White Paper 2025: 01

Validation of wheel flat detection and characterization using wayside data

Vanessa Meulenberg, Aron Laurell Håkansson, Jesper Westerberg, Wolfgang Birk (Predge);
Mirko Leiste (VTG Rail Europe GmbH)

AUTHORED BY PREDGE AB

IHHA 2025
THE BROADMOOR, COLORADO SPRINGS, CO, USA
NOVEMBER 17-21, 2025

VALIDATION OF WHEEL FLAT DETECTION AND CHARACTERISATION USING WAYSIDE DATA

V. Meulenberg[1], A. Laurell Håkansson[1], J. Westerberg[1] & W. Birk[1,2]
[1]Predge AB, Västra Varvsgatan 11, 972 36 Luleå, Sweden
[2]Automatic Control, Luleå University of Technology, 971 87 Luleå, Sweden

M. Leiste
VTG Rail Europe GmbH, Nagelsweg 34, 20097 Hamburg, Germany

ABSTRACT: This paper discusses the validation of an analytics solution that uses measured wheel impact load data from multiple wayside train monitoring systems detect and estimate the length of wheel flats. While the solution was implemented as a decision support providing actionable insights to different roles in the maintenance organization, the validation needed to consider real-life inspection context in both workshops and field. To validate a larger case study was setup where more than 5000 wheels were inspected both in the field and in workshops by maintenance personnel using standard inspection guidelines and reporting principles. The inspection information was then combined with the actionable insights on the inspected wagons and the performance of the flat detection and flat length estimation was assessed. The results indicate that wheel flat detection and wheel flat length estimation can be accomplished using wheel impact load data condition-based maintenance and predictive maintenance.

1 INTRODUCTION

Wheel flats or out-of-roundness are a common type of wheel damage that has the potential of causing severe damage to the rail and the wagon. To manage these risks, infrastructure managers, train operating companies and wagon owners have agreed on a maximum permissible wheel flat length. There are both regulations and practices in place which should both protect the infrastructure and the rolling stock, like e.g. the General Contract of Use (GCU) (2025) and the technical regulations in place in Sweden, Swedish Transport Administration (2014). According to the GCU, wheel flats of 60 mm and larger in length are not admissible for operation. There are also other properties of a flat that affects the impact forces, such as shape and depth, but those are not part of the contractual requirements.

It is well known that wheel flats induce large wheel-rail contact forces measured by wheel impact load detectors (WILD) (Dong et.al. (1994), Kalay et.al. (1995), Johansson et.al. (2003)) and do increase with the length of the wheel flat. Thus, WILD detectors can be used to identify wheels with damages and if the measured forces exceed predefined thresholds, wagons can be stopped from operation. The open questions are therefore which forces are associated with a certain wheel flat length and how such an analytics solution can be validated in real life.

To answer the first part of the question a method was proposed by Birk, et. al. (2025) to identify wheel flats as a damage category and subsequently determine the length of the flat, which was only validated on a limited set of data.

The main goal of this investigation is therefore to validate the solution in a larger case study where many wheels are inspected both in the field and in workshops by maintenance personnel. Standard inspection guidelines and reporting principles should also be used to acquire the inspection data a realistic manner but supported by hand-held devices to capture the information in digitalized and structured form. The inspection information can then be combined with the outcome of the analytics approach for the inspected wagons and the performance of the flat detection and flat length estimation can be quantified.

Previous research has mainly been focused on the method development and the field trials are often small in scale and rarely reflect the operational aspects of train operations and the monitoring of it.

Nevertheless, there are important studies that have inspired this work and support the underlying performance assessment of the analytics.
Bogdevicious et.al. (2016) and Johansson et.al. (2003) show that impact loads from wheel flats are a function of train speed which is not linearly increasing where also field tests are conducted. The tests conducted in these studies have also shown that a simple mapping is not feasible and the investigation by Inwicki et.al. (2023) and Ye, et.al. (2020), which has inspired the method development that is subject of the validation. While research in this topic has been ongoing for more than 30 years, there are still very recent approaches for the detection of wheel flats like e.g. Cui et.al. (2025) and Zhou et.al. (2024). It is the believe of the authors that there is a lack of formal validation and trials reinforcing hesitation to trust theoretical methods. It also means that the implementation and real-life use still pose an obstacle, even though platforms enabling swift and easy real-life deployment for test, validation and operation have already been proposed and realized in 2015 as stated by Karim et.al. (2015).
The main contribution of this paper is therefore to show how a large-scale test can be designed and conducted without larger disruptions to normal operations for the involved stakeholders. Further, the acquired data sets are also beneficial for the development, benchmarking, and validation of other methods.
The paper is now organized as follows. First the problem statement is given and followed by a summary of the methods that are target of the validation in Section 3. Section 4 then describes the validation approach, and the results are summarised in section 5. The paper is then concluded with conclusions and recommendations.

2 THE PROBLEM DEFINITION

Nowadays, railway undertakings and even wagon keepers—as entities responsible for maintenance—have little to no visibility into the condition of rolling stock outside of scheduled workshop visits. The extent of the damage and its evolution remain unknown. As a result, wheelset faults often occur unexpectedly, at inconvenient times, in unsuitable locations, and under challenging conditions. Typically, this happens during the manual predeparture check just before train departure, during transit, on loaded wagons, or far from workshops or areas accessible to mobile service teams.
These incidents lead to significant operational disruptions for all stakeholders—wagon keepers, railway undertakings, infrastructure managers, as well as the recipients and senders of freight. Bringing the wagon back into service often requires high manual effort and coordination. Naturally, this has a major impact on wagon availability and negatively affects customer satisfaction and the overall perception of rail freight transport. Despite this, the status quo is accepted by all parties involved.
From the perspective of the freight wagon, the wheelset is the most maintenance-intensive component according to Hecht, et.al. (2018). To mitigate the impact of spontaneous failures, many wheelsets must be stored at various locations across Europe to ensure rapid response times.
Over the past 15 years, the frequency of wheel defects has increased significantly. At the same time, rising material and logistics costs are putting additional pressure on maintenance budgets. Given these developments, it is important to gain transparency into the condition of wheelsets and implement data-driven maintenance strategies to:

- reduce unplanned downtime,
- optimize spare parts inventory,
- extend the service life of the wheelset and
- automize the predeparture check.

Only through greater visibility and predictive maintenance can the rail freight sector improve reliability, reduce costs, and enhance customer confidence in the long term. With a track-side system such as WILD, it is possible to detect damages of the wheel surface. Rather than impact forces, it is important to determine the length of the damage, as this is the current threshold for

- taking the wagon out of service during the predeparture check (threshold: greater than 60 mm according to the GCU (2025) and
- the initial inspection in a workshop for reprofiling the wheel (threshold: greater than 30 mm according to the widely accepted European Maintenance Guide, VPI-EMG by Perrey (2025).

The estimation of wheel flat length offers an approach for deriving quantifiable insights that enhance the transparency of wagon condition monitoring. This supports the resolution of the previously identified challenges.
It is expected that the implementation of data-driven diagnostics will, by itself, result in a measurable reduction in the frequency of wheel reprofiling's. This is primarily due to the decreased reliance on subjective and often undocumented assessments. Consequently, the approach enables a more objective evaluation of component condition and fosters continuous learning processes aimed at extending wheelset service life and reducing maintenance and stock costs for wagon keepers. Aggregated with the mileage of the wagon the development of the wheel flat could be

predicted in residual kilometres instead of time. Maintenance actions could be aligned with the operational plans of the railway undertaking and or with already planned workshop measures.

3 METHODS TARGETED FOR VALIDATION

The wheel flat detection and wheel flat length estimation are already discussed in detail by Birk et.al. (2025), but for the sake of completeness a short summary is given here.
Since the algorithms consider multiple detector system which are managed independently in different countries the observation reported in Öhman et.al. (2023) and Birk et.al. (2019) are considered in their implementation and operation on the acquired WILD data.

3.1 *Wheel flat detection*

The detection of a wheel flat and the estimation of its length are based on the observation that a wheel flat typically develops on both wheels of a wheelset simultaneously, resulting in an increase in wheel-rail contact forces. The process for detecting a wheel flat consists of four steps:

1. Track the dynamic force over time for left and right wheel
2. Establish a lower bound as the "normal" force level for each wheel and track it over time
3. Detect the occurrence of a simultaneous jump in the force levels on both left and right wheel
4. Flag for a wheel flat after the occurrence of two subsequent jumps on both left and right wheel

3.2 *Flat length estimation*

When a wheel flat is flagged, the length of the flat is estimated for each wheel individually. The idea is that the recorded forces relate to the length of the flat. Then a reverse calculation from the forces to the wheel flat length in line with Ye, et.al. (2020) is performed as follows:

1. Represent the wagon as a partial car model using a linear multi-body system
2. Simulate the forces during a passage of the wheel over the detector for different flat lengths
3. Select the flat length which fits the recorded forces best
4. Redo the estimate for subsequent passages to update the flat length and keep the largest estimate as the new estimate

The approach has certain limitations, as wheel flat hits the rail differently on every passage leading to underestimation. However, by updating the flat length estimate at subsequent WILD passages, the estimate will converge to the correct flat length over time. One-sided flats will not be detected, those may occur in uneven load scenarios or when a stumbling block is not correctly used. Moreover, multiple damages and especially RCF damage may occlude the presence of a wheel flat.
These aspects need to be considered in the validation and the quantification of the performance.

4 VALIDATION APPROACH

In simple terms, the validation of the wheel flat detection and flat length estimation was conducted by inspecting wheelsets and measuring the flat lengths on a large number of wheels.

4.1 *Test setup*

The validation project involves assessing wagons independently of their type and usage profile, either in workshops or in the field. The focus is on wagons that are equipped with RFID tags and have available wheel impact force data from Sweden, the Netherlands, and France, as well as dynamic ratio data based on impact forces from Switzerland.
Currently, these are the only countries in Europe where such specific data is accessible to wagon keepers, making them central to the project's data-driven approach.
The population of wagons that are targeted are selected as a subset of the VTG owned wagon fleet. The selected of the workshops and geographical field was aligned with the availability of WILD data on the wagons. Clearly, this limits the number of workshops and places, facilitating the acquisition of the inspection reports, and in the end four workshops were identified.
Irrespectively of the data that is acquired by the WILD, any wagons that were in the workshops were examined for damages on the wheels. Thus, inspection reports on wheels with and without damages are acquired.
The project was then conducted until enough wheels were inspected to reach statistical significance. The time frame was also selected such that the climatic conditions in the geographical region were quite similar. Thus, the spring and summer period was chosen, as winter in Scandinavia leads to a situation where the damage frequency is very high and workshops are quite overloaded. Such a situation would lead to an imbalance in the data set due to seasonal aspects.

4.2 *Digitalized acquisition of inspection data*

In a practical implementation of a decision support solution which also manages inspections and maintenance actions, the manual activities need to systematically be captured and designed such that human errors are minimized.

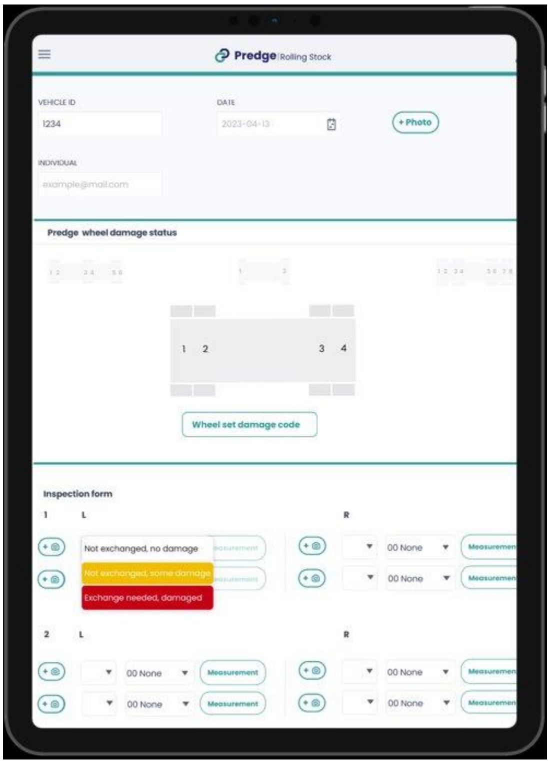

Figure 1 Online form for capturing of the wheel inspections according to GCU and VPI. It has the capability to photo document the damages

For this end, a dedicated online form, see Figure 1, has been developed to support and standardize the wheel assessment process, which can function reliably on a rugged hand-held device. All four workshops involved in the project have been equipped with a tablet and an RFID reader, see Figure 2, to facilitate the digital data collection. The form includes photo documentation, with a magnetic ruler placed next to the wheel damage, enabling the identification and verification of outliers in wheel damage measurements at a later stage

The assessment follows a standardized routine, which includes:

- Verifying the RFID tagging of each wagon
- Conducting the assessment in accordance with GCU and VPI standards
- Categorizing damage using a traffic light system: Red, Yellow, Green
- Capturing photo documentation for each wheelset

Figure 2 RFID reader and tablet used for workshop data collection

4.3 *Aligning wheelset assessments and analytics*

When performing analytics and performance assessments it is crucial to systematically and in an automated way to connect the different data sources with the physical asset to high data quality.

Therefore, all relevant data from workshop assessments and field inspections are collected and systematically stored together with the wheel impact load

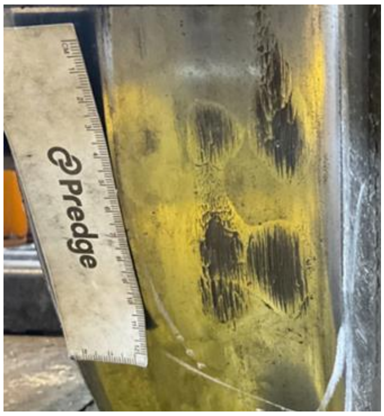

Figure 3 Multiple wheel flats on a single wheel

data from the WILD systems.

The envisioned platform discussed by Karim et.al. (2015) has been implemented by Predge as Predge RollingStock, where not only the data from the forms and the WILD systems is collected, but also the analytics methods for the wheel flat detection and the wheel flat length estimation are implemented.

As a result, the analytics outcomes are generated automatically and stored aligned with all the other relevant data on the physical asset.

In addition, it also enables to revisit the inspection reports when outliers are detected where the report is not in line with the analytics outcome to reconfirm the reporting using the photo documentation. Obviously, a wheel can be damages in various ways which means multiple damages of potentially different categories can be present at the same time, see Figure 3.

5 RESULTS

Over a four-month period, 680 wagons were assessed, comprising 2,720 axles and 5,440 wheels. Of these, 86% were covered by wayside monitoring data (WTMS), forming the basis of the analysis. Workshop evaluations indicated that approximately 13% of the wheels exhibited visible damage, with 78% of those identifiable via wheel force data. The majority of visible damage was classified as shelling or wheel flats (see Figure 4), with some cases of material build-up. Multiple damage types were often observed per wheel. Although assessments followed VPI and GCU guidelines, the manual nature of measurements introduces subjectivity and potential variability in reported damage severity.

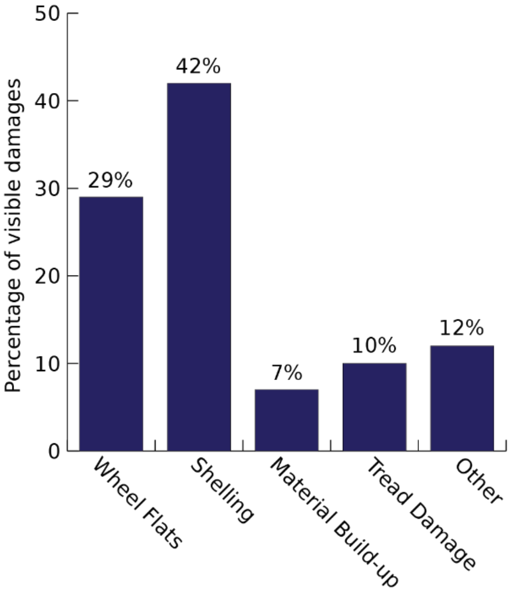

Figure 4 Damage classification by the workshops

The flat length predictions were done with respect to the daily mileage data from each wagon. An example of this can be seen in Figure 5, where a time series is shown to exemplify the assessment date. There is a sharp increase in the data indicating the detection of a wheel flat. On the right, the image taken of the flat on the assessment date can be seen, showing a flat length of ~35x35mm.

Overall results of the flat detection algorithm can be seen in Figure 6. Here the absolute difference between the measured flat length and the predicted flat length is shown. The majority of error (~45%) is below 2 mm whereas there are a few cases (~20%) where the flat length error was between 8 mm and 10mm. This can be explained by several smaller flats being measured as one large flat, or a wagon type where the dynamic model is not fitted to in the algorithm. In addition to this, the flat length estimation algorithm provided an accuracy of 99% and a precision of 67%.

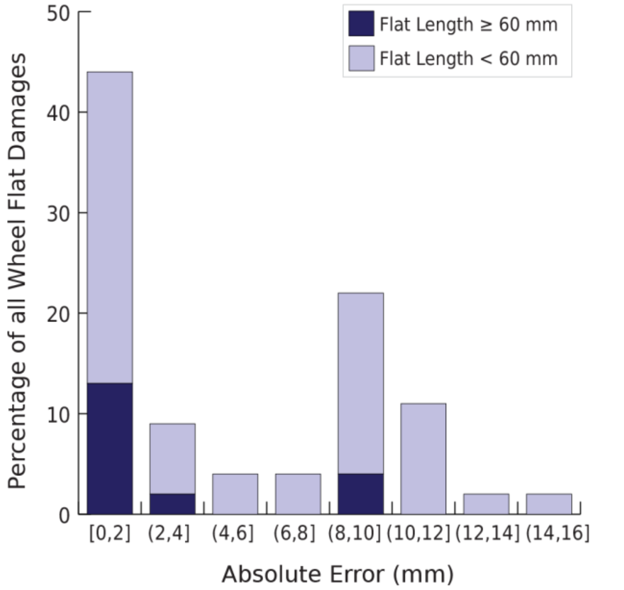

Figure 6 Absolute error of flat length estimation |Predicted - Measured|

The confusion matrix in Figure 7 compares predicted and assessed damage categories ('Red', 'Yellow', 'Green'). The top right corner of the matrix, bound by a green box, are axles classified as good enough for continued operation, whereas the bottom left corner (bounding-box red) indicates axles that were damaged and were predicted to be damaged. Most wagons

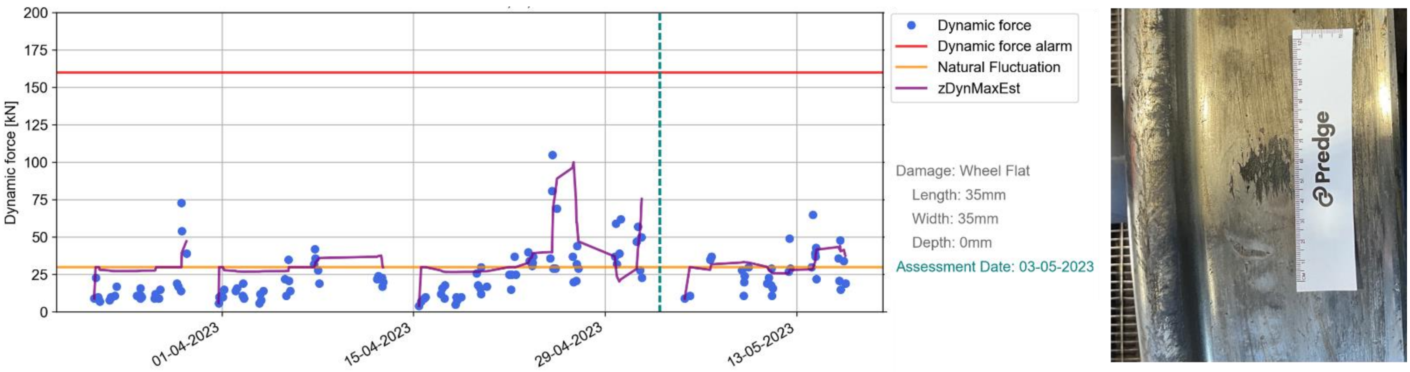

Figure 5 Raw data, predicted values, the assessment date and measured damage

were correctly classified, with 3.8% missed (false negatives) and 0.8% falsely flagged (false positives).

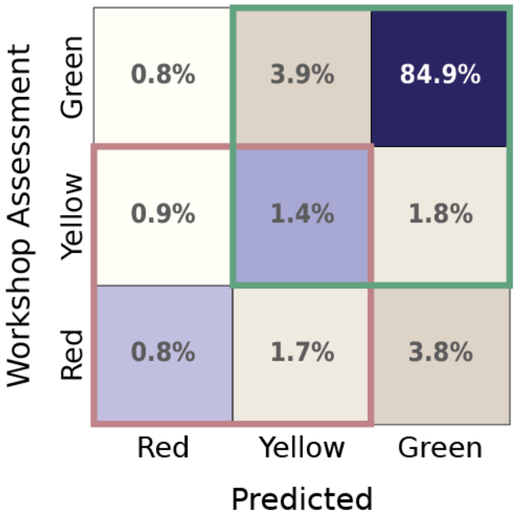

Figure 7 Confusion matrix of predictions and workshop assessments

6 CONCLUSIONS & RECOMMENDATIONS

This paper discusses the validation of wheel flat detection and wheel flat length estimation using data from wheel impact load detectors and inspection information from workshops and field.
The validation considers wagon operating in Sweden, The Netherlands, and Switzerland, and that are in normal operation. In total 680 wagons were inspected rendering 5740 wheels. For the validation population an accuracy of 99% and precision of 67% was achieved. Since the number of wheels is well beyond 100 the results can be seen as statistically significant.

Thus, the methods for detection and length estimation of wheel flats show promising results and have proven to provide the desired outcomes.
It is recommended to use these methods in real life by train operating companies and wagon keepers, as it makes maintenance more plannable and aids in shifting from a reactive regime to a condition based and predictive maintenance regime.
Currently, the solution is deployed in several train operations and more than 10,000 vehicles are continuously monitored and decision support is provided to maintenance roles in user organizations.

7 REFERENCES

Birk, W., Dittman, T., Karim, R., & Westerberg, J. (2019). Experiences from the detection and prediction of wheel damages on railway vehicles in operation. In *Proceedings of the 13th international heavy haul association STS conference. Narvik*.

Birk, W., Håkansson, A.L., Lindström, H., & Westerberg, J. (2025, June). Using Data from Multiple Wayside Train Monitoring Systems to Detect and Estimate the Size of Wheel Flats on Railway Vehicles. In *International Congress and Workshop on Industrial AI*. Cham: Springer Nature Switzerland.

Bogdevicius, M., Zygiene, R., Bureika, G., & Dailydka, S. (2016). An analytical mathematical method for calculation of the dynamic wheel–rail impact force caused by wheel flat. *Vehicle system dynamics*, *54*(5), 689-705.

Cui Y, Kim E, Yan S, Yu Q. (2025). Wheel flat detection using long short-term memory and transformer models with a 1:10 scale railway test rig. *Advances in Mechanical Engineering*. 2025;17(1).

Dong, R. G., & Sankar, S. (1994). The characteristics of impact loads due to wheel tread defects. *RTD Rail Transpn ASME*, *8*, 23-30.

General Contract of Use Bureau (2025), "General contract of use forwagons." [Online]. Available: https://gcubureau.org/

Hecht, M., Leiste, M., Jobstfinke, D., Deghela, U. (2018). Roadmap zur Digitalisierung der Wagentechnischen Untersuchung (Technical Report 16/2018), Faculty of Verkehrs- und Maschinensysteme, TU Berlin.

Iwnicki, S., Nielsen, J. C., & Tao, G. (2023). Out-of-round railway wheels and polygonisation. *Vehicle system dynamics*, *61*(7), 1787-1830.

Johansson, A., & Nielsen, J. C. (2003). Out-of-round railway wheels—wheel-rail contact forces and track response derived from field tests and numerical simulations. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *217*(2), 135-146.

Kalay, S., Tajaddini, A., Reinschmidt, A., & Guins, A. (1995, June). Development of a performance-based wheel-removal criteria for North American railroads. In *Proceedings of 11th International Wheelset Congress, Paris, France* (pp. 7-18).

Karim, R., Birk, W., & Larsson-Kråik, P. O. (2015). Cloud-based emaintenance solutions for condition-based maintenance of wheels in heavy haul operation. In *International Heavy Haul Association: The 11th International Heavy Haul Association Conference will be held 21-24 June 2015 in Perth 21/06/2015-24/06/2015*. International Heavy Haul Association.

Öhman, J., Birk, W., & Westerberg, J. (2023, June). Wheel Damage Prediction Using Wayside Detector Data for a Cross-Border Operating Fleet with Irregular Detector Passage Patterns. In *International Congress and Workshop on Industrial AI* (pp. 491-501). Cham: Springer Nature Switzerland.

Perrey, A. (2025), VPI European Maintenance Guide, https://vpihamburg.de/en/vers/vpi-emg/overview

Swedish Transport Administration (2014). Bvf 592.11 – detektorer. hantering av larm från stationära detektorer samt åtgärder efter upptäckta skador vid manuell avsyning. (tdok 2014:0689).

Ye, Y., Shi, D., Krause, P., & Hecht, M. (2020). A data-driven method for estimating wheel flat length. *Vehicle System Dynamics*.

Zhou, Y., Vuitton, J., Tian, Q., & Hecht, M. (2024). Wheel flat detection by using the angular domain synchronous averaging method and axle box acceleration: Simulation and experiment. *Measurement*, *230*, 114508.