

# Rail4Future



Projekttitel:	<b>Resilient Digital Railway Systems to enhance performance</b>
Start Datum:	<b>01/04/2021</b>
Durchlaufzeit:	<b>42 Monate</b>
Projektnummer:	<b>882504</b>
Ausschreibung:	<b>8. Ausschreibung COMET Projekte 2019</b>

## D1.3.3 Report - Enabling large-scale simulations (M18)

Fälligkeitsdatum	31.12.2022
Einreichungsdatum	
Eingereicht von	

<b>Version</b>	<b>Datum</b>	<b>Bearbeitet von</b>	<b>Beschreibung</b>
0.1	22.02.23	Shiyang Zhou	Erstellung
0.2	06.03.23	Manfred Grafinger	Ergänzungen
1.0	06.03.23	Shiyang Zhou	Überarbeitung

# 1 Executive Summary

The holistic railway infrastructure digital twin (DT) platform is sophisticated and consists of a series of submodels (e.g., turnouts, tracks, vehicles, etc.) that are built through various methodologies and software. However, integrating these submodels into the DT platform is tremendously challenging due to considerable computational complexity, software and interface restrictions. In this report, we present a machine learning (ML) based surrogate modelling methodology for the submodel integration in the holistic railway infrastructure DT platform to enable the large scale simulations and illustrate the methodology through a case study. The main contribution of our work lies in the well-built ML-based surrogate modelling methodology for reducing the computation complexity and time of different submodels, which facilitates the unification and integration of different submodels. Furthermore, this approach can also be applied to other submodels and help to build the holistic railway DT platform collaboratively.



## 2 Table of Contents

- 1 EXECUTIVE SUMMARY ..... 3
- 3 DESIGN PROCESS ..... 5
  - 3.1 MBD MODEL OF THE RAILWAY VEHICLE-TRACK DYNAMICS ..... 5
  - 3.2 DEVELOPMENT PROCESS OF ML-BASED SURROGATE MODELING METHODOLOGY ..... 6
  - 3.3 CASE STUDY: BUILDING, TESTING AND VALIDATION ..... 7
  - 3.4 CONCLUSION AND FUTURE WORK ..... 8



## 3 Design Process

In this document, we will introduce an ML-based surrogate modeling methodology for replacing the railway vehicle-track MBD model as a use case. In subsection 3.1, we will briefly introduce the MBD model for railway vehicle-track dynamics analysis. In subsection 3.2, the general development process of ML-based surrogate modeling methodology will be demonstrated. Then we will analyze and validate the surrogate model with the dataset generated from the MBD Model and the measurement dataset collected from the "OBB-Infrastructure AG (Infrastructure Manager of Austrian Federal Railways) track system in subsection 3.3. In the end, the integration process of the sub-model into the holistic railway DT platform will be briefly discussed in subsection 3.4.

### 3.1 MBD Model of the railway vehicle-track dynamics

The interaction between the wheels and the track determines the dynamic behavior of railway vehicles. The geometry irregularities of the track can result in large dynamic forces on the vehicles, which can damage the vehicle wheels, leading to derailments in the worst cases. Therefore, understanding vehicle-track dynamics is vital for the RIS. The commercial software SimPACK, which can provide a reliable way to understand the dynamic forces acting on the track elements and the vehicles, is used to build the MBD model. SimPACK can generate different types of track irregularities using the Power Spectral Density (PSD) function, which describes the frequency-specific power distribution of the track excitation. The track excitation is deployed to the predefined track. Usually, the calculation of the PSD is based on the relevant regulations (commonly known as the ERRI B176 [15]), which are established from various parameters measured by the European railway operators. In the end, the irregularity data can be utilized as the input for the MBD model. The calculated forces on the wheelset can be extracted from the MBD Model as outputs.

### 3.2 Development process of ML-based surrogate modeling methodology

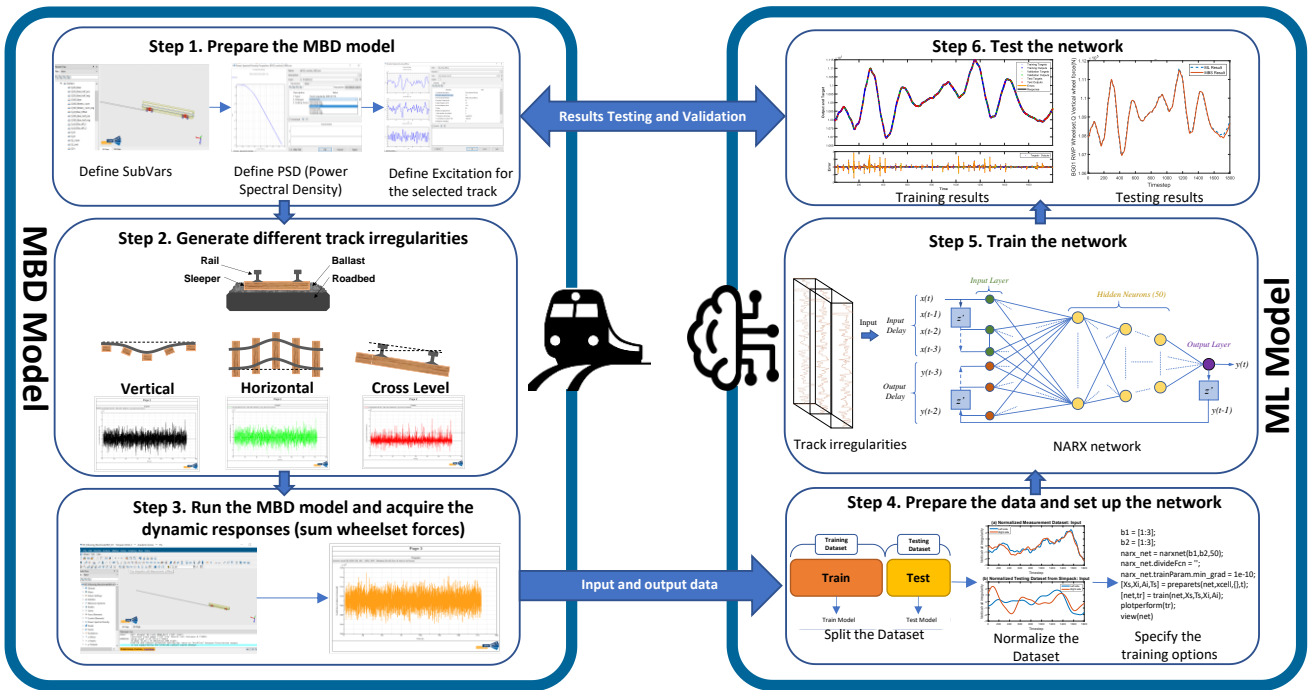


Fig. 1. Landscape of the ML-based surrogate modeling methodology

Consequently, establishing the above-mentioned ML-based surrogate modeling methodology is a significant part of the R4F research program. It will provide plenty of opportunities for future work, pushing sustainable development of the holistic large-scale railway DT system. Figure 1 presents an overview of the development process of the ML-based surrogate modeling methodology.

Step 1, the essential sub-variables, such as track length, velocity, and vehicle mass, are defined in the MBD model. After that, the PSD functions are created based on practical requirements (frequency and amplitude, etc.) in SimPACK. Accordingly, the track excitation is generated, and its related parameters are also defined (frequency interval type, start and end distance, etc.).

Step 2, different types of track irregularities are created by the corresponding track excitation based on the actual demands (e.g., vertical, lateral, cross-level track irregularities or any of their combinations). These irregularities work correspondingly together to determine the forces on the wheelset. Then, the MBD model is run with a solver implemented through a SODASRT integration method in SimPACK.



Step 3. After the simulation is done, the dynamic responses, i.e., the sum of wheelset forces, are collected from the post-processing software SimPACK-Post as output data. Different types of track irregularities as input data and the wheelset forces as output data are sent to the ML model for further ML training and testing.

Step 4. In the ML Model, all the input and output data from the MBD Model are first processed. To reduce the computation complexity, all the input data (i.e., different track irregularities) and the corresponding output data (i.e., wheelset forces) are equally divided into multiple datasets. Then, these datasets are randomly split into two groups: one group for network training and the other group for network testing. After that, the input datasets are normalized and standardized. The purpose of normalization is to make the data homogeneous across all different track irregularity dimensions, which helps to improve the data quality. Standardization means putting various features of all types of input data on the same scale. In other words, standardized data may be described as rescaling the characteristics of the input data so that their mean is 0 and the standard deviation is 1. After normalization and standardization, the datasets are fed to the ML Network for training and testing. The ML network we used in this study is a nonlinear autoregressive network with exogenous inputs (NARX), which is a recurrent dynamic neural architecture. It is one type of feed-forward time delay neural networks (TDNN) and is commonly used in time series data prediction.

Step 5. The landscape of the NARX model utilized in this paper is presented. The training algorithm applied in this model is the Levenberg-Marquardt algorithm. For regular networks in which the number of weights is under a few hundred, the Levenberg-Marquardt algorithm has the fastest convergence speed in function approximation problems [11]. Furthermore, with this algorithm, the squared errors and weights may be continuously reduced before reaching the optimum combination for the best-performing network so that overtraining can be avoided. During training, the number of neurons and delays needs to be determined from experimental iterations.

Step 6. The network's performance is evaluated using the normalized Root Mean Square Error (RMSE). After the training process is done, the testing datasets are applied to the well-trained ML model to evaluate the performance (Step 6). Finally, for validation, measurement data are applied to both the MBD Model and the ML Model as input. The results from both models are compared to validate the fidelity of the ML-based surrogate model.

### 3.3 Case study: building, testing and validation

We built an MBD model to simulate the railway vehicle-track dynamics, which is based on the Manchester Benchmarks Passenger Vehicle. The reliability and fidelity of surrogate models are of great importance. A well-built surrogate model should ensure the integrity

of the original resources and the reliability of the optimization, prediction and feasibility evaluation. Therefore, it is a vital task to validate the trained ML Model properly. In this study, the measurement data from the Austrian national railway track system are employed for the final validation of the fidelity of the ML-based surrogate model. Due to the measurement restrictions, only the vertical track irregularities are obtained. These measured vertical track irregularities are composed of two dimensions, i.e., irregularities from the left-side track and right-side track. In line with the measurement data, two-dimensional input datasets are created following Step 1-3 in Figure 1. The vertical wheelset forces were extracted from SimPACK-Post as the output data.

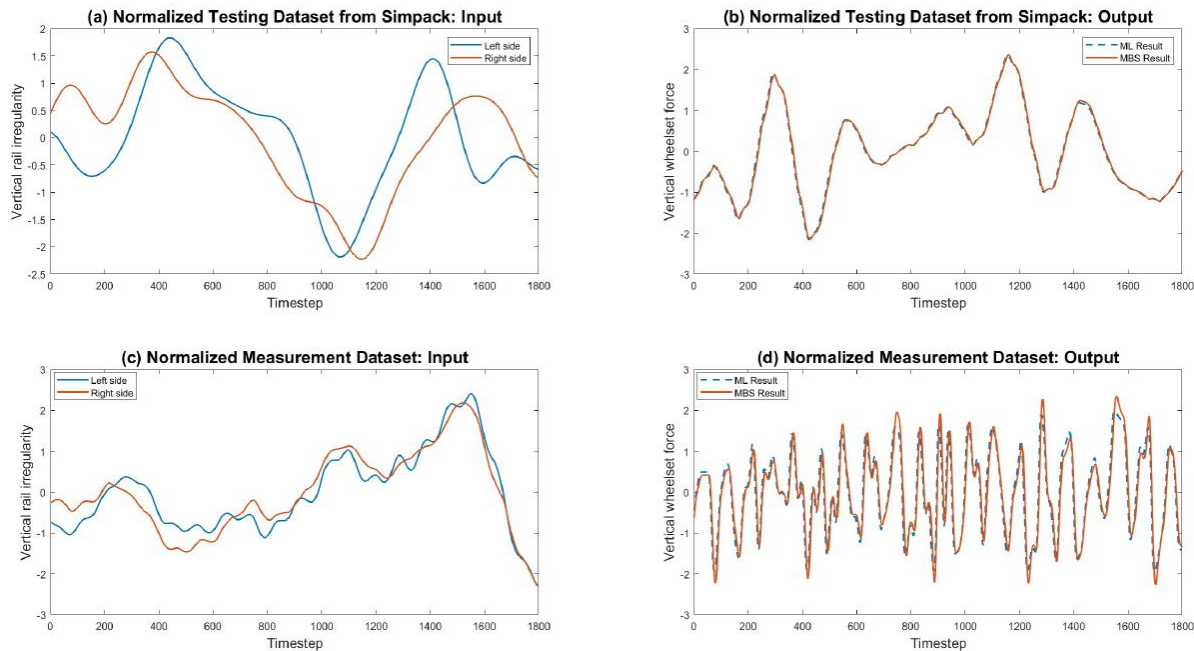


Fig. 2. Comparison between the output of MBD simulation (MBS) and ML surrogate model. The input datasets are either the testing dataset (upper two panels) or the measurement dataset (lower two panels).

All the presented data are normalized (with center 0 and standard deviation 1). Figure 2 (a) shows the normalized testing input dataset of the irregularities of the left and right tracks generated from MBD simulation for 1800 timesteps. Figure 2 (b) shows the normalized simulated output data from the MBD model and the predicted results of the ML-based surrogate model. It can be seen that the predicted results are in perfect coincidence with the simulation results from MBD Model. The corresponding normalized RMSE value between the two results is 0.48%, which indicates the ML-based surrogate model can perfectly replicate the MBD model with high precision. Then, we used the measurement datasets provided by the Austrian national railway track system to prove the fidelity of the ML surrogate model. Figure 2 (c) shows the irregularities of the left and

right tracks that are measured from a 25m-long track in a part of the Austrian railway system.

### 3.4 Conclusion and Future Work

In conclusion, the well-built ML-based surrogate model can make quick and precise predictions of the vertical dynamic responses based on different track irregularities. The surrogate model can replace the MBD simulation efficiently and be easily integrated into the holistic railway DT systems, as it has much less computational complexity than the traditional MBD simulation. The calculation efficiency is also greatly improved. For a 5 km long railway, it only takes about 8 seconds for the ML-based surrogate model to finish the calculation, a value that is three orders lower than the time needed for the MBD simulation (30 minutes). Furthermore, the proposed methodology can also enable the integration of different surrogate models into the holistic railway DT Platform in a fast and reliable way. The presented study outlines a broad spectrum of conceptual architecture for an entirely virtual validation platform for the integration of models and data of railway infrastructure subsystems. More use cases shall be integrated into this platform to validate its reliability and fidelity in future publications. Visualization research is conducted according to the principles of user-centered design and improved iteratively based on the properties of available data and the feedback of domain scientists.

Consequently, establishing the above-mentioned ML-based surrogate modeling methodology is a significant part of the R4F research program. It will provide plenty of opportunities for future work, pushing sustainable development of the holistic large-scale railway DT system.