Towards automated detection and diagnosis of suspension system defects in passenger railway vehicles

Nikolaas N. Oosterhof[1], Margot F.E. Peters[1]

[1] *NS, Utrecht, The Netherlands*
nick.oosterhof@ns.nl
margot.peters@ns.nl

ABSTRACT

Suspension defects in passenger trains reduce safety and comfort in rail transport systems. We investigated the feasibility of detecting such defects using dynamic wheel load differences (DWLDs), which are measured from trains in operation using sensors in the track.

We found that DWLD data shows considerable variability but in a consistent manner, and developed a method that purifies the data to improve signal-to-noise. Further, we developed an algorithm that uses the purified DWLD data to detect anomalous events in suspension imbalance, which is indicative of a defect or repair event. Our algorithm further provides a diagnosis of an anomalous event as related to either the primary or the secondary suspension system. We validated our algorithm using a limited set of maintenance records and found a high detection and correct classification rate, although more extended validation is needed with more maintenance records and DWLD data.

In sum, our work indicates a promising avenue of high-frequent, automated detection and diagnosis of suspension defects, which would contribute to efficient, economical and save operation of railway vehicles.

1. INTRODUCTION

With more than a million passengers every day, railways are a popular means of passenger transport in the Netherlands. This critically depends on efficient, economical, reliable and safe operation of railway vehicles, which requires minimal probability of derailment and high passenger comfort. Those may negatively be affected by defects on suspension systems used in wheel sets and bogies of passenger coaches. To ensure a safe and comfortable ride, the vibrations that originate from irregularities on the track are damped by the suspension system of each coach. This suspension is shown in Figure 1:

Nikolaas N. Oosterhof 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.

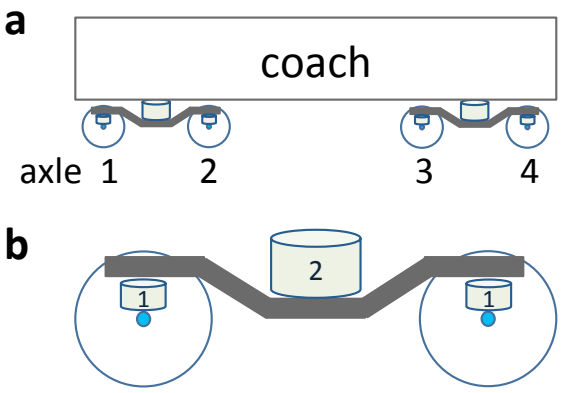

Figure 1. Simplified side view of a coach (a) and two wheelsets in a bogie (b), with primary (b.1) and secondary (b.2) springs.

every bogie consists of two wheelsets that are connected to the bogie with primary springs (item 1 in Figure 1). The bogie itself is connected to the coach with secondary springs (item 2 in Figure 1), which are supplied with compressed air.

The suspension system is inspected visually every few days because of its impact on safety. However, defects or wrong settings are not always visible with the naked eye due to the location of the springs and the placement of surrounding components. Therefore it would be of great value to have an alternative detection method that is automated depends less on human factors.

In this paper we aim at developing such a detection method based on dynamic wheel load differences (DWLDs), which will be introduced in Section 2. Furthermore, in Section 2 the physical interpretation of DWLDs in relation to suspension defects is explained. In Section 3, a corrective method is developed to decrease variability in DWLDs, of which a significant amount can be attributed to systematic factors. In Section 4, we propose an algorithm to detect the sudden changes

in DWLDs which might indicate defects or repairs in wheelsets or bogies. In Section 5, we present the results of verification of the algorithm, indicating a high detection rate. Section 6 concludes the paper.

2. SUSPENSION DEFECTS & WHEEL LOAD DIFFERENCES

The Dutch rail infrastructure manager ProRail has installed the 'Gotcha' measurement system (Boom, 2007) at about fifty locations in infrastructure across the Netherlands. Primary goals of this system are the measurement of wheel and axle loads and the detection of wheel defects. NS, the primary railway operator for passenger transport in the Netherlands, uses these measurements to reprofile the wheels in case of wheel defects.

In the present work a potential secondary purpose of the Gotcha system is explored: the detection of suspension defects. Since suspension defects typically lead to an unbalanced wheel load between the left and right wheel of a wheelset, the monitoring of wheel load differences with Gotcha could provide an alternative detection method of such defects.

A wheel load difference (WLD) is defined in Commission Regulation EU (2014) No 1302/2014 as follows:

$$\text{WLD} = \frac{Q_r - Q_l}{Q_r + Q_l},$$

where Q_i represents the wheel load of the left ($i = l$) or right ($i = r$) wheel of the same axle. A WLD that deviates from zero indicates an imbalance in the load distribution. A negative WLD means represents an imbalance to the left side of the train; a positive WLD an imbalance to the right side.

The *static* wheel load difference of new, overhauled and refurbished trains is measured inside the depot. This static WLD is limited by European standards (Council of European Union, 2014) and therefore checked before the train is admitted to the track. With the Gotcha measurement system, the wheel load differences are computed using *dynamically* measured wheel loads during operation; thus we refer to these dynamic WLDs as DWLDs. Hunting oscillations can cause small DWLDs but the amplitude of this oscillation is limited. Large DWLDs can reasonably be assumed to have been caused by a new defect, since structural skewness in the construction of the train body is limited by the European standard for static WLDs before admittance. Therefore, in search for an alternative detection method for suspension defects, we start with exploring DWLDs measured with Gotcha.

We expect a change in this measure when a suspension defect occurs. Moreover, because of the type of defects (a broken spring; air leakage or a broken valve in the suspension system), we expect this change to be sudden. Apart from defects, a wrong setting of the height control bars in the secondary suspension system could also cause an imbalance in the suspension system. This will generate a sudden change in DWLDs as well, right after adjustment of the settings. Therefore, we will first focus on sudden jumps in the measurements of DWLDs.

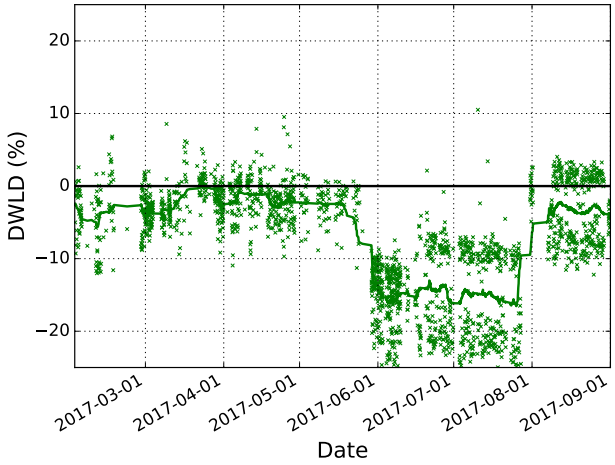

Figure 2. DWLDs of one of the axles in a bogie with an air leakage in the secondary suspension. Each dot is a DWLD measurement in Gotcha, the line is the moving average DWLD over 50 measurements.

As a first step, we investigate DWLDs of a coach with a visually confirmed air leakage in the secondary suspension that was repaired subsequently. DWLDs of one of the affected axles are presented in Figure 2. Also, this figure shows the substantial variance in DWLD measurements, which will be further addressed in Section 3.

According to DWLDs in Figure 2, the defect is likely to have occurred in the second half of May 2017. Furthermore, DWLDs show an excessive jump back to values near zero when the defect is repaired, as maintenance records show at the end of July 2017.

Due to the stiffness of the coach and the construction of the suspension system, a secondary suspension defect causes torsion over the whole coach; the first bogie is unevenly loaded in opposite direction from the second bogie. Therefore a defect in one secondary spring is expected to affect DWLDs of all axles of a coach in a systematic way. This is illustrated by the data shown in Figure 3, which shows DWLDs of all axles of this coach.

Secondly, we investigate DWLDs of a coach with a visually confirmed primary suspension defect. Because the primary suspension connects each axle to the bogie, a primary suspension defect is expected to cause torsion within one bogie, i.e. to affect the DWLDs of two axles. In Figure 4 the behavior of DWLDs of a coach with a primary suspension defect in axle 1 are shown. From this figure it is presumed that the spring broke around the 10th of October, where a sudden change in

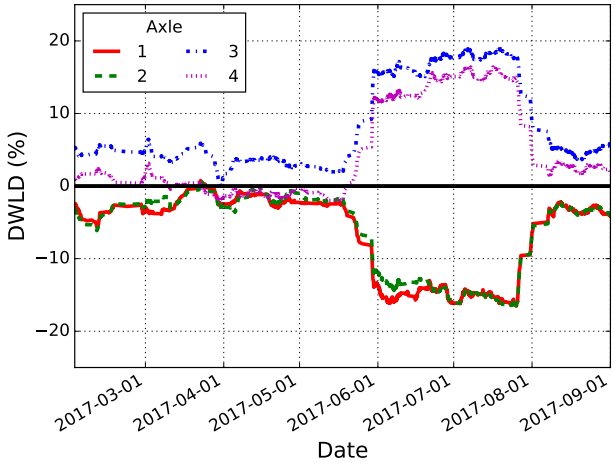

Figure 3. DWLDs for all axles of one coach in which a secondary suspension defect occurs (end of May) and is repaired (end of July).

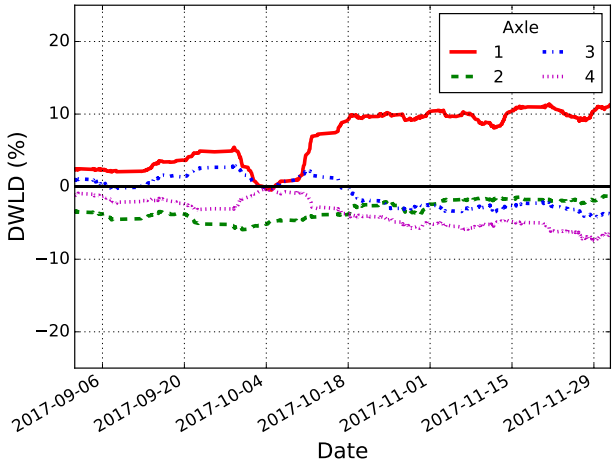

Figure 4. Moving average DWLDs (over 50 measurements) for a coach with a primary suspension defect.

DWLDs of axle 1 occurs, while DWLDs of axle 2, 3 and 4 remain unaffected.

The repair date of the defects above coincides with a sudden decrease of DWLDs in the affected axles. Although the date of initiation of the defects is not known, a highly educated guess can be made based on the sudden increase in DWLDs some time period before the repair date.

Figures 2 to 4 indicate that detecting sudden changes in DWLDs could be a feasible approach to detect suspension defects. Furthermore, combining DWLDs of different axles in one coach, enables us to indicate the location of origin of the defect (primary or secondary suspension). Other investigated cases show similar results, as shown more elaborately in the verification in Section 5. However, as illustrated in Figure 2, DWLDs show considerable variance which could hamper detection of changes in DWLDs. In the next Section we explore systematic sources of variance in more detail in an attempt to reduce the variance, which we is needed for accurate detection of anomalous events in DWLDs.

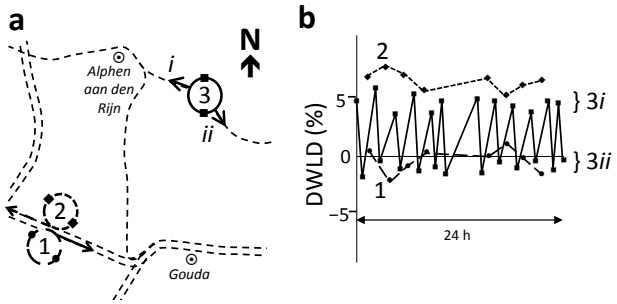

Figure 5. Figure 5: (a) Schematic map showing the locations and train direction with typical train directions for double tracks (detector 1, 2) and single track (detector 3). (b) Illustration of DWLDs during one day measured by three detectors numbered 1, 2 and 3. N indicates North.

3. REDUCING VARIANCE IN DWLDS

As an example of the variance in DLWDs, Figure 5 shows a train operated during one day, passing two detectors on a double track (detectors 1 and 2) and a third detector on a single track (detector 3) in two orientations (3*i* and 3*ii*). Importantly, this train does not turn, meaning that eastward passages along detectors 1 and 3 are with cabin A leading, whereas westwards passages along detectors 2 and 3 are with cabin B leading. It can be observed that for the same orientation and detector, measurements are fairly consistent. They are not expected to be identical due to other factors including variation in of passenger load and detector measurement errors. Yet there is substantial variability for different measurements, not only as a function of detector but also of orientation. This indicates a more complex interaction between train, infrastructure and measurement system that goes beyond trivial calibration issues.

Such differences due to detector and/or orientation could result in excessive and sudden changes in DWLDs in a coach if the set of detectors that are passed by the coach changes abruptly. This is not a mere theoretical possibility: a change of train schedule or train set assignments may result in a train operating in another geographic region, and thus a change in the set of detectors it passes. To reduce the occurrence of sudden changes in DWLDs caused by a change of the set of detectors that a train passes, the data can be purified by correcting for these kinds of systematic factors. In order to do that accurately, we will first assess the consistency of variance over different time intervals and different trains in Section 3.1. Second, we will consider the amount of explained variance by these systematic factors in Section 3.2, to con-

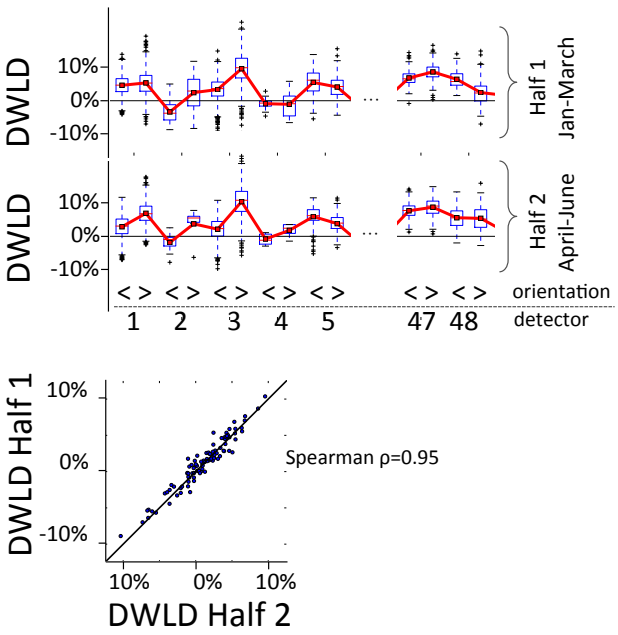

Figure 6. (a) DWLDs measured for two halves of the VIRM-IV fleet measured for different time periods (top and bottom plots). Each bar represents measurements for a combination of detector and train orientation. (b) Scatter plot of average DWLDs for each combination of detector and orientation across the two halves.

clude on the purification method in Section 3.3.

3.1. Consistency of variability over time and trains

First we assessed the consistency of measurements over time as a function of detector and orientation (see Figure 6a). As an internal consistency check, we split all trains from one sub-series (VIRM-IV) in two halves pseudo-randomly. To be able to assess the temporal consistency, the first half of trains used data from a three month interval (January-March 2017) and the second half of trains from another non-overlapping three month interval (April-June 2017). Thus we divided the data into four subsets (half 1 Jan-Mar; half 1 Apr-Jun; half 2 Jan-Mar; half 2 Apr-Jun), and compared two sets that have no train and no timestamp in common.

Then we considered, for each sub-series half, DWLDs from the first axle, and split these according to detector and train orientation. We only considered passages in the 'positive' track direction, as this is the common direction for passages, and we wanted to exclude the possibility of a small number of passages skewing our results. For each combination of sub-series half, detector and train orientation we computed the median DWLD (we used the median and not the mean to avoid effects from outliers caused by possible unknown defects).

Finally, we computed the Spearman correlation between the median values in the first half and the second half to assess whether DWLD values were consistent across train sets and over time. As shown in Figure 6b, DWLD values were highly consistent with a Spearman correlation of 0.95 across the two halves.

In addition we conducted a similar split-half analysis that considered axle position, which also showed high consistency for DWLDs over detectors and orientation (but somewhat puzzling, consistent anti-correlations between some pairs of axles); details are shown in Appendix A.

3.2. Explained variance by systematic factors

Second, we considered the effects of variability more systematically by computing how much variance is explained by each of four factors: detector, axle number, train orientation, and train direction. We used a training set with three months of DWLD data for one half of the VIRM-IV train sets, and a validation set using the other half. We took any non-empty subset of the four factors, resulting in $2^4 - 1 = 15$ combinations. In the training set we split the data according to the combination of factors, and computed the median value of DWLDs. Then in the validation set we also split the data according to the combination of factors, and in each split subtracted the median value of DWLDs obtained from the training data to obtain purified testing data.

We then computed the variance of both the original validation data and the variance of the purified validation data. Finally we computed the amount of explained variance of the purification process by subtracting the ratio of variance of purified data to the variance of the original validation data from one.

Results for each of the 15 combinations of factors are shown in Figure 7 for the VIRM-IV. These indicate that:

- the combination of orientation, axle and detector explains around 50% of variance;
- the explained variance by these three variables is higher than the explained variance by each variable alone, or by a pair of two variables;
- the factor of train direction does not increase the amount of variance explained.

We also considered three other train types operated by NS, namely the ICM, SGM and SLT. For these series we observed a very similar pattern of results. We note that a full understanding of the physical mechanism underlying this pattern remains elusive. We have consulted experts from within NS, ProRail and several other companies but did not manage to reach a satisfying physical explanation yet. although we cannot exclude the possibility of hardware-software interactions in the Gotcha detectors (that we do not fully understand). We hope that future research will lead to a better understanding of the physical mechanisms underlying the pattern of data that

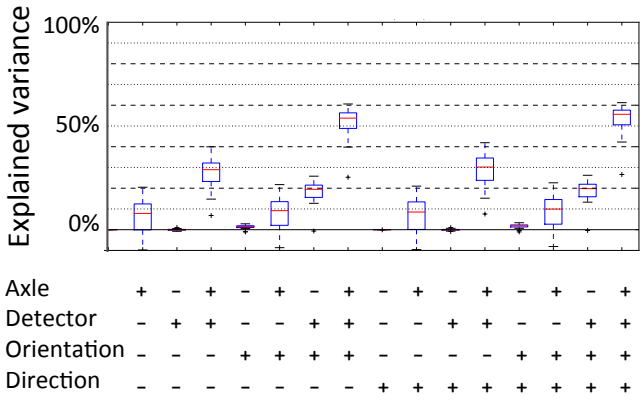

Figure 7. (a) Explained variance in dynamic wheel load difference values as a function of combinations of orientation, axle, detector, and train direction.

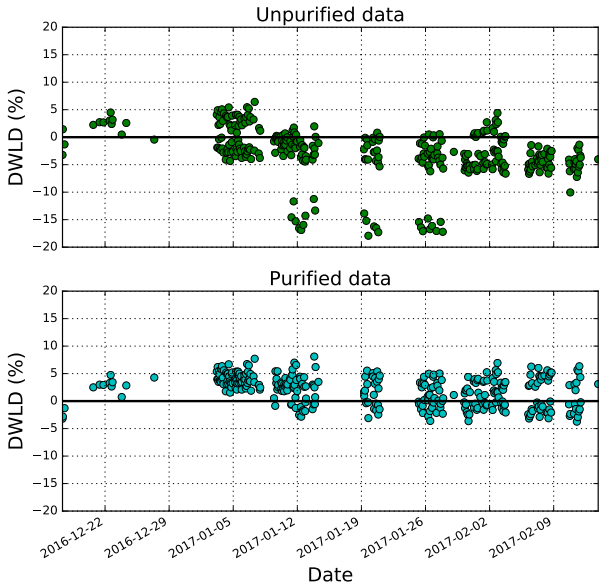

Figure 8. (a) Unpurified and (b) purified DWLD data of one axle of one train over time.

we observe.

3.3. Purification method

Despite currently lacking a full physical understanding of the variability in DWLDs, our analyses support the notion that DWLDs can be purified by computing a correction term for each combination of orientation, axle and detector based on a large set of measurements from multiple train sets and over a long temporal interval. An illustration of the effect of purification is presented in Figure 8, where the purified data shows fewer outliers and less variance than the unpurified data. This purified data can be used for a more sensitive detection of suspension defects through Dynamic Suspension Imbalance anomaly events, as described in the next section.

4. ANOMALOUS EVENTS IN DYNAMIC SUSPENSION IMBALANCE

We propose an algorithm to detect anomalous events in DWLDs, which might indicate defects or repairs in wheelsets or bogies. By integrating information across wheel-axles, as illustrated in Figure 2, this algorithm can also differentiate between defects in primary and secondary suspension. Next to a high DWLD value, these defects can cause bogie diagonal imbalance or coach diagonal imbalance. All imbalance values are grouped under the term Dynamic Suspension Imbalance.

A preparation step in this algorithm is purifying the data as described in Section 2. After purifying the data, we group DWLDs based on a single passage of each coach; thus each group consists of DWLDs of all axles during the passage of a coach along a detector. In order to diagnose the cause of a sudden change in DWLDs, we introduce two new quantities. Using the purified data, we compute bogie diagonal imbalance values $I^{\text{diag}}_{\text{bogie}}$ through

$$I^{\text{diag}}_{\text{bogie 1}} = \frac{\text{DWLD}_1 - \text{DWLD}_2}{2}$$

$$I^{\text{diag}}_{\text{bogie 2}} = \frac{\text{DWLD}_3 - \text{DWLD}_4}{2}$$

for two bogies in a coach, and the coach diagonal imbalance $I^{\text{diag}}_{\text{coach}}$ values as

$$I^{coach}_{diag} = \frac{\text{DWLD}_1 + \text{DWLD}_2 - \text{DWLD}_3 - \text{DWLD}_4}{4}$$

where $\text{DWLD}_{1\ldots4}$ are DWLDs of the four axles in that coach.

Note that these formulas are very similar but not exactly identical to the definition in the Commission Regulation EU (2014) as that definition uses wheel loads instead of wheel load differences. We cannot use the Commission Regulation EU definition in combination with the purification process described in section 3, as the purification terms are based on DWLD values, not on wheel loads. However, we compared the values from our definition with those from the Commission Regulation EU's definition (using unpurified data), and found that the bogie and coach diagonal imbalance values show a correlation between 99.8% and 100.0%, supporting our formulas for the diagonal imbalance values.

In summary, we use three types of quantities (all based on the purified data): DWLDs, which provide information about (im)balance in individual wheel sets; $I^{\text{diag}}_{\text{bogie}}$, quantifying torsion at the bogie level and indicative of primary suspension defects; and $I^{\text{diag}}_{\text{coach}}$, quantifying torsion at the coach level and

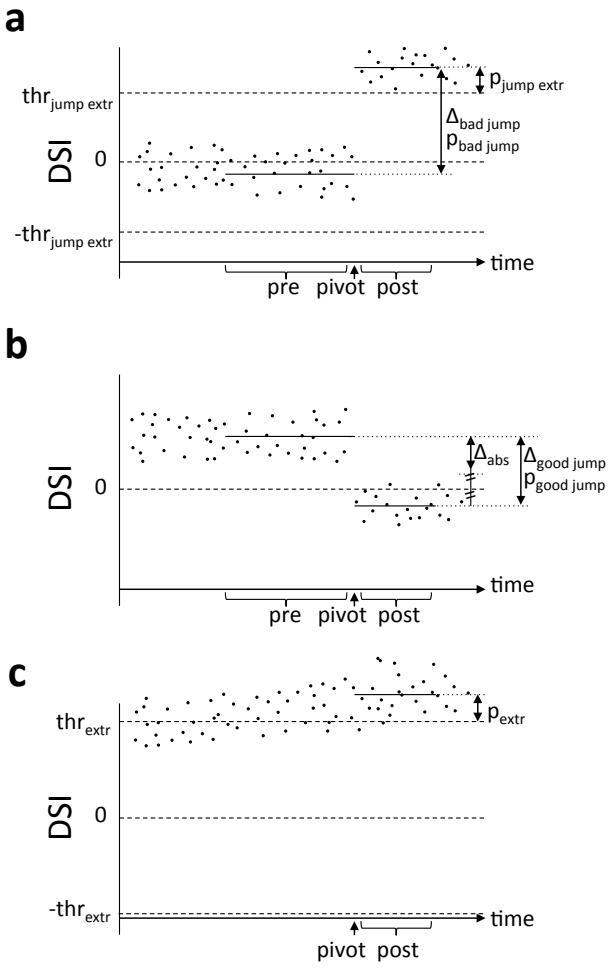

Figure 9. (a) Examples of (a) bad jump, (b) good jump, and (c) extreme value event. Each plot shows dynamic suspension imbalance (DSI) values (vertical axis) as a function of time (horizontal axis). See the main text for details.

indicative of secondary suspension defects. Collectively we refer to DWLDs, $I^{\text{diag}}_{\text{bogie}}$ and $I^{\text{diag}}_{\text{coach}}$ as Dynamic Suspension Imbalances, or DSIs. For all three types of DSIs, a value near zero indicates good balance, whereas a value deviating strongly from zero may indicate a suspension defect.

For each type of DSI, we consider three possible anomalous events of interest: (a) a bad jump, (b) a good jump, and (c) an extreme value; see Figure 9.

First, a bad jump event (Figure 9a) occurs when there is a sudden increase of the absolute DSI value. This may indicate that a defect occurred, such as a broken spring in the primary suspension system (leading to a high DWLD and/or $I^{\text{diag}}_{\text{bogie}}$ value), or a leak in the secondary suspension system (leading to a high DWLD and/or $I^{\text{diag}}_{\text{coach}}$ value). To detect a jump, we take a 'pivot' time point, and then take a certain number of measurements before the pivot point (the pre-period) and a certain number of measurements after the pivot point (the post-period). For a bad jump event to occur, we define the jump as the difference between the DSI values in the pre-period and the post-period and require that:

- The absolute DSI value in the post-period must exceed a threshold value $\text{thr}_{\text{jump extr}}$. Specifically, we require that the DSI values deviate significantly more from zero than a threshold DSI, according to a probability threshold $\text{p}_{\text{jump extr}}$ from a Wilcoxon signed-rank test;
- The jump must be large enough in effect size $\Delta_{\text{bad jump}}$ so that the absolute difference between the medians of the pre- versus post-period DSI values exceeds an effect size threshold; and statistically significant so that the pre- versus post-period show different medians according to a probability threshold $\text{p}_{\text{bad jump}}$ from a Wilcoxon-Mann-Whitney test.

Second, a good jump event (Figure 9b; within NS also known as a cake event to symbolize improvements) occurs when there is a sudden decrease of an absolute DSI value. For a good jump event to occur, we require that:

- The jump must be sufficiently large enough in effect size $\Delta_{\text{good jump}}$ and statistically significant according to a probability threshold $\text{p}_{\text{good jump}}$ from a Wilcoxon-Mann-Whitney test (in similar fashion as a bad jump, although different thresholds can be used);
- The improvement must be large enough in effect size: the difference of the absolute median DSI value during the post-period minus the that of the pre-period Δ_{abs} must exceed an a threshold value.

Third, an extreme value event (Figure 9c) occurs when a DSI value deviates strongly from zero. It is aimed to serve as an additional safety net for cases where a defect developed gradually thus a bad jump was not detected. Extreme value events are detected using only data from the post-period. For an excessive value event to occur, we require that:

- The absolute DSI value in the post-period must exceed a threshold value thr_{extr} according to a probability threshold p_{extr} from a Wilcoxon signed-rank test.

We note that detection of these events depends on setting various threshold parameters, which in turn affects the balance between false positive and false negative rate. We have chosen initial threshold parameters that result in the detection results reported in this manuscript, and plan to further optimize these values based on analyses on future maintenance reports and DWLD data.

For time series analysis, we move the pivot point over time, and for each temporal position see if an event was detected. Since neighboring pivot points also show considerable overlap of data points, detected events tend to cluster temporally. To reduce such clusters to a single event, we perform the following steps:

	Truth		DSIs	
Case	Type of issue	Suspension	Anomalous event	Suspension
1	defect	primary	yes	secondary[1]
2	defect	secondary	yes	secondary
3	defect	secondary	yes	secondary
4	defect	primary	yes	primary[2]
5	defect	secondary	yes	secondary
6	setting	secondary	yes	secondary
7	setting	secondary	yes	secondary
8	setting	primary	yes	primary

Table 1. Results of verification

[1] wrong diagnosis of defect. [2] anomalous event detected on neighboring axle. Abbreviation: DSI, Dynamic Suspension Imbalance.

1. We consider all time points and find the strongest event based on effect size.
2. We then eliminate neighboring time points within a certain temporal distance of the same event type from consideration.
3. Iteratively we find the next strongest event.
4. We eliminate its neighboring time points from consideration.

These four steps are repeatedly performed until no more events are found. This analysis is done separately for each coach; event type (bad jump, good jump, and extreme value); DSI type and position number (i.e. each wheel set's DWLD, bogie's $I^{\text{diag}}_{\text{bogie}}$ and coach's $I^{\text{diag}}_{\text{coach}}$ values are analyzed separately). The anomalous events that are obtained in this manner can then be used for subsequent analyses and/or as the basis to send alarms to maintenance personnel.

5. VERIFICATION

The original DWLDs have already been available in a real-time monitor to maintenance and reliability engineers in 2017. However in this monitor the proposed purification and automatic detection of sudden changes or excessive DSIs described earlier were not applied, because these were not yet available at that time. Eight cases of extreme DWLDs in this monitor have been investigated in the depot. The maintenance personnel has judged visually in DWLD monitor what they considered 'extreme' values. The result of this investigation in the depot is shown as 'Truth' in column 2 and 3 of Table 1. In all cases of extreme DWLDs, issues were found in the suspension system. Two types of issues are distinguished in both the primary and secondary suspension: defects (broken spring, valve defect, air leakage) and settings (uneven length of primary springs, wrong setting of height control bars of secondary springs).

For each of these cases we tested our algorithm, resulting in the last two columns of Table 1 as DSIs (Dynamic Suspension Imbalances). For the computation of the anomalous events (bad jumps, good jumps and excessive values), the purified DSIs were used. In one case (case 4 in Table 1) the suspension defect did result in an anomalous event detected on the wrong axle (axle 3 instead of axle 2) in our algorithm, possibly caused by a coach imbalance in combination with an axle effect that was too small to be detected with our current threshold settings. Furthermore, in one case (case 1 in Table 1) the DSI type seems to have been diagnosed wrongly: a secondary suspension defect instead of a primary suspension defect. However, according to the high $I^{\text{diag}}_{\text{coach}}$ this coach is very likely to have had a secondary defect as well, that obscured the primary defect. The secondary suspension system was not further investigated once the primary defect was found due to time constraints, so this secondary defect is not (dis)confirmed yet.

These results show that the use of purified DWLDs and detecting anomalous events based on DSIs are a promising method for detecting suspension defects. We plan to test and verify the algorithm for all train types for a longer period of time, so that its settings can be optimized to detect and diagnose anomalous events more accurately.

6. CONCLUSION & DISCUSSION

In summary, our results indicate a promising avenue for detection of suspension system defects. More work is needed to understand and reduce the observed variance in measurements. Furthermore, we will validate our approach further by matching excessive changes in dynamic suspension imbalance values against suspension system defects or maintenance events. This could form the basis of automated detection and diagnosis of suspension system defects during service, which would contribute to efficient, economical and save operation of railway vehicles.

After further verification and optimization, our method could be used to decrease the frequency of visual inspections of the suspension system.

ACKNOWLEDGEMENT

We thank our colleagues Bob Huisman and Inka Locht for useful comments on an earlier draft of this paper.

ABBREVIATIONS & SYMBOLS

DWLD	dynamic wheel load difference
DSI	dynamic suspension imbalance
$I^{\text{diag}}_{\text{bogie}}$	dynamic bogie diagonal imbalance
$I^{\text{diag}}_{\text{coach}}$	dynamic coach diagonal imbalance
ICM	intercity materieel
NS	Nederlandse Spoorwegen
SGM	stadsgewestelijk materieel
SLT	spinter light train
VIRM	verlengd interregion materieel
WLD	wheel load difference

REFERENCES

Boom, P. (2007). Gotcha! intelligent maintenance and network asset management system. In *AusRAIL PLUS 2007, 4-6 December 2007, Sydney, NSW, Australia.*

Council of European Union. (2014). *Council regulation (EU) no 1302/2014 concerning a technical specification for interoperability relating to the 'rolling stock — locomotives and passenger rolling stock' subsystem of the rail system in the european union.*

APPENDIX

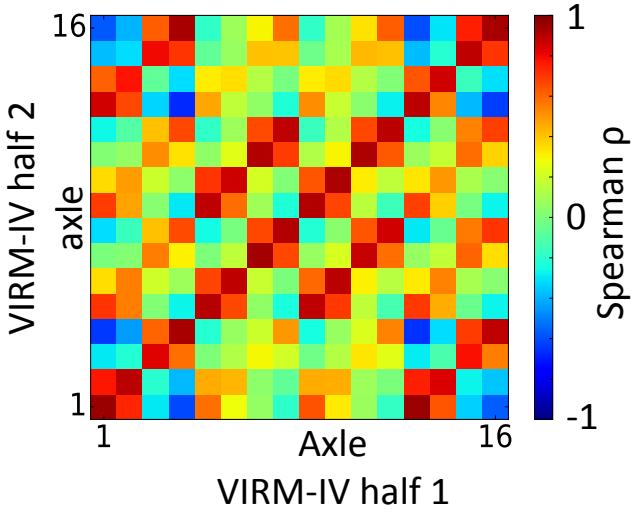

Figure 10. cross-correlation heat map showing the correlations for all axle-combinations. Each cell represents the correlation of the DWLDs formed by all combinations of detectors and orientation (c.f. Figure 6b). Note the curiously strong positive and negative correlations between, for example, axles 1 to 4.

BIOGRAPHIES

Nikolaas N. Oosterhof (Groningen, The Netherlands, 1982) has a PhD in psychology from Bangor University, Bangor, U.K. He has three master's degrees: in Cognitive Science from the University of Amsterdam, Amsterdam, The Netherlands; and in computer science and in philosophy of science, technology and society from Twente University, Enschede, The Netherlands. Formerly a postdoctoral researcher in cognitive neuroscience at the University of Trento, Italy; Dartmouth College, U.S.A.; and Harvard University, U.S.A., he has been working the field of Prognostics and Health Management at NS for half a year.

M.F.E. Peters (Elst, The Netherlands, 1988) has a Master's degree in Theoretical Physics from Radboud University Nijmegen. She has been working in the field of Prognostics and Health Management at NS for 6 years and is local chair of the European Conference of the PHM Society in Utrecht in 2018.