Application of machine learning techniques to build digital twins for long train dynamics simulations

Application of machine learning techniques to build digital twins for long train dynamics simulations

N. Bosso[a]*, M. Magelli[a], R.Trinchero[b] and N. Zampieri[a]

[a]*Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Torino, Italy;*

[b]*Department of Electronics and Telecommunications, Politecnico di Torino, Torino, Italy*

*corresponding author: Nicola Bosso, e-mail: nicola.bosso@polito.it, Tel.: +390110906952, Fax.: +390110906999

Application of machine learning techniques to build digital twins for long train dynamics simulations

The paper shows the feasibility of building closed-form and fast-to-evaluate surrogate models via supervised kernel-based machine learning regressions behaving as digital twins for computationally expensive multibody simulations. The aforementioned surrogate models are adopted to predict the railway vehicle dynamics safety indexes defined in the international standards, depending on the wheel-rail forces, directly from the results of longitudinal train dynamics simulations. The digital twin models are trained with the outputs of Simpack multibody simulations of a reference freight wagon, to which the in-train forces calculated by an in-house MATLAB longitudinal train dynamics simulator are applied. Two machine learning techniques are considered: the support vector machine and the least-squares support vector machine regressions. Both techniques ensure a good accuracy even with a limited number of training samples. The derivation of the surrogate models can strongly enhance the modelling capabilities of common longitudinal train dynamics simulators, that cannot evaluate the wheel-rail contact forces. At the same time, the method shown in the paper allows to strongly reduce the total computational times, as the evaluation of the closed-form surrogate models allows to effectively estimate the safety indexes with no need for computationally demanding multibody simulations.

Keywords: longitudinal train dynamics; digital twin; machine learning; kernel-based regression; multibody simulation

1. Introduction

The use of machine learning (ML) techniques to build digital twins surrogate models (or metamodels) in the railway sector can find application in different research areas. Several examples of digital twins are used to better target the maintenance action on the infrastructure [1-5], or to predict track degradation using models derived from the analysis of experimental data collected in operating conditions [6-10]. At the same time, other applications of digital twin models concern the management of electrical power

[11-13].

ML digital twin models were also developed to predict the degradation of vehicle components, such as bearings [14,15] and suspensions [16-19], as well as to estimate adhesion or condition of wheel-rail rolling surfaces [20-23], based on accelerometric or other measurements carried out with monitoring systems installed on board the vehicle. Therefore, in this context, the development of digital twin models is an important support for the development of diagnostic systems.

Commonly, the approach adopted to build digital twin surrogate models relies on the use of experimental data measured on the vehicle to define an equivalent model describing the behaviour of the vehicle, so that the comparison of the model outputs with measurements carried out on the vehicle during operation allows to detect anomalies or to classify different types of faults. In this sense, one of the approaches often used in the railway sector, which can be considered a precursor of digital twins and of modern and more advanced ML techniques is the use of the Kalman filter to estimate the variations undergone by certain parameters of the system [24-29].

Therefore, it can be stated that traditionally digital twin techniques in the railway field are applied starting from an experimental database to generate a model of the system, which can be used for prediction or diagnostic purposes.

The present work proposes instead a different use of ML techniques, in which the database used to derive the surrogate models is built from the results of complex simulations of the vehicle dynamic behaviour. These simulations are carried out using a multibody (MB) model that allows to accurately simulate the railway vehicle dynamics, but with a high a computational effort. The results of the complex MB model are used to build surrogate models, acting as digital twins that are capable of predicting the results with very low calculation time. Examples of this strategy can be found in the

railway literature, dealing with the optimization of wheel and rail profiles [30-32] and with the development of reliable monitoring algorithms [33,34].

2. Method

The paper adopts the approach of using a database obtained from the outputs of complex MB simulations to build fast-to-evaluate kernel-based regressions allowing to accurately estimate the main safety indexes describing the running behaviour of railway vehicles, with no need for computationally expensive MB simulations. The surrogate models can then be integrated as an add-on module in a longitudinal train dynamic (LTD) simulator, so that the modelling capabilities of the LTD code can be enhanced, as common LTD simulators are not able to estimate the safety indexes, since they do not compute the values of the wheel-rail forces. In fact, common LTD simulators only consider the longitudinal degree of freedom (d.o.f) of the vehicles, and thus provide an overall train model made up of lumped masses, i.e., the vehicles, connected via nonlinear springs and dampers, featuring the typical hysteretic behaviour of coupling systems [35-37]. On the other hand, MB codes are provided with modules for the estimation of the wheel-rail contact forces [38-40], but they are typically used to model isolated vehicles, neglecting the in-train forces, or short groups of vehicles with an increase of the computational effort.

Large values of the in-train forces can increase the derailment risk, nonetheless, the EN14363 international standard for railway vehicle acceptance [41] does not explicitly consider the in-train forces as parameters to assess the safety of the vehicle running behaviour, as the criteria considered by the standards are derived from the forces acting at the wheel-rail interface. These criteria include the derailment safety ratio, i.e., the ratio between lateral and vertical forces Y/Q, the wheel unloading ratio DQ/Q and the sum of the lateral forces acting on the wheelset $\sum Y$.

Therefore, the construction of digital twin models, replacing the complex MB model of the railway vehicle, can strongly cut down the computational times for the evaluation of the safety indexes, which can be calculated directly from the results of an LTD simulation. In fact, the complex MB simulations are only launched in an initial stage to build the database for training and testing of the surrogate models. In the end, the proposed strategy allows to evaluate the safety indexes by means of the closed-form surrogate models, that require a negligible computational time.

The approach is considered as promising, since it is by far faster compared to the intermediate strategy proposed by Bosso et al. [42,43], which relies on a MB simulation of a whole train in which most of the vehicles are modelled in a simplified manner, only considering the longitudinal d.o.f, while a limited number of vehicles is modelled in detail. A similar intermediate strategy is directly available in the in the commercial multibody software Universal Mechanism [44,45]. At the same time, the proposed strategy is also faster compared to the method adopted by Krishna et al. [22,46], which relies on MB simulations of individual wagons carried out by applying the in-train and tractive/braking forces predetermined by an LTD simulation. In fact, even in case parallel-computing techniques are implemented, the computational time can still be huge if a long train with many vehicles is considered.

In conclusion, the goal of the paper is to show the feasibility of deriving digital twins able to replace the MB simulations in the calculation of the vehicle safety indexes with big benefits in terms of reduction of the computational times while ensuring a good accuracy. The proposed method can lead the way to several applications, such as using the digital twin surrogate models in numerical optimization schemes, for instance to optimize the distribution of the wagon mass along the train composition, so that safety during braking and traction operations can be maximized. The proposed strategy is

extremely suitable for the simulation of freight vehicles since freight wagons in a train usually adopt the same type of bogie (Y25 or three-piece) and a similar dynamic behaviour can be expected. The MB model used in the training phase can hence be modified by simply updating the carbody mass and thus the axle-load when different wagons are investigated.

The paper is organised as follows. The next section shows the general architecture of the computational framework built to derive the metamodels, with focus on the sampling strategy, on the LTD code and on the MB model of the reference vehicle. Then, a background on the ML techniques adopted in the paper is provided. Next, the main results of the numerical activity are presented and discussed, highlighting the statistical accuracy and the performance of the proposed surrogate models. Finally, the last section summarizes the conclusions of the work and suggests future developments of the methods proposed in the paper.

3. Computational framework

The strategy adopted to build the metamodels via ML techniques implies the following steps, see Figure 1:

(1) LTD simulations for the computation of the in-train forces on several different track layouts. The main simulation inputs are drawn using a Latin hypercube sampling (LHS) [47].

(2) MB simulation of individual wagons in the train composition, with application of the coupler forces computed by the LTD module, and calculation of the main output quantities of interest, namely the safety parameters defined in the international standards.

(3) Training, testing and validation of the metamodels with dedicated MATLAB routines.

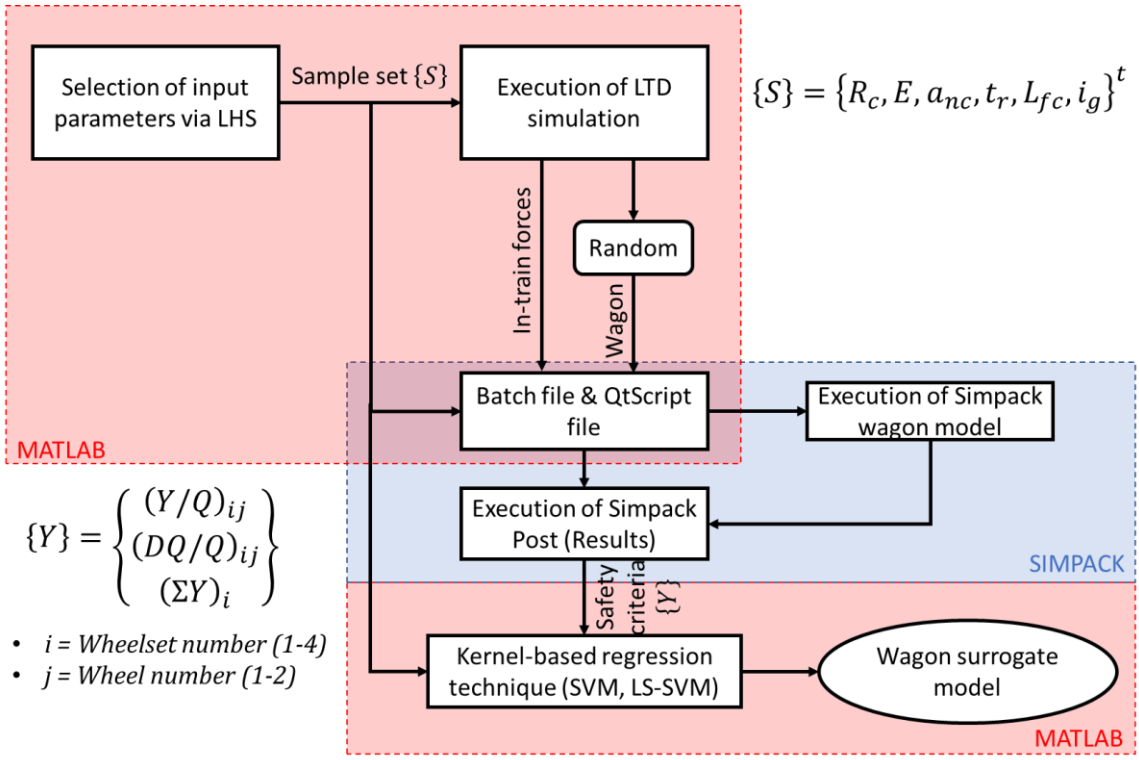

Figure 1: Sketch of the computational framework implemented to build the surrogate models.

Each LTD simulation is performed considering a train including a head locomotive followed by 18 wagons, running on a track featuring a single right-handed curve and a constant slope (different types of curves and slopes are considered). As the vehicle MB model is symmetric with respect to the lateral axis, the results for left-handed curves would be specular to the outputs corresponding to right-handed curves, hence only right-handed curves are investigated. For each simulation, the main track parameters are selected via LHS. Specifically, the inputs obtained with LHS include curve radius R_c, superelevation E, unbalanced acceleration a_{nc}, rail twist t_r, length of the full curve section L_{fc} and track slope i_g. Table 1 shows the main input parameters drawn

via LHS and their corresponding range. From this data, the speed of the leading locomotive $\dot{x}_{loco}$, which is kept constant during the LTD simulation, can be computed according to Equation 1, where g is gravity and d_{cp} is the nominal distance between the contact point positions (1.5 m for a standard gauge of 1435 mm). Only sample sets that ensure a speed below the vehicle speed limit of 120 km/h are considered, and other combinations obtained via LHS are discarded.

$$\dot{x}_{loco} = \sqrt{R_c \cdot \left(a_{nc} + g\frac{E}{d_{cp}}\right)} \quad (1)$$

Once the LTD simulation is carried out in MATLAB, a random wagon of the train is selected, and a detailed dynamic simulation is launched using the vehicle MB model implemented in Simpack, applying the in-train forces computed by the LTD code on the front and rear coupling systems. The MB simulations are run in batch from a MATLAB script, calling dedicated QtScript routines that manage the Simpack solver and post-processor. For each MB simulation, the output quantities of interest are collected to train the corresponding metamodels. More in detail, the output quantities include: (i) the derailment coefficient Y/Q of each wheel, (ii) the unloading ratio DQ/Q of each wheel and (iii) the sum of the wheel-lateral forces for each wheelset ∑Y, for a total of 20 outputs. All outputs are post-processed according to the prescriptions of the EN14363 standard, i.e., by applying a low-pass filter with cut-off frequency of 20 Hz and a sliding mean with window length of 2 m. For each simulated track, the maximum absolute value between percentiles 99.85% and 0.15% is collected for each output.

Table 1. Parameters drawn via LHS in MATLAB and corresponding range.

Parameter	**Symbol**	**Lower bound**	**Upper bound**
Curve radius	R_c	30 m	3000 m

Superelevation	E	0 m	0.16 m
Unbalanced acceleration	a_{nc}	0.1 m/s^2	0.6 m/s^2
Rail twist	t_r	0.1‰	4‰
Length of full curve	L_{fc}	30 m	500 m
Track slope	i_g	-15‰	15‰

For each combination of the parameters drawn via LHS, a random wagon is selected automatically for the MB simulation, to ensure that the metamodels obtained during the training stage can be used for each vehicle. Obviously, a higher accuracy could be achieved if different surrogate models were derived for each wagon, especially if the train is composed by vehicles with different characteristics (axle-load, bogie spacing, etc.), but in this work the vehicles in the train composition are identical. Nonetheless, with the definition of separate models for the different wagons, the total computational time required in the training phase might become huge. Furthermore, metamodels obtained for each single wagon could be less robust and would not be applicable to other possible configurations of the train consist. Therefore, as the aim of the present work is the demonstration of the feasibility of applying kernel-based regressions as a digital twin of MB simulations to enhance the modelling capabilities of LTD simulators, it is the authors' belief that a slight loss of accuracy of the surrogate models can be tolerated, as long as the general validity of the proposed approach is ensured.

After running the LTD and MB simulations and extracting the main outputs, the desired metamodels are obtained via kernel-based ML techniques for different sets of training samples. The inputs of the surrogate models do not exactly coincide with the parameters drawn via LHS and reported in Table 1. The selected surrogate model inputs

include: i) curve radius, ii) superelevation, iii) rail twist, iv) length of full curve section, v) mean vehicle speed while negotiating the curve, vi) median of front coupler force, vii) median of the rear coupler force, viii) maximum value of front coupler force and ix) maximum value of rear coupler force. Input parameters i)-iv) are directly obtained from the LHS, while inputs v)-ix) are computed during the post-processing stage of the LTD simulation. The track slope is not considered as an input parameter of the surrogate models, as the effect of the track slope is accounted for by the coupler forces. The median and maximum values of the coupler forces are computed after post-processing the results of the LTD simulation according to the same strategy adopted for the safety criteria. After the application of a low-pass filter and a smoothing mean, the median value is calculated as the 50-th percentile of the signals, while the maximum value is the maximum absolute value between the 99.85-th and 0.15-th percentiles of the signals.

3.1 LTD model of the train consist

The longitudinal dynamics of the train consist is simulated using the dedicated in-house LTDPoliTo code, which was validated in previous works [48-50] against the dataset of the international benchmark of LTD simulators [51,52]. The LTDPoliTo code only models the longitudinal d.o.f of the wagons, and the effect of grades and curves is considered by introducing extra resistant forces. The longitudinal dynamics of each vehicle is expressed by Equation 2, in which m_v is the vehicle mass, $\ddot{x}$ is the vehicle acceleration, $F_{c,f}$ is the force of the front coupler, $F_{c,r}$ is the force of the rear coupler, F_{ord} is the term accounting for ordinary resistances due to rolling and aerodynamic drag, F_g is the force due to track grade, F_{crv} is the resistant force due to track curvature and finally $F_{t/DB}$ is the tractive/braking force produced by the electric motors, which is clearly zero for non-powered vehicles. Obviously, the force on the front coupler of the leading locomotive and the force on the rear coupler of the tail-end wagon are set to

zero.

$$m_v\ddot{x} = F_{c,f} - F_{c,r} + F_{t/DB} - F_{ord} - F_g - F_{crv} \quad (2)$$

To simplify the training of the surrogate models, LTD simulations are run setting a constant speed of the leading locomotive. All tractive/braking efforts are provided by the leading locomotive in the train consist, which includes a locomotive followed by 18 four-axle wagons with an axle-load of 22.5 ton. A short European freight train is considered, to ensure low computational times in the training and testing of the surrogate models. Nonetheless, the approach can be easily extended to longer and heavier freight trains.

The tractive/braking force $F_{t/DB}$ during the LTD simulation is calculated by introducing a proportional controller as stated by Equation 3, in which K_{pos} is the term proportional to the position error, K_{velo} is a term proportional to the speed error, x_{loco} is the locomotive position, $\dot{x}_{loco}$ is the locomotive speed and finally x_{set} and $\dot{x}_{set}$ are the reference values of position and speed, respectively, in each time step.

$$F_{t/DB} = K_{pos}(x_{loco} - x_{set}) + K_{velo}(\dot{x}_{loco} - \dot{x}_{set}) \quad (3)$$

The effect of possible air brake forces is neglected in this work, since the force needed to balance the motion resistances is totally provided by the head locomotive in the form of a traction/dynamic braking force. In fact, in view of the goal of the present paper is to prove the feasibility of building surrogate models acting as digital twins of complex MB simulations to evaluate the major safety indexes defined in the international standards, the air brake forces are not considered for the sake of simplicity and to limit the number of parameters drawn via LHS. Furthermore, the LTDPoliTo code was developed and validated in the context of the international benchmark of LTD simulators [51,52],

which neglected the application of air brake forces. Nonetheless, since the derailment risk is higher during emergency braking operations, the simulation of the air brake forces will be addressed in future upgrades of the activity.

The force due to track grade F_g is computed according to Equation 4, in which i_g is the track slope and g is gravity, while several expressions exist to calculate ordinary and curve resistances [53,54]. In the present work, the ordinary resistant forces for the 18 wagons are calculated according to the expression suggested by Deutsche Bahn for full freight wagons, see Equation 5, while the ordinary resistant force acting on the locomotive is calculated according to the expression proposed by British railways, see Equation 6. On the other hand, the resistant force due to track curves is calculated using the expression given in the international benchmark, see Equation 7. Please note that in Equations 5-7, V is the vehicle speed in km/h, g is gravity, and R_c is the curve radius.

$$F_g = m_v g i_g \tag{4}$$

$$F_{ord} = \frac{m_v g}{1000} \cdot \left[1 + 0.02 \cdot \left(\frac{V}{10}\right)^2\right] \tag{5}$$

$$F_{ord} = \frac{m_v g}{1000} \cdot [4.587 + 0.0245V + 3.6697 \cdot 10^{-4} \cdot V^2] \tag{6}$$

$$F_{crv} = \frac{m_v}{1000} \cdot \frac{6116}{R_c} \tag{7}$$

Finally, the in-train forces are calculated in the LTDPoliTo code using a using a look-up table (LUT) approach. The LUTs store the loading and unloading curves of the coupling systems as a function of the coupler deflection. The paper deals with European freight trains, which are commonly equipped with the buffer-hook coupling element. Therefore, the LUTs for the computation of the in-train forces were built starting from the experimental characteristics of a single hook and a single buffer with a stroke of 105

mm [55]. Figure 2a shows the original experimental impedance characteristics of buffer and hook, which feature an initial preload and the typical hysteretic behaviour, while Figure 2b presents the final LUT characteristic obtained by considering the series connection of two hooks and the parallel connection of two pairs of buffers in series. Although other strategies exist for the calculation of the in-train forces [37], the LTDPoliTo code relies on the LUT approach to ensure fast computational times and to quickly adapt to different types of coupling elements.

During the LTD simulation, the force acting on each coupler F_c is calculated according to Equation 8, in which Δx is the coupler deflection, $\Delta\dot{x}$ is the deflection speed, F_L is the loading force, F_U is the unloading force and finally v_{th} is a threshold speed value which smooths the transition between the loading and unloading curves, according to the strategy proposed by Zhang et al. [56]. Therefore, in each time-step of the LTD simulation, the in-train force on each connection system is calculated by indexing the LUT according to the deflection on the coupling system. When the absolute value of the deflection speed is greater than the speed threshold value, a force equal to either the loading or the unloading force is applied, depending on the current operating conditions. On the other hand, when the absolute value of deflection speed is lower than the threshold value, a smoothing transition occurs, and the in-train force lies in-between the loading and unloading force values. This strategy is essential to avoid numerical instabilities that may be generated by abrupt step transitions between the loading and unloading curves. In the simulations shown in the paper, the threshold speed value was set equal to 5e-4 m/s.

$$F_c = \begin{cases} F_L(\Delta x), & |\Delta\dot{x}| > v_{th} \ \wedge Loading \\ F_U(\Delta x), & |\Delta\dot{x}| > v_{th} \ \wedge Unloading \\ \frac{F_L(\Delta x)+F_U(\Delta x)}{2} + \frac{|F_L(\Delta x)-F_U(\Delta x)|}{2} \cdot \frac{\Delta\dot{x}}{v_{th}}, & |\Delta\dot{x}| \leq v_{th} \end{cases} \quad (8)$$

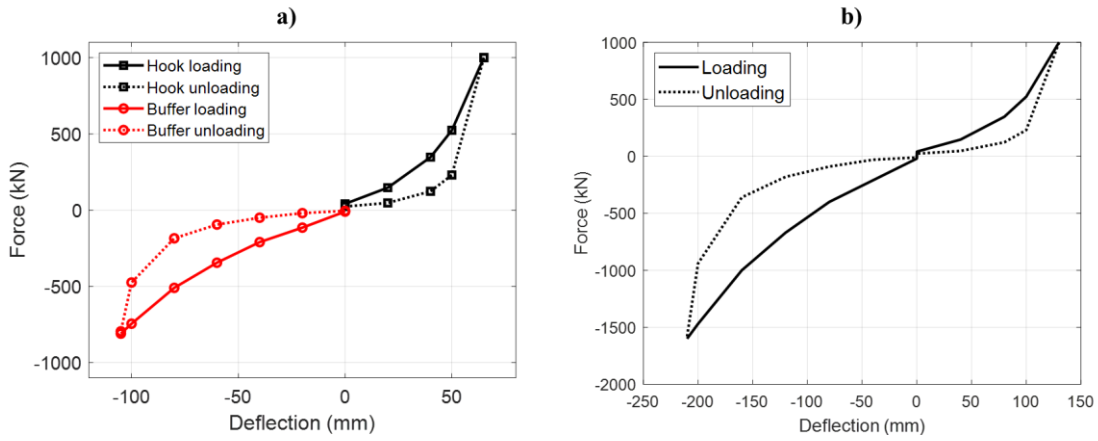

Figure 2: Original experimental characteristics of buffer and hook (a) and final LUT curves (b).

As already mentioned, the LTD simulations are run on simplified tracks featuring a constant grade and including a single curve with clothoid inlet and outlet transitions. Figure 3 shows the slope and curvature characteristics of the simulated track, highlighting the length of the main track sections. The length of the clothoid transitions L_{cl} is calculated from superelevation E and rail twist t_r, according to Equation 9, for each combination of the input parameters obtained from LHS.

$$L_{cl} = \frac{E}{t_r} \tag{9}$$

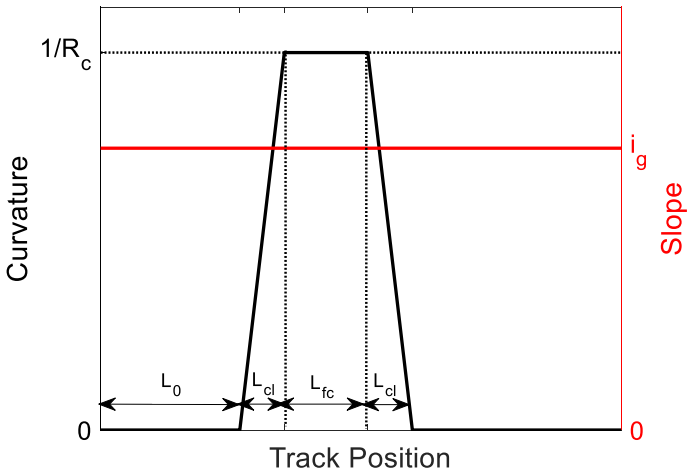

Figure 3: Curvature and slope characteristics of the simulated track layouts.

3.2 MB model of the railway wagon

The freight wagon considered in this paper is equipped with the Y25 bogie, which can be considered as the European standard freight bogie, provided with friction damping elements on both primary and secondary suspension levels. An additional non-linearity is provided by the primary suspension, which includes two groups of two concentric helical springs with different nominal length for each axle box. The longer spring is called the “tare spring” while the shorter one is the “load spring”, which acts only when a certain deflection of the suspension is overcome due to the payload.

The vehicle model includes 16 rigid bodies: the vehicle carbody which is divided into two half-structures, to consider the wagon torsional stiffness in railway twist, two bogie frames, four wheelsets and eight axle-boxes. The main inertial properties of each body are given in Table 2. The position of the centre of gravity is calculated according to a simplified three-dimensional geometric model of each body, and the joint position of each body is modified accordingly.

The three-dimensional (3D) wagon model is implemented in Simpack, by following the same modelling approach as in [57], which allows to reproduce in detail the strongly non-linear behaviour of both primary and secondary suspension stages, also accounting for the effect of the friction damping elements.

Table 2. Inertial properties of the bodies in the wagon MB model.

Body	**Mass (kg)**	**I_{xx} (kg·m^2)**	**I_{yy}(kg·m^2)**	**I_{zz} (kg·m^2)**
Half body frame	40710	59188	721517	725122
Axle-box	170	1.48	6.58	5.23
Bogie frame	1450	1103	715	1783
Wheelset	1080	903	119	903

To model the interaction with the adjacent wagons, two force elements, which simulate the actions of the buffer-hook coupling system, are applied to the vehicle body at a height of 1060 mm. Each force is applied between a marker defined on the wagon structure and a “Follow-Track” marker defined on the main inertial reference system. This marker type is a peculiar Simpack marker which is located and oriented according to the position and orientation of a specific joint. This strategy assumes that the in-train forces are always aligned to the ideal track. The force magnitude is calculated from an input function, storing the results of the LTD simulation, as a function of the vehicle position along the track. During this initial stage of the activity, the rotation of the couplers in curves and the consequent generation of lateral coupler forces were neglected for the sake of simplicity. Nonetheless, since the derailment risk is strongly related to the rise of lateral components of the in-train forces, that can cause coupler jack-knifing, the authors are planning to consider these effects in future upgrades of the work.

The simulated track is modelled without track irregularities, since track irregularities are random spectra that can modify the vehicle dynamics, depending on their frequency and amplitude, despite not being explicit vehicle parameters. In fact, the main goal of the numerical activity proposed in the present work is the investigation of the ideal behaviour of the model and the subsequent assessment of the validity of the proposed approach, with no influence of non-controllable variables, like irregularities.

A preliminary simulation was run to ensure that the speed profile obtained from the MB simulation launched on an isolated vehicle with application of the in-train forces is close to the speed predicted by the LTD model. In the simulated scenario, a train made up of 28 wagons and 1 head locomotive faces a right-handed curve of radius equal to 500 m and superelevation of 120 mm, see Figure 4a. This preliminary simulation was run on a longer and heavier train compared to the one used in the simulations launched for the training and testing of the surrogate models, which includes only 18 wagons. At the same time, in the preliminary simulation, a variable reference speed along the track was applied, while all simulations run to train and test the metamodel were launched with a constant reference speed signal. This was done to ensure that the speed profiles obtained in the LTD and MB simulations are close to each other even in operating conditions that are more complex with respect to the one considered in the simulations launched for the derivation of the surrogate models. Figure 4b shows the reference speed for the leading locomotive, variable along the track, the speed of the leading locomotive calculated by the MATLAB LTD simulator and the speed of a reference wagon in the middle of the train composition obtained from the LTD and MB simulations. As noticeable from the plot, a good agreement exists between the results of the LTD simulation and the speed calculated by the Simpack

model. Slight differences can be due to discrepancies in the curve resistance calculated by the two models.

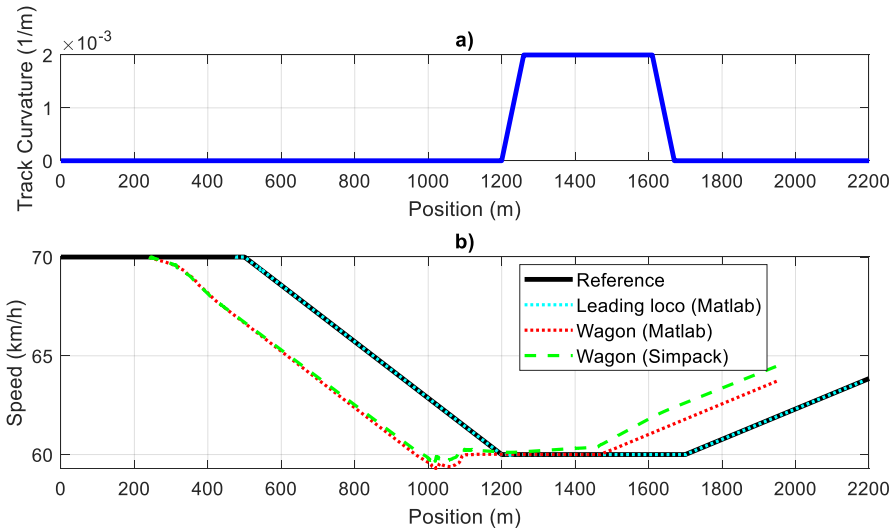

Figure 4: Preliminary simulation: track curvature (a) and comparison between the wagon speed calculated by the LTD simulator and the MB model (b).

3.3 Background on surrogate models and supervised Machine Learning regression techniques

The aim of this section is twofold. First, it discusses the benefits of adopting digital twins built via supervised ML regression techniques within the simulation scenarios described before. Moreover, it provides a brief overview of the mathematical background of two advanced kernel-based regression techniques: the Support Vector Machine (SVM) and the Least-Squares Support Vector Machine (LS-SVM) regressions, respectively, also highlighting their key features and stressing their advantages w.r.t other state-of-the-art techniques.

3.3.1 Digital twins and surrogate models

Usually, design space exploration tasks of complex structures are performed by relying on the results of computational expensive simulation experiments via the so-called

"computational model". The computational model can be seen as the most accurate synthetic description of the actual behaviour of the system under modelling, able to virtually predict the output quantities of interest for several configurations of the system parameters. Computational models can have several levels of complexity, going from a simple closed-form approximation to the most complicated full-body one. Although full-body simulations are clearly more accurate, they can be extremely expensive in terms of computational costs, thus making design space exploration tasks extremely inefficient when many simulation runs are needed. In the current work, the computational model is represented by the vehicle MB model.

In the above scenario, digital twins based on surrogate models can be seen as an effective and efficient alternative to plain simulations based on the computational model [58]. The surrogate model is "a model of a model", which provides a closed-form and fast-to-evaluate mathematical relationship able to accurately mimic the complex and usually non-linear parametric behaviour of the computational model. Such models are constructed via generic regression or interpolation techniques by using data obtained from a limited set of simulations run with the expensive computational model. Such data used to train the surrogate model is referred to as training set. The samples belonging to the training set must be carefully chosen to explore the input space as much as possible [47,58]. Once the surrogate model has been trained, it can be used to predict the parametric behaviour of the output of interest for a generic configuration of the system parameters, and therefore it can be suitably embedded within statistical analysis and optimization schemes [58-60].

This paper focuses on two different supervised ML kernel-based regressions, namely the SVM [61] and LS-SVM [62] regressions. Compared to classical parametric model formulations based on basis expansion (e.g., polynomial expansion), in which the

model complexity (i.e., the number of basis functions and model unknows) grows exponentially with the number of input parameters [60], the SVM and LS-SVM regressions allow to obtain non-parametric models, for which the model complexity is independent from the number of input parameters. This allows to overcome curse-of-dimensionality and possible overfitting issues [60-62]. Furthermore, w.r.t to other non-parametric techniques, like neural networks (NNs), kernel-based regressions are less "data hungry", since the model training for such techniques relies on the solution of a convex optimization problem via low-complexity algorithm. Therefore, kernel-based regressions can be considered good candidates for constructing accurate and efficient digital twins of the expensive computational model from a small set of training samples [60-62].

The remaining of this section briefly presents a quick overview of the mathematical background of such techniques with specific emphasis on their application to surrogate model construction.

3.3.2 SVM and LS-SVM regression: mathematical background

Let us consider the problem of approximating the output y ∈ ℝ of a generic non-linear vector-valued function $\mathcal{M}$: $\mathbb{R}^d \rightarrow \mathbb{R}$ as a function of its parameters $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^d$, such as:

$$y = \mathcal{M}(\mathbf{x}) \tag{10}$$

where y is the output of the system under modelling and $\mathbf{x} = [x_1, \dots, x_d]^T \in \mathbb{R}^{\mathrm{d}}$ is a vector collecting the system parameters. For instance, within the present case study, the variable y can be interpreted as one of the 20 main outputs corresponding to the safety indexes defined by the international standards and the vector **x** represents the

configurations of the main metamodels inputs declared in the previous subsections, which are obtained from the LHS and from the results of the LTD simulations.

Kernel-based regressions can be suitably adopted to build a fast-to-evaluate surrogate model $\widetilde{\mathcal{M}}$ able to approximate the actual nonlinear input-output behaviour of the function $\mathcal{M}(\mathbf{x})$ for any $\mathbf{x} \in \mathcal{X}$. The surrogate model is constructed via the training set $\mathcal{D} = \{(\mathbf{x}_l, \mathrm{y}_l)\}_{l=1}^{L}$, computed via a set of simulations with the computational model, such as $\mathrm{y_l} \approx \mathcal{M}(\mathbf{x}_l)$.

Like standard regression techniques, such as the OLS, the primal space formulation of the SVM and LS-SVM regressions is based on a simple linear expansion of basis functions [61,62], see Equation 11, in which $\langle\cdot,\cdot\rangle$ is the inner product, $\mathbf{\Phi}(\mathbf{x}) = [\phi_1(\mathbf{x}), \dots, \phi_D(\mathbf{x})]^T$ is a vector collecting the basis functions $\phi_{\mathrm{i}}(\mathbf{x})$, which maps the parameter space of dimension d into the corresponding feature space of dimension D (i.e., $\mathbf{\Phi}(\cdot)$: $\mathbb{R}^d \rightarrow \mathbb{R}^D$,), b is the bias term and $\mathbf{w} = [w_1, \dots, w_D]^T$ is a vector collecting the basis functions.

$$y(\mathbf{x}) \approx \mathcal{M}_{(LS-)SVM}(\mathbf{x}) = \sum_{d=1}^{D} w_d \phi_{\mathrm{d}}(\mathbf{x}) = \langle \mathbf{w}, \mathbf{\Phi}(\mathbf{x}) \rangle + b \qquad (11)$$

For SVM and LS-SVM regressions, the unknowns $\mathbf{w}$ and b are estimated by solving the optimization problem, known as the empirical risk minimization [61], see Equation 12, where $\ell(\cdot)$ is a loss function providing the model "error" computed over the training samples, $\|\boldsymbol{w}\|_{L_2}^2$ is the Tikhonov regularizer [63], which is intended to keep the model as flat as possible, and finally γ is a regression hyperparameter tuned during the training phase to reduce the model variance and to prevent overfitting.

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|_{L_2}^2 + \gamma \sum_{l=1}^{L} \ell(y_l, \langle \mathbf{w}, \mathbf{\Phi}(\mathbf{x}_l) \rangle + b) \qquad (12)$$

In the empirical risk minimization in Equation 12, the main difference between the SVM and LS-SVM regression comes from the adopted loss-function $\ell(\cdot)$. Specifically, the LS-SVM regression uses a traditional squared loss function, whilst the SVM regression uses the linear ε-insensitive loss, which adds a penalization equal to the excess model $\ell 1$-error w.r.t. ε [61]. For such reason the region $[-\varepsilon, +\varepsilon]$ is called ε-insensitive zone. Further details can be found in the literature [60-62].

It is important to remark that the primal space formulation in Equation 11 provides a parametric model in which the number of unknown coefficients (i.e., the cardinality of the vector $\boldsymbol{w}$) is equal to the number of basis functions D. On the other hand, both regression techniques allow an equivalent dual space formulation, as stated by Equation 13, in which the vector $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_L]^{\mathrm{T}}$ collects a new set of regression unknowns for the above dual space formulation and $k(\cdot\,,\cdot)\colon \mathbb{R}^{\mathrm{D}} \times \mathbb{R}^{\mathrm{D}} \to \mathbb{R}$ is the so-called kernel function.

$$y \approx \mathcal{M}_{(LS)SVM}(\mathbf{x}) = \sum_{l=1}^{L} \alpha_l \, k(\mathbf{x}, \mathbf{x}_l) + b \tag{13}$$

Please note that in the above formulation of the SVM and LS-SVM regression is a non-parametric model in which the number of unknown coefficients in the vector $\boldsymbol{\alpha}$ is equal to the number of samples L. Therefore, the surrogate model built using such regressions turns out to be completely independent from the model complexity and from the number of input parameters. In this paper, a radial basis function (RBF) kernel function is used to train both the SVM and the LS-SVM regressions, see Equation 14, in which σ is the kernel hyperparameter defining the width of the RBF [61]. For the case of the dual-space formulation of the LS-SVM regression, the vector of coefficients $\boldsymbol{\alpha}$ and the bias coefficient b can be obtained in a closed-form via the solution of a linear system of equations [62]. On the other hand, for the SVM regression, the optimal set of unknown

coefficients in the vector $\boldsymbol{\alpha}$ is computed numerically as the solution of a quadratic convex optimization problem [61]. It is important to remark that thanks to the properties of the ε-insensitive loss function, this optimization leads to a sparse solution (i.e., most coefficients α_l are equal to zero), thus providing a cheap model implementation in terms of memory resources.

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x} - \mathbf{x}')\|^2\right) \tag{14}$$

4. Results

This section investigates the performance of the SVM and LS-SVM regressions for the construction of a digital twin able to inexpensively predict the 20 outputs of interest described in the previous section, which correspond to the main safety criteria defined in the EN14363 standard. Due to the inherently scalar nature of such regression techniques, the problem can be reformulated to be equivalent to train 20 scalar-output independent surrogate models (i.e., one for each output dimension). For the generic i-th output, a set of training samples $\mathcal{D}^{(i)} = \left\{\left(\mathbf{x}_l, \mathrm{y}_l^{(i)}\right)\right\}_{l=1}^{L}$ has been generated from the results of parametric simulations with the computational model. Both the output and input samples have been standardized so that they are centred around the mean with a unit standard deviation. The surrogate models constructed via the SVM and the LS-SVM regressions are trained in MATLAB via built-in functions [64] and the LS-SVMlab v1.8 toolbox [65], respectively. The hyperparameters of the SVM regression are optimized using a Bayesian optimizer with 5-fold cross-validation (CV) error [63] for a maximum of 30 iterations, while the LS-SVM technique uses the same Bayesian optimizer with a leave-one-out CV error for a maximum of 25 optimization steps.

The accuracy of the surrogate models trained via the SVM and LS-SVM regressions is investigated for an increased number of training samples. Specifically, Figure 5 shows the mean and the standard deviation of the R^2 score computed from the 20 outputs under modelling for 500 test samples by considering an increasing number of training samples L = {48,98,198,296,344,496}. The curves show a monotonic increase of the test R^2 score, until its mean value gets very close to 1, w.r.t the size of the training set up to L=344, as well as a reduction of the standard deviation of the R^2 score. This means that the proposed approaches can learn the complex non-linear behaviour of the outputs w.r.t the input parameters. On the other hand, the flatness of the learning curve after L=344 can be easily explained by considering the strategy adopted to run the MB simulation after each LTD simulation. In fact, as already mentioned, after each LTD simulation is over, a random wagon in the train consist is selected for the next MB simulation. This choice was made to derive surrogate models that can be adopted for all wagons in the train consist, as considering different wagons allows to better investigate the input space of the in-train forces. Higher accuracy could be obtained if different metamodels for the different wagons in the train consist were trained, but this would drastically increase the time required to run the simulations needed to train the surrogate models.

From the results, the surrogate models trained via the LS-SVM regression turn out to be slightly more accurate than the corresponding ones constructed via the SVM regression. On the other hand, the SVM regression leads to a compact surrogate model with an average sparsity around 12%. Regarding the time costs, the training time goes from 144 s for L =48 to 620 s for L=496 for the SVM regression, and from 14 s to 181 s for the LS-SVM regression, respectively. After the training phase, the longest time required to evaluate the surrogate models on 500 test configurations is still less than 1 s.

On the other hand, when the safety parameters were estimated from the outputs of a dynamic simulation, the total computational time on all 500 test configurations was equal to approximately 23 hours and 30 minutes (almost 1 day), on a notebook with 16 GB of internal memory and equipped with an Intel Core i7-8750H CPU (6 cores, 12 threads, base frequency: 2.20 GHz). Therefore, the availability of ML surrogate models for the prediction of safety parameters leads to a huge reduction of the computational times, since computationally expensive MB simulations are avoided. Furthermore, since the surrogate models are closed-form models, they slightly affect the total computational time required by an LTD simulation. In fact, each stage of the LTD simulation requires the following amounts of the total time: 77% for the numerical integration of the LTD equations, 17% for basic post-processing operations on the simulation outputs, 5% for additional post-processing operations needed to evaluate the surrogate models and finally less than 2% for the evaluation of the surrogate models. Therefore, as the modules for the numerical integration of the LTD equations and for the basic post-processing operations were already required in the original version of the LTD code, the add-ons for the evaluation of the safety parameters with the dedicated surrogate models requires an increase of the LTD simulation computational time lower than 8%.

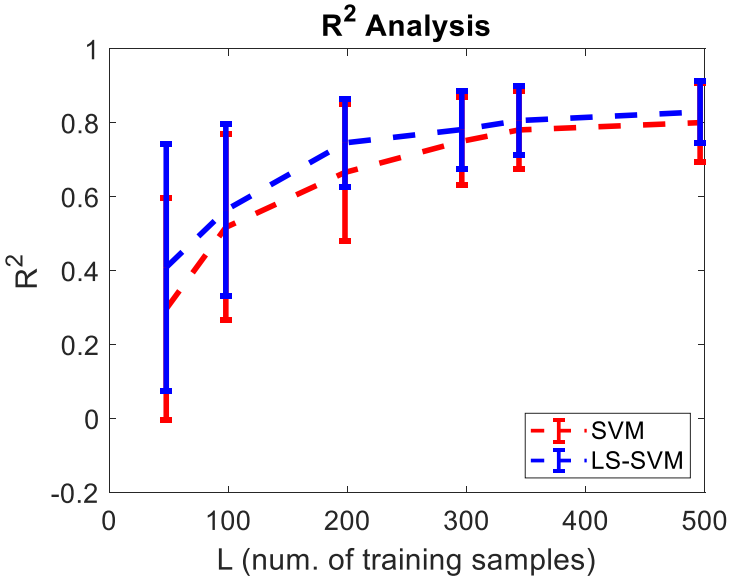

Figure 5: R^2 score of the surrogate models for different numbers of training samples.

To better highlight the accuracy of the surrogate models, Table 3 shows the mean value of the R^2 score for the three safety criteria, calculated on the test samples for a number of training samples equal to 496. As visible in Table 3, the surrogate models built for the Y/Q parameter feature the highest accuracy, while the lowest accuracy is related to the surrogate models for the ΣY parameter.

Table 3. Mean value of the R^2 score for the safety criteria.

Output	**SVM**	**LS-SVM**
Y/Q	0.889	0.901
DQ/Q	0.780	0.804
ΣY	0.663	0.737

As a further comparison Figure 6 shows the scatter plots for some of the key output variables, i.e., the Y/Q and DQ/Q parameters calculated on the wheels of leading wheelset of the leading bogie, and the ΣY computed on the leading wheelsets of the two

bogies. For the sake of ease of understanding, the parameters are plotted in absolute value. The wheelsets are numbered in ascending order starting from the leading wheelset, while wheels are numbered using two digits: the first digit refers to the wheelset, while the second digit is related to the side of the wheel (1 for right wheels and 2 for left wheels).

The scatter plots emphasize the remarkable correlation between the predictions of the proposed surrogate models and the ones obtained via the computational model collected in the test set, as the samples are very close to the dashed lines, which represent a perfect agreement between the surrogate models and the reference samples. As previously mentioned, the ΣY parameter features the lowest R^2 score. A possible explanation can be given by observing the corresponding scatter plot, which shows a limited variation of this output, due to the absence of irregularities on the track and to the relatively large values of curve radius. In future works, the effect of the track irregularities will be investigated, but in the present work they are neglected, since they are random spectra that may change the vehicle behaviour, despite not being a physical vehicle parameter.

In conclusion, the proposed approach is promising, as the surrogate models can give good predictions of the chosen outputs. Nonetheless, the sample sets are selected from large ranges of the input parameters, as a general scenario in terms of track layout and train consist is considered. Obviously, the model structure, as well as its input parameters, could be further improved to account for the effect of additional parameters, such as the wagon payload, and to better suit the characteristic of a specific real-life track and vehicle scenario.

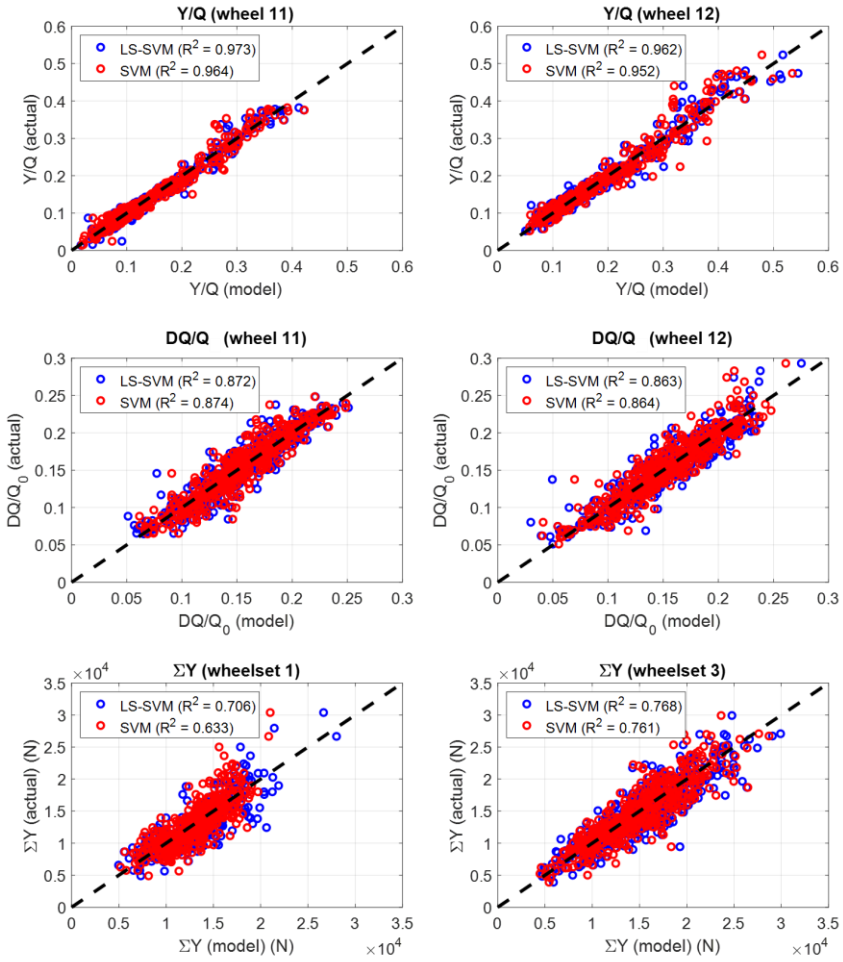

Figure 6: Scatter plots for key output parameters.

5. Conclusions

The paper shows an innovative simulation technique to improve the capability of LTD simulators. Specifically, the approach is based on the development of surrogate models of the wagons composing the train, acting as a digital twin of the MB simulation. These models are obtained from kernel-based regressions and trained using the outputs of complex and computationally expensive MB simulations of the wagon. This approach allows to drastically reduce the computational time required to calculate the safety indexes of all the wagons of the train. In fact, when the surrogate models are available, there is no need to launch long-lasting MB simulations. It is the authors' belief that the proposed method is mainly suited to freight vehicles, which are equipped with standard

bogies, hence different types of wagons can be modelled by simply changing the properties of the carbody.

The results are promising and demonstrate the general feasibility of the proposed approach. The surrogate models, in fact, replace with good accuracy the detailed MB model for all the considered simulation scenarios. The strategy leads the way to LTD simulators able to evaluate not only the in-train forces but also the main quantities that describe the safety and running behaviour of each wagon in the train consist. On a test sample including 500 configurations, the surrogate models evaluated the safety parameters in less than 1 s, while the estimation via MB simulation would take approximately 1 day. Furthermore, the additional routines for the evaluation of the surrogate models have a limited impact on the computational times of the original LTD code, with an increase in the total computational time less than 8%.

Since the aim of the present work is to present the simulation techniques and to evaluate the nominal behaviour of the vehicle, irregularities are neglected, but the effect of irregularities will be investigated in future works. Clearly, it is important to consider a wide range of irregularity spectra to develop surrogate models able to describe the behaviour of the vehicle under different track conditions. Moreover, air brake forces were neglected at this initial stage of the activity to limit the number of parameters to be drawn via LHS. Nonetheless, since often derailments occurs during emergency braking operations, whereby large in-train compressive forces arise, such conditions will be addressed in future upgrades of the work, that will also consider jack-knifing effects and the generation of lateral in-train forces during curve negotiation. Furthermore, future works will deal with the development of surrogate models able to evaluate the transient response of the vehicle, that would improve the modelling capabilities and the accuracy of the derived digital twin.

References

1. Dimitrova E, Tomov S. Digital Twins: An Advanced technology for Railways Maintenance Transformation. 2021 13th Electrical Engineering Faculty Conference (BulEF); 8-11 September; Varna, Bulgaria: IEEE; 2021.
2. Yang J, Sun Y, Cao Y, et al. Predictive Maintenance for Switch Machine Based on Digital Twins. Information. 2021;12(11):485.
3. Doubell GC, Kruger K, Basson AH, et al. The Potential for Digital Twin Applications in Railway Infrastructure Management. In: Pinto JOP, Kimpara MLM, Reis RR, et al., editors. 15th WCEAM Proceedings; 15-18 August; Campo Grande, Brazil: Springer Cham; 2022. p. 241-249.
4. Futai MM, Bittencourt TN, Santos RR, et al. Utilization of Digital Twins for Bridge Inspection, Monitoring and Maintenance. In: Pellegrino C, Faleschini F, Zanini MA, et al., editors. Proceedings of the 1st Conference of the European Association on Quality Control of Bridges and Structures; 29 August-1 September; Padova, Italy: Springer Cham; 2022. p. 166-173.
5. Pappaterra MJ. A Literature Review for the Application of Artificial Intelligence in the Maintenance of Railway Operations with an Emphasis on Data. In: Marrone S, De Sanctis M, Kocsis I, et al., editors. Dependable Computing – EDCC 2022 Workshops; 12 September; Zaragoza, Spain: Springer Cham; 2022. p. 59-75.
6. Kampczyk A, Dybeł K. The Fundamental Approach of the Digital Twin Application in Railway Turnouts with Innovative Monitoring of Weather Conditions. Sensors. 2021;21(17):5757.
7. Meixedo A, Santos J, Ribeiro D, et al. Data-driven approach for detection of structural changes using train-induced dynamic responses. International Conference on Structural Health Monitoring of Intelligent Infrastructure: Transferring Research into Practice, SHMII; 30 June - 2 July; Porto, Portugal 2021. p. 441-448.
8. Hamarat M, Papaelias M, Kaewunruen S. Fatigue damage assessment of complex railway turnout crossings via Peridynamics-based digital twin. Scientific reports. 2022 Aug 23;12(1):14377.
9. Jiang R, Wang W, Xie Y, et al. Research and Design of Infrastructure Monitoring Platform of Intelligent High Speed Railway. 2022 IEEE 6th Information Technology and Mechatronics Engineering Conference (ITOEC); 4-6 March; Chongqing, China: IEEE; 2022. p. 2096-2099.
10. Kulkarni R, De Rosa A, Qazizadeh A, et al. Monitoring of Alignment Level (AL)and Cross Level (CL) Track Geometry Irregularities from Onboard Vehicle Dynamics Measurements Using Probabilistic Fault Classifier. In: Orlova A, Cole D, editors. Advances in Dynamics of Vehicles on Roads and Tracks II; 17-19 August; Saint Petersburg, Russia: Springer Cham; 2022. p. 479-487.
11. Wang Y, Zhang G, Chen R, et al. Analysis of digital twin application of urban rail power supply system for energy saving. 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI); 15 July - 15 August; Beijing, China: IEEE; 2021. p. 29-32.
12. Zheng X, Xu B, Zhao W, et al. Modeling Method and Application of Metro Power Supply System Based on Digital Twin. 2021 IEEE 2nd China International Youth Conference on Electrical Engineering (CIYCEE); 15-17 December; Chengdu, China: IEEE; 2021. p. 1-7.
13. Chen L, Yu Y, Gong Z, et al. Multi-scale Modeling and Simulation Method of Urban Rail Traction Power Supply System for Digital Twin. In: Jia L, Qin Y,

Liang J, et al., editors. Proceedings of the 5th International Conference on Electrical Engineering and Information Technologies for Rail Transportation (EITRT) 2021; 21-23 October; Singapore: Springer Singapore; 2022. p. 666-675.
14. Li Y, Zhang W, Xiong Q, et al. A Novel Fault Diagnosis Model for Bearing of Railway Vehicles Using Vibration Signals Based on Symmetric Alpha-Stable Distribution Feature Extraction. Shock and Vibration. 2016.
15. Kreuzer M, Schmidt A, Kellermann W. Novel features for the detection of bearing faults in railway vehicles. The 50th International Congress and Exposition on Noise Control Engineering; 1-5 August; Online Conference 2021. p. 3833-3844.
16. Jesussek M, Ellermann K. Fault detection and isolation for a full-scale railway vehicle suspension with multiple Kalman filters. Vehicle System Dynamics. 2014;52(12):1695-1715.
17. Wei X, Jia L, Guo K, et al. On fault isolation for rail vehicle suspension systems. Vehicle System Dynamics. 2014;52(6):847-873.
18. Karlsson H, Qazizadeh A, Stichel S, et al. Condition Monitoring of Rail Vehicle Suspension Elements: A Machine Learning Approach. In: Klomp M, Bruzelius F, Nielsen J, et al., editors. Advances in Dynamics of Vehicles on Roads and Tracks; 12-16 August 2019; Gothenburg, Sweden: Springer Cham; 2020. p. 119-127.
19. Hesser DF, 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. 2022;164.
20. Zhao Y, Liang B, Iwnicki S. Friction coefficient estimation using an unscented Kalman filter. Vehicle System Dynamics. 2014;52(sup1):220-234.
21. Bernal E, Spiryagin M, Vollebregt E, et al. Prediction of rail surface damage in locomotive traction operations using laboratory-field measured and calibrated data. Engineering Failure Analysis. 2022;135:106165.
22. Krishna VV, Wu Q, Hossein-Nia S, et al. Long freight trains & long-term rail surface damage – a systems perspective. Vehicle System Dynamics. 2022:1-24.
23. Spiryagin M, Wu Q, Polach O, et al. Problems, assumptions and solutions in locomotive design, traction and operational studies. Railway Engineering Science. 2022;30(3):265-288.
24. Leonardi F, Alfi S, Bruni S. An application of Kalman filtering estimation to the condition monitoring of the geometric quality of the railway track. In: Zobory I, editor. Proceedings of the Mini Conference on Vehicle System Dynamics, Identification and Anomalies; 7-9 November; Budapest, Hungary. Budapest: Budapest University of Technology and Economics; 2016. p. 163-172.
25. Yang Z, Lu ZG, Wang XC, et al. Wheelset states estimation using unscented Kalman filter. In: Spiryagin M, Gordon T, Cole C, et al., editors. The Dynamics of Vehicles on Roads and Tracks; 14-18 August 2017; Rockhampton, Queensland, Australia. London, UK: CRC Press; 2018. p. 1071-1076.
26. Deng W, Huang J, Zhang Q. A Method of the Locomotive Speed Estimation Based on Fuzzy Logic and Extended Kalman Filter. 15th IEEE Conference on Industrial Electronics and Applications (ICIEA); 9-13 November; Kristiansand, Norway: IEEE; 2020. p. 1353-1358.
27. Ulum Z, Nur Aini S. Model predictive control based Kalman filter for active suspension design of light rail vehicles. Journal of Advanced Research in Dynamical and Control Systems. 2020;12(3):261-267.

28. Xiao X, Sun Z, Shen W. A Kalman filter algorithm for identifying track irregularities of railway bridges using vehicle dynamic responses. Mechanical Systems and Signal Processing. 2020 2020/04/01/;138:106582.
29. Pichlík P, Bauer J. Adhesion Characteristic Slope Estimation for Wheel Slip Control Purpose Based on UKF. IEEE Transactions on Vehicular Technology. 2021;70(5):4303-4311.
30. Karttunen K, Kabo E, Ekberg A. Numerical assessment of the influence of worn wheel tread geometry on rail and wheel deterioration. Wear. 2014;317(1):77-91.
31. Wang J, Ren Z, Chen J, et al. Study on Rail Profile Optimization Based on the Nonlinear Relationship between Profile and Wear Rate. Mathematical Problems in Engineering. 2017.
32. Ye Y, Sun Y, Dongfang S, et al. Optimizing wheel profiles and suspensions for railway vehicles operating on specific lines to reduce wheel wear: a case study. Multibody System Dynamics. 2021;51(1):91-122.
33. Tsunashima H. Condition Monitoring of Railway Tracks from Car-Body Vibration Using a Machine Learning Technique. Applied Sciences. 2019;9(13).
34. Li C, He Q, Wang P. Estimation of railway track longitudinal irregularity using vehicle response with information compression and Bayesian deep learning. Computer-Aided Civil and Infrastructure Engineering. 2022.
35. Cole C, Sun YQ. Simulated comparisons of wagon coupler systems in heavy haul trains. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit. 2006;220(3):247-256.
36. Massa A, Stronati L, Aboubakr AK, et al. Numerical study of the noninertial systems: application to train coupler systems. Nonlinear Dynamics. 2012;68(1):215-233.
37. Wu Q, Cole C, Luo S, et al. A review of dynamics modelling of friction draft gear. Vehicle System Dynamics. 2014;52(6):733-758.
38. Kalker JJ. A Fast Algorithm for the Simplified Theory of Rolling Contact. Vehicle System Dynamics. 1982;11(1):1-13.
39. Polach O. A Fast Wheel-Rail Forces Calculation Computer Code. Vehicle System Dynamics. 1999;33(sup1):728-739.
40. Polach O. Creep forces in simulations of traction vehicles running on adhesion limit. Wear. 2005;258(7):992-1000.
41. EN. Railway applications - Testing and Simulation for the acceptance of running characteristics of railway vehicles - Running Behaviour and stationary tests. Standard No.: 14363.
42. Bosso N, Zampieri N. Long train simulation using a multibody code. Vehicle System Dynamics. 2017;55(4):552-570.
43. Bosso N, Gugliotta A, Zampieri N. A Mixed Numerical Approach to Evaluate the Dynamic Behavior of Long Trains. Procedia Structural Integrity. 2018;12:330-343.
44. Petrenko V. Simulation of Railway Vehicle Dynamics in Universal Mechanism Software. Procedia Engineering. 2016;134:23-29.
45. Pogorelov D, Yazykov V, Lysikov N, et al. Train 3D: the technique for inclusion of three-dimensional models in longitudinal train dynamics and its application in derailment studies and train simulators. Vehicle System Dynamics. 2017;55(4):583-600.
46. Krishna VV, Wu Q, Hossein-Nia S, et al. Long freight trains & Long-term rail surface damage: Towards digital twins that enable predictive maintenance of

track. World Congress on Railway Research 2022 (WCRR2022); June 6-10; Birmingham, UK 2022.
47. McKay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics. 2000;42(1):55-61.
48. Bosso N, Magelli M, Zampieri N. Long train dynamic simulation by means of a new in-house code. In: Passerini G, Mera JM, Takagi R, editors. WIT Transactions on The Built Environment.; 1-3 September; Online Conference. Southampton, UK: WIT Press; 2020. p. 249-259.
49. Bosso N, Magelli M, Zampieri N. Development and validation of a new code for longitudinal train dynamics simulation. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit. 2021;235(3):286-299.
50. Bosso N, Magelli M, Zampieri N. Validation of a new longitudinal train dynamics code for time domain simulations and modal analyses. International Journal of Transport Development and Integration. 2021;5(1):41-56.
51. Spiryagin M, Wu Q, Cole C. International benchmarking of longitudinal train dynamics simulators: benchmarking questions. Vehicle System Dynamics. 2017;55(4):450-463.
52. Wu Q, Spiryagin M, Cole C, et al. International benchmarking of longitudinal train dynamics simulators: results. Vehicle System Dynamics. 2018;56(3):343-365.
53. Bosso N, Magelli M, Rossi Bartoli L, et al. The influence of resistant force equations and coupling system on long train dynamics simulations. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit. 2022;236(1):35-47.
54. Wu Q, Wang B, Spiryagin M, et al. Curving resistance from wheel-rail interface. Vehicle System Dynamics. 2022;60(3):1018-1036.
55. UIC. Wagons - Buffers with a stroke of 105 mm. Standard No.: 526-1.
56. Zhang Z, Li G, Chu G, et al. Compressed stability analysis of the coupler and buffer system of heavy-haul locomotives. Vehicle System Dynamics. 2015;53(6):833-855.
57. Bosso N, Gugliotta A, Soma A. Multibody simulation of a freight bogie with friction dampers. ASME/IEEE Joint Railroad Conference; 23-25 April; Washington, DC, USA IEEE; 2002. p. 47-56.
58. Fang K-T, Li R, Sudjianto A. Design and modeling for computer experiments. London: Taylor & Francis Group, LLC; 2016.
59. Trinchero R, Larbi M, Torun HM, et al. Machine Learning and Uncertainty Quantification for Surrogate Models of Integrated Devices With a Large Number of Parameters. IEEE Access. 2019;7:4056-4066.
60. Trinchero R, Canavero F. Machine Learning Regression Techniques for the Modeling of Complex Systems: An Overview. IEEE Electromagnetic Compatibility Magazine. 2021;10(4):71-79.
61. Vapnik VN. Statistical Learning Theory. New York, NY, USA: Wiley; 1998.
62. Suykens JAK, Van Gestel T, De Brabanter J, et al. Least Squares Support Vector Machines. World Scientific; 2002.
63. Ghojogh B, Crowley M. The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. arXiv preprint arXiv:190512787. 2019.

64. Statistics and Machine Learning Toolbox. 11.7. Natick, MA, USA: The MathWorks, Inc.
65. LS-SVMlab. 1.8. Leuven, Belgium: Department of Electrical Engineering (ESAT), Katholieke Universiteit Leuven; 2011. Available online: http://www.esat.kuleuven.be/sista/lssvmlab.