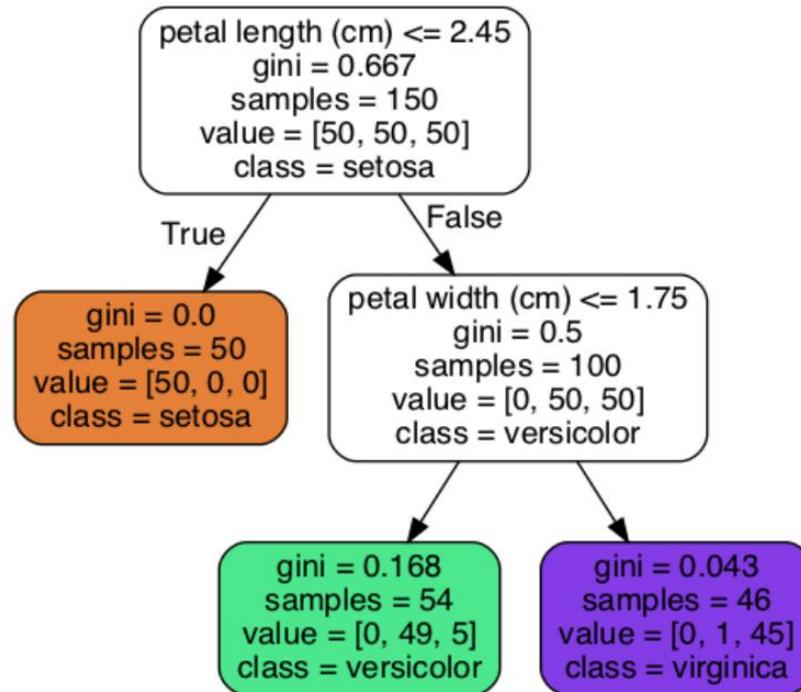


Figure 32: Example of a decision tree for multivariate splits - (GÉRON, 2019).

2.8.3 Model ensembling

Model ensembling (grouping) is the combination of different models to overcome the limited generalization performance of each model, thus, generating more accurate predictions than these individual models. After optimizing the hyperparameters of different machine learning models and guaranteeing an optimal configuration for the problem, ensembling them will likely perform better than the individual models (FERREÑO et al., 2021; LASISI; ATTOH-OKINE, 2019). Even if the models are weak learners (only slightly better than random guessing), the ensemble can still achieve high accuracy, provided that there are a sufficient number of weak learners and they are sufficiently diverse.

It is important to keep in mind that there is no universal algorithm that is guaranteed to perform well for all types of problems and datasets, which is also known as the “no free lunch theorem” (DE ROSA et al., 2020; LASISI; ATTOH-OKINE, 2019; SHAFIULLAH et al., 2010). Different algorithms will perform better on different parts or characteristics of the dataset. The diversity of solutions that comes with combining these algorithms can increase

robustness and generalization. Therefore, these methods work best when the predictors are as independent of one another as possible, like training very different algorithms.

Figure 33 and Figure 34 exemplify ensembling for a classification problem. Suppose from Figure 33 that different classification algorithms were trained and each one achieved about 80% accuracy. The predictions of each model can be grouped up to predict the class that gets the most votes. In Figure 34, three out of the 4 models predicted class 1, therefore, the final output would be class 1. Since three out of the 4 models returned 1, it is more likely to be correct than by using a singular model.

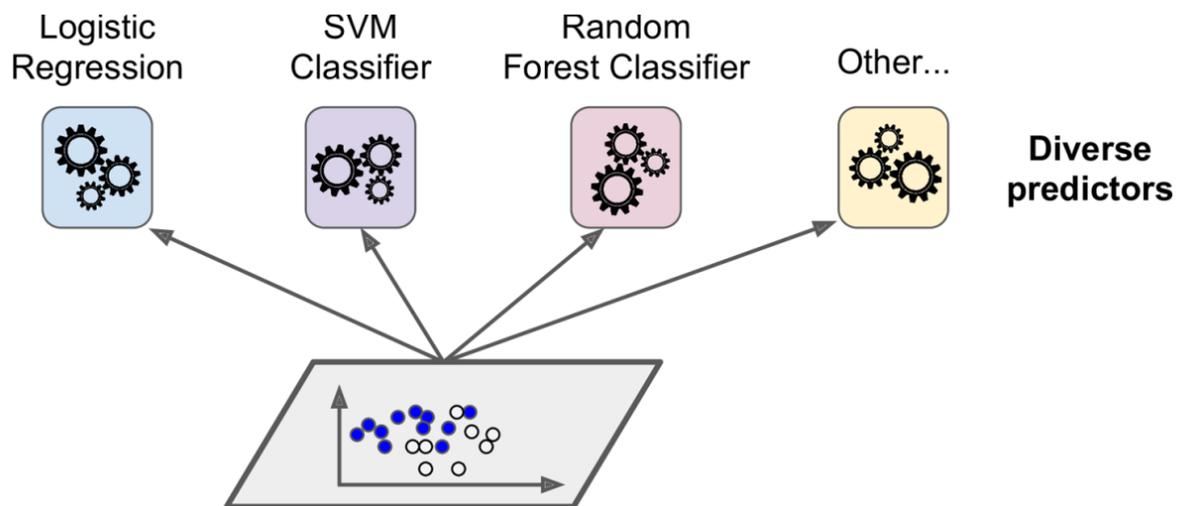


Figure 33: Training diverse classifiers - (GÉRON, 2019).

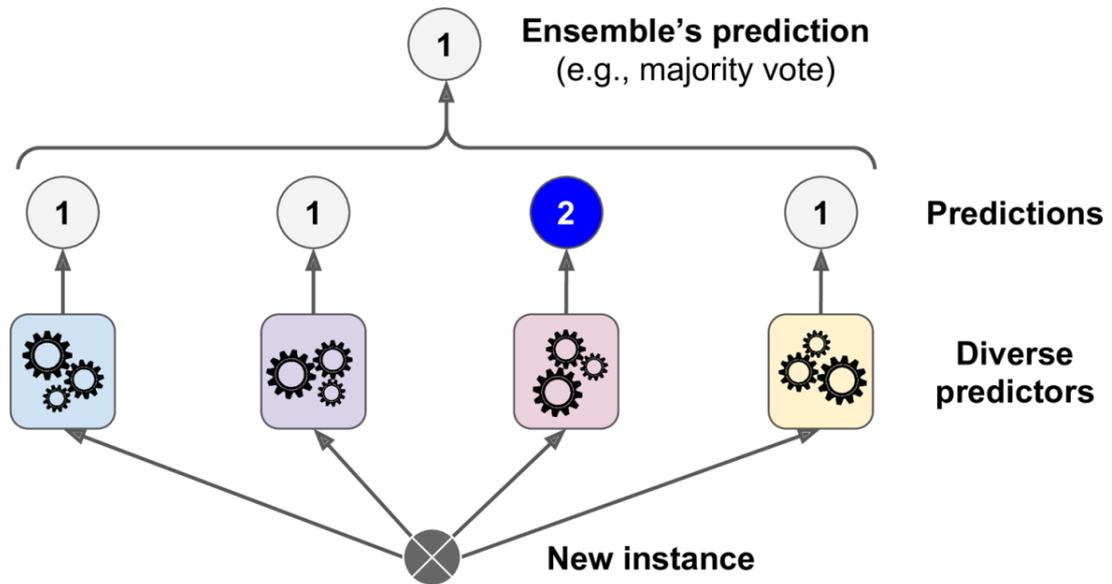


Figure 34: Combining models with a hard voting classifier - (GÉRON, 2019).

There are many ways to combine different machine learning models. The main approaches will be presented in the following sections.

2.8.3.1 Bootstrap aggregation (Bagging)

Instead of training a diverse set of models, Bagging uses the same algorithm but trains it on different random subsets of the training set, as shown in Figure 35 (GÉRON, 2019). Once all models are trained, the ensemble aggregates these models, and their combination produces the final output.

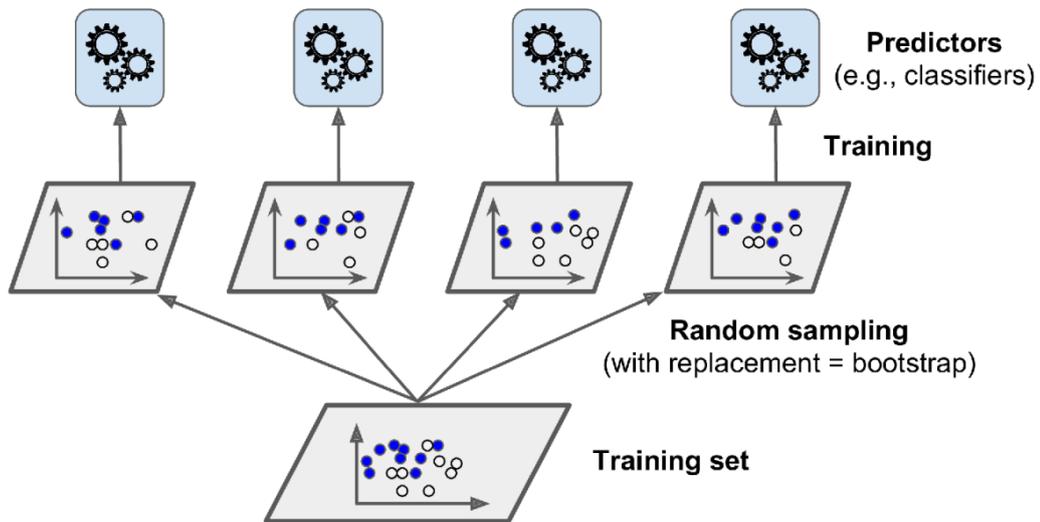


Figure 35: Illustration of bagging - (GÉRON, 2019).

Figure 36 compares the decision boundary of a single decision tree with the decision boundary of a bagging ensemble of 500 trees, both trained on the same dataset. The ensemble's prediction will likely generalize better than the single decision tree, as seen by the less irregular green decision boundary separating the blue class from the yellow class.

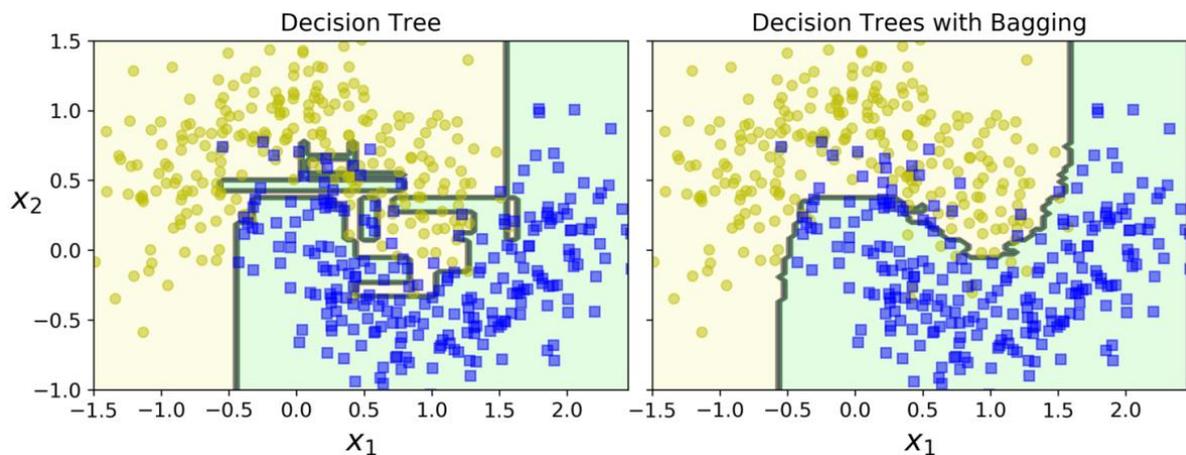


Figure 36: A single decision tree (left figure) versus a bagging ensemble of 500 trees (right figure) - (GÉRON, 2019).

Bagging is also used as a nonparametric way of determining a model's confidence interval (bootstrap confidence intervals) (BROWNLEE, 2019). With just the original dataset, there will be only one single estimate of the population parameter with no information about the

uncertainty of the estimation. By taking random samples of the dataset using Bagging, there is now available data to estimate a population parameter such as the mean of the estimate or a confidence interval for quantifying uncertainty. A downside of resampling methods is that they can be computationally expensive due to requiring hundreds to thousands of resamples to develop a robust estimate of the populating parameter.

Forest algorithms such as RandomForest and ExtraTrees have this sampling strategy known as bootstrap sampling and the maximum number of samples that are used by each tree can be defined as a hyperparameter.

2.8.3.2 Boosting

Boosting methods involve training models sequentially, that is, the next model tries to correct the residual of its predecessor (AGGARWAL, 2015; AGGARWAL; SATHE, 2015; FERREÑO et al., 2021; LASISI; ATTOH-OKINE, 2019). Famous algorithms such as XGBoost, AdaBoost, CatBoost, LightGBM, etc all employ this concept. This is better explained in the example shown in Figure 37 for a classification problem.

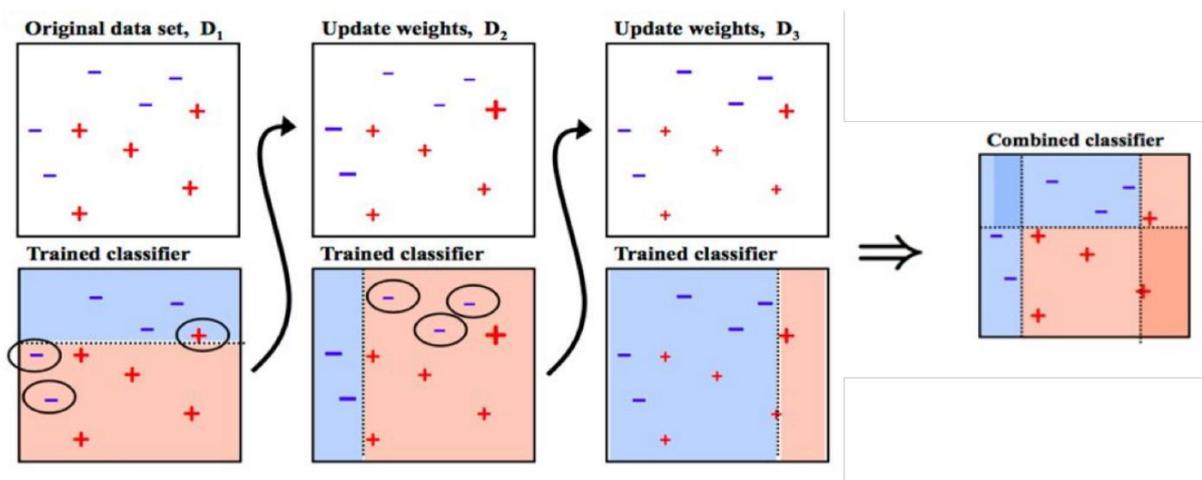


Figure 37: Illustration of how boosting works, exemplified by an AdaBoost classifier - (MARSH, 2016).

From Figure 37, a model is fitted to the original dataset D_1 , which lead to the creation of a decision boundary after training. Initially, all samples have the same weight on the error. The

highlighted black circles show the error in the model. Incorrect predictions are assigned to the next base learner with larger importance (larger weights or boosted weights). By giving more importance to the incorrect predictions, the next model will prioritize correcting these terms. This can be seen by the slightly larger symbol size of the incorrect points after updating the weights. The second model D_2 is trained and the steps are repeated. For this figure, the ensemble of three models was enough to correctly classify the dataset. More accurate models will have a bigger weight for the final output.

Note that the sequential nature of the boosting algorithm makes them more prone to overfitting compared to traditional machine learning models. Since model errors made by previous iterations are fitted by the subsequent model, over-specialization to the training data can occur, which would reduce the model's ability to generalize. So it is important to use techniques that prevent overfitting such as early stopping and regularization. Another option is to reduce the number of models subsequently trained and verify their effect using cross validation.

2.8.3.3 Stacking

Instead of using a trivial function such as a voter or average, a machine learning model is responsible for the aggregation (GÉRON, 2019; OZDEMIR; SUSARLA, 2018). Stacking utilizes the output of simpler models (base learners) to create a new dataset that will serve as the output for a second-level model, also known as a meta-model or blender (PIRES et al., 2021). This second model will use the individual model predictions as the input and give the final estimate as the output. A simple flowchart is shown in Figure 38 while an example of a regression problem is shown in Figure 39.

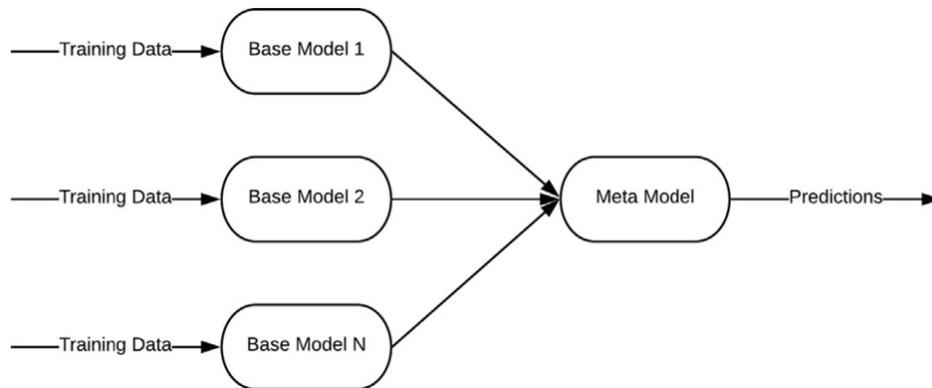


Figure 38: Simple flowchart of how model stacking works - (PIRES et al., 2021).

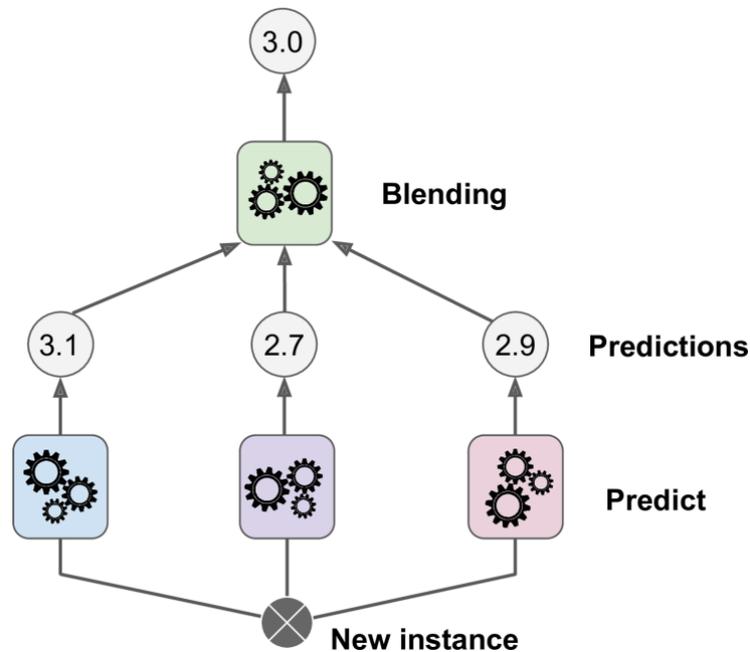


Figure 39: Aggregating predictions using stacking - (GÉRON, 2019).

Gerum et al. (2019) stacked a random forest model with a recurrent neural network (RNN) to predict track defects and minimize underestimations. The overall average false negative rate decreased from 15% to less than 5% while the false positive rate decreased from 20% to 8%. The Partial Swarm Optimization (PSO) algorithm was implemented to find the best weights for the linear combination of the models. In the case of this thesis, model stacking was not necessary due to the good error metrics, but was included for completeness.

2.8.4 Neural networks

Artificial neural networks are a type of deep learning model that is capable of generating nonlinear mappings between the input and the output through the connection of layers of nodes or neurons (GULER, 2014; KARIS, 2018; ZHAO et al., 2018). These neurons receive an input from some other neurons and a weight value is attributed to the input connections. The neuron then applies a mathematical function to the input and returns an output to be used by subsequent neurons (AGGARWAL, 2015). For the network to work properly, the weights of the neural network are optimized by minimizing a cost function. The effectiveness of the neural network is heavily reliant on its arrangement of neurons or architecture.

An example of a neural network is shown in Figure 40. This neural network was used to obtain wheel contact forces from track geometry and vehicle data to quantify safety through the Nadal index (LI et al., 2006).

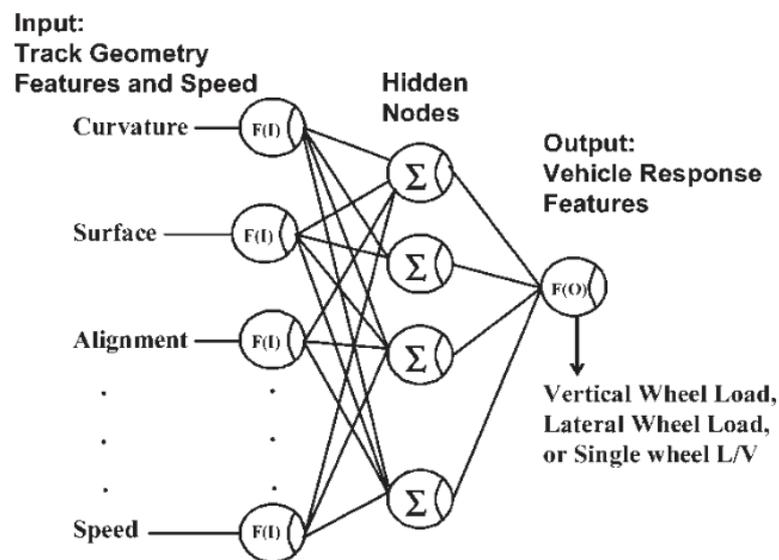


Figure 40: Simplified neural network architecture for obtaining wheel loads from track geometry data - (LI et al., 2006).

In this thesis, a deep neural network was used, characterized by its multiple hidden layers, each equipped with a substantial amount of neurons. While deeper neural networks, meaning those with a larger amount of hidden layers, have the potential to capture more intricate relationships, they are also more susceptible to overfitting due to the larger number of

hyperparameters. To mitigate the risk of overfitting, regularization techniques such as early stopping, dropout, and the application of L1 or L2 norms to penalize hidden layers, are often utilized. Deeper neural networks also demand a higher computational cost during training and necessitate a larger dataset to avoid overfitting. It is important to note that deepening a neural network does not guarantee improved performance, and thus, great caution must be exercised when designing the network architecture and training the model.

3 METHODOLOGY

Figure 42 shows the flowchart of the methodology adopted in this thesis. Five main parameters are necessary to evaluate the dynamic behavior, which are the track macrogeometry, wheel and rail profile, velocity profile, wagon model, and geometric track irregularities. The wagon's dynamical model must be capable of measuring the variables that the BRA1 railway uses in its day-to-day activities and also the variables used by the IRVs of other researchers to visualize which potential sensors would be important for this application. The dataset used in this thesis can be divided into three categories:

- **IRV sensor data:** variables that the BRA1 IRV can measure with its sensors;
- **Virtual sensor data:** variables that are measured by the IRVs of other researchers but not the BRA1 due to lack of instrumentation;
- **Target variable:** vertical and lateral rail irregularities of the left and right rail.

The raw dataset was created inside *SIMPACTTM*, which is a commercial program used for multibody dynamic simulation and is commonly used by railway researchers. The BRA IRV model was validated with measured data and will act as the digital mockup of the real instrumented wagon. The model was created with three substructures: two bogies and the wagon, as shown in Figure 41. More details on the model are available in Silva (2022).

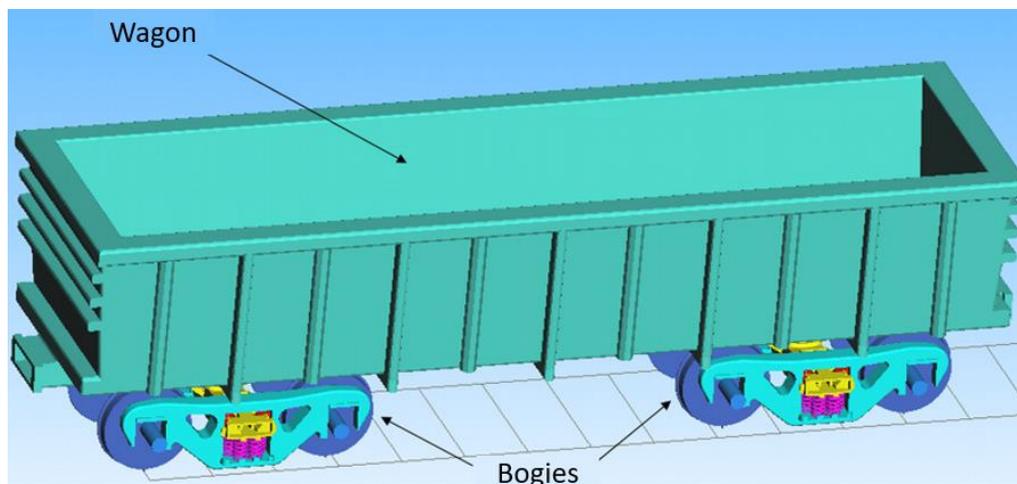


Figure 41: SIMPACK model – modified from (SILVA, 2022).

Preprocessing is done to rename the columns and create categorical columns to facilitate the exploratory data analysis (EDA) step. Once the data is preprocessed, the EDA step consists of visualizing the dataset with lineplots, boxplots, and correlation metrics to highlight the relationship between input and target output (track irregularities). After establishing a baseline model in the EDA phase, feature engineering is done to better find the mapping between the dataset and the target variable and improve upon the baseline model. Once the feature engineering step is concluded, feature selection was done to remove redundant variables. Different machine learning models were then trained and optimized using Optuna (AKIBA et al., 2019; OPTUNA, 2020) and compared to one another. The best model in terms of error metrics was chosen for further investigation involving model interpretability and feature importance. In the last topic, the variables that most contribute to the model output were defined.

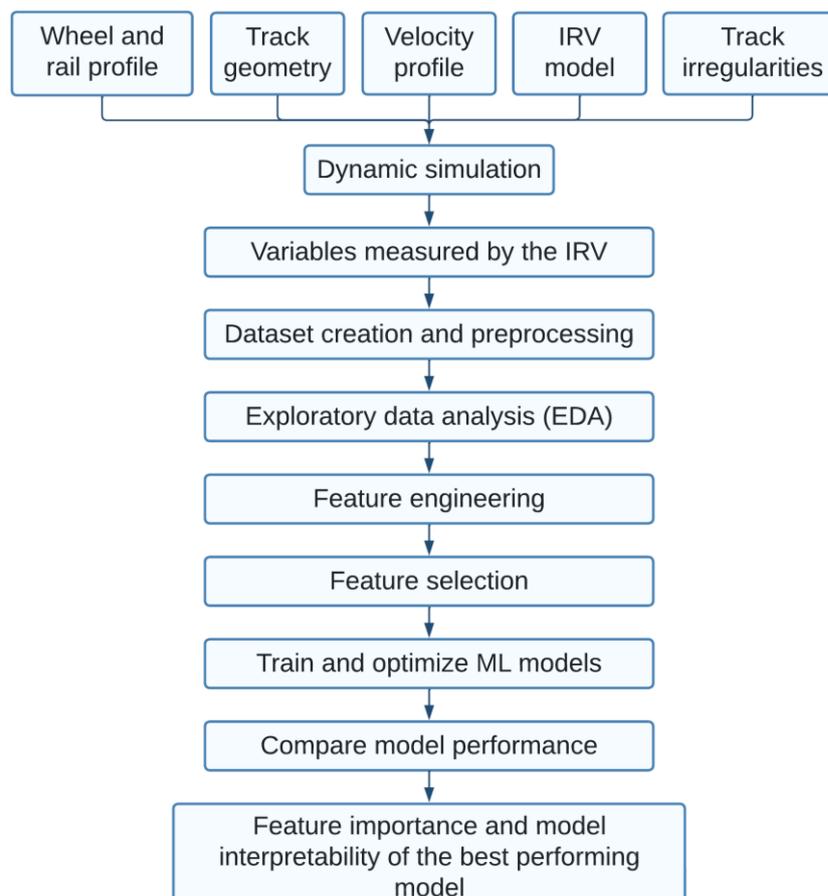


Figure 42: Methodology flowchart.

3.1 Dataset creation and preprocessing

The current instrumentation of the IRV may be insufficient for measuring track irregularities with reliable results. Therefore, the dataset created from simulation shall have the same sensors as the real IRV, but also sensors that do not exist in the real IRV. By simulating with both real and virtual sensors, it is possible to do the following:

- Test different sensor positions to determine an optimal location from the already existing instrumentation;
- Test different sensor types that are not used in the current IRV (gyroscope for example) and analyze their viability.

The raw dataset consists of 5 .csv files containing simulation data from 4 different FRA class irregularities and a no irregularity case. Each of these files contains 68352 rows and 56 columns. There are a large number of columns because there are 11 real sensors, 38 virtual sensors, 4 target columns (track irregularities), and 3 auxiliary columns (distance, curvature, and irregularity class). From the curvature of the track, 4 categorical variables were created to add information about the track segment such as, for example, the type of track (tangent or curve), and track radius. The virtual sensor positions were based on the study of other authors such as (BLEAKLEY, 2006; COWIE et al., 2015; HARDIE et al., 2016; PIRES et al., 2021). Therefore, the final raw dataset has 341760 rows and 60 columns.

The scope of part of this work is to verify if it is possible to find the nonlinear mapping between the track irregularities and the real IRV variables using a machine learning model, so the numerical variables of the dataset were reduced to the variables that this IRV can measure. This was done because, if the mapping proves to be sufficiently accurate with the current instrumentation, then there will be no need to add other sensors and increase costs. Table 4 shows a brief description of this filtered dataset that now has 341760 rows and 22 columns.

Data	Localization		Variable	Component	Units	Description	Name				
Input: numerical variables	Carbody	Middle	Acceleration	X	m/s ²	Triaxial accelerometer variables	TA_X				
				Y			TA_Y				
				Z			TA_Z				
	Leading bogie sideframe	Left side	Acceleration	Z	m/s ²	Vertical uniaxial accelerometer	UA_Z_LR				
		Right side					UA_Z_LL				
	Trailing bogie sideframe	Left side					UA_Z_TL				
		Right side					UA_Z_TR				
	Leading bogie suspension	Left side	Displacement	Z	m	Displacement of spring attached to a load cell	SS_Z_LR				
		Right side					SS_Z_LL				
	Trailing bogie sideframe	Left side					SS_Z_TL				
		Right side					SS_Z_TR				
	-	-					Distance	-	m	Distance traveled by the IRV	Distance
		-					Curvature	-	1/m	Track curvature	Curvature
	Input: categorical variables	-	-	Class	-	-	FRA irregularity class	Class			
-			Type of segment	-	-	Binary variable to distinguish curved from tangent track	Type of segment				
-			Segment	-	-	Specific track segment	Segment				
-			Radius	-	-	Radius of the segment	Radius				
-			Radius class	-	-	Classify track if a segment is tangent, has a radius greater or less than 500 m	Radius class				
Target variables	Track	Right rail	Vertical irregularity	Z	mm	Geometric track irregularities	Irr_Z_RR				
		Left rail					Irr_Z_LR				
		Right rail	Lateral irregularity	Y			Irr_Y_RR				
		Left rail					Irr_Y_LR				

Table 4: Preprocessed dataset.

3.2 Exploratory data analysis

Exploratory data analysis (EDA) is a common procedure for visualizing and exploring the dataset before model training to better understand the problem and identify possible challenges. The EDA step in this study involves:

- Removing outliers using boxplots based on the expected sensor values obtained from real IRV data;
- Visualizing the relationship between variables using lineplots, boxplots, and correlation metrics to remove useless features;
- Establishing a baseline model.

Once the data is preprocessed, the EDA step consists of visualizing the dataset with lineplots, boxplots, and correlation metrics to highlight the relationship between input and target output (track irregularities). After establishing a baseline model in the EDA phase, it is also possible to create expectations about modeling success by looking at univariate and multivariate correlations between the input dataset and target variable. The exploratory data analysis (EDA) step will focus on using only the variables that exist in the BRA IRV. This means that the studied dataset is filtered to now have 341760 rows and 22 columns.

3.2.1 Outlier removal

To better represent the regression problem, it is important to remove extreme values that are outside the range of what is expected from the data since this is likely due to an error unrelated to the problem. These types of outliers could appear due to measurement error or data corruption (BROWNLEE, 2019). To identify if a data point is an outlier, a mixture of domain knowledge and simple statistical methods were used. In the BRA1 railway, the uniaxial accelerometer values range up to around 70 m/s^2 . Looking at the boxplot of the acceleration variables in Figure 43, the collected data far exceeds this threshold. This is likely due to the closer proximity to the wheel-rail interface and the difficulty in simulating the impact loads that

occur there. Due to the large number of plots necessary to visualize the outliers for each variable, this thesis will show an example of how outlier removal was performed. Of all the variables, only the uniaxial accelerometer variables showed values outside the expected range.

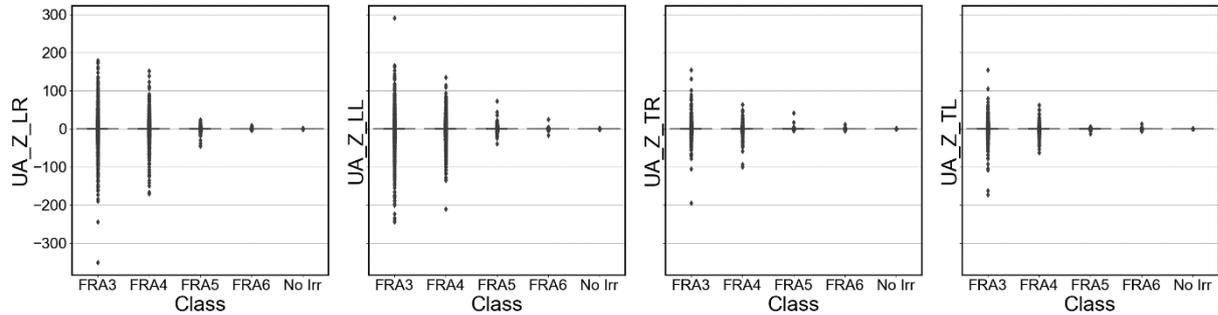


Figure 43: Before outlier removal.

The Interquartile Range (IQR) method was used to remove these outliers. The IQR starts obtaining the 25th and 75th percentile of a variable to calculate their IQR. The new limits to identify outliers are a factor k multiplied by the IQR above the 75th and below the 25th percentile. Brownlee (2019) recommends a factor $k = 1.5$ for common applications or a factor $k = 3$ or more to identify extreme outliers. In this work, the factor k was selected in such a way that simulation values would be in the expected range of measured acceleration values, which would be around 55 m/s² for a FRA 4 class irregularity. The other sensor variables such as the triaxial accelerometer and the spring sensors did not exhibit absurd values. So the rows that had an absolute uniaxial acceleration measurement above 55 m/s² were removed.

The boxplots of the most critical variables after outlier removal are shown in Figure 44. Note in Figure 43 that the range of outlier values is so large that the boxplot cannot be visualized. After outlier removal, the boxplots of FRA3 and FRA4 are barely visible, highlighting that the dataset still contains outliers. These outliers, however, are not outside the expected range of measured values and are likely due to the physical characteristics of the system instead of numerical simulation error. No filtering techniques were applied.

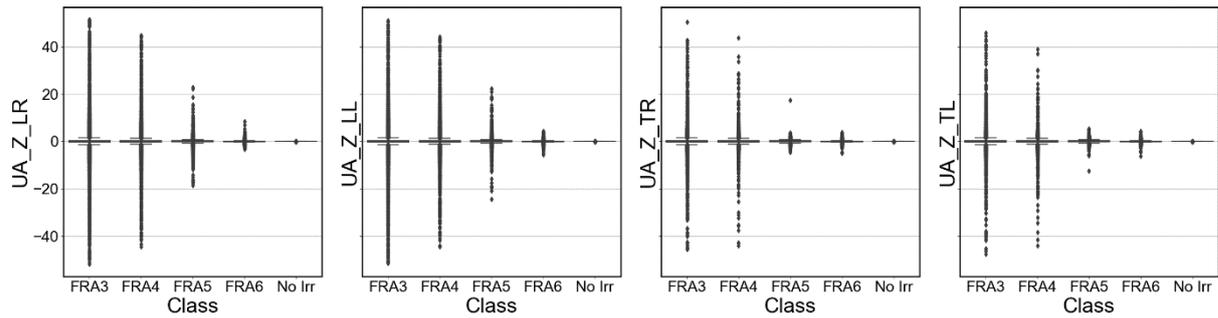


Figure 44: After outlier removal.

3.2.2 The effect of track severity on measured signals

Track irregularity severity has a large impact on the measured signal, as shown by the Y component of the triaxial accelerometer in Figure 45. A track with no irregularities allows the user to see the effect that track curvature has on the signal, as seen by the zoomed section in Figure 46 between 4100 m and 4800 m. However, class FRA6 irregularities make it harder to visualize this while FRA3 irregularities are even more so. This suggests that information with regards to the track segment is masked by track irregularities. The categorical variables that describe track design characteristics (view Table 4 for more information) will not be useful because the data will be too similar. To exemplify in terms of feature selection, the information of whether a track segment is a tangent or a curve will add nothing new for the model to learn this distinction due to their similarity. Consequently, a useless feature will be added that will hinder model training.

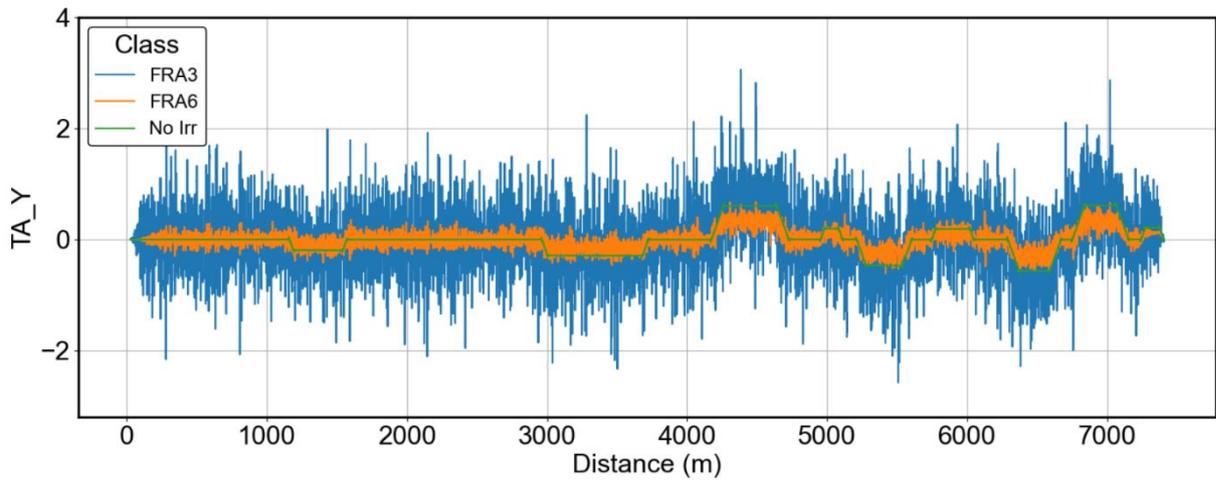


Figure 45: Y component of the triaxial accelerometer.

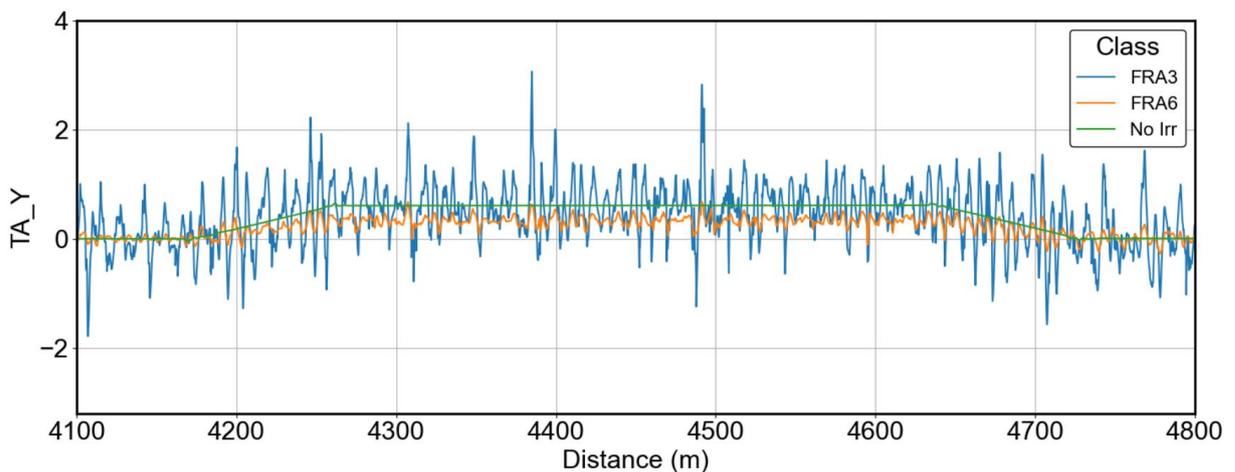


Figure 46: Zoom of the curved section between 4100 m and 4800 m.

The hypothesis that track characteristics are unimportant for large track irregularities can be better explained by using Figure 47. Looking at the boxplot range of the second row for the suspension variables, no irregularities clearly show a distinction between tangent and curved sections. When the track severity increases from FRA6 to FRA3, the boxplot ranges between tangent and curved sections become closer. Up to FRA6, the categorical variables previously defined are likely to be important since there is a noticeable difference in the boxplots. However, for FRA5 and lower, the boxplots are very similar. This means that the track design features created are unlikely to improve model performance for large track irregularities. Since this study prioritizes detecting severe track irregularities over smaller ones, these categorical variables were removed. With this, the model will have to generalize for any type of track

segment. After removing the 4 categorical variables and 2 auxiliary plotting variables (distance and curvature), the dataset was reduced from 22 columns to 16.

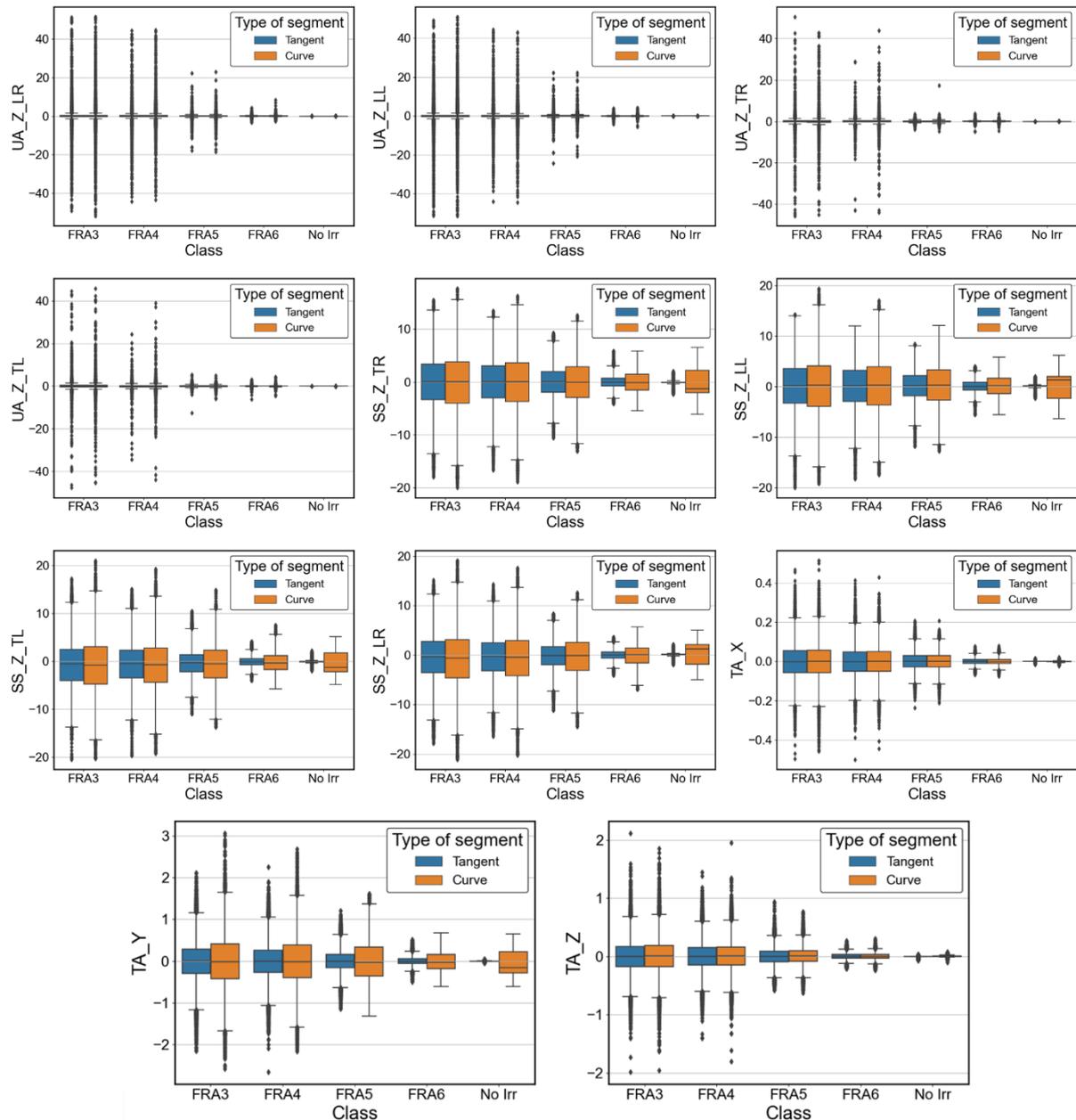


Figure 47: Boxplot of the sensor variables.

Figure 48 shows the pairplot of the irregularities for all FRA classes. Note that, by using all of the data, there will be a data imbalance. The majority of the dataset would be of small irregularities while the main objective is to prioritize the detection of severe irregularity cases. Since the BRA railway is classified as FRA4 by the maintenance engineers, the dataset is

filtered to only contain this class. With this, the data that will be analyzed represents a similar condition to what is measured by the BRA IRV. The dataset now has 68155 rows and 15 columns since the "Class" column is useless now that it only contains class FRA 4 irregularities.

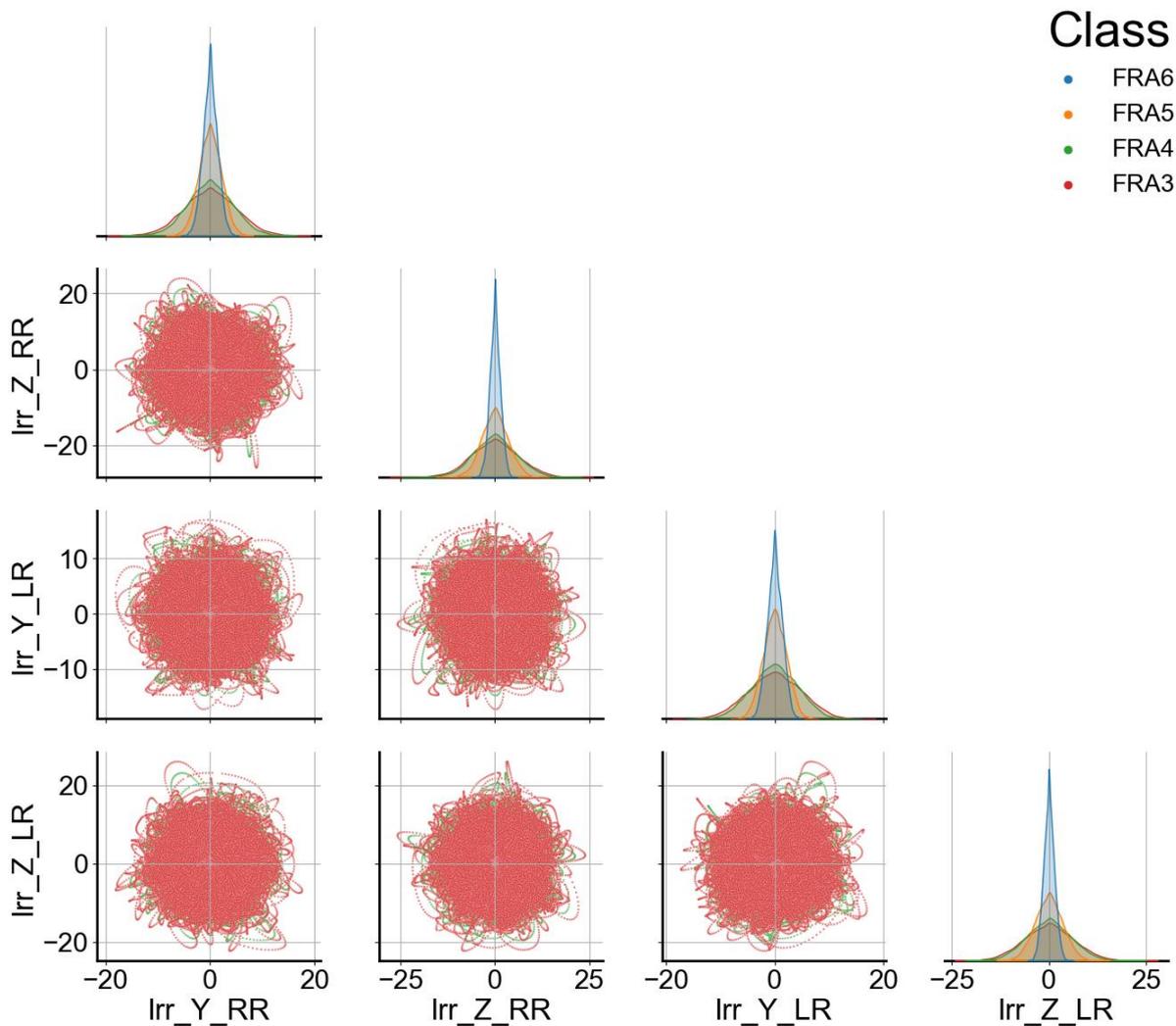


Figure 48: Distribution plot of the track irregularities.

3.2.3 Correlation analysis

There are two main approaches for studying correlation in machine learning: univariate correlation and multivariate correlation (BROWNLEE, 2019; CHRIST et al., 2018). The objective of correlation is to have a metric that describes the relationship between the input

dataset and the target variable. With these metrics, the data scientist can create an expectation of how the model will likely perform.

Univariate feature selection considers each feature individually while multivariate correlation (wrapper methods) relies on an external prediction model that provides coefficients indicating the importance of each input feature towards the output (KIMOTHO; SEXTRO, 2014). While univariate correlation shows individual features with redundant information or seemingly uninformative by themselves, multivariate correlation sees if their combination with other variables proves to be useful. Both types of correlation were used in the present study.

Figure 49 shows the correlation heatmap for the Pearson coefficient (linear correlation) and Spearman coefficient (monotonic correlation) for the filtered dataset containing FRA4 irregularities. From this figure, it is clear that the current instrumentation of the BRA IRV has a very small correlation with lateral irregularities and a reasonable correlation with vertical irregularities at around 0.47. Note the Pearson correlation of 0.52 between the vertical irregularities of the left rail (bottom row) and the spring displacement on the left side of the leading bogie (second column).

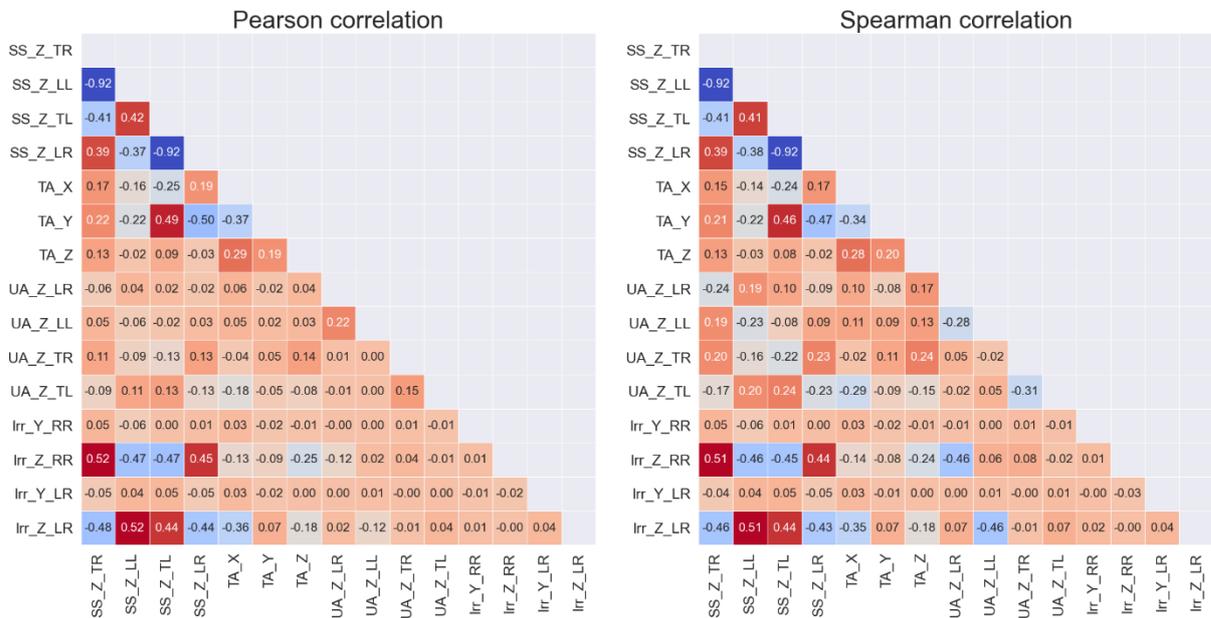


Figure 49: Correlation heatmaps.

These results seem to make intuitive sense due to how the current instrumentation of the BRA1 IRV is implemented. The vast majority of the IRV sensors measure vertical excitations while only the Y component of the triaxial accelerometer measures the lateral movement. Even so, the correlation of the lateral triaxial component with the lateral irregularities is practically

zero. This highlights the possible need for a different instrumentation configuration to accurately monitor the lateral movements of the wagon. With this in mind, the present thesis will focus on obtaining the vertical track irregularities from the IRV data.

In Figure 49, all four suspension sensors have a solid correlation with the vertical track irregularities while the triaxial accelerometer components have a low correlation. It is likely that the Z component of the triaxial accelerometer has a low correlation because of the filtering effect of the secondary suspension, which changes the signal significantly (TSUNASHIMA; NAGANUMA; KOBAYASHI, 2014). The uniaxial accelerometers in the leading bogie have a good monotonic correlation with the vertical irregularities while the sensors in the trailing bogie do not. This indicates that sensor placement would be better on the leading bogie when compared to the trailing bogie if the objective was to capture vertical track excitations.

Six different machine learning models with default hyperparameters were used to determine the multivariate correlation between the input and vertical track of the right rail, as shown in Figure 50. All models used are shown in the title and they were created using their default hyperparameters with no hyperparameter optimization. By using a diverse set of models and observing the coefficient of determination (R^2) metric, it is possible to confirm the observations done in the univariate correlation and determine if the mapping produces sufficiently accurate results.

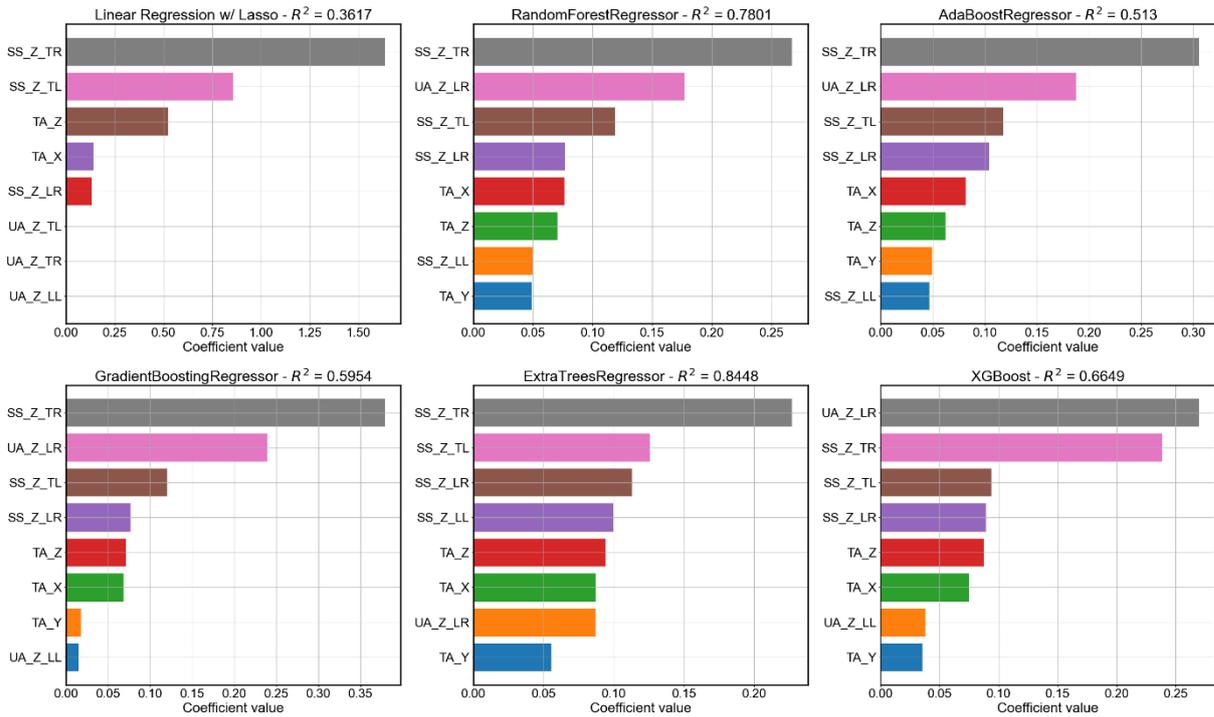


Figure 50: Multivariate correlation using a wrapper method for 6 different models.

From Figure 50, the objective is to find models that show some accurate mapping between the input and the output and analyze their feature importance. Looking at the R^2 metric, these models can either be the RandomForestRegressor or the ExtraTreesRegressor. From these 2 models, the leading uniaxial accelerometer on the right sideframe and the 4 trailing suspension sensors can be seen in the top 4, which is expected based on the univariate correlation with the right rail (third to bottom row in Figure 49). An interesting observation is that the trailing uniaxial accelerometer variables do not appear in the top 8 features of the models while the triaxial variables do. Even though the Y component of the triaxial sensor has about the same univariate correlation as the trailing uniaxial accelerometers, it is more important when considering all the variables together. The argument that positioning the sensors on the trailing bogie is not ideal seems more plausible.

To continue, the vertical track irregularity of the left rail was selected as the target variable to prove that the mapping is possible. The other track irregularity columns are removed, leading to an output dataset of 68155 rows and 12 columns.

3.2.4 Conclusions from EDA

The main conclusions of the EDA step can be summarized as follows:

- Outliers appeared only on the uniaxial accelerometers due to their proximity to the wheel-rail contact interface. The impact loads that occur in this interface likely lead to numerical errors which caused these outliers;
- Larger track irregularities (FRA3 - FRA5) mask the effect that the track curvature has on the measured IRV variables;
- Categorical variables that contain track design characteristics will likely not contribute to increasing model accuracy due to the masking effect of the more severe irregularities (FRA3-FRA4). Thus, they were removed;
- After removing the categorical variables, maintaining only class FRA4, and selecting a target variable, the output dataset has 68155 rows and 12 columns (11 sensor variables and the target variable);
- The baseline model performance is $R^2 = 0.8448$, which is far below the desired accuracy for practical implementations.

3.3 Feature engineering

Traditional data-driven methods likely require the design of new features (feature engineering) to better find the mapping between the dataset and the target variable. This procedure is very common in diagnostic and prognostic modeling (LEI et al., 2020; TSUI et al., 2015). Feature selection/extraction is also a key part of this process; therefore, it depends on the domain knowledge and expertise of the data scientist. The following sections show the creation of new columns based on domain knowledge and feature extraction using statistical metrics.

3.3.1 Domain knowledge of wagon movements

A typical bogie model with six degrees of freedom is shown in Figure 51. Due to the track excitations, the movement experienced by the vehicle will represent either a translation or rotation about these axes. Thompson et al. (2016) and Darbie et al. (2003) highlighted that the secondary suspension deflection (suspension travel) and accelerometers can be used to detect track geometry irregularities and also movements that correlate with oscillatory modes of the wagon (rotations around the 3 axes). Bleakly's IRV, for example, had the objective of capturing roll, pitch, and yaw from vertical and lateral accelerations of the wagon body (BLEAKLEY, 2006). This is shown in Figure 52.

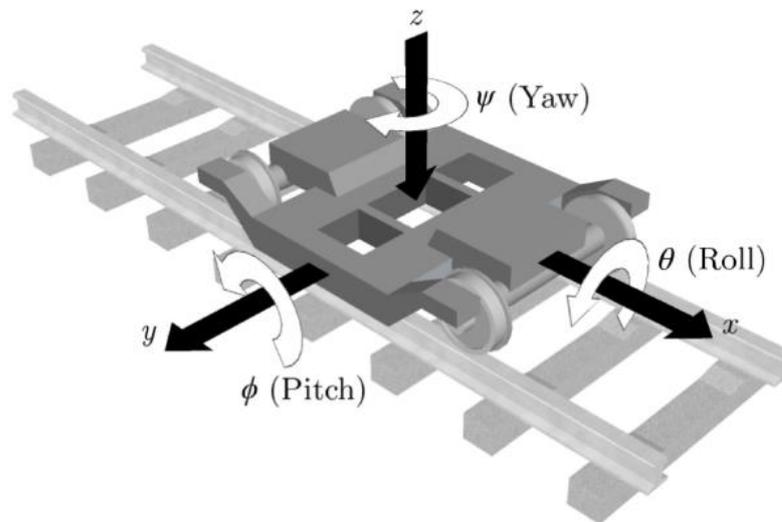


Figure 51: Representation of a bogies' six degrees of freedom - (YEO, 2017).

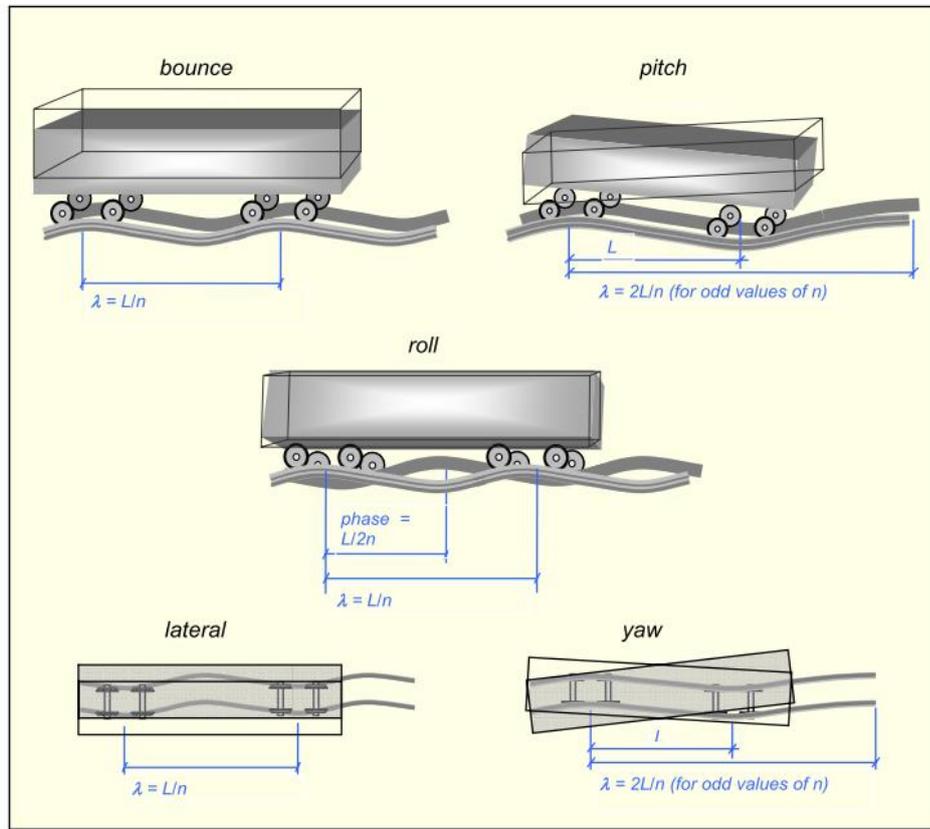


Figure 52: Excitation of oscillatory modes in the wagon body - (BLEAKLEY, 2006).

To represent these types of movements, the main references used were from Bleakly (2006) and Thompson et al. (2016). The sensor arrangement of Bleakly (2006) consists only of accelerometers and is shown in Figure 53. The variables that capture these degrees of freedom alongside their mathematical expressions used by Bleakly are shown in Figure 54.

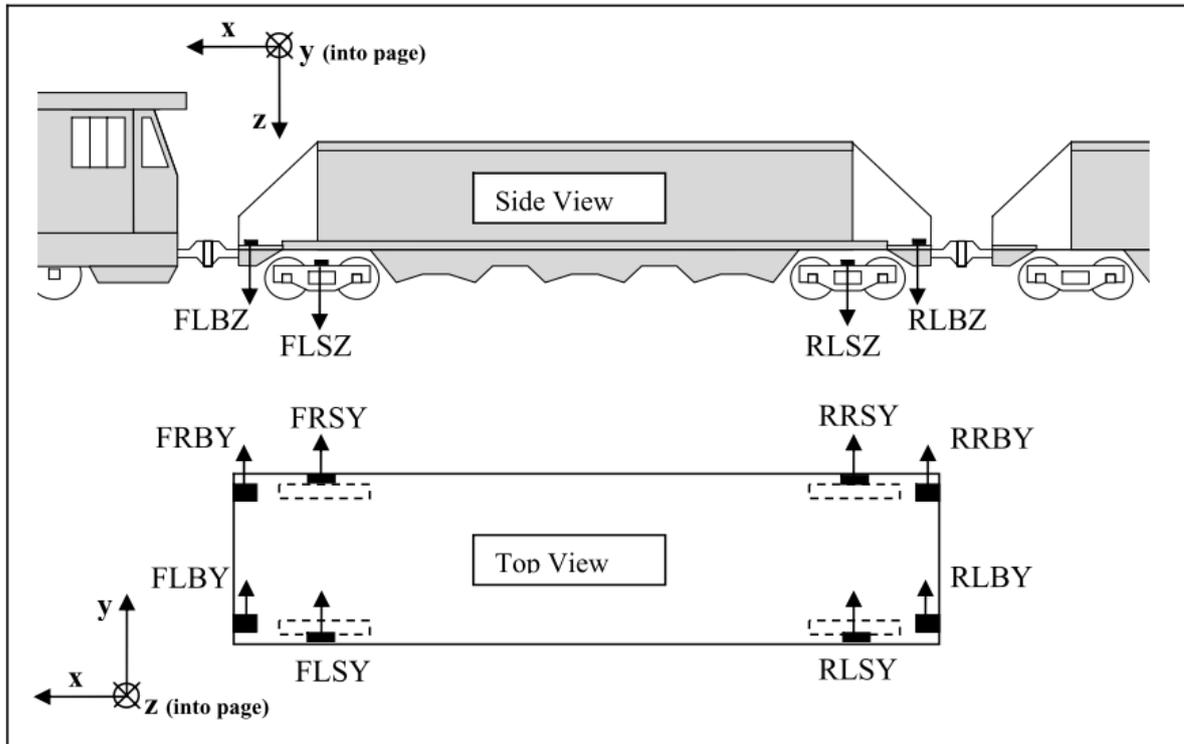


Figure 53: Accelerometer locations and axis naming convention - (BLEAKLEY, 2006).

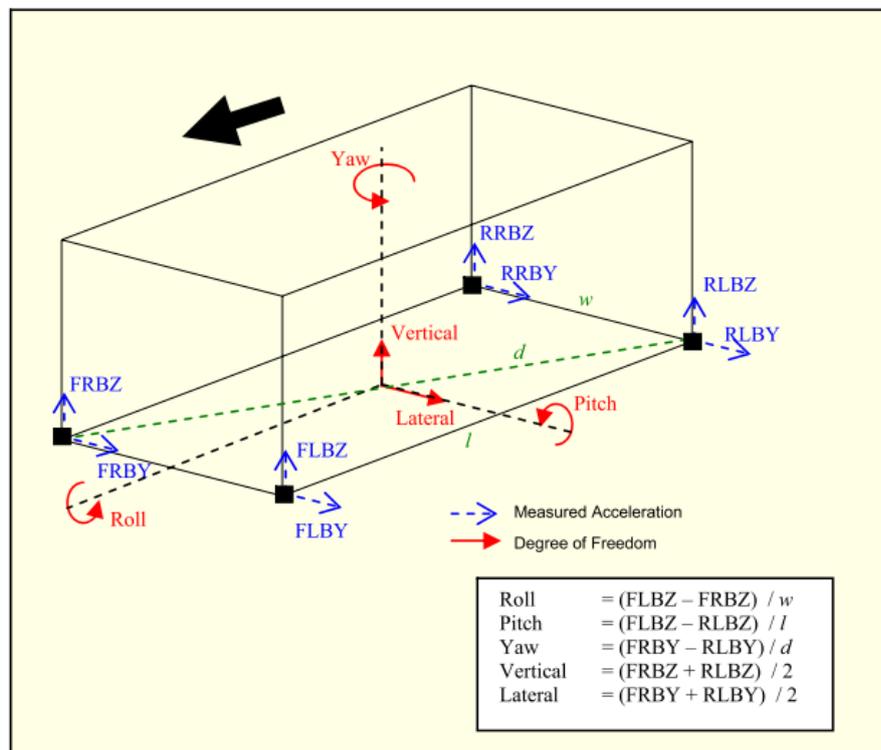


Figure 54: Relating corner accelerations to 5 degrees of freedom - (BLEAKLEY, 2006).

Comparing the IRV of Bleakly with the BRA IRV of this thesis, the main differences are:

- **Carbody:** BRA only has a triaxial IRV instead of Bleakly's setup in Figure 53;
- **Bogie:** BRA measures acceleration only in the vertical direction for similar sensor locations. BRA also measures secondary spring displacement while Bleakly measures lateral acceleration;
- **Axlebox:** Both IRVs do not have instrumentation on the axlebox.

Due to the BRA IRV having spring displacement sensors at approximately the same location as Bleakly's accelerometers, the mathematical expressions used in this thesis are different. The rigid body motions captured in this thesis are illustrated in Figure 55. The mathematical expressions used to create these new variables from domain knowledge are available in Appendix C.

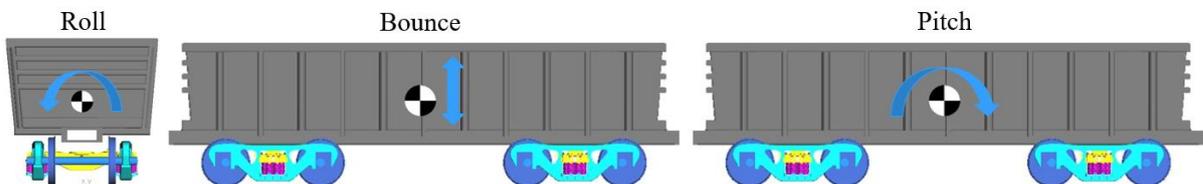


Figure 55: Example of rigid body motions that can be recreated using feature engineering on the IRV measurements.

3.3.2 Integrating accelerometer signals

As stated in Section 2.5, vertical irregularity values can be calculated by double integrating the vertical acceleration of the bogie or axlebox mounted sensors (BLEAKLEY, 2006; LEE et al., 2012; THOMPSON et al., 2016). Although there are disadvantages to simply using this integration to measure track geometry, it is still possible to take advantage of the natural correlation between the double-integrated uniaxial acceleration signal and the track irregularity for data-driven models. In this study, the two uniaxial accelerometer variables of the leading bogie were integrated numerically due to their good correlation with the output. The

integration was done using the `scipy` library in Python by applying Simpson's rule (`scipy.integrate.simps`).

3.3.3 Statistical metrics

Statistical metrics in the time domain, frequency domain, or time-frequency domain (wavelets) are commonly used for feature extraction from raw signal data (ANKRAH; KIMOTHO; MUVENGEI, 2020; LEI et al., 2020; TSUI et al., 2015). Larger track irregularities may cause the time-domain signal to change in amplitude and distribution compared to the signal excited by small irregularities. The frequency spectrum and its distribution may also change, meaning that new frequency components may appear associated with track severity (LEI et al., 2007). Only time domain metrics were used here due to their simplicity in implementation. These metrics were: mean, standard deviation, root mean absolute, mean absolute, root mean square, maximum absolute value, skewness, kurtosis, and peak-to-peak value.

A rolling window was used to create the new variables. To illustrate the procedure, consider a column vector T , a target variable and a window size (w) of 4. The first step is to create new columns that are the lagged features of T . Since the window size is 4, there would be 3 lagged features, as shown in Table 5 by columns $T-1$, $T-2$, and $T-3$. Note the appearance of NaNs due to the shift in the time series or because the mean could not be calculated with a NaN term. The next step is to calculate one of the statistical metrics stated previously, which is the mean in this illustration. The final step is to remove the rows with NaNs and the columns containing the lagged features. This results in a final dataset with fewer rows (size of the window) and an additional column for each statistical metric applied to T . All statistical metrics were applied to all numerical variables after the domain knowledge feature engineering step. It is important to note that only past information is being used to predict the present value. In Python, the Pandas package provides a function called “rolling” that does the steps shown previously (BROWNLEE, 2020).

Index	T	T-1	T-2	T-3	Mean (T,T-1,T-2,T-3)	Target
1	10	NaN	NaN	NaN	NaN	4.5
2	20	10	NaN	NaN	NaN	5.5
3	30	20	10	NaN	NaN	6.5
4	40	30	20	10	25	7.5
5	50	40	30	20	35	8.0
6	30	50	40	30	37.5	9.0
7	40	30	50	40	40	10.0

Table 5: Illustration of the rolling window process employed.

Index	T	Mean (T,T-1,T-2,T-3)	Target
4	40	25	7.5
5	50	35	8.0
6	30	37.5	9.0
7	40	40	10.0

Table 6: Dataset after the rolling window process

Since raw sensor data is a time series signal, a rolling window can be applied and generate new columns of data from these metrics. Tsunashima et al. (2019) used a rolling window of 4 samples to calculate the root mean square of the carbody vertical acceleration as a metric for track quality, although he stated that the exact size of the window is a parameter to be optimized. The window size was optimized and from a range between 5 and 100, a window size of 55 samples gave the best results. The ExtraTrees model was trained on multiple datasets obtained with different window sizes and the one that returned the best R^2 on the test set was used.

The final result of the feature engineering step was a dataset of 68101 rows and 410 columns. Note that some rows were discarded due to the rolling window of 55 samples. With the increase in column numbers from 12 to 410 columns, many of them likely present redundant or irrelevant features. These types of columns must be discarded to speed up model training, avoid overfitting and the curse of dimensionality (LEI et al., 2007, 2020). Therefore, feature selection must be done to only maintain the most important variables while minimizing model accuracy loss.

3.4 Feature selection

Taking advantage of the wrapper methods for multivariate correlation presented in Section 3.2.3, the ExtraTreesRegressor model with default hyperparameters was fitted on this dataset since it has shown to have the best performance out of all the 6 baseline models, as shown in Figure 50. All columns that have an importance coefficient lower than a constant a multiplied by the mean importance of the model is removed. This constant a is a hyperparameter to be tuned based on the amount of precision and number of variables the user wants to be removed. For this study specifically, it was determined that $a = 1.1$ gives satisfactory results, reducing the dataset from 410 columns to 30 while R^2 was reduced from 0.9738 to 0.9408. It is important to note that no hyperparameter tuning was done yet, so the drop-off in model performance is likely to be recovered.

3.5 Machine learning models

Once the final dataset is obtained, machine learning algorithms are trained to solve the regression problem of mapping the IRV variables to the vertical track irregularities. It is important to keep in mind that there is no universal algorithm that is guaranteed to perform well for all types of problems or datasets, also known as the “no free lunch theorem”. Different algorithms will perform better on different parts or characteristics of the dataset. With this in mind, a diverse set of different algorithms were selected to be trained and optimized.

If performance is inadequate, model stacking (GÉRON, 2019; PIRES et al., 2021) would be done to increase generalization and robustness. Stacking is a type of ensemble (model grouping) where different machine learning models are combined to overcome the limited generalization performance of each model. As a result, ensembles are capable of more accurate predictions than each model individually. More information is presented in Section 2.8.3.3.

The following machine learning models were used in this work (more information about these models is in Section 2):

- Linear regression with regularization (Lasso, Ridge, ElasticNet);
- Stochastic gradient descent with regularization – SGDRegressor;
- Forest algorithms: RandomForestRegressor and ExtraTreesRegressor;
- Boosting algorithms: AdaBoostRegressor, XGBRegressor, CatBoostRegressor, and LightGBM;
- Deep Neural Network.

The machine learning models were evaluated using k -fold cross validation with ten splits (KFold). This process involves splitting the dataset into k non-overlapping groups. A training dataset is created with all but one group while the model is evaluated on the leftover group. This process is repeated until each group is given a chance to be used as the holdout test set (which is 10 in this study). The average performance across all evaluations represents the model's performance. This work goes further and establishes a 95% confidence interval for the error metrics to assess the reliability of the predictions. Due to the restricted number of samples (10 in this case), a t-student distribution was used to obtain the confidence intervals (BROWNLEE, 2019).

3.6 Hyperparameter optimization

Hyperparameter optimization was done using the open-source package Optuna (AKIBA et al., 2019) due to its pruning strategy that avoids combinations of parameters that will not lead to model improvement. This makes the searching process faster than the traditional GridSearch and RandomSearch. The dataset was divided so that 70% would be used as the training set while the remainder for the test set.

One of Optuna's most attractive features is its visualizations for hyperparameter optimization. Figure 56 shows how the objective function of the neural network (MSE) changes over the number of trials. Each blue point is the objective function at a specific trial (combination of hyperparameters from the search space in Appendix A). The red line shows the best objective value over all trials. If a specific trial is better than all previous ones, the red line is updated. This helps show how many trials are needed for the objective function to

converge. After each trial, the search space is reduced by avoiding combinations of hyperparameters that will not perform well based on their history.

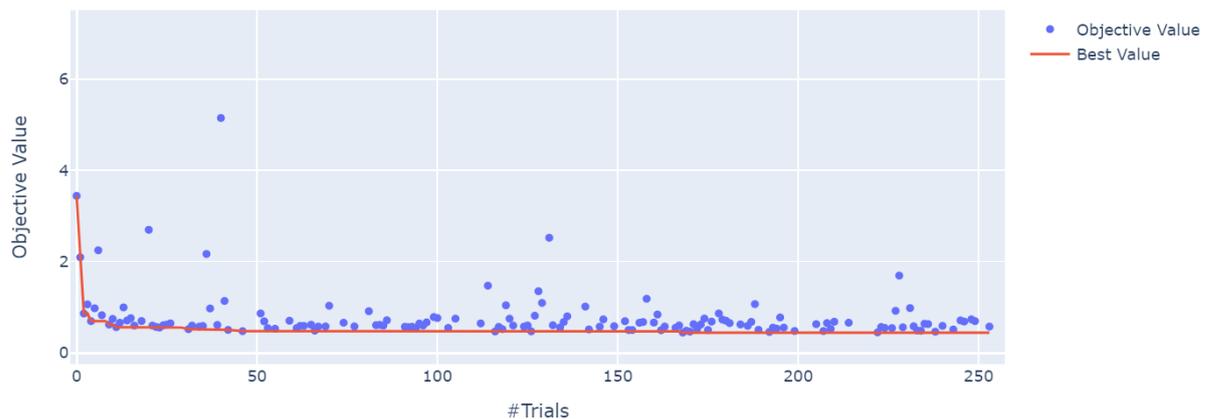


Figure 56: Objective history plot.

Figure 57 and Figure 58 show the objective function value attributed to each hyperparameter. The color is given by the trial number. In the case of Figure 57, it shows how three different types of activation functions (relu, tanh, and linear) affect the MSE. It can be seen that the tanh, relu, and tanh activation functions had the best objective function values for the 1st, 2nd, and 3rd layers respectively (from left to right). For the number of neurons in each layer, the optimal number is not as clear compared to Figure 58, however, it is possible to limit the hyperparameter range to focus on the more promising regions. An example of this would be limiting the search space that was initially 500-1500 neurons to 1000-1500 neurons in the 3rd layer (leftmost side of Figure 58). Note that the number of layers is also a hyperparameter, but to illustrate some of Optuna's visualization tools, a neural network with three layers was used, which is also the number of layers of the best performing model.

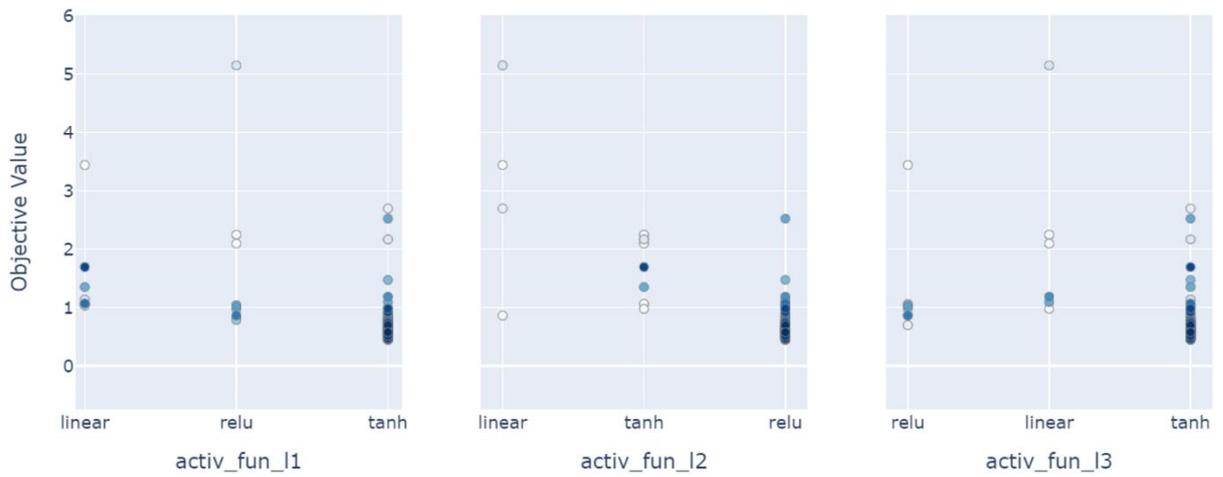


Figure 57: The effect that the activation function has on the objective function.

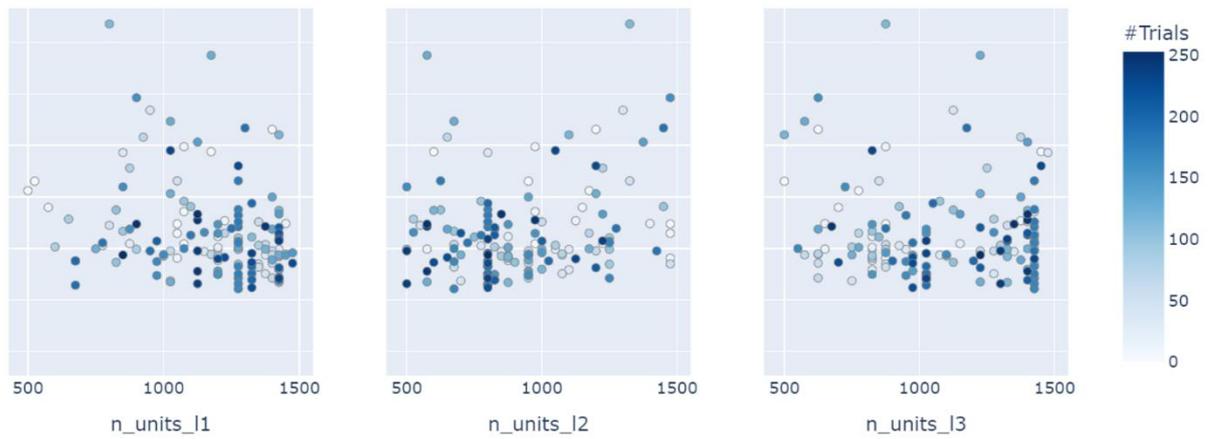


Figure 58: The effect that the number of neurons of each layer has on the objective function.

4 RESULTS

In this chapter, the results are presented following the methodology proposed in chapter 3. Table 7 shows the performance of the regression models used in this study after hyperparameter optimization. The top 4 rows show that linear models are not capable of finding an accurate mapping. RandomForestRegressor and ExtraTreesRegressor showed solid error metrics with a RMSE of 2.197 mm and 2.368 mm respectively, but not good enough for practical applications due to the order magnitude. Of all the boosting algorithms, CatBoostRegressor showed the best results with an RMSE of 1.125 mm. Overall, a deep neural network showed the best metrics with a RMSE of 0.556 mm. The rightmost column of Table 7 shows the average run time of each hyperparameter optimization run done in Optuna, considering one standard deviation. A total of 30 iterations were done to provide an initial estimate. Note that the neural network has a large standard deviation due to the number of layers being a hyperparameter.

Model	MAE	RMSE	R2	Time (s)
Lasso	2.885 ± 0.014	3.630 ± 0.024	0.591 ± 0.004	0.151 ± 0.054
Ridge	2.885 ± 0.020	3.630 ± 0.035	0.591 ± 0.006	0.009 ± 0.002
ElasticNet	2.885 ± 0.026	3.630 ± 0.033	0.591 ± 0.006	0.131 ± 0.032
SGD_Reg	2.887 ± 0.018	3.633 ± 0.027	0.591 ± 0.006	0.151 ± 0.054
RandomForest	1.641 ± 0.011	2.197 ± 0.032	0.850 ± 0.002	10.788 ± 5.585
ExtraTrees	1.095 ± 0.025	2.368 ± 0.013	0.926 ± 0.004	1.619 ± 0.865
LightBoost	0.818 ± 0.009	1.141 ± 0.026	0.960 ± 0.001	4.941 ± 2.018
XGBR	0.912 ± 0.008	1.271 ± 0.012	0.950 ± 0.001	23.733 ± 19.349
Adaboost	2.821 ± 0.018	3.514 ± 0.032	0.617 ± 0.005	40.166 ± 17.308
CatBoost	0.814 ± 0.029	1.125 ± 0.025	0.961 ± 0.006	12.325 ± 10.011
Neural Network	0.393 ± 0.027	0.556 ± 0.025	0.990 ± 0.005	780.008 ± 369.898

Table 7: Model results.

Figure 59 shows the optimal neural network architecture and its learning curves can be seen in Figure 60. The optimal models are shown in Table 7 and can be seen in Appendix B.

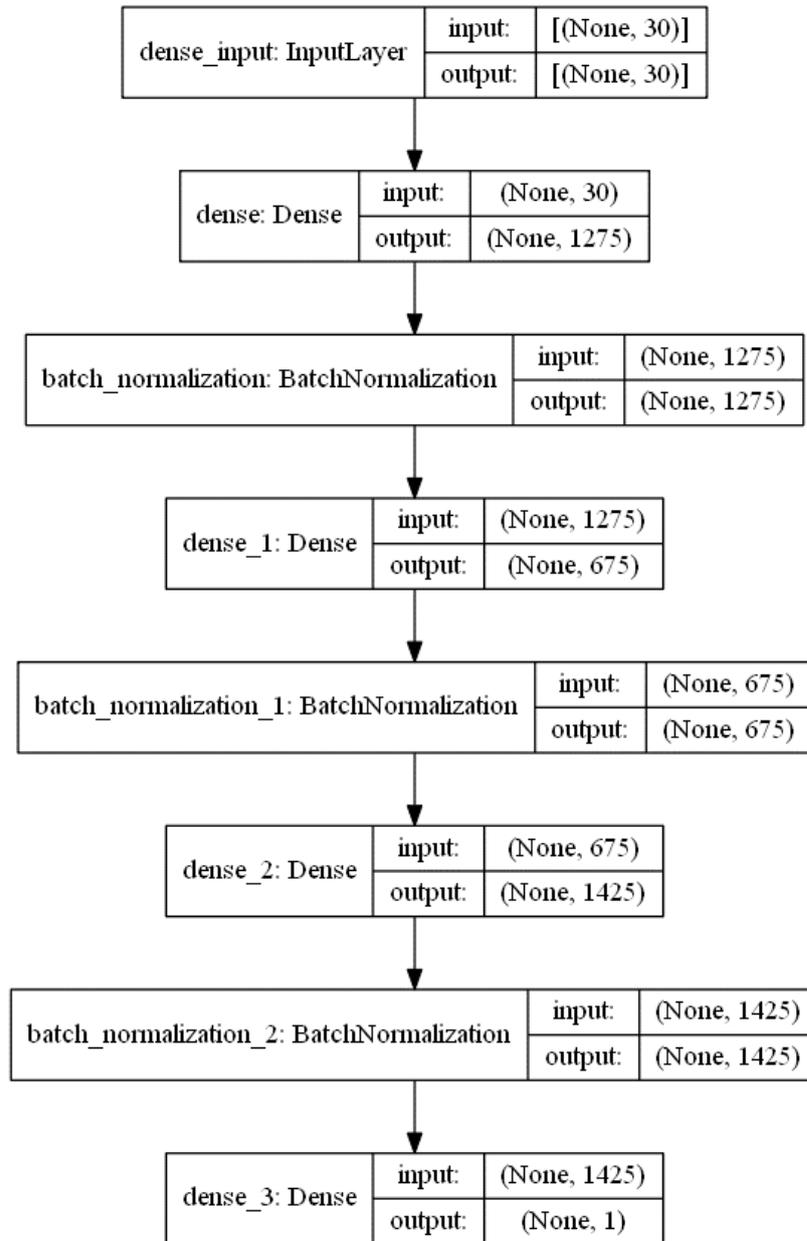


Figure 59: Optimized neural network architecture.

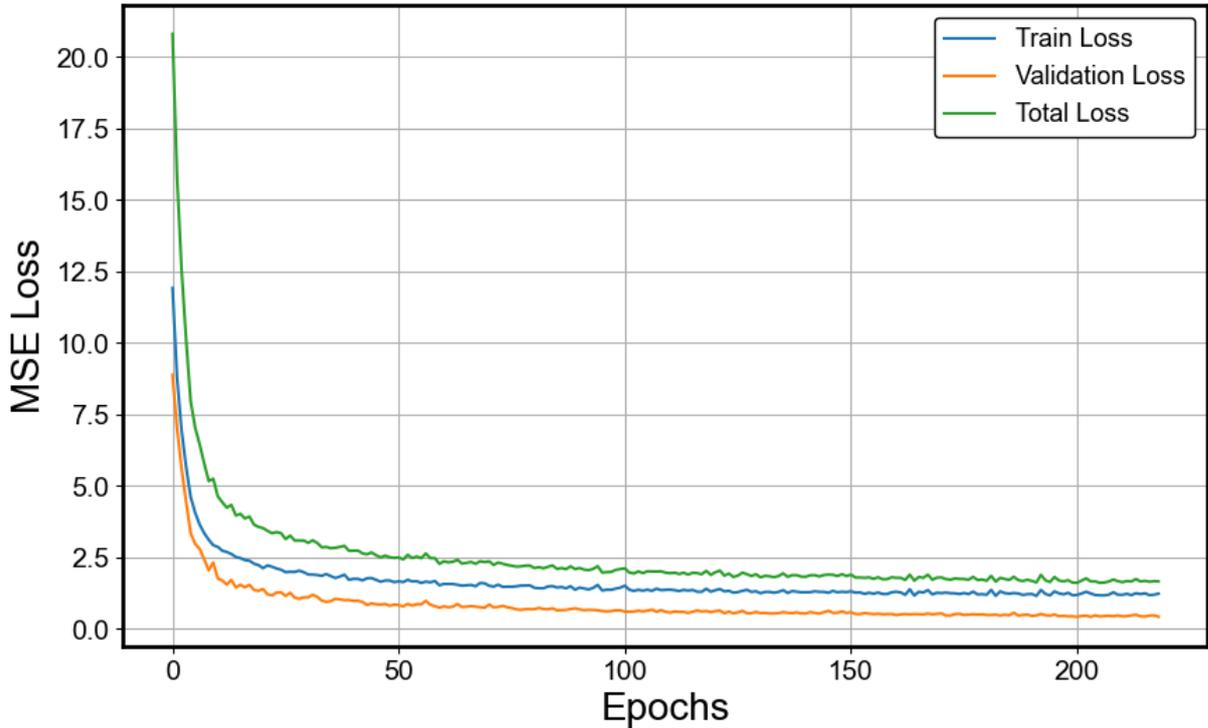


Figure 60: Neural network learning curve.

Compared to other researchers, Urda et al. (2021) estimated track irregularities using a scaled IRV and obtained an overall RMSE of 0.439 mm for vertical track irregularities. Although the model in this thesis provided an RMSE of 0.556 mm, postprocessing of the model's output signal allowed the reduction of the RMSE to 0.410 mm, making it comparable. The postprocessing step will be explained in Section 4.1.

4.1 Data postprocessing

As stated in Section 2.6.4, the artificial irregularities generated in *SIMPACT*TM should contain wavelengths between 3 m and 25 m, therefore, the output of the models should also be in this range. Figure 61 shows the power spectral density (PSD) of the target and predicted vertical irregularities from the best-performing model (neural network). Both target and prediction are inside the expected D1 wavelength range of the FRA 4 class; however, wavelengths lower than 3 m show a considerable difference. Since small wavelengths are the

same as large spatial frequencies, the prediction likely has a noise term that should be filtered out. Wavelengths larger than 25 m are of little concern since their order of magnitude is much larger than the length of the wagon, meaning that it is unlikely to excite the wagon dynamics.

Following the example of Yeo (2017), a fourth-order Butterworth low pass filter was applied to the predicted vertical irregularity to minimize the noise term. For this work specifically, a cutoff spatial frequency of 0.35 1/m (wavelength of about 2.86 m) was used since it is sufficiently close to the lower wavelength limit of 3 m while also showing better error metrics. From the purple line of Figure 61, the cutoff provided by this filter is much more abrupt.

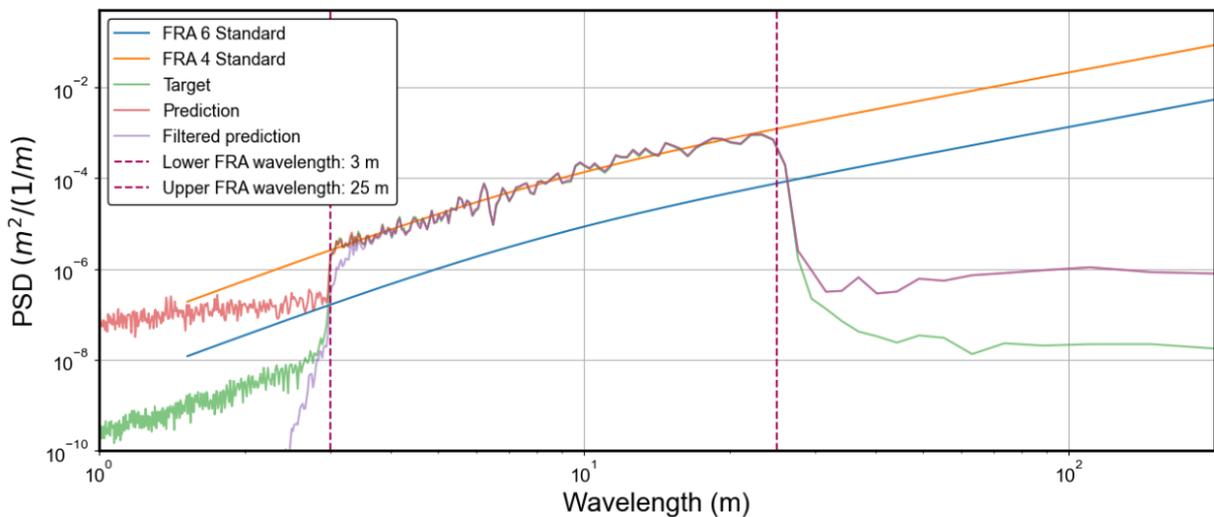


Figure 61: PSD of the target variable and model output compared to the FRA norm specifications.

The application of the filter leads to an improvement in the RMSE from 0.556 mm to 0.410 mm for the neural network, improving by around 26.3%. This increase in performance can be explained with the help of Figure 62. This figure shows the target variable with the filtered and unfiltered model prediction in the spatial domain for the test set. The smoothing effect caused by noise removal was responsible for improving the error metrics, which can be seen from index 61600 to 61800. The application of this filter is intended to be an option that the user can adopt to see if it delivers better performance or as a way to check if the model output makes sense.

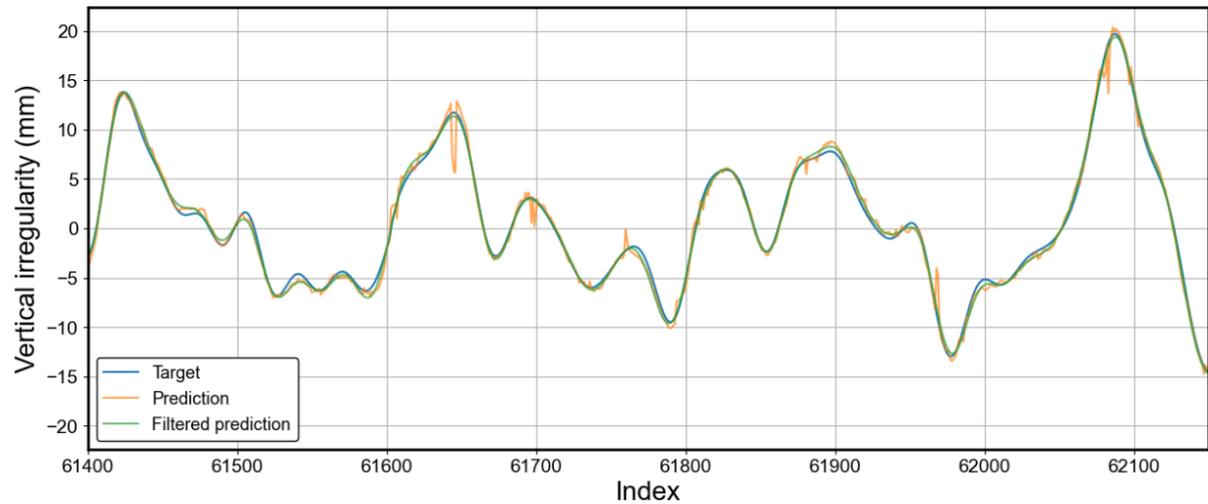


Figure 62: Comparison between the model prediction and filtered prediction for a section of the test set.

4.2 Model interpretability

The vertical track irregularities obtained from measured IRV data will be used to monitor the track and plan maintenance. For these decision-making applications, model interpretability is crucial so that domain experts responsible for maintenance planning trust the information provided by the ML model in their day-to-day operations (ALLAH BUKHSH et al., 2019). Specifically, it would be possible to obtain detailed information with regards to feature importance and critical instances where the IRV data measured very large values.

Figure 63 shows a flowchart of how model interpretability could complement a machine learning model for applications involving decision-making. Currently in the BRA railway, measured IRV data is visualized in reports and dashboards and compared to the tolerable limits for maintenance planning. The machine learning model would add another metric which would be the geometric irregularity values. Model explainability techniques would add a branch to the process for detailed explanations relating the measured dynamics to the track excitations. This work uses Shapely Additive Explanation (SHAP) as the model interpretability technique (LUNDBERG; LEE, 2017).

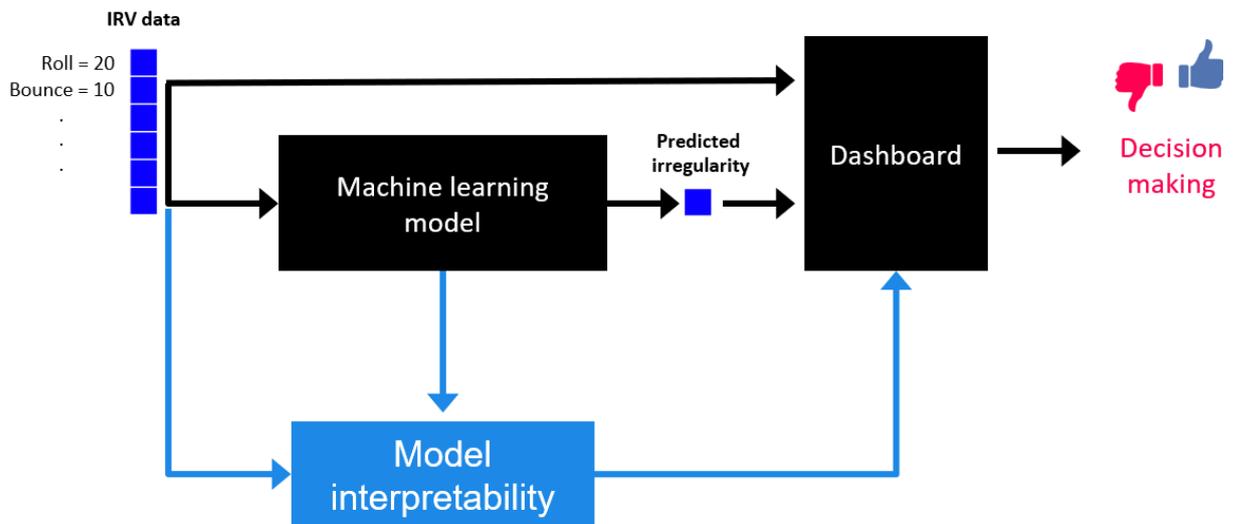


Figure 63: Flowchart of how to aggregate the machine learning with explainability techniques for decision making.

Developed by Lundberg and Lee (2017), SHAP uses game theory to explain the output of any machine learning model by calculating Shapley values. These values explain why the prediction of a specific point differs from the model average and how much each variable contributed to the final output. The Shapley value is defined as the average marginal contribution of a feature value across all possible feature combinations. For each combination, the marginal contribution is computed by taking the difference between the predicted variable with and without the specific feature value.

Figure 64 shows two ways of visualizing the SHAP values for the neural network model that allows different types of analysis to be done. The left figure shows the global feature importance of the top 15 variables, which is given by the mean absolute SHAP value for that feature over all given samples. The meaning behind each of these 15 variables is available in Appendix D. This means that, for this specific model, the top 3 variables that affect the output of the model the most are:

- The mean value of the suspension on the right side of the trailing bogie;
- The mean value of wagon pitch obtained from the suspension sensors on the left side of the bogie.
- Suspension displacement on the right side of the trailing bogie.

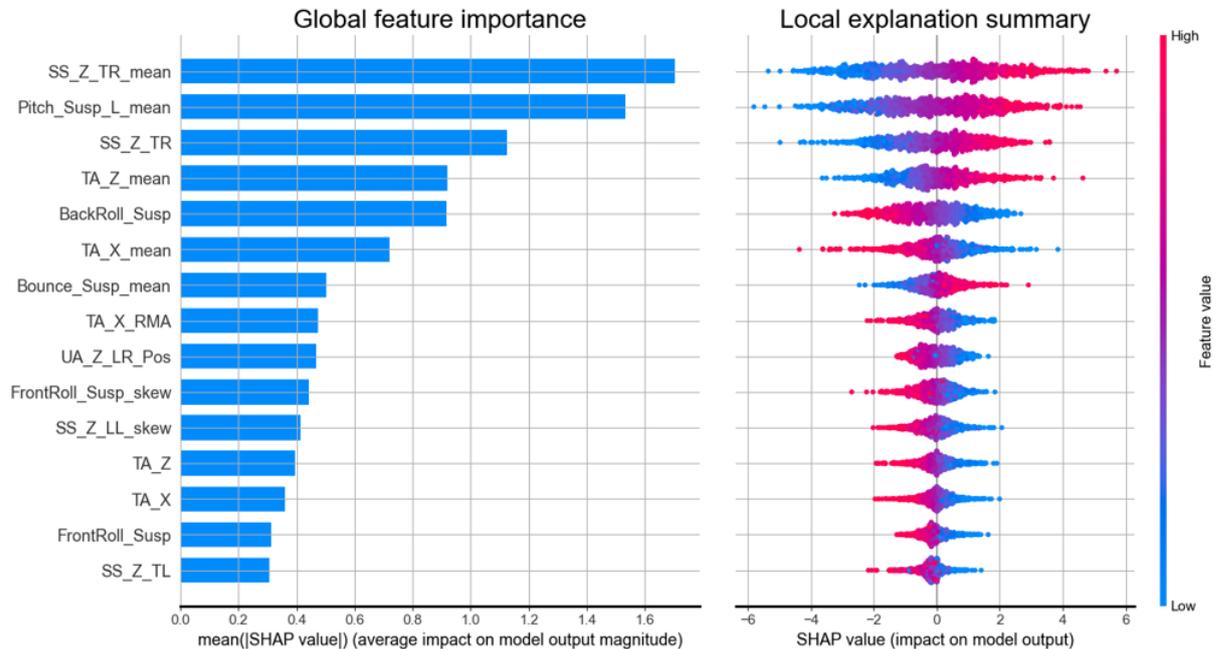


Figure 64: SHAP values for determining feature importance.

Note that variables related to the triaxial sensors have a considerable impact on the model output; something that the univariate correlation tests didn't reveal. Surprisingly, the only variable related to the uniaxial accelerations that made it into the top 15 was “UA_Z_LR_Pos” which represents the double integration of the accelerometer positioned on the right side of the leading bogie. This shows that the suspension sensors variables have the largest impact on the model output, followed by the triaxial accelerometer and the uniaxial accelerometers. The reason that features created from the uniaxial accelerometer measurements are not important is likely due to a large number of outliers still being present compared to other sensors. Another possible issue is the noise that appeared during the numerical integration of these variables. Future studies will focus on better understanding, defining these outliers, and reducing noise by integrating in the frequency domain.

The right side of Figure 64 highlights details that the global feature importance plot does not show, allowing an explanation of the individual data points (local explanation). The density of the data points can be seen in this plot, where the cold colors (blue) mean that the variable has small values while warm colors (red) represent large values. To better illustrate the information provided by this plot, the 5th most impactful variable will be used as an example due to roll being a type of wagon movement that will facilitate interpretation (Figure 55). The

local explanation reveals that large values of this variable negatively affect the model output while small values have a positive impact. This can be better visualized in Figure 65.

A linear relationship is observed between the roll of the trailing bogie (“BackRoll_Susp”) and its SHAP values. The question now is if this result makes intuitive sense from a physical standpoint. For the back roll variable, positive values represent a clockwise spin around the wagon's x-axis while negative values represent a counterclockwise spin. Keep in mind that the objective variable is the vertical irregularity on the right rail. If the wagon has a negative roll, the right vertical irregularity is likely to be larger than the left vertical irregularity since this creates the torque necessary for the roll movement. Therefore, large positive irregularity values on the right rail cause a counterclockwise spin (negative roll values), and this relationship was captured by the neural network.

Looking now at the color of Figure 65, it can be seen that there is a clear interaction term between the roll of the trailing bogie and the roll of the wagon as a whole. This makes sense because the wagon roll variable was created as the most critical case of roll observed considering both bogies. From this result, it can be said that the model is capable of capturing the underlying dynamics of the wagon and its relationship to track excitations.

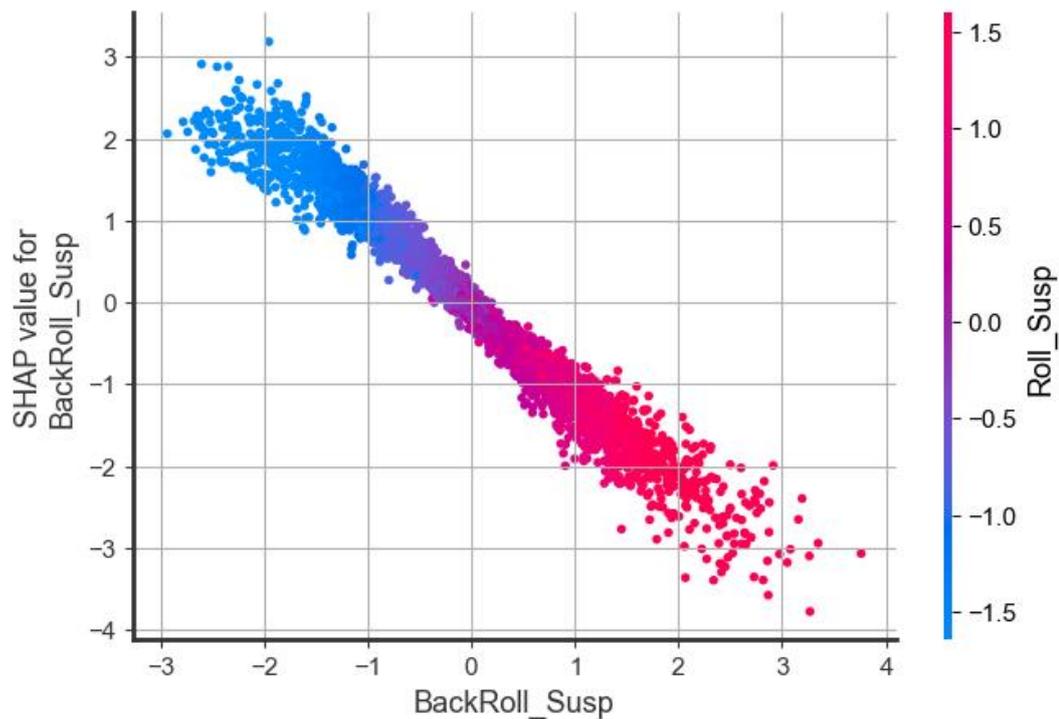


Figure 65: Shap dependence plot of the variable representing the wagon’s roll movement.

5 CONCLUSION

This thesis proposes a data-driven approach to obtaining geometric track irregularities from instrumented railway vehicle (IRV) data. Machine learning models are trained to find the nonlinear mapping between the IRV sensor data (input) and the track irregularities (output). A virtual dataset was generated using dynamic simulation of a BRA1 IRV model inside SIMPACK™. The dataset contains sensor variables measured by the BRA1 IRV and by other IRVs found in the literature. The target variable is class D1 (wavelength between $3 \leq \lambda \leq 25$ m) track irregularities due to their relationship with vehicle safety. With this dataset, nine different machine learning models were used and optimized using Optuna.

Exploratory data analysis revealed that, with the current instrumentation of the BRA1 IRV, it is possible to find the mapping between vertical track irregularities and IRV sensor measurements due to the Pearson and Spearman correlation coefficients of around 0.5 for some of the sensor variables. The Spearman correlation of the uniaxial vertical accelerometers on the trailing bogie was close to zero compared to the 0.46 observed for the same sensors on the leading bogie. This means that sensor placement on the trailing bogie for obtaining vertical track irregularities is not recommended.

For lateral track irregularities, accurate mapping is unlikely due to the correlation coefficients being close to zero for all variables. This is due to IRV instrumentation being tailored to monitor vertical excitations since most of the measured sensor components are vertical. A different instrumentation layout with sensors measuring lateral movements will likely lead to better correlation results. Therefore, machine learning models were trained to obtain the vertical track irregularities.

The results show that the best performing model was a deep neural network with a root mean squared error (RMSE) of 0.556 ± 0.025 mm and a coefficient of determination R^2 of 0.990 ± 0.005 . Inside the D1 wavelength range, the model output accurately follows the target variables, however, there was noise present in the model output. Applying a 4th order Butterworth lowpass filter on the model output to filter the desired wavelength range D1 caused the RMSE to improve from 0.556 mm to 0.410 mm.

Model interpretability techniques using SHAP values revealed that the top 3 most important features were created using suspension sensor variables while the 4th and 6th are the mean Z and X components of the triaxial accelerometer. Of the top 15 variables, only one of

them was obtained using uniaxial accelerometer measurements which was the double integration of the signal. Therefore, the suspension and triaxial sensors are the most important for measuring vertical track irregularities.

6 FUTURE STUDIES

- Improve upon the double integration done in feature engineering by detrending the data and removing noise by applying a bandpass filter after each integration step, as done in Haji Abdulrazgh et al. (2021) and Malekjafarian et al. (2021) while also analyzing the signal-to-noise ratio. Another option is integrating in the frequency domain;
- Simulate segments with different velocity profiles to verify model robustness;
- Better define outlier thresholds for the uniaxial accelerometer variables;
- Apply a statistical test to better determine and quantify the effect that track irregularities have on measured sensor variables;
- Study different instrumentation arrangements for monitoring lateral track irregularities to find a mapping between lateral track irregularities and sensor variables;
- Validate methodology with experimental data. For practical implementation, it is necessary to have IRV and track geometry data measured at a close enough date. In the work of Haji Abdulrazgh et al. (2021), acceleration data was measured by an instrumented car on July 18, 2018, and track geometry parameters were recorded by a track geometry car on July 15, 2018. A similar timespan would be required.

Bibliography

AGGARWAL, C. C. **Data Mining: The Textbook**. 1. ed. New York: Springer International Publishing, 2015.

AGGARWAL, C. C.; SATHE, S. Theoretical Foundations and Algorithms for Outlier Ensembles. **ACM SIGKDD Explorations Newsletter**, v. 17, n. 1, p. 24–47, 2015.

AKIBA, T.; SANO, S.; YANASE, T.; OHTA, T.; KOYAMA, M. Optuna: A next-generation hyperparameter optimization framework. **arXiv**, p. 2623–2631, 2019.

ALCOCER, D. C. Track Data-Oriented Maintenance Intervention Limit Determination for Ballasted Light Rail Tracks through Multibody Simulations. 2019.

ALLAH BUKHSH, Z.; SAEED, A.; STIPANOVIC, I.; DOREE, A. G. Predictive maintenance using tree-based classification techniques: A case of railway switches. **Transportation Research Part C: Emerging Technologies**, v. 101, n. June 2018, p. 35–54, 2019.

ANKRAH, A. A.; KIMOTHO, J. K.; MUVENGEI, O. M. Fusion of Model-Based and Data Driven Based Fault Diagnostic Methods for Railway Vehicle Suspension. **Journal of Intelligent Learning Systems and Applications**, v. 12, n. 03, p. 51–81, 2020.

ATTO-OKINE, N. O. **Big Data and Differential Privacy: Analysis Strategies for Railway Track Engineering**. 1st. ed. Newark, Delaware: 2017 John Wiley & Sons, Inc., 2017.

BAASCH, B.; ROTH, M.; HAVRILA, P.; GROOS, J. C. Detecting singular track defects by time-frequency signal separation of axle-box acceleration data. **Proceedings of the World Congress on Railway Research (WCRR)**, p. 1–6, 2019.

BAI, L.; LIU, R.; SUN, Q.; WANG, F.; WANG, F. Classification-learning-based framework for predicting railway track irregularities. **Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit**, v. 230, n. 2, p. 598–610, 2016.

BAI, L.; LIU, R.; LI, Q. Data-driven bias correction and defect diagnosis model for in-service vehicle acceleration measurements. **Sensors (Switzerland)**, v. 20, n. 3, p. 1–16, 2020.

BARBOSA, R. S. New method for railway track quality identification through the safety dynamic performance of instrumented railway vehicle. **Journal of the Brazilian Society of Mechanical Sciences and Engineering**, v. 38, n. 8, p. 2265–2275, 2016.

BERAWI, A. R. B. Improving Railway Track Maintenance Using Power Spectral Density (PSD). **Doctoral Dissertation, University of porto**, p. 224, 2013.

BLEAKLEY, S. S. **Time Frequency Analysis of Railway Wagon Body Accelerations for a**

- Low-Power Autonomous Device.** Queensland: Central Queensland University, 2006.
- BOWEY, R. **Monitoring and control of in-train forces in heavy haul operations.** Conference On Railway Excellence Adelaide. **Anais...**2018a.
- BOWEY, R. Monitoring and control of in-train forces in heavy haul operations. **CORE 2018: Rail: Smart, Automated, Sustainable, Conference on Railway Excellence**, v. 1, n. 30 April–2 May, 2018b.
- BROWNLEE, J. *Statistical Methods for Machine Learning.* p. 291, 2019.
- BROWNLEE, J. **Introduction to Time Series Forecasting with Python.** v1.9 ed. [s.l: s.n.].
- CARLVIK, F. Detection of Rail Squats from Axle Box Acceleration Optimization of a Machine Learning Algorithm. **Sound and Vibration**, 2020.
- CHONG, T. L. et al. Defining Rail Track Input Conditions Using an Instrumented Revenue Vehicle. **Procedia Engineering**, v. 188, p. 479–485, 2017.
- CHRIST, M.; BRAUN, N.; NEUFFER, J.; KEMPA-LIEHR, A. W. Time Series Feature Extraction on basis of Scalable Hypothesis tests (tsfresh – A Python package). **Neurocomputing**, v. 307, p. 72–77, 2018.
- COWIE, J.; TAYLOR, L.; FELSNER, B.; STRAMARE, A.; HARDIE, G. Use of instrumented revenue vehicle to manage 40 tonne axle load operation at Fortescue Metals Group Ltd. **Proceedings of the International Heavy Haul Association Conference**, p. 870–878, 2015.
- DARBY, M.; ALVAREZ, E.; MCLEOD, J.; TEW, G.; CREW, G. The development of an instrumented wagon for continuously monitoring track condition. **AusRAIL PLUS 2003**, v. 1, n. November, p. 17–19, 2003.
- DE ROSA, A.; KULKARNI, R.; QAZIZADEH, A.; BERG, M.; DI GIALLEONARDO, E.; FACCHINETTI, A.; BRUNI, S. Monitoring of lateral and cross level track geometry irregularities through onboard vehicle dynamics measurements using machine learning classification algorithms. **Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit**, v. 0, n. 0, p. 1–14, 2020.
- ESCALONA, J. L.; URDA, P.; MUÑOZ, S. A track geometry measuring system based on multibody kinematics, inertial sensors and computer vision. **Sensors (Switzerland)**, v. 21, n. 3, p. 1–27, 1 fev. 2021.
- FERREÑO, D.; SAINZ-AJA, J. A.; CARRASCAL, I. A.; CUARTAS, M.; POMBO, J.; CASADO, J. A.; DIEGO, S. Prediction of mechanical properties of rail pads under in-service conditions through machine learning algorithms. **Advances in Engineering Software**, v. 151, n. July 2020, 2021.

- GÉRON, A. **Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow - Concepts, Tools, and Techniques to Build Intelligent Systems**. 2. ed. Canada: O'Reilly Media, Inc., 2019.
- GERUM, P. C. L.; ALTAY, A.; BAYKAL-GÜRSOY, M. Data-driven predictive maintenance scheduling policies for railways. **Transportation Research Part C: Emerging Technologies**, v. 107, p. 137–154, 2019.
- GULER, H. Prediction of railway track geometry deterioration using artificial neural networks: A case study for Turkish state railways. **Structure and Infrastructure Engineering**, v. 10, n. 5, p. 614–626, 2014.
- HAJI ABDULRAZAGH, P.; HENDRY, M. T.; ROGHANI, A.; TOMA, E. Evaluating Rail Surface Roughness from Axle-Box Acceleration Measurements: Computational Metrology Approach. **Journal of Transportation Engineering, Part A: Systems**, v. 147, n. 12, p. 04021087, 2021.
- HARDIE, G.; ZHENG, D.; WALSH, A.; CRUMP, C. The Use of Instrumented Revenue Vehicle Data To Calculate Track Geometry Parameters. **CORE 2016, Maintaining the Momentum, Conference on Railway Excellence**, 2016.
- HESSER, D. F.; ALTUN, K.; MARKERT, B. Monitoring and tracking of a suspension railway based on data-driven methods applied to inertial measurements. **Mechanical Systems and Signal Processing**, v. 164, n. May 2021, p. 108298, 2022.
- KARIS, T. **Correlation between Track Irregularities and Vehicle Dynamic Response Based on Measurements and Simulations**. [s.l.] KTH Royal Institute of Technology, 2018.
- KASRAEI, A.; ZAKERI, J. A.; BAKHTIARY, A. Optimal track geometry maintenance limits using machine learning: A case study. **Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit**, 2020.
- KHAJEHEI, H. **Data-driven models for railway track geometry maintenance**. [s.l.] Luleå University of Technology, 2021.
- KHOSRAVI, M.; SOLEIMANMEIGOUNI, I.; AHMADI, A.; NISSEN, A. Reducing the positional errors of railway track geometry measurements using alignment methods: A comparative case study. **Measurement: Journal of the International Measurement Confederation**, v. 178, n. March, p. 109383, 2021.
- KIMOTHO, J. K.; SEXTRO, W. An approach for feature extraction and selection from non-trending data for machinery prognosis. **Proceedings of the Second European Conference of the Prognostics and Health Management Society**, v. 5, n. 4, p. 1–8, 2014.

- KOJIMA, T.; TSUNASHIMA, H.; MATSUMOTO, A. Fault detection of railway track by multi-resolution analysis. **WIT Transactions on the Built Environment**, v. 88, p. 955–964, 2006.
- LASISI, A.; ATTOH-OKINE, N. Machine Learning Ensembles and Rail Defects Prediction: Multilayer Stacking Methodology. **ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering**, v. 5, n. 4, p. 04019016, 2019.
- LEDERMAN, G.; CHEN, S.; GARRETT, J.; KOVAČEVIĆ, J.; NOH, H. Y.; BIELAK, J. Track-monitoring from the dynamic response of an operational train. **Mechanical Systems and Signal Processing**, v. 87, n. June 2016, p. 1–16, 2017a.
- LEDERMAN, G.; CHEN, S.; GARRETT, J. H.; KOVAČEVIĆ, J.; NOH, H. Y.; BIELAK, J. A data fusion approach for track monitoring from multiple in-service trains. **Mechanical Systems and Signal Processing**, v. 95, p. 363–379, 2017b.
- LEE, J. S.; CHOI, S.; KIM, S. S.; PARK, C.; KIM, Y. G. A mixed filtering approach for track condition monitoring using accelerometers on the axle box and bogie. **IEEE Transactions on Instrumentation and Measurement**, v. 61, n. 3, p. 749–758, 2012.
- LEI, Y.; HE, Z.; ZI, Y.; HU, Q. Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs. **Mechanical Systems and Signal Processing**, v. 21, n. 5, p. 2280–2294, 2007.
- LEI, Y.; YANG, B.; JIANG, X.; JIA, F.; LI, N.; NANDI, A. K. Applications of machine learning to machine fault diagnosis: A review and roadmap. **Mechanical Systems and Signal Processing**, v. 138, p. 106587, 2020.
- LI, D.; MEDDAH, A.; HASS, K.; KALAY, S. Relating track geometry to vehicle performance using neural network approach. **Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit**, v. 220, n. 3, p. 273–281, 2006.
- LINGAMANAİK, S. N.; THOMPSON, C.; NADARAJAH, N.; RAVITHARAN, R.; WIDYASTUTI, H.; CHIU, W. K. Using Instrumented Revenue Vehicles to Inspect Track Integrity and Rolling Stock Performance in a Passenger Network during Peak Times. **Procedia Engineering**, v. 188, p. 424–431, 2017.
- LUNDBERG, S. M.; LEE, S.-I. A Unified Approach to Interpreting Model Predictions. **NeurIPS Proceedings**, 2017.
- MALEKJAFARIAN, A.; OBRIEN, E. J.; QUIRKE, P.; CANTERO, D.; GOLPAYEGANI, F. Railway track loss-of-stiffness detection using bogie filtered displacement data measured on a passing train. **Infrastructures**, v. 6, n. 6, p. 1–17, 2021.

- MARSH, B. Multivariate Analysis of the Vector Boson Fusion Higgs Boson. n. September, 2016.
- MOHAMMADI, R.; HE, Q.; GHOFRANI, F.; PATHAK, A.; AREF, A. Exploring the impact of foot-by-foot track geometry on the occurrence of rail defects. **Transportation Research Part C: Emerging Technologies**, v. 102, n. September 2018, p. 153–172, 2019.
- MONAKALI, B.; RAVITHARAN, R. An Approach to Improving the Efficiency and Effectiveness of Maintenance through In-service Continuous Dynamic Performance Data In South Africa. **Maintaining the momentum: CORE 2016: conference on railway excellence**, 2016.
- MORI, H.; TSUNASHIMA, H.; TAKASHI, K.; MATSUMOTO, A.; MIZUMA, T. Condition Monitoring of Railway Track Using In-Service Vehicle. **Journal of Mechanical Systems for Transportation and Logistics**, v. 3, n. 1, p. 154–165, 2010.
- NADARAJAH, N.; SHAMDANI, A.; HARDIE, G.; CHIU, W. K.; WIDYASTUTI, H. Prediction of railway vehicles' Dynamic behavior with machine learning algorithms. **Electronic Journal of Structural Engineering**, v. 18, n. 1, p. 38–46, 2018.
- NEUHOLD, J.; VIDOVIC, I.; MARSCHNIG, S. Preparing Track Geometry Data for Automated Maintenance Planning. **Journal of Transportation Engineering Part A: Systems**, v. 146, n. 5, p. 1–11, 2020.
- ODASHIMA, M.; AZAMI, S.; NAGANUMA, Y.; MORI, H.; TSUNASHIMA, H. Track geometry estimation of a conventional railway from car-body acceleration measurement. **Mechanical Engineering Journal**, v. 4, n. 1, p. 16- 00498-16–00498, 2017.
- OPTUNA. Optuna Documentation. 2020.
- OZDEMIR, S.; SUSARLA, D. **Feature Engineering Made Easy**. [s.l.] Packt, 2018.
- PIRES, A. C.; MENDES, G. R.; SANTOS, G. F. M.; DIAS, A. P. C.; SANTOS, A. A. Indirect identification of wheel rail contact forces of an instrumented heavy haul railway vehicle using machine learning. **Mechanical Systems and Signal Processing**, v. 160, p. 107806, 2021.
- PODWÓRNA, M. Modelling of random vertical irregularities of railway tracks. **International Journal of Applied Mechanics and Engineering**, v. 20, n. 3, p. 647–655, 2015.
- RAVITHARAN, R. Safer Rail Operations: Reactive to Proactive Maintenance Using State-of-the-Art Automated In-service Vehicle-Track Condition Monitoring. **2018 International Conference on Intelligent Rail Transportation, ICIRT 2018**, p. 5–8, 2019.
- REZAEIANJOUYBARI, B.; SHANG, Y. **Deep learning for prognostics and health management: State of the art, challenges, and opportunities** Measurement: Journal of the

International Measurement Confederation, 2020.

SANTOS, G.; REICHL, P. Visualizing Rail Data Using Integrated Tools To Enhance Understanding and Planning. **Conference On Railway Excellence Adelaide**, n. July 2016, p. 5–7, 2014.

SHAFIULLAH, G. M.; ALI, A. B. M. S.; THOMPSON, A.; WOLFS, P. J. Predicting vertical acceleration of railway wagons using regression algorithms. **IEEE Transactions on Intelligent Transportation Systems**, v. 11, n. 2, p. 290–299, 2010.

SILVA, A. S. **SENSORIAMENTO OTIMIZADO DE VAGÕES COMO ALTERNATIVA AO USO DE RODEIROS INSTRUMENTADOS**. Campinas: Campinas, 2022.

SOBRINHO, O. G. et al. Big data analytics in support of the under-rail maintenance management at Vitória - Minas Railway. **Proceedings - 2021 IEEE International Conference on Big Data, Big Data 2021**, p. 6026–6028, 2021.

SOLEIMANMEIGOUNI, I.; AHMADI, A.; KUMAR, U. Track geometry degradation and maintenance modelling : A review. v. 0, n. 0, p. 1–30, 2016.

TAKIKAWA, H. T. AND M. Monitoring the Condition of Railway Tracks Using a Convolutional Neural Network. **Intech**, p. 13, 2012.

TAN, C. W.; WEBB, G. I.; PETITJEAN, F.; REICHL, P. Tamping effectiveness prediction using supervised machine learning techniques. **ICRT 2017: Railway Development, Operations, and Maintenance - Proceedings of the 1st International Conference on Rail Transportation 2017**, v. 2017- July, p. 1010–1023, 2018a.

TAN, C. W.; HERRMANN, M.; FORESTIER, G.; WEBB, G. I.; PETITJEAN, F. Efficient search of the best warping window for Dynamic time Warping. **SIAM International Conference on Data Mining, SDM 2018**, p. 225–233, 2018b.

THOMPSON, C.; REICHL, P.; ZENG, D.; WHITE, J.; AHMED, F.; SETHI, H. **Predictive maintenance approaches based on continuous monitoring systems at Rio Tinto**. Conference On Railway Excellence. **Anais...Australia**: 2016.

TSUI, K. L.; CHEN, N.; ZHOU, Q.; HAI, Y.; WANG, W. Prognostics and health management: A review on data driven approaches. **Mathematical Problems in Engineering**, v. 2015, 2015.

TSUNASHIMA, H.; NAGANUMA, Y.; MATSUMOTO, A.; MORI, H. Condition monitoring of railway track using in-service trains. In: PERPIÑA, X. (Ed.). **Reliability and Safety in Railway**. 1st. ed. [s.l: s.n.]. v. 1.

TSUNASHIMA, H. Condition Monitoring of Railway Tracks from Car-Body Vibration Using a Machine Learning Technique. **Applied Sciences**, v. 9, n. 13, p. 2734, 2019.

- TSUNASHIMA, H.; HIROSE, R. Condition monitoring of railway track from car-body vibration using time–frequency analysis. **Vehicle System Dynamics**, 2020.
- TSUNASHIMA, H.; NAGANUMA, Y.; KOBAYASHI, T. Track geometry estimation from car-body vibration. **Vehicle System Dynamics**, v. 52, n. SUPPL. 1, p. 207–219, 2014.
- URDA, P.; ACEITUNO, J. F.; MUÑOZ, S.; ESCALONA, J. L. Measurement of railroad track irregularities using an automated recording vehicle. **Measurement**, v. 183, n. January, p. 109765, 2021.
- WEI, L.; ZENG, J.; WU, P.; GAO, H. Indirect method for wheel-rail force measurement and derailment evaluation. **Vehicle System Dynamics**, v. 52, n. 12, p. 1622–1641, 2014.
- WESTON, P.; ROBERTS, C.; YEO, G.; STEWART, E. Perspectives on railway track geometry condition monitoring from in-service railway vehicles. **Vehicle System Dynamics**, v. 53, n. 7, p. 1063–1091, 2015.
- WESTON, P. F.; LING, C. S.; GOODMAN, C. J.; ROBERTS, C.; LI, P.; GOODALL, R. M. Monitoring lateral track irregularity from in-service railway vehicles. **Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit**, v. 221, n. 1, p. 89–100, 2007a.
- WESTON, P. F.; LING, C. S.; ROBERTS, C.; GOODMAN, C. J.; LI, P.; GOODALL, R. M. Monitoring vertical track irregularity from in-service railway vehicles. **Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit**, v. 221, n. 1, p. 75–88, 2007b.
- YEO, G. J. **Monitoring railway track condition using inertial sensors on an in-service vehicle**. [s.l.] University of Birmingham, 2017.
- YI, Q.; ZHAO, C. Y.; WANG, P.; GAO, X.; WANG, L. C.; KE, W. H. Estimation of rail axial force in high-speed railway ballastless track based on wave modes. **Mathematical Problems in Engineering**, v. 2019, 2019.
- YUEN, A.; ZHENG, D.; MUTTON, P.; YAN, W. Classification of impact signals from insulated rail joints using spectral analysis. **Notes on Numerical Fluid Mechanics and Multidisciplinary Design**, v. 139, p. 771–780, 2018.
- ZHAI, W. **Vehicle–Track Coupled Dynamics Theory and Applications**. 1. ed. Singapore: Springer Nature Singapore Pte Ltd., 2020.
- ZHANG, L.; REN, C.; YUAN, X.; ZHANG, W. Ride comfort control of in-wheel motor drive unmanned ground vehicles with energy regeneration. **Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering**, v. 235, n. 4, p. 1057–

1069, 2021.

ZHAO, H.; LIU, H.; HU, W.; YAN, X. Anomaly detection and fault analysis of wind turbine components based on deep learning network. **Renewable Energy**, v. 127, p. 825–834, 2018.

ZHAO, R.; YAN, R.; CHEN, Z.; MAO, K.; WANG, P.; GAO, R. X. Deep learning and its applications to machine health monitoring. **Mechanical Systems and Signal Processing**, v. 115, p. 213–237, 2019.

ZHU, M.; CHENG, X.; MIAO, L.; SUN, X.; WANG, S. Advanced stochastic modeling of railway track irregularities. **Advances in Mechanical Engineering**, v. 2013, 2013.

APPENDIX A – Optuna hyperparameter search space

This section contains the hyperparameter searchspace used by Optuna to optimize the performance of the machine learning models.

Lasso:

- alpha = range(0,2,0.05)

Ridge:

- alpha = range(0,2,0.05)

ElasticNet:

- alpha: range(0,2,0.05)
- l1_ratio: range(0,1,0.05)

SGDRegressor:

- alpha: range(0,2,0.05)
- penalty: ['l1', 'l2', 'elasticnet']
- loss: ['squared loss', 'huber', 'epsilon insensitive']

KNeighborsRegressor:

- n_neighbors: range(1,8,1)
- weights: ['uniform','distance']
- leaf_size: range(30,500,10)
- p: [1,2]
- distance metric: ['euclidean','manhattan']

RandomForestRegressor:

- n_estimators: range(60,600,20)
- min_samples_leaf: range(1,8,1)
- min_samples_split: range(1,8,1)
- bootstrap: [True,False]

ExtraTreesRegressor:

- n_estimators: range(40,800,20)
- min_samples_leaf: range(1,8,1)
- min_samples_split: range(1,8,1)
- bootstrap: [True,False]

AdaBoostRegressor:

- n_estimators: range(60,1000,20)
- learning_rate: range(0.001,0.1,0.005)
- loss: ['linear','square','exponential']

XGBRegressor:

- n_estimators: range(60,600,20)
- max_depth: range(2,20,1)
- learning_rate: range(0.001,0.1,0.005)
- min_child_weight: range(1,8,1)
- colsample_bytree: range(0,1,0.1)
- min_child_weight: range(4,12,1)
- subsample: range(0,1,0.1)
- gamma: range(0,10,1)

CatBoostRegressor:

- n_estimators: range(20,1000,20)
- loss: ['squared loss', 'MAPE', 'Poisson']
- l2_leaf_reg: [0.001,0.01,1,3,5,11,20,50,100]
- depth: range(4,12,1)

LightGBM:

- n_estimators: range(60,800,20)
- max_depth: range(4,100,2)
- num_leaves: range(50,300,20)
- min_data_in_leaf: range(2,20,2)

Deep neural network:

- Number of layers: 1,2,3
- Number of neurons: range(300,1500,25)
- Activation function: ['relu','linear','tanh']
- learning rate: loguniform(1e-4,1e-2,5e-4)

APPENDIX B – Optimal model parameters

This section shows the optimal models obtained after using Optuna for hyperparameter optimization. The neural network optimal

- Lasso(alpha=0.02)
- Ridge(alpha = 1.06)
- ElasticNet(alpha=1e-05, l1_ratio=0.262)
- SGDRegressor(loss = 'squared_loss', penalty='elasticnet', alpha=0.001)
- KNeighborsRegressor(n_neighbors=2, weights='distance', metric='manhattan', leaf_size=473, p=1)
- RandomForestRegressor(n_estimators = 170, max_depth = 24, min_samples_leaf = 1, min_samples_split=5, max_features=24)
- ExtraTreesRegressor(bootstrap=True, min_samples_split=7, n_estimators=200, min_samples_split = 7, min_samples_leaf=1)
- AdaBoostRegressor(learning_rate=0.846, loss='square', n_estimators=500)
- XGBRegressor(max_depth = 20, learning_rate = 0.04, n_estimators = 230, colsample_bytree = 0.85, min_child_weight=11, subsample=0.77, gamma= 3)
- CatBoostRegressor(n_estimators = 790, loss="MAPE", depth=12, l2_leaf_reg= 100)
- Deep neural network:
 - Layer 1: 1275 neurons, linear activation function
 - Layer 2: 925 neurons, relu activation function
 - Layer 3: 1150, tanh activation function
 - Output layer: 1 neuron, linear activation function
 - Learning rate: 1e-3

APPENDIX C – Feature engineering equations for wagon movements

This appendix contains all formulas used for feature engineering based on domain knowledge. Since the EFVM has 4 suspension displacement sensors and 4 sideframe accelerometers positioned in different locations, a nomenclature code of 3 letters will be used to describe these sensors:

1st letter: represents which bogie the sensor is located (F for front bogie and T for trailing bogie)

2nd letter: represents on which side the sensor is located (R for right and L for left)

3rd letter: represents the sensor being used (Z for vertical displacement of the suspension sensor and A for the uniaxial accelerometer)

Example: FLZ means the suspension sensor (Z) located on the left side (L) of the front bogie (F).

Most of the formulas as based on Boolean operations that follow the frame below:

$$X = (\text{logical test}, \text{value if true}, \text{value if false})$$

Maximum acceleration:

$$A_{max} = IF(\max(FLA, FRA, TLA, TRA) \geq \min(FLA, FRA, TLA, TRA), \max(FLA, FRA, TLA, TRA), \min(FLA, FRA, TLA, TRA))$$

Where FLA is the accelerometer positioned on the left side of the front bogie, FRA is the accelerometer positioned on the right side of the front bogie, TLA is the accelerometer positioned on the left side of the trailing bogie, TRA is the accelerometer positioned on the right side of the trailing bogie and A_{max} is the maximum acceleration value.

Suspension travel of the left side of the IRV

$$ST_{left} = IF(\max(FLZ, TLZ) \geq \min(FLZ, TLZ), \max(FLZ, TLZ), \min(FLZ, TLZ)),$$

Where FLZ is the displacement sensor positioned on the left side of the front bogie, TLZ is the displacement sensor positioned on the left side of the trailing bogie and ST_{left} is the suspension travel calculated on the left side of the bogie.

Suspension travel of the right side of the IRV

$$ST_{right} = IF(\max(FRZ, TRZ) \geq \min(FRZ, TRZ), \max(FRZ, TRZ), \min(FRZ, TRZ))$$

Where FRZ is the displacement sensor positioned on the right side of the front bogie, TRZ is the displacement sensor positioned on the right side of the trailing bogie and ST_{right} is the suspension travel calculated on the right side of the bogie.

Largest suspension travel

$$ST = IF(\max(ST_{left}, ST_{right}) \geq \min(ST_{left}, ST_{right}), \max(ST_{left}, ST_{right}), \min(ST_{left}, ST_{right}))$$

Where ST is the largest suspension travel value considering both left and right sides.

Pitch (measured with accelerometers)

Right pitch

$$P_{accel_{left}} = \frac{FRA - TRA}{M}$$

Where M is the distance between accelerometers FRA and TRA in meters and $P_{accel_{left}}$ is the pitch movement calculated on the left side of the vehicle using the accelerometers.

Left pitch

$$P_{accel_{right}} = \frac{FLA - TLA}{M}$$

Where $P_{accel_{right}}$ is the pitch movement calculated on the right side of the vehicle using the accelerometers.

Largest Pitch value

$$P_{accel} = IF(\max(P_{accel_{left}}, P_{accel_{right}}) \geq \min(P_{accel_{left}}, P_{accel_{right}}), \max(P_{accel_{left}}, P_{accel_{right}}), \min(P_{accel_{left}}, P_{accel_{right}}))$$

P_{accel} is the largest pitch movement of the vehicle calculated using accelerometers.

Pitch (measured with spring displacement):**Left pitch**

$$P_{susp_{left}} = \frac{FRZ - TRZ}{N}$$

Where N is the distance between displacement sensors FRZ and TRZ in meters and $P_{susp_{left}}$ is the pitch movement calculated on left side of the railway vehicle using the displacement sensors.

Right pitch

$$P_{susp_{right}} = \frac{FLZ - TLZ}{N}$$

$P_{susp_{right}}$ is the pitch movement calculated on right side of the railway vehicle using the displacement sensors.

Largest pitch value

$$P_{susp} = IF(\max(P_{susp_{left}}, P_{susp_{right}}) \geq \min(P_{susp_{left}}, P_{susp_{right}}), \max(P_{susp_{left}}, P_{susp_{right}}), \min(P_{susp_{left}}, P_{susp_{right}}))$$

Where $P_{susp_{right}}$ is the largest pitch value calculated from the displacement sensors.

Bounce (measured with accelerometers):**Bounce of the leading bogie**

$$B_{accel_{front}} = \frac{FLA - FRA}{2}$$

Where $B_{accel_{front}}$ is the bounce movement of the leading bogie calculated from the accelerometers.

Bounce of the trailing bogie

$$B_{accel_{back}} = \frac{TLA - TRA}{2}$$

Where $B_{accel_{back}}$ is the bounce movement of the trailing bogie calculated from the accelerometers.

Largest bounce value

$$B_{accel} = IF(\max(B_{accel_{front}}, B_{accel_{back}}) \geq \min(B_{accel_{front}}, B_{accel_{back}}), \max(B_{accel_{front}}, B_{accel_{back}}), \min(B_{accel_{front}}, B_{accel_{back}}))$$

Where B_{accel} is the largest bounce movement of the railway vehicle calculated from the accelerometers.

Bounce (measured with displacement sensors):**Bounce of the leading bogie**

$$B_{susp_{left}} = \frac{FLZ - FRZ}{2}$$

Where $B_{susp_{left}}$ is the bounce movement of the leading bogie calculated from the displacement sensors.

Bounce of the trailing bogie

$$B_{susp_{right}} = \frac{TLZ - TRZ}{2}$$

Where $B_{susp_{right}}$ is the bounce movement of the leading bogie calculated from the displacement sensors.

Largest bounce value

$$B_{susp} = IF(\max(B_{susp_{left}}, B_{susp_{right}}) \geq \min(B_{susp_{left}}, B_{susp_{right}}), \\ \max(B_{susp_{left}}, B_{susp_{right}}), \min(B_{susp_{left}}, B_{susp_{right}}))$$

Where B_{susp} is the largest bounce movement of the vehicle calculated from the displacement sensors.

Roll (measured with accelerometers):**Roll of the leading bogie**

$$R_{accel_{front}} = \frac{FRA - FLA}{P}$$

Where P is the distance between accelerometers FRA and FLA in meters and $R_{accel_{front}}$ is the roll movement calculated on the leading bogie using the accelerometers.

Roll of the trailing bogie

$$R_{accel_{back}} = \frac{TRA - TLA}{P}$$

Where $R_{accel_{back}}$ is the roll movement calculated on the trailing bogie using the accelerometers.

Largest roll value of the bogie

$$R_{accel} = IF(\max(R_{accel_{front}}, R_{accel_{back}}) \geq \min(R_{accel_{front}}, R_{accel_{back}}), \\ \max(R_{accel_{front}}, R_{accel_{back}}), \min(R_{accel_{front}}, R_{accel_{back}}))$$

Where R_{accel} is the largest roll value of the vehicle calculated from the accelerometers.

Roll (measured with displacement sensors):

Roll of the leading bogie

$$R_{susp_{front}} = \frac{FRZ - FLZ}{Q}$$

Where Q is the distance between displacement sensors TRZ and TLZ in meters and $R_{accel_{front}}$ is the roll movement calculated on the leading bogie using the accelerometers.

Roll of the trailing bogie

$$R_{susp_{back}} = \frac{TRZ - TLZ}{P}$$

Where $R_{susp_{back}}$ is the roll movement calculated on the trailing bogie using the accelerometers.

Largest roll value of the bogie

$$R_{susp} = IF(\max(R_{susp_{front}}, R_{susp_{back}}) \geq \min(R_{susp_{front}}, R_{susp_{back}}), \max(R_{susp_{front}}, R_{susp_{back}}), \min(R_{susp_{front}}, R_{susp_{back}}))$$

Where R_{susp} is the largest roll value of the vehicle calculated from the displacement sensors.