Early Warnings for failing Train Axle Bearings based on Temperature

M.F.E. Peters[1]

[1]*Netherlands Railways (NS), PO Box 2167, 3500 GD Utrecht, Nederland*
Margot.Peters@ns.nl

ABSTRACT

By studying the temperature behavior of axle bearings both statistically and physically, the research and development (R&D) department of Netherlands Railways (NS) has successfully developed and implemented a health monitoring system for bearings. In an early stage of degradation, temperature deviations are detected and the level of severity of the degradation is identified through a decision tree. This method enables us to detect bearing failures one to three months earlier than any other method in use, in more than half of the cases. Different handling scenarios per type of temperature behavior have been designed in a way that minimizes impact on train service.

1. INTRODUCTION

NS exploits and maintains approximately 3,000 carriages. The maintenance program encompasses first line service, running maintenance and overhaul. The latter two types of maintenance focus on preventive tasks in order to improve the reliability of the trains, whereas first line service focuses on routine checks for the components that are directly critical for safety (Apallius de Vos, J. & van Dongen, L.A.M., 2015). The safety checks are performed during the night at shunting yards, by inspecting the outside of the train visually and auditory.

The most important components for safety and thus subject for safety checks are the axle bearings in the wheelset. Functional failure can be caused by different mechanical and electrical failure mechanisms, which in most cases result in an increasing temperature of the bearing at a certain rate. A hot axle bearing is proven to be increasingly hard to recognize visually from the outside of the axle box because of the higher resistance to discoloration of recently developed conservation materials. Therefore the axle box temperature of passing trains is measured by the infrastructure manager ProRail, at 27 locations in the Dutch network. This system uses infrared sensors and is called HotBox detection.

In the current setting, when the measured temperature of a passing axle box exceeds the limit of 115°C the driver of the train is commanded by phone to stop immediately. The timing of this alarm is late in the process of failure: at this temperature, the probability of the bearing causing secondary damage to the wheelset has already increased significantly. Moreover, the immediate stop of the train causes large disturbances in the dense railway network. Despite the fact that occurrence of axle bearing failures is rare, the impact on service is large enough to justify the investment in a predictive method for this type of failure in addition to the safety checks.

Therefore, we have set up a study on the temperature measurements of the HotBox detection system. The analysis focused on one type of train with bearing problems in a period of two years. The failing bearings were noticed by train drivers who heard a humming sound or by mechanics during routine safety checks, before the temperature of the bearing had exceeded the alarm limit of HotBox. The degradation that was found for these bearings did not cause functional failure of the wheelset yet, but would have led to functional failure within time.

In this paper first the failure behavior of train axle bearings and the method for detection are explained. Secondly, we present the design of the handling scenarios for alarms. In the end the results of this alarm system are discussed.

2. TEMPERATURE BEHAVIOR OF BEARINGS

Failure of train axle bearings can be caused by several different failure mechanisms: fatigue, spalling, creep, stress corrosion, fretting corrosion, false brinelling, contamination of grease, or by craters because of electrical discharge through the bearing (Dasgupta, A., & Pecht, M., 1991 and Vale, C. et al, 2016). These mechanisms increase the friction within the bearing by deforming the components of the bearing, by generating debris material, or by changing the properties of the grease. Increased friction always results in an increase of the temperature in the bearing.

However, temperature variations are also part of normal behavior of the bearings in the train. These variations are due to operational conditions, such as environmental temperature, sunshine (see Figure 1), and by the time schedule: the bearing cools significantly by the wind while driving and warms up while standing still. Furthermore, some bearings can seem slightly warmer than others because of an engine or gearbox in their direct neighborhood that influences the HotBox measurement. Also, the variance as well as the magnitude of temperature measurements of intercity and commuter trains differ significantly because of differences in operational use and in construction of the axle box, as illustrated in Figure 2.

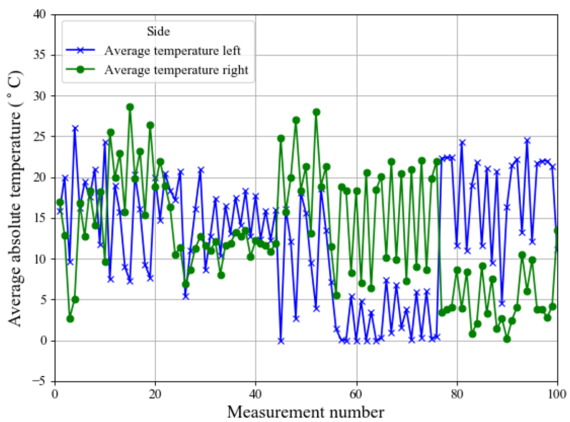

Figure 1: Average temperature of all bearings on one side of a train per measurement. A train is measured several times a day by HotBox.

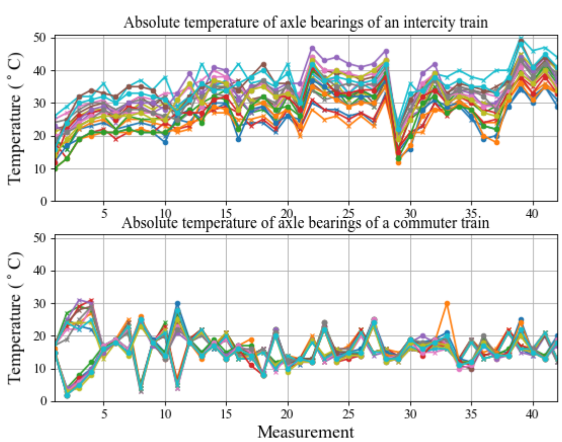

Figure 2: Temperature measurements of an intercity train and a commuter train. Each line represents one axle bearing.

In order to explore the possibilities for a predictive module, the temperature behavior has been studied considering the trains in which a bearing failure was found.

One of the bearing failures is shown in Figure 3, where the absolute value of the measured temperature in degrees Celsius is set against the amount of days before the bearing failure was noticed at the right-hand-side of axle 7.

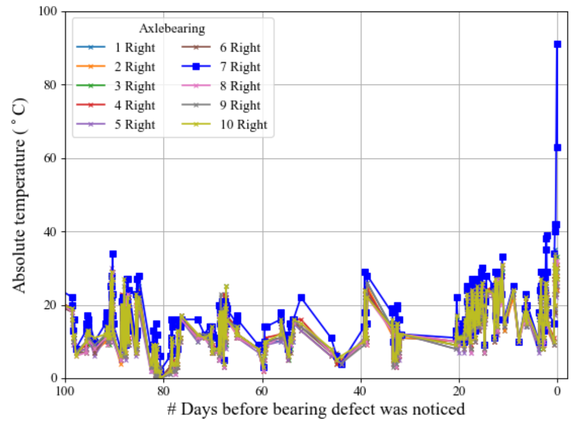

Figure 3: Temperature measurements of all bearings of a train. The bearing marked with a square is degrading.

As this figure suggests, the failing bearing is already showing a slightly higher temperature than the other bearings up to 100 days before the defect was noticed in a safety check. Similar results are found for other trains with a bearing that had started to degrade: in more than half of the cases the temperature of that bearing was structurally higher than temperatures of other bearings on that train, for a period of at least a month.

This is remarkable, since in literature temperature measurements are commonly considered as unreliable and not timely enough to be used for predictive maintenance on bearings. This statement is recently anew confirmed by the extensive research of Vale, C. et al (2016). Different types of vibration measurements, on-board or wayside, are generally preferred as predictors for bearing failure. Hence, a failure detection method that gives reliable and early predictions based on temperature is an important step forward.

3. FAILURE DETECTION METHOD

Having understood the temperature behavior of the normal and failing bearings, we designed an automatic failure detection system. Because of the large variance in temperature behavior, some dynamic thresholding is needed to detect the first signs of a warming axle bearing adequately.

Our first attempt at a dynamic detection system was a neuro-fuzzy system with a self-constructing rule generation algorithm (Lee, W.J. et al, 2002). This algorithm clusters similar data points and detects deviating clusters automatically. However, the occurrence of failures proved to be too infrequent to determine the correct values for the required thresholds. Because the thresholds have no obvious physical meaning, their value is difficult to estimate. Therefore, instead a rule-based classifier was developed in the form of a decision tree. Decision tree classifiers are capable of breaking down a complex decision-making process into a collection of simpler decisions, thus providing a solution that is often easier to interpret (Racoul Safavian, S., & Landgrebe, D., 1991).

Before any decision rules are designed, the detectability of the elevated temperature of the failing bearing is firstly enlarged by calculating the difference between the absolute temperature of the bearing with the median of other bearing temperatures on the same side of the train. We call this temperature difference the *side-difference* (ΔT). The side-difference is illustrated for the case that was discussed previously, in Figure 4 and Figure 5.

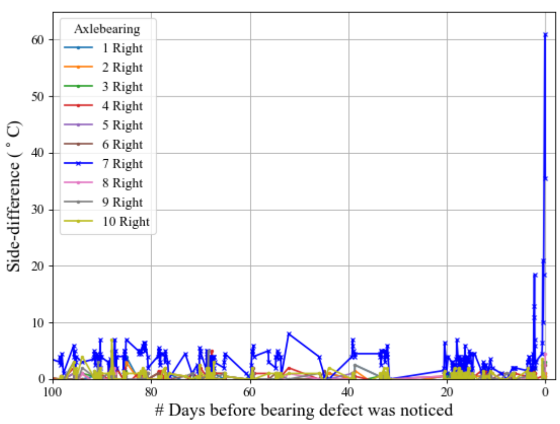

Figure 4: Side-difference for the bearings of a train with a failing bearing (7 Right, marked with a cross).

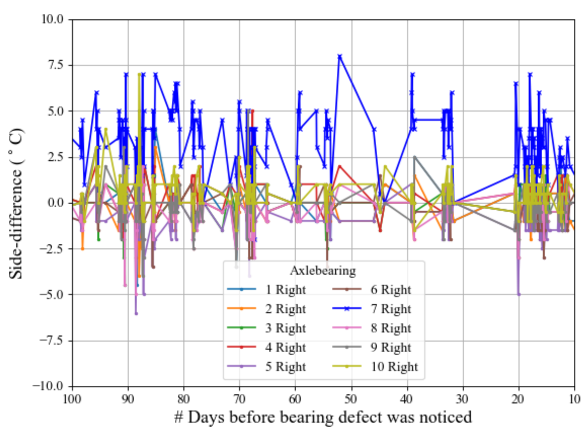

Figure 5: Zoomed version of figure 4, that highlights the large deviations in side-difference.

We have designed a set of rules that indicate the severity of the failure from low (alarm level 1) to high (alarm level 4). The rules are shown in a decision tree in Figure 7. The decision tree applies to slowly as well as rapidly increasing temperatures. Firstly the alarm setting for rapid increase in temperature and secondly the setting for slow increase in temperature is explained.

To detect rapid increase of temperature, we have set limits to the side-difference directly. Since the variance under normal behavior is relatively large for the reasons mentioned before, the first limit is set to a rather high side-difference of 30□C. Depending on the frequency of occurrence, exceeding this limit generates alarms of level 1, 2 or 3. When the side-difference raises to more than 50□C, the alarm level is immediately set to 3. Besides that, an absolute temperature of 80□C generates an alarm of level 4 directly. This set of rules detects the rapidly developing bearing failures adequately, without generating false alarms.

For the slow increase of temperature that is shown in the graphs, the rules described above only account for the end of the failure development. To detect the first part of failure development as well, extra rules are added to generate alarms of level 1 and 2. Since the limit of the side-difference cannot be lowered without generating false alarms, a different method for detection has been designed. The early failure development is characterized by the fact that the failing bearing has the highest temperature of all bearing temperatures in each measurement. The side-difference might only be a few degrees Celsius, but this bearing is consistently the hottest one within a measurement, over a period of several weeks. Therefore, for each bearing side-difference, the deviation with the median of all side-differences of the train is calculated. The median is used instead of the average to exclude the high side-difference of the failing bearing. If the side-difference of one bearing deviates more than 3.5 standard deviations from this median for at least 10 measurements within 30 days, alarm level 1 is generated for that bearing. A doubling of the frequency results in alarm level 2.

With this addition to the set of rules, both the early and the later stage of failure development are included in the decision tree. Furthermore, slow as well as fast developing failures are both detected as soon as possible. If one temperature measurement leads to more than one alarm level in the decision tree, the alarm with the highest level is generated.

The results of this set of rules on the case that is shown in the graphs in this paper, are displayed in Figure 6. The first alarm is generated more than 90 days before the bearing failure was noticed during the safety checks.

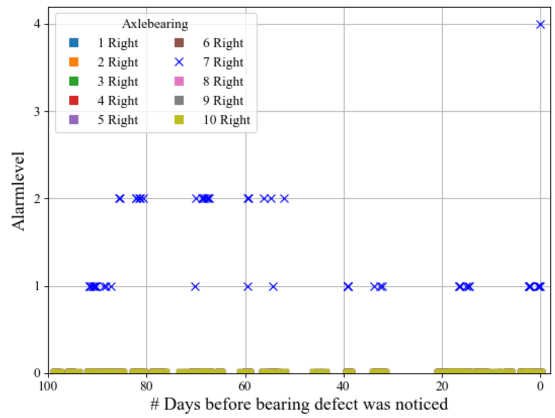

Figure 6: Alarms for a train with a failing bearing (7 Right)

4. DESIGN OF HANDLING SCENARIOS

For each alarm level a handling scenario has been designed in a way that minimizes the disturbance on the train service schedule and the unavailability of trains. In the PHM system's architectural framework that Kunche S., Chen, Ch., and Pecht, M. (2012) describe, the design of handling scenarios is included in the steps of Prognostics assessment and Advisory generation.

To simplify the handling of this rare-occurring failure, the handling scenarios are equal for each alarm level, except for the pre-warning level 1. The handling is as follows: take the train out of service within time *t*, replace the axle in the depot and send the train back on track. The bearings in the axle are investigated by bearing engineers and the outcomes are evaluated in order to monitor the general behavior of the bearings. Moreover, this feedback is essential for continuous validation and improvement of the developed decision tree.

The only difference between the handling scenarios is the time t at which the train is taken out of service to replace the axle. The alarm of level 1 would have occurred falsely three times in the last two years, so this alarm level is not yet operationalized. The possibility of only inspecting the axle at alarm level 1 (without replacing it) has been taken into consideration, but this alarm level occurs at such a timely state of degradation that visual or auditory inspection will not show anything. A test drive with accelerometers is also

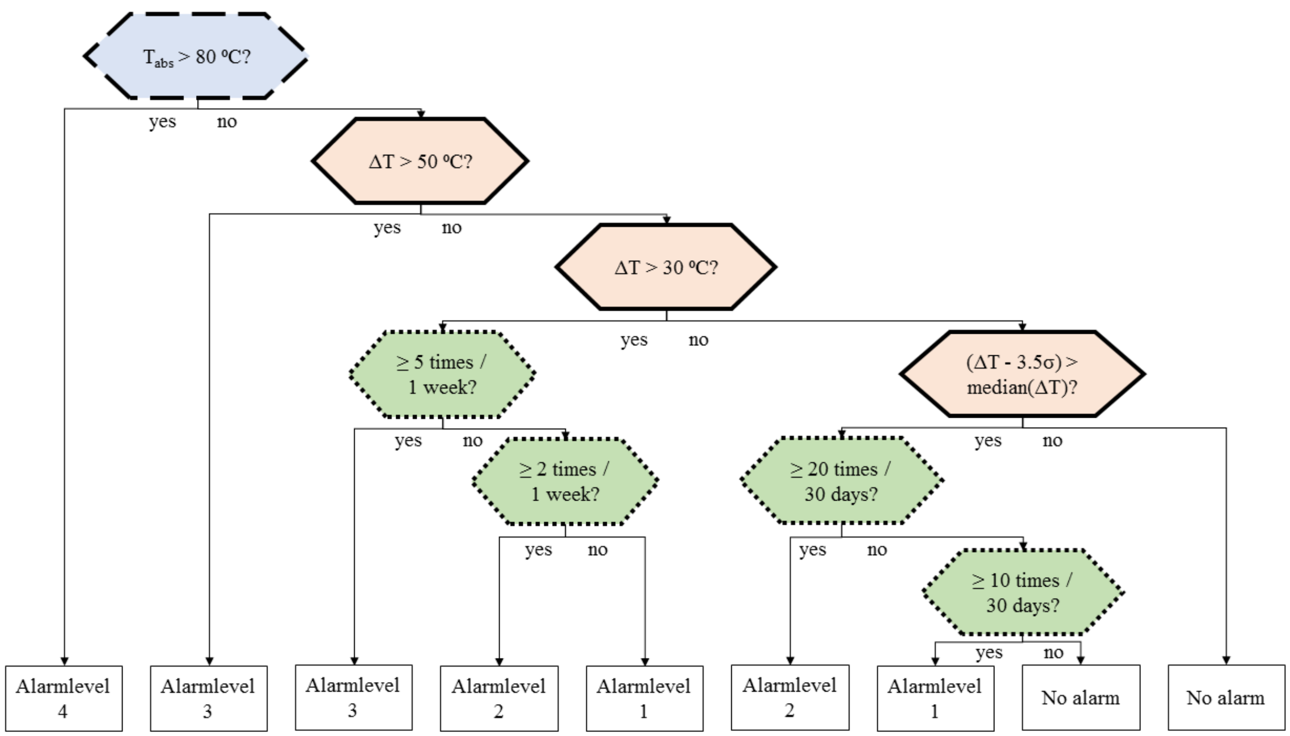

Figure 7: Decision tree of the rule-based classifier. The hexagons are nodes and the squares are the leaves of the tree, which show the outcome of the decisions. The top hexagon is the root. A blue hexagon (large dashed lines) concerns absolute temperature T_{abs}, a red hexagon (solid line) contains rules on the side-difference ΔT and the green hexagons (small dashed lines) contain rules regarding the frequency. When a measurement leads to more than one alarm level, the highest alarm is generated.

considered, but costs in terms of engineering capacity and unavailability of the train are high. The occurrence of failures is too low to optimize the decision tree for alarm level 1 in order to improve the performance. Therefore alarm level 1 is currently evaluated for further optimization, but not yet operationalized.

To optimize the benefits of this alarm system, the time between detection and functional failure should be used to minimize the operational impact of the occurrence of the failure. Within five days, the operational control center is always able to find a convenient timeslot to arrange the logistics of taking the train out of service without impact to the general service. Hence, at alarm level 2 the train is taken out of service at the most convenient time within five days. At alarm level 3, the train is taken out of service before the end of the day. At alarm level 4, the velocity of the train is restricted immediately and the train is taken out of service at the next intercity station.

The occurrence of lower alarm levels shortly after an alarm is generated does not influence the handling: the bearing could have cooled down in the meantime but could still be degrading. Of course a higher alarm does influence the handling by shortening time *t* to *t'* that suits the higher alarm level.

When four or more bearings of one train during one measurement register a temperature of more than 50°C, the cause is attributed to the measurement system itself and the alarms are suppressed. This phenomenon has unfortunately happened several times in the last year for some locations, and would generate false alarms if not suppressed.

5. RESULTS

Finally we investigate the reliability of our algorithm. The decision tree without alarm level 1 is applied to HotBox data of all 131 trains of the same train series over two years. This resulted in 60 million temperature measurements which yielded 7 cases of alarms. The results are listed in Table 1.

Table 1: Results of decision tree level 2 to 4

	Failure	**No failure**
Alarm	*Correct result: 6x* Highest level: 4, 3, 3, 2, 2, 2	*False positive: 1x* Highest level: 2
No alarm	*False negative: 5x*	*Correct result*

The results show that the decision tree enables us to predict more than half of the bearing failures. The first alarm occurred in every case one to three months before the current detection methods had found the defect.

Additionally, eleven cases of strange behavior of a measurement station were found, in which four or more bearings registered a temperature of more than 50°C. Furthermore two cases of bearing failure occurred in a period of time in which the concerned train was not driving past any HotBox measurement location for more than three weeks, so these bearings are not considered in the results. This point is further discussed in section 6.

Although the HotBox data for the false negatives is examined elaborately, the failure of these bearings was not detectable in the data, no matter what kind of rule or module would have been designed. Perhaps an increase in density of HotBox measurement stations would have helped, but it is likely that some specific failure mechanisms lead to such a rapid temperature increase that a system based on way-side temperature measurements will not be able to detect them early in the failure process. These failing bearings were discovered by mechanics or train drivers. It shows that this sensor system cannot replace human inspections yet.

6. DISCUSSION

The decision tree that has been developed in this work performs very well for the examined train series. However, the application of this method to other series requires some further research tailored specifically to that type of train. Preliminary investigations show that relatively more alarms occur while hardly any bearings have shown to be defect. Analysis of the data behind these alarms suggests that the alarms are not really false: the involved bearings do show a structurally elevated temperature for a longer period of time. The fact that this behavior was unnoticed in operation is not surprising, since the alarm level does not exceed level 2 and at some point in time the deviations always disappear. In most cases, this disappearance coincides with the general preventive replacement of the axle. This phenomenon is discovered recently and more research is needed to explain this behavior before the alarm system can be operationalized for these train series. Other causes of elevated temperature, such as failing gearboxes, cables or brakes, will be considered as well as the possibility of extremely slowly heating bearings.

Secondly, in order to improve the detection capability of the HotBox system, more measurement locations should be installed. It regularly occurs that trains are not registered by any measurement location for a week. The HotBox system was installed mainly for cargo trains and is thus mostly available on cargo tracks. The business case and location strategy for an extension of the system will be investigated. Moreover, the addition of other types of sensors, such as vibration, is considered in order to improve detection capability.

Lastly, bearing failures are currently registered more elaborately in order to be able to connect each failure mechanism to specific temperature behavior in the future. With these improvements, the remaining useful life could be estimated more accurately.

7. CONCLUSION

A decision tree based on way-side measurements of axle bearing temperatures has proven to be able to generate correct and timely warnings for bearing failures. Combined with handling scenarios that were designed with the operational process in mind, the implementation of this alarm system will lead to less impact on service in the case of bearing failure. However, human inspections are still required to detect all types of failure mechanisms in bearings. Further research will be done to extend the implementation towards other train series.

REFERENCES

Apallius de Vos, J.I., & van Dongen, L.A.M., (2015), Performance Centered Maintenance as a core policy in strategic maintenance control. *Proceedings of the Fourth International Conference on Through-life Engineering Services* doi:10.1016/j.procir.2015.08.016

Vale, C., Bonifácio, C., Seabra, J., Calçada, R., Mazzino, N., Elisa, M., Terribile, S., Anguita, D., Fumeo, E., Saborido, C., Vanhonacker, T., De Donder, E., Laeremans, M., Vermeulen, F., & Grimes, D. (2016) Novel efficient technologies in Europe for axle bearing condition monitoring – the MAXBE project. *Proceedings of the sixth Transport Research Arena.* April 18-21. doi:10.1016/j.trpro.2016.05.313

Dasgupta, A., & Pecht, M., (1991) Material Failure Mechanisms and Damage models. *IEEE Transactions on Reliability,* vol. 40, no. 5, pp. 531-536.

Rasoul Safavian, S., & Landgrebe, D., (1991) A Survey of Decision Tree Classifier Methodology. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 3, pp. 660-674.

Lee, W.J., Ouyang, Ch.Sh., & Lee, Sh.J., (2002) Constructing Neuro-Fuzzy Systems with TSK Fuzzy Rules and Hybrid SVD-Based Learning. *Proceedings of the IEEE International Conference on Fuzzy Systems,* Feb 2002

Kunche, S., Chen, Ch., & Pecht, M., (2012) A review of PHM system's architectural frameworks *MFPT 2012: Proceedings of the Prognostics and Health Management Solutions Conference,* April 24-26, Dayton, OH, 2012

BIOGRAPHY

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 Netherlands Railways for 5 years and is local chair of the European Conference of the PHM Society in the Netherlands in 2018.