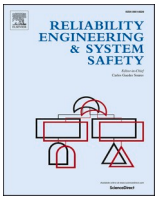
Safety assessment using computer experiments and surrogate modeling: Railway vehicle safety and track quality indices

Reliability Engineering and System Safety 229 (2023) 108856

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress



Safety assessment using computer experiments and surrogate modeling: Railway vehicle safety and track quality indices

João Neves Costa [a,*], Jorge Ambrósio [a], António R. Andrade [a], Daniel Frey [b]

[a] LAETA, IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[b] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

ARTICLE INFO

Keywords:
Design of experiments
Computer experiments
Gaussian processes
Track geometry quality
Rail vehicle safety

ABSTRACT

Mathematical modeling and advances in computation allow exploring multiple scenarios and studying the reliability and safety of transportation systems. Although track geometry directly impacts vehicle safety, the track quality indices used by infrastructure managers to assess tracks seldom consider vehicle dynamics. This work provides a design and analysis of computer experiments framework to model the relationships between track quality and vehicle safety. The framework considers input selection and pre-processing, vehicle responses and post-processing, input screening, surrogate modeling, sensitivity analysis, and safety assessment. This approach allows studying how track geometry parameters and other variables influence safety quantities. The framework is demonstrated with a case study that combines two European standards: the standard for track geometry quality, EN 13848, and the standard for vehicle acceptance, EN 14363. The case study considers different vehicle types, vehicle speed, track curvature, track flexibility, and track irregularities. The results show, for each safety quantity, which inputs are relevant. In particular, the sensitivity analysis indicates two influential inputs not considered in EN 13848 that could help assess track condition. Finally, an example illustrates how these surrogates can be used to find which safety quantities govern safety and define track geometry limits directly linked to vehicle safety.

1. Introduction

Policymakers are searching for cleaner transportation alternatives to reduce traffic congestion and decarbonize energy supply. Consequently, there is renewed interest in railways as a sustainable and energy-efficient transportation mode. However, turning railways into the backbone of a sustainable and multimodal transportation system requires guaranteeing track quality and safety of passengers and goods.

In a rail vehicle, the geometry of the wheels and the rails govern the motion of the wheelset and, consequently, the vehicle, while the wheel-rail contact forces provide guidance, traction, and braking. Therefore, ensuring track geometry quality is essential for safe and comfortable rides. Rail infrastructure managers periodically inspect tracks with vehicles that record track geometry parameters. Although the designations and characteristics of the geometry parameters change depending on country and standards, they usually involve the lateral and vertical displacements of the rails. The parameters consist of the longitudinal level, alignment, cross-level, twist, and gauge. The longitudinal level and alignment are the vertical and lateral displacements of the rail relative to its design geometry. The cross-level is the height difference between the rails. The twist is the difference between two cross-level measurements divided by the distance separating the measurements. In other words, the twist measures the rate of change of the cross-level. Finally, the gauge is the smallest distance between the inner edges of the rail heads.

Since rail networks span over km of track and geometry parameters are usually sampled every 0.25 m, or 1 ft, infrastructure managers use Track Quality Indices (TQIs) to assess the track quality. A TQI is a numerical value that characterizes the quality of a track section. Commonly, TQIs involve the standard deviations (SDs) of the geometry parameters. However, they can also consider other statistical quantities, the ratio between the length of the curve traced by the geometry parameter and the length of the track section (length-based TQI), and the proportion of isolated defects (threshold exceedance) in a section. For example, the European standard EN 13848, China Railway, Canadian National Railway, Polish State Railways, Banverket (Sweden), ProRail (Netherlands), and Indian Railways have their own TQIs involving SDs. On the other hand, the Federal Railroad Administration

* Corresponding author.
E-mail address: joao.n.costa@tecnico.ulisboa.pt (J. Neves Costa).

https://doi.org/10.1016/j.ress.2022.108856
Received 30 November 2021; Received in revised form 18 September 2022; Accepted 22 September 2022
Available online 24 September 2022

(USA) uses a length-based TQI [1], and the Polish State Railways has another TQI based on threshold exceedances [2]. The reader is referred to Ref. [3] for an updated and comprehensive overview of TQIs. Additionally, TQIs are classified as individual or aggregate indices. Individual TQIs involve a single geometry parameter, e.g., the SD of the longitudinal level, while aggregate TQIs involve several geometry parameters, e.g., the average of the SDs of the geometry parameters.

TQIs summarize the quality of a track section, which is appealing to both practitioners and researchers. Practitioners can compare them to limit values to plan maintenance actions [4], and researchers use them as a base to develop track degradation models or maintenance strategies. Andrade and Teixeira assume TQIs (EN 13848) are indicators of track geometry degradation and proposed representing them using hierarchical Bayesian models. This approach captures cross and spatial correlations of quality parameters and deterioration rates among consecutive track sections [5,6]. Bai et al. model the track as consecutive maintenance units, whose defects are quantified in terms of TQIs (China Railway), and use a Markov process with heterogeneous factors (structure and tonnage) to evaluate track deterioration [7]. Cárdenas-Gallo et al. proposed forecasting track degradation using an ensemble classifier that encompasses defect growth over time as a gamma process and defect classification using binary logistic regression and support vector machines [8].

Andrade and Teixeira point out that TQIs limits are usually seen as constraints in optimization and decision-making and are seldom studied. Thus, they explore the costs and impacts associated with track maintenance and renewal actions for different alert limits of TQIs, and present new limits with lower lifecycle costs [9]. Sharma et al. devise a data-driven condition-based policy for inspection and maintenance of track geometry. They consider the track a Markovian system, use TQIs (Federal Railroad Administration) as track section state indicators, and define a Markov decision process that considers maintenance and geometry defect repair costs to obtain optimal maintenance policies [10]. Sresakoolchai and Kaewunruen used machine learning to identify track component defects based on track geometry parameters. They used deep neural networks to predict different types of track defects, *k*-means clustering to group track component defects according to curve radius and track gradient, and association rules to analyze the likelihood of two defects on 50 m long track sections [11]. Mohammadi et al. explored machine learning techniques to predict rail defects based on tonnage, geometry parameters, and other quantities involving the geometry parameters, such as spectral density, TQIs (Federal Railroad Administration), maxima, SDs, and time-trends [12]. The authors found that dimensionality reduction using recursive feature elimination, adaptive synthetic sampling to overcome the imbalance of the dataset, and extreme gradient boosting (with Bayesian optimization of the hyperparameters) provides acceptable results, given the challenging data set.

Despite their usefulness for infrastructure managers, TQIs still present limitations, and researchers keep proposing new ones. Chudzikiewicz et al. proposed a TQI that uses axle box accelerations and GPS to locate track faults [13]. This TQI allows continuous track monitoring using service vehicles equipped with accelerometers, unlike most TQIs, which rely on periodic measurements from inspection vehicles. Offenbacher proposed a new aggregate index, the TUG-TQI [3]. The TUG-TQI is the average of the Federal Railroad Administration's TQIs after normalizing the longitudinal level, alignment, cross-level, twist, and gauge by predefined thresholds. The authors show that this TQI is less sensitive to gauge widening and transition curves and claim that normalizing the geometry parameters avoids assigning weights to the different parameters. Lasisi and Attoh-Okine criticize the subjectiveness of aggregate TQIs and propose a data-driven approach using principal components as an aggregate TQI [14]. Their analysis shows that the first three principal components outperform other aggregate TQIs (every geometry parameter, Polish State Railways, Indian Railways, China Railway, and Canadian National Railway) when training models to identify track defects.

Something the literature often disregards is the relationship between TQIs and vehicle response. Validation of new TQIs usually consists of a mix between the researchers' expectations and a benchmark against existing TQIs, which makes it unclear if the new metric provides any tangible benefit from the point of view of vehicle performance. Moreover, neither the TQI limits nor their rationale is usually addressed. The reasons for the omissions could be twofold. First, the available data is highly biased towards safety because infrastructure managers maintain the track geometry parameter within acceptable limits. Second, although the geometry parameter data is available, data concerning the response of service vehicles is unavailable. The relationships between TQIs and the vehicle response can be obtained by changing the track geometry parameters, and other relevant variables, and assessing the vehicle response. However, physical experiments to this end are unfeasible. Since there is an extensive and mature literature on computational methods for railway dynamics, whose results show excellent agreement with those of physical experiments, computer experiments are a reasonable option to seek these relationships. Although computer experiments allow studying a system in ways that would otherwise be difficult or impossible, they still present limitations, e.g., computing power and time are finite resources. Design and Analysis of Computer Experiments (DACE) provide strategies to find reliable and accurate predictors using computer experiments.

The DACE workflow can be summarized into three steps: (i) choosing a design, (ii) evaluating the computer code for the design, and (iii) fitting a model to the data. Although there are many options for choosing designs and fitting models, the canonical approach in the DACE is to use variations of Latin Hypercube Designs (LHDs) [15,16] and fit Gaussian Process (GP) models [17–19]. LHDs are attractive because they are easy to compute, provide uniform marginals, and can be combined with distance or discrepancy metrics to obtain space-filling designs. GPs are popular for several reasons. They are linear predictors capable of representing highly non-linear functions and provide uncertainty estimates of predictions that can be used for sequential design/active learning [20–24]. GPs provide efficient and reliable surrogates [25–27], enabling sensitivity analysis [26–28] and black-box optimization [19,27]. Finally, they can be calibrated using both computer and physical experiments [20,27,29,30].

Since TQIs limits and the relationships between TQIs and the vehicle are rarely addressed, this work proposes a DACE framework to find and represent the relationships between track quality and vehicle safety. The framework is divided into six steps: (i) input selection and pre-processing, (ii) vehicle responses and post-processing, (iii) input screening, (iv) surrogate modeling, (v) sensitivity analysis (and estimation of alternative models), and (vi) safety assessment. The framework is demonstrated with a case study that uses TQIs from the European standard for track geometry quality and safety quantities from the European standard for vehicle acceptance. This framework allows (i) identifying, depending on the safety quantity, which inputs are relevant and could be used as TQIs, and (ii) finding TQIs limits related to vehicle safety.

The structure is this paper is the following. This section presented the literature review and relevance of the topic. Section 2 overviews the relevant parts of the standards used in this work. Additionally, it describes input selection and pre-processing, and vehicle response and post-processing. Section 3 covers the methods and models used in the framework. Section 4 discusses the case study results, namely, non-influential inputs, surrogate models, sensitivity analysis, alternative models, and safety assessment. Finally, Section 5 presents the conclusions and future work.

2. Input selection and vehicle responses

This section overviews the relevant parts of the European standards for track geometry quality and vehicle acceptance and explains the data generation process for the surrogate models. It covers input selection,

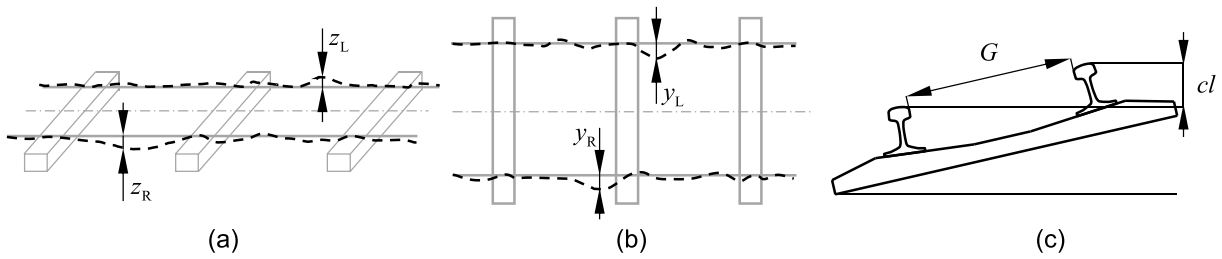


Fig. 1. (a) Longitudinal level of the left and right rail, (b) alignment of the left and right rail, (c) gauge and cross-level.

Table 1
Cumulative distribution function of the TQCs [4].

Track Quality Class	A	B	C	D	E
Cumulative distribution function	$F(\sigma) < 0.1$	$0.1 \leq F(\sigma) < 0.3$	$0.3 \leq F(\sigma) < 0.7$	$0.7 \leq F(\sigma) < 0.9$	$F(\sigma) \geq 0.9$

Table 2
TQCs limit values for σ_z and σ_y [4].

v [km/h]	Track Quality Classes [mm]									
	σ_z					σ_y				
	A	B	C	D	E	A	B	C	D	E
(0, 80]	1.25	1.75	2.75	3.75	> 3.75	0.90	1.25	1.95	2.70	> 2.70
(80, 120]	0.75	1.10	1.80	2.50	> 2.50	0.50	0.70	1.05	1.45	> 1.45
(120, 160]	0.65	0.85	1.40	1.85	> 1.85	0.45	0.55	0.75	1.00	> 1.00
(160, 230]	0.60	0.75	1.15	1.60	> 1.60	0.40	0.50	0.70	0.90	> 0.90
(230, 300]	0.40	0.55	0.85	1.15	> 1.15	0.35	0.40	0.50	0.65	> 0.65

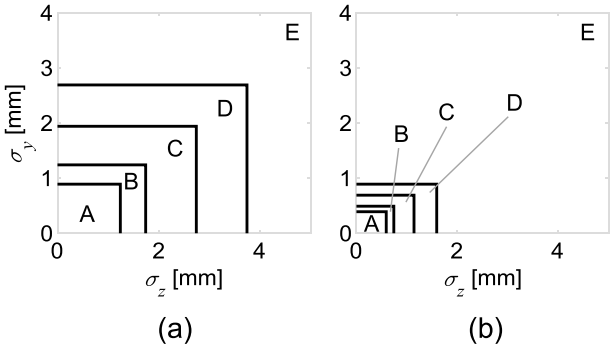


Fig. 2. TQCs envelopes for speed intervals (a) (0, 80] and (b) (160, 230] km/h.

pre-processing, vehicle responses, and post-processing for the computer experiments. Note that some aspects of the framework depend on the author's decisions and available data, and can be changed to fit the researcher's needs. For example, the authors chose European standards because they apply to several countries, but the framework is equally valid for other standards and maintenance guidelines.

2.1. Overview of EN 13848-1, EN 13848-6, and EN 14363

In Europe, standards or codes establish normative and safety limits for railways. Two different standards of particular interest to this work are EN 13848 [4,31,32] and EN 14363 [33]. EN 13848 defines minimum requirements for track geometry quality and covers several aspects of the track. EN 13848-1 defines six track geometry parameters that need to be collected by track recording vehicles [31]: longitudinal level, z_r, alignment, y_r, cross-level, cl, gauge, G, and twist, *twist*. The subscript r can be L to denote the left rail or R to denote the right rail. Fig. 1(a) and (b) depict the longitudinal level and alignment of the rails, while Fig. 1 (c) illustrates the cross-level and gauge. Recall that twist is the difference between two cross-levels divided by their distance apart (3 m in this work).

EN 13848-6 characterizes track geometry quality using Track Quality Classes (TQCs) [4]. TQCs are functions of speed, v, and TQIs. In other words, TQCs define intervals of allowable TQIs. Although the standard

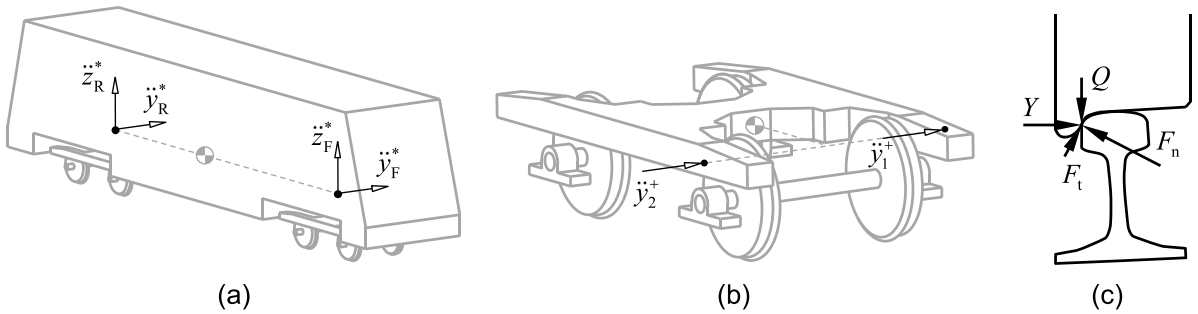


Fig. 3. (a) Lateral and vertical acceleration of the vehicle body above the bogies, (b) lateral accelerations of the bogie above the axle boxes, and (c) lateral and vertical decomposition of the wheel-rail forces.

Table 3
Radii intervals of the curve types.

Curve	R [m]	Designation	Ref.
Straight, very large, and large radius curves	[900, 10000]	I_1	
Medium radius curves	[600, 900)	I_2	[35]
Small radius curves	[400, 600)	I_3	[33, 35]
Very small radius curves	[250, 400)	I_4	[33, 35]

Table 4
Statistics of the cross-level regression coefficients.

	Estimate	t-statistic	p-value
Intercept	-7.5157E-05	-1.1764E-01	9.0636E-01
$\kappa = 1/R$	3.7365E+01	7.6147E+01	<1.0000E-16
v	5.0930E-04	5.0571E+01	<1.0000E-16
R^2		0.9041	
MSE		0.0014	

overviews several TQIs used by European railway networks or infrastructure managers, it uses SDs because of their simplicity and understandability. Particularly, it uses the SD of the longitudinal level and the SD of the alignment, referred to hereafter as σ_z and σ_y. TQCs in EN 13848-6 were defined based on a survey of track irregularities to railway infrastructure managers across Europe [34], and to the best of the author's knowledge, there is no direct relationship between TQCs and vehicle safety. The TQCs were defined according to the cumulative distribution in Table 1, which resulted in the limits in Table 2. Fig. 2 shows the TQCs envelopes formed by σ_z and σ_y for speed intervals (0, 80] and (160, 230] km/h. In general, TQCs envelopes shrink as speed increases.

Standard EN 14363 defines tests, measurements, post-processing, and limit values for the acceptance of railway vehicles. This work does not follow EN 14363 exactly because it has a different goal. Instead, it borrows quantities from the standard to measure safety, namely, the lateral and vertical accelerations of the vehicle body above the bogie, $\ddot{y}^*$ and $\ddot{z}^*$, respectively, and the lateral acceleration of the bogie frame above the axle boxes, $\ddot{y}^+$. Fig. 3(a) and (b) illustrate the position where these quantities are measured. Another quantity borrowed from the standard is the ratio between the lateral wheel force, Y, and the vertical wheel force, Q, depicted in Fig. 3(c). The quantity Y/Q is referred to hereafter as wheel force ratio. The wheel force ratio characterizes the potential for flange climb derailment, which occurs when the wheel flange climbs on top of the rail. Intuitively, high values of wheel force ratio are associated with high lateral contact force, reduced vertical force, or both. Vehicles are prone to wheel flange derailment on curves because their inertia pushes them against the outer (high) rail of the curve.

2.2. Input selection and pre-processing

The inputs of the experiment are the vehicle speed, v, curve radius, R, vertical stiffness of the track, k_{vert}, and the SDs of the longitudinal level, alignment, twist, and gauge irregularity, σ_z, σ_y, σ_{twist}, and $\sigma_{\Delta G}$, respectively. The gauge irregularity is the difference between the measured and nominal gauge. v, σ_z, and σ_y are included because EN 13848-6 defines σ_z and σ_y as TQIs, and TQCs as a function of the three variables. Despite EN 13848 defining limit values for isolated defects of twist and gauge irregularity, it does not consider the SDs of these irregularities to assess track geometry quality. Therefore, σ_{twist} and $\sigma_{\Delta G}$ are included in this study to determine if and how they influence safety.

Curve radius, R, is included because EN 14363 requires testing the vehicles on different curves. The standard defines that vehicles must be tested on straight segments, very large radius, large radius, small radius, and very small radius curves. However, the standard only provides radii intervals for the small radius and very small radius curves. Leaflet UIC 518 [35], which is similar to EN 14363, defines a radii interval for medium radius curves. Since there are no reference limits for straights, very large radius, and large radius curves, this work considers a single radii interval for these. Table 3 lists the four radii intervals considered in this work, which are denoted from now on as I_i, with $i \in \{1,2,3,4\}$. The passenger vehicle study considers the interval $I_1 \cup I_2 \cup I_3$, while the locomotive and freight studies consider the interval $I_1 \cup I_2 \cup I_3 \cup I_4$. Interval I_4 is not included in the analysis of the passenger vehicle because combinations of high speed and small radius would lead to derailments in which the wheelset goes entirely off the rails. The quantities used to assess derailment measure the imminence of derailment and are not expressive after it has happened. Moreover, these simulations may harm the model because they end prematurely and, consequently, their outputs are not meaningful.

Designs with uniform marginals are common in computer experiments. However, uniformly sampling the radius leads to a different number of samples in each radii interval because the intervals have different lengths. Therefore, the marginal associated with the radius is transformed to obtain an equal number of samples in each radii interval. The cumulative distribution function of the radius, $U = F(R)$, is defined as a linear piecewise function where each piece has equal probability. Finally, the radius sample is obtained by applying the inverse of the cumulative distribution to the marginal, i.e., $R = F^{-1}(u)$, where u is the marginal.

Besides vehicle speed, the other variables introduced so far are related to track geometry. This work also hypothesizes that track flexibility-related variables could be influential and, thus, also considers the vertical track stiffness, k_{vert}, as an input.

The track model has an Iberian gauge and new UIC60 rail profiles with a 1/20 rail inclination. The geometry of the track starts with a 100 m long straight segment, followed by a 500 m transition curve, and ends with a 5.4 km long right-hand curve with constant curvature $\kappa = 1/R$. In the straight segment, the stiffness of the track changes linearly from 100 kN/mm (VAMPIRE Pro's default value) to k_{vert}.

The cross-level is defined using a multiple linear regression model estimated using speed and curvature data from the Portuguese railway network and limits from EN 13803 [36]. Table 4 summarizes the statistics of the regression. The intercept was dropped from the regression because it is not significant at the 0.05 level. Therefore, the cross-level is given by

$$cl_{\text{reg}} = 3.7365 \times 10^{1}(1/R) + 5.0930 \times 10^{-4}v, \tag{1}$$

where, inputs R and v are in m and km/h, respectively. The cross-level in the constant curvature segment is computed as

$$cl = \begin{cases} \min\left(cl_{\text{reg}}, 0.9 \times 10^{-3}(R-50)\right), & R < 320\text{m} \\ \min\left(cl_{\text{reg}}, 0.185\right), & R \geq 320\text{m} \end{cases}, \tag{2}$$

to ensure compliance with EN 13803. The upper limit of cross-level for $R < 320$ m corresponds to the cross-level limit for torsionally-stiff freight wagons [36], while the upper limit of cross-level for $R \geq 320$ m corresponds to the cross-level normal limit [36]. Both limits are for the Iberian gauge. Finally, the transition curve is a clothoid long enough to satisfy EN 13803's requirements for the range of radii involved in the experiment.

The track irregularities, z, y, Δcl, and ΔG are generated using a multivariate ARMA process [37] and appropriately scaled to the target SDs. Note that controlling σ_{twist} requires transforming the cross-level irregularity, Δcl, to *twist*, scaling it, and transforming it back to Δcl. Moreover, the irregularities start at the 700 m mark and cover the remaining length of the track. Fig. 4(a–d) depicts the vehicle speed, track curvature, track stiffness, and the longitudinal level of the left rail as functions of track length.

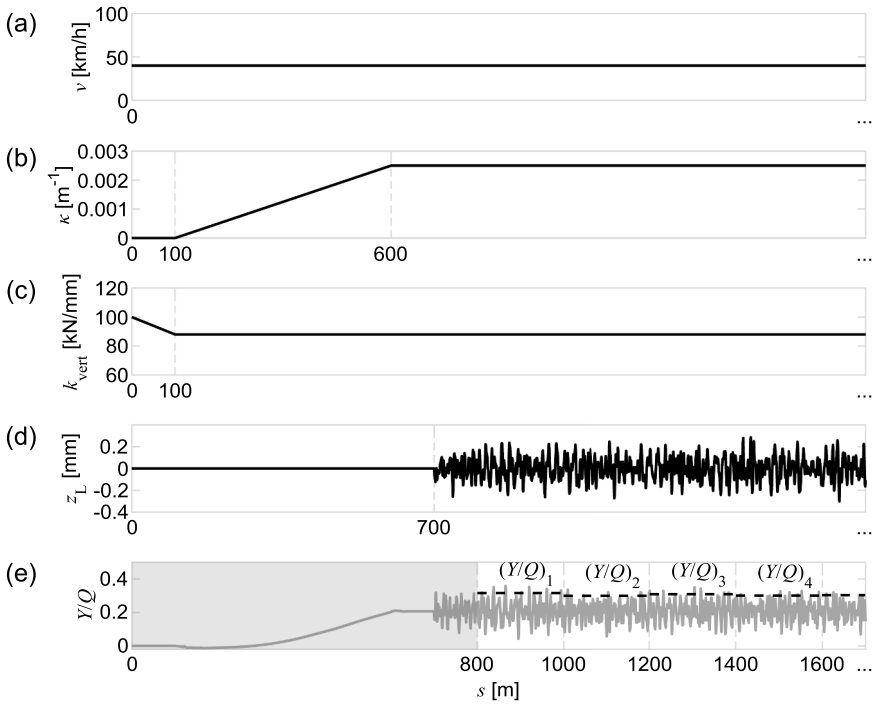


Fig. 4. Simulation 4 of the elementary effects analysis of the passenger vehicle: (a) speed, (b) curvature, (c) track stiffness, (d) longitudinal level of the left rail, and (e) wheel force ratio as functions of track length.

2.3. Vehicle responses and post-processing

The limits values established in EN 14363 are directly related to safety and comfort and could help model the relationship between track quality and vehicle safety. Therefore, the selected responses include the quantities from EN 14363, introduced in Section 2.1, and wheel unloading, $\Delta Q/Q_0$. The lateral accelerations of the bogie frame above the left and right axle boxes of the leading wheelset, $\ddot{y}_{\mathrm{L}}^{+}$ and $\ddot{y}_{\mathrm{R}}^{+}$, are referred to hereafter as lateral accelerations of the left and right side of the bogie. The lateral and vertical accelerations of the vehicle body above the front bogie, $\ddot{y}^{*}$ and $\ddot{z}^{*}$, are referred to from now on as the lateral and vertical accelerations of the vehicle body. The last quantity from EN 14363 is the wheel force ratio on the left (outer) wheel Y/Q. Since wheel force ratio is not a meaningful derailment measure on straight and large radius curves because sustained flange contact is less likely, wheel unloading of the left and right rail, $(\Delta Q/Q_0)_{\mathrm{L}}$ and $(\Delta Q/Q_0)_{\mathrm{R}}$, are also included as quantities. Wheel unloading characterizes the reduction of vertical support on a wheel and is given by $(Q_0 - Q)/Q_0$, where Q_0 is the static wheel load. High values of wheel unloading make the vehicle more susceptible to sudden lateral forces due to discrete track irregularities. This set of quantities is referred to hereafter as safety quantities. The safety quantities are filtered according to EN 14363. Y/Q is filtered with a low-pass 20 Hz filter and a 2 m sliding mean, $\ddot{y}^{+}$ with a 10 Hz low-pass filter, $\ddot{y}^{*}$ with a low-pass 6 Hz filter, and $\ddot{z}^{*}$ with a $[0.4, 4]$ Hz band-pass filter. $\Delta Q/Q_0$ is filtered with a low-pass 20 Hz filter and a 2 m sliding mean, the same filters used in EN 14363 for wheel forces related quantities.

Since the goal is to build surrogate models that represent safety quantities, a reasonable measure for worst-case scenarios is extreme values in the response. However, choosing the extreme value of the response discards most of the data and might overestimate safety. For example, due to the stochastic nature of track irregularities, an isolated irregularity with an exceptionally high value might lead to an outlier in the dynamic response and, consequently, an incorrect severity assessment. Thus, a strategy is employed to avoid this misrepresentation. The process consists of (i) dividing the track into 200 m sections, (ii) calculating the extreme value for each section as the maximum of the absolute values of the 2.5% and 97.5% percentiles, i.e., $y_i = \max(|P^{i}_{2.5\%}|, |P^{i}_{97.5\%}|)$, where y is the quantity under study and i is the section, and (iii) calculating the maximum among all sections, $y_{\max}$. The track sections are 200 m long to match the distance used to calculate the SDs of the longitudinal level and alignment in EN 13848-6.

Fig. 4 depicts the inputs and one output (wheel force ratio) of a simulation. The goal of gradually introducing the inputs one at a time over the track length is to avoid sudden changes in the responses and ensure the vehicle is close to steady-state operation in the constant curvature section, which is the relevant track length for this study. Data before the 800 m mark are considered transient and discarded. In Fig. 4 (e), the grayed-out area represents the discarded section, and the dashed lines mark the extreme values in the 200 m sections, i.e., $(Y/Q)_i$.

The end goal is to find the following surrogate models

$$\begin{aligned} (Y/Q)_{\max} &= y_1(\mathbf{x}) & \ddot{y}^{+}_{\mathrm{R,max}} &= y_5(\mathbf{x}) \\ (\Delta Q/Q_0)_{\mathrm{L,max}} &= y_2(\mathbf{x}) & \ddot{y}^{*}_{\max} &= y_6(\mathbf{x}) \\ (\Delta Q/Q_0)_{\mathrm{R,max}} &= y_3(\mathbf{x}) & \ddot{z}^{*}_{\max} &= y_7(\mathbf{x}) \\ \ddot{y}^{+}_{\mathrm{L,max}} &= y_4(\mathbf{x}) \end{aligned}, \tag{3}$$

where, $\mathbf{x} = [v, R, k_{\mathrm{vert}}, \sigma_z, \sigma_y, \sigma_{twist}, \sigma_{\Delta G}]$. Note that the "computer code" in this work encompasses the vehicle simulation and the post-processing step to find extreme values.

3. Methods and models

This section covers the methods and models used in the framework. Specifically, (i) the elementary effects method for input screening, (ii) GPs for surrogate modeling, and (iii) Sobol indices for the sensitivity analysis.

3.1. Elementary effects

The framework screens inputs using the factorial sampling plan introduced by Morris [38] with the enhancements proposed by Campolongo et al. [39]. The plan consists of multiple randomized

Table 5
Lower and upper bounds of the input space for the elementary effects method.

Vehicle	v [km/h]	R [m]	k_{vert} [kN/mm]	σ_z [mm]	σ_y [mm]	σ_{twist} [mm/m]	$\sigma_{\Delta G}$ [mm]
Passenger vehicle	[40, 200]	[500, 10,000]	[88, 316]	[0.1, 3.5]	[0.1, 2.5]	[0.1, 2.5]	[0.1, 2.5]
Locomotive, loaded and empty freight	[40, 100]	[250, 10,000]					

Table 6
Lower and upper bounds of the input space for GP models estimation.

Vehicle	v [km/h]	R [m]	σ_z [mm]	σ_y [mm]	$\sigma_{\Delta twist}$ [mm/m]
Passenger vehicle	[40, 220]	[400, 10,000]	[0.1, 4]	[0.1, 3]	[0.1, 3]
Locomotive, loaded and empty freight	[40, 100]	[250, 10,000]			

one-factor-at-a-time designs. This way, changes in the output are unambiguously attributed to changes in the individual inputs. This method requires the function being studied to be deterministic.

Assume the input space, $\mathscr{D}$, is restricted to a p-level grid discretization of a k-dimensional hypercube, $[0,1]^k$, i.e.,

$$x_j \in \left\{0, \frac{1}{p-1}, \frac{2}{p-1}, \ldots, 1\right\}, \text{for} j = 1, \ldots, k. \tag{4}$$

The elementary effect of the jth input is defined as [38]

$$d_j(\mathbf{x}) = \frac{y(x_1, x_2, \ldots, x_j + \Delta, \ldots, x_k) - y(\mathbf{x})}{\Delta}, \tag{5}$$

where, Δ is a multiple of $1/(p-1)$ and $\mathbf{x} \in \mathscr{D}$ (except that $x_j \leq 1 - \Delta$). In other words, if the output y is at least once differentiable, the derivative provides information on the effect of x_j on y. A distribution of the absolute values of the elementary effects associated with input j can be obtained by random sampling $\mathbf{x}$ on the p-level grid, i.e., $|d_j(\mathbf{x})| \sim G_j$. The mean of G_j, μ^*, measures the importance of input j and is a good proxy of the total sensitivity index [39].

Morris' sampling plan is economical and uses some evaluations of y to compute more than one elementary effect. The plan generates r trajectories of $k+1$ points, where each trajectory is represented by [38]

$$\mathbf{B}^* = \left(\mathbf{J}_{k+1,1}\mathbf{x}^* + \Delta/2\left[\left(2\mathbf{B} - \mathbf{J}_{k+1,k}\right)\mathbf{D}^* + \mathbf{J}_{k+1,k}\right]\right)\mathbf{P}^*, \tag{6}$$

where, $\mathbf{J}_{k+1,1}$ is a $(k+1)$ column vector of 1s, and $\mathbf{x}^*$ is a randomly chosen value of $\mathbf{x}$ such that the j^{th} input is defined as

$$x_j^* \in \left\{0, \frac{1}{p-1}, \frac{2}{p-1}, \cdots, 1-\Delta\right\}, \text{for } j = 1, \cdots, k. \tag{7}$$

$\mathbf{B}$ is a matrix of 0s and 1s, with the key property that for every column $j = 1, \ldots, k$ there are two rows of $\mathbf{B}$ that only differ in their jth entry (a convenient choice for $\mathbf{B}$ is a strictly lower triangular matrix of 1s). $\mathbf{D}^*$ is a k-dimensional diagonal matrix whose elements are either 1 or –1 with equal probability, and $\mathbf{P}^*$ is a $k \times k$ random permutation matrix in which each row contains one element equal to 1, and all others equal to 0, and no two columns have 1s in the same position (random permutation of the columns of the identity matrix).

Poor coverage of the input space due to random sampling is avoided by generating a high number of trajectories (between 500 and 1000) and selecting the subset of r trajectories with the maximum spread over the input space [39]. The final sampling plan, or design, is obtained by concatenating r trajectories, i.e.,

$$\mathbf{X} = \begin{bmatrix} \mathbf{B}_1^* \\ \mathbf{B}_2^* \\ \vdots \\ \mathbf{B}_r^* \end{bmatrix}. \tag{8}$$

3.2. Gaussian process model

The output of a computer code can be represented by a stochastic model as [18]

$$y(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T\boldsymbol{\beta} + z(\mathbf{x}), \tag{9}$$

where, $\mathbf{f}(\mathbf{x})^T = [f_1(\mathbf{x}), \cdots, f_p(\mathbf{x})]$ is a vector of known functions, $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_p]^T$ is a vector of unknown regression coefficients, and $z(\mathbf{x})$ is a zero-mean stationary GP. The regression term, $\mathbf{f}(\mathbf{x})^T\boldsymbol{\beta}$, is often assumed constant or zero because the true flexibility of the model comes from the GP, $z(\mathbf{x})$.

Let $\mathbf{X}$, $\mathbf{y}_\mathbf{X}$, $\mathbf{F}_\mathbf{X}$, which denote respectively a design with n points, a vector comprising the outputs of the code at $\mathbf{X}$, and a matrix comprising the evaluations of $\mathbf{f}(\mathbf{x})^T$ at $\mathbf{X}$, be represented by

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}, \ \mathbf{y} = \begin{bmatrix} y(\mathbf{x}_1) \\ \vdots \\ y(\mathbf{x}_n) \end{bmatrix}, \ \mathbf{F}_\mathbf{X} = \begin{bmatrix} \mathbf{f}(\mathbf{x}_1)^T \\ \vdots \\ \mathbf{f}(\mathbf{x}_n)^T \end{bmatrix}. \tag{10}$$

The best linear unbiased predictor of $y(\mathbf{x})$ is given by [17,18]

$$\hat{y}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^T\hat{\boldsymbol{\beta}} + \mathbf{k}(\mathbf{x})^T\mathbf{K}_\mathbf{X}^{-1}(\mathbf{y}_\mathbf{X} - \mathbf{F}_\mathbf{X}\hat{\boldsymbol{\beta}}), \tag{11}$$

where, $\hat{\boldsymbol{\beta}} = (\mathbf{F}_\mathbf{X}^T\mathbf{K}_\mathbf{X}^{-1}\mathbf{F}_\mathbf{X})^{-1}\mathbf{F}_\mathbf{X}^T\mathbf{K}_\mathbf{X}^{-1}\mathbf{y}_\mathbf{X}$ is the generalized least squares estimator of $\boldsymbol{\beta}$, $\mathbf{k}(\mathbf{x})^T = [C(\mathbf{x}, \mathbf{x}_1), \ldots, C(\mathbf{x}, \mathbf{x}_n)]$ is a vector comprising the correlations between $\mathbf{x}$ and $\mathbf{X}$, and $\mathbf{K}_\mathbf{X}$ is a matrix comprising the correlations among $\mathbf{X}$ defined as [19]

Table 7
Mean squared prediction error of the linear models and ratio of the mean squared prediction errors.

	Passenger vehicle		Locomotive		Loaded freight wagon		Empty freight wagon	
	$MSPE_{\text{linear}}$	$\frac{MSPE_{\text{linear}}}{MSPE_{\text{GP}}}$	$MSPE_{\text{linear}}$	$\frac{MSPE_{\text{linear}}}{MSPE_{\text{GP}}}$	$MSPE_{\text{linear}}$	$\frac{MSPE_{\text{linear}}}{MSPE_{\text{GP}}}$	$MSPE_{\text{linear}}$	$\frac{MSPE_{\text{linear}}}{MSPE_{\text{GP}}}$
$(Y/Q)_{\max}$	0.0029	0.9	0.0023	1.1	0.0028	2.4	0.0026	1.3
$(\Delta Q/Q)_{\text{L,max}}$	0.0094	5.8	0.0013	1.8	0.0061	1.2	0.0026	0.9
$(\Delta Q/Q)_{\text{R,max}}$	0.0047	3.4	0.001	2.3	0.004	1.2	0.0018	0.7
$\ddot{y}^+_{\text{L,max}}$	0.3926	1.9	0.2307	1.8	0.0767	0.9	0.0914	1
$\ddot{y}^+_{\text{R,max}}$	0.3856	1.7	0.2254	1.8	0.0736	0.8	0.0875	0.9
$\ddot{y}^*_{\max}$	0.6425	1.6	0.0951	2	0.0833	1	0.0678	1.6
$\ddot{z}^*_{\max}$	0.0027	0.6	0.0075	0.7	0.0265	0.8	0.0625	0.8

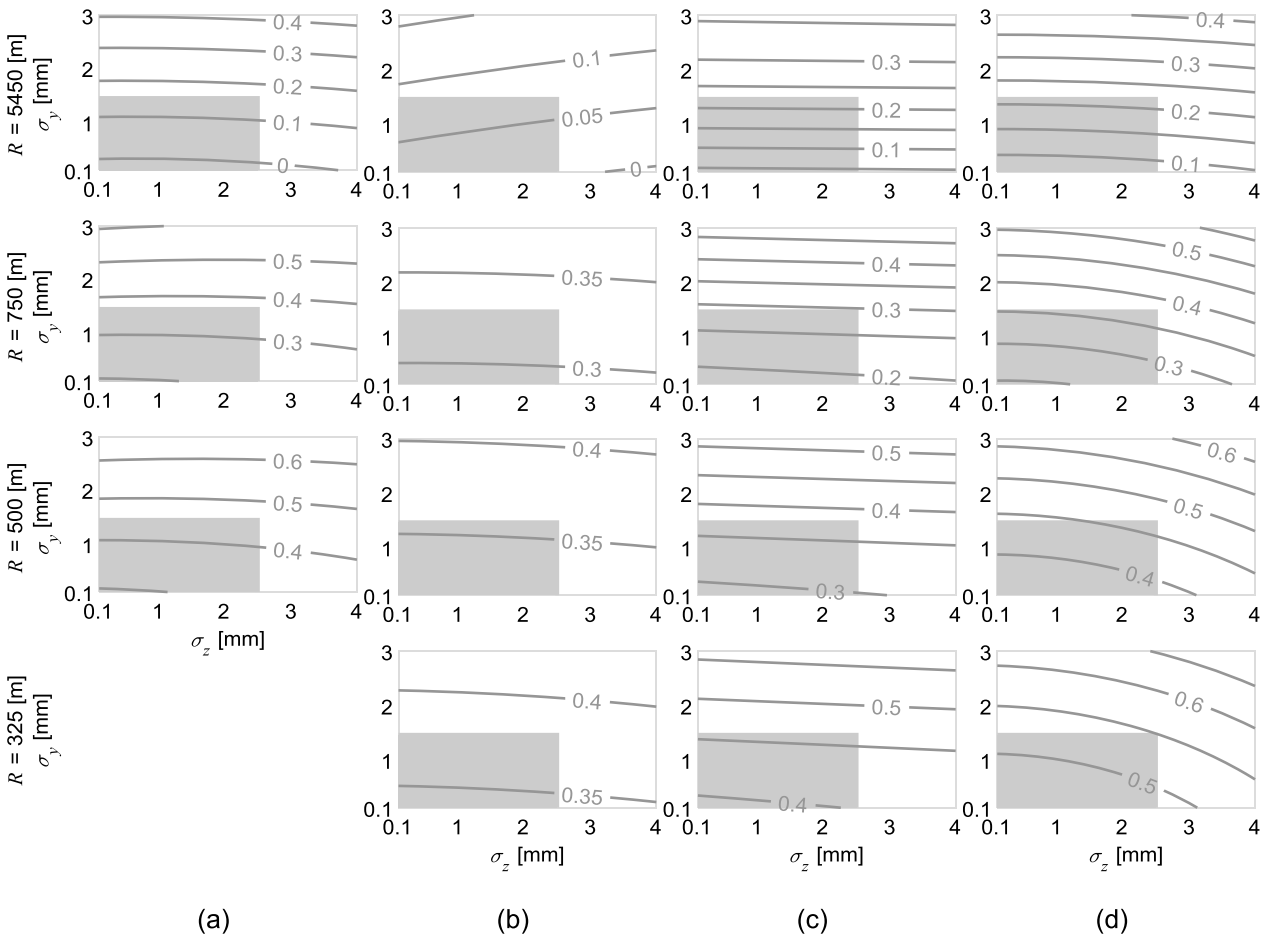


Fig. 5. $(Y/Q)_{\max}$ of the (a) passenger vehicle, (b) locomotive, (c) loaded freight wagon, and (d) empty freight wagon at $v = 90$ km/h.

Table 8
Limit values of the safety quantities.

	$(Y/Q)_{\max}$	$(\Delta Q/Q_0)_{\max}$	$\ddot{y}^{+}_{\max}$ [m/s^2]	$\ddot{y}^{*}_{\max}$ [m/s^2]	$\ddot{z}^{*}_{\max}$ [m/s^2]
Passenger Locomotive	0.8	0.6, 0.9	≈ 11	$\begin{cases} 2.6, & R \in [250, 400) \\ 2.8, & R \in [400, 600) \\ 3.0, & R \in [600, \infty) \end{cases}$	3.0
Freight Wagon				3.0	5.0
Ref.	[33]	[49]	[33]	[33]	[33]

$$K(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} C(\mathbf{x}_i, \mathbf{x}_j) + g, & \text{for } i = j \\ C(\mathbf{x}_i, \mathbf{x}_j), & \text{for } i \neq j, \text{even if } \mathbf{x}_i = \mathbf{x}_j \end{cases}, \tag{12}$$

where, $g > 0$ is a parameter that smooths out noise. Finally, $C(\mathbf{x}, \mathbf{x}')$ is the separable Gaussian correlation function, which is given by [40]

$$C(\boldsymbol{\theta}, \mathbf{x}, \mathbf{x}') = \exp\left(- \sum_{j=1}^{k} \theta_j \left| x_j - x_j' \right|^2 \right), \tag{13}$$

where, $\theta_j > 0$ is the rate parameter of the jth input and controls the decay of the correlation. Intuitively, points close together are more correlated than points apart. This correlation function implies a smooth and continuous process.

The minimum Mean Squared Prediction Error (MSPE) of $y(\mathbf{x})$ is given by [17,18]

$$\widehat{s}^2(\mathbf{x}) = \sigma^2 \left(1 + g - \mathbf{k}(\mathbf{x})^T \mathbf{K}_{\mathbf{X}}^{-1} \mathbf{r}(\mathbf{x}) + \mathbf{h}^T \mathbf{Q}^{-1} \mathbf{h}\right), \tag{14}$$

where, σ^2 is the variance, $\mathbf{h} = \mathbf{f}(\mathbf{x}) - \mathbf{F}_{\mathbf{X}}^T \mathbf{K}_{\mathbf{X}}^{-1} \mathbf{k}(\mathbf{x})$, and $\mathbf{Q} = \mathbf{F}_{\mathbf{X}}^T \mathbf{K}_{\mathbf{X}}^{-1} \mathbf{F}_{\mathbf{X}}$. The square root of Eq. (14) provides a measure of the uncertainty of the prediction, which increases quadratically with the distance between $\mathbf{x}$ and $\mathbf{X}$. The unknown parameters σ^2, $\boldsymbol{\theta}$, and g are estimated using maximum likelihood [19].

3.3. Sobol indices

This work uses the variance-based method proposed by Sobol, which decomposes a deterministic function $y(\mathbf{x})$ into summands of functions of increasing dimension as [19]

$$y(\mathbf{x}) = y_0 + \sum_{j=1}^{k} y_j(x_j) + \sum_{i=1}^{k} \sum_{j>i}^{k} y_{ij}(x_i, x_j) + \ldots + y_{1\ldots k}(x_1, \ldots x_k). \tag{15}$$

If $y(\mathbf{x})$ is square-integrable, all the terms in the decomposition are orthogonal and have zero mean, and the inputs are uncorrelated, the unconditional variance of $y(\mathbf{x})$ can be decomposed as [19]

$$\mathbb{V} = \sum_{j=1}^{k} \mathbb{V}_j + \sum_{i=1}^{k} \sum_{j>i}^{k} \mathbb{V}_{ij} + \ldots + \mathbb{V}_{1\ldots k}, \tag{16}$$

where,

$$\begin{aligned} \mathbb{V} &= \mathbb{V}[y(\mathbf{x})] \\ \mathbb{V}_j &= \mathbb{V}\left[\mathbb{E}\left[y \middle| x_j\right]\right] \\ \mathbb{V}_{ij} &= \mathbb{V}\left[\mathbb{E}\left[y \middle| x_i, x_j\right]\right] - \mathbb{V}_i - \mathbb{V}_j \end{aligned}, \tag{17}$$

and $\mathbb{E}$ and $\mathbb{V}$ denote expectation and variance, respectively. Higher-order variances are derived recursively in a similar fashion. In sum-

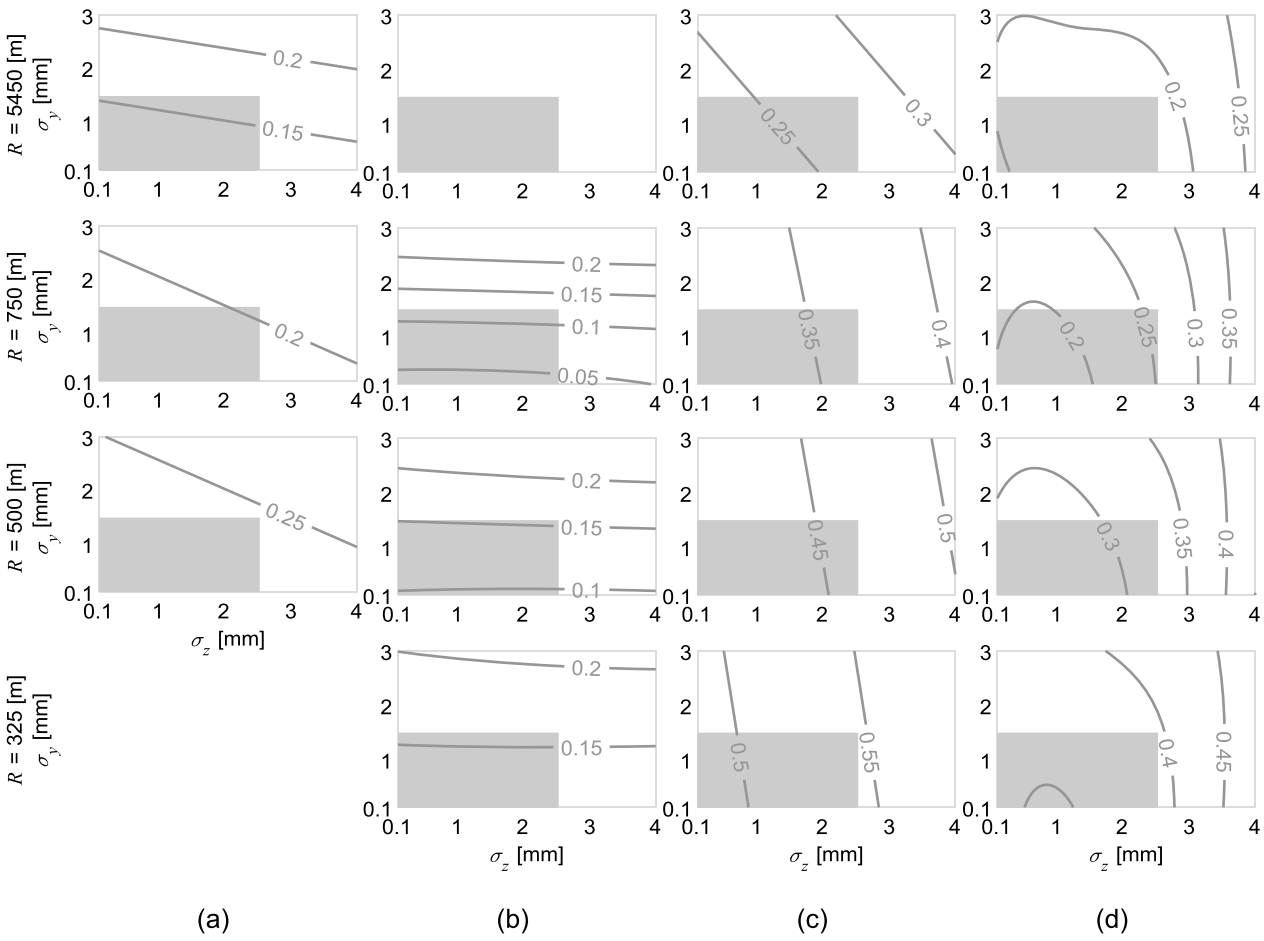


Fig. 6. $(\Delta Q/Q_0)_{L,\max}$ of the (a) passenger vehicle, (b) locomotive, (c) loaded freight wagon, and (d) empty freight wagon at $v = 90$ km/h.

mary, the variance of $y(\mathbf{x})$ is decomposed in terms attributable to each input, $\mathbb{V}_j$, and their interactions, $\mathbb{V}_Q$, where $Q \subseteq \{1,\ldots,k\}$.

Based on the variance decomposition, the first-order sensitivity indices are defined as [19]

$$S_j = \frac{\mathbb{V}_j}{\mathbb{V}}, \tag{18}$$

and correspond to the proportion of variance due to the jth input. Similarly, the second-order sensitivity indices are defined as [19]

$$S_{ij} = \frac{\mathbb{V}_{ij}}{\mathbb{V}}, \tag{19}$$

and correspond to the proportion of variance due to the joint effects of inputs x_i and x_j. The total sensitivity indices are defined as [19]

$$T_j = \frac{\mathbb{V} - \mathbb{V}\left[\mathbb{E}\left[y|x_{-j}\right]\right]}{\mathbb{V}}, \tag{20}$$

where, subscript $-j$ denotes every element that does not include j. In other words, the total sensitivity index corresponds to the sum of all sensitivities involving the jth input. Note that the difference between the total sensitivity index and first- and second-order sensitivity indices of the jth input, $T_j - S_j - \sum_{i=1}^{k}\sum_{j>1}^{k} S_{ij}$, corresponds to the proportion of the variance due to joint effects of more than two inputs involving j.

The sensitivity indices are evaluated using a quasi-Monte Carlo integration algorithm proposed by Gramacy [19]. The algorithm is modified to calculate second-order indices. Depending on the sensitivity analysis results, it might be worth exploring alternative models that explain the data.

4. Results and discussion

The case study involves railway vehicle dynamic analyses using VAMPIRE Pro [41], a multibody dynamics package specifically developed to solve rail vehicle dynamic problems quickly and cost-effectively. Each simulation consists of a vehicle running at constant speed on a 6 km long track with constant elevation. The setup considers four vehicles: (i) a passenger vehicle [42], (ii) a locomotive [43], and a freight wagon [44] in (iii) loaded and (iv) empty configurations. Appendix A compiles the data concerning the multibody models. The maximum allowed speed of the passenger vehicle is 220 km/h, while the maximum allowed speed of the locomotive and freight wagon is 100 km/h. The vehicles are equipped with new S1002 wheels, and they represent a small sample of the different vehicle types operating on the Portuguese rail network. A budget of 800 simulations (200 per vehicle) is set, where 25% is allocated to input screening, and 75% is assigned to GP model estimation. The results from the GPs are compared to the alert limits established in EN 13848-5.

4.1. Input screening

Since the elementary effects method uses r trajectories of $k+1$ points, the number of trajectories must be six to respect the simulation budget. Thus, the number of simulations used for input screening is 48 per vehicle (24% of the budget instead of 25%). Table 5 lists the lower and upper bounds of the input space used to analyze the elementary effects. The same draw from the multivariate ARMA process is used to model the irregularities in each trajectory. The goal is to guarantee that changes within each trajectory are unambiguously attributed to changes

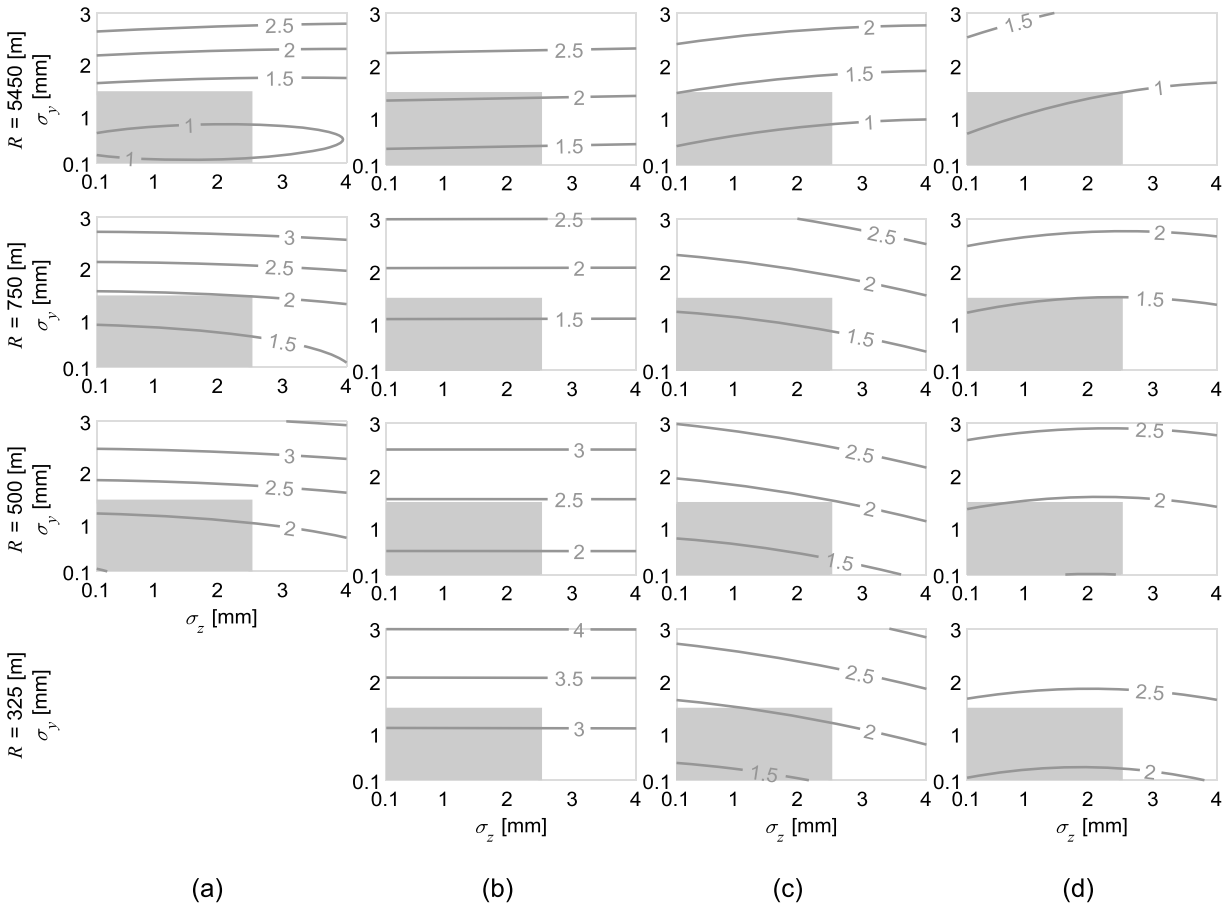


Fig. 7. $\ddot{y}^{+}_{\mathrm{L,max}}$ of the (a) passenger vehicle, (b) locomotive, (c) loaded freight wagon, and (d) empty freight wagon at $v = 90$ km/h.

in the inputs. Using different draws inside a trajectory could confound the effects of the input changes with the stochasticity of different draws. The optimization to find the subset r that maximizes distances among the trajectories was performed using Cariboni's function [39].

Appendix B depicts bar graphs of the elementary effects of $(Y/Q)_{\max}$, $(\Delta Q/Q_0)_{\mathrm{L,max}}$, $(\Delta Q/Q_0)_{\mathrm{R,max}}$, $\ddot{y}^{+}_{\mathrm{L,max}}$, $\ddot{y}^{+}_{\mathrm{R,max}}$ $\ddot{y}^{*}_{\max}$, and $\ddot{z}^{*}_{\max}$. Recall that the sensitivity measure is μ^* and high values indicate important inputs. Each Figure has four panels, one for each vehicle. The goal is to qualitatively compare μ^* across vehicles and see if a set of inputs is unimportant for all vehicles. A quantitative assessment across vehicles is not helpful or fair at this stage. Although the importance of the inputs changes with the safety quantity and vehicle considered, k_{vert} and $\sigma_{\Delta G}$ rank the lowest consistently. Therefore, k_{vert} and $\sigma_{\Delta G}$ are excluded from the next step, the estimation of the GP models.

4.2. Surrogate models

The designs are generated using the R package DiceDesign [45]. Random LHDs with $n = 152$ and $k = 5$ are generated and optimized using simulated annealing to minimize the L_2 discrepancy [46]. Table 6 lists the lower and upper bounds of the input space. Each simulation uses a different draw from the multivariate ARMA process to model the irregularities and ensure the GP model represents the SDs of irregularities instead of a specific draw. The models are estimated using a version of the DACE toolbox [47] modified for GP regression. The goal of regression is to filter noise due to using a different draw to model the irregularities in each simulation. However, note that there is no random error in this experiment. The regression term in Eq. (9) is a zero-order polynomial, i.e., β_0. In the numerical optimization to estimate the unknown parameters, g is constrained to the interval $[\varepsilon, \mathbb{V}[\mathbf{y}_{\mathbf{X}}]]$, where $\varepsilon = 2^{-52}$ is MATLAB's floating-point relative accuracy, and θ_j is constrained to the interval $[10^{-6}, 100]$.

The quantities reported for the GP models, namely, the predicted coefficient of determination, R^2_{pred}, and Mean Square Prediction Error, *MSPE*, are cross-validated metrics. In short, the quantities reported are the averages of k metrics evaluated using k-fold cross-validation [48]. The number of folds is $k = 10$, which is a typical number of folds [48]. Appendix C lists, for each safety quantity, the model parameters and the cross-validated metrics of the different vehicles.

4.3. Sensitivity analysis of the GP models

Appendix D summarizes the first- and second-order sensitivity indices, S_i and S_{ij}, and the variance proportion due to interaction involving more than two inputs, $T_j - S_j - \sum_{i=1}^{k}\sum_{j>i}^{k} S_{ij}$, rounded to two decimals. The sample size, N, used in the quasi-Monte Carlo integration is 15×10^5. Table D.1 lists the results of the sensitivity analysis of the wheel force ratio, $(Y/Q)_{\max}$. The most influential input for the locomotive and freight wagon (loaded and empty) is R explaining 79, 66, and 58% of the variance. For the passenger vehicle, the most influential inputs is v, which explains 57% of the variance. Depending on the vehicle, σ_y explains between 6% and 24% of the variance, while σ_z seems non-influential. σ_{twist} explains little variance for the passenger vehicle, locomotive, and loaded freight wagon, between 1% and 2%, but explains 14% for the empty freight wagon. Overall, the two-way sensitivity indices are very low or negligible. Interactions involving v can explain between 1% and 3%, and interactions involving R between 1% and 4%. The explained variance of higher-order interactions is negligible.

Table D.2 summarizes the sensitivity analyses of the wheel unload-

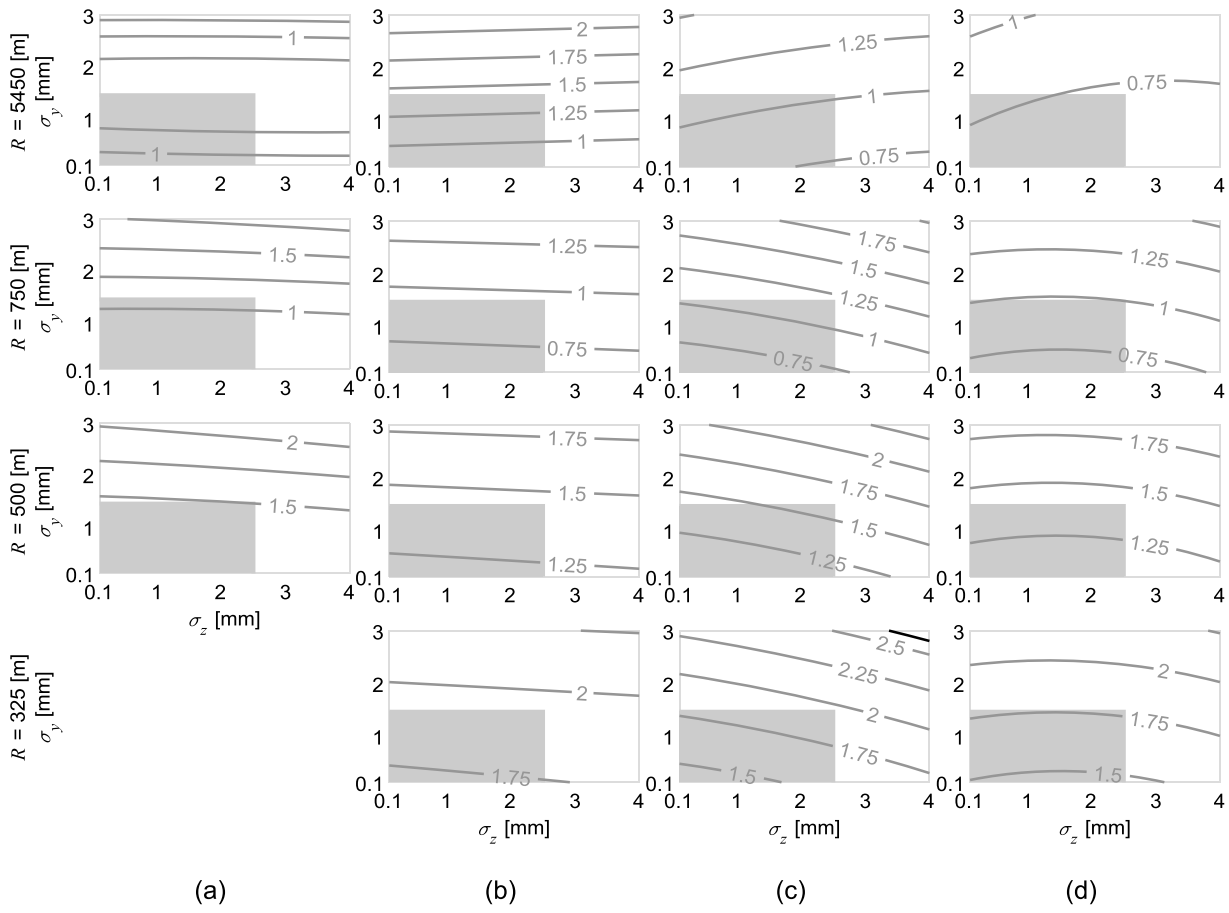


Fig. 8. $\ddot{y}^{*}_{\max}$ of the (a) passenger vehicle, (b) locomotive, (c) loaded freight wagon, and (d) empty freight wagon at $v = 90$ km/h.

ing, $(\Delta Q/Q_0)_{r,\max}$. The most influential input for the passenger vehicle and freight wagon (loaded and empty) is v, explaining between 33% and 56% of the variance, and the most influential input for the locomotive is R, explaining between 41% and 46%. σ_z and σ_y contribute little to the variance (between 0% and 3%), except for the locomotive, where σ_y explains between 27% and 30%. Conversely, σ_{twist} explains between 9% and 39% of the variance depending on the vehicle, except for the locomotive, where its contribution is between 0% and 2%. It is unexpected that v does not influence wheel unloading on the left wheel of the locomotive and that it lacks the wheel symmetry observed for the other vehicles. This discrepancy could be due to the asymmetry of the experimental setup (a single right-hand curve) and the suspension of the locomotive. Concerning two-way sensitivities, the interaction v, R explains between 7% and 15% of the variance, depending on the vehicle, and the interaction R, σ_y explains between 4% and 13% for the locomotive. Some other interaction involving v and R explain very little variance, 2% at most. Some variance is explained by higher-order interactions for the passenger vehicle, locomotive, and empty freight, but it is negligible for the loaded freight.

Table D.3 lists the sensitivity indices concerning the lateral acceleration of the bogie, $\ddot{y}^{+}_{r,\max}$. The first remark is that the sensitivities of the left and right wheels are remarkably close for this output, and differences between the wheels never go beyond 1%. The most influential input is v, explaining between 48% and 67% of the variance depending on the vehicle. Depending on the vehicle, R explains between 8% and 18% of the variance, σ_y between 9% and 12%, and σ_{twist} between 3% and 13%. σ_z does not explain any variance for this output. Regarding two-way sensitivities, depending on the vehicle, the interaction v, R explains between 3% and 17% of the variance, and other interactions involving v and R explain very little variance. The variance explained by higher-order interactions is very low.

Table D.4 comprises the sensitivity analysis of the lateral acceleration of the vehicle body, $\ddot{y}^{*}_{\max}$. The results are somewhat similar to those of $\ddot{y}^{+}_{r,\max}$, with v explaining most of the variance, σ_y and σ_{twist} explaining a smaller percentage, and σ_z explaining no variance. Concerning the two-way sensitivities, depending on the vehicle, interaction v, R explain between 7% and 10% of the variance, interactions involving v, σ_y and v, σ_{twist} explains between 1% and 3%. The variance explained by higher-order interactions is very low.

Table D.5 summarizes the sensitivity indices of the vertical acceleration of the vehicle body, $\ddot{z}^{*}_{\max}$. The most influential input is σ_z, explaining between 48% and 92% of the variance. Depending on the vehicle, v explains between 20% and 31% of the variance, except for the passenger vehicle, where it only explains 3%. σ_y does not explain any variance, and σ_{twist} explains between 1% and 3% for the freight wagon (loaded and empty). Regarding two-way sensitivities, interaction v, σ_z explains between 3% and 16% depending on the vehicle, interaction v, σ_{twist} explains 1% for the empty freight wagon, R, σ_z explains 1% for the locomotive, and σ_z, σ_y explains 1% for the passenger vehicle. The variance explained by higher-order interactions is very low.

Although the sensitivity analysis does not provide insight into the parametric form of the outputs, it reveals that most of the variance is explained by first- and second-order sensitivities. Consequently, a decomposition of the form

$$y(\mathbf{x}) = y_0 + \sum_{j=1}^{k} y_j(x_j) + \sum_{i=1}^{k} \sum_{j>i}^{k} y_{ij}(x_i, x_j), \tag{21}$$

should explain the outputs reasonably well. Therefore, second-order linear models with intercept, linear, interaction, and squared terms

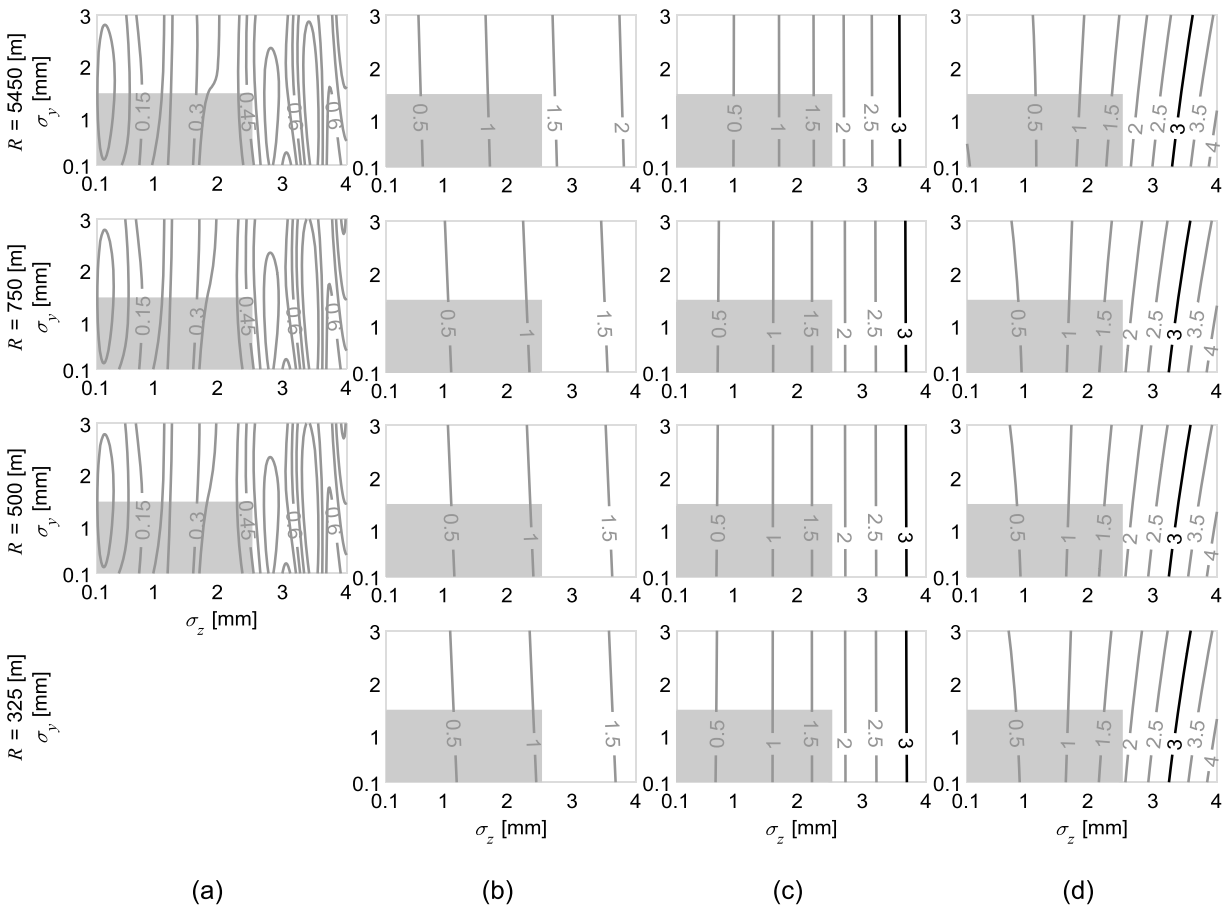


Fig. 9. $\ddot{z}^{*}_{max}$ of the (a) passenger vehicle, (b) locomotive, (c) loaded freight wagon, and (d) empty freight wagon at $v = 90$ km/h.

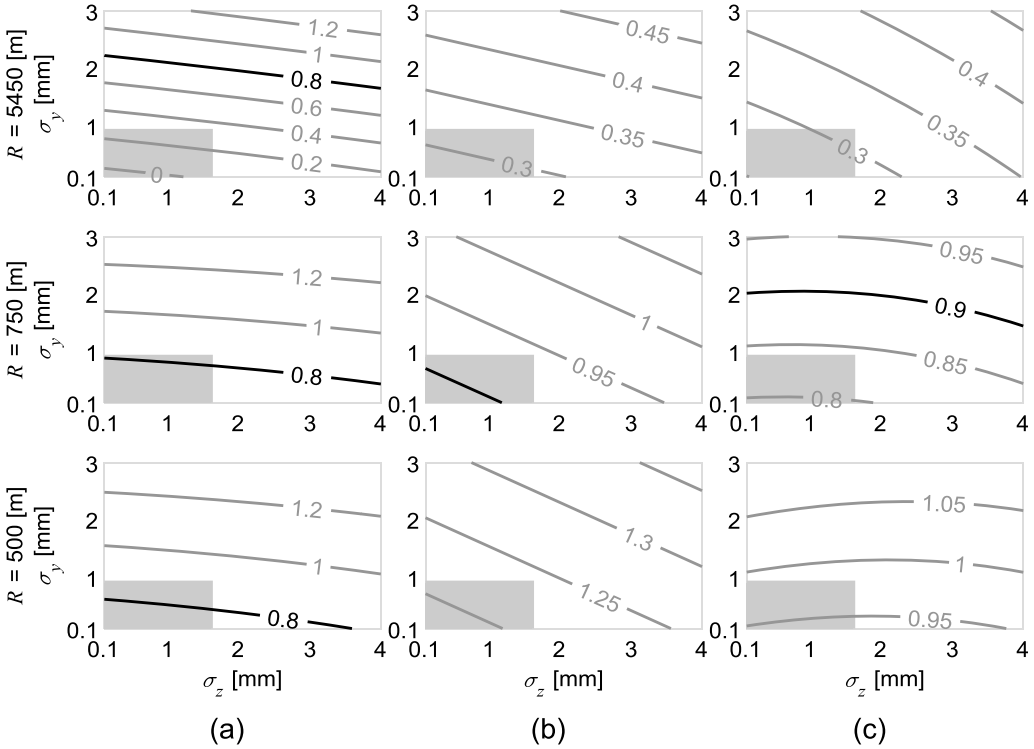


Fig. 10. (a) $(Y/Q)_{max}$, (b) $(\Delta Q/Q_0)_{L,max}$, (c) $(\Delta Q/Q_0)_{R,max}$ of the passenger vehicle at $v = 210$ km/h.

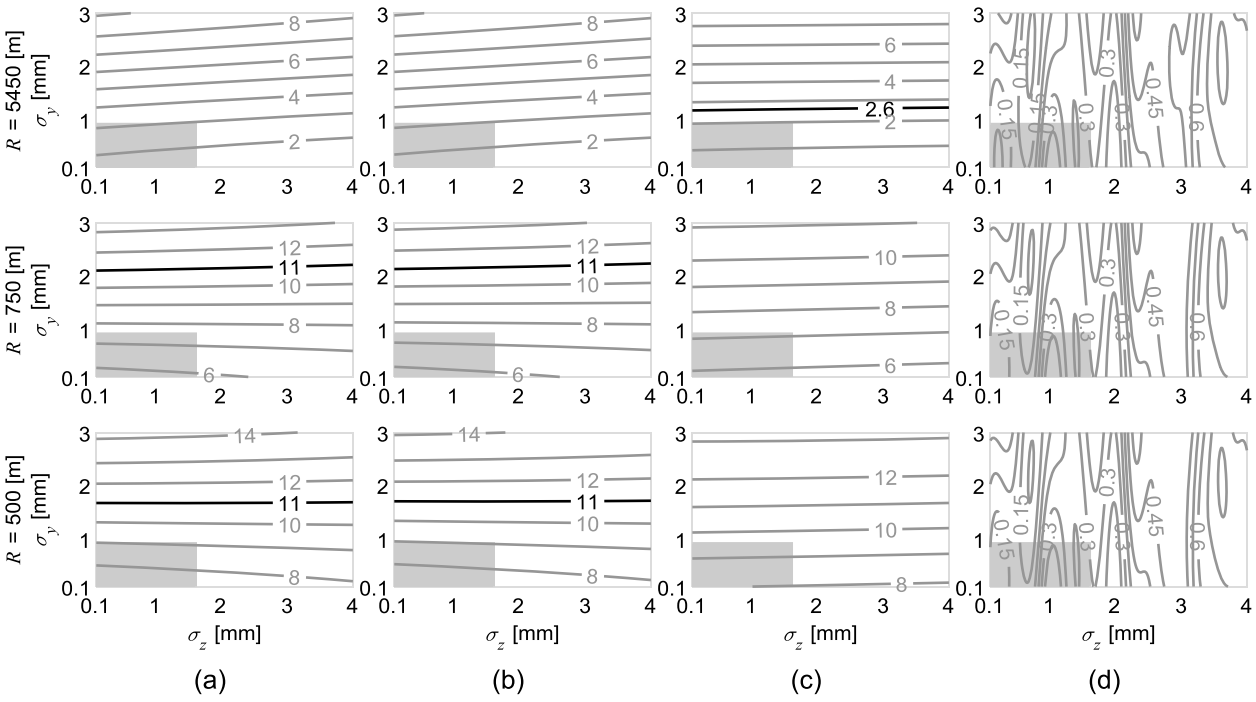


Fig. 11. (a) $\ddot{y}^{+}_{\mathrm{L,max}}$, (b) $\ddot{y}^{+}_{\mathrm{R,max}}$, (c) $\ddot{y}^{*}_{\mathrm{max}}$, and (d) $\ddot{z}^{*}_{\mathrm{max}}$ of the passenger vehicle at $v = 210$ km/h.

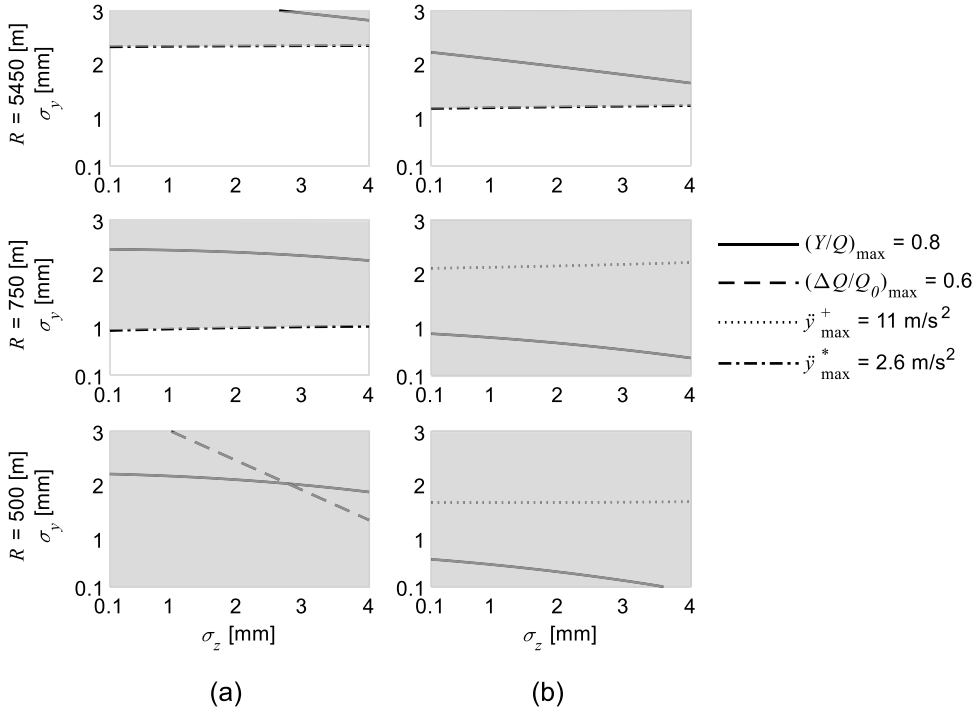


Fig. 12. Safety envelopes based on the safety quantities of the passenger vehicle at (a) 140 and (b) 210 km/h.

are estimated using stepwise regression and explored as an alternative to the GP models. The significance level to add or remove terms is 0.05 and 0.10, respectively. Appendix E lists the parameters of the linear models.

The mean squared prediction error of the linear model, denoted $MSPE_{\mathrm{linear}}$, is also calculated using k-fold cross-validation and compared to that of the GP models. Table 7 lists the ratio of the mean square prediction error of the linear models to that of the GP models, i.e., $MSPE_{\mathrm{linear}}/MSPE_{\mathrm{GP}}$. Intuitively, a ratio less than or equal to one indicates that the linear model predicts, on average, better than or as well as the GP model and could be used as a surrogate that has both predictive and explanatory power. For $MSPE_{\mathrm{linear}}/MSPE_{\mathrm{GP}} \leq 1.1$, the table shows that linear models might be suitable to represent (i) $(Y/Q)_{\mathrm{max}}$ for the passenger vehicle and locomotive, (ii) $(\Delta Q/Q)_{\mathrm{L,max}}$, and $(\Delta Q/Q)_{\mathrm{R,max}}$ for the empty freight wagon, (iii) $\ddot{y}^{+}_{\mathrm{L,max}}$, and $\ddot{y}^{+}_{\mathrm{R,max}}$ for the freight wagon (loaded and empty), (iv) $\ddot{y}^{*}_{\mathrm{max}}$ for the loaded freight wagon, (v) and $\ddot{z}^{*}_{\mathrm{max}}$ for the for every vehicle.

4.4. Safety assessment – surrogate model projections vs. TQCs

This subsection compares the alert limits suggested in EN 13848 for σ_z and σ_y with the results obtained from surrogate models. EN 13848-5 does not define alert limits for σ_z and σ_y, but specifies that they should be within TQC D. Therefore, the TQC D envelope is used throughout this subsection as a reference. The comparisons only involve the GP surrogate models.

Since the surrogates are 5-dimensional models, a 2-dimensional representation is obtained by evaluating σ_z and σ_y on a rectangular grid with 62,500 points covering the input space, i.e., $[0.1, 4] \times [0.1, 3]$, and holding the remaining variables constant. σ_{twist} is held at 0.7, which is approximately the SD that the multivariate ARMA process produces for the twist irregularity. Initially, v is held constant at 90 km/h to "stress" the locomotive and freight wagon because it is close to their maximum allowed speed. Then, following the same rationale, v is held constant at 210 km/h to "stress" the passenger vehicle. The locomotive and freight wagon are not analyzed at $v = 210$ km/h because it is beyond the domain of the models. R is held constant at 325, 500, 750, and 5450 m for the locomotive and freight wagon, and at 500, 750, and 5450 m for the passenger vehicle. These radii correspond to the middle of the radii intervals defined in Table 3.

The contour plots for $v = 90$ km/h are depicted in Figs. 5–9 and are organized in a 4×4 grid, where the rows represent fixed radii values, and the columns represent vehicles. In each panel, the horizontal and vertical axes are σ_z and σ_y, respectively. Solid dark gray lines represent the contours from the surrogate models, and the light gray area represents the TQC D envelope. Table 8 lists the limit values of the safety quantities, which, if inside the domain, are represented by solid black lines. The limit value for the wheel force ratio is 0.8. There is no consensus in the literature on a limit value for wheel unloading, and thus, two values are used, 0.6 and 0.9. The limit value for the lateral acceleration of the bogie is a function of the bogie mass, m^+. Since the limits for the passenger vehicle, locomotive, and freight wagon are 11.40, 11.38, 11.62 m/s^2, they are rounded down to 11 m/s^2 to use a single contour. The limits for the accelerations are directly retrieved from EN 14363.

Fig. 5 depicts the wheel force ratio, $(Y/Q)_{\max}$, at $v = 90$ km/h. Overall, the lower the radius, the higher the wheel force ratio, and every contour crossing TQC D is below the limit. Except for the loaded freight, whose contours are straight, the contours of the remaining vehicles are concave and resemble elliptic shapes. Interestingly, this output is lower when the freight wagon is loaded than when it is empty. Compared to the loaded freight wagon, contours of the same value on the empty freight wagon admit lower σ_y values and curve as σ_z increases. This result is not unreasonable because the suspension of the freight wagon is load-dependent and might perform better when the vehicle is loaded. The contours of the locomotive for $R = 5450$ m have a different shape than other radii and vehicle combinations. However, the unusual shape is likely due to noise and of no consequence. Note that the function varies little, and its values are low and far from the limit. Moreover, for $R = 5450$ m, the contours of the remaining vehicles are not very sensitive to σ_z.

Fig. 6 presents the wheel unloading of the left wheel, $(\Delta Q/Q_0)_{\mathrm{L,max}}$, at $v = 90$ km/h. In general, the lower the radius, the higher the wheel unloading, and the contours crossing the TQC D are all below the limit. The contours for the passenger, locomotive, and loaded freight wagon are relatively straight. In contrast, the contours of the empty freight wagon resemble ellipses with the major axis aligned along σ_y. The contours of the locomotive are almost horizontal, indicating this output is not very sensitive to σ_z. For the locomotive and $R = 5450$ m, the wheel unloading is approximately constant and equal to 0.13, hence the lack of contours. For the freight wagon, this time, the output relaxes when the vehicle is empty. For example, for $R = 750$ m, the 0.35 contour admits higher σ_z values for the empty wagon than the loaded one. Although the experimental setup is not symmetric, the contours of the left and right wheels are similar, with only a few changes in curvature in the passenger vehicle and orientation of the ellipses in the empty freight wagon. The results of the $(\Delta Q/Q_0)_{\mathrm{R,max}}$ at $v = 90$ km/h are in Appendix F (Fig. F.1).

Fig. 7 displays the lateral accelerations of the left side of the bogie frame, $\ddot{y}^+_{\mathrm{L,max}}$, at $v = 90$ km/h. In general, the lower the radius, the higher the lateral acceleration, and the contours crossing the TQC D are all below the limit. The contours of the passenger vehicle are relatively horizontal for higher values of σ_y, but tend to curve down as σ_y decreases and σ_z increases. The contours for the locomotive are practically horizontal, i.e., almost no dependence on σ_z. It is hard to summarize the difference between the loaded and empty freight wagons for this output because their behavior depends on the region being analyzed. Still, the loaded freight wagon is more sensitive to changes in σ_z and σ_y than the empty one. The results for the right side of the bogie are remarkably close to those of the left side. The results of the $\ddot{y}^+_{\mathrm{R,max}}$ are in Appendix F (Fig. F.2).

Fig. 8 presents the lateral acceleration of the vehicle body, $\ddot{y}^*_{\max}$, at $v = 90$ km/h. Like the previous output, the lower the radius, the higher the acceleration, and the contours inside TQC D are below the limit. The contours of the passenger vehicle and freight wagon resemble parts of ellipses. Once more, the change in the curvature of the contours of the loaded and empty freight wagon configurations highlights the influence of load. Although far from the TQC D envelope, the 2.6 m/s^2 limit contour is visible in the upper right corner of the loaded freight for $R = 325$ m. For $R = 5450$ m, the contours of the passenger vehicle and freight wagon show an unusual shape, likely due to the lateral acceleration not being as relevant in curves with this radius and speed combination.

Fig. 9 depicts the vertical acceleration of the vehicle body, $\ddot{z}^*_{\max}$, at $v = 90$ km/h. Despite what looks like noise in the acceleration of the passenger vehicle, the contours are mostly vertical. The passenger vehicle presents the lowest values of the vehicles, which is expected because it must comply with passenger safety and comfort requirements. The difference between the loaded and empty freight wagon is that the gradient of the latter is higher, and the contours start curving to the right as σ_y increases. For example, for low values of σ_y, the 3 m/s^2 contour moves from $\sigma_z \approx 3.7$ mm on the loaded freight wagon to $\sigma_z \approx 3.2$ mm on the empty one. Regardless of the vehicle, this quantity does not seem very sensitive to either σ_y nor R.

Figs. 10 and 11 display the safety quantities of the passenger vehicle at $v = 210$ km/h. In this Figure, rows correspond to fixed radii values, and columns correspond to safety quantities. Fig. 10(a) shows that the $(Y/Q)_{\max}$ limit shifts with R. For $R = 500$ and 750 m, the 0.8 limit contour intersects the TQC D envelope, which means the envelope overestimates safety in these cases. Fig. 10(b) and (c) show that the $(\Delta Q/Q)_{r,\max}$ limit also moves with R. For both wheels, for $R = 5450$ m, the contours are below the two limits. For $R = 750$ m, every contour is above the first limit, 0.6, but not the second one, 0.9. For $R = 500$ m, all contours are above the two limits. Fig. 11(a) and (b) show no significant difference between the $\ddot{y}^+_{\mathrm{L,max}}$ and $\ddot{y}^+_{\mathrm{R,max}}$, and the vehicle is safe for all radii. Fig. 11(c) presents the $\ddot{y}^*_{\max}$, and shows that, except for $R = 5450$ m, all contours are above the 2.6 m/s^2 limit. Fig. 11 (d) shows the $\ddot{z}^*_{\max}$. The contours are still reasonably vertical, and their values are almost the same as at $v = 90$ km/h. This output is not very sensitive to v and R for the passenger vehicle.

These surrogate models allow exploring the input space and seeing how vehicle safety changes. For example, the surrogate models can be used to determine which outputs govern safety in different situations. Fig. 12 overlaps the surrogate models of the safety quantities of the passenger vehicle for three different radii, $R = 500$, 750, and 5450 m, at two different speeds, $v = 140$ and 210 km/h. For clarity, the Figure only shows the limit contours and represents function values above the limits in gray. Therefore, the gray area corresponds to the union of the areas above the limit of each output. The visualization shows that the safety of

this vehicle is governed by $\ddot{y}^{*}_{\max}$ regardless of radii and speed. Although a single input determines safety in this case, the boundary can differ for other input combinations or vehicles. Assuming for the sake of argument that $\ddot{y}^{*}_{\max}$ is not a relevant quantity, Fig. 12 shows that the safety boundary for $R = 500$ m and $v = 140$ km would be a combination of $(Y/Q)_{\max}$ and $(\Delta Q/Q)_{\max}$.

5. Conclusions

This work proposed a DACE framework to model relationships between track quality and vehicle safety. The framework consists of six steps: (i) input selection and pre-processing, (ii) vehicle responses and post-processing, (iii) input screening using the elementary effects method (iv) surrogate modeling using GPs, (v) sensitivity analysis using Sobol indices, and (vi) safety assessment.

The case study demonstrating the framework considered three different vehicles and the following inputs: vehicle speed, track curvature, vertical track stiffness, and the SDs of the longitudinal level, alignment, twist, and gauge irregularity. The input screening process excluded track stiffness and the SD of the gauge irregularity because they did not significantly influence the safety quantities. The sensitivity analysis of the GP models estimated with the remaining inputs showed that the radius and SD of the twist contribute, in some instances, to the variance of the safety quantities. Although the European standard for track quality does not consider these quantities, the results suggest they could be helpful TQIs. Moreover, inputs that influence one vehicle may be insignificant to another, which is why defining TQIs limits is challenging. The sensitivity analysis also showed that first- and second-order sensitivities explained most of the variance, and second-order linear models might be suitable to represent some safety quantities. Although the 2-dimensional representations of the surrogates on the SD of the longitudinal level and SD of the alignment plane showed curved contours, the sensitivity analysis indicates this effect is not due to interaction between these two variables. In some instances, the contours resemble ellipses whose axes are so large that they seem horizontal or vertical lines. EN 13848-6 does not provide a reason for the TQCs rectangular envelopes. However, this shape seems reasonable if it results from rectangles circumscribed by quarter-elliptic contours, or an overlap of two different safety quantities, one with vertical contours and another with horizontal contours.

Although this work focused on EN 13848, it does not preclude other standards or maintenance guidelines. For example, this framework can be used to objectively benchmark aggregate TQIs according to vehicle safety instead of a subjective judgment of how TQIs should behave. As shown in the introduction, researchers are attempting to shift from periodically inspecting the track using recording vehicles, which results in service downtime, to continuously monitoring the track using service vehicles equipped with accelerometers. This new paradigm and the proposed framework complement each other. On the one hand, the surrogate models can provide information regarding the vehicle that would otherwise be unavailable because measured data is biased towards safety. On the other hand, measured data from inspection and service vehicles can be used to calibrate the surrogate model. A shortcoming of this study is that the input space of the locomotive and freight wagon was not broad enough, and the limit contours are often outside the domain of the surrogate. Future work could consider an active learning approach that sequentially queries the computer code around the desired contour (safety limit) based on the expected improvement. Nonetheless, the results from the passenger vehicle show the potential of the proposed approach to find track geometry limits directly related to safety.

CRediT authorship contribution statement

João Neves Costa: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Jorge Ambrósio:** Supervision, Funding acquisition. **António R. Andrade:** Writing – review & editing, Supervision, Conceptualization. **Daniel Frey:** Writing – review & editing, Supervision, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The first author expresses his gratitude to the Portuguese Foundation for Science and Technology (Fundação para a Ciência e a Tecnologia) and the Luso-American Development Foundation (Fundação Luso-Americana para o Desenvolvimento) through the grants PD/BD/128138/2016 and Project – 140/2019, respectively. This work was supported by the Portuguese Foundation for Science and Technology, through IDMEC, under LAETA, project UIDB/50022/2020.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ress.2022.108856.

References

[1] El-Sibaie M, Zhang YJ. Objective track quality indices. Transp Res Rec J Transp Res Board 2004;1863(1):81–7. https://doi.org/10.3141/1863-11.
[2] Madejski J. Autonomous track geometry diagnostics system. J Mater Process Technol 2004;157–158:194–202. https://doi.org/10.1016/J.JMATPROTEC.2004.09.029.
[3] Offenbacher S, Neuhold J, Veit P, Landgraf M. Analyzing major track quality indices and introducing a universally applicable TQI. Appl Sci 2020 2020;10. https://doi.org/10.3390/APP10238490.
[4] CEN. EN 13848-6 railway applications - track - track geometry quality part 6: characterisation of track geometry quality. BSI Standards Limited; 2014.
[5] Andrade AR, Teixeira PF. Hierarchical Bayesian modelling of rail track geometry degradation. Proc Inst Mech Eng Part F J Rail Rapid Transit 2013;227(4):364–75. https://doi.org/10.1177/0954409713486619.
[6] Andrade AR, Teixeira PF. Statistical modelling of railway track geometry degradation using hierarchical Bayesian models. Reliab Eng Syst Saf 2015;142:169–83. https://doi.org/10.1016/J.RESS.2015.05.009.
[7] Bai L, Liu R, Sun Q, Wang F, Xu P. Markov-based model for the prediction of railway track irregularities. Proc Inst Mech Eng Part F J Rail Rapid Transit 2013;229(2):150–9. https://doi.org/10.1177/0954409713503460.
[8] Cárdenas-Gallo I, Sarmiento CA, Morales GA, Bolivar MA, Akhavan-Tabatabaei R. An ensemble classifier to predict track geometry degradation. Reliab Eng Syst Saf 2017;161:53–60. https://doi.org/10.1016/J.RESS.2016.12.012.
[9] Andrade AR, Teixeira PF. Exploring different alert limit strategies in the maintenance of railway track geometry. J Transp Eng 2016;142:04016037. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000867.
[10] Sharma S, Cui Y, He Q, Mohammadi R, Li Z. Data-driven optimization of railway maintenance for track geometry. Transp Res Part C Emerg Technol 2018;90:34–58. https://doi.org/10.1016/J.TRC.2018.02.019.
[11] Sresakoolchai J, Kaewunruen S. Railway defect detection based on track geometry using supervised and unsupervised machine learning. Struct Health Monit 2022;21(4). https://doi.org/10.1177/14759217211044492.
[12] Mohammadi R, He Q, Ghofrani F, Pathak A, Aref A. Exploring the impact of foot-by-foot track geometry on the occurrence of rail defects. Transp Res Part C Emerg Technol 2019;102:153–72. https://doi.org/10.1016/J.TRC.2019.03.004.
[13] Chudzikiewicz A, Bogacz R, Kostrzewski M, Konowrocki R. Condition monitoring of railway track systems by using acceleration signals on wheelset axle-boxes. Transport 2018;33:555–66. https://doi.org/10.3846/16484142.2017.1342101.
[14] Lasisi A, Attoh-Okine N. Principal components analysis and track quality index: a machine learning approach. Transp Res Part C Emerg Technol 2018;91:230–48. https://doi.org/10.1016/J.TRC.2018.04.001.
[15] McKay MD, Beckman RJ, Conover WJ. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code.

Technometrics 1979;21:239–45. https://doi.org/10.1080/00401706.1979.10489755.
[16] Damblin G, Ghione A. Adaptive use of replicated latin hypercube designs for computing Sobol' sensitivity indices. Reliab Eng Syst Saf 2021;212:107507. https://doi.org/10.1016/J.RESS.2021.107507.
[17] Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. Stat Sci 1989;4:409–23.
[18] Santner TJ, Williams BJ, Notz WI. The design and analysis of computer experiments. New York, US: Springer; 2018.
[19] Gramacy RB. Surrogates: Gaussian process modeling, design and optimization for the applied sciences. Boca Raton, Florida: Chapman Hall/CRC; 2020.
[20] Yuan J, Ng SH. A sequential approach for stochastic computer model calibration and prediction. Reliab Eng Syst Saf 2013;111:273–86. https://doi.org/10.1016/J.RESS.2012.11.004.
[21] Fruth J, Roustant O, Kuhnt S. Sequential designs for sensitivity analysis of functional inputs in computer experiments. Reliab Eng Syst Saf 2015;134:260–7. https://doi.org/10.1016/J.RESS.2014.07.018.
[22] Zhang X, Wang L, Sørensen JD. REIF: a novel active-learning function toward adaptive Kriging surrogate models for structural reliability analysis. Reliab Eng Syst Saf 2019;185:440–54. https://doi.org/10.1016/J.RESS.2019.01.014.
[23] Zuhal LR, Faza GA, Palar PS, Liem RP. On dimensionality reduction *via* partial least squares for Kriging-based reliability analysis with active learning. Reliab Eng Syst Saf 2021;215:107848. https://doi.org/10.1016/J.RESS.2021.107848.
[24] Ameryan A, Ghalehnovi M, Rashki M. AK-SESC: a novel reliability procedure based on the integration of active learning kriging and sequential space conversion method. Reliab Eng Syst Saf 2022;217:108036. https://doi.org/10.1016/J.RESS.2021.108036.
[25] Bichon BJ, McFarland JM, Mahadevan S. Efficient surrogate models for reliability analysis of systems with multiple failure modes. Reliab Eng Syst Saf 2011;96:1386–95. https://doi.org/10.1016/J.RESS.2011.05.008.
[26] Qian G, Massenzio M, Brizard D, Ichchou M. Sensitivity analysis of complex engineering systems: approaches study and their application to vehicle restraint system crash simulation. Reliab Eng Syst Saf 2019;187:110–8. https://doi.org/10.1016/J.RESS.2018.07.027.
[27] Ray D, Ramirez-Marquez J. A framework for probabilistic model-based engineering and data synthesis. Reliab Eng Syst Saf 2020;193:106679. https://doi.org/10.1016/J.RESS.2019.106679.
[28] O'Hagan A. Bayesian analysis of computer code outputs: a tutorial. Reliab Eng Syst Saf 2006;91:1290–300. https://doi.org/10.1016/J.RESS.2005.11.025.
[29] Campbell K. Statistical calibration of computer simulations. Reliab Eng Syst Saf 2006;91:1358–63. https://doi.org/10.1016/J.RESS.2005.11.032.
[30] Perrin G. Adaptive calibration of a computer code with time-series output. Reliab Eng Syst Saf 2020;196:106728. https://doi.org/10.1016/J.RESS.2019.106728.
[31] CEN. EN 13848-1 railway applications - track - track geometry quality part 1: characterization of track geometry. BSI Standards Limited; 2019.
[32] CEN. EN 13848-5 railway applications - track - track geometry quality part 5: geometric quality levels - plain line, switches and crossings. BSI Standards Limited; 2017.
[33] CEN. EN 14363 railway applications - railway applications - testing and simulation for the acceptance of running characteristics of railway vehicles - running behaviour and stationary tests. BSI Standards Limited; 2016.
[34] CEN. FprCEN/TR 16513 railway applications - track - survey of track geometry quality. BSI Standards Limited; 2014.
[35] UIC. UIC Code 518 testing and approval of railway vehicles from the point of view of their dynamic behaviour - safety - track fatigue - running behaviour. International Union of Railways; 2009.
[36] CEN. EN 13803-1 railway applications — track — track alignment design parameters — track gauges 1435 mm and wider part 1: plain line. BSI Standards Limited; 2010.
[37] Costa JN, Ambrósio J, Frey D, Andrade AR. A multivariate statistical representation of railway track irregularities using ARMA models. Veh Syst Dyn 2021;60(7):2494–510. https://doi.org/10.1080/00423114.2021.1912365.
[38] Morris MD. Factorial sampling plans for preliminary computational experiments. Technometrics 1991;33:161–74. https://doi.org/10.1080/00401706.1991.10484804.
[39] Campolongo F, Cariboni J, Saltelli A. An effective screening design for sensitivity analysis of large models. Environ Modell Softw 2007;22:1509–18. https://doi.org/10.1016/j.envsoft.2006.10.004.
[40] Forrester A, Sobester A, Keane A. Engineering design *via* surrogate modelling: a practical guide. Chichester, West Sussex, UK: Wiley; 2008.
[41] ENSCO: VAMPIRE Pro, 7.0 [Software]. 2020, SNC-Lavalin, Montreal, Canada*Software name, version, release date, publisher, location*.
[42] Antunes P, Magalhães H, Ambrósio J, Pombo J, Costa JN. A co-simulation approach to the wheel–rail contact with flexible railway track. Multibody Syst Dyn 2019;45:245–72. https://doi.org/10.1007/s11044-018-09646-0.
[43] Millan P. Dynamics of freight trains under realistic operation conditions. Instituto Superior Técnico; 2021. Master's thesis.
[44] Pagaimo J, Magalhães H, Costa JN, Ambrósio J. Derailment study of railway cargo vehicles using a response surface methodology. Veh Syst Dyn 2020. https://doi.org/10.1080/00423114.2020.1815810.
[45] Dupuy D, Helbert C, Franco J. DiceDesign and DiceEval: two R packages for design and analysis of computer experiments. J Stat Softw 2015;65:1–38. https://doi.org/10.18637/jss.v065.i11.
[46] Morris MD, Mitchell TJ. Exploratory designs for computational experiments. J Stat Plan Inference 1995;43:381–402. https://doi.org/10.1016/0378-3758(94)00035-T.
[47] Nielsen HB, Lophaven SN, Söndergaard J. DACE - a matlab kriging toolbox. Informatics and Mathematical Modelling, Technical University of Denmark, DTU; 2002.
[48] Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York: Springer New York; 2009.
[49] Iwnicki S, Spiryagin M, Cole C, McSweeney T. Handbook of railway vehicle dynamics. Boca Raton, Florida: CRC Press; 2019.