Railway Engineering Science
https://doi.org/10.1007/s40534-026-00446-3

Assessing freight train derailment risk: comparison of European rules integrated with digital twins for running safety indexes

N. Bosso[1] · L. Cantone[2] · M. Magelli[1] · R. Trinchero[3] · N. Zampieri[1]

Received: 25 February 2026 / Revised: 24 May 2026 / Accepted: 25 May 2026

Abstract
This paper compares the assessment of freight train running safety according to two European rules, namely the UIC IRS 40421 and the EN 14363, under emergency braking operations. The paper relies on a combination of longitudinal train dynamics (LTD) simulations for the evaluation of the in-train forces and multibody (MB) simulations for the calculation of the running safety indexes derived from the wheel-rail contact forces. The results confirm that the IRS 40421 is more conservative when assessing the safety of families of freight trains with a statistical perspective. Nonetheless, in individual operating scenarios, the IRS 40421 can predict safe conditions when MB simulations calculate running safety indexes above the limits prescribed by the EN 14363. To improve the running safety assessment from the outputs of LTD simulations, the paper suggests the introduction of kernel regressions, acting as digital twins of MB simulations, for the prediction of the running safety indexes. Digital twins improve the consistency with the classification of derailment conditions based on the outputs of the MB simulations. The findings highlight the complementary value of the two rules and the potential of digital twins to support the reliable assessment of freight train running safety.

Keywords Railways · Running safety · Derailment · Longitudinal train dynamics · Multibody dynamics · Machine learning · Digital twins

1 Introduction

Derailment is the biggest threat to the safety of railway systems, since it can cause loss of human lives, as well as damages to the railway infrastructure and surroundings, with high associated costs [1]. Although derailment of passenger trains can immediately result in the loss of human lives, freight train accidents can lead to human fatalities, too, due to triggered side phenomena, especially when hazardous materials are being carried. Derailments mainly stem from track defects, human factors, and operational issues. European data show that signal-passed-at-danger incidents dominated until train protection systems approximately halved them in the 2010s [2, 3]. In the USA, broken rails and track defects prevail on main lines, while switches dominate in yards [4, 5]. Ebrahimi et al. [6] report decreasing technical failures but rising human-factor contributions to derailments in Canada.

When dealing with the operation of long freight trains, the derailment risk is strongly affected by the longitudinal train dynamics (LTD) [7, 8]. Precisely, during air brake operations [9], delays in the pressure drop wave propagation cause compression of the coupling elements with the rise of large longitudinal compressive forces (LCFs). The lateral and vertical components of such forces are eventually transferred to the wheel-rail contact and can then trigger different derailing modes, as identified by Cole et al. [10].

Assessing and quantifying the derailment risk is pivotal to improve the safety of railway operations. However, experimental tests are extremely difficult to perform, due to the high danger represented by derailment conditions. In the European scenario, the running behaviour of vehicles is commonly assessed as prescribed by the EN 14363 standard [11], which nowadays has paved the way for virtual homologation of railway vehicles. The EN 14363 standard

✉ N. Zampieri
nicolo.zampieri@polito.it

1 Department of Mechanical and Aerospace Engineering, Politecnico Di Torino, 10129 Turin, Italy

2 Department of Enterprise Engineering, Tor Vergata University of Rome, 00133 Rome, Italy

3 Department of Electronics and Telecommunications, Politecnico Di Torino, 10129 Turin, Italy

defines different assessment criteria with their corresponding limits values. These criteria require to derive indexes calculated from the wheel–rail contact forces, such as the derailment ratio and the sum of lateral forces. The former index, calculated as the ratio of lateral over vertical wheel-rail forces, should be below the Nadal limit to avoid flange climbing, while the latter, related to the track lateral shift, is limited according to the Prud'homme's criterion, depending on the vehicle nominal axle-load [12]. In the framework of virtual homologation, the Nadal and Prud'homme's indexes can be easily calculated using multibody (MB) simulations, as most software packages include routines for the calculation of the wheel–rail contact forces. However, due to the high computational demands of MB simulations, usually the virtual homologation is carried out for isolated vehicles or small groups of vehicles, either completely neglecting the influence or LTD or accounting for it with oversimplified approaches.

On the other hand, to consider the effect of train composition on the running safety of freight train, the UIC IRS 40421 [13], defining the rules governing the composition and braking of international freight trains, establishes a relative approach to assess the safety of families of new freight trains not explicitly admitted to the railway traffic. By 'not explicitly admitted', it is referred to families of train configurations that are not directly covered by the pre-defined admissible domains specified in the current version of IRS 40421 (e.g., in terms of total mass, train length, and braking regime). These domains describe sets of configurations sharing common characteristics; if a train family falls outside these boundaries, it must be assessed through a relative approach by comparison with reference systems that have already demonstrated safe operation. The relative approach is currently based on the calculation of the in-train LCFs and their comparison against permissible values. This is done comparing safe train families, that are explicitly defined in the IRS 40421, and new train families for which traffic admission is demanded. The permissible LCFs are commonly determined based on the testing procedure defined in the UIC 530-2 leaflet [14] and EN 15839 standard [15]. The testing procedure involves the negotiation of an S-shaped curve with 150 m radius under the application of constant compression forces on the empty test wagon. Although the IRS 40421 defines the extrapolation rules to account for different values of curve radius and loading condition, the testing procedure is based on a standard configuration and does not allow to consider several train heterogeneities that can exist in real operations [16]. To address this issue, Krishna et al. [17] built a computational framework relying on MB simulations for the identification of permissible LCFs. Since the IRS 40421 only requires the calculation of the LCFs, it allows to perform simulations using simplified LTD codes (it is recommended the use of TrainDy software [18] ratified by the UIC TrainDy special group), which model a train consist as a system of lumped masses, with a single degree of freedom (DOF) along the track, thus ensuring better computational efficiency [8, 19–22]. Nonetheless, such codes cannot calculate the wheel-rail contact forces and hence do not allow to calculate the running safety indexes defined in the EN 14363.

Different methods exist to combine the two simulation frameworks [23, 24]. Pshinko et al. [25] proposed a computational framework for the identification of the derailment causes, based on a combination of LTD simulations, for the calculation of the in-train forces, and detailed MB simulations to obtain the wheel–rail contact forces and related safety coefficients. Di Gialleonardo et al. [26] used an advanced LTD code, including a detailed model of the buffers, to identify the most dangerous train configuration and the critical wagon for a trainset with low-flatcar wagons during braking on a S-shaped curve. Subsequently, experimental tests were performed to obtain the wheel-rail contact forces on the critical wagon. However, combining LTD simulations with MB simulations or experimental tests for many train configurations with several vehicles raises some feasibility concerns related to computational demands.

The goal of the present paper is to compare the running safety criteria of EN 14363 and IRS 40421 for a wide range of dynamic operating conditions, simulated with numerical codes combining LTD and MB models. One of the key novelties is that simulations are performed under realistic dynamic operations of railway freight trains. The simulations are run for a reference train configuration under emergency braking during negotiation of tight curves, considering a wide variety of the major inputs related to wagon configuration, track layout and train operations. To the best of the authors' knowledge, no existing work in the literature has previously dealt with an extensive numerical comparison of the derailment risk identification according to these standards over such a broad range of operating parameters. Additionally, as a further novelty, the present paper suggests the introduction of surrogate models built with machine learning (ML) least square-support vector machine (LS-SVM) kernel regressions in previous activities [27, 28], within LTD codes. The surrogate models can act as digital twins of MB simulations and provide predictions for the running safety indexes from the outputs of LTD simulations. This can enforce the capabilities of properly assessing the derailment risk with no need for computationally expensive MB simulations, thus contributing to an overall improvement of the running safety and reliability of railway systems. In this way, the freight trains can increase their hauled mass also in P-mode with a level of safety provided by the EN standards, via MB digital twins.

The paper structure is as follows. The Methods section describes the implemented computational framework and provides details on the reference simulation scenarios.

Furthermore, a quick overview on the previously developed surrogate models is provided. The Results section presents the main outcomes of the numerical simulations and compares the derailment classifications performed according to the approaches of the EN 14363 and IRS 40421 rules, also showing the application of digital twins of MB simulations. The Discussion section critically analyses the implications of the outputs of the numerical simulations and elaborates on the limitations of the modelling approach. Finally, the Conclusions section summarizes the main outcomes of the activity and suggests possible future developments.

2 Methods

2.1 Simulation scenario and input data definition

The comparison between the safety assessments carried out according to the IRS 40421 and EN 14363 is performed through numerical simulation using modules of the computational framework developed by the authors in previous activities dealing with the development of digital twins of MB simulations [27, 28]. The computational framework, sketched in Fig. 1, includes four main stages, namely:

1) Input data generation: definition of the inputs for different simulation samples through Latin hypercube sampling (LHS) implemented in a MATLAB script.
2) LTD simulations with an upgraded version of the UIC-approved TrainDy software [18]: calculation of the in-train forces and classification according to the IRS 40421.
3) MB simulations of a group of three wagons with a previously developed SIMPACK model [29]: calculation of the wheel-rail contact forces and classification according to the Nadal and Prud'homme's criteria as prescribed by the EN 14363.
4) LS-SVM kernel regressions with MATLAB user-functions: prediction of the running safety indexes from the outputs of the LTD simulations with previously trained surrogate models and classification according to the criteria of the EN 14363.

All simulation samples consider the same reference train and the same operation type, that is selected to increase the derailment risk, so that the comparison between the safety assessments according to the IRS 40421 and EN 14363 is performed for critical, but realistic, operating scenarios. The reference train includes 40 wagons equipped with a pair of Y25 bogies trailed by a head locomotive, see Fig. 2a. Except for the middle wagon trio (vehicles 20–22 with counting from the head locomotive), all other wagons feature the same values of length over buffers (18 m) and axle-load (22.5 tons), which are representative of fully loaded freight wagons running on European lines. For the wagons in the middle trio, the values of axle-load and length are determined from parameters

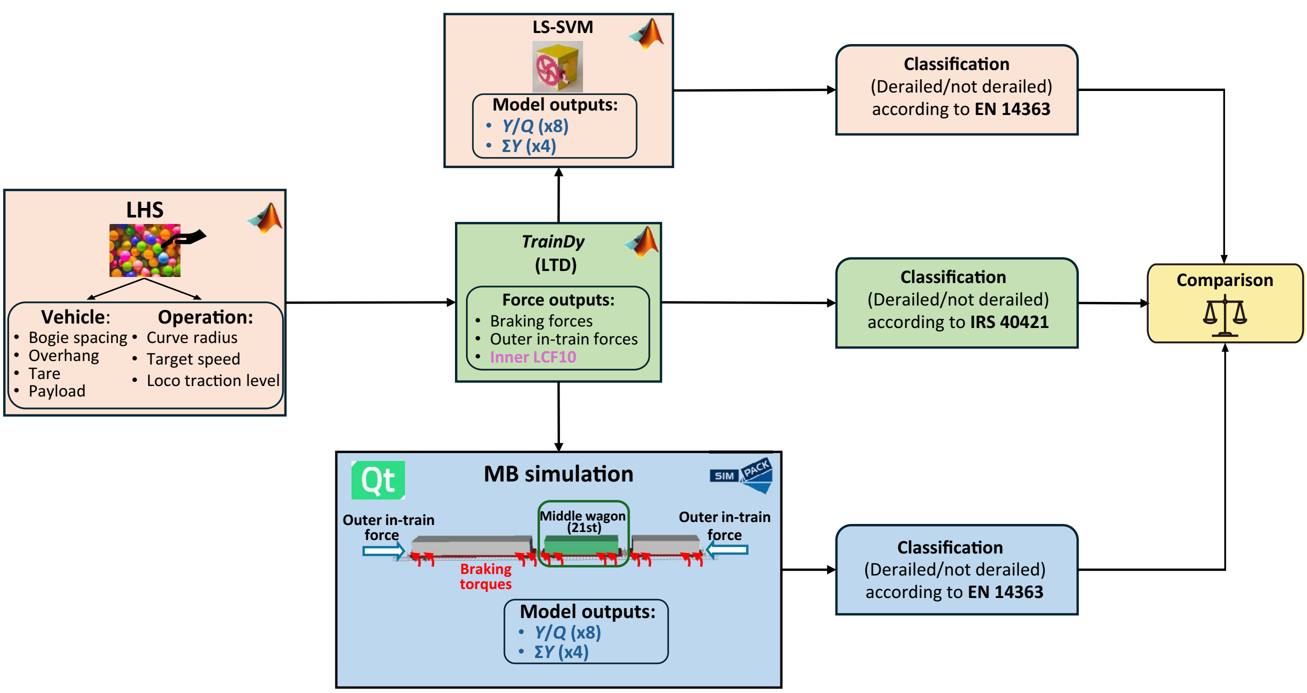

Fig. 1 Flow-chart of the computational framework

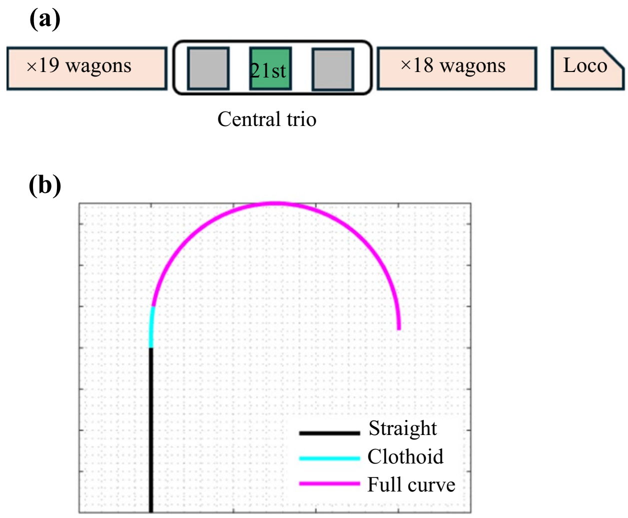

Fig. 2 Reference scenario: **a** reference train and **b** reference track layout

drawn via LHS, always ensuring that the middle wagon (21st vehicle in the train) is the lightest in the composition. For all configurations of the input space parameters, the simulations deal with an emergency braking operation after acceleration from 5 km/h to a target speed value in the range 20–40 km/h. The chosen speed range is based on the friction characteristics of the cast iron (LL-type brake blocks are similar) brake blocks considered in the simulations. Since the coefficient of friction (CoF) increases as the sliding speed decreases, the most critical LCFs are typically generated at low speeds. Consequently, this range is representative of the conditions under which derailments caused by excessive LCFs are most likely to occur and are typically investigated in European safety assessments [30], while higher speeds are primarily used to assess braking performance (e.g., stopping distances). However, the proposed methodology remains general and not inherently limited to this specific speed range. The reference train runs on a track including an initial straight section, followed by a clothoid transition and a full-curve section with radius drawn via LHS in the range 150–300 m, see Fig. 2b. The values of superelevation and clothoid transition length are calculated as a function of the curve radius to obtain a nominal twist of 2‰. In all simulations, the middle wagon reaches the end of the clothoid transition when subjected to the largest LCF. In the MB environment, a twist defect is defined at the end of the clothoid transition. The twist defect is such that the total twist is equal to 6.5‰ over a base length of 3 m, which corresponds to the limit value prescribed by Rete Ferroviaria Italiana (RFI), i.e., the Italian railroad administration [31]. The presence of the twist defect at the end of the clothoid transition further justifies the low speed values sampled via LHS. In fact, running over the twist defect at low speed ensures the generation of wheel-rail forces with quasi-static components that can really cause wheel climbing and derailment.

The major inputs for the simulations, drawn through LHS with a dedicated MATLAB script, are related to the characteristics of the wagons in the middle trio and to the operation type. All inputs drawn via LHS are assumed to feature a uniform probability distribution within a pre-defined variation range. Precisely, for each wagon in the middle trio, four parameters are drawn, namely bogie spacing, overhang, tare load and payload. This allows to explore different values of axle-load and length over buffers. The bogie spacing is drawn in the range 5–20 m, while the vehicle overhang is determined to ensure compliance with the UIC 530–2 leaflet [14]. Tare is selected from the range 20–36 ton, while the payload is defined to guarantee that the maximum weight on rails of the outer vehicles is limited to 90 ton, whereas for the central wagon, the weight on rails is limited so that this vehicle is always the lightest (up to 60 ton) in the composition and hence experiences the highest derailment risk.

Focusing on the operation, the radius of the full-curve section is drawn from the range 150–300 m, with application of gauge widening as prescribed by RFI [31]. Furthermore, the target speed at braking start is determined from the range 20–40 km/h. Finally, the last parameter drawn via LHS is related to the locomotive traction level for the initial acceleration to the target speed. This makes it possible to obtain different values of in-train forces on the wagons of the middle trio, thus making the analysis independent of the reference train configuration.

2.2 LTD simulations: TrainDy

The TrainDy software [18] was developed by the University of Rome Tor Vergata with financial and technical support from Faiveley Transport (a Wabtec Group company). In 2025, UIC published the 2nd version of the International Railway Solution (IRS) 40,421 [13], in which the TrainDy software, as validated by the UIC TrainDy Special Group, is recommended for LTD simulations. TrainDy simultaneously solves the pneumatic and dynamic problems under various operational scenarios.

The software employed for the LTD simulations presented in this paper is an extended version of the UIC-approved TrainDy code, specifically enhanced to analyse train configurations with user-modified inputs, including those generated through LHS. For the purposes of this study, the initial position of the train is treated as an input parameter. This approach allows, for each set of input parameters generated by the LHS, the imposition of the previous condition that the peak of the LCFs acting on the 21st vehicle occurs when the wagon is located at the clothoid transition outlet. In this manner, all simulations share a common characteristic, irrespective of the specific input parameters. To enforce this condition, multiple preliminary simulations are performed for each input set prior to the final simulation, with the aim of adjusting the initial train position to satisfy the condition on the peak LCFs. To this end, a dedicated algorithm, conceptually similar to the MATLAB *fzero* function, was developed to identify the zeros of a nonlinear numerical equation. This iterative procedure typically converges in approximately three iterations, with a maximum of six in rare cases, ensuring that the target position is reached with a residual error below ten centimetres (the chosen threshold). It is important to note that this adjustment only modifies the initial train position to satisfy a prescribed geometric constraint at the onset of braking. Since it does not constrain the dynamic evolution of the system or the sampled input parameters, no systematic bias is introduced towards specific longitudinal force histories in the explored operating space. Furthermore, to improve computational efficiency, TrainDy was restructured as a MATLAB function capable of parallel execution and of dynamically modifying input data written to disc, such as payload distribution and wagon tare and length, without requiring distinct input files for each case. Consequently, it was not necessary to define a separate train configuration, test, or operational scenario for each set of input parameters.

Two types of analyses are performed in the context of this paper. For the former, TrainDy simulates trains composed of wagons characterised by virtual geometries and inertial properties derived from the LHS samples, thus considering "realistic" wagon types. For the latter, the characteristics of the wagon trio are adjusted to reproduce those of existing wagons, specifically in terms of length, bogie spacing (which is not directly used by TrainDy), tare mass, and maximum payload. In this latter analysis, each wagon in the middle trio corresponds to an existing wagon. For all cases, cast iron has been used as friction material for tread blocks of all wagons.

2.3 MB model of the wagon trio

The MB simulations are run with a detailed model of the middle wagon trio built in SIMPACK, and derived from a well-established model of the Y25 bogie developed by the railway research team from Politecnico di Torino (PoliTo) [29], see Fig. 3a for the model assembly and Fig. 3b for a detailed view of the central wagon. Please note that to consider the wagon torsional stiffness, each vehicle is modelled with the definition of two half-carbody bodies, connected with a force element with lumped stiffness (bushing), see Fig. 3b.

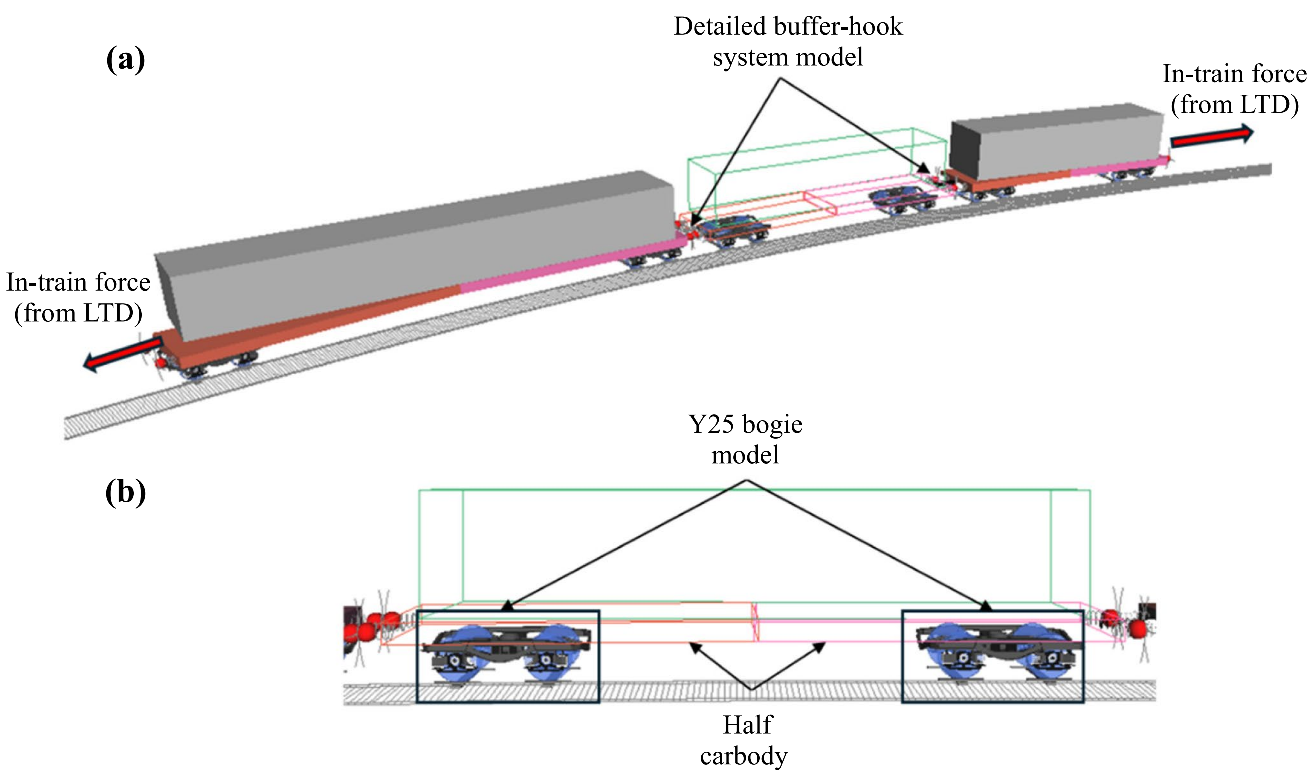

Fig. 3 SIMPACK MB model: **a** model of the trio and **b** detail of an individual wagon

The buffer-hook systems connecting the central wagon to the outer wagons are modelled in detail, thus allowing to determine the lateral and vertical components of the in-train forces. The model of each wagon in the trio includes four additional bodies for the buffers. The inner buffer bodies, transmitting the forces between the central wagon and the outer ones, have a single DOF in longitudinal direction with respect to the half-carbody. The front buffers of the front wagon and the rear buffers of the rear wagon (outer buffers) are fixed to the corresponding half-carbody. Each buffer is connected to the corresponding half-carbody with a SIMPACK built-in force element with hysteresis, which implements the buffer mechanical impedance characteristic, see Fig. 4. Furthermore, to model the contact between adjacent buffers, Hertzian contact force elements are defined in the model, which solve the contact between pairs of identical spheres with radius of 2.75 m. The hook connection is modelled through the definition of nonlinear point-to-point (PtP) force elements with hysteresis, defined between markers on two adjacent half-carbodies and implementing the mechanical characteristics of the series of two hooks, see Fig. 4 for the mechanical characteristics of a single hook. It is worth pointing out that the detailed model of the buffer-hook system leads to the generation of lateral and vertical components of the in-train forces, that mainly originate from curve negotiation and differences in weight-on-rails of the vehicles. The magnitude of the lateral and vertical components that are eventually transferred to the wheel-rail contact, possibly contributing to flange climbing, depends on the considered characteristics for buffers and hook. In this work, the mechanical impedance characteristics are piecewise linear functions extracted from the TrainDy database. From previous activities, these characteristics are known to ensure a good match with experimental data. Nonetheless, future upgrades of the work could consider the application of different mechanical impedance curves, to investigate the effect of different types of buffers and hooks, as well as of different modelling strategies. This could thrust the development of

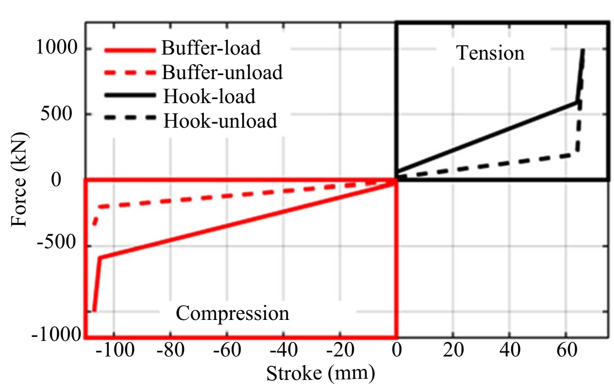

Fig. 4 Mechanical impedance characteristics with hysteresis for hook and buffer

surrogate models able to estimate the uncertainties related to the mechanical behaviour of different types of buffers and hooks or different applied preloads.

On the other hand, the connection of the outer vehicles to the rest of the train composition is simulated through the application of the in-train forces calculated from the TrainDy simulations, as sketched in Fig. 3a. These forces act along the longitudinal direction only, to ensure that the speed profile calculated from the MB simulations well matches the one computed from the LTD simulations. In fact, it is believed that neglecting the lateral and vertical components of these in-train forces, which connect the outer wagons to the rest of the composition, has a marginal effect on the running safety indexes calculated on the middle wagon under analysis. To further mitigate this effect, it could be possible to build a model with five wagons, however, in the authors' opinion, this would significantly increase the computational effort while minimally impacting the values of the running safety indexes of interest.

The air brake forces calculated with TrainDy are applied as equivalent braking torques on the wheelsets. Additionally, dedicated expression forces are defined in the model to account for the ordinary and curving resistances applied in the TrainDy simulations.

The wheel-rail contact problem is solved using the SIMPACK equivalent elastic method for the calculation of the contact patch size and normal load, while the FASTSIM algorithm [32] is applied to calculate the tangential forces. Simulations are run considering the nominal S1002 and UIC60 wheel and rail profiles, respectively. The track is modelled as a moving elastic foundation with sleeper mass and stiffness values obtained from the literature [33].

The MB model is built with a parametric approach, in order to update the main model variables based on the values of the simulation inputs drawn via LHS for each sample. As the model of each wagon includes two carbodies connected with a lumped stiffness element to consider the torsional stiffness of the carbody, the value of the torsional stiffness is adjusted based on the value of the bogie spacing. Similarly, the model automatically updates the value of the wheelset pitch inertia based on the tare value of each wagon, to ensure a constant ratio between the rolling inertias and tare mass. The cascade of MB simulations is managed via a MATLAB script that launches batch sessions calling SIMPACK QtScript files.

2.4 LS-SVM regression for metamodeling

LS-SVM regression is a kernel-based supervised learning method that enables the learning of complex input–output relationships from limited training data by solving a least-squares optimization problem [34, 35]. In the primal space, LS-SVM regression is expressed as

$$\mathcal{M}_{\mathrm{LS-SVM}}(\boldsymbol{x};\boldsymbol{w},b) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Phi}(\boldsymbol{x}) + b, \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^d$ is the model input vector, $\boldsymbol{\phi}:\mathbb{R}^d \rightarrow \mathbb{R}^D$ is a nonlinear mapping function vector into the feature space, while $\boldsymbol{w}$ and b are the regression coefficients and the bias term, respectively. The model parameters are determined in the training stage from the solution of a regularized optimization problem [34, 35] expressed as

$$\min_{\boldsymbol{w},b} \frac{\gamma}{2} \sum_{l=1}^{L} \left(y_l - \mathcal{M}_{\mathrm{LS-SVM}}(\boldsymbol{x}_l;\boldsymbol{w},b)\right)^2 + \frac{1}{2}\|\boldsymbol{w}\|^2, \tag{2}$$

where $\boldsymbol{x}_l \in \mathbb{R}^d$ are the training data input vectors, $y_l \in \mathbb{R}$ are the outputs, γ is the regression hyperparameter [36] and the second term acts as a regularizer that prevents overfitting while keeping the model flattness.

To make the model complexity independent of input and feature space dimensionalities, the LS-SVM regression can be expressed in its dual formulation as [34, 35, 37]:

$$\mathcal{M}_{\mathrm{LS-SVM}}(\boldsymbol{x}) = \textstyle\sum_{l=1}^{L} \alpha_l k\left(\boldsymbol{x}_l, \boldsymbol{x}\right) + b, \tag{3}$$

where $k(\cdot,\cdot)$ is the kernel function, and the α_l terms refer to the dual coefficients. The kernel function implicitly maps data into a higher-dimensional space, thus not requiring to explicitly calculate feature transformations $\boldsymbol{\phi}$ [34], hence enabling high computational efficiency.

In past activities [27, 28], the authors of the present paper built multiple scalar LS-SVM surrogate models for a quick evaluation of the running safety indexes of all wheelsets and wheels of the middle wagon, using Gaussian radial basis function kernels. In the training stage, the coefficients for the LS-SVM regressions were obtained from the outputs of MB simulations run for a limited dataset considering randomly generated wagons.

The surrogate models are closed-form expressions acting as DTs of MB simulations, allowing to evaluate the running safety indexes with very high accuracy computational accuracy. More in detail, the surrogate models require as inputs:

- Vehicle parameters, such as tare load, payload, length over buffers and bogie spacing of all wagons in the simulated trio.
- Operating parameters, namely the curve radius and the target speed at the onset of braking.
- LTD-derived quantities, i.e., the LCFs at the front and rear of the middle wagon.

Overall, 20 scalar LS-SVM regressors are employed to model and estimate the derailment index and unloading ratio at each of the eight wheels of the middle wagon, together with the sum of the lateral forces for each of the four wheelsets, thereby avoiding the need for computationally demanding MB simulations. The models were trained using the MATLAB LS-SVM Lab Toolbox [38], while the hyperparameters were optimized through leave-one-out cross-validation combined with simplex optimization. It is worth noting that each surrogate model was trained independently and involves tuning only two hyperparameters: the parameter γ, see Eq. (2), and the width of the Gaussian kernel.

2.5 Classification of derailment conditions

2.5.1 Classification according to IRS 40421

The IRS 40421 explicitly specifies the characteristics of freight trains admitted to operation, but it also establishes a methodology for the statistical assessment of the safety level of new train compositions in comparison with the established safety level of trains already in service. This is achieved with a relative approach, which is also recalled in European legislation, namely Implementing Regulation (EU) No. 402/2013 of 30 April 2013 on the common safety method for risk evaluation and assessment [39], providing acceptable means of compliance. The relative approach consists in comparing a new train family with an accepted train family in terms of LCFs and in evaluating the corresponding number of potential derailments. If the number of potential derailments associated with the new train family is lower than that of the accepted train family, the new family is deemed suitable for safe operation. Once the two train families have been defined, both are simulated following the selected train operating conditions. The resulting LCFs for both train families are then computed and compared with the corresponding permissible LCF values. For each simulation, it is possible to define a ratio of the in-train forces $\mathrm{DF}_{\mathrm{IRS}}$, as stated by

$$\mathrm{DF}_{\mathrm{IRS}} = \max\left|\frac{\mathrm{LCF}_k}{\mathrm{PLCF}_k}\right|, \tag{4}$$

where k is the counter of the vehicle, *LCF* is the maximum value of LCF acting on vehicle k, and *PLCF* is the corresponding permissible value. The permissible LCFs are defined at the wagon level and depend on key parameters, such as wagon type (two-axle or bogie), payload condition, and horizontal curve radius. These values are originally derived from experimental tests, defined in the UIC 530-2 [14] leaflet and subsequently generalized through extrapolation rules in the IRS 40421 [13]. For instance, according to this methodology, a maximum permissible LCF of 400 kN is defined for an empty bogie wagon negotiating an S-shaped curve with a 190 m radius. For safety assessments, wagons are typically evaluated under these critical conditions to ensure a conservative estimation of the risk.

The number of virtually derailed trains is determined by counting the cases in which the $\mathrm{DF}_{\mathrm{IRS}}$ ratio exceeds unity.

Subsequently, the confidence intervals of the derailment probability for both train families are evaluated. IRS 40421 recommends a confidence level of 95% and the use of the Clopper–Pearson interval. If the upper bound of the confidence interval for the new train family is lower than the lower bound of the confidence interval for the accepted train family, the new train family may be considered at least as safe as the existing one and can therefore be admitted to operation in accordance with IRS 40421.

Simulations run in this paper do not intend to admit new families of freight trains to traffic, rather they have the goal of evaluating the running safety of individual scenarios. Therefore, the values of the LCFs at the front and rear of the central wagon estimated with TrainDy are directly compared against the permissible values extracted from the IRS 40421, with application of Eq. (4). Therefore, when for a specific simulation scenario either of the LCFs on the middle wagon is above the permissible value, the sample is classified as "derailed" according to the IRS 40421.

2.5.2 Classification according to EN 14363

Concerning the classification based on the criteria defined in the EN 14363, at the end of each MB simulation, a representative value of the Nadal factor (Y/Q) is extracted for each of the 8 wheels of the central wagon, using the post-processing strategy and percentile selection defined in the EN 14363. The same procedure is applied to extract representative values of the Prud'homme's factor ΣY for the four wheelsets of the central wagon. Furthermore, for each sample, it is possible to estimate the same quantities from the outputs of the TrainDy simulations using the LS-SVM regressions. To quantify the derailment risk, a risk ratio is defined in this paper, for both Nadal and Prud'homme's factors obtained with either method (MB simulation or LS-SVM regressions), as stated by:

$$\mathrm{DF_N} = \max \frac{|Y/Q|_{ij}}{0.8} \quad i = 1-4, j = 1-2, \tag{5}$$

$$\mathrm{DF_{Ph}} = \max \frac{|\sum Y|_i}{0.85 \cdot (10 + P_{F0}/3)} \quad i = 1-4, \tag{6}$$

where $\mathrm{DF_N}$ and $\mathrm{DF_{Ph}}$ are the risk ratios for Nadal and Prud'homme's factors, Y/Q is the derailment ratio, ΣY is the sum of guiding forces on the wheelset, P_{F0} is the axle-load and finally i is the wheelset counter, while j is the wheel counter, which is 1 for right-hand side wheels and 2 for left-hand side wheels. Please note that the denominators in Eqs. (5) and (6) correspond to the limit values for the derailment ratio and sum of guiding forces as prescribed by the EN 14363 for freight wagons for dynamic tests. In fact, the simulated operating scenario resembles braking manoeuvres that can take place during common operations on standard tracks. Similarly, a global risk ratio $\mathrm{DF_{EN}}$ is defined as

$$\mathrm{DF_{EN}} = \max\left[\mathrm{DF_N};\mathrm{DF_{Ph}}\right]. \tag{7}$$

Therefore, when the global risk ratio is above 1, the simulation scenario is classified as "derailed" based on the prescription of the EN 14363, as either of the two safety indexes is above the prescribed limits.

It is important to stress that in all simulations analysed in this paper, no real derailment was actually observed in the MB framework. The "derailed"/ "not derailed" terms are used in the paper only to refer to conditions where the limits of the international rules are exceeded. In these terms, MB derailments can be considered virtual, as is the case for IRS derailment. In fact, running safety indexes above their corresponding limit values do not directly imply a derailment, but can be considered as a symptom of anomalous running conditions that could potentially bring to flange climbing.

3 Results

3.1 Randomly generated wagons

To compare the classifications performed according to the two European rules, 394 simulation samples with randomly generated wagons were originally considered. Figure 5a shows the cumulative distribution function for the global

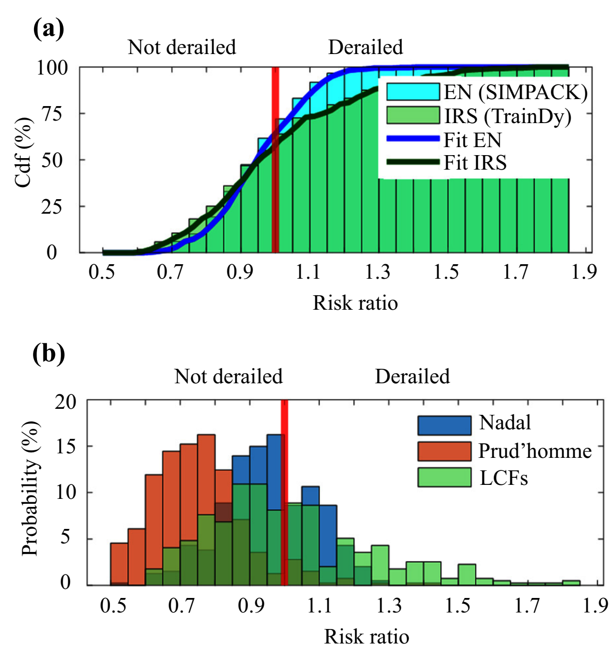

Fig. 5 Derailment risk ratios calculated from MB simulations: **a** cumulative distribution function (Cdf) of the global risk ratio; **b** probability distribution for Nadal and Prud'homme's factors

risk ratio calculated according to Eq. (7) from the results of the MB simulations, as well as for the risk ratio based on the LCFs, calculated with Eq. (4). It can be observed that out of all simulations, slightly less than 40% is classified as "derailed" based on the EN 14363 criteria, and that the global risk ratio is always above 0.6. This confirms that the selected simulation scenario leads to risky running conditions, as designed. Furthermore, it can be observed that the IRS 40421 tends to predict higher values of the risk ratio, thus proving to be more conservative overall with respect to the EN 14363. From Fig. 5b, which displays the probability distributions of the derailment risk ratios calculated with Eqs. (4)–(6), it can be noticed that on average, the Nadal's criterion is more critical with respect to Prud'homme's one, i.e., flange climbing is more likely to occur with respect to the track lateral shift.

In this paper, the comparison between classifications performed with different methods are shown as in Fig. 6. Figure 6a is a confusion chart that clearly highlights the agreements and discrepancies in labelling of running conditions. On the other hand, Fig. 6b is a scatter plot that shows the values of the risk ratio. The closer a point is to the bisector line, the better the agreement between the two methods in calculating the risk ratio. Concerning the comparison between the IRS 40421 and EN 14363, as shown in Fig. 6a, in 323 cases out of the 394 simulation samples, corresponding to 82%, the two rules lead to the same classification. For 44 simulation samples (11%), instead, the IRS 40421 proves to be more conservative than the EN 14363, with prediction of "derailed" conditions when the running safety indexes calculated from the wheel–rail contact forces are below their limits. This is in line with the expectations. In fact, since the IRS 40421 does not consider the wheel–rail contact forces, it is desired that the derailment classification with this rule is more conservative. Unfortunately, in the remaining 27 tested configurations (7%), the IRS 40421 rule predicts false negatives, i.e., it detects safe conditions when a detailed MB simulation shows instead that the running safety indexes exceed their corresponding limit values. This is the most undesired condition, as usually when assessing the running behaviour according to the IRS 40421 rule, the wheel–rail contact forces are not calculated and this means that unsafe train operation could be undetected. Nonetheless, according to the IRS 40421, 168 simulations (43%) out of the total 394 cases are labelled as "derailed", while the running safety indexes defined by the EN 14363 exceed their limit values only in 151 simulations (38%). This means that, despite possible disagreements on individual simulations, when performing statistical evaluations of the running safety of freight trains, the number of samples labelled as "derailed" by the IRS 40421 method is in line with the number of simulations classified as "derailed" according to the EN 14363 prescriptions applied to the outputs of the MB simulations. For better graphical clarity, the scatter plot in Fig. 6b shows the values of risk ratios for the two methods in all simulation scenarios. Points with x-axis values above 1 and y-axis value below 1 are classified as false negative values, since they are labelled as "not derailed" by the IRS 40421 method when instead the MB simulation leads to running safety indexes above the prescribed limit values defined in the EN 14363. Similarly, points with x-axis values below 1 and y-axis values above 1 are false positives, since they correspond to simulations where, despite not featuring running safety indexes above the limits established in the EN 14363, the IRS identifies derailment. Concerning instead points with y- and x-values above 1, identified as true positives, it can be noticed that the risk ratio calculated with the IRS 40421 criterion tends to be higher than the one obtained from the EN 14363 limit values, hence resembling the trends observed in Fig. 5.

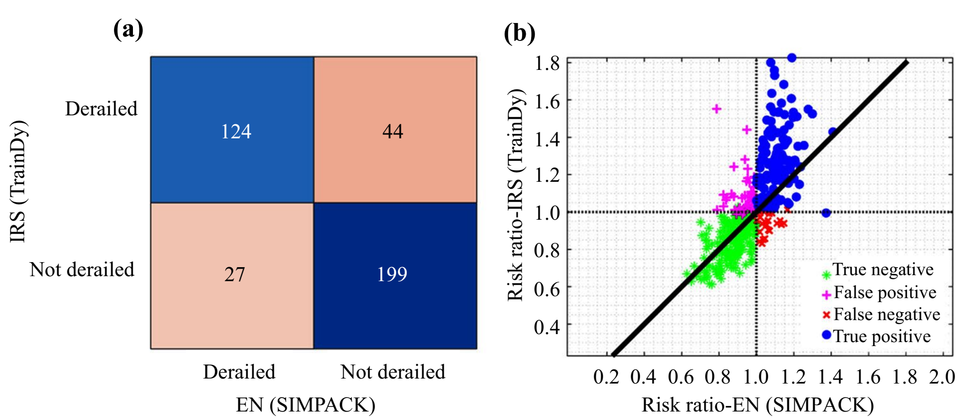

Fig. 6 Classification according to the EN 14363 and IRS 40421 for randomly generated wagons: **a** confusion chart and **b** scatter plots for the values of risk ratio

3.2 Existing wagons

The original 394 simulation samples considered wagons with realistic values of bogie spacing and overhang, complying with requirements set by international rules, but not specific existing wagon families. Therefore, it was decided to repeat all simulations for real wagon types running on European tracks. The existing wagons for each case were selected from the TrainDy database, ensuring main features close to those of the originally simulated vehicles. The simulations were repeated considering the same value of weight on rails for the vehicles, thus modifying the payload values according to the tare mass, so that the central wagon is always the lightest in the composition. This additional simulation set is essential to ensure that the results obtained in the first stage are not biased by using randomly generated wagons. Figure 7a provides the confusion chart for the simulations repeated with existing wagons. It can be noticed from a comparison with Fig. 6a that the values in the cells of the confusion charts are similar, meaning that when using real wagons, the classifications performed with the two methods do not feature significant changes. Overall, with real wagons, the number of "derailed" cases according to the EN 14363 slightly increases to 157 (40%), while the IRS 40421 classifies 166 samples (42%) as "derailed". Although generally more conservative, the IRS 40421 still yields 35 false negatives (9%), confirming that relying solely on its criteria may be misleading. Similarly, the scatter plot in Fig. 7b resembles the one in Fig. 6b, with true positive points tending to be well above the bisector line since the IRS 40421 approach is more conservative overall.

Figure 8a provides the confusion chart for comparison of the labelling according to the EN 14363 criteria applied on the outputs of the MB simulations and on the predictions of the LS-SVM models, built in past activities [28] considering randomly generated wagons and trained with 400 samples. A comparison with Fig. 7a highlights that the total number of disagreements drops from 79 (20%) to 44 (11%), with the number of false negatives approximately halving from 35 (9%) to 20 (5%). Still, labelling of the running conditions

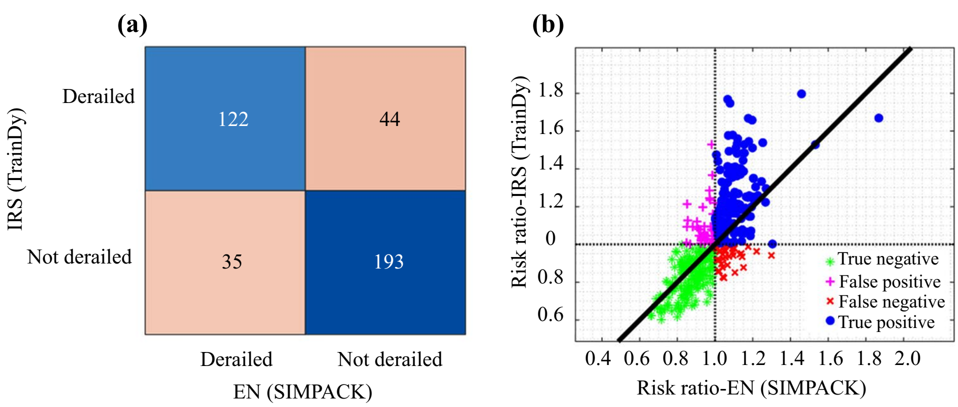

Fig. 7 Classification according to the EN 14363 and IRS 40421 for existing wagons: **a** confusion chart and **b** scatter plots for the values of risk ratios

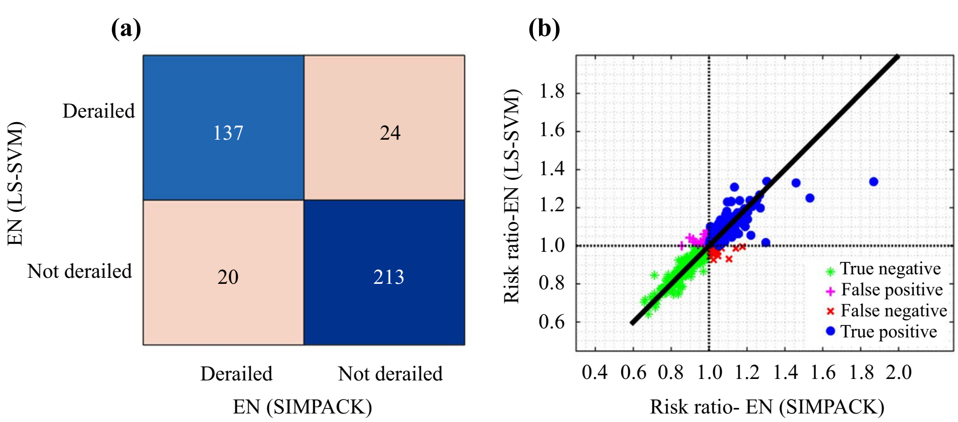

Fig. 8 Comparison of for the classification according to the EN 14363 for results of MB simulations and predictions of LS-SVM surrogate models: **a** confusion chart and **b** scatter plot of risk ratios

based on the application of the LS-SVM models on the outputs of the LTD simulations leads to the identification of a larger number of derailments (161) with respect to classifying the outputs of the MB model according to the EN 14363 criteria (157). This means that the introduction of digital twins and their application to the outputs of LTD simulations can ensure an additional improvement of the classification of the simulation scenarios with no need for extra MB simulations, whilst at the same time being more conservative overall. Nonetheless, it should be clarified that this outcome was observed on the specific dataset considered in this work, but no constraints are applied in the training stage to ensure that the LS-SVM are indeed more conservative. Focusing on the scatter plot in Fig. 8b, points tend to be closer to the bisector line, thus proving that, in addition to improved labelling of running operations, the LS-SVM models can calculate values of the derailment risk ratio close to the one obtained from the MB simulations.

4 Discussion

The results of the simulations show that both with realistic and existing wagons, the IRS 40421 predicts a higher number of derailment cases with respect to the EN 14363, thus proving to be a conservative and reliable method when performing statistical evaluations on families of trains with the relative approach. On the other hand, when considering specific cases, the IRS 40421 can lead to false negatives, predicting safe running conditions when derailment is detected by the application of the EN 14363 criteria on the outputs of MB simulations. The calculation of the running safety indexes with LS-SVM surrogate models can almost halve the number of false negatives, although not completely eliminating them. In this regard, it should be stressed that the simulations carried out in the present paper consider conditions that are on the edge of derailment, as the derailment risk factor calculated with respect to the EN 14363 limits is always above 0.6 when considering existing wagons, see Fig. 9. Therefore, it is expected that the labelling of the simulation scenarios could improve when considering either typical operations, with very small derailment risk, or risky conditions. At the same time, the LS-SVM regressions predict numerical values for the EN 14363 safety indicators, which can be compared against their limit values. This allows to define safety margins that can help detect on-the-edge conditions that, despite not being classified as "derailed", feature running safety indexes close to their limit values. Similarly, future developments of the activity could focus on building probabilistic surrogate models, capable of providing not only a mean prediction but also the associated uncertainty. In this way, the running safety index could be described through a probability distribution rather than as a single deterministic value. Consequently, such models would make it possible to estimate the probability of derailment, instead of relying solely on a deterministic derailment risk ratio.

The classification discrepancies between IRS 40421 and EN 14363 could be due to the method used for the determination of the permissible LCFs. In fact, the testing procedure suggested in the UIC 530-2 leaflet considers steady-state conditions, while typical operations including the simulation scenarios analysed in this paper are dynamic manoeuvres. Moreover, the UIC 530-2 leaflet defines specific operating conditions, in terms of neighbouring wagons and track layout, while freight wagons actually operate under very heterogenous conditions. The definition of permissible LCFs for more heterogenous conditions could help improve the agreement between the IRS 40421 and the EN 14363. Due to the high cost of experimental testing, this could be obtained through detailed numerical simulations with MB codes as already done in the existing literature [17]. Nonetheless, running MB simulations for a wide range of operating conditions, also considering multiple dynamic scenarios could be

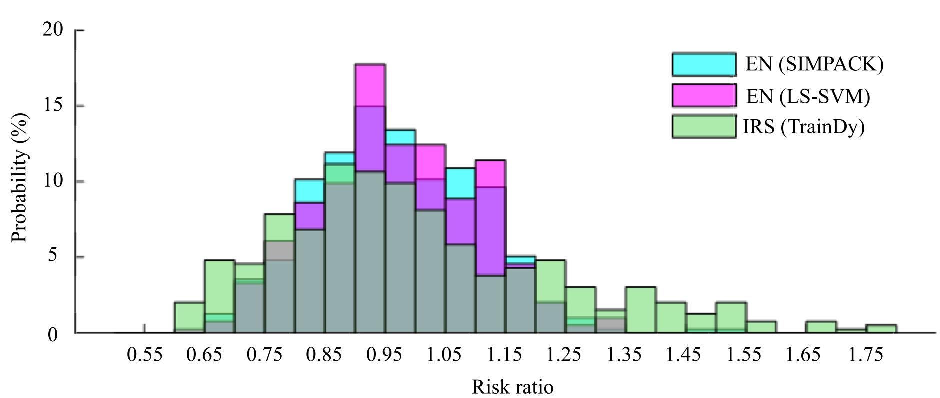

Fig. 9 Probability distribution of the derailment risk ratio from TrainDy, SIMPACK and LS-SVM surrogate model outputs

practically unfeasible. Therefore, in the authors' opinion, the derivation of surrogate models can be the most convenient approach, since it limits the number of training data whilst ensuring good prediction accuracy.

It is important to highlight that, despite their general validity, the specific results obtained in this paper are affected by certain underlying assumptions and hypotheses. First, when considering different wagons, the MB model updates the value of carbody torsional stiffness with a simplified approach, as a function of the bogie-spacing only. Therefore, the model does not currently account for the specific mechanical structure of the carbody, which can change significantly from one wagon type to another. Future activities should address the improvement of the modelling for a broad range of wagon types to increase the accuracy of the calculated outputs. Similarly, in the simulations launched in the framework of the present paper, gauge widening was applied in tight curves as prescribed by RFI [31]. However, rules on gauge widening are currently being relaxed across Europe, hence in future activities the effect of gauge widening should be analysed more deeply. Additionally, it is clear that the running safety indexes depend on the method selected for the solution of the wheel-rail contact problem. In this paper, the MB simulations rely on a golden standard approach, namely the combination of the SIMPACK equivalent elastic semi-Hertzian method and Kalker's FASTSIM algorithm [32]. This approach allows to achieve a good compromise between computational efficiency and accuracy. However, especially for simulation samples with derailment risk ratio close to 1, it could be of interest to run simulations with refined methods for the calculation of the wheel-rail contact forces.

5 Conclusions

The activity shown in the paper allows to throw a light on the discrepancies in the evaluation of the running safety conditions according to the prescriptions of the IRS 40421, based on the in-train forces calculated via LTD simulations, and those of the EN 14363, which relies on running safety indexes calculated from the wheel-rail contact forces computed with MB simulations. The main outcomes of the activity are given in the following bulleted list.

- Overall, the IRS 40421 identifies a higher number of derailments with respect to the EN 14363 when considering both "realistic" vehicles and existing wagon types. This confirms that the prescriptions of the IRS 40421 are conservative when applying a statistical assessment on families of freight trains with a fleet-oriented investigation.
- Under specific individual scenarios, the IRS 40421 predicts false negatives, hence suggesting that some degree of discrepancy exists between methods based on the in-train forces and methods that consider the wheel-rail contact forces. Therefore, it is recommended not to rely entirely on the permissible in-train compressive forces calculated according to IRS 40421 beyond the statistical context for which it was designed.
- The introduction of digital twins of MB simulations based on kernel regressions predicting the running safety indexes from the outputs of LTD simulations can significantly reduce the number of false negatives at a negligible computational cost.
- The study confirms the complementary role of the two European rules, as combining LTD and MB simulations is indeed the most reliable method for the assessment of the running safety of freight trains. However, such strategy features practical feasibility concerns when dealing with large numbers of train configurations and operating scenarios. The introduction of digital twins is promising as it can improve accuracy whilst at the same time not impacting on the computational burden.

Future upgrades of the study could address the following research areas:

- Moving towards a probabilistic assessment of running safety under individual operating conditions by adopting ML models that provide a predictive distribution for derailment, rather than a single deterministic estimate.
- Upgraded MB modelling, to consider the real torsional stiffness of existing wagon types, explore different wheel-rail contact algorithms and investigate the effect of different gauge widening strategies.

Acknowledgements This work was supported by Ministero dell'Università e della Ricerca, FISA-2023-00314, Nicolò Zampieri.

References

1. Wilson N, Wu H, Klopp A et al (2019) Railway vehicle derailment and prevention. In: Iwnicki S, Spiryagin M, Cole C et al (eds) Handbook of railway vehicle dynamics, 2nd edn. CRC Press, Boca Raton
2. Evans AW (2011) Fatal train accidents on Europe's railways: 1980–2009. Accid Anal Prev 43(1):391–401
3. Evans AW (2021) Fatal train accidents on Europe's railways: an update to 2019. Accid Anal Prev 158:106182
4. Liu X, Saat MR, Barkan CPL (2012) Analysis of causes of major train derailment and their effect on accident rates. Transp Res Rec J Transp Res Board 2289(1):154–163
5. Liu X (2017) Statistical causal analysis of freight-train derailments in the United States. J Transp Eng A Syst 143(2):04016007
6. Ebrahimi H, Sattari F, Lefsrud L et al (2021) Analysis of train derailments and collisions to identify leading causes of loss incidents in rail transport of dangerous goods in Canada. J Loss Prev Process Ind 72:104517
7. Cole C et al (2019) Longitudinal train dynamics and vehicle stability in train operations. In: Iwnicki S, Spiryagin M, Cole C (eds) Handbook of railway vehicle dynamics, 2nd Edition. CRC Press, Boca Raton
8. Wu Q, Spiryagin M, Cole C (2016) Longitudinal train dynamics: an overview. Veh Syst Dyn 54(12):1688–1714
9. Wu Q, Cole C, Spiryagin M et al (2023) Freight train air brake models. Int J Rail Transp 11(1):1–49
10. Cole C, McClanachan M, Spiryagin M et al (2012) Wagon instability in long trains. Veh Syst Dyn 50(sup1):303–317
11. European Committee for Standardization (2016) Railway applications: testing and simulation for the acceptance of running characteristics of railway vehicles—running behaviour and stationary tests. EN 14363:2016
12. Schmid R, Micić Batka V, Pospischil F (2026) Prud'homme criterion: a review of its application in railway vehicle approval. Int J Rail Transp 14(1):1–16
13. International Union of Railways (2021) Rules for the consist and braking of international freight trains. IRS 40421/E/2:2021
14. International Union of Railways (2012) Wagons: running safety. UIC 530–2:2008
15. Deutsches Institut für Normung e.V. (2024) Railway applications: testing and simulation for the acceptance of running characteristics of railway vehicles — running safety under longitudinal compressive force. EN 15839:2024
16. Krishna VV, Berg M, Stichel S (2020) Tolerable longitudinal forces for freight trains in tight S-curves using three-dimensional multi-body simulations. Proc Inst Mech Eng Part F J Rail Rapid Transit 234(5):454–467
17. Krishna VV, Jobstfinke D, Melzi S et al (2021) An integrated numerical framework to investigate the running safety of overlong freight trains. Proc Inst Mech Eng Part F J Rail Rapid Transit 235(1):47–60
18. Cantone L (2011) TrainDy: the new Union Internationale des Chemins de Fer software for freight train interoperability. Proc Inst Mech Eng Part F J Rail Rapid Transit 225(1):57–70
19. Cole C, Spiryagin M, Wu Q et al (2017) Modelling, simulation and applications of longitudinal train dynamics. Veh Syst Dyn 55(10):1498–1571
20. Spiryagin M, Wu Q, Cole C (2017) International benchmarking of longitudinal train dynamics simulators: benchmarking questions. Veh Syst Dyn 55(4):450–463
21. Wu Q, Spiryagin M, Cole C et al (2018) International benchmarking of longitudinal train dynamics simulators: results. Veh Syst Dyn 56(3):343–365
22. Bosso N, Magelli M, Zampieri N (2020) Long train dynamic simulation by means of a new in-house code. WIT Trans Built Environ 199:249–259
23. Magelli M, Zampieri N (2025) A novel approach for longitudinal train dynamics simulations with multibody codes. Veh Syst Dyn 63(5):978–996
24. Magelli M, Corrêa PHA, dos Santos AA (2025) Assessing the effect of modelling approaches for longitudinal train dynamics simulations with a multibody code. Proc Inst Mech Eng Part K J Multi-body Dyn 239(4):326–338
25. Pshinko OM, Ursulyak LV, Zhelieznov KI et al (2020) To the problem of train running safety. IOP Conf Ser Mater Sci Eng 985(1):012014
26. Di Gialleonardo E, Cazzulani G, Melzi S et al (2017) The effect of train composition on the running safety of low-flatcar wagons in braking and curving manoeuvres. Proc Inst Mech Eng Part F J Rail Rapid Transit 231(6):666–677
27. Bosso N, Magelli M, Trinchero R et al (2024) Application of machine learning techniques to build digital twins for long train dynamics simulations. Veh Syst Dyn 62(1):21–40
28. Bosso N, Cantone L, Gugliotta A, et al (2024) Introduction of digital twins in the longitudinal train dynamics simulation of freight train air brake operations. Advances in dynamics of vehicles on roads and tracks III. Springer Nature Switzerland, pp 297–306
29. Bosso N, Magelli M, Zampieri N (2023) Dynamical effects of the increase of the axle load on European freight railway vehicles. Appl Sci 13(3):1318
30. Cantone L (2022) Simulation of freight trains with up to three traction units in radio communication. IOP Conf Ser Mater Sci Eng 1214(1):012039
31. RFI Disposizione di esercizio N° 27:2007. Allegato 1: Standard di qualità geometrica del binario con velocità fino a 300 km/h
32. Kalker JJ (1982) A fast algorithm for the simplified theory of rolling contact. Veh Syst Dyn 11(1):1–13
33. International Union of Railways (1993) Benchmark problem results and assessment. Bogies with steered or steering wheelsets. Technical Report B 176/3
34. Suykens JAK, Van Gestel T, De Brabanter J et al (2002) Least squares support vector machines. World Scientific, Singapore
35. Trinchero R, Larbi M, Torun HM et al (2019) Machine learning and uncertainty quantification for surrogate models of integrated devices with a large number of parameters. IEEE Access 7:4056–4066
36. Ghojogh B, Crowley M (2019) The theory behind overfitting, cross validation, regularization, bagging, and boosting: tutorial. https://doi.org/10.48550/arXiv.1905.12787
37. Soleimani N, Trinchero R, Canavero FG (2023) Bridging the gap between artificial neural networks and kernel regressions for vector-valued problems in microwave applications. IEEE Trans Microw Theory Tech 71(6):2319–2332
38. De Brabanter K, Karsmakers P, Ojeda F et al (2011) LS-SVMlab toolbox user's guide version 1.8. ESAT-SISTA, Katholieke Universiteit Leuven, Belgium
39. European Committee for Standardization (2013) On the common safety method for risk evaluation and assessment and repealing regulation (EC) No 352/2009. EU 402/2013