

Validation of a Physics-based Prognostic Model with Incomplete Data: A Rail Wear Case Study

Annemieke Meghoe, Richard Loendersloot and Tiedo Tinga

University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands

a.a.meghoe@utwente.nl

r.loendersloot@utwente.nl

t.tinga@utwente.nl

ABSTRACT

While the development of prognostic models is nowadays rather feasible, its implementation and validation can still create many challenges. One of the main challenges is the lack of high-quality input data like operational data, environmental data, maintenance data and the limited amount of degradation or failure data. The uncertainty in the output of the prognostic model needs to be quantified before it can be utilised for either model validation or actual maintenance decision support. This study, therefore, proposes a generic framework for prognostic model validation with limited data based on uncertainty propagation. This is realised by using sensitivity indices, correlation coefficients, Monte Carlo simulations and analytical approaches. For demonstration purposes, a rail wear prognostic model is used. The demonstration concludes that by following the generic framework, the prognostic model can be validated, and as a result, realistic maintenance advice can be given to rail infrastructure managers, even when limited data is available.

1. INTRODUCTION

Prognostic models are used to predict the future behaviour of engineering systems, including their Remaining Useful Life (RUL). The RUL is used in the maintenance decision-making process, and it facilitates just-in-time maintenance. To enable prognostics-based maintenance decision-making, it is important to include all sources of uncertainty in the RUL prediction (Sankararaman, 2015). These sources of uncertainty, which include measurement, modelling and usage uncertainty, have been reviewed by several researchers including Sankararaman (2015), Atamuradov, Medjaher, and Noureddine (2017) and Corbetta, Kulkarni, Banerjee, and Robinson (2021). Corbetta et al. discusses four main categories of uncertainty sources, namely model, method, mea-

sure and input. The sources in the model category include the uncertainty in the process of model development. Uncertainties arise from the assumptions made to represent physical processes as realistically as possible through various mathematical formulations. Furthermore, the method category includes algorithms, computation tools and numerical errors. While the uncertainties of the measured category arise from incomplete measurements, measuring equipment, human errors, measuring processes and sensor installation and calibration. Finally, the input category consists of external variables such as environmental conditions, operational conditions and initial and boundary conditions. These four main categories of uncertainty sources are matched with the PHM challenges as illustrated by (Atamuradov et al., 2017), see Figure 1. The main challenges are related to either the model or the data such as varying operational conditions, lack of data collection, efficient approach of model validation, limited amount of data, quality of data and improper usage of incomplete data (Atamuradov et al., 2017; Jardine, Lin, & Banjevic, 2006; Heng, Zhang, Tan, & Mathew, 2009).

A very often occurring problem is that only some of the input and output parameters are monitored, and therefore quite some others are not readily available. This makes especially the prognostic model validation challenging, as model predictions based on an incomplete range of input parameters must be compared to a limited set of measured results (e.g. time to failure). Zhang, Xiong, He, and Pecht (2019) and Tamssaouet, Nguyen, Medjaher, and Orchard (2021) have discussed the validation of their proposed prognostic models including model parameter update methods based on the Bayesian learning framework and the particle filter method. Both studies discuss how to deal with a limited amount of data in the process of model parameter determination, and how to preserve part of that data to be used for model validation. However, in these approaches, the prognostic models are either data-driven or degradation-based, which anyhow need a considerable data set for model training and parameter setting (or updating). Validation is then in princi-

Annemieke Meghoe et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://doi.org/10.36001/IJPHM.2023.v14i1.3283>

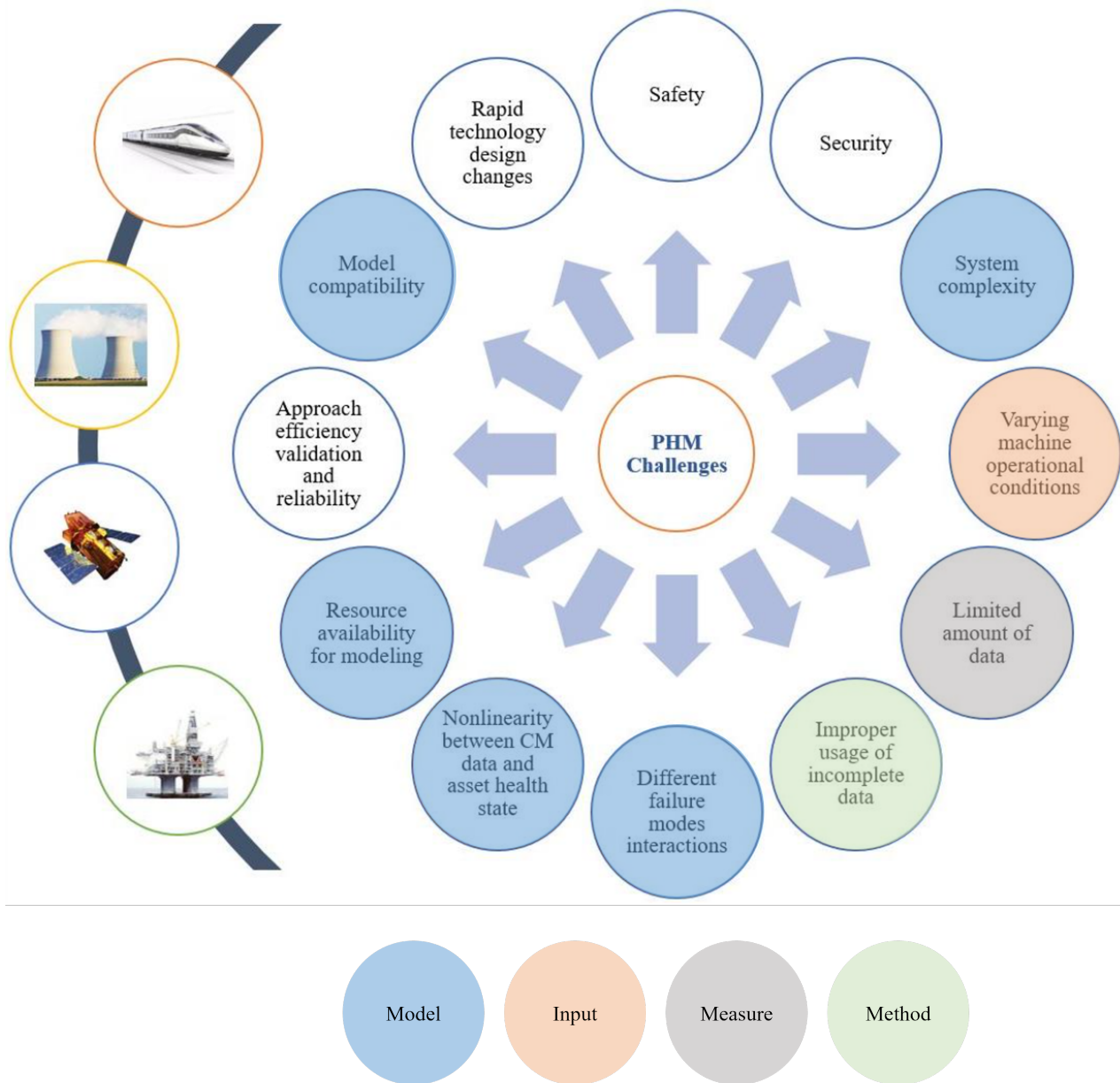


Figure 1. PHM challenges adopted from (Atamuradov et al., 2017) including macro-categories of uncertainty sources.

ple always possible, although the total amount of data required for optimal training and validation might not always be available. In the specific case of the current study, only two measuring points are available. This directly excludes any data-driven prognostic approach, since the training data set is much too small. The proposed physics-based prognostic model is still feasible in such a case, but with only two observed ‘failures’, the data available for parameter determination and model validation is very limited. This situation typically occurs in general when less than 10-20 ‘failures’ are available, and most data-driven methods are unfeasible. As a

solution to this problem, stochastic models are used to evaluate the uncertainty propagation in the model response due to uncertainty in the input parameters (Nejadseyfi, Geijselaers, & van den Boogaard, 2018). Uncertainty propagation analyses have been performed in various scientific research fields, ranging from healthcare (Barchiesi, Kessentini, & Grosge, 2011) and environmental modelling (Uusitalo, Lehtikoinen, Helle, & Myrberg, 2015) to engineering design (Chen, Jin, & Sudjianto, 2005; Nejadseyfi et al., 2018) and maintenance modelling (Gao & Zhang, 2008). Over the past decades, various uncertainty analysis methods have been developed.

These methods include perturbation methods using Taylor-series expansions, Monte Carlo, and analytical methods using integrals and tensor product basis functions. The approximation with Taylor-series expansion predicts the mean and variance of the model response and is used due to its simplicity (Parkinson, Sorensen, & Pourhassan, 1993), but for nonlinear response functions, when the uncertainty in the input parameter is relatively large, this method is not sufficiently accurate (Wu, Millwater, & Cruse, 1990). On the other hand, the Monte Carlo method can handle both linear and nonlinear models. Hills and Trucano (1999) concluded that the Monte Carlo method is a robust method when considering highly nonlinear models. However, the disadvantage is that the results are not always reproducible, fast and accurate. The computational expense can be reduced when Monte Carlo is used together with meta-models. For non-complex models that are represented by well-established meta-models like polynomials, Kriging, Radial Basis Function, etc., the analytical approach is suitable (Chen, Baghdasaryan, Buranathiti, & Cao, 2004).

This paper proposes a generic method to validate a physics-based prognostic model with a limited or incomplete data set using uncertainty propagation analysis. As a case study, the rail wear prognostic meta-model as previously developed by the authors in (Meghoe, Loendersloot, & Tinga, 2019) will be used. The following section will first illustrate the proposed generic framework followed by the rail wear model. Thereafter the sensitivity and correlation analysis that determines whether probability density functions for the input parameters must be used will be discussed in Section 2.3. Then the uncertainty propagation analysis using the Monte Carlo method is introduced in Section 2.3 and the formulations for the analytical approach are presented in Section 2.5. Section 3 starts with an introduction to the case study, followed by the results of the several analyses (see Section 3.2 - 3.4). Finally, the obtained probability density functions for the rail wear response using numerical and analytical approaches are compared with field measurement data in Section 3.5 and some conclusions are drawn in Section 4.

2. METHODOLOGY

This section presents the proposed generic framework, the rail wear model and the formulations for the numerical and analytical methods for uncertainty propagation. To successfully validate a prognostic model, the observed or measured results should be within the confidence bounds of the predicted results obtained from the numerical and analytical methods. This confidence bound is derived from the uncertainty in the response value obtained by considering stochastically determined input parameters.

2.1. Generic framework

The proposed generic framework of this study is presented in Figure 2. A prognostic model and input data from field measurements that will be used to evaluate the prognostic model, are the pillars of the framework. First a sensitivity analysis is performed. In this sensitivity analysis the minimum and maximum value of each input parameter set is chosen to check which parameter has the highest influence on the response value. Also the correlation of these parameters with the response value is checked to conclude which parameters should be considered as deterministic or as stochastic parameters. Thereafter, either a numerical or an analytical analysis is carried out to quantify the uncertainty in the response value. This response value along with its uncertainty is then compared with field measurements. If the resulting output of the analysis is in agreement with the field measurements then the prognostic model is considered to be validated. If not, then the output reveals which input parameter(s) should be measured more accurately and new data should be gathered from field measurements, or the prognostic model should be improved. In this study, both a numerical and an analytical analysis are carried out in order to investigate which method is computationally efficient with a sufficient level of accuracy. It should be mentioned that the infrastructure manager should determine the threshold for validation in order to decide on how to proceed further, i.e. to collect more data or to modify the model.

2.2. Rail wear model

From literature and maintenance experience, it can be concluded that the RUL of railway tracks is mainly determined by either the wear (Enblom, 2009; Sheinman, 2012) or Rolling Contact Fatigue (RCF) mechanism (Soleimani & Moavenian, 2017). RCF is kept under control by preventive grinding (also referred to as artificial wear) of rails. Rails are replaced when the amount of material removed from the railhead, resulting from natural and artificial wear, exceeds a certain critical value implying the RUL equals zero. The amount of railhead material removal is measured by using a laser scanning system mounted on measurement trains. Maintenance interventions follow from these measurements, supported by a linear wear rate model as obtained from previous measurements. Thus, the rail wear prediction is based on historic data and does not include different usage profiles and changing environmental conditions. Accurate prediction of the total amount of rail wear by considering the varying operational conditions is expected to considerably improve maintenance planning and reduce maintenance costs.

Several wheel/rail wear prediction models have been developed in the past decades, such as the models of Zobory (1997) and Dirks (2015). These models include vehicle dynamics, local contact theory, and wear calculation based on Archard's

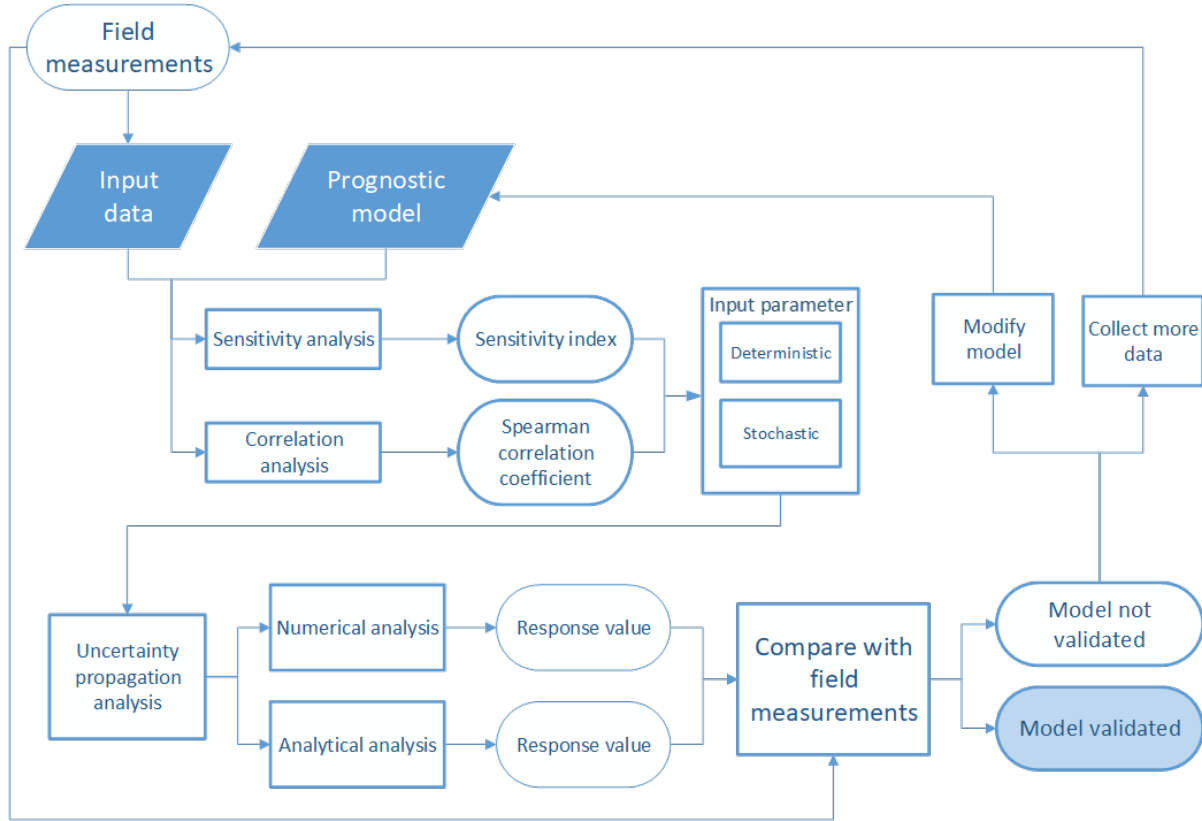


Figure 2. Generic framework for model validation.

wear law (Archard, 1953) and energy dissipation (Pearce & Sherratt, 1991). However, a relatively limited amount of attention has been paid to rail wear prediction as compared to wheel wear prediction. This is due to the fact that the wear rate of wheels is higher. Hence, extensive descriptions on the wheel wear prediction procedure can be found in literature (Zobory, 1997; Jendel, 2002; Enblom, 2009; Ignesti, Innocenti, Marini, Meli, & Rindi, 2014; Ramalho, 2015). (Enblom, 2009) proposed a procedure to predict rail wear similar to that of wheel wear. The amount of rail wear was overestimated when compared to field measurements and the authors therefore suggested to proportionally scale down the wear coefficients. (Dirks, 2015) also developed a model to predict wheel and rail wear, but this model was only verified for the wheels. It should be noted that all the methods developed so far propose a numerically expensive procedure for the wheel/rail wear prediction. Despite their proven validity, they have not been implemented by infrastructure managers in the maintenance decision-making process due to their high computational costs and difficulty in monitoring the required parameters. Furthermore, the list of parameters that could have a considerable influence on rail wear is extensive, and not all of these parameters are monitored.

The rail wear prognostic model used in this study is based on

Archard's wear law which indicates that the amount of wear volume loss is related to sliding distance s , normal force F_N and hardness of material H (Archard, 1953):

$$V = k \frac{s F_N}{H} \quad (1)$$

The wear coefficient k depends on the surface conditions and is usually determined empirically e.g. by a pin on disk configuration measurements. For simple calculations, the wear coefficient for rail and wheel steels can be derived from the wear map of Jendel (2002).

Using Archard's law to calculate the rail wear requires detailed numerical modelling of the dynamic wheel-rail contact (to obtain the normal force and sliding distance values). To reduce computation time, a set of meta-models has been created, that directly relate the wear rate to a set of operational input parameters, see Figure 3. The meta-models in this process are in the form of second-order polynomials and are defined as follows:

$$y(\mathbf{x}) = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{k+i} x_i^2 + \sum_{i=1}^{k-1} \sum_{j>i}^k \beta_{i,j} x_i x_j \quad (2)$$

where y is the rail wear area in mm^2 , x is the vector of x_i which are the various input parameters and β_i are the fitted

model parameters or fit coefficients (similar to those in equation (7) and (8)).

There are nine input parameters ($k=9$) in the rail wear prediction model which include axle load, curve radius, vehicle speed, the longitudinal and lateral stiffness of the primary bogie, rail profile geometry, material hardness, friction coefficient and rail cant. Furthermore, the rail profile geometry is represented by the vertical wear depth at the rail head.

Although the computationally expensive rail wear models have been replaced by meta-models, proper validation of this prognostic model with field measurements and quantification of the effect of input parameters' uncertainty on the maintenance decision-making process is still lacking. As the present paper forwards a generic framework to validate a prognostic model with a limited or incomplete data-set, the rail wear prognostic model is considered to be a suitable case study. The input parameters required for this model are derived from (incomplete) field observations, and the output of the model is then compared to periodic field measurements of the rail wear. Therefore, various sources of uncertainty are present that are unknown to the rail infrastructure managers. These variations are eventually included in the RUL prediction.

2.3. Sensitivity and correlation analysis

A sensitivity analysis followed by a correlation analysis is necessary prior to performing an uncertainty propagation analysis. The outcome of the sensitivity analysis includes the ranking of the input parameters according to the degree of influence on the response value. Whereas, the outcome of the correlation analysis reveals the input parameters which have a strong linear or non-linear correlation with the response value and, therefore, should be considered as stochastic variables. In this way, the analysis time of the uncertainty propagation analysis is minimised since the computational efficiency depends on the number of stochastic variables. The scatter in the stochastic input parameters is then represented by their probability distribution function (pdf).

The Sensitivity Index (SI) is calculated from two scenarios per input parameter i to estimate its individual sensitivity (Pannell, 1997):

$$SI_i = \frac{y_{max,i} - y_{min,i}}{y_{max,i}} \quad (i = 1, \dots, k) \quad (3)$$

where $y_{min,i}$ and $y_{max,i}$ represent the response values for the input parameter at its minimum and maximum setting, respectively.

Next, n Monte Carlo (MC) simulations with the prognostic model are performed to determine the Spearman correlation coefficients that can be calculated with the following equation

(Mukaka, 2012):

$$\rho(x_i, y) = 1 - \frac{6 \sum_{j=1}^n d_j}{n(n^2 - 1)} \quad (4)$$

where x_i is the input vector where only the i^{th} input parameter is varied such that n different values are taken in the MC simulation and thus n values of the response y are obtained. The n input (x_i) and n output (y) values are then ranked from highest to lowest and thereafter d_j is calculated as the difference in ranking. If x_i and y are ordered identically, d_j is always 0, and $\rho = 1$, which indicates a maximum positive correlation. The value of the correlation coefficient is generally between -1 and +1, where a correlation coefficient of 0 indicates that no correlation exists between the input parameter and response value. A strong correlation is represented by a relatively high (absolute) value for the correlation coefficient, i.e. close to ± 1 . If the input parameter has both a high correlation coefficient and a high sensitivity index then the associated input parameter has to be considered as a stochastic variable.

2.4. Numerical uncertainty propagation analysis

For the numerical uncertainty propagation analysis similar to the correlation analysis, MC simulation is applied to generate a number of samples from the input parameter's data sets, which are then used to evaluate the prognostic model. The first step includes stochastic selection of samples from the input's distribution, and then the resulting set of model outputs can be seen as a stochastic sample of the distribution of the output (Kennedy & O'Hagan, 2001). The mean, standard deviation and shape of the distribution can be used to quantify the uncertainty propagation. This usually requires a large number of model runs.

2.5. Analytical uncertainty propagation analysis

In the analytical approach of the uncertainty propagation analysis, the mean and variance of the response are calculated. Given a response $y = f(\mathbf{x})$ with \mathbf{x} being the vector with all input parameters which can be either stochastic or deterministic and \mathbf{x}_R being the vector with only the stochastic input parameters such that $\mathbf{x}_R \in \mathbb{R}$, the mean and variance of the response are defined as follows (Chen et al., 2005):

$$\mu(\mathbf{x}) = \int f(\mathbf{x}) p_R(\mathbf{x}_R) d\mathbf{x}_R \quad (5)$$

$$\sigma^2(\mathbf{x}) = \int [f(\mathbf{x}) - \mu(\mathbf{x})]^2 p_R(\mathbf{x}_R) d\mathbf{x}_R \quad (6)$$

where $p_R(\mathbf{x}_R)$ is the probability distribution function of the stochastic input parameters. The analytical approach assumes

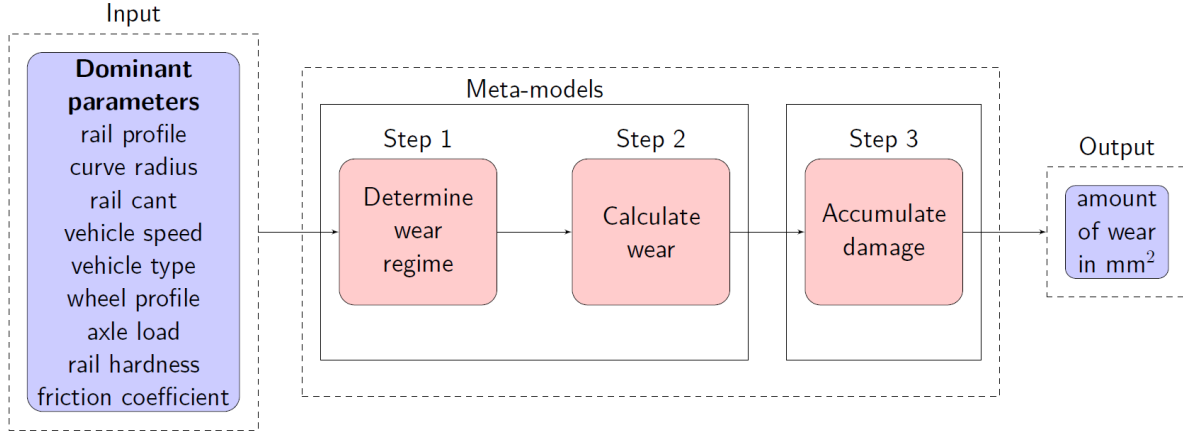


Figure 3. Rail wear estimation process by means of meta-models (Meghoo et al., 2019).

the response to be normally distributed as the input parameters are also assumed to be normally distributed.

Based on the concept of tensor product basis functions, the formulations for the mean and variance of a second-order polynomial model, which will be used in the case study of rail wear, and normally distributed input parameters are as follows (Chen et al., 2005):

$$\mu(\mathbf{x}) = \beta_0 + \sum_{i \in \mathcal{R}} \beta_i \mu_i + \sum_{i \in \mathcal{R}} \beta_{ii} \sigma_i^2 + \sum_{i \in \mathcal{R}} \sum_{j \in \mathcal{R}, j \geq i} \beta_{ij} \mu_i \mu_j + \sum_{i \notin \mathcal{R}} (\beta_i + \sum_{j \in \mathcal{R}} \beta_{ij} \mu_j) x_i + \sum_{i \notin \mathcal{R}} \sum_{j \notin \mathcal{R}, j \geq i} \beta_{ij} x_i x_j \quad (7)$$

$$\sigma^2(\mathbf{x}) = \sum_{i \in \mathcal{R}} \beta_{ii}^2 \sigma_i^4 + \sum_{i \in \mathcal{R}} \sum_{j \notin \mathcal{R}, j \geq i} \beta_{ii}^2 \sigma_i^2 \sigma_j^2 + \sum_{i \in \mathcal{R}} (\beta_i + \beta_{ii} \mu_i + \sum_{j \in \mathcal{R}} \beta_{ij} \mu_j + \sum_{j \notin \mathcal{R}} \beta_{ij} x_j)^2 \sigma_i^2 \quad (8)$$

where $\beta_{ji} = \beta_{ij}$ if $i > j$, μ_i and σ_i^2 are the mean and variance of the stochastic individual input parameters, x_i the values of the deterministic input parameters and the β 's are the model parameters (i.e. fit coefficients of the second order polynomial model).

3. DEMONSTRATION

In this section the demonstration of the generic framework as depicted in Figure 2 is discussed. First the case study is introduced, followed by the application of the several analyses and concluding with the results and discussion on these analyses.

3.1. Case study introduction

As mentioned in the introduction the rail wear prognostic model as described in Section 2.2 will be used as a case study to demonstrate the proposed framework. The field data which

is available for rail wear is not well defined. The environmental conditions, reflected in the friction coefficient, and the operational conditions, represented by the type of trains, are both unknown. This level of uncertainty is quite common when only observational field data is available instead of test data from well-controlled experiments. Therefore, the rail wear case is suitable to demonstrate the framework on how to validate a prognostic model with limited data.

The railway track chosen for the case study is located between the Dutch cities of Weesp and Almere. This rail section contains curvatures with radii $R = 1500$ metres and $R = 1800$ metres. The case study is further based on two measurement sets from two different time periods, namely from March 2018 to October 2018 for $R = 1500$ metres and from November 2018 to March 2019 for $R = 1800$ metres. These cases will be further regarded as case study 1 and 2, respectively.

The uncertainty propagation in the wear prediction model is calculated using both the numerical and analytical approach as described in section 2. Using the input parameter distributions, both approaches provide a distribution of the accumulated wear, which can finally be compared to the rail wear inspection data. From the accumulated wear distributions, the RUL can then be predicted with confidence intervals. After the model has been validated in this way, the rail wear prediction model will enable the rail infrastructure manager to make future RUL predictions with only limited data sets. Thus, initial estimates can be made before future usage profiles and environmental conditions are known. These estimates are valuable for initial (long-term) planning.

3.2. Sensitivity and correlation analysis

As depicted in Figure 2, a sensitivity and correlation analysis is performed and therefore, the minimum and maximum value for each input parameter of the wear prognostic model are given in Table 1. The distribution is assumed to be a nor-

mal distribution, where the minimum and maximum values of the input parameters represent all possible values that can occur in practice. For example, a wide range of friction coefficient values is given in Table 1 because the friction in wheel-rail contact depends on several operating environmental factors such as the ambient temperature, atmospheric humidity, rail surface condition, friction modifiers, water, sand, grease (lubrication), contamination, leaves, leaf mulch and water with iron oxides and wear debris (Trummer, Lee, Lewis, & Six, 2021; Rong et al., 2021).

Furthermore, it should be noted that some of the geometrical and operational parameters are exactly known and are therefore fixed. For example, the curve radius is known from the track geometry, and the vehicle velocity and the axle load are measured by a wayside measuring system called Quo Vadis. Therefore, for these three input parameters, no distribution is provided in Table 1 and they are considered as deterministic input parameters. The Quo Vadis system can also determine the passing vehicle type based on the distance between the shafts, the speed and the length of the vehicle. From the information about the vehicle type, the longitudinal and lateral stiffness of the primary suspension of the bogie can be extracted. However, not for all vehicle types the information could be retrieved and therefore, for the longitudinal and lateral stiffness of those vehicle types a normal distribution is used.

The calculated sensitivity indices and correlation coefficients, which are required to determine which of the input parameters are dominant and which parameter should be treated as a stochastic variable, are presented in Table 2. The sensitivity indices show that the friction coefficient and the hardness are the most dominant parameters. The Spearman correlation coefficients suggest that the friction coefficient has the strongest correlation with the response value followed by the longitudinal stiffness, the rail cant and the hardness. Therefore, it is decided to combine the results and thus to consider the friction coefficient, material hardness, rail cant (measured as the difference in height between the two rails), and lateral stiffness as stochastic variables. The pdf of the friction coefficient is obtained from (Popović, Lazarević, Brajović, & Vilotijević, 2015) and the pdfs of the next three parameters are obtained by field measurements. A normal distribution bounded by the minimum and maximum value (see Table 1) is chosen for the lateral stiffness, as the vehicle types that have passed the track are unknown.

Furthermore, the sensitivity indices and correlation coefficients show which parameters are positively or negatively correlated with the response value. For example, an increase in the friction coefficient leads to an increase in rail wear area, and an increase in the material hardness leads to a decrease in rail wear area. This is as expected according to Archard's wear law. The wear area is inversely proportional to the mate-

rial hardness. Furthermore, a higher friction coefficient leads to an increase in slip velocity and contact pressure, which are both proportional to wear according to Archard's wear law. The vertical wear depth, which represents the present condition of the rail profiles, is negatively correlated with the response value. This means that less wear is generated by a vehicle that travels over worn rail profiles. This is also as expected since the considered wheel profile in the analysis is assumed to be worn. According to Meghoo, Loendersloot, Bosman, and Tinga (2018) the combination of worn wheels and worn rail profiles generates a minimum amount of wear.

3.3. Numerical analysis

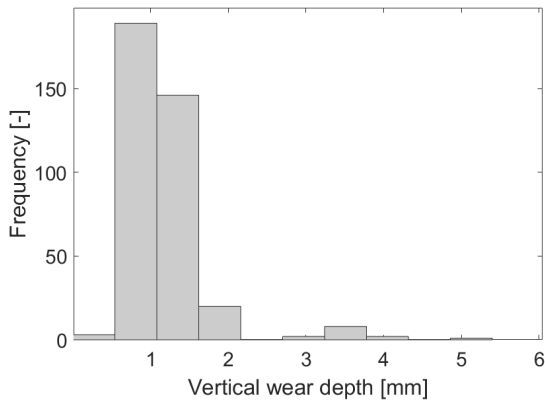
The input parameters that are used for the evaluation of rail wear are mainly obtained from field measurements. The histogram for the vertical wear depth, see Figure 4a, is plotted from the rail profile measurements. The histogram for the rail cant is also derived from measurements conducted by the same measuring device that was used to measure the rail profiles. Furthermore, the histogram for the hardness is plotted from measurements taken by using a portable hardness tester, at random locations of the specific track, see Figure 4b. As mentioned before, the histogram for the friction coefficient is taken from Popovici (2010), see Figure 4c. And for the vehicle types of which the stiffness values were unknown, a uniform distribution is assumed for both the longitudinal and the lateral stiffness with lower and upper bounds as given in Table 1.

The uncertainty in the wear model prediction using the numerical approach is obtained by evaluating the rail wear prognostic model as described in Section 2.2. To have significant wear on moderate curved tracks, e.g. on tracks with a curve radius larger than 1000 meters, approximately one million wheels need to pass the track. The evaluation of this takes hours of computation time and including the variation in the input parameters further increases the computational effort. Thus, considering the pdfs of the input parameters for each passing wheel in the Monte Carlo method will result in n evaluations for each wheel multiplied by one million wheels, where n is the number of MC samples. As this is way too much, the number of evaluations will be reduced from one million to 40 by clustering the wheel passages based on the type of wheel, type of train and axle load. Hence, the following steps are taken (also depicted in Figure 5):

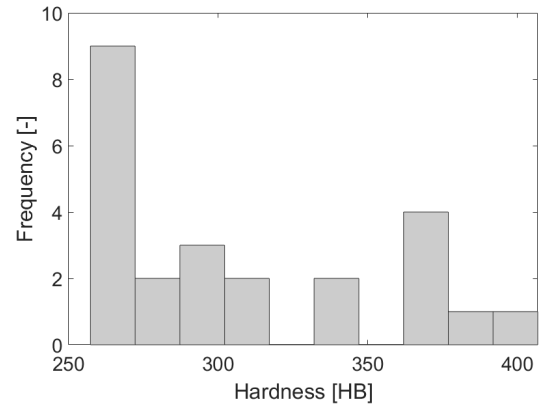
1. Split the data set of wheel passages based on the wheel type. In this case, two groups are created, namely for wheel type s1002 and HIT.
2. For each wheel type group, split the data set into known and unknown longitudinal and lateral stiffness values. This yields another two groups for each wheel type.
3. Split the data in each of the four groups in ten bins (or clusters) based on the axle load, which results in a total

Table 1. Input parameters and their ranges.

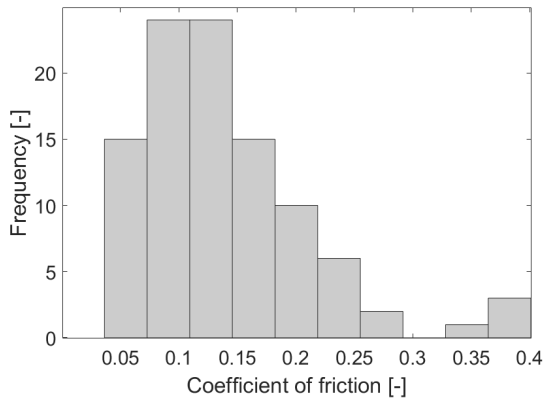
Input parameters	Min	Max	Mean	Std	Unit
Vertical wear depth	0	12	6	2	[mm]
Curve radius	1500	1500	1500	0	[m]
Rail cant	90	110	100	3	[mm]
Friction coefficient	0.02	0.6	0.3	0.1	[-]
Vehicle velocity	33.33	33.33	33.33	0	[m/s]
Axle load	15	15	15	0	[ton]
Longitudinal stiffness	2.56e5	4e7	2e7	8e6	[N/m]
Lateral stiffness	2.56e5	4e7	2e7	8e6	[N/m]
Hardness	260	340	289	14.45	[HB]



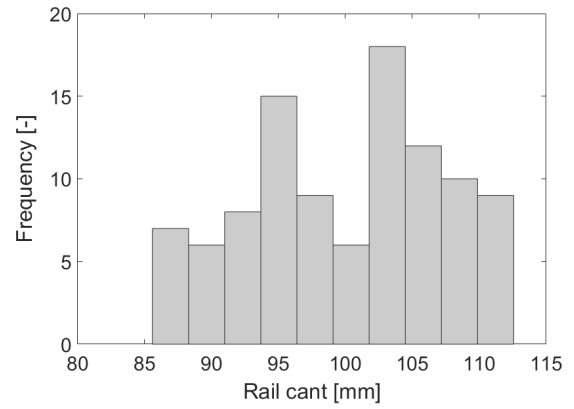
(a) Measured vertical wear depths



(b) Measured hardness



(c) Measured friction coefficients



(d) Measured rail cant

Figure 4. Histograms for stochastic input parameters.

Table 2. Sensitivity indices and Spearman correlation coefficients for wheeltype s1002.

Input parameters	Sensitivity Index	Spearman coeff.
Vertical wear depth	-0.07	-0.06
Rail cant	0.03	0.34
Friction coefficient	0.98	0.53
Longitudinal stiffness	0.07	0.40
Lateral stiffness	0.03	0.08
Hardness	-0.28	-0.20

of 40 bins. Determine the number of wheel passings in each bin.

- For each bin, sample the input parameter distribution functions n times using the MC method and define n input vectors with values for the k parameters.
- Evaluate the rail wear prognostic model (Equation 2) for each input vector in each bin. This process ultimately yields the probability density function for the wear area for each bin. After that, multiply the obtained pdf of the specific bin with the number of wheels in that bin.

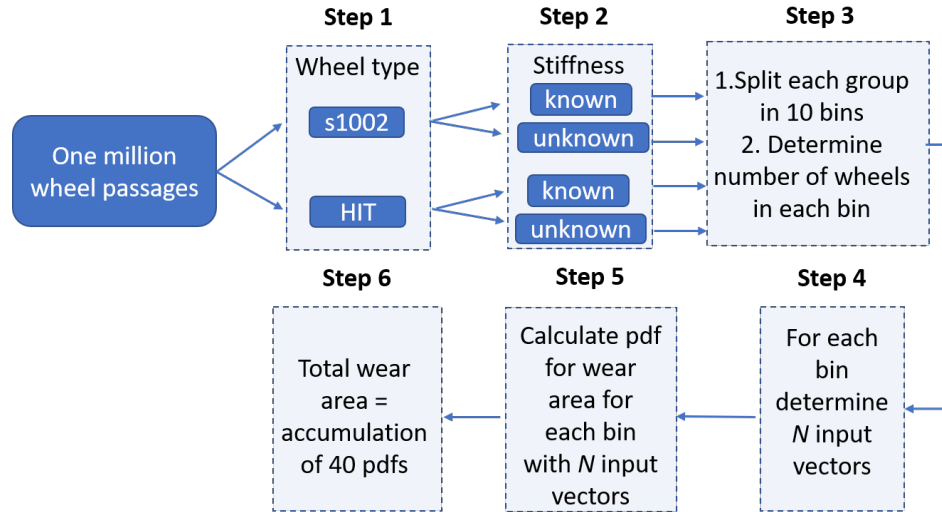


Figure 5. Reduction of number of wear evaluations.

- Accumulate all the results (i.e. pdfs) from the 40 bins to obtain the total response value, which in this case is the total pdf of the wear area.

In this way, the total number of evaluations is reduced from $1.10^6 \times n$ to $40 \times n$, where n is the number of MC simulations. From field observations, it can be concluded that the vehicle velocity for different wheel passages does not vary significantly from the average allowable speed on a specific track section. Therefore, the vehicle velocity is taken as a constant.

The approach for reducing the number of prognostic model evaluations is considered to be robust and reliable as the results obtained from following these steps were compared with the conservative method of evaluating each of the one million wheels separately. The resulting difference turned out to be only 7.3% (see Table 3), whereas the gained computational efficiency is enormous. Note that the number of bins is chosen based on a convergence study, indicating that using 40 bins yields the proper balance between model accuracy and computation time. The input parameters that are used in this convergence study are similar to the mean values for the input as presented in Table 1.

Table 3. Convergence study results.

Number of bins	Wear area [mm ²]	Difference [%]
40	22.99	7.3
400	22.28	4.0
20000	21.98	2.6
Actual number of wheels	21.43	0

3.4. Analytical analysis

The analytical approach as described in section 2.4 is applicable when the input parameters are normally distributed. Therefore, the mean and standard deviation of the actual distributions for the vertical wear depth, rail cant, material hardness and friction coefficient, as depicted in Figure 4, were calculated such that the pdfs are approximated as normal distributions. For the pdfs that were far from normal, multiple normal distributions with weight factors were used. Finally, the mean and standard deviation for the lateral stiffness were obtained from its uniform distribution. Using these inputs, equation 7 and 8 are then used to translate these uncertainties to the mean and variance of the prognostic model response, i.e. the amount of wear.

In Meghoo et al. (2019), it was shown that a single prognostic model did not fully cover all operating conditions. Therefore, multiple sets of model parameters (β_{ij}) were derived. In this work, these different parameter sets have been used to propagate the uncertainties (equation 7 and 8), and in the end one resulting mean and variance value have been determined. These overall values are obtained as the weighted average of the various models, using the number of passing wheels associated to each of them as weight factor.

3.5. Results and discussion

Table 4 shows the uncertainty propagation analysis results for case study 1 and 2, obtained from various MC simulations (with 100, 500 and 1000 samples), as well as from the analytical analysis. Table 4 also presents the measured wear area in terms of the mean and standard deviation, based on the measurements conducted at randomly selected locations along the curved tracks. The predicted wear area for a track with a cer-

tain curve radius is assumed to represent the average wear over its entire length.

Table 4. Wear area results in [mm²]

Method	Case study 1		Case study 2	
	Mean	Std	Mean	Std
Measurement	14.31	5.97	5.04	2.18
MC100 (normal distribution)	23.18	11.62	-	-
MC100 (actual distribution)	20.52	12.15	6.47	3.97
MC500 (actual distribution)	19.38	12.08	6.39	3.80
MC1000 (actual distribution)	19.49	11.21	6.01	3.73
Analytical	20.81	11.77	6.83	3.59

The results in Table 4 show that the predicted wear area is closer to the measured wear area for an increasing number of MC simulations. However, the difference between these simulations is not very large, so 100 simulations are considered to be an adequate number. For the MC simulations of case study 1 with 100 samples, both the actual distribution and an assumed normal distribution for the input parameters are used. It can be concluded that the obtained mean and standard deviation for the response value are closer to the measured results when using the actual pdfs for the input parameters than when using the normal distributions. The relative difference error (calculated with respect to the mean of the measured results) between the two methods is 18.58%. This means that in case of limited data, assuming a normal distribution can lead to a prediction with approximately 20% offset. This also implies that by using a normal distribution in this specific case a conservative maintenance decision will be taken, which means that the rail is replaced while it was still in relatively good condition.

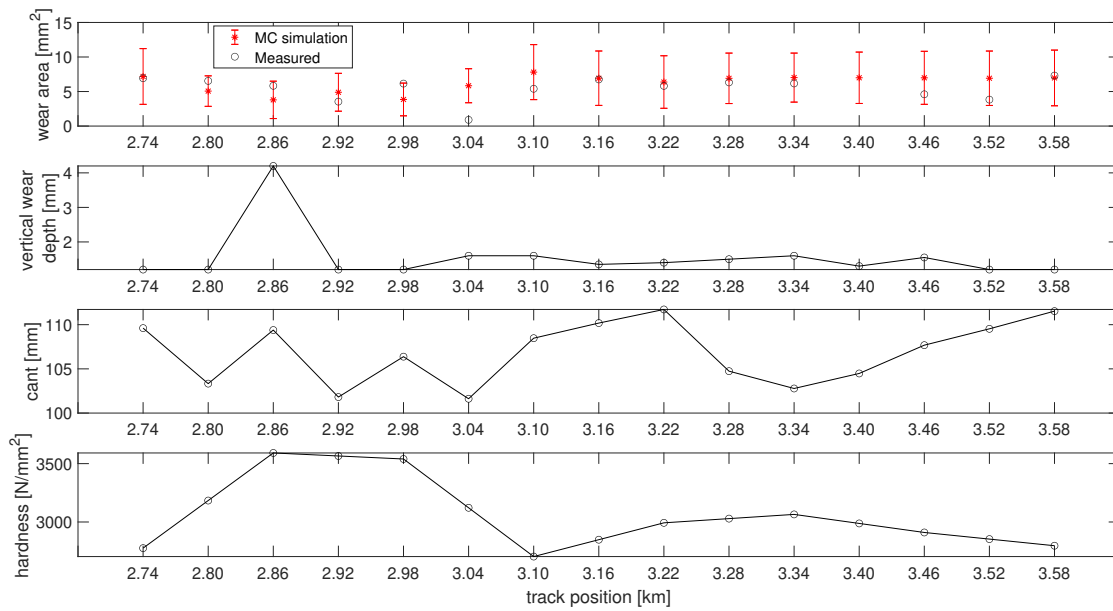
Table 4 also presents the results for case study 2. For case study 2, only the friction coefficient and the lateral stiffness are considered as stochastic input parameters as the remaining parameters, such as the vertical wear depth, the rail cant and the material hardness, were locally measured. The relative difference between the mean of the MC simulations with 100 samples and the mean of the measured results for case study 2 is 28.37% whereas the relative difference for case study 1 is 43.40%. These results show the significance of an improved data set. But even knowing three additional input parameters, the relative difference of approximately 30% is still rather high. This may be considered unacceptable, but a direct comparison of the mean values of these two distributions is not appropriate, as will be discussed below. However, the deviation of this specific measurement from the simulated results means that either the friction coefficient or the lateral stiffness (which are both uncertain in this case) have a significant influence on rail wear. From the sensitivity and correlation analysis results (see Table 2) it can be seen that the friction coefficient has both a higher sensitivity index and a higher correlation coefficient value compared to the lateral

stiffness.

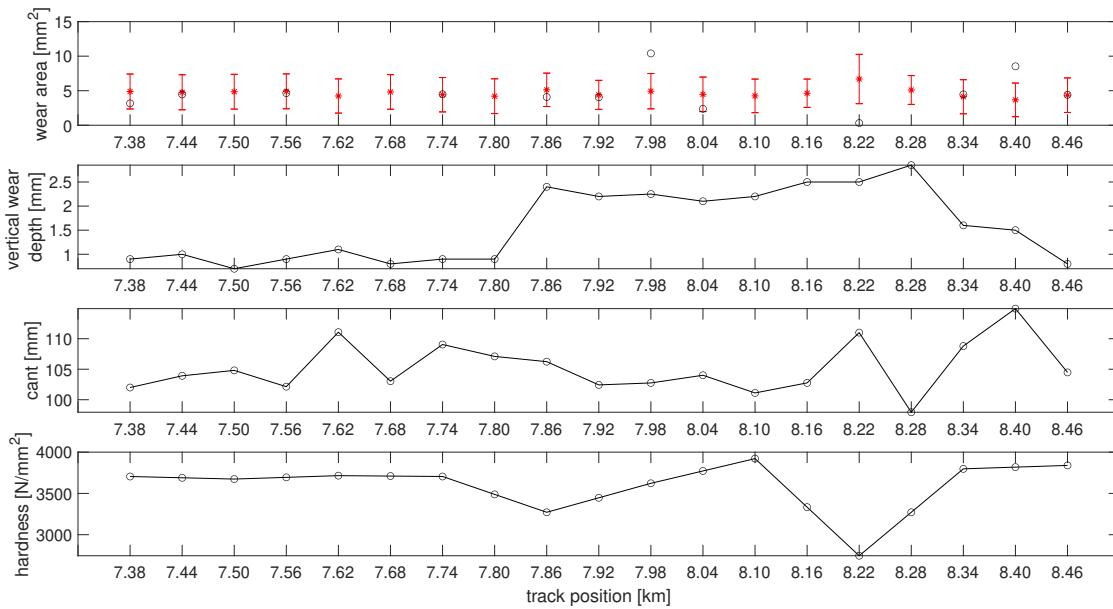
Figure 6 presents the local wear results for the tracks with curve radius $R=1500$ meters and $R=1800$ meters. The measurement numbers (or locations) are six meters apart from each other. Furthermore, the spatially dependent parameters such as the vertical wear depth, rail cant and material hardness are also plotted in this figure. The influence of the material hardness is clearly visible in this figure. Whenever the hardness is at its maximum, the wear area calculated from the MC simulation is at its minimum. This conclusion was also drawn from the sensitivity analysis results. Furthermore, the majority of the measured values are within the confidence bounds of the simulation results. For $R=1500$ m, 14 out of 15 measurements are within the confidence bounds and for $R=1800$ m, 8 out of 11 measurements. Note that for a couple of measurements performed on $R=1800$ m, incorrect measurements have been discarded from the data set.

Figure 7 shows the pdfs of the response value for case studies 1 and 2, for both the numerical and analytical approach compared with the field measurements. Despite the overlap of the measured and simulated results, the variance of the simulated results is relatively large. The reason for this is the variance of the friction coefficient as the only two parameters that were considered to be stochastic were the friction coefficient and the lateral stiffness (case study 2). And as mentioned before, the influence of the friction coefficient on rail wear is higher compared to the influence of the lateral stiffness, see Table 2. Due to the significant variation of the friction coefficient, the variance in the predicted amount of wear is large. Thus to minimise the uncertainty in response value and allow a direct comparison with the measurements, the uncertainty in the friction coefficient should be reduced. This theoretically could be done by measuring the friction coefficient for each wheel when it passes the track at a specific location, as the friction coefficient is both time and space-dependent. However, this is not feasible in practice and a friction model would have to be developed to be included in the rail wear prognostic model (see "Modify model" in Figure 2).

On the other hand, it should be realised that the measurement is just a small sample from the large distribution that the prognostic model produces. For example, a series of measurements are made in a certain period on a specific track section, see Figure 6. However, the prognostic model simulates 100 different situations with the complete (possible) range of input parameters. Therefore, it is to be expected that this results in a much wider spread because there are (most likely) many more (extreme) situations than in the measurement set. This is also apparent from Figure 7: the predictions give a wider distribution than the measurement. It is therefore not possible to fully validate the prognostic model based on just this comparison. A full validation can only be done when measurements would be available for all scenarios included



(a) $R = 1500\text{m}$ (case study 1)

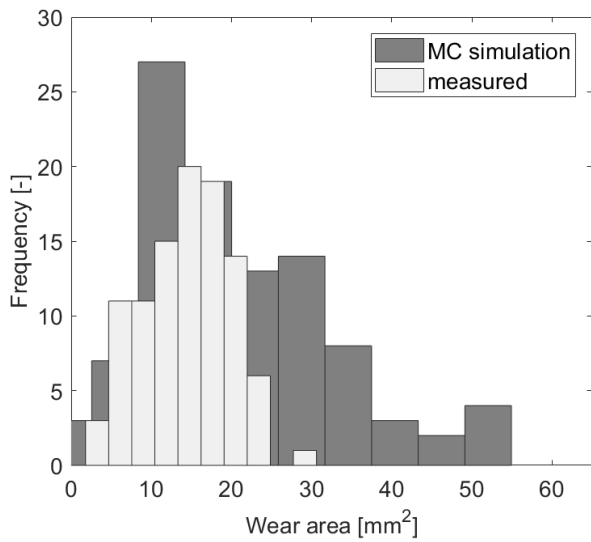


(b) $R = 1800\text{m}$ (case study 2)

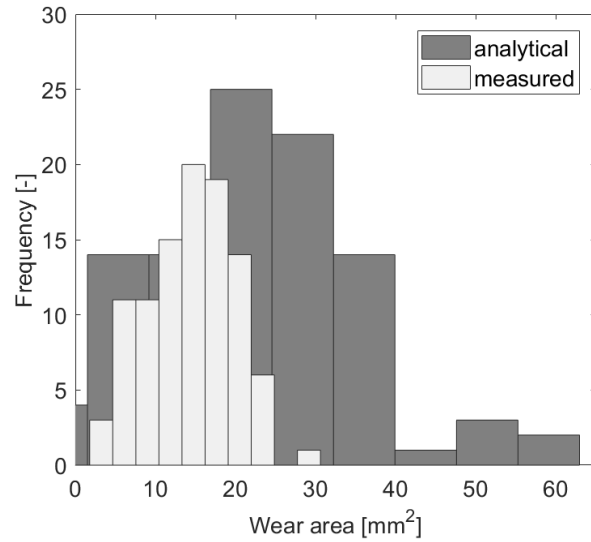
Figure 6. Local wear results for case study 1 and 2.

in the simulations. However, that will never be the case when using field data. Still, the results do demonstrate that this ‘random’ sample is fully covered by the prognostic model, which means that the model is not proven wrong. Compared to a situation in which no comparison with field data can be done, a partial validation as proposed in this work already yields a huge improvement: it provides the user trust in the

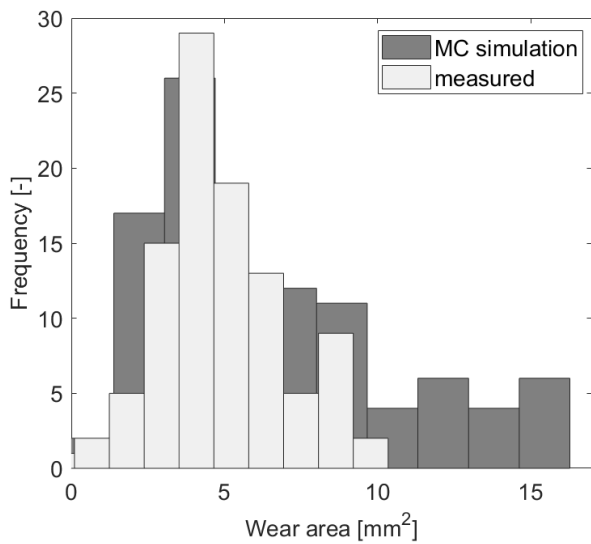
model, showing that at least the situations that are present in the (limited) field data set are correctly covered by the model. This will, in absence of a full validation, at least give a certain amount of trust in the prognostic model, which will proportionally increase if it could be done for additional series of measurements. It also means that this specific measurement set can only be reproduced by the model if the uncertainty is



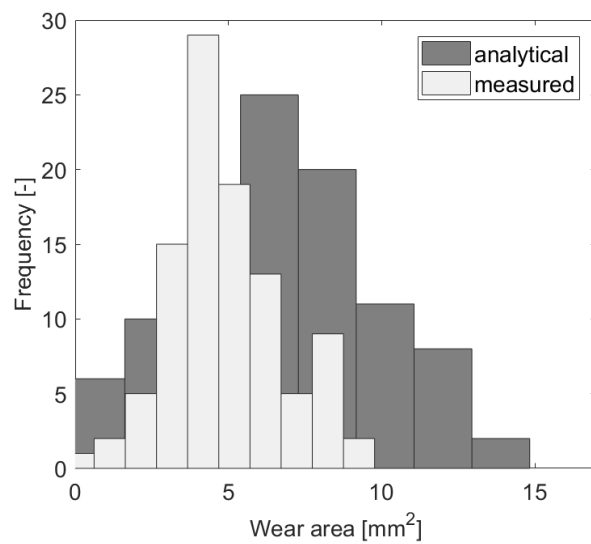
(a) Monte Carlo simulation (case study 1)



(b) Analytical results (case study 1)



(c) Monte Carlo simulation (case study 2)



(d) Analytical results (case study 2)

Figure 7. Wear area results.

reduced, as is demonstrated in case study 2, see Table 5 and Figure 7c and d. With more available input parameters (i.e. the wear depth, cant and hardness are fully known here) the predictions of case study 2 show less variance which are quite close to the variance of the measurements. Table 5 shows that the fit for case study 2 is much better (see the percentage of overlap, mean, most occurring value etc.).

Even though case study 2 shows prediction results closer to the measurements, case study 1 also shows 65% overlap and the measurements are within the distribution of the predicted results. Thus, these two measurements (samples) give confi-

dence in the proposed framework. Furthermore, the framework does not only support the validation but also provides insight in which parameter should be monitored or how the model could be modified. In this case, it became clear that the friction coefficient needs to be monitored more accurately or that the prognostic model should be modified by including a friction model to increase the model accuracy. The proposed method allows to quantify how large the effect of an unknown or uncertain variable is (e.g. the friction coefficient). Both collecting additional data and modifying the model are solutions to reduce such an effect. Although the method is not able to directly determine which of those two approaches is

Table 5. Similarity measures

Similarity measure	Field data	MC	Diff. [%]
Average	14.31	20.52	43.40
Standard deviation	5.97	12.15	103.52
Minimum value	1.79	-0.35	119.55
Maximum value	30.60	52.00	69.93
Most occurring value	14.60	11.30	22.60
Skewness	-0.15	0.79	-634.57
Kurtosis	2.25	2.63	16.79
ks2stat value	0.37		

(a) Case study 1: MC simulation

Similarity measure	Field data	Analytical	Diff. [%]
Average	14.31	20.81	45.42
Standard deviation	5.97	11.77	97.15
Minimum value	1.79	-2.33	230.17
Maximum value	30.60	59.10	93.14
Most occurring value	14.60	20.70	41.78
Skewness	-0.15	0.28	-286.56
Kurtosis	2.25	3.25	44.43
ks2stat value	0.46		

(b) Case study 1: Analytical

Similarity measure	Field data	MC	Diff. [%]
Average	5.04	6.47	28.37
Standard deviation	2.18	3.97	82.11
Minimum value	0.67	0.56	16.42
Maximum value	9.77	15.40	57.63
Most occurring value	4.08	3.87	5.15
Skewness	0.29	0.79	169.15
Kurtosis	2.95	2.63	10.60
ks2stat value	0.29		

(c) Case study 2: MC simulation

Similarity measure	Field data	Analytical	Diff. [%]
Average	5.04	6.83	35.52
Standard deviation	2.18	3.59	64.68
Minimum value	0.67	0.67	0.00
Maximum value	9.77	13.90	42.27
Most occurring value	4.08	6.34	55.39
Skewness	0.29	-0.30	200.71
Kurtosis	2.95	2.78	5.74
ks2stat value	0.50		

(d) Case study 2: Analytical

the best, it does allow to make a business case of additional monitoring or modelling. For example, how much will the infrastructure manager gain by improving the model, and does that outweigh the required effort for model improvement? By also analysing this for additional data collection, a proper decision can be taken.

Table 5 also presents the ks2stat value which is a statistic measure from the Kolmogorov-Smirnov test. As the distributions given in this study are non-parametric, the Kolmogorov-Smirnov test has been selected to compare the probability distribution functions resulting from the field measurements, the analytical method and the numerical method. The ks2stat value quantifies the distance between the cumulative distribution functions. The distributions are different if the ks2stat value is not close to zero. From Table 5, it can be concluded that the numerical simulations with the exact distribution has a lower ks2stat value compared to the analytical distribution which is as expected and also discussed from the results of the other similarity measures. The ks2stat value can be used by the user to specify if the model is validated or not. The closer the ks2stat value is to zero, the more similar the two distributions are.

3.6. Application in maintenance decision making

Figure 8 depicts the wear evolution of the track considered in case study 2 until its end of life. Through this period three different scenarios are taken into account: 1) It is assumed that the operational conditions remain unchanged, 2) a different distribution is used for the friction coefficient as measured by (Chollet, 2017) which is representative for dry contact, 3)

train traffic is doubled, hence the number of trains is doubled. These three scenarios are compared with the extrapolated wear rate based on previous measurements. From these results, it can be concluded that the wear rate prediction from historic data will overestimate the remaining useful life of the rail (≈ 80 years). The model for the regular scenario predicts a slightly shorter RUL (≈ 70 years), but in the scenarios with changing conditions, the RUL deviates by roughly a factor of two ($\approx 40 - 45$ years), from the extrapolated value. This demonstrates the added value of the model, as asset managers can now use the predicted RUL values for different track sections to schedule future maintenance tasks based on variable conditions, instead of relying on only extrapolating the measurements from the past. Note that the confidence regions for each of the simulated cases in Figure 8 are very wide. As was discussed in section 3.5, this is caused by uncertainties in the input parameters, especially the friction coefficient. Taking decisions based on these results will be difficult, as for the ‘dry condition’ case the wear threshold is expected to be exceeded somewhere between 30 and 100 years of operation. This illustrates once more that for useful predictions the actual friction coefficient must be measured or calculated. Note further that the wear process modelled here is not the only rail degradation mechanism. As was mentioned in section 2.2, the formation of cracks due to RCF is counteracted by periodically (e.g. every 6 months) grinding the rails. This is an artificial wear process with a much higher wear rate and the effect should be added to the natural wear process in Figure 8, e.g. by adding a step-wise increase of the wear area every time a grinding action takes place. This means that the time at which the wear area threshold is reached will in prac-

tice be much shorter than the 40 – 80 years in Figure 8. Still, the slopes of the lines in Figure 8 can be used to predict the moment in time when full replacement of the rails will be necessary. This is a decision that has to be taken far ahead, so a reliable prediction is required for that. The simulations do show the relative effects of changing scenarios: doubling the amount of traffic will on average lead to a 50% reduction of the track lifetime. Finally, note that also more complex (and more realistic) scenarios, where conditions change per periods of e.g. five years, can be simulated with the model.

4. CONCLUSION

This study proposed a generic framework for the validation of prognostic models in situations with limited data. A rail wear prediction case was chosen for the demonstration. The input parameter's uncertainties were propagated such that the resulting output (in this case, wear area) can be utilised in the process of maintenance decision-making. The uncertainties have been evaluated through numerical and analytical analysis, and it can be concluded that both analyses yield similar results. However, the shape of the pdf obtained with the (numerical) MC simulation corresponded better with the measured results than the analytical results. On the other hand, the analytical evaluation is ten times faster than the numerical evaluation. Hence, the user can decide which one to use for maintenance decision purposes.

Furthermore, it can be concluded that it is possible to validate a prognostic model with the proposed general framework. The case study has shown that in situations where only field data of a failure or degradation process is available, prognostic model validation is challenging. The reason is that typically not all possible combinations of input parameters are included in the data set, and also the precise values of several input parameters might be unknown. This is in contrast to situations where well-defined (laboratory) experiments can be used to validate a model. The proposed framework allows to include these uncertainties and propagate them through the model. A limited set of measurements (i.e. field results) then enables to check the validity of the developed model, where the level of validation depends on the amount and variation of measurement data.

Moreover, the rail wear case study has demonstrated that such a prognostic model can be used for maintenance decision making, provided that the variation in predicted wear area can be reduced. A more accurate prediction is possible when sufficient usage data and environmental parameters become available, allowing a more detailed planning shortly before the actual maintenance activities.

The proposed framework then provides insight in how to minimise the variance. In this case it can be recommended to use actual pdfs for the friction coefficient or a friction model that can provide friction coefficient values for various operational

and environmental conditions.

ACKNOWLEDGMENT

This research has been conducted within the Shift2Rail project In2Smart (European Union (EU) Horizon 2020 research and innovation program, grant agreement 730569). The authors would like to thank Strukton Rail for the participation in the project and acknowledge David Vermeij and Marc de Wolf from Strukton Rail and their team for the fruitful discussions on rail maintenance in practice and for providing actual track data. Furthermore, the authors would like to thank Maurice van Olderen from ProRail for providing the Quo Vadis data.

REFERENCES

- Archard, J. F. (1953). Contact and rubbing of flat surfaces. *Journal of Applied Physics*, 24(8), 981-988. doi: 10.1063/1.1721448
- Atamuradov, V., Medjaher, K., & Nouredine, Z. (2017). Prognostics and health management for maintenance practitioners-review, implementation and tools evaluation. *International Journal of Prognostics and Health Management*, 60, 1-32.
- Barchiesi, D., Kessentini, S., & Grosjes, T. (2011). Uncertainty analysis of nanoparticles for cancer photothermal therapy. In *Advances in safety, reliability and risk management: Proceedings of the european safety and reliability conference (esrel 2011)* (p. 2197–2204).
- Chen, W., Baghdasaryan, L., Buranathiti, T., & Cao, J. (2004). Model validation via uncertainty propagation and data transformations. *AIAA Journal*, 42(7), 1406-1415. doi: 10.2514/1.491
- Chen, W., Jin, R., & Sudjianto, A. (2005). Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty. *Journal of Mechanical Design*, 127(5), 875. doi: 10.1115/1.1904642
- Chollet, H. (2017). Estimation of the distribution of the wheel-rail friction coefficient in curves during the dynotrain project [Conference Proceedings]. In C. C. T. M. Maksym Spiriyagin Timothy Gordon (Ed.), *25th international symposium on dynamics of vehicles on roads and tracks*. doi: <https://doi.org/10.1201/9781351057189>
- Corbetta, M., Kulkarni, C., Banerjee, P., & Robinson, E. (2021). Uncertainty quantification framework for autonomous system tracking and health monitoring. *International Journal of Prognostics and Health Management*, 12(3). doi: 10.36001/ijphm.2021.v12i3.2936
- Dirks, B. (2015). *Simulation and measurement of wheel on rail fatigue and wear* (Unpublished doctoral disserta-

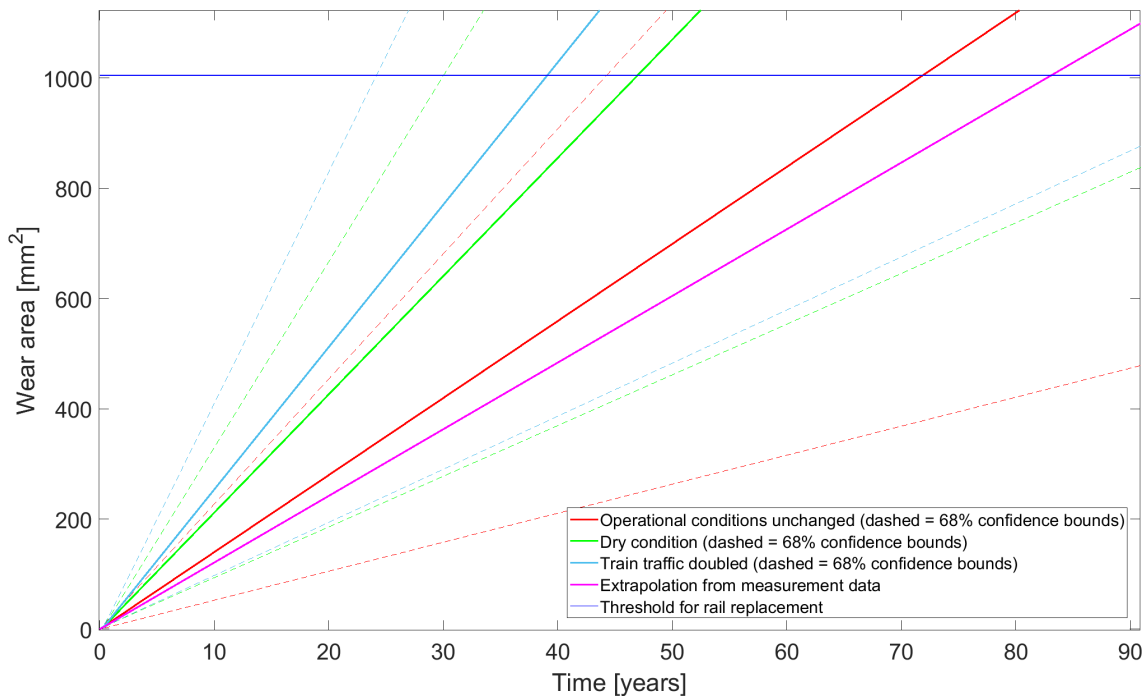


Figure 8. Rail wear evolution for case study 2 considering different scenarios.

tion). KTH Royal Institute of Technology.

Enblom, R. (2009). Deterioration mechanisms in the wheel-rail interface with focus on wear prediction: a literature review. *Vehicle System Dynamics*, 47(6), 661-700. doi: 10.1080/00423110802331559

Gao, L., & Zhang, Z. (2008). Robust optimization for managing pavement maintenance and rehabilitation. *Transportation Research Record*, 2084(1), 55-61. doi: 10.3141/2084-07

Heng, A., Zhang, S., Tan, A. C. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*, 23(3), 724-739. doi: j.ymsp.2008.06.009

Hills, R. G., & Trucano, T. G. (1999). *Statistical validation of engineering and scientific models: background* (Tech. Rep. No. SAND99-1256). Sandia National Laboratories.

Ignesti, M., Innocenti, A., Marini, L., Meli, E., & Rindi, A. (2014). Development of a model for the simultaneous analysis of wheel and rail wear in railway systems. *Multibody System Dynamics*, 31(2), 191-240. doi: 10.1007/s11044-013-9360-0

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483-1510. doi:

j.ymsp.2005.09.012

Jendel, T. (2002). Prediction of wheel profile wear-comparisons with field measurements. *Wear*, 253(1-2), 89-99. doi: Pii S0043-1648(02)00087-X Doi 10.1016/S0043-1648(02)00087-X

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3), 425-464. doi: 10.1111/1467-9868.00294

Meghoe, A., Loendersloot, R., Bosman, R., & Tinga, T. (2018). Rail wear estimation for predictive maintenance: a strategic approach. In *European conference of the prognostics and health management society 2018*.

Meghoe, A., Loendersloot, R., & Tinga, T. (2019). Rail wear and remaining life prediction using meta-models. *International Journal of Rail Transportation*, 1-26. doi: 10.1080/23248378.2019.1621780

Mukaka, M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal*, 24(3), 69-71.

Nejadseyfi, O., Geijselaers, H. J. M., & van den Boogaard, A. H. (2018). Evaluation and assessment of non-normal output during robust optimization. *Structural and Multidisciplinary Optimization*, 59(6), 2063-2076. doi: 10.1007/s00158-018-2173-2

Pannell, D. (1997). Sensitivity analysis of normative economic models: Theoretical framework and practical

- strategies. *Agricultural Economics*, 16, 139-152.
- Parkinson, A., Sorensen, C., & Pourhassan, N. (1993). A general approach for robust optimal design. *Journal of Mechanical Design*, 115(1), 74-80. doi: 10.1115/1.2919328
- Pearce, T. G., & Sherratt, N. D. (1991). Prediction of wheel profile wear. *Wear*, 144(1-2), 343-351. doi: 10.1016/0043-1648(91)90025-P
- Popovici, R. (2010). *Friction in wheel - rail contacts* (Unpublished doctoral dissertation). University of Twente.
- Popović, Z., Lazarević, L., Brajović, L., & Vilotijević, M. (2015). The importance of rail inspections in the urban area -aspect of head checking rail defects. *Procedia Engineering*, 117, 596-608. doi: 10.1016/j.proeng.2015.08.220
- Ramalho, A. (2015). Wear modelling in rail-wheel contact. *Wear*, 330, 524-532. doi: 10.1016/j.wear.2015.01.067
- Rong, K.-j., Xiao, Y.-l., Shen, M.-x., Zhao, H.-p., Wang, W.-J., & Xiong, G.-y. (2021). Influence of ambient humidity on the adhesion and damage behavior of wheel-rail interface under hot weather condition. *Wear*, 486-487, 204091. doi: <https://doi.org/10.1016/j.wear.2021.204091>
- Sankararaman, S. (2015). Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction. *Mechanical Systems and Signal Processing*, 52-53, 228-247. doi: j.ymsp.2014.05.029
- Sheinman, E. (2012). Wear of rails. a review of the american press. *Journal of Friction and Wear*, 33(4), 308-314. doi: 10.3103/S1068366612040101
- Soleimani, H., & Moavenian, M. (2017). Tribological aspects of wheel-rail contact: A review of wear mechanisms and effective factors on rolling contact fatigue. *Urban Rail Transit*, 3(4), 227-237. doi: 10.1007/s40864-017-0072-2
- Tamssaouet, F., Nguyen, K. T. P., Medjaher, K., & Orchard, M. (2021). Fresh new look for system-level prognostics. , 12(2). doi: 10.36001/ijphm.2021.v12i2.2777
- Trummer, G., Lee, Z. S., Lewis, R., & Six, K. (2021). Modelling of frictional conditions in the wheel-rail interface due to application of top-of-rail products. *Lubricants*, 9(10), 100. doi: 10.3390/lubricants9100100
- Uusitalo, L., Lehtikainen, A., Helle, I., & Myrberg, K. (2015). An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environmental Modelling & Software*, 63, 24-31. doi: 10.1016/j.envsoft.2014.09.017
- Wu, Y. T., Millwater, H. R., & Cruse, T. A. (1990). Advanced probabilistic structural analysis method for implicit performance functions. *AIAA Journal*, 28(9), 1663-1669. doi: 10.2514/3.25266
- Zhang, Y., Xiong, R., He, H., & Pecht, M. (2019). Validation and verification of a hybrid method for remaining useful life prediction of lithium-ion batteries. *Journal of Cleaner Production*, 212, 240-249. doi: <https://doi.org/10.1016/j.jclepro.2018.12.041>
- Zobory, I. (1997). Prediction of wheel/rail profile wear. *Vehicle System Dynamics*, 28(2-3), 221-259. doi: 10.1080/00423119708969355