PREDICTION OF GROUND-VIBRATIONS INDUCED BY RAILWAY TRAFFIC USING A MACHINE LEARNING MODEL

A. Ramos[1], A. Colaço[1], J. Lázaro[1,2], P. Alves Costa[1]

[1]CONSTRUCT – LESE, Faculty of Engineering, University of Porto, Portugal
[2]Amplitude Acoustics, Maia, Portugal
{aramos@fe.up.pt, aires@fe.up.pt, joao.lazaro@amplitude-ac.eu, pacosta@fe.up.pt}

Resumo

Nos últimos anos, tem havido uma procura crescente no avanço de técnicas numéricas sofisticadas para prever com precisão as vibrações no maciço causadas pelo tráfego ferroviário. Apesar da eficácia desses modelos na abordagem de tais fenómenos, a sua utilidade prática ao oferecer uma avaliação abrangente no âmbito de potenciais impactos em novos projetos ferroviários ou em projetos de reabilitação tem sido um desafio. Este artigo apresenta uma metodologia inovadora que utiliza a inteligência artificial para prever vibrações na superfície do maciço resultantes da passagem de comboios. A partir de uma base de dados gerada através da abordagem Método dos Elementos Finitos 2.5D com PML (FEM-PML), foram desenvolvidos diferentes modelos de *machine learning* (redes neurais artificiais e *random forest*) para prever essas vibrações de forma eficiente. Através de testes rigorosos considerando diferentes base de dados, o modelo demonstrou taxas de erro notavelmente baixas, ressaltando a sua eficácia na precisão preditiva em cenários do mundo real.

Palavras-chave: vibrações, 2.5D FEM-PML, modelo de *machine learning*.

Abstract

In recent years, there has been a growing demand for the advancement of sophisticated numerical techniques to accurately predict ground-borne vibrations caused by railway traffic. Despite the effectiveness of these models in addressing such phenomena, their practical utility in offering a comprehensive assessment of the potential impacts on new or updated railway projects has been challenging. This paper introduces a pioneering methodology that harnesses artificial intelligence for predicting ground-borne vibrations at the surface resulting from train passage. Leveraging a database generated through the 2.5D Finite Element Method with *PML* (FEM-PML) approach, different machine learning models (artificial neural networks and random forest) were developed to efficiently forecast these vibrations. Through rigorous testing considering different datasets, the model demonstrated remarkably low error rates, underscoring its efficacy for predictive accuracy in real-world scenarios.

Keywords: vibrations, 2.5D FEM-PML, machine learning model.

PACS no. 43.40.-r

1 Introduction

The effective management of densely populated urban areas, their mobility, and the imperative to address climate pose significant challenges to society. These challenges underscore the wide expansion of railway projects in modern cities. While the benefits - economic, social, and environmental – associated with rail transport are evident, their exploration rise to environmental concerns. This concern is primarily rooted in the generation and propagation of vibrations and noise, impacting the comfort and quality of the residents close to railway infrastructures.

In recent years, there has been a growing demand for the development of advanced numerical techniques spanning from the vibration source (vehicle–track interaction) to the receiver (building). These models typically consider essential aspects associated with the physical problem: i) the train's movement on the track as the source of vibrations; ii) the dispersion of energy in the ground; iii) the vibration field reaches nearby buildings giving rise to vibrations and noise [1-8].

While these models are effective in addressing these phenomena, predicting vibration levels demands high-standard numerical models that come with significant computational costs. This condition is enhanced when it is intended to have a general assessment of the potential impacts of new/updated railway projects. In such cases, there is a pressing need for the development of expedited prediction methods. These methods would enable the identification of cases requiring deeper analysis using advanced numerical models those that can be immediately promptly dismissed, offering advantages in terms of cost and time.

Indeed, in recent years, there have been different works in the scope of the prediction of ground-borne vibration using machine learning models [9, 10]. However, there are very few in the scope of railways [11] and also considering the vibrations at the receiver, the building.

Hence, the objective is to develop a prediction tool, empowered by an efficient and intelligent calculation engine based on surrogate modelling. This tool aims to facilitate an efficient assessment of ground-borne vibrations at the free-field surface and at the building resulting from railway operation.

2 Methodology

The developed methodology includes the train-track-ground response as well as the building response. Regarding the track-ground response, a database is generated considering the railway infrastructure modelling and multiple scenarios. Based on the database (generated using the Monte Carlo method), the vibration response on the ground surface is calculated using the 2.5D FEM-PML approach for each scenario. This database is then used to find the best machine learning model capable of predicting the vibration level at the ground generated by the passage of the train. The idea is to compute the free-field response in the form of 1/3 octave bands. The building's response is then computed using a 3D FEM model, taking into account the dynamic soil-structure interaction. Although this paper focused only on the train-track-ground system, Figure 1 presents a general scheme of the entire methodology.

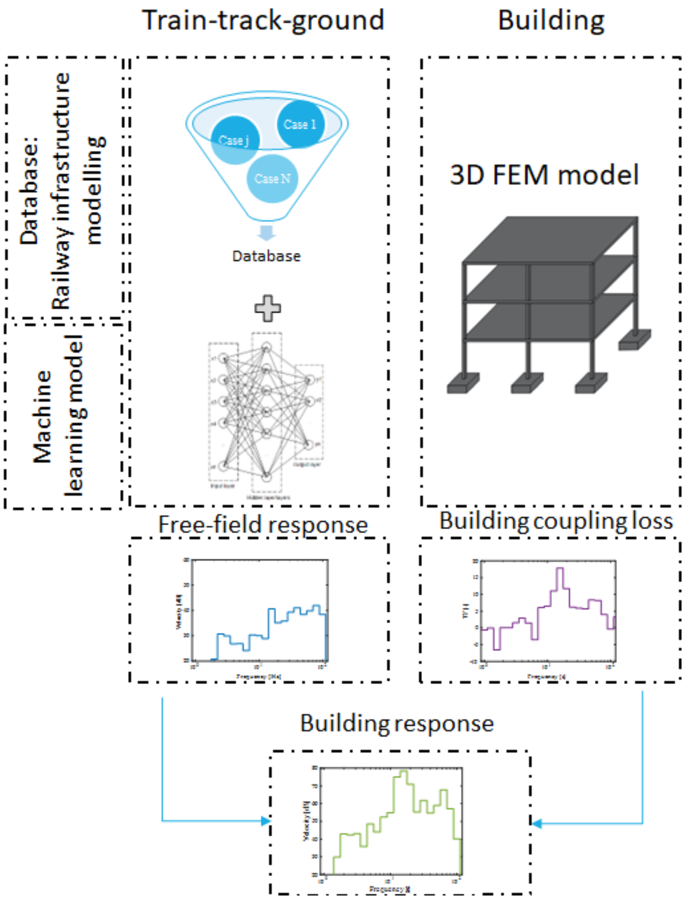


Figure 1 – General overview of the machine learning model and the developed methodology.

3 Surrogate model for the train-track-ground system

3.1 Generalities

The predictive tool for the response of the train-track-ground system is structured in two layers:

- Semi-analytical/numerical modelling of the subsystems of the railway infrastructure:

This initial layer involves detailed modelling and a large database is generated through extensive simulations. These simulations are designed to capture the most typical scenarios encountered in railway infrastructure projects. The simulations draw upon advanced numerical models previously developed by the authors [12-14], namely the 2.5D FEM-PML approach.

- Development of a machine learning model to have a ‘quick-to-compute’ engine.

For the second layer, instead of performing the physical modelling of the system (which requires a huge computational effort), the so-called machine learning models will be created to easily obtain results through a relatively simple and ‘quick to-compute’ emulator of results. It functions as an efficient and intelligent calculation engine, providing a balance between computational speed and accuracy.

Thus, advanced numerical models previously developed by the research group and experimentally validated are used in this task for the simulation of the track-ground system. Based on that, it was possible to compute the free-field response for different scenarios. The parametric studies performed attend to different characteristics of rolling stock, railway infrastructure, and geotechnical conditions. The wide nature of the analysis is evident in the consideration of several parametric studies for the development of the extensive database. The modelling parameters considered for each case are explained in detail. Through these efforts, the study aims to provide a reliable and efficient tool for the prediction of the track-ground system response considering different railway structures. In this paper, the focus is the slab track in a tunnel.

3.2 Generation of the database

In order to predict the vibrations induced in a homogeneous ground, a database was generated considering a certain typology of the railway structure: slab track in a tunnel. Thus, different variables were selected. The description of the variables of the database is presented in Table 1.

Regarding the database, certain soil characteristics are essential for accurate modeling. These variables include soil shear wave velocity (C_s), Poisson ratio of the ground (ν) and damping of the ground (ξ). Other variables were also included regarding the vehicle (type and train speed) and the condition of the track (irregularity class of the track) as well as the distance between the symmetric plane of the model and the points/nodes where the vibrations are calculated.

For the scenario of the slab track in a tunnel, the focus of the superstructure is the slab thickness as well as tunnel depth. The geometry of the tunnel was not considered as a variable. Regarding the properties of the support layer of the slab, they are considered constants, and, because of that, they were not included as variables.

Table 1 – Description of the variables and database of slab track in a tunnel

System	Variables	Description
Ground	C_s (m/s)	Discrete variable: Uniform distribution considering the values: 100 m/s, 150 m/s, 200 m/s, 300 m/s, 400 m/s and 800 m/s.
	Poisson ratio (-)	Discrete variable: Conditioned uniform: - when C_s is between 100 m/s and 400 m/s, the ν follows uniform with values 0.3 and 0.49; - when C_s is 800 m/s, ν has a unique value of 0.25.
	Damping of the soil (%)	Continuous variable: Conditioned normal distribution: - when C_s=100 m/s, the ξ follows a normal distribution with a mean value of 6% and a minimum and maximum between 4% and 8%; - when C_s=150 m/s, the ξ follows a normal distribution with a mean value of 4 % and a minimum and maximum between 3% and 6%; - when $C_s \geq 200$ m/s and < 800 m/s, the ξ follows a normal distribution with a mean value of 2.5% and a minimum and maximum between 3% and 4%;

		Uniform distribution with a unique value of 2% when C_s=800 m/s.
Vehicle	Type of train (-)	Discrete variable - Uniform distribution considering the trains ML90 and EuroTram.
	Train speed (km/h)	Discrete variable: Uniform distribution considering the values 60 km/h, 80 km/h, 100 km/h and 120 km/h.
Track	Slab thickness (m)	Discrete variable: The slab thickness follows a uniform distribution with the values 0.25 and 0.35.
	Tunnel depth (m)	Discrete variable: Uniform distribution considering the values 9 m; 12 m; 15 m; 18 m; 24 m 30 m.
	Class of the irregularities (-)	Discrete variable - Uniform distribution considering the values: class 3 and class 6.
Point response	Distance (m)	Discrete variable - Uniform distribution considering certain distances according to the mesh that varies between 0 m and 50 m.

For this analysis, the Monte Carlo Method was used to generate the databases. The Monte Carlo method is a computational algorithm that relies on repeated random sampling to obtain numerical results. Thus, this method involves generating random variables according to a specified probability distribution. These random variables represent possible values of uncertain inputs. The algorithm performs a large number of simulations or iterations. In each iteration, it calculates outcomes using a different set of random variables. After performing numerous simulations, the results are analyzed statistically to estimate the probability distribution of the outcome. This allows for the assessment of mean values, variances, confidence intervals, and other statistical measures.

Regarding the generation of the database, 7004 scenarios were generated.

In each scenario, the response of the track was recorded considering the vertical vibration velocity in 1/3 octave bands using the 2.5D FEM-PML approach.

3.3 Application of the data mining techniques

In this study, different algorithms were employed to predict the vertical vibration velocity in 1/3 octave bands: artificial neural network (ANN), and Random Forest (RF). Thus, the models were trained, validated and tested.

ANN is characterized by the input, hidden and output layers with connected neurons. As in the human brain, the existing nodes process and transmit input signals to the next nodes.

The random forest is part of the tree-based methods that employ ensembling. It consists of a number of decision trees each trained on a random bootstrapped sample of both observations and the features. The concept of bagging is employed by combining the random trees in the forest to decrease the variance of the predictions.

The ANN usually needs more data to achieve the same level of accuracy. And the random forest offers a performance gain when a certain amount of data is reached. This means that ANN usually benefits from large amounts of data and continuously improves the accuracy. Thus, the main difference is the nature of the algorithm since the ANN is part of the deep learning group of algorithms. Moreover, the number of hidden layers and the activate function were selected to find the best fit.

It is important to mention that the cross-validation method (k=10) was used to divide the database into test, train and validation.

3.4 Evaluation metrics

To evaluate the performance of the models, the Root Mean Square Error (RMSE) was employed, as well as the scatter plots and the R^2.

The RMSE allows to measure the accuracy of the model and the respective predictions, comparing with the observed (numerical) values. This metric measures the errors that are the differences between the predicted values (values predicted by the machine learning models) and the actual values of the variables (numerical values, in this case). The RMSE is obtained through the following expression:

$$RMSE = \sqrt{\frac{\sum(y_i - y_p)^2}{n}}, \tag{1}$$

where y_i corresponds to the actual value, y_p is the predicted value and n is the number of observations/rows of the database.

While the RMSE can measure numerically the error and its magnitude, the scatter plot is a visualization tool that allows to quickly understand how close the predicted and the observed values are, using the line $y=x$. The closer the values are to the line, the more accurate the prediction.

On the other hand, the Coefficient of Determination was also adopted and R^2 represents the proportion of variance in the dependent variable that can be explained by the independent variables:

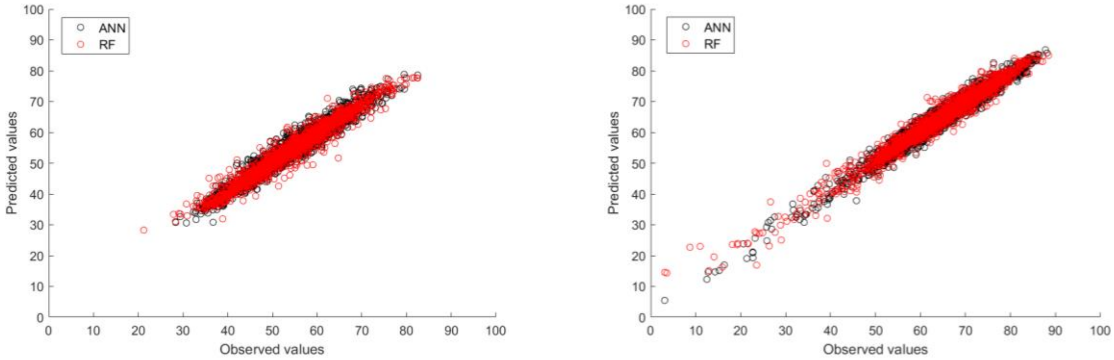
$$R^2 = 1 - \frac{\sum_i^n\left(x_i^{real} - x_i^{pred}\right)}{\sum_i^n\left(x_i^{real} - \overline{\overline{x^{real}}}\right)} \tag{2}$$

Where n is the total number of data points, x_i^{real} and x_i^{pred} are the observed and predicted values, respectively, and $\overline{\overline{x^{real}}}$ is the mean of the observed values.

3.5 Results

To fulfill the objectives initially stated, two different machine learning algorithms, and techniques, are applied to this problem. The expected outputs correspond to the predicted vibration in 1/3 octave bands. To estimate the vibrations, 16 models (mean values of the respective 1/3 octave bands) were generated, and each model corresponds to one frequency from 3.55 Hz and 112.2 Hz. Since the 2.5D FEM-PML approach solves the dynamic equations in the frequency domain and considers that the responses in this domain are independent of f_{i-1} and f_{i+1} (unlike in the time domain and implicit methods), the machine learning models follow the same path. This means that each model corresponds to one frequency and the models are independent.

For the frequency bands equal to 22.39 Hz-28.18 Hz and 56.23 Hz-70.79 Hz, the scatter plots considering the testing dataset are presented in Figure 2. These results are helpful to understand the dispersion of the predicted values when compared with the observed values. Figure 2 shows very good results, with very low values of dispersion. The results demonstrate higher dispersion considering the lower frequencies. Moreover, the Random Forest (RF) shows better results than the artificial network. In order to quantify the results, the RMSE and R^2 were evaluated and presented in Table 2.



a) b)

Figure 2 – Scatter plots: a) band 22.39 Hz – 28.18 Hz; b) band 56.23 Hz – 70.79 Hz

Table 2 - Evaluation metrics (RMSE and R^2) for the testing dataset considering different models

Model	Metrics	Slab track in a tunnel	
		ANN	RF
22.39 Hz-28.18 Hz	RMSE	2.065	1.889
	R^2	0.945	0.956
56.23 Hz-70.79 Hz	RMSE	1.908	1.884
	R^2	1.908	0.966

To show the ability of the model and confirm the very good results, a result of the velocity in 1/3 octave bands is shown considering a case of the testing dataset (Figure 3).

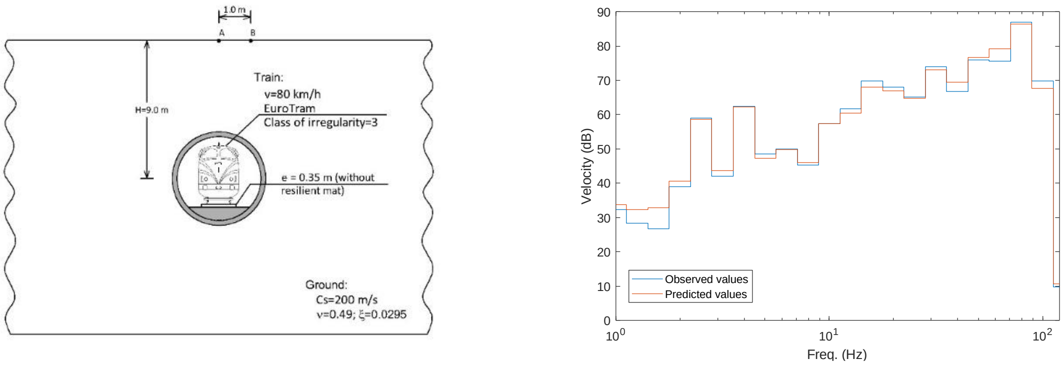


Figure 3 – Comparison between the predicted and observed values.

4 Application example

To validate the results, a different scenario from the database was defined. Considering the validation scenario defined in Table 3, the results obtained by the Random Forest models are compared with numerical results (observed values).

The results regarding the vibrations on the ground are depicted in Figure 4. It is possible to observe a positive correlation. It is necessary to highlight that each frequency band corresponds to one model of the Random Forest. These results were obtained considering that the vibrations are measured at the ground surface in the tunnel axle center. For this study, it was considered the irregularity class 6 and Eurotram at a speed of 100 km/h.

Table 3 - Validation example properties

C_s (m/s)	ν (–)	ξ (%)	Slab thickness (m)	Train speed (km/h)	Tunnel depth (m)	Irregularity class (-)	Train (-)	Distance (-)
500	0.35	2	0.3	100	9	6	EuroTam	0

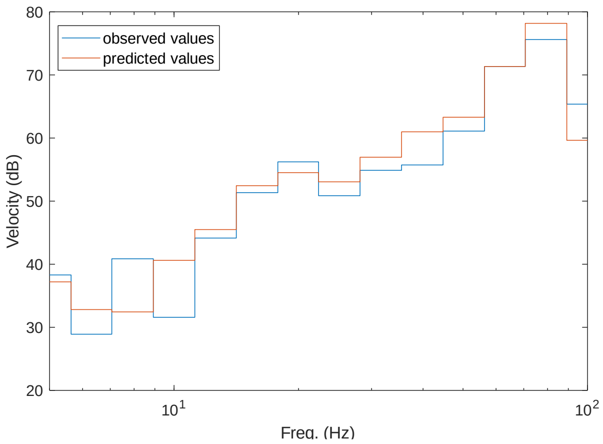


Figure 4 – 1/3 octave bands for the validation example.

Figure 4 shows that for frequencies between 10 Hz and 100 Hz, the predicted values are very close to the observed values and the differences between them are perfectly acceptable.

5 Conclusions

From the results achieved, it is possible to conclude that the presented methodology is able to be used in the prediction of the dynamic response of the track-ground system, allowing a computationally efficient assessment of free-field vibrations due to railway operation. In order to make the prediction tool able to be used by technicians who, although have basic knowledge about the topic, are not experts on railway noise and vibrations, it is intended the development of a user-friendly framework of wider practical application. The integration of the prediction tool with GIS (Geographical Information Systems) can also be useful for the visualization of problematic regions, allowing focus on those problematic areas, either through additional and more complex numerical studies or the design of mitigation measures.

Acknowledgements

This work is financially supported by: Base Funding - UIDB/04708/2020 with DOI 10.54499/UIDB/04708/2020 and Programmatic Funding - UIDP/04708/2020 with DOI 10.54499/UIDP/04708/2020 of the CONSTRUCT - Instituto de I&D em Estruturas e Construcoes - funded by national funds through the FCT/MCTES (PIDDAC); by national funds through the FCT/MCTES (PID- DAC), under the project PTDC/ECI-EGC/3352/2021 – IntRAIL – Intelligent prediction tool for ground-borne noise and vibrations induced by railway traffic; Individual Grant no.

2022.00898. CEECIND (Scientific Employment Stimulus - 5th Edition) pro- vided by FCT (10.54499/2022.00898.CEECIND/CP1733/CT0005).

References

[1] Lombaert, G., G. Degrande, and D. Clouteau, Numerical modelling of free field traffic-induced vibrations. Soil Dynamics and Earthquake Engineering, 2000. 19(7): p. 473-488.

[2] Sheng, X., C.J.C. Jones, and D.J. Thompson, A comparison of a theoretical model for quasi-statically and dynamically induced environmental vibration from trains with measurements. Journal of Sound and Vibration, 2003. 267(3): p. 621-635.

[3] Colaço, A., P.A. Costa, and D.P. Connolly, The influence of train properties on railway ground vibrations. Structure and Infrastructure Engineering, 2016. 12(5): p. 517-534.

[4] Galvín, P., et al., A 2.5D coupled FE-BE model for the prediction of railway induced vibrations. Soil Dynamics and Earthquake Engineering, 2010. 30(12): p. 1500-1512.

[5] François, S., et al., The influence of dynamic soil–structure interaction on traffic induced vibrations in buildings. Soil Dynamics and Earthquake Engineering, 2007. 27(7): p. 655-674.

[6] Lopes, P., et al., Vibrations inside buildings due to subway railway traffic. Experimental validation of a comprehensive prediction model. Science of The Total Environment, 2016. 568: p. 1333-1343.

[7] Colaço, A., et al., Prediction of Vibrations and Reradiated Noise Due to Railway Traffic: A Comprehensive Hybrid Model Based on a Finite Element Method and Method of Fundamental Solutions Approach. Journal of Vibration and Acoustics, 2017. 139(6).

[8] Nagy, A.B., et al., Prediction of interior noise in buildings generated by underground rail traffic. Journal of Sound and Vibration, 2006. 293(3-5): p. 680-690.

[9] Zhou, J., et al., Prediction of ground vibration induced by blasting operations through the use of the Bayesian Network and random forest models. Soil Dynamics and Earthquake Engineering, 2020. **139**: p. 106390.

[10] Xie, L., et al., Prediction of Ground Vibration Velocity Induced by Long Hole Blasting Using a Particle Swarm Optimization Algorithm. 2024. **14**(9): p. 3839.

[11] Liang, R., et al., *A novel efficient probabilistic prediction approach for train-induced ground vibrations based on transfer learning.* Journal of Vibration and Control, 2023. **30**: p. 107754632211487.

[12] Alves Costa, P., R. Calçada, and A. Silva Cardoso, Track-ground vibrations induced by railway traffic: In-situ measurements and validation of a 2.5D FEM-BEM model. Soil Dynamics and Earthquake Engineering, 2012. 32(1): p. 111-128.

[13] Lopes, P., et al., Numerical modeling of vibrations induced by railway traffic in tunnels: From the source to the nearby buildings. Soil Dynamics and Earthquake Engineering, 2014. 61-62: p. 269-285.

[14] Colaço, A., et al., Vibrations induced by railway traffic in buildings: Experimental validation of a sub-structuring methodology based on 2.5D FEM-MFS and 3D FEM. Engineering Structures, 2021. 240.