

## Using data analytics to understand why certain rail sections at the Dutch high-speed line are affected by RCF

Schalk, Ricks; Zoeteman, Arjen; Nunez, Alfredo; Wolfert, Rogier

**DOI**

[10.25084/raileng.2017.0096](https://doi.org/10.25084/raileng.2017.0096)

**Publication date**

2017

**Document Version**

Accepted author manuscript

**Published in**

Proceedings of 14th International Railway Engineering-2017conference in Edinburgh, UK: 2017

**Citation (APA)**

Schalk, R., Zoeteman, A., Nunez, A., & Wolfert, R. (2017). Using data analytics to understand why certain rail sections at the Dutch high-speed line are affected by RCF. In *Proceedings of 14th International Railway Engineering-2017conference in Edinburgh, UK: 2017* <https://doi.org/10.25084/raileng.2017.0096>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

## USING DATA ANALYTICS TO UNDERSTAND WHY CERTAIN RAIL SECTIONS AT THE DUTCH HIGH-SPEED TRACK ARE AFFECTED BY RCF

R. Schalk<sup>1,2</sup>, A. Zoeteman<sup>1</sup>, A. Nuñez<sup>3</sup> and R. Wolfert<sup>1</sup>

<sup>1</sup> Delft University of Technology, Civil Engineering and Geosciences, Section of Integral Design and Maintenance, Stevinweg 1, 2628CN Delft, The Netherlands

<sup>2</sup> Mott MacDonald B.V., Amsterdamseweg 15, 6814CM Arnhem, The Netherlands

<sup>3</sup> Delft University of Technology, Civil Engineering and Geosciences, Section of Railway Engineering, Stevinweg 1, 2628CN Delft, The Netherlands

**KEYWORDS:** High-Speed Rail, Rail Maintenance, Rolling Stock, Data Analytics, Rolling Contact Fatigue

### ABSTRACT

This paper describes the use of big data for analysing Rolling Contact Fatigue (RCF) phenomena at the High Speed Line (HSL Zuid) in The Netherlands. The authors developed a data model to investigate the impacting parameters in train-track interaction. This has been done to gain more insight in the circumstances where RCF occurs and to conclude why some track sections are severely affected and others not.

To evaluate the worst affected areas by RCF, the methodology included a bottom-up approach which focuses at the worst affected areas by RCF, developing a set of characteristic parameter values regarding different types of hotspots. The methodology has been applied for the Dutch High-speed line, where certain sections had been heavily affected by RCF. Findings concluded that slow running traffic through curves on a high-speed line is likely to contribute to the appearance of RCF.

### INTRODUCTION

Data analytics are a common way to study relationships between one or more data parameters. They can be used to gain insight in the connections between these parameters in order to support decision making for rail infra managers. The aim of this study was to use data analytics to study the cause of Rolling Contact Fatigue (RCF) at the Dutch High-Speed track (HSL-Zuid).

Rolling Contact Fatigue is an issue affecting the integrity of the rails which is the result of the stress cycle between wheels rolling over the rails (Dollevoet, 2010). This stress cycle eventually leads to material fatigue in the rails which can result into various defects. Among these defects, squats, head checks and various forms of corrugations are examples. These defects affect the integrity of the rails, hereby also affecting the track availability and safety. Therefore its valuable to know the causes for RCF at certain track sections to be able to address the cause of RCF at these sections and improve the overall track performance.

The HSL-Zuid has been used as a case for this study to test two approaches which have been developed by the authors to study the causes of RCF. The approaches were named 'bottom-up-' and 'top-down approach', which can be used separately or together to study the cause of RCF. Among these approaches the bottom-up approach and the results from its application at HSL-Zuid will be presented in this paper.

### HSL-ZUID

The HSL-Zuid is the only high-speed track in The Netherlands. The construction of the HSL-Zuid was intended to connect the Netherlands with the European high-speed rail network, establishing a high-speed connection between Schiphol Amsterdam Airport, Rotterdam, Breda and Belgium. Construction started in 2001 and was finished in 2006, the track was opened for commercial traffic in 2009.

The HSL-Zuid consists of two main sections; the North Track, running from Hoofddorp to Rotterdam and the South Track, running from Barendrecht to the Belgian border with turnouts halfway from- and to Breda. At Hoofddorp, Rotterdam and Breda the HSL-Zuid connects to the Dutch conventional track. Both sections consist of about 45km double track.



Fig. 1 Track overview of the HSL-Zuid

The HSL-Zuid has been designed for speeds of up to 300km/h, the curves at the maximum speed sections have been designed for a speed range of 220-300km/h. The track is constructed in Rheda2000 slab track with transition zones to Ballast 160 and Ballast 300 (Belgian Border). Among these transition zones also voltage locks are installed whereas among the HSL-Zuid 25kV ac is the electrification, the regular Dutch tracks are powered by 1500kV dc. The 60E1 rail profile has been installed with 60E2 among the high rails throughout the curves, which has been chosen as an anti-headcheck profile. These profiles have either a regular 260 rail grade or a 350HT rail grade (curves). Several special assets have been constructed among the HSL-Zuid like, Tunnels, Viaducts, flyovers and a bridge.

The HSL-Zuid has been designed for high-speed traffic. Two types of rolling stock were intended to use the tracks commercially. One type with a maximum speed of 300km/h and another type with a maximum speed of 250km/h. However, the second type was initially delayed and temporary replaced by conventional rolling stock with a maximum speed of 160km/h. But, when the second type was finally scheduled after a long period of delays in December 2012, it was again cancelled in January 2013 due to technical difficulties. This caused the temporary conventional trains to be used permanently. The 300km/h train has two locs, one in the front and one in the back, each having four powered axles. The conventional type was up to 2015 using a single pulling loc and since 2015 used in a so-called sandwich configuration. The conventional train is being scheduled more frequently than the high-speed train 33x against 14 times a day (each direction). However, both trains are not using the same track sections, the conventional train is scheduled between Breda and Amsterdam Schiphol Airport, whereas the high-speed train is scheduled between the Belgian Border and Amsterdam Schiphol Airport.

**ROLLING CONTACT FATIGUE AT THE HSL-ZUID**

In November 2014, during a visual inspection along the tracks of the HSL-Zuid, unexpected severe damages had been found. These damages were unexpected as regular measuring campaigns didn't indicate these were to be found. The regular measurements executed were eddy current gauge corner measurements and ultrasonic measurements. Following these findings the whole track was inspected. This procedure resulted in the finding of five other major affected areas, so-called hotspots. Fig. 2 (a) and Fig. 2 (b) show the damage among one of these hotspots which were typical for the findings among the HSL-Zuid among the transition zone ballast sections.

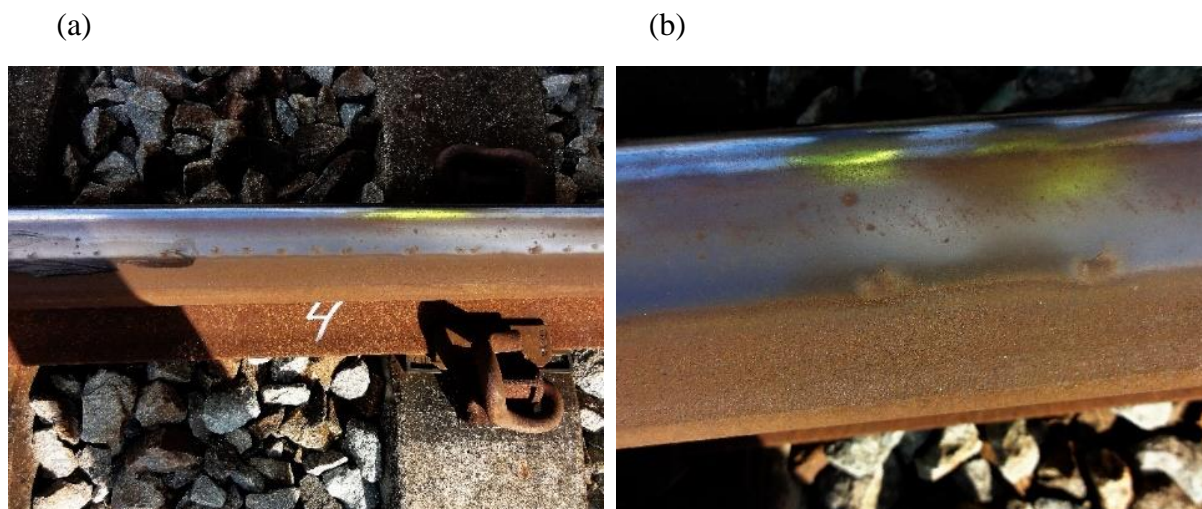


Fig. 2 (a) showing lots of damage at one of the Hotspots at Hoofddorp, a ballast track transition zone. (b) close up at the same hotspot, damage is located at the rail head.

Characteristic to the RCF found at the HSL-Zuid is their location on top of the rail head. Therefore a new measurement method was introduced; the Sperry eddy current walking stick. Using this new tool the whole track had been measured between July and October 2015. Using the walking stick the cracks were measured. An examined piece of rail with the damage is shown in Fig. 3 (a). Fig. 3 (b) shows the mismatch in the measuring campaign for the smallest cracks by eddy current. Eddy current measures cracks up to 5.00 mm of depth.

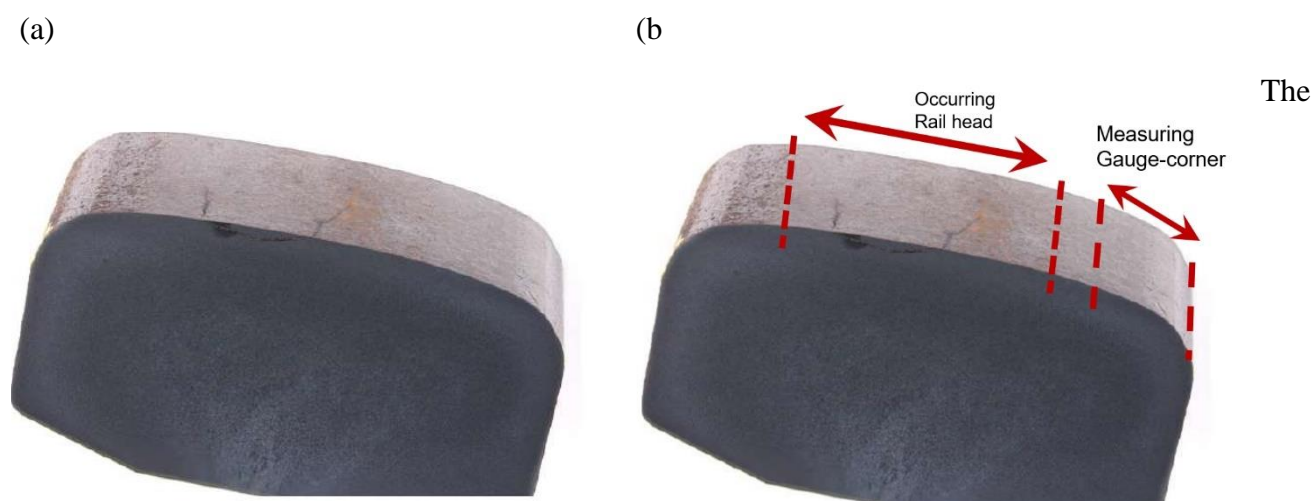


Fig. 3 (a) examined rail with damage on top of rail. (b) Same rail showing the mismatch in the measuring campaign.

damages found occurred in hotspots ranging in lengths from 700m up to 5km. Material investigation showed the rails were according specifications and free of irregularities. The damages show most

similarities with ‘studs’ or ‘spalling defects’, they don’t meet the ‘lung-shape’ and the depression among the surface of squats. They show similarities with the studs described in (S.L. Grassie, 2012), (S.L. Grassie, 2015) and (Stuart L Grassie, Fletcher, Hernandez, & Summers, 2011) in particular: the occurrence on top of the rail and growing across the rail towards the field side. Also striking is the occurrence of the damages only in the open areas and occurrence in only heat treated rail sections among the HSL-Zuid. Also, the RCF at the HSL-Zuid occurred only after 30MGT.

**METHODOLOGY**

Parameters

In order to find the causes for the RCF at the HSL-Zuid an approach called the ‘bottom up’ (B-U) approach has been developed. Using: eddy current measurements, design and alignment data, maintenance data and rolling stock data. In this approach the rail condition, whether its affected by RCF or not- and by what amount was approached by the interaction of track geometry, rolling stock and maintenance parameters as shown in Fig. 4.

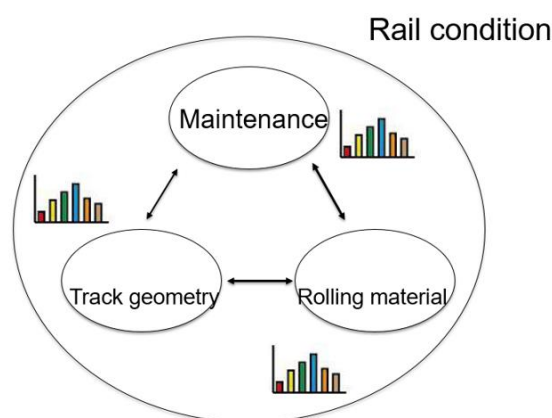


Fig. 4 visual representation of the rail condition approached as the interaction between groups of parameters.

Using this approach the following parameters with their respective variables have been used, which are shown in Table 1.

Table 1 Overview of the parameters used and their variables

Category	Track	Rolling stock	Maintenance	Damage
Parameter	Superstructure (type)	Speed (km/h)	Grinding type (type)	Damage depth (mm)
	Rail grade (type)	Traction (% of max)	Grinding depth (mm)	
	Rail profile (type)	Tonnage cumulative (MGT)	Grinding date (d/m/y)	
	Assets (type)	Tonnage per vehicle (MGT)		
	Design speed (km/h)	Cant deficiency (mm)		
	Curve radius (m)			
	Curve cant (mm)			
	Height difference (m)			

The parameters have been chosen according the possible influence they can have on the appearance of RCF. No direct data from the trains regarding forces on the track nor wheel maintenance was available for this study. Regarding the rolling stock for both types of train the average tractive effort per axle and the average speed profiles were available.

### Intensity

To study the condition of the rail, a new parameter was introduced to represent both the depth of the defects and the amount of defects in a track partition. This has been done using eddy current measurements. The introduced parameter has been named ‘intensity’ and functions as a KPI for rail condition. Regarding railway operations KPI’s govern the way maintenance operations are governed KPI’s in railways have been reported in (Åhrén & Parida, 2009), (Parida & Chattopadhyay, 2007), (Stenström, Parida, Lundberg, & Kumar, 2015) and (Stenström, Norrbin, Parida, & Kumar, 2016).

Connecting measurements with a KPI, using ABA, defining robust and predictive KPI’s which consider stochasticity of the defects and predicting over maintenance time horizons has been reported in (A. Jamshidi et al., 2016), (Ali Jamshidi, Nunez, & Li, 2015) and (Ali Jamshidi, Núñez, Li, & Dollevoet, 2015). For this study only one round of eddy current measurements was available. The intensity has been calculated for each leg separately. Threshold values have been introduced for each mm of depth. Cracks smaller than 0,1 have been filtered out due to the accuracy of the measurements. For the calculation of intensity at a track partition the following formula has been introduced:

$$I_X(t) = \sum_{c=1}^5 \lambda_c * n_{c,X}(t)$$

In which:

$I_X$  = Intensity at rail partition  $X$

$X$  = Interval position; km  $x$  to  $x+500$ m

$t$  = Time of measurement

$c$  = Category

$\lambda_c$  = Category coefficient

$n_{c,X}(t)$  = Number of defects in category  $c$  at partition  $X$  at time  $t$

### Partitioning

With the introduction of the intensity parameter the track can be modelled dividing it in a number of partitions. Each of these partitions will be characterized by its respective parameter values. In general, the smaller the partitions the more accurate the results are to be expected. The partitioning is proposed to be done for each leg separately in order to evaluate the respective values throughout curves. The partitioning process has been visualized in Fig. 5.

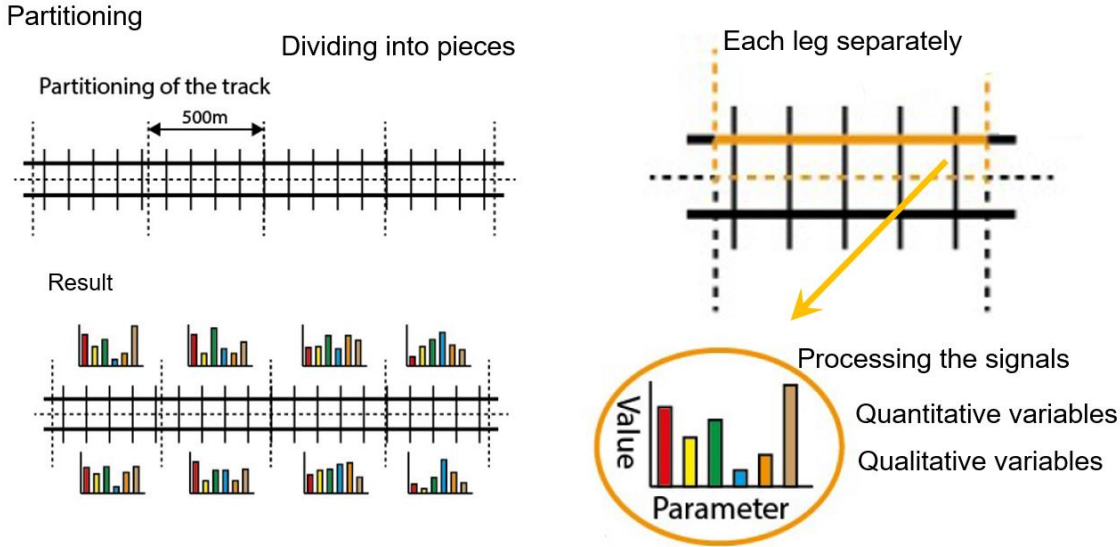


Fig. 5 Visualization of the partitioning process

Processing parameters

Using the modelled track divided in partitions and with the introduction of the intensity parameter the identification of the hotspots is the next proposed step. The aim of this approach is to find influencing parameters regarding the rail condition (intensity). The hotspots can be identified as the partitions with the highest intensity values, setting a threshold value. It's recommended to also have visual evidence of an inspection to support the measurements.

As both quantitative and qualitative variables are being processed these signals should be treated differently. For the qualitative variables a 'mixed' variable has been introduced for when a transition point is present among the partition. An example regarding rail grade is

mathematically formulated as:

$$\delta_x^{rail}(k) = \left\{ \begin{array}{l} 260 \text{ if } \delta^{rail}(x, k) = 260 \text{ for all } x \in X \\ 350HT \text{ if } \delta^{rail}(x, k) = 350HT \text{ for all } x \in X \\ mix \text{ if } \delta^{rail}(x_1, k) = 260, \delta^{rail}(x_2, k) = 350HT \end{array} \right\}$$

For:  $x_1 \neq x_2 \in X$

In which:

$\delta^{rail}(k)$  : value of the parameter ( $\delta$ ) rail at moment of measurement  $k$

$X$ : partition

$k$ : moment of the measurement

$x$ : location

For the quantitative variables (speed, radius, cant, etc.) the average value of the different signals within the 500m partition has been used. This is formulated with the example of the speed of the TRAXX as:

$$\delta_X^{VTRAXX}(k) = \frac{1}{N_X^{VTRAXX}(k)} \sum_{x \in X} \delta^{VTRAXX}(x, k)$$

For:  $\delta^{VTRAXX}(x, k) \neq null$

In which:

- $\delta^{VTRAXX}(k)$ : Value of the parameter ( $\delta$ ) speed TRAXX at moment of measurement  $k$
- $x$ : location
- $k$ : moment of measurement
- $X$ : Partition
- $N_X(k)$ : number of signals within partition  $X$  at moment of measurement  $k$

Similarity

The next step of the process is to find similarities among the hotspots. These are parameters with the same values for the nominal parameters and closely lying values for the numerical ones. So here the parameter values among the different hotspots are compared to each other. This in order to be able to pinpoint the parameters which should be investigated more closely. The other argument would also be able to exclude a number of parameters as the cause for the damages at the hotspots.

The similarity function to describe the similarity of one parameter at two hotspots is:

$$V(\delta_{X_{h1}}(k), \delta_{X_{h2}}(k)) = \|\delta_{X_{h1}}(k) - \delta_{X_{h2}}(k)\|^2$$

In which:

- $V$ : similarity function
- $\delta$ : parameter value
- $X_h$ : partition of hotspot
- $k$ : moment of measurement

he condition for similarity will be described according a similarity threshold  $\epsilon_\delta$ .

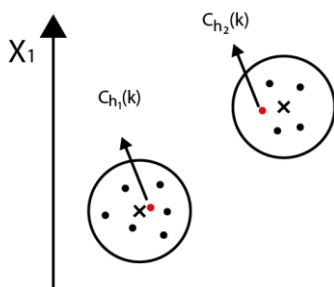
If:  $V(\delta_{X_{h1}}(k), \delta_{X_{h2}}(k)) \leq \epsilon_\delta$  we will say:  $\delta_{X_{h1}}(k) \approx \delta_{X_{h2}}(k)$  thus similar.

Clustering

In of the possibilities is that not all the hotspots among a track are affected by the same damage mechanism causing the RCF. In order to distinguish these clustering is introduced to group characteristic parameter values for each type of hotspot. Clustering is a measure of classification, more specifically ‘unsupervised classification’ which aims at discovering groups in data (Govaert, 2009).

Regarding the clusters its required to be homogenous and well separated (Hansen & Jaumard, 1997). The sample for the clustering will be the set of hotspots which have been found earlier using the identification of the hotspots.

The clusters will consist of sets of characteristic similar parameters for a certain hotspot type according the formulas and **Error! Reference source not found.:**



$$\overline{C_{h_1}}(k) = \begin{bmatrix} \delta_{h_1}^1 \\ \cdot \\ \cdot \\ \cdot \\ \delta_{h_1}^m \end{bmatrix}$$

Fig. 6 How the clustering works according two hotspot types.



$$\overline{C}_{h_2}(k) = \begin{bmatrix} \delta_{h_2}^1 \\ \cdot \\ \cdot \\ \cdot \\ \delta_{h_2}^m \end{bmatrix}$$

$$C_i(k) \in C_{h_1} \text{ if } V(\overline{C}_{h_1}(k), C_i(k)) \leq \varepsilon_c$$

When there are five hotspots evaluated regarded clustering, the output can for instance be that two hotspots types are found which divide the five hotspots, which can be described mathematically as:

$$C_{h_1} = \{C_2(k), C_3(k)\}$$

$$C_{h_2} = \{C_1(k), C_4(k), C_5(k)\}$$

Where  $C_{h_1}(k)$  is the selection of characteristic parameter values which are similar for a certain hotspot type  $h_1$  at moment of measurement ( $k$ ). The hotspot types are thus described as a vector which makes existing of a set of parameter values.

Hypothesis

The bottom-up approach results in a set of characteristic parameter values for each hotspot type. These values should when able, be linked to the RCF among these hotspots setting a hypothesis for each type of hotspot regarding the most likely cause of RCF.

However, this set of characteristic parameter values for each hotspot can also be used checking the hypothesis, testing it against the rest of the track to identify other areas which share the same characteristic parameter values. A flow chart is introduced to help checking the set of characteristic parameter values, as shown in Fig. 7.

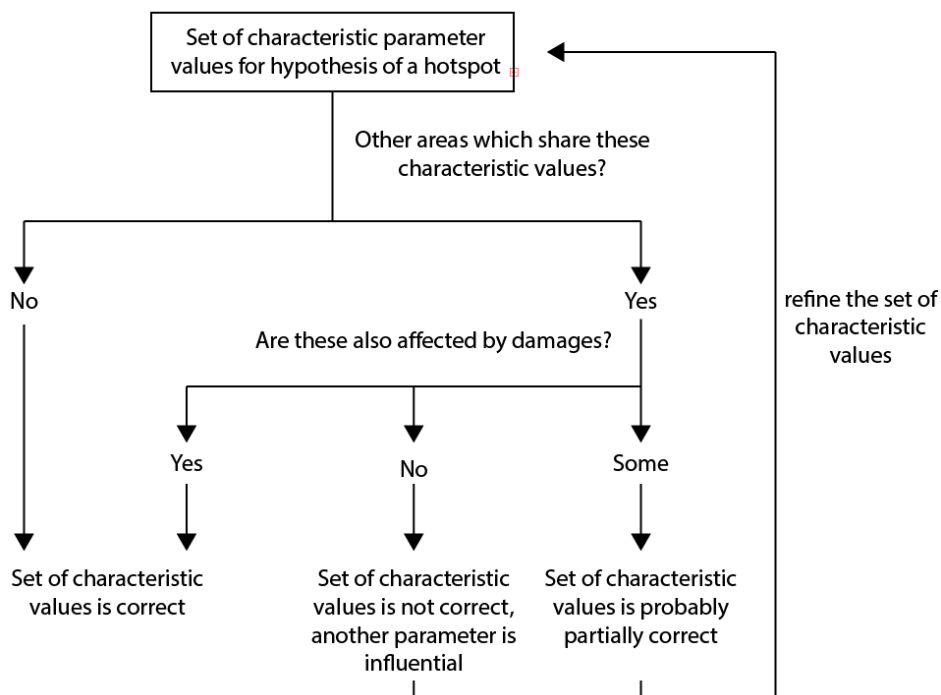


Fig. 7 Flow chart regarding the checking of the hypothesis for the hotspots

## RESULTS

The application of the B-U approach resulted in some unexpected findings for the HSL-Zuid. Partitioning values of 500m have been used, resulting in over 700 partitions. Intensity has been introduced with two additional thresholds to identify hotspots, filtering cracks larger than 1.00mm and 3.00. Results from this application are shown in Fig. 8.

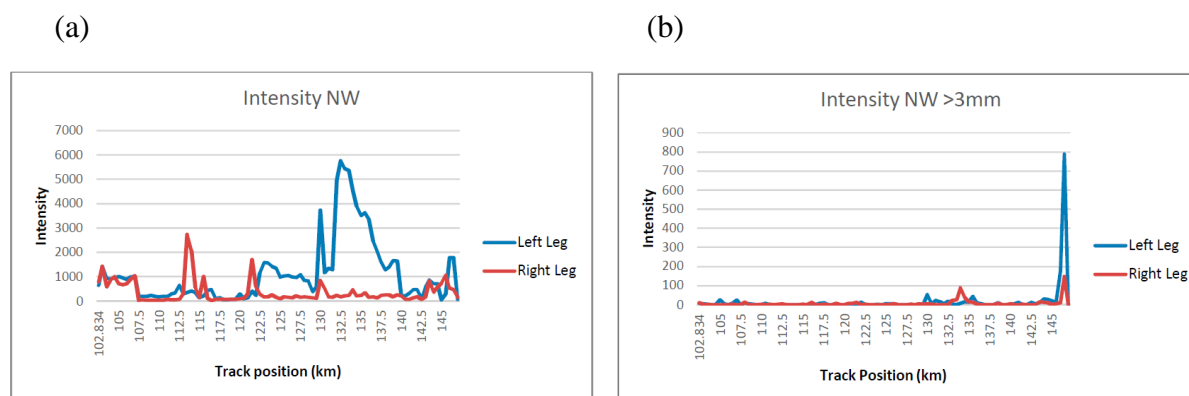


Fig. 8 (a) showing the intensity of the North-West track for all cracks (b) showing the intensity for cracks larger than 3.00mm, showing one single peak which is the hotspot near Hoofddorp.

Similarities among all the hotspots were that they occurred only among sections where the 350HT rail grade was installed – thus among curves. These curves had cant of at least 75mm. Dominant load by vehicles came from the conventional train, one of the hotspots occurs on a section where only the conventional train is scheduled, near Breda. There are no hotspots at the tracks where only the high-speed train is scheduled. Also, there are no hotspots among the tunnel sections.

Clustering resulted in the identification of two types of hotspots. Which can be best differentiated among their respective locations among the ‘open track’, where the design speed is at maximum and among the ‘entry zones’ where the trains enter the HSL-Zuid. The characteristic parameter values of these two hotspot types are presented in Table 2.

Table 2 Overview of the characteristic parameter values for both types of hotspots

Open track hotspots		Entry zone Hotspots	
Parameter	Value	Parameter	Value
Superstructure	Rheda 2000	Asset	Voltage lock
Rail grade	350HT	Superstructure	Ballast 160
Rail profile	60E2 upper leg	Rail grade	350HT
Design speed	300	Rail profile	60E2 upper leg
Speed high-speed train	300	Vertical curves	yes
Speed conventional train	160	Design speed	160 km/h
Cant excess conventional train	>50	Cant Excess conventional train	Theoretical canting ~0
Dominant load	Conventional train	Dominant load	Conventional train
Traction conventional train	Yes	Traction conventional train	No
Traction high-speed train	Yes	Traction high-speed train	No

The dynamic effects through curves are much different from the high-speed train among both areas. At both areas the conventional train drives below design speed, resulting in undesired effects through the curves. For instance, having different steering moments through the curves and other vertical loading on the rails. At the open track hotspots this results in large cant excess of up to 100mm. Whereas, at the entry zone hotspots this results in zero cant excess/deficiency. The latter, causing theoretically, the train having no leading leg through the curve, thus unpredictable dynamic effects. It is expected to see the differentiations between both hotspot types in which leg throughout is affected mostly in the open track hotspots. Among the open track hotspots, at all, the lower leg was affected by RCF and in 2 out of 4 only the lower leg was affected.

Regarding checking of the hypothesis, among the entry zones there were only the two initially identified hotspots. The other entry zones didn't meet the same characteristic parameter values. One of the entry zones was at the Belgian border, where no conventional trains are scheduled and it lies in the maximum speed area. The two other entry zones are located in tunnels, thus not meeting the criteria (and show no high intensity values).

Whereas for the open track characteristic parameter values. An additional thirteen curves were located which met the characteristic cant excess value of at least 50mm for the conventional train. Two of these curves had a 260 rail grade installed, these showed no RCF. One of these curves was located along a voltage lock where both trains don't have any tractive efforts these rails were also unaffected. For three curves, only the conventional train is giving tractive efforts, these showed very small intensities. At two curves only the high-speed train gives tractive efforts these showed very small intensities. Four out of the remaining five showed larger damage concentrations, here all the characteristic parameter values were met.

## CONCLUSIONS

The application of data analytics showed promising results regarding its application at the HSL-Zuid. The proposed KPI intensity could locate all the hotspots of RCF among the HSL-Zuid. Also, two types of hotspots have been identified using the methodology consisting of very different parameter values. For the HSL-South further study should focus on the rail grades and the wheel-rail behaviour of both types of rolling stock through curves. Also, the introduction of more maintenance related-parameters can open new perspectives upon the appearance of RCF.

Proposed further application of data analytics with regard to the other parameters are numerical modelling and statistically identifying the set of characteristic parameter values of each hotspot type. Another consideration is the introduction of homogenous partitioning instead of fixed partition, this will reduce the number of samples but will remove the 'mixed values' which can provide more accurate results. Also, introducing more eddy current measurements is expected to give better results as growth rates among different partitions can be monitored and linked to possible causes of RCF and eventually improve the performance of the rails.

## ACKNOWLEDGEMENT

This research was supported by Infraspied Maintenance B.V.

## REFERENCES

- Åhrén, T., & Parida, A. (2009). Maintenance performance indicators (MPIs) for benchmarking the railway infrastructure: a case study. *Benchmarking: An International Journal*, 16(2), 247-258.

- Dollevoet, R. P. B. J. (2010). *Design of an Anti Head Check profile based on stress relief*. University of Twente.
- Govaert, G. (2009). *Data analysis*. London: ISTE.
- Grassie, S. L. (2012). Squats and squat-type defects in rails: the understanding to date. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 226(3), 235-242.
- Grassie, S. L. (2015). *Studs and Squats: the evolving story*. Paper presented at the 10th International Conference on Contact Mechanics, Colorado Springs, Colorado, USA.
- Grassie, S. L., Fletcher, D. I., Hernandez, E. G., & Summers, P. (2011). Studs: a squat-type defect in rails. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 0954409711421462.
- Hansen, P., & Jaumard, B. (1997). Cluster analysis and mathematical programming. *Mathematical programming*, 79(1-3), 191-215.
- Jamshidi, A., Faghih-Roohi, S., Núñez, A., Babuska, R., de Schutter, B., Dollevoet, R., & Li, Z. (2016). A big data analysis approach for rail failure risk assessment. *Under review*.
- Jamshidi, A., Nunez, A., & Li, Z. (2015). *Maintenance decision indicators for treating squats in railway infrastructures*. Paper presented at the 94th Annual Meeting Transportation Research Board, Washington, USA, 11-15 January 2015; Authors version.
- Jamshidi, A., Núñez, A., Li, Z., & Dollevoet, R. (2015). *Fuzzy Maintenance Decision Support for Treating Squats in Railway Infrastructures*. Paper presented at the 2015 Joint Rail Conference.
- Parida, A., & Chattopadhyay, G. (2007). Development of a multi-criteria hierarchical framework for maintenance performance measurement (MPM). *Journal of Quality in Maintenance Engineering*, 13(3), 241-258.
- Stenström, C., Norrbin, P., Parida, A., & Kumar, U. (2016). Preventive and corrective maintenance–cost comparison and cost–benefit analysis. *Structure and Infrastructure Engineering*, 12(5), 603-617.
- Stenström, C., Parida, A., Lundberg, J., & Kumar, U. (2015). Development of an integrity index for benchmarking and monitoring rail infrastructure: application of composite indicators. *International Journal of Rail Transportation*, 3(2), 61-80.