Rail RCF damage quantification and comparison for different damage models

Visakh V. Krishna¹ · Saeed Hossein-Nia¹ · Carlos Casanueva¹ · Sebastian Stichel¹ · Gerald Trummer² · Klaus Six²

Received: 24 May 2021 / Revised: 6 August 2021 / Accepted: 7 August 2021 / Published online: 27 September 2021
© The Author(s) 2021

Abstract There are several fatigue-based approaches that estimate the evolution of rolling contact fatigue (RCF) on rails over time and built to be used in tandem with multi-body simulations of vehicle dynamics. However, most of the models are not directly comparable with each other since they are based on different physical models even though they shall predict the same RCF damage at the end. This article studies different approaches to quantifying RCF and puts forward a measure for the degree of agreement between them. The methodological framework studies various steps in the RCF quantification procedure within the context of one another, identifies the ‘primary quantification step’ in each approach and compares results of the fatigue analyses. In addition to this, two quantities—‘similarity’ and ‘correlation’—have been put forward to give an indication of mutual agreement between models. Four widely used surface-based and sub-surface-based fatigue quantification approaches with varying complexities have been studied. Different operational cases corresponding to a metro vehicle operation in Austria have been considered for this study. Results showed that the best possible quantity to compare is the normalized damage increment per loading cycle coming from different approaches. Amongst the methods studied, approaches that included the load distribution step on the contact patch showed higher similarity and correlation in their results. While the different approaches might qualitatively agree on whether contact cases are ‘damaging’ due to RCF, they

might not quantitatively correlate with the trends observed for damage increment values.

Keywords Rolling contact fatigue · Rail surface damage · Wedge model · KTH model · Surface fatigue index · Multi-body simulation · Track-friendliness

1 Introduction

Rail surface damage caused by rolling contact fatigue (RCF) is a major problem faced by several railway networks across the world with increasing operational requirements such as heavier axle loads, higher line capacities and vehicle running speeds. Management of rail surface damage has only been getting more critical to guarantee safety at still economically feasible cost. Study of RCF has assumed even more importance in recent times since RCF mitigation measures are driving a lot of track maintenance and wheel-reprofiling activities [1].

Several studies modelling the formation of surface cracks on rails exist. Considering that the physics behind the development of head checks is complex and still very much a topic of research, various models have been developed. They often employ multi-body simulations (MBS) of railway vehicles to obtain the dynamic wheel–rail loads and kinematics modelled for various running conditions. However, the treatment of results coming from MBS varies significantly amongst the approaches when it comes to quantifying RCF initiation in the wheel–rail contact. For instance, as Six et al. [2] pointed out in their work, some models take into account loads/stresses only

✉ Visakh V. Krishna
visakh@kth.se

¹ Department of Engineering Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden

² Virtual Vehicle Research GmbH, Graz, Austria

without consideration of wear, which continuously removes the material from the rail surface while some also take wear into account. However, at the end, they should predict the same results and agree with field observations. Therefore, the modelling principles behind different approaches need to be understood and the similarities and differences between them need to be identified.

This article studies in detail four widely used damage models in railway applications namely Wedge model [3], KTH model [4], surface fatigue index [5] and the whole life rail model (WLRM) [6]. For each method, the damage quantification is done for varying service conditions and a measure for the degree of agreement between the methods is put forward. This will allow researchers working on damage models to compare their results with other models in a systematic way. It also helps to understand the key differences between the results from different methods and what they signify within the context of one another. In doing so, the influence of parameters at different stages of modelling can be evaluated and a meaningful basis for comparison is established. All four methods have been calibrated with the help of experimental investigations or on-track measurements as described in the following sections.

A major factor to be considered when studying different approaches is the feasibility w.r.t. the scale of calculation. Damage quantification should be fast and preferably rely on simple analytical solutions when done for a large number of load cycles. With increasing wheel–rail system complexity in the form of various factors to consider (such as vehicle suspensions, wheel profiles, curve radii, and friction levels), it becomes unfeasible to use very detailed models to quantify RCF damage due to the associated long processing times. At the same time, the use of a simpler model for a relatively small parameter space over-generalizes the wheel–rail contact problem by neglecting specific phenomena at play that influence damage (e.g. partial slip) and therefore might not be accurate enough. This is particularly relevant when studying RCF damage models because of the differing physical phenomena they are modelled on.

This article is structured in the following way. Section 2 classifies and describes the models used for damage quantification. Section 3 describes the simulation model and the operating scenarios used to generate data for the RCF assessment. Section 4 lists and discusses the damage quantification results for different models, which is followed by discussion on their similarities and differences in Sect. 5. Section 6 summarizes the conclusions from the work.

2 Conceptual framework

2.1 System complexity vs. damage model complexity

Figure 1 depicts the conflict between system complexity and complexity of the damage model marked by the X- and Y-axis, respectively, arising due to calculation constraints depicted qualitatively by the dotted line. The extent of application for some RCF damage quantification approaches is also depicted on the plot.

System complexity pertains to the number of load cycles in a given vehicle–track operation and the underlying spectrum of varying parameters such as friction levels, wheel/rail profiles, and axle load that need to be considered. The quantification of damage is of practical significance only when it is expressed for a given running distance (wheels) or tonnage passing (rails). For this, the damage at each occurring wheel–rail contact point, referred to as *contact patch* hereon, is accumulated for a given set of loading cycles. This increases the system complexity, broadly classified in the figure with three levels. In the short-term level, the running distance/tonnage is small enough (e.g. 1 km/0.1 MGT (million gross tonnes)) such that the wheel–rail combination essentially remains the same, implying a single MBS case. In the long-term level, the prolonged loading cycles mean that the wheel–rail combination evolves due to the removal of material (e.g. 10,000 km/5 MGT). This is commonly modelled with iterative MBS where the wheel/rail profiles are updated after each iteration. Finally, the rail/wheel life extends further where the effect of intermediate maintenance activities needs to be considered when quantifying damage (e.g. 100 MGT). With increasing system complexity, the

Fig. 1 Different approaches for quantifying RCF damage

number of accumulated contact patches (and therefore calculations) increases.

Damage model complexity pertains to how detailed the physical phenomena behind RCF are modelled. The complexity is again depicted in three levels in the figure. ‘Global’ methods quantify damage by considering the contact patch as a whole. The whole life ail model (WLRM) proposed in [6] is an example of a simple ‘global’ approach where the overall contact energy dissipation is determined directly from MBS results and fit in a function to determine the damage increment for the whole contact patch (see Fig. 2a). Yet another example of a global approach is the surface fatigue index (FI_{surf}) put forward by Ekberg et al. [5] that considers the traction coefficient of the contact (contact friction force divided by contact normal force) and the maximum Hertzian contact pressure to map a working point (WP) on a shakedown diagram to quantify damage increment while assuming full-slip conditions (see Fig. 2b).

‘Local’ methods go a step further by discretizing the contact patch’s surface area into numerous elements and quantifying the corresponding damage for each element. This requires the extra step of determining normal and tangential pressure distributions on the contact patch from wheel–rail forces using suitable contact models, which means increasing the damage model complexity. The contact models can vary within individual approaches such as the common *Hertz* (Normal) + *Fastsim* (Tangential) approach, more sophisticated approaches like the extended creep force (ECF) model [7–9] that additionally take third body layers and local temperature effects into account or *Hertz* + *FaStrip* in the KTH model [10]. Here, the local stresses determine the working point of each contact patch element instead of the whole contact patch to quantify damage. The local approach as work done in [11] shows higher damages when compared with the global approach.

If the global FI_{surf} value is distributed elliptically over the contact patch, then comparing it with the values obtained by local methods that use contact models such as *Fastsim* or *FaStrip*, results of the global method show an underestimation of the calculated damage indices. It should also be noted that amongst the contact models used, the *Fastsim* algorithm tends to overestimate the damage indices when compared to the *FaStrip* algorithm since it uses a parabolic traction bound compared to the elliptical traction bound of *FaStrip*.

Methods falling under ‘local sub-surface’ level extend the discretization beneath the contact surface. In doing so, they also consider sub-surface stresses in varying degrees to quantify damage. The Wedge model [3, 12], for instance, considers a thin layer on the rail surface up to a depth of about 0.13 mm to include the effect of plastic deformation. The Brick model [13] considers a sub-surface layer discretized into tiny rectangular ‘bricks’, each accumulating shear strain according to the applied shear stress and the material strength. Once the accumulated plastic shear strain exceeds the critical shear strain for failure, the initially ‘healthy’ brick is deemed to be ‘weak’ and eventually removed, hence quantifying damage. More complex methods involving finite element methods (FEMs) and fracture mechanics can also be used but do not fall under purview of this work since they do not have an analytical solution and the ensuing calculation times make them unsuitable to be used in tandem with iterative MBS calculations.

Different models quantify the damage differently due to the scales and the different modelling approaches. However, at the end, they should give the same results because they all try to predict the same damage phenomenon. Thus, a systematic methodology is needed to make a quantitative comparison possible.

Fig. 2 Global methods formulating RCF damage: **a** energy dissipation approach [6]; **b** shakedown map (surface fatigue index) approach [5]. $F\gamma$ is the energy dissipated at contact point, FI_{surf} is surface fatigue index, p_0 is the maximum pressure on the contact point, k is the yield strength of the material, BC is the boundary curve, WP is the working point, and BC-WP is the distance of the working point from the boundary curve

The visualization in Fig. 2 does not necessarily indicate that only models with higher damage model complexity give more accurate results, but rather the amount of detailing considered in the approach as indicated by the Y-axis. A more complex model would be required only in systems that require more physical insight/prediction detail. In this article, global methods employing energy dissipation and shakedown map, and local methods such as the KTH model and Wedge models are studied. They quantify RCF in the form of surface-initiated crack initiation due to rolling contact fatigue, commonly referred to as head checks. RCF arising due to other modes such as sub-surface-initiated fatigue and squats is not modelled by these methods.

2.2 Model descriptions for local approaches

Brief descriptions of the local approaches used in the current work are given in this subsection. The damage calculation starts with MBS of various vehicle-track combinations in a commercial software package such as GENSYS [14] or SIMPACK [15]. Upon its completion, some key outputs are used to model the local characteristics of the contact patches generated. They include parameters such as geometry of the contact area, friction, creepages, normal forces, and relative position of the wheel with respect to rail. Following these results from MBS, normal and tangential pressure distributions on the contact patch are calculated using suitable contact models.

2.2.1 Wedge model

Wedge model [3, 12] is a ‘local sub-surface’ approach that has been developed for improving the prediction of head check crack initiation by considering the influence of severe plastic shear deformation at the surfaces of wheels and rails. Such layers are frequently observed in railway operation and exceed the thickness of deformed surface layers in many other classical tribological problems [16]. Severe plastic shear deformation leads to microstructural alignment with the shear plane. Fatigue crack growth experiments in plastically shear-deformed rail steel samples have shown that fatigue crack growth rates are the highest in the direction parallel to the plane of plastic shear deformation [17]. If such favourable crack initiation planes are oriented at an oblique angle to the surface, microscopic cracks may remain in the material and eventually form fatigue cracks instead of generating a wear particle that is removed from the surface.

To identify favourable conditions for crack initiation in the Wedge model, calculated plastic shear strain distributions in a near-surface layer are compared to reference distributions by means of a similarity parameter. These

reference distributions have been postulated to be favourable for crack initiation based on experimental observations. The similarity parameter in combination with the maximum principal (tensile) stress of the loading history and a factor addressing the influence of traction and braking determines an “effective stress” which is used subsequently to calculate damage increments by means of a power law. The stress state of the contact loading is determined by a strip-wise normal and tangential contact model preceding the damage calculation [3, 12].

The Wedge model has been calibrated with test results for rail material R350HT loaded by wheels made of material R7. The reference condition for crack initiation in the Wedge model as well as the fatigue properties of the model has been derived from full-scale wheel–rail test rig experiments by analysing crack orientations and preferential orientations of the material at the gauge corner of rails [3, 16, 17]. A stress–strain curve for the prediction of plastic shear deformation near the surface of rolling contacts has been derived from measurements of microstructure orientations due to plastic shear deformation and hardness profiles on specimens of twin disc experiments [16, 17]. The model has been validated for a metro operation scenario [3, 18].

2.2.2 KTH model

The KTH model is a ‘local surface’ approach which is derived from Johnson’s shakedown theory [19]. The theory is based on the ratchetting phenomenon on the surface of a material. Ratchetting is a phenomenon in which the material subjected to cyclic loading undergoes a cyclic “creep” deformation which is small and slow plastic deformation. It does not always occur in conditions where the loading is above the yield limit of the material. After the first few cycles, the material gets hardened and with the help of residual stresses the response of the material becomes elastic. This behaviour of the material is called Shakedown since the material ‘shakes’ down to a lower state of deformation. According to Johnson, assuming steady-state rolling, plane strain and under the presence of residual stresses, the material under cyclic loading ratchets if and only if the surface shear stresses exceed the material shear strength.

The mentioned idea has been used by Dirks et al. [4] and Hossein-Nia et al. [10] at KTH. Dirks’ model considers Hertz + Fastsim and assumes a power-law relationship between the surface shear stresses exceeding the shear strength (termed as the stress index, SI) and the depth of the RCF cracks. The calibration experience using the SI for the KTH model is described in [4]. Material parameters in the model were calibrated against crack measurements in a curve on the Dutch railways over a period of 5 years. An

exponential power law was empirically created that linked crack depth with the SI. Two different rail types were installed in the curve: 260 Mn-grade rail and a harder MHH-grade rail. The crack measurements were made visually by measuring the surface crack length and ultrasonically by measuring the crack depth at twelve different positions in the curve. Hossein-Nia et al. [10] have improved the accuracy of the shear stress estimation by using FaStrip instead of Fastsim for the tangential contact problem. In this work, Dirks' model with the improved FaStrip contact model has been used.

The effect of wear is considered differently by Dirks et al. [4] and Hossein-Nia et al. [10]. Dirks et al. calculate the wear depth using the KTH wear calculation methodology [20] separately, which is then subtracted from the RCF crack depth. However, instead of the crack depth, Hossein-Nia et al.'s model estimates the number of wheel rotations (N_r) until the cracks are visible on the surface while considering the direct effect of wear on RCF using an energy dissipation-based method at the contact patch. Using Hossein-Nia et al.'s method on the simulation cases considered in this work, the influence of wear was found to be negligible and only applied for very few cases with high energy dissipation values.

The modelling approaches are very different when comparing Wedge model with the KTH approach. It needs to be kept in mind that both approaches try to predict the same damage pattern (RCF cracks). This highlights again how important a clear methodology to compare different models is.

2.3 Comparison framework

In the previous subsections, the approaches for damage quantification were briefly described. The workflows of the four models for RCF quantification used in this work namely WLRM, surface fatigue index, KTH model and Wedge model are described side by side for comparison in Fig. 3. The steps along the workflow for all models can broadly be categorized in three stages: contact response, load distribution and fatigue modelling as described in the figure. For global methods (WLRM and the surface fatigue index approach), load distribution on the contact patch is lacking due to the assumptions of the models.

A salient feature in all the approaches barring the WLRM approach is that they record the exceedance of a quantity beyond an acceptable threshold which can be regarded as the primary quantification step along the workflow. The thresholds listed in Table 1 are usually based on well-known criteria concerning the strength of materials such as Tresca criterion, von Mises criterion, and critical strain theory. Take the surface fatigue index for instance, the Tresca criterion corresponds to the calculation

of FI_{surf} . Since the models are based on different assumptions, the exceedance criteria are often different quantities and therefore they are not comparable directly at the primary quantification step.

Moving further, the workflow converges for the methods at a common point of fatigue modelling as described in Fig. 3 where suitable exponential laws are used to determine the number of cycles for crack initiation (or crack initiation damage increment per loading cycle for the WLRM approach) from values obtained at the primary quantification step. This step is often complemented with laboratory measurements [3, 21] or on-track measurements [6, 22] based on which empirical exponent coefficients are derived to calculate life. At this step, the approaches converge to give similar physical quantities and are therefore more comparable as shown in Table 2.

It is important to note that an absolute comparison of damage increments from different approaches would not necessarily yield equal results even though they physically mean the same since the coefficients are calibrated to the specific phenomena modelled (certain length of crack on the surface or certain depth of cracks, etc.). Therefore, this work does not seek to establish which model is the accurate one since all models are built and calibrated differently according to the requirements. However, comparison of normalized values from different models and studying their correlation would help understand how different parameters influence damage increments in each approach. The normalized damage values are visualized in a logarithmic scale. The following steps are suggested in this comparison framework:

- (1) Define simulation cases that contain representative wheel–rail contact scenarios such as flange contact, curve radii, and friction levels.
- (2) Perform damage quantification using all approaches to obtain damage increment values ΔD_{ij} per loading cycle, where i refers to the quantification approach and j indicates the simulation case.
- (3) Normalize ΔD_{ij} w.r.t the maximum damage increment value obtained for the given approach $\Delta D_{i-\text{max}}$ to obtain $\Delta D_{ij-\text{norm}}$.
- (4) Compare normalized results $\Delta D_{ij-\text{norm}}$ in the logarithmic axis.
- (5) Compare trends formed by $\Delta D_{ij-\text{norm}}$ for different approaches ($j = 1, 2, \dots$) in a logarithmic scale.

The following section describes simulation cases which is followed by the corresponding results and detailed discussion on model similarities in Sect. 4 and Sect. 5, respectively.

Fig. 3 Comparing workflows of different RCF modelling approaches

3 Simulation cases

A previously created database of contact patch data from [23] has been used as input for the rolling contact fatigue investigations in this paper. These contact patch data have been generated by MBS of railway vehicles operating in a metro network in the commercial software SIMPACK [15]. The contact patch data include the contact patch position, the contact dimension, the normal and tangential forces, and the longitudinal, lateral and spin creeps. These results are available for all contact patches of all wheel/rail contact pairs of the vehicle.

Figure 4 presents a range of operating conditions for the considered underground metro system. Three vehicle types that are in service in the metro train system in Graz have been considered: one vehicle type with a conventional

(non-steering) bogie and two vehicle types with self-steering bogies. Three tractive conditions typical for metro operation have been considered:

- (1) Traction less rolling,
- (2) Application of traction to the powered axles of the vehicles corresponding to an acceleration of $a = 1.21 \text{ m/s}^2$, and
- (3) Braking equivalent to a deceleration of $a = -0.87 \text{ m/s}^2$.

The vehicle speed was kept constant at 19.7 m/s (71 km/h) in all simulations to ensure comparability. Simulations were carried out for three different curve radii (300, 600 and 1100 m). In all simulations, an unworn rail profile of type S48 has been used. Track irregularities (vertical, lateral, cross-level, and gauge irregularities) are

Table 1 Primary quantification steps for each approach

Model	Description of primary quantification
WLRM approach	Calculation of <i>energy dissipation</i> from forces (F_x and F_y) and creepages (γ_x and γ_y) $T\gamma = F_x\gamma_x + F_y\gamma_y$
Surface fatigue index approach	Calculation of surface fatigue index FI (<i>average shear stress at full slip</i>) from wheel–rail forces (F_x , F_y and F_z) and shakedown map (k/p_0) $FI = \frac{\sqrt{F_x^2 + F_y^2}}{F_z} - \frac{k}{p_0}$
KTH approach	Calculation of stress index SI (<i>maximum shear stress</i>) from contact stresses (τ_{zx} and τ_{zy}) and material yield limit in shear (k) $SI(x, y) = \sqrt{\tau_{zx}(x, y)^2 + \tau_{zy}(x, y)^2} - k$
Wedge model approach	Calculation of an effective stress for crack initiation σ_{eff} based on maximum principal stress value σ^* , a factor f_a assessing the near-surface plastic shear deformation condition, and a factor f_t addressing the tangential loading direction $\sigma_{eff} = \max_x(\sigma^*) \times f_a \times f_t$

Table 2 Fatigue modelling in different approaches along with materials used in calibration of model parameters

WLRM approach	Surface fatigue index approach	KTH Approach	Wedge model approach
$\Delta D = f(T\gamma)$	$\Delta D_i = \frac{(FI_{surf})^4}{10}$	$\Delta D_i = \alpha(SI_{max})^\beta$	$\Delta D_A = \left(\frac{\sigma_a}{\sigma_f}\right)^{-\frac{1}{b}}$
For f refer to Fig. 2a, where ΔD is the damage increment per loading cycle	Results from [5], where ΔD_i is the damage increment per loading cycle	Results from [4], where ΔD_i is the damage increment per loading cycle, SI_{max} is the maximum stress index, $\alpha = 4 \times 10^{-6}$, and $\beta = 3.7$	Results from [3, 12], where ΔD_A is the damage increment per loading cycle, σ_a is the modified maximum principal stress, $\sigma_f = 2,025 \times 10^6$ is a material constant, and $b = 0.085$ is the Basquin exponent value
Rail material R220 measured on track	Rail material R260 loaded by wheels of R7 material	260 Mn-grade rail and a harder MHH-grade rail on track	Rail material R350HT loaded by wheels of R7 material

Fig. 4 Distribution of parameters in the MBS simulations: **a** vehicle loading l (0: empty, 1: fully laden); **b** coefficient of friction μ ; **c** lateral acceleration a_q ; **d** profile set number p

applied to the constant radius part of the simulation model. The track irregularities have been measured on a 300 m radius curve of the simulated metro system. For each combination of operating condition, curve radius and vehicle type, 50 simulations have been carried out:

- In 23 of the simulation runs, the following parameters have been drawn from Gaussian distributions characterized by a mean value m and the standard deviation (SD) σ :
 - Vehicle loading l ($m = 0.5$, $\sigma = 0.1725$) ($l = 0$ representing an empty vehicle, $l = 1$ representing a fully laden vehicle),
 - Coefficient of friction f ($m = 0.35$, $\sigma = 0.05$), and
 - Lateral acceleration a_q ($m = 0.2 \text{ m/s}^2$, $\sigma = 0.067 \text{ m/s}^2$) (prescribed as cant deficiency in the simulation).

For each vehicle type, 10 sets of measured wheel profiles p that represent different wear states were used in the simulation. One set of wheel profiles consists of individual wheel profiles for all wheels of the vehicle model. The actual wheel profile set used in a particular simulation has been randomly chosen based on a uniform distribution.

- The remaining 27 simulation runs vary the vehicle loading l , the coefficient of friction f , the lateral acceleration a_q , and the profile set number p systematically with respect to the reference condition ($l = 0.5$, $\mu = 0.35$, $a_q = 0.2 \text{ m/s}^2$, and $p = 1$). Figure 4 shows the distribution of the parameters.

The MBS results of the wheel–rail contact cases remain common for RCF quantification using all the four models considered in this work. In the following section, the RCF results for the cases using all the four models are presented. Similarities and differences in results are further discussed in Sect. 5.

4 Results

The damage quantification results from all four approaches for the studied contact patch cases on the high rails are plotted over energy dissipation $T\gamma$ (as done in the WLRM approach) in Fig. 5. The damage increment per load cycle (ΔD) for each case has been plotted on a logarithmic scale after normalizing the value w.r.t the highest damage increment obtained for the particular approach. They are marked in accordance to curve radii with different symbols

Fig. 5 Overall comparison of normalized damage increments of four methods: **a** WLRM; **b** surface fatigue index model; **c** KTH model; **d** Wedge model

and colours. Results in this section are depicted in the form of $\log_{10}(\Delta D_{\text{Norm}})$. The $\log_{10}(\Delta D_{\text{Norm}})$ notation is hereon referred to as ' nD_{log} ' for brevity.

In some of these plots, several cases can be seen to depict a nD_{log} value of -10 . These contact cases were identified as 'non-damaging' by the models and had negative or zero damage values. Therefore, these cases were manually assigned a very small damage increment value (10^{-10}) so that they are visible distinctly on the logarithmic plots. This is particularly relevant for the WLRM, surface fatigue index, and KTH models.

Also, approximate trend lines have been traced for the different approaches akin to the WLRM damage function depicted in the top-left corner with the green line. Additionally, the maximum cases for each approach have been indicated by larger markers filled with yellow colour on the maximum line at $nD_{\text{log}} = 0$. The local approaches (Wedge and KTH models) indicate the maximum damage for the same case, depicted by the red circles ($R = 300$ m) on the zero-line. The maximum values for the global methods on the other hand occur for different cases, both occurring for a curve radius of 600 m (black triangles).

With increasing model complexity, the number of cases quantified as damaging increases. Starting from the observations in the surface fatigue index approach in

Fig. 5b, there are cases from all three curve radii reaching up to an energy dissipation value of 220 Nm/m yet classified as non-damaging. When we proceed one step further to the KTH model in Fig. 5c, the number of cases decreases and is mainly restricted to cases from the higher curve radius of 1100 m for energy dissipation values below 100 Nm/m. Moving further to Wedge model, fewer cases have been quantified as non-damaging. While the KTH model places a threshold in the form of yield limit (k) at this step for damage quantification, the Wedge model formulates the same in the form of a multiplication factor. In other words, a non-damaging case would be indicated by a negative number or zero ($-\infty$ on the logarithmic scale) by the KTH model at the primary quantification step. For the Wedge model, however, the same would be indicated by a very small ratio at the primary quantification step. This is consistent with nD_{log} values observed at lower energy dissipation values in Fig. 5. The KTH model and the surface fatigue index model are more binary in this region wherein they quantify damage or outright classify them as non-damaging. For the Wedge model, however, a dense cloud of points at very low values (-10 to -7) implies a very small but nevertheless a damage increment.

As expected, a tighter curve radius results in higher energy dissipation values, indicated by the horizontal

Fig. 6 Classification of cases based on traction for local methods

position of the red dots in all approaches. Amongst these, the flange contact cases have higher energy dissipation values. The WLRM approach assigns no damage increment to almost all cases pertaining to the 300 m radius case, signifying the dominating effect of wear at higher energy dissipation values. Other approaches, however, show damage increments for all curve radii albeit at different magnitudes. The surface fatigue index approach assigns comparatively higher normalized damage to all cases (indicated by the curve), especially so for cases with larger curve radii. Coming to the KTH approach, the trends are similar to the surface fatigue index but show greater spread of damage increment values for cases at the lower end of energy dissipation (therefore, the larger curve radii). Also, the rough damage function indicates higher damage for the 600 m radius cases when compared to the 1100 m radius case. This inference has earlier been observed in [1] where

higher concentrations of RCF were observed for curve radii of 600 m. The spread of damage increment values for the cases at lower energy dissipation only increase further for the more complex Wedge model. This is particularly significant when analysis is restricted to curves with larger radii where the energy dissipation values lie below 100–120 Nm/m.

When the local methods are compared w.r.t each other, cases corresponding to $R = 300$ m show significantly lower damage increment values for the Wedge model w.r.t the KTH model. Also, the KTH model predicts less damage at larger curve radii ($R = 1100$ m) with respect to the Wedge model. However, the shape and (therefore) the relative trend between individual cases for a given curve radius stay the same in both the local methods.

Figure 6 compares $nD_{\log}$ of all cases for the local models against the specific wear number (i.e. the energy

Fig. 7 General comparison of damage results from different models: **a** percentage of nonzero damage cases; **b** similarity percentage of damaging casesets; **c** cross-correlation of damage quantification in all cases

dissipated per unit area) of the contact patch. They are presented individually for each curve radius and classified w.r.t traction, free-rolling and braking cases, respectively (symbol/colour). Also, the trailing wheelset contact cases are classified separately at the lower energy dissipation values. It can be surmised that in most of the cases for a given damage increment value, the tractive contact cases have the highest energy dissipation values, followed by the rolling and braking contact cases, respectively. Conversely, for contact cases at a given energy dissipation value, generally higher damage increments can be expected for the braking case which is then followed by the rolling and tractive cases, respectively. This relative trend w.r.t traction remains the same for all curve radii cases. This trend can also be observed in contact cases arising on the trailing wheelset. However, the damage increments are substantially lower and concentrated on the lower energy dissipation spectrum as seen in the diagram.

5 Model similarities and discussion

In the previous section, results from different methods for various scenarios were compared to each other, within the context of their location on the energy dissipation spectrum. In this section, the agreement between various models is further discussed. This is done in two steps:

- General comparison of damage quantification results for all contact cases, i.e. the similarity of casesets identified as damaging by each model and the correlation of the respective values.
- Parametric influence on the degree of agreement.

5.1 General comparison

The percentages of nonzero damage cases for each model, P_i , are plotted in Fig. 7a. For a given model ‘ i ’ and a given number of contact cases studied ‘ N_0 ’, they are calculated as

$$P_i = \left(\frac{|\{\Delta D_{\text{Norm}} > 0\}|}{N_0} \right)_i, \quad (1)$$

The percentages of damaging cases increase in the order of Surface fatigue index < WLRM < KTH < Wedge model, signifying a general trend of how more complex models assign damage, albeit in very small increments, already reiterated from the previous section. The WLRM captures about 78% of the cases as damaging, the remaining 22% ($\Delta D_{\text{Norm}} < 0$) are not all ‘non-damaging’ from a strict sense since some of these points are captured at high energy dissipation values, due to flange contact. This only signifies that wear is dominating over RCF for these cases, but these cases are still classified as ‘non-damaging’ with respect to RCF.

5.1.1 Similarity of damaging casesets

It is necessary to find out whether models agree with each other in identifying the same set of contact cases qualitatively as ‘damaging’ for an overall comparison. This is achieved by examining the proportion of damaging contact cases that overlap amongst model combinations. For this, similarity percentage of caseset between two models ‘ i ’ and ‘ j ’ is given by the Jaccard similarity coefficient ‘ S_{i-j} ’ in in Fig. 7b as

$$S_{i-j} \leftrightarrow S_{j-i} = \frac{|\{\Delta D_{\text{Norm}} > 0\}_i \cap \{\Delta D_{\text{Norm}} > 0\}_j|}{|\{\Delta D_{\text{Norm}} > 0\}_i \cup \{\Delta D_{\text{Norm}} > 0\}_j|}, \quad (2)$$

where the numerator signifies the number of cases identified as damaging by both models and the denominator signifies the total number of cases identified as damaging by either model. As an example, $S_{\text{Wedge-KTH}} = 85\%$ indicates that amongst the total number of unique contact cases classified as damaging by Wedge and KTH models combined, there is an 85% overlap. This implies that the remaining 15% cases were classified as damaging by only one of the models (where the models' assessment disagrees). This index does not indicate whether the damage increment values for a given case obtained from both models are equal, but rather the 'similarity' of the caseset in being qualitatively classified as 'damaging'. In other words, the 'similarity' between two models is a measure of how much the assessment of a given contact case as 'damaging' by a given model agrees with the other. This is an important measure since models giving similar P_i may not necessarily agree for the same set of contact cases, making them very 'dissimilar' in their assessment despite similar percentage of nonzero damage cases.

The highest similarity is between the Wedge and KTH models amongst the combinations of four models. At the same time, the lowest similarity is seen between the WLRM and the surface fatigue index model. Model combinations involving either the WLRM or the surface fatigue index model generally exhibited lower similarity. This is broadly due to the higher number of contact cases identified as 'damaging' by the more detailed 'local' models (Wedge and KTH).

5.1.2 Correlation of damage increment values

Moving further, a measure for joint variability of results from both models is checked using correlation analysis in Fig. 7c. Higher correlation denotes higher values from one model mainly corresponding with higher values of the other and vice versa for lower values. For this, $nD_{\log}$ results from different models are checked for cross-correlation.

This is given by the Pearson correlation coefficient (ρ) for models 'i' and 'j' as

$$\rho_{i-j} \leftrightarrow \rho_{j-i} = \frac{\text{cov}\left((nD_{\log})_i, (nD_{\log})_j\right)}{\sigma_{(nD_{\log})_i} \sigma_{(nD_{\log})_j}}, \quad (3)$$

where 'cov' denotes covariance function and σ denotes standard deviation. This value varies between -1 and 1 , where a negative value implies an inverse relationship, 0 implies very low correlation (i.e. random behaviour), and a value of 1 suggests linear proportionality.

Higher correlation is observed for the Wedge-KTH combination than for all other model combinations. This implies that the relative trend of results w.r.t varying contact cases is more similar between these models. At the same time, both Wedge-WLRM and KTH-WLRM combinations are poorly correlated. Keeping similarity analysis conducted in the previous subsection in mind, even though Wedge-WLRM and KTH-WLRM combinations might be reasonably 'similar' in qualitatively identifying cases as damaging, the results are not quantitatively correlated (i.e. do not agree w.r.t the trend). Amongst the model combinations studied, higher correlation in results is noticed for models lying next to each other in the damage complexity axis depicted in Fig. 1.

5.2 Parametric influence

While overall trends were discussed in the previous subsection, the micro-trends pertaining particularly to the influence of curve radius and traction are discussed in this subsection. Only the overlapping damaging cases across models ($\{\Delta D_{\text{Norm}} > 0\}_i \cap \{\Delta D_{\text{Norm}} > 0\}_j$) are considered. For this, the difference in logarithmic normalized damage increments ' $\Delta nD_{\log}$ ' from two models 'i' and 'j' is given by

$$(\Delta nD_{\log})_{ij} = (nD_{\log})_i - (nD_{\log})_j, \quad (4)$$

Fig. 8 Comparison of damage values between KTH and Wedge models: **a** comparison; **b** spread in difference

Fig. 9 Comparison of damage values between WLRM and Wedge model: **a** comparison; **b** spread in difference

and the corresponding spread of this difference is visualized using mean and standard deviations given by ' $\mu(\Delta nD_{\log})_{ij}$ ' and ' $\sigma(\Delta nD_{\log})_{ij}$ ', respectively. The analysis of $\Delta nD_{\log}$ values helps to visualize the factor by which the damage increments from both models (ΔD_i and ΔD_j) differ.

This is visualized in Fig. 8 for results from KTH and Wedge models. In the left plot (Fig. 8a), normalized damage is plotted for each case with the x - and y -axes indicating $nD_{\log}$ for the models being compared. Each point describes a unique contact case marked w.r.t curve radius (colours) and traction (symbol). The diagonal green line in the middle represents the 'equivalence' line. The $nD_{\log}$ is more equivalent for both models if the point is closer towards the equivalence line. The equivalence between two models is analysed further on the plot towards the right. Representative Gaussian distributions are plotted using $\mu(\Delta nD_{\log})_{ij}$ and $\sigma(\Delta nD_{\log})_{ij}$. High equivalence between models is achieved when $\mu(\Delta nD_{\log})_{ij}$ and $\sigma(\Delta nD_{\log})_{ij}$ have a value close to 0.

In Fig. 8, the distribution curves of $\Delta nD_{\log}$ are plotted w.r.t curve radius and traction for $\Delta nD_{\log-ij}$ between Wedge and KTH models ($\Delta nD_{\log-KTH-Wedge}$). The distributions of contact cases at $R = 600$ m (indicated by black markers) are close to the equivalence line, implying high equivalence between models. For lower curve radius ($R = 300$ m) represented by the red curves, the distribution curves are offset towards the right (i.e. the position of $\mu(\Delta nD_{\log})_{ij}$), implying that the KTH model assigns higher damage increment values than the Wedge model. For larger curve radius ($R = 1100$ m) represented by blue distributions, the distribution curves are offset towards the left, implying that the Wedge model assigns higher damage increment values than the KTH model. The offsets in distributions due to the influence of traction are much smaller and not as strongly noticed as for curve radius.

Overall, the combinations involving the WLRM show higher offset $\mu(\Delta nD_{\log})_{i-WLRM}$ as the Wedge-WLRM combination in Fig. 9 indicates. The offset for the cases at

$R = 300$ m indicated with red distributions are particularly high (~ 6), indicating how the WLRM tends to predict damage increments several magnitudes higher than the Wedge model. Similar observations also hold for WLRM-KTH, WLRM-surface fatigue index combinations, respectively. The distribution plots for the other four model combinations have also been analysed and are listed in the appendix.

The summary of the parametric influence amongst all six model combinations is characterized by the mean and standard deviations ($\mu(\Delta nD_{\log})_{ij}$ and $\sigma(\Delta nD_{\log})_{ij}$) as visualized for two model combinations in Fig. 8 and Fig. 9. These plots help to visualize the influence of individual parameters while comparing two different RCF quantification models. As indicated in the figures, these plots need to be read within the context of similarity values as described in Sect. 5.1.1 since a higher similarity indicates that the distributions are constructed with a higher number of (overlapping) contact cases.

5.3 Overall evaluation

General statistics from Sect. 5.1 and parametric analyses from Sect. 5.2 are consolidated and presented together for all six model combinations in Table 3. The values are colour-coded w.r.t the degree of agreement in the combinations with green indicating a high degree of agreement while red indicating the opposite. The 'degree of agreement' indicates the relative closeness of results from the models studied. This is associated with higher similarity and correlation values. It also corresponds to high equivalence in the form of low offset $\mu(\Delta nD_{\log})_{ij}$ and low spread $\sigma(\Delta nD_{\log})_{ij}$ in the comparison of results between the two models.

Amongst the six combinations studied, the Wedge-KTH combination shows the best overall agreement. The similarity and correlation are the highest in the combination, implying a general agreement in both identifying the

Table 3 Overall and parameter-wise evaluation of RCF damage quantification results in $(\Delta nD_{\log})_{ij}$ for various model combinations ' $i-j$ ' (green–agreement; red–disagreement)

Model combinations ($i-j$)	Wedge-KTH	Wedge-FI	Wedge-WLRM	KTH-FI	KTH-WLRM	FI-WLRM
Parameter for ($nD_{\log}$)						
	Overall statistics					
Similarity	85%	61%	77%	57%	68%	50%
Correlation	0.51	0.35	0.06	0.17	0.05	0.17
Overall mean $\mu_T(\Delta nD_{\log})_{ij}^*$	0.45	0.69	3.32	1.51	2.89	1.05
Overall standard deviation $\sigma_T(\Delta nD_{\log})_{ij}^*$	2.69	1.42	3.50	2.24	2.67	1.77
Parameter Influence of individual parameters						
Mean offset in results*						
$(\mu(\Delta nD_{\log})_{ij})^*$						
$R = 300$ m	3.00	1.80	5.48	-0.16	2.62	0.60
$R = 600$ m	-0.26	0.93	1.99	1.26	2.25	1.19
$R = 1100$ m	-2.94	-0.13	1.73	2.63	4.53	1.50
Tractive	-0.64	0.51	-0.95	0.54	-0.25	0.20
Braking	0.56	-1.40	0.60	-1.77	0.00	-0.14
Rolling	0.09	-0.18	0.32	0.07	0.21	-0.01
Standard deviation*						
$(\sigma(\Delta nD_{\log})_{ij})^*$						
$R = 300$ m	1.46	1.71	3.74	2.19	2.63	1.87
$R = 600$ m	1.46	1.03	2.98	1.58	2.32	1.61
$R = 1100$ m	1.58	1.30	1.81	2.46	3.04	2.10
Tractive	1.32	0.88	3.41	1.72	2.90	1.73
Braking	1.72	1.71	3.07	3.22	2.44	1.71
Rolling	1.21	0.98	2.67	1.45	2.49	1.97
<div> <div>Low agreement</div> <div> </div> <div>High agreement</div> </div>						
* A low mean value and standard deviation value only when taken together denotes high agreement FI denotes the surface fatigue index model						

damaging cases and the relative trends in the damage increment values. Even though the overall offset $\mu_T(\Delta nD_{\log})_{ij}$ is low, the overall spread $\sigma_T(\Delta nD_{\log})_{ij}$ is still high as we detect a strong radius-dependent mean offset in Fig. 8, visible in Table 3. However, if the influence of traction on the degree of agreement between KTH and Wedge models is considered, there is high agreement between the models, indicated by the low $\mu(\Delta nD_{\log})_{ij}$ and $\sigma(\Delta nD_{\log})_{ij}$ values for all tractive scenarios. This implies that if the strong radius-dependent mean offset is accounted for, the damage functions for both models such as the ones

plotted in Fig. 5 will coincide to a large extent, making the model predictions equivalent. This emphasizes the need to study parametric influence in addition to the overall comparison.

Following the Wedge-KTH combination, the next best combination is the Wedge-FI (fatigue index) combination that has lesser similarity and correlation compared to the former. This is followed by the KTH-FI combination. Even though these two combinations show slightly better values in the parametric influence section of Table 3, the

overlapping cases (indicated by similarity percentage) are considerably lower than for the KTH-Wedge combination.

The lowest overall agreement was found for the Wedge-WLRM combination. Even though it has a rather high similarity percentage (i.e. overlapping cases classified as damaging), the results are poorly correlated. Also, it is characterized by high overall offset $\mu_T(\Delta nD_{\log})_{ij}$ and spread $\sigma_T(\Delta nD_{\log})_{ij}$. Even when the parametric influences are individually isolated and studied, there is still a very low degree of agreement with high $\mu(\Delta nD_{\log})_{ij}$ and $\sigma(\Delta nD_{\log})_{ij}$ values regardless of the radius or tractive scenario. The KTH-WLRM combination follows next with a similar trend.

One of the main reasons why both KTH and Wedge models have higher similarity and correlation is the inclusion of the load distribution step, indicated by Fig. 3. This step is lacking in the other two models—leading to lower correlation values in KTH-WLRM and Wedge-WLRM combinations for instance. Another reason is the way how RCF is modelled in the approaches. For instance, the surface fatigue approach considers only surface-initiated RCF cracks for quantification while the WLRM also considers the effect of wear in decreasing cracks, leading to very low agreement between the models at higher energy dissipation values (see Fig. 5). The framework provides a way to address the effects of various aspects of RCF modelling when comparing results as described.

On evaluating overall agreement amongst the six combinations, models placed next to each other on the damage complexity axis described in Fig. 1 were more likely to show higher agreement, understandably so due to more commonalities in their approaches. However, this does not necessitate equal results from both models as the strong parametric influence of radius show for the KTH-Wedge model combination. However, there is relative agreement in the similarity of damaging casesets and the trends formed by the corresponding models.

6 Conclusions

There are different models available to predict/quantify RCF crack initiation. Some of the well-known and widely used models have been studied in the present work. The conclusions from this work are listed as follows:

- The selection of a suitable damage model is limited by calculation constraints arising due to the trade-off between damage model complexity (and related prediction quality) and system complexity (size of the wheel–rail contact parameter space).
- The RCF damage models are based on different approaches making it difficult to compare the

predictions quantitatively since they use different physical quantities.

- The best possible quantity to compare is the normalized damage increment per loading cycle which is more common amongst different approaches.
- However, comparison of this quantity needs to be visualized on a logarithmic scale since different models use different empirically derived power laws.
- In keeping with the previous conclusions, a new methodology has been put forward allowing to compare prediction results from different RCF damage models. It can be summarized by the following sequence of steps:
 - Calculate damage increment values using each model for a set of simulation cases including a wide range of contact conditions and normalize them.
 - Calculate similarity coefficients to identify unique contact cases qualitatively identified as ‘damaging due to RCF’ by both models ($\{\Delta D_{\text{Norm}} > 0\}_i \cap \{\Delta D_{\text{Norm}} > 0\}_j$).
 - Conduct correlation analysis between the results of models to evaluate joint variability of results.
 - Evaluate the equivalence of damage increments of the overlapping cases w.r.t different parameters to check for parametric sensitivity in results.
 - High similarity and correlation of normalized damage increments denote high degree of agreement in results. Low offset in difference of results from both models ($\mu(\Delta nD_{\log})_{ij}$) and low spread $\sigma(\Delta nD_{\log})_{ij}$ denotes high agreement between the models (i.e. high equivalence).
- The methodology was demonstrated on four existing RCF quantification approaches.
 - The Wedge-KTH combination showed the best overall agreement. However, the differences in results are very sensitive to curve radius. This was followed by the Wedge-FI and KTH-FI combinations, respectively, that had lesser similarity and correlation compared to the former.
 - The models placed next to each other w.r.t damage model complexity (refer to Fig. 1) were more likely to show higher agreement in results.
- Damage models might qualitatively agree on whether contact cases are ‘damaging’ due to RCF but not quantitatively correlate with the damage increments.

Appendix: Parametric influence in model combinations

See Figs. 10, 11, 12, and 13.

Fig. 10 Comparison of damage values between Wedge and surface fatigue index models: **a** comparison; **b** spread in difference

Fig. 11 Comparison of damage values between KTH and surface fatigue index models: **a** comparison; **b** spread in difference

Fig. 12 Comparison of damage values between KTH model and WLRM: **a** comparison; **b** spread in difference

Fig. 13 Comparison of damage values between surface fatigue index model and WLRM: **a** comparison; **b** spread in difference

Acknowledgements This project has received funding from the Shift2Rail Joint Undertaking (JU) under the European Union's Horizon 2020 research and innovation programme under grant agreement (No. 826206). The content of this paper reflects only the authors' view, and the JU is not responsible for any use that may be made of the information it contains. Klaus Six and Gerald Trummer would like to acknowledge the financial support within the COMET K2 Competence Centres for Excellent Technologies from the Austrian Federal Ministry for Climate Action (BMK), the Austrian Federal Ministry for Digital and Economic Affairs (BMDW), the Province of Styria (Dept. 12) and the Styrian Business Promotion Agency (SFG). The Austrian Research Promotion Agency (FFG) has been authorized for the programme management.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Krishna VV, Hossein-nia S, Casanueva C, Stichel S (2020) Long term rail surface damage considering maintenance interventions. *Wear* 460–461:203462
- Six K, Mihalj T, Marte C et al (2020) Rolling contact fatigue behaviour of rails: Wedge model predictions in T-Gamma world. *Proc Inst Mech Eng Part F J Rail Rapid Transit* 234:1335–1345
- Trummer G, Marte C, Dietmaier P et al (2016) Modeling surface rolling contact fatigue crack initiation taking severe plastic shear deformation into account. *Wear* 352–353:136–145
- Dirks B, Enblom R, Ekberg A, Berg M (2015) The development of a crack propagation model for railway wheels and rails. *Fatigue Fract Eng Mater Struct* 38:1478–1491
- Ekberg A, Åkesson B, Kabo E (2014) Wheel/rail rolling contact fatigue - Probe, predict, prevent. *Wear* 314:2–12
- Burstow MC (2003) Whole life rail model application and development for RSSB—development of an RCF damage parameter. RSSB Report, AEATR-ES-2003–832, Issue 1, Rail Safety & Standards Boards, London, UK
- Meierhofer A (2015) A new wheel-rail creep force model based on elasto-plastic third body layers. Dissertation, Graz University of Technology
- Meierhofer A, Müller G, Six K, Dietmaier P (2016) A wheel-rail creep force model considering normal stress and temperature dependent plasticity effects. In: *The Dynamics of Vehicles on Roads and Tracks, Proceedings of the 24th Symposium of the International Association for Vehicle System Dynamics (IAVSD 2015)*. CRC Press, Graz, Austria
- Six K, Meierhofer A, Müller G, Dietmaier P (2015) Physical processes in wheel–rail contact and its implications on vehicle–track interaction. *Veh Syst Dyn* 53:635–650
- Hossein-Nia S, Sichani MS, Stichel S, Casanueva C (2018) Wheel life prediction model – an alternative to the FASTSIM algorithm for RCF. *Veh Syst Dyn* 56:1051–1071
- Hossein-Nia S (2017) On heavy-haul wheel damages using vehicle dynamics simulation. Dissertation, KTH Royal Institute of Technology
- Trummer G, Six K, Marte C et al (2014) An approximate model to predict near-surface ratcheting of rails under high traction coefficients. *Wear* 314:28–35
- Franklin FJ, Chung T, Kapoor A (2003) Ratcheting and fatigue-led wear in rail–wheel contact. *Fatigue Fract Eng Mater Struct* 26:949–955
- Persson I (2020) GENSYS. <http://www.gensys.se/index.html#>. Accessed 25 Feb 2021
- Simpack (2020) Dassault Systèmes, Vélizy-Villacoublay Cedex, France. <https://www.3ds.com/products-services/simulia/products/simpack/>. Accessed 25 Feb 2021
- Six K, Meierhofer A, Trummer G et al (2017) Plasticity in wheel–rail contact and its implications on vehicle–track interaction. *Proc Inst Mech Eng Part F J Rail Rapid Transit* 231:558–569
- Leitner T, Trummer G, Pippan R, Hohenwarter A (2018) Influence of severe plastic deformation and specimen orientation on the fatigue crack propagation behavior of a pearlitic steel. *Mater Sci Eng A* 710:260–270
- Trummer G, Marte C, Scheriau S et al (2016) Modeling wear and rolling contact fatigue: Parametric study and experimental results. *Wear* 366–367:71–77
- Johnson KL (1985) *Contact Mechanics*. Cambridge University Press, Cambridge
- Jendel T (2002) Prediction of wheel profile wear—comparisons with field measurements. *Wear* 253:89–99

21. Ekberg A, Kabo E, Andersson H (2002) An engineering model for prediction of rolling contact fatigue of railway wheels. *Fatigue Fract Eng Mater Struct* 25:899–909
22. Dirks B, Enblom R, Berg M (2016) Prediction of wheel profile wear and crack growth – comparisons with measurements. *Wear* 366–367:84–94
23. Trummer G, Scheriau S, Dietmaier P, Six K (2021) Reproducing rolling contact fatigue relevant loading conditions of railway operation on a test rig. *Tribol - Mater Surfaces Interfaces* 15:127–137