Proceedings of the Sixth International Conference on
Railway Technology: Research, Development and Maintenance
Edited by: J. Pombo
Civil-Comp Conferences, Volume 7, Paper 8.13
Civil-Comp Press, Edinburgh, United Kingdom, 2024
ISSN: 2753-3239, doi: 10.4203/ccc.7.8.13

Machine Learning-Based Parametric Analysis of Railway Systems

J. A. Sainz-Aja[1], D. Ferreño[1], J. Pombo[2,3], I. Carrascal[1], J. Casado[1], S. Diego[1], J. Castro[1] and I. Rivas[1]

[1]University of Cantabria, University of Cantabria, Spain
[2]Institute of Railway Research, School of Computing and Engineering, University of Huddersfield, UK
[3]IDMEC, Instituto Superior Técnico, Universidade de Lisboa Lisboa, Portugal

Abstract

Effective and meticulous management of railway infrastructure is essential to prevent accidents and minimize operation and maintenance costs. This requires comprehensive knowledge of the assets, their interactions, and the impact of each track parameter on the overall performance of the infrastructure. This study conducts extensive analyses using a previously calibrated finite element model of slab track, varying key track parameters within their typical ranges. The resulting data is then used to train and validate predictive models employing machine learning algorithms. This approach provides deeper insights and improves the prediction of track behavior, which involves numerous variables such as soil/subgrade, supporting layers, sleepers, pads, and rails. Additionally, the study considers the impact of train axle loads and service speeds, which are crucial factors affecting track performance. The findings highlight that the most influential parameters on railway infrastructure are soil properties, rail pad characteristics, and axle loads. This research can facilitate the implementation of predictive maintenance strategies for railway tracks and the development of innovative technological solutions, addressing industrial needs for cost reduction and enhancing the competitiveness of railway transport.

Keywords: Railway tracks; Infrastructure assets; Predictive models; Machine learning algorithms; Monte Carlo method; Predictive maintenance.

1 Introduction

Modern societies demand efficient transport systems for both passengers and goods, prioritizing speed, comfort, safety, and environmental sustainability. The railway stands out among transportation options due to its high safety, reliability, cost-effectiveness, and low CO2 emissions [1–3]. To capitalize on these advantages and increase railway usage, there has been significant investment in research and infrastructure development, particularly in high-speed rail.

Understanding the behavior of track components and their interactions under varying loads and environmental conditions is crucial for optimizing infrastructure management and maintenance. Numerous factors affecting railway performance can be grouped into train-related loads and frequencies, environmental conditions like temperature, and material properties of track components [4–8]. Despite extensive research, no comprehensive study currently predicts the overall track response to external actions and component interactions. However, advancements in computational methods such as FE and Machine Learning (ML) offer new opportunities in this area. Recent examples include ML models for predicting mechanical behavior and assessing structural responses to different conditions.

This work aims to analyze four key quantities for assessing dynamic track behavior: rail and slab displacements and accelerations. A literature review identified variables influencing dynamic behavior, including axle loads, speed, wheel passing frequency, wheel-rail contact forces, and environmental temperature in Seville (Spain) and Moscow (Russia). Material properties considered include the density, Young's modulus, and Poisson's ratio of fastening system components, concrete slab characteristics, and soil geotechnical properties.

Using statistical distributions of infrastructure variables, 5400 random samples were generated via the Monte Carlo method [9–11]. These scenarios were simulated using an experimentally validated FE model [12,13] to obtain rail and slab displacements and accelerations. The resulting datasets were analyzed with ML algorithms, including multilinear regression, K-nearest neighbors, decision trees, random forest, gradient boosting, and neural networks. The best models for each assessment variable were selected and interpreted using permutation importance and partial dependence plots, establishing recommended operational ranges for track features.

The remainder of the paper is organized as follows: Section 2.1 defines the FE model and material properties, Section 2.2 describes the material properties, Section 2.3 defines the procedure to generate the syntetic data, and Section 2.4 outlines the ML and statistical methods. Section 3 presents the results and analysis, and Section 4 discusses the interpretation and relevance of the findings.

2 Methods

2.1. FE Model of the track

The FE track model utilized in this study was calibrated using laboratory tests [12,13]. Developed with the Harmonic Response module of ANSYS, this dynamic model underwent three key modifications to reduce computational cost. Firstly, symmetry conditions were applied, allowing the use of just one-quarter of the original model. Secondly, to accommodate the analysis of various fastening systems, the original fastening elements (EPDM elastic pad, steel plate, and rubber pad) were replaced by a single entity with equivalent mechanical properties. Finally, since the two soil layers in the original model (subgrade and frost protection layer) are both compacted sands with differing compaction levels, they were merged into one layer with equivalent properties. Fig. 1 depicts the final system configuration and the various components, with their dimensions detailed in Table 1.

Table 1: Dimensions of the track model

ID	Layer	Material	Width [mm]	Length [mm]	Height [mm]
1	Subgrade + FPL	Compacted sand	6000	2200	1200
3	HBL	Concrete layer (low quality)	3000	2100	300
4	Grout	Bituminous grout	2550	2100	40
5	Slab	Concrete (HA-35)	2550	1930	200
6	Fastening system	EVA / EPDM / TPE	150	160	6
7	Rail (UIC 60)	Steel	--	--	---

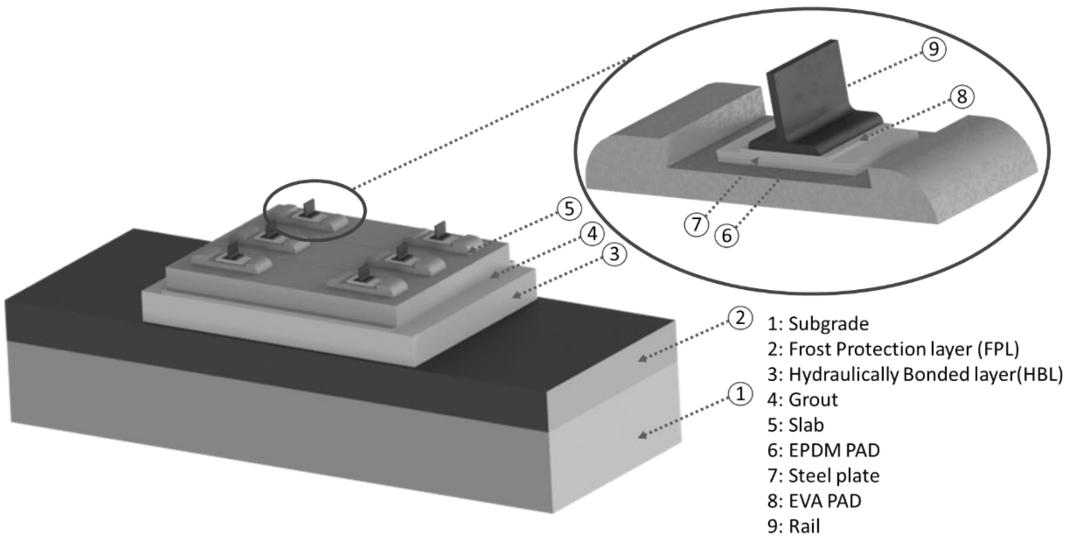

Fig. 1: Elements of the FE track model

2.2. Material properties

The mechanical properties of the soil, Hydraulically Bonded Layer (HBL), grout, and slab were obtained through an extensive literature review, allowing for the estimation of a statistical distribution for each property. For estimating the properties of the fastenings, an algorithm designed by Ferreño et al. [14]was used. This algorithm estimates the mechanical response of baseplates based on several factors: material, temperature, train load, frequency, clamping force (dependent on the condition of the fastening), and temperature (historical temperature data from two distinct regions, Seville and Moscow was used to analyse this variable). For a more detailed explanation of this methodology, please refer to reference [15].

2.3. Generation of synthetic data

Using the Monte Carlo method [9–11], 5400 random samples were generated to represent specific operating conditions. These samples correspond to 5400 distinct scenarios, which were then analyzed with the track FE model to determine the resulting displacements and accelerations.

2.4. ML Algorithms

The dataset for the ML analysis consists of 5400 samples, each with 27 features (19 inputs and 8 outputs). The inputs are divided into two categories:

- **Train type-specific variables**: train axle load, train speed, wheel passing frequency, load amplitude on rail fastening, and forces on inner and outer FE fastenings.
- **Track location-dependent variables**: city, rail pad material, toe load, city temperature, modulus of elasticity (sand, HBL, slab, rail pad), density (sand, HBL, slab, seat plate), and Poisson's ratio (sand, HBL, slab).

The mechanical behavior of the slab track is defined by eight outputs: acceleration and vertical displacement at the railhead and sleeper of two segments.

Data standardization is performed using the StandardScaler algorithm from Scikit-Learn. The dataset is then split into 75% for training (4049 samples) and 25% for testing (1350 samples). Six ML algorithms are used for regression modeling: Logistic Regression (LR), K-Nearest Neighbors (KNN), Decision Tree (DT), Random Forest (RF), Gradient Boosting (GB), and Artificial Neural Network (ANN). Model performance is evaluated using R2, RMSE, MAE, and MAPE.

ML algorithms can assess and compare the importance of each variable in predicting outcomes. This study uses impurity-based and permutation-based algorithms from Scikit-Learn to estimate feature importance. Partial Dependence Plots (PDPs) are employed to analyze the effect of each variable on the predicted values.

3 Results

3.1. Variable correlation

In Fig. 2, the correlation matrix is shown. This correlation matrix allows us to check for significant linear correlations between the different variables analyzed. From this matrix, it was possible to determine that there was a high correlation between some of the inputs, leading to the decision to eliminate some variables in the predictive model generation process. Additionally, it was possible to observe a high correlation between some inputs and some outputs. This implies that these variables are likely to be important for predicting the outputs.

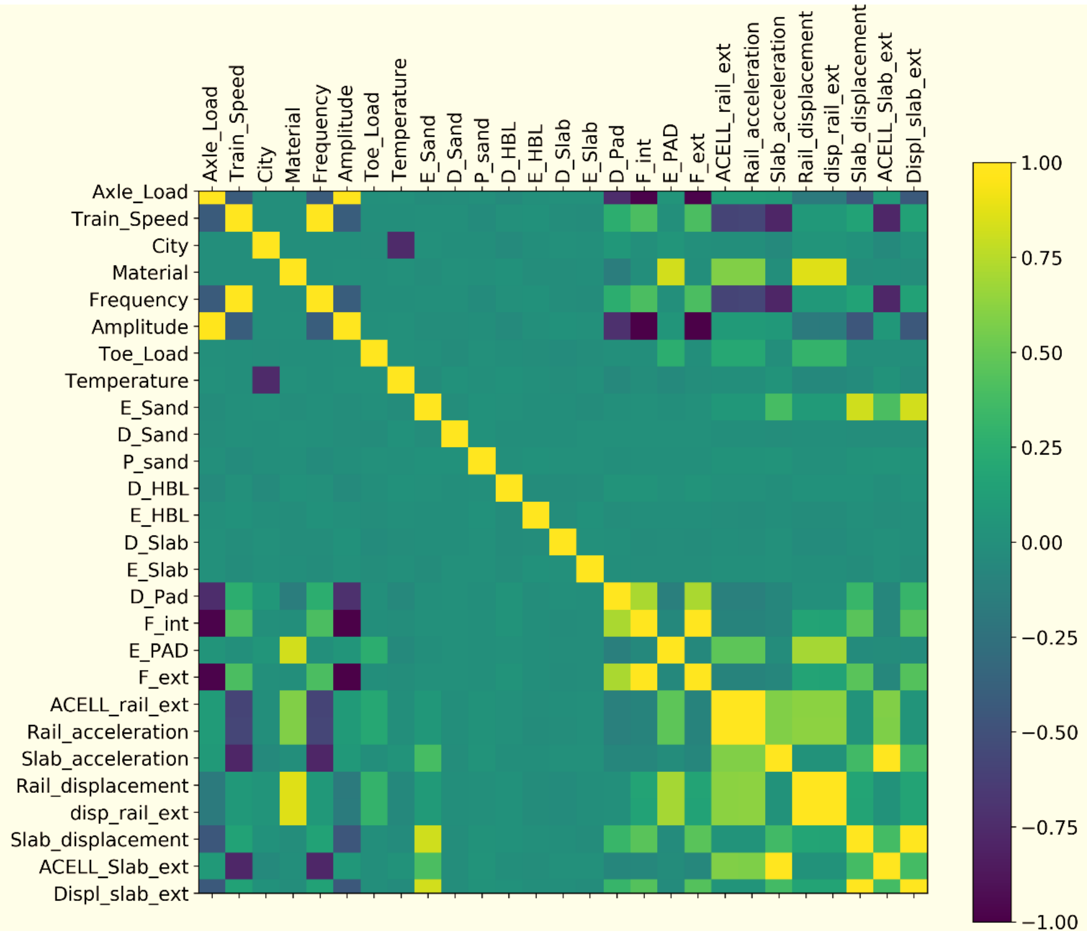

Fig. 2: Correlation matrix

In this initial phase of data inspection, another aspect analyzed is the distribution of the different outputs. Fig. 3 illustrates, as an example, the distribution of vertical displacement values of the slab, showing that this value can range between 0.02 and 0.14 mm depending on the operating conditions. The following sections define the procedure for accurately estimating the various outputs based on the values of the operating conditions.

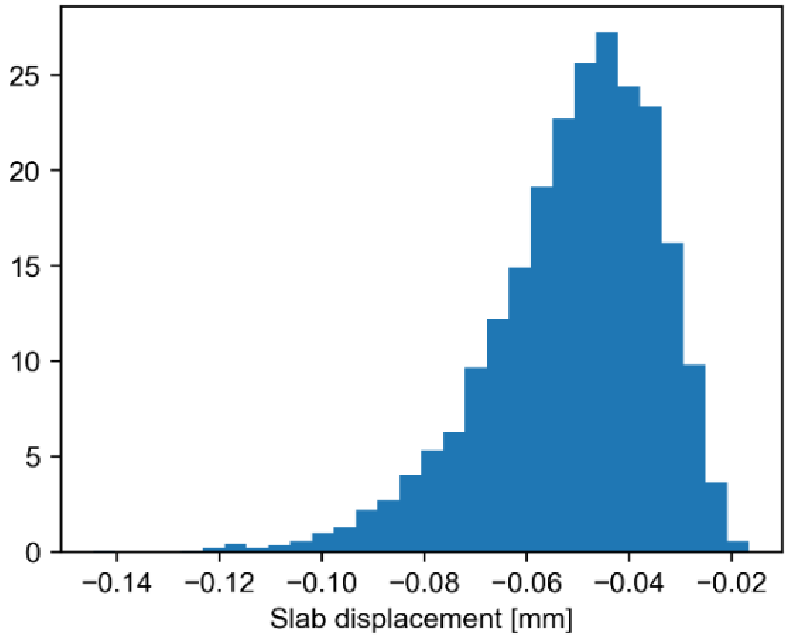

Fig. 3: Example of the slab displacement distribution.

3.2. Model calibration

In this section, the objective is to determine which of the trained models with optimized hyperparameters provides the most accurate results for data not used in training, i.e., the test data. Fig. 4 illustrates the correlation between experimental values and the values predicted by the model. Ideally, a 1:1 slope line would indicate perfect results. It can be observed that the model's predictions align very well with the experimental data, thus validating the model. This allows for proceeding to the next step, which involves extracting information from the models (a model has been selected for each of the outputs studied).

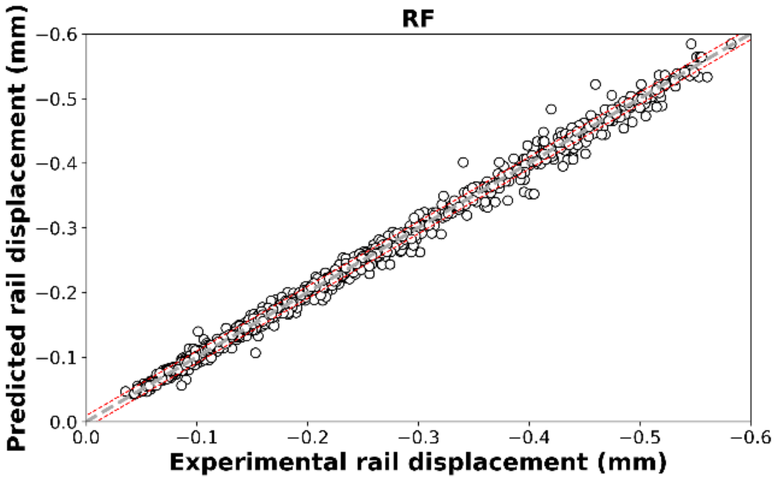

Fig. 4: Example of the correlation between the experimental results and the results obtained by the model

3.3. Feature importance

Once it was confirmed that a sufficiently accurate predictive model was available, the next step was to extract information from it. Firstly, the focus was on identifying the most relevant variables for each of the four case studies, one for each output: case study 1 (slab displacement), case study 2 (rail displacement), case study 3 (slab acceleration), and case study 4 (rail acceleration). Fig. 5 illustrates the importance of each variable for each of the case studies.

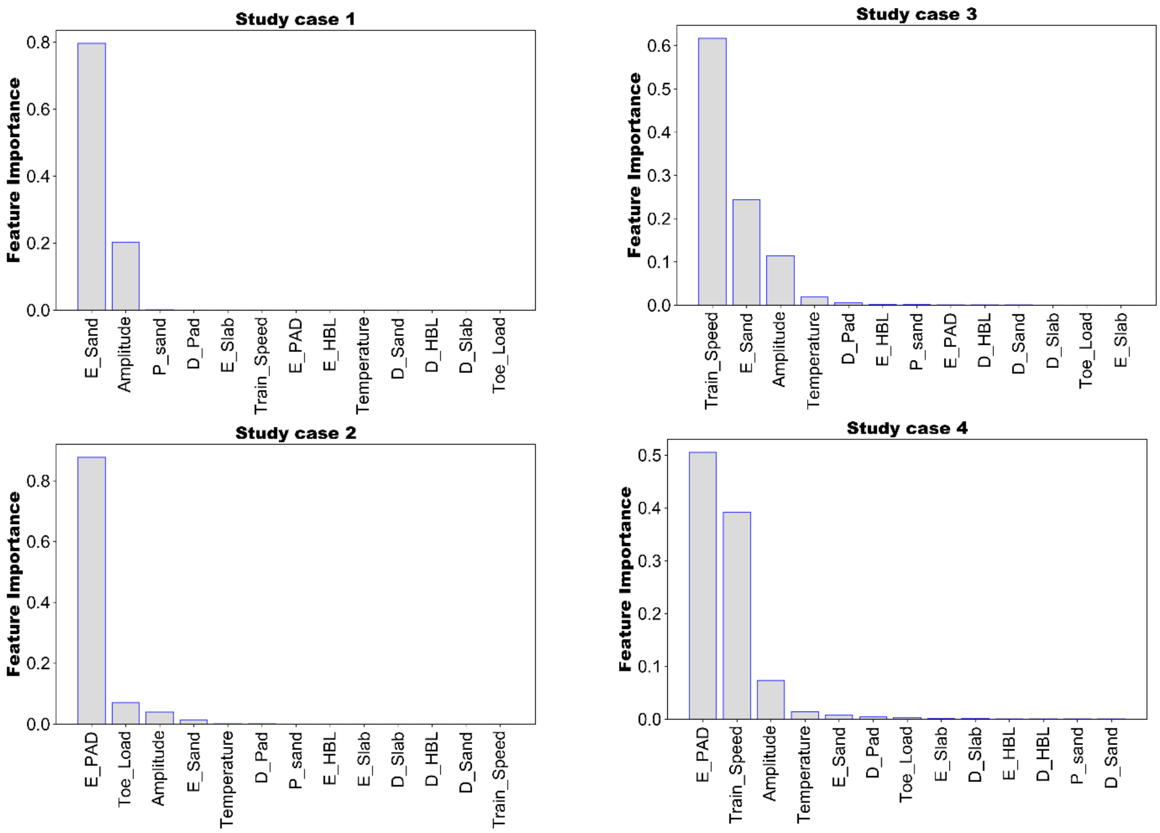

Fig. 5: Importance of each of the variables for each of the models

From Fig. 5, it is clear that for all four case studies, the behavior can be defined based on 2 to 4 parameters. These results indicate that the parameters governing the behavior of the slab and rail head are the same for both displacements and accelerations when the variable Train_speed is included. For the slab behavior, the most influential parameters are primarily E_sand and Amplitude. For the rail head behavior, the most significant parameters are E_pad, Amplitude, and Toe_Load.

3.4. Partial Dependence Plots

Apart from identifying the most important variables for predicting each of the outputs, it is also possible to estimate the influence of each variable across the studied range using partial dependence plots. Fig. 6 provides an example of these partial dependence plots.

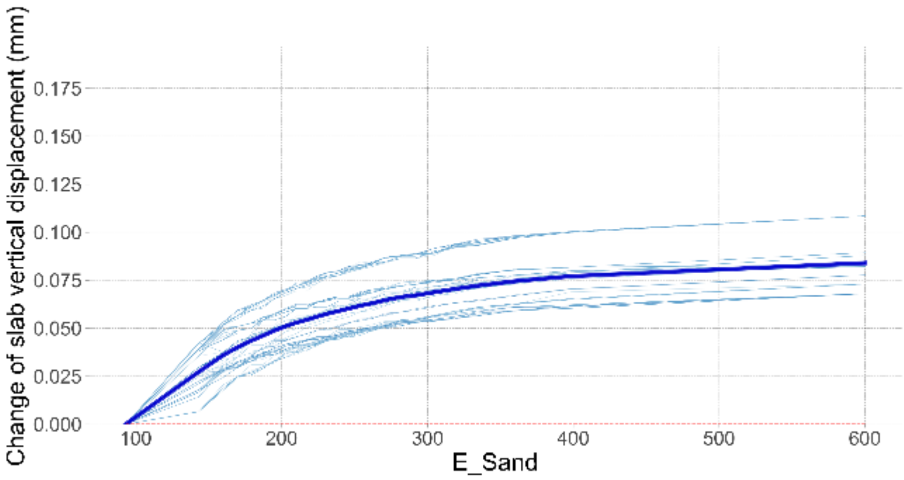

Fig. 6: Elements of the FE track model

From Fig. 6, it can be observed that variations in E_sand at lower values result in a significantly greater variation in the vertical displacement of the slab compared to when E_sand is higher. In fact, for E_sand values above 500, it appears to have no significant influence.

4 Conclusions and Contributions

In this study, 5400 simulations of a pre-calibrated FE track model were conducted, adjusting each of its 27 features within their typical variation ranges. From these simulations, various machine learning predictive algorithms were trained and validated, enhancing the understanding of complex track systems. This approach aimed to identify key parameters affecting track performance and to plan maintenance interventions based on the actual conditions of critical assets. The findings led to the following conclusions:

- Several predictive models were calibrated, with Random Forest (RF) performing the best, achieving an R2 consistently above 0.979 and a MAPE below 6.23%.
- The parameters most influencing the vertical displacements of the slab are E_Sand and Axle load.
- The parameters most affecting the vertical displacements of the rail are E_PAD, Toe_load, Axle load, and E_Sand.
- The parameters most impacting the vertical accelerations of the slab are Train_Speed, Axle load, and E_Sand.
- The parameters most influencing the vertical accelerations of the rail are E_PAD, Axle_Load, and Train_Speed.
- The influence of E_Sand and E_PAD is more significant when their values are lower.

The results provide valuable insights into the track features and operating conditions most relevant for designing new railway tracks or developing specific predictive maintenance strategies. Additionally, the study identifies parameters that should be adjusted if problems arise during the operation of an existing track section.

Acknowledgements

This work was supported by the “SMART-ALGORITHMS” project (PID2021-128031OB-I00) funded by the MCIN/AEI /10.13039/501100011033/ and by the European Regional Development Fund (ERDF). A way to build Europe. This work was supported by FCT, through IDMEC, under LAETA, project UIDB/50022/2020, and to the funding from the State Plan for Scientific,

References

1. To, W.M.; Lee, P.K.C.; Yu, B.T.W. Sustainability Assessment of an Urban Rail System – The Case of Hong Kong. *J Clean Prod* **2020**, *253*, doi:10.1016/j.jclepro.2020.119961.
2. He, G.; Mol, A.P.J.; Zhang, L.; Lu, Y. Environmental Risks of High-Speed Railway in China: Public Participation, Perception and Trust. *Environ Dev* **2015**, *14*, 37–52, doi:10.1016/j.envdev.2015.02.002.
3. Senaratne, S.; Mirza, O.; Dekruif, T.; Camille, C. Life Cycle Cost Analysis of Alternative Railway Track Support Material: A Case Study of the Sydney Harbour Bridge. *J Clean Prod* **2020**, *276*, doi:10.1016/j.jclepro.2020.124258.
4. Zhu, S.Y.; Cai, C.B.; Luo, Z.; Liao, Z.Q. A Frequency and Amplitude Dependent Model of Rail Pads for the Dynamic Analysis of Train-Track Interaction. *Sci China Technol Sci* **2014**, doi:10.1007/s11431-014-5686-y.
5. Fenander, Å.; Fenander, A.; Fenander, Å. Frequency Dependent Stiffness and Damping of Railpads. *Proc Inst Mech Eng F J Rail Rapid Transit* **1997**, *211*, 51–62, doi:10.1243/0954409971530897.
6. Wei, K.; Wang, F.; Wang, P.; Liu, Z.X.; Zhang, P. Effect of Temperature- and Frequency-Dependent Dynamic Properties of Rail Pads on High-Speed Vehicle–Track Coupled Vibrations. *Vehicle System Dynamics* **2017**, *55*, 351–370, doi:10.1080/00423114.2016.1267371.
7. Wei, K.; Zhang, P.; Wang, P.; Xiao, J.; Luo, Z. The Influence of Amplitude- and Frequency-Dependent Stiffness of Rail Pads on the Random Vibration of a Vehicle-Track Coupled System. *Hindawi Publishing Corporation Shock and Vibration*, doi:http://dx.doi.org/10.1155/2016/7674124.
8. Wei, K.; Zhao, Z.; Ren, J.; Ou, L.; Wang, P. High-Speed Vehicle–Slab Track Coupled Vibration Analysis of the Viscoelastic-Plastic Dynamic Properties of Rail Pads under Different Preloads and Temperatures. *Vehicle System Dynamics* **2019**, *3114*, doi:10.1080/00423114.2019.1673444.
9. Metropolis, N.; Ulam, S. The Monte Carlo Method. *J Am Stat Assoc* **1949**, *44*, 335–341.
10. Rubinstein, R.Y.; Kroese, D.P. *Simulation and the Monte Carlo Method*; John Wiley & Sons, 2016; Vol. 10; ISBN 1118632206.
11. Hénon, M. The Monte Carlo Method. In Proceedings of the International Astronomical Union Colloquium; Cambridge University Press, 1971; Vol. 10, pp. 151–167.

12. Sainz-Aja, J.A.; Pombo, J.; Tholken, D.; Isidro Carrascal; POLANCO, J.; Ferreno, D.; Casado, J.; Diego, S.; Pérez, A.; Filho, J.A.; et al. Dynamic Calibration of Slab Track Models for Railway Applications Using Full- Scale Testing. *Comput Struct Inpress*.
13. Thölken, D.; Abdalla Filho, J.E.; Pombo, J.; Sainz-Aja, J.; Carrascal, I.; Polanco, J.; Esen, A.; Laghrouche, O.; Woodward, P. Three-Dimensional Modelling of Slab-Track Systems Based on Dynamic Experimental Tests. *Transportation Geotechnics* **2021**, *31*, doi:10.1016/j.trgeo.2021.100663.
14. 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* **2021**, *151*, doi:10.1016/j.advengsoft.2020.102927.
15. Sainz-Aja, J.A.; Ferreño, D.; Pombo, J.; Carrascal, I.A.; Casado, J.; Diego, S.; Castro, J. Parametric Analysis of Railway Infrastructure for Improved Performance and Lower Life-Cycle Costs Using Machine Learning Techniques. *Advances in Engineering Software* **2023**, *175*, doi:10.1016/j.advengsoft.2022.103357.