Data-driven condition now- and forecasting of railway switches for improvement in the quality of railway transportation

Daniela Narezo Guzman[1], Edin Hadzic[2], Robert Schuil[2], Eric Baars[2], Jörn Christoffer Groos[1]

[1]*German Aerospace Center (DLR), Institute for Transportation Systems, Berlin, 12489, Germany*
Daniela.NarezoGuzman@dlr.de, Joern.Groos@dlr.de

[2]*Strukton Rail, 3600BA Maarssen (Utrecht), Netherlands*
Edin.Hadzic@strukton.com
Robert.Schuil@strukton.com
Eric.Baars@strukton.com

ABSTRACT

This contribution presents a data-based model that exploits the power consumed by point engines during blades movement of railway switches to detect relevant anomalies in switch behavior. The model incorporates local air temperature at the time of the measurement to account for the significant influence it has on normal switch behavior. The anomaly detection capability of the model is validated against alerts triggered by the state-of-the-art monitoring system POSS®, which is based on switch-specific and manually selected reference curves. The data-based model leads to less in number and more reliable alerts in comparison to the current version of POSS®. Especially false alerts caused by temperature effects are significantly reduced. Furthermore, the high sensitivity of the model proves to be capable of detecting emerging switch failures at an early stage of development. The detection capabilities of switch condition (nowcast) and identification of emerging failures at an early stage (required for failure forecast) proves that the model is useful for traffic interference prevention, condition-based predictive maintenance and switch health enhancement.

1. INTRODUCTION

Railway switches are crucial for guiding trains to tracks or platforms and allow them to take alternative routes in case of disruption. Switches are costly assets since the components and functions require frequent inspection, maintenance and renewal. The switch moving parts are subject to high deterioration and prone to malfunctioning, posing, in the worst case, a safety hazard if no action is taken. Nowadays online condition monitoring, inspection vehicles, standardization of both inspection and maintenance actions, as well as data-based models are tools supporting decision making for optimizing preventive and condition-based maintenance plans. These efforts shall lead to asset life extension, cost reduction and an overall improvement in the quality of railway transportation.

Automated switch status forecasting systems based on continuous switch current consumption (or other comparable measurements such as from a force sensor at the switch-blades) are not yet seen in 24/7 operation. (Camci et al., 2016) provides a comprehensive overview of existing efforts at research institutions and companies to develop forecasting models. The main challenge that such systems pose is the numerous failure types, which can occur simultaneously, and that are inherent to railway switches as complex electro-mechanical systems. Physical models show poor performance even under well controlled laboratory conditions with simulated failure development (Camci et al., 2016). Recent efforts have focused on developing data-driven models for monitoring the function of the switch and diagnosing failures e. g. by (Eker et al., 2010; Letot et al., 2015). (Böhm, 2017) applied different supervised classification techniques to predict the remaining useful time of switches. The main advantage of data-driven methods is that models with good apparent prediction performance can be derived from example data sets. The main remaining challenges are over-fitting, the creation of complete (containing all relevant switch failures types) training data sets with correct labelling, and the generalization of derived models for a large amount of switches.

Strukton Rail (SR) uses POSS®[1], a state-of-the-art system to monitor critical assets such as switch engines. Over 10,000 assets worldwide, most of them switches, are equipped with sensors and monitored by this system. POSS® measures the engine current (proportional to the engine power consumption (Stoll and Bollrath, 2002)) during the switch blades movement (see Figure 1). Switch malfunctioning, mostly of mechanical nature, can lead to irregularities in the power consumed during this movement. When these irregularities exceed certain thresholds defined by maintenance experts from manually selected reference curves[2], POSS® alerts are triggered (Dutschk et al., 2017). These alerts can indicate that the current state of the switch is different than expected (based on reference curves and thresholds derived from them). Moreover, weather conditions such as temperature and precipitation also play a role on the typical shape and the characteristics of such current measurements. Reference curves in POSS® are updated about every half a year in order to reduce the influence of seasonal temperature variation on switch condition monitoring and to prevent this from causing false alerts. Even for maintenance experts it is challenging to identify the source of detected irregularities or differences with respect to a selected reference curve, and to decide whether these are of concern or not (e.g. when a threshold is set too low). In case an alert is triggered in POSS®, maintenance experts assess weather conditions, switch history, threshold levels and the corresponding measured current, and decide whether the alarm is true or false as well as on the urgency of inspection.

Frequent manual selection of up-to-date reference curves for every switch and every direction of blade movement, in addition to the assessment of every POSS® alert, represents a significant work load for the condition monitoring operators. The selection of relatively large thresholds avoids false alerts but likewise hampers the early detection of degrading switch condition and emerging switch failures. The objective of the data-based model for anomaly detection presented here is to significantly reduce the work load by disposing the need for manual reference curve selection, while reducing the amount of false alerts and enhancing early detection, which is necessary for failure forecasting and to prevent complete switch failure.

In the following section the input data to the model, as well as POSS® output data used for validating the model are described. In section 3 the data-based model is discussed (see also (Böhm et al., 2016)). In section 4 the model output is validated against POSS® alerts (in what follows called just alerts) additionally assessed by a maintenance expert from seven switches and found to provide temperature-robust anomaly detection. It is also shown that the model is capable of detecting evolving failures, which can ultimately lead to failure forecast. In section 5 the validation results as well as future efforts to develop automated methods for providing more reliable diagnostic and prognostic information to support condition-based and predictive maintenance are discussed.

2. DATA SET

The data considered in this paper consists of current curves measured with a frequency of 50 Hz at seven switches for blades movement in direction 1 only. The current curves were acquired between January 2012 and February 2017. The air temperature at the relay house (located between 30 m and 2.5 km away from the switches) at the time each current curve was measured is available. Table 1 contains an overview of the available data for each switch identified with an ID-number.

In its output, POSS® provides a status description for each measured current curve. The large majority of curves have an "okay" status. However when one or more of the set thresholds based on the reference curve are exceeded the status is different than "okay" (i.e. an alert is triggered). In the data considered here, all alerts belong to the most common four alert types generated in POSS®. Two are based on the total power consumed by the engine and can lead to "power too high" or "power too low" status. Another type is related to the total duration of the current curves and leads to "time too long" or "time too short" type of alert. Some curves may lead to both "time too long" and "power too high" alerts, since these quantities are correlated. However the status only provides one alert type. When the measured current reaches high values and exceeds corresponding thresholds a "current too high" alert is triggered. Table 1 shows for each switch the number of total alerts triggered in the considered time period, as well as the number of alerts for which the reference curve corresponding temperature is available.

A SR maintenance expert assessed each single alert a-posteriori. The goal of this assessment was to categorize all alerts into false, true or undefined according to domain knowledge. For this assessment the maintenance expert considered the following: the validity of the reference curve (i.e. the temperature difference with respect to the curve that triggered the alert), the set threshold and by how much it was exceeded, the current curve shape, and the switch history. Alerts classified as undefined correspond to cases in which e.g. a current curve presented a small deviation from the expected shape but the current curve of the next switchblades movement looked completely normal.

---

[1] In this paper the most common version of POSS® is used as reference. During 2018 a new POSS® release will become available with improved functionality to manage thresholds.

[2] For a few switches the summer/winter reference curve is automatically selected depending if the temperature at the relay house at the time a current curve is acquired is larger/smaller than a manually selected critical temperature. This was the case for switch 2604, which had a critical temperature of 5°C.

Switch ID	Total no. of c. c.	No. of c. c. in Tr. set	Tr. set length in months	No. of alerts with available τ_{ref} (alerts in total)	Distance to relay house in meters
2604	5543	1041	12	42 (42)	1015
2606	14326	1766	12	12 (76)	910
3015	2535	444	12	38 (38)	2410
3069	9617	1810	12	24 (24)	282
3076	10653	1801	12	73 (189)	175
3083	10213	2239	12	125 (125)	30
3090	4968	729	9	39 (39)	770

Table 1. Overview of data set considered in the analysis. Training set abbreviated as Tr. set, current curves as c. c.

Furthermore historical data sets of all reported failures (or technical malfunctions) are available. This data set contains time and date when failures were reported by the network operator as well as when the switch was made available again. It also contains additional remarks provided by the maintenance team on site. These remarks are not stored in a systematic way and descriptions may strongly vary from operator to operator. The data set of reported failures can be incomplete due to e.g. malfunctions that were fixed without reporting or were only temporary. It is not unusual that malfunctions are not reported by the railway operators if the switch moves finally to a safe end position after several retries. Failures are represented in Figure 2a, Figure 3a and Figure 6 by vertical dashed bars; their duration is proportional to the bar thickness (which is only noticeable in Figure 6).

Planned maintenance activities related to each maintenance campaign, as well as planned execution dates (actual execution of maintenance might differ from the planned timetable) are registered in a separate data set for each switch. Details of the maintenance performed are not included in this data set, i.e. it is not possible to know exactly which switch parts were subjected to which specific actions. The duration of maintenance activities is unknown. For the sake of visualization it is assumed here that maintenance actions spanned 24 hours and they are included as vertical solid bars in Figure 6 (also in Figure 2a and Figure 3a but given the wide time range shown in them these lines can barely be seen). Note that failures and maintenance information are used as additional information for interpreting the model results (as in Figure 6) and not included in the data-based model.

3. DATA-BASED MODEL

Supervised learning strategies require high quality training data sets for their success. As described in section 2, such data sets are not (yet) available for switch condition since relevant influences are unknown (e.g. influence of a specific type of maintenance action, weather variables, etc.) and corresponding data is missing or stored in a non-systematic way. Therefore a data-driven approach based on current curves and air temperature simultaneously measured is presented in this paper. One assumption is made; that the training set used to build the model represents switch normal behavior. This approach is work in progress and the results preliminary.

Every switch behaves in a unique way as the switchblades are unlocked, then moved from the start to the end position, and finally locked in their set end position (see Figure 1). These phases of the switch movement leave a typical (but not identical among different switches) trace on the measured current at the engine. Therefore the model is necessarily switch and direction-specific.

3.1. Selection of training set

First the model is trained with features extracted from a selection of current curves (so-called training set) that are assumed to predominantly represent normal switch behavior. Then the trained model is applied to the same features extracted from other current curves (from the same switch and in the same direction).
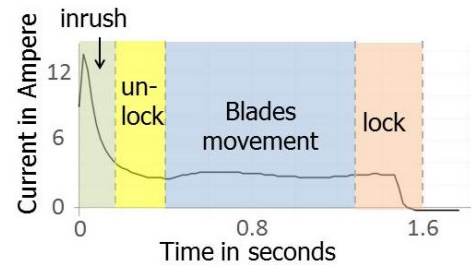

Figure 1. Current measured at the engine during all phases (each separated by a vertical dashed line: current inrush, blades unlock, blades movement and blades lock in end position) involved in the switchblades movement.

The selection of the training set is a non-trivial aspect of the model development, as it defines the model output and is the base for detecting abnormal switch behavior. Different approaches for selecting samples representing normal operation are possible. The method applied here consists of identifying beforehand a timeframe in which it is assumed that the switch predominantly functioned normally (e.g. the time between a pair of consecutive reported failures). This is possible since information on historical reported failures is available. Current curves measured in this timeframe are analyzed in order to remove the ones that are statistical outliers from the training set based on two criteria: total

duration of switch blades movement and area under the curve. The number of current curves in the training set depends on switch usage and the time-window chosen for training (typically one year), see Table 1.

3.2. Feature extraction

Features are derived from each current curve and defined such that they represent the switch behavior. Feature selection derives from data science, asset domain knowledge and explorative data analysis (see (Dutschk et al., 2017)). Here we consider a subset of the features identified in (Dutschk et al., 2017): 1) area under the curve, 2) maximum, 3) median, 4) kurtosis, 5) skewness, 6) duration, 7) mean value during blades movement, and 8) standard deviation during blades movement.

The model input consists of features from training set curves, including the temperature measured at the relay house at the time current curves were acquired. The switch behavior temperature dependence is reflected in the features. For example the area under the curve (or total power consumed by the engine) and the total duration of the curve systematically decrease with increasing temperature, up to a certain limit (Böhm and Doegen, 2010). Each feature is scaled to have zero mean and standard deviation equal to one; see (Kuhn and Johnson, 2016). This transformation is separately applied to feature values from current curves, which were measured at a temperature within the same one Kelvin bin. With this scaling the temperature dependence is removed from the features that are the input to the model. In what follows we refer to the scaled features as features.

3.3. Model training

The model is built by taking the training set features as input and applying the Principal Component Analysis (PCA) to them (Jackson and Mudholkar, 1979; Sotiris and Pecht, 2017). Because PCA is sensitive to feature ranges it requires their previous normalization or scaling. PCA consists of finding a basis defined by orthonormal vectors (i.e. Principal Components or PCs) that minimize redundancy among the training set features, while maximizing their variance. The dimension of this new basis is determined by the amount of variance the PCs are chosen to retain; in our case the retained variance is 90% and the dimension of the basis is between 2 and 4, depending on the switch. The PCs form the model subspace. The orthonormal vectors that are not retained form the residuals subspace. At this point the model is said to be trained or built. Now, a point in the features space (i.e. features extracted from a single current curve) is projected into both the model and the residual subspaces. Its squared Euclidean distance to the origin in the model subspace projection is the Hotelling's parameter (T^2), and the Euclidean distance to the origin in the residual subspace projection equals the Square Prediction Error (SPE). Therefore for each current curve the model output consists of two parameters (T^2 and SPE) (Böhm et al., 2016).

3.4. Range of normal switch behavior

For each of the model output parameters we obtain the 90% quantile (defines the probability that the parameter takes a value less than or equal to 90%) of the training data set distribution. The 90% quantile $q_{0.9}(T^2)$ is defined with relatively high accuracy given the relatively high density of data points in the probability distribution. This value is then scaled by 1.2 and used to define the range of normal switch behavior: $[0, 1.2 \cdot q_{0.9}(T^2)]$. The factor 1.2 is somewhat arbitrary and represents a first approach to define a threshold for anomaly detection based on a statistical quantity defined with relatively high accuracy (90% quantile). Note that choosing alternatively e.g. the 99% quantile is less accurate given the low density of points in the right tail of the T^2 distribution. The T^2 value from the i-th current curve T_i^2 that is not part of the training set is evaluated and identified to represent *significant* abnormal switch behavior if it fulfills:

$$T_i^2 \gg T^2_{thre} \coloneqq 1.2 \cdot q_{0.9}(T^2)$$

Due to the somewhat arbitrary definition of the upper threshold T^2_{thre}, T_i^2 values close to it cannot be strictly associated to normal or abnormal switch behavior. Therefore T_i^2 values slightly larger than T^2_{thre} are referred here to as *mild* anomalies. A similar situation is encountered by maintenance experts when they need to decide whether a current curve that triggered an alert is mildly or significantly different from expected normal switch behavior. In the former case they will probably not react to it, while in the latter they will.

The lower threshold of T^2 related to normal switch behavior equals zero since the features are normalized and centered (previous to applying the PCA), and thus distributed around zero. In consequence, small values of T^2 and SPE correspond to current curves with feature values close to the feature mean values as obtained from the training set. These mean values represent normal switch behavior by definition and selection of the training set. We note that the training set features (after the PCA transformation) are not necessarily standard distributed. If, however, that was the case the T^2 parameter would be chi-squared distributed with a degree of freedom equal to the number of PCs (Sotiris and Pecht, 2017).

So far it is assumed that normal behavior of a switch does not change in time. That is, for most switches the model is trained with current curves measured in one year and applied to current curves measured in four other years. However this assumption might be violated when e.g. a high-impact maintenance action is performed on the switch, modifying the switch behavior under normal operation and thus the range of normal behavior (T^2_{thre} is based on the

training set distribution). The topic of the constancy of normal switch behavior is an important aspect of the method proposed here but is out of scope in this paper.

4. RESULTS

Figure 2a and Figure 3a show each the Hotelling's parameter (in log scale since T^2 values cover up to 5 orders of magnitude) over time for two different switches. Data points larger than T^2_{thre} of the training set are outside the range of assumed normal behavior, thus called mild/significant anomalies here. Switch 3083 (Figure 2) experienced many more switchblades movements than switch 2604 (Figure 3). The training set, spanning a whole year in both cases, consists of 2239 and 1041 current curves (see Table 1), respectively.

In Figure 2a most anomalies detected by the model coincide with two types of alerts. On the other hand, only a few "power too high" alerts are not strictly detected as anomalies but are close to T^2_{thre}. Some alerts and anomalies occurred briefly before a failure was reported, likely indicating the compromised functionality of the switch. In spite of the fact that the seasonal temperature variation (see Figure 2b) is compensated through the scaled features (see section 3), T^2 values tend to be slightly smaller in the winters than in the summers, except for years with reported failures in the winter (2015 and 2017). This indicates that the temperature influence on the features is not fully accounted for in the normalization, which is not surprising given that the air temperature is only a proxy of the asset actual temperature.

In Figure 3a all alerts, which incurred in the cold months (see Figure 3b showing the temperature at the time of blades movements), are within the range of normal behavior. In fact, none of these alerts seem to have been crucial: there is no failure (indicated by vertical dashed-lines) reported after they were triggered, except for one in 2015 (a loose clamp/bolt between the blade and the stock rail was reported and fixed, eventually preventing the complete failure) and its shape was identified as normal by experienced POSS® operators (see Figure 3c and

Figure 4). T^2 values for this switch do not show a particular correlation with cold/warm months.

Figure 2c and Figure 3c present the difference between $\log(T^2)$ values of alerts and $\log(T^2_{thre})$. This quantity is plotted as a function of the difference between the temperature when the alert-triggering curve was measured (τ) and the one of the reference curve associated to that alert (τ_{ref}). Thus, according to the model and the detection rule chosen (see section 3.4), small/large positive $\log(T^2 / T^2_{thre})$ values correspond to mild/significant anomalies and negative (enough) ones to normal switch behavior.

Adding on to Figure 2c and Figure 3c, Figure 5 shows all alerts from seven switches (see Table 1) provided the corresponding τ_{ref} is available (for some years and switches it is not). It is found that temperatures at the time of switch movements were up to 20 Kelvin larger or smaller than τ_{ref} used in POSS® to detect alerts.
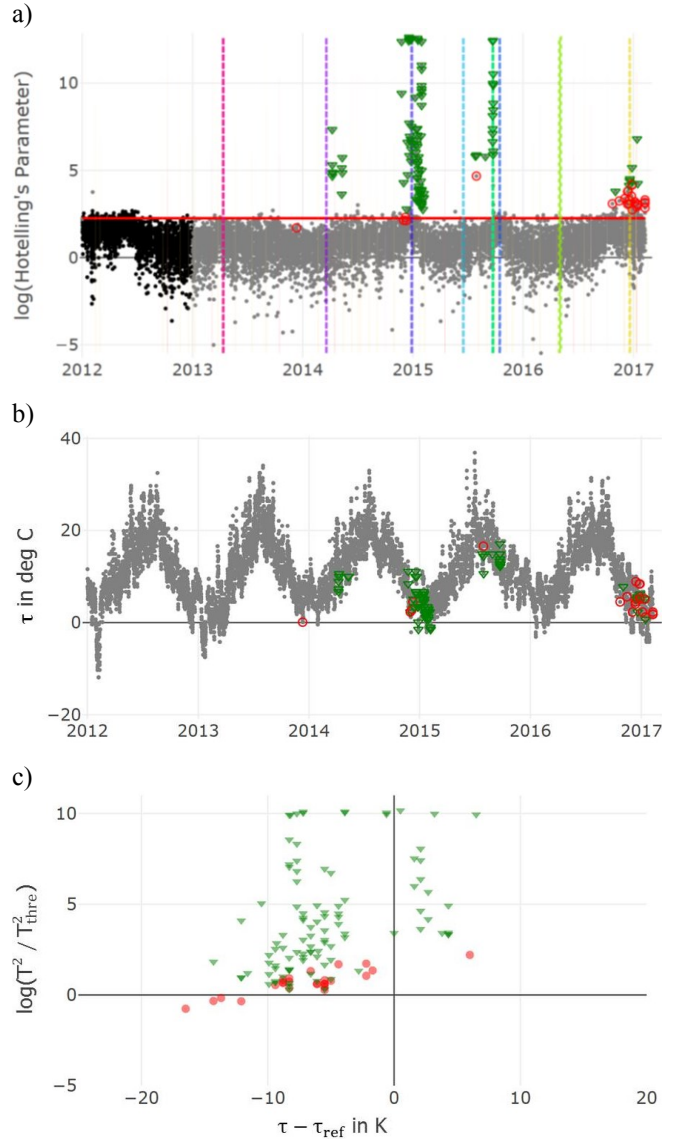

Figure 2. Switch 3083 in direction 1. POSS® alerts: "time too long" (triangles) and "power too high" (circles). a) Logarithm of T^2 as a function of time. Current curves in training set and outside it: black and grey points, respectively. Horizontal (red) solid line: T^2_{thre}. Dashed vertical bars: reported failures. Solid vertical bars (barely noticeable): maintenance. b) Temperature at the time of current measurement as a function of time. c) Logarithm of T^2 from alerts divided by T^2_{thre} as a function of the difference between temperature at the time of movement (τ) and of reference curve (τ_{ref}). Maintenance expert assessment

of alerts: true (solid symbols), undefined (open symbols), false (symbols with cross or dot inside).

The maintenance expert found that 75% of 346 analyzed alerts were true, 23% were false and 2% undefined. The assessment results (see Table 2) are displayed through crossed or dotted, solid, and open symbols for false, true and undefined alerts, respectively (see Figure 2c, Figure 3c and Figure 5). The results for switches 3083 and 2604 exemplify two extreme cases. In the first case 121 out of 125 alerts detected by POSS® (all classified as true alerts by the expert) coincide with anomalies detected by the model. In the second case none of the 42 alerts (all classified as false by the expert) are detected as anomalies, see Figure 4.

Switch 2604 provides evidence that the state-of-the-art system raises false alerts if a current curve is measured at a temperature that differs significantly from τ_{ref}. The ultimate goal is to develop a reliable monitoring system that does not depend on manual selection and assessment due to the significant workload this represents, and which raises only true alerts automatically. To gain insight into the model performance to detect current curves that triggered POSS® alerts, we consider the model results and argue about the validity of the alerts in view of the difference in temperature at which they were triggered with respect to τ_{ref}.

As previously mentioned, current curves are influenced by temperature: the total power consumed by the engine, the maximal current value during switchblades movement and the total current curve duration, are quantities that decrease with increasing temperature until they reach their temperature-independent nominal value. For current curves under normal operation measured at a temperature $\tau < \tau_{ref}$, these quantities are necessarily larger than the corresponding values of the reference curve. Thus under such circumstances, even if the switch is behaving normally, the thresholds (derived from the reference curve) can be exceeded, triggering "power too high", "current too high" and "time too long" false alerts in POSS®. In this context, a negative enough $\tau - \tau_{ref}$ can lead to alerts purely caused by a non-valid reference curve. However not all "power too high", "current too high" and "time too long" alerts with $\tau - \tau_{ref} < 0$ are necessarily false; they too can point to compromised switch functionality, as confirmed by the expert findings (see left quarters in Figure 5).

To quantitatively differentiate between a true and a false "power too high", "current too high" or "time too long" alert with $\tau - \tau_{ref} < 0$, one would have to consider by how much the threshold set in POSS® is exceeded. Moreover according to maintenance experts, the threshold value for a given switch is not necessarily constant (e.g. might vary from year to year); nevertheless there is no recording available of the thresholds. For current curves that triggered a "power too high", "current too high" or "time too long"

alert and that fulfill $\tau - \tau_{ref} < 0$, we can state that it is more likely that the alert is false the larger the difference between τ and τ_{ref}. Based on analogous arguments, a "power too low" alert triggered when $\tau < \tau_{ref}$ points out to compromised switch functionality and is considered a true alert, coinciding with the expert's assessment (see square in upper left quarter of Figure 5).
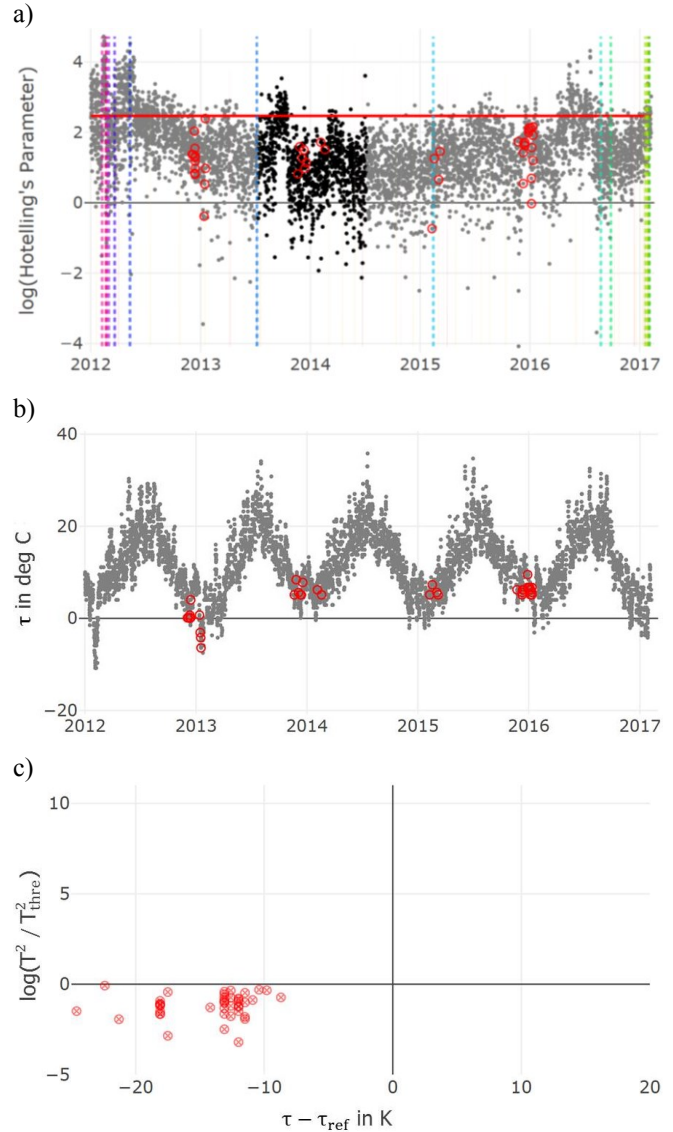

Figure 3. Switch 2604, direction 1. See Figure 2 caption.

On the other hand, for cases where $\tau > \tau_{ref}$, "power too high", "time too long" and "current too high" alerts are to be treated as serious warnings or true alerts (as corroborated by the expert assessment - see upper right quarter of Figure 5), while "power too low" alerts are likely false.
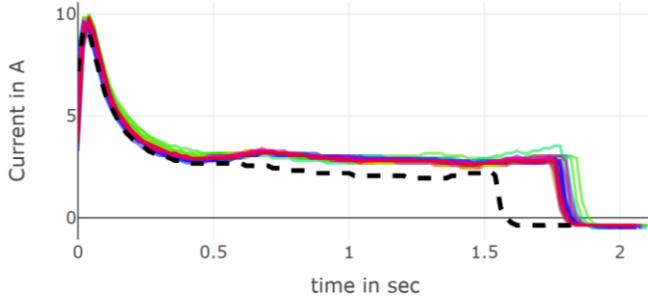
Figure 4. All current curves of switch 2604 in direction 1 that triggered "power too high" alerts in POSS® (solid curves) and corresponding reference curve (dashed). These curves show no abnormal behavior but triggered false alerts due to $\tau - \tau_{ref} < -10$ K.

Switch ID	Number of alerts with available temperature	Number of true alerts	Number of false alerts	Number of undefined alerts
2604	42	0	42	0
2606	12	12	0	0
3015	38	3	34	1
3069	24	125	0	0
3076	73	73	0	0
3083	125	23	0	1
3090	39	24	3	5
Total number (%)	346 (100%)	260 (75%)	79 (23%)	7 (2%)

Table 2. Results of POSS® alerts assessment by maintenance expert.

The argumentation in the previous two paragraphs is applied to the POSS® alerts and, together with the expert's assessment, used to verify the data-based model results. "Power too high", "time too long" and "current too high" alerts found for $\tau > \tau_{ref}$ (top right quadrant in Figure 5) as well as the "power too low" alert with $\tau < \tau_{ref}$ (square in top left quadrant) are detected by the model as anomalous and, based on previous argumentation, considered to be true alerts, in full agreement with the expert's assessment. "Power too high", "time too long" and "current too high" alerts found for $\tau < \tau_{ref}$ and detected as anomalies by the model (top left quadrant) were mostly assessed as true alerts by the expert in spite of the fact that the reference curves used to trigger alerts were doubtfully valid. In total the model identified 270 alerts as anomalies, out of which 253 (or 94%) were assessed by the expert as true alerts, 15 (5%) as false and 2 (less than 1%) as undefined. From the 76 "Power too high", "time too long" and "current too high" alerts with $\tau < \tau_{ref}$ detected by the model as normal (bottom left quadrant), 64 (or 84%) were identified as false, 6 as uncertain (8%) and only 6 (8%) as true according to the expert assessment. These 6 alerts with contradictory

findings (true alert and detected as normal by the model) are however very close to T^2_{thre} (which is not to be considered a strict division between normal and abnormal, as argued in section 3.4). Based on the temperature argumentation, these alerts are more likely to be false the larger $\tau - \tau_{ref}$ is which is in accordance with the findings in Figure 5, where less and less true alerts and anomalies are found the larger $\tau - \tau_{ref}$ becomes.
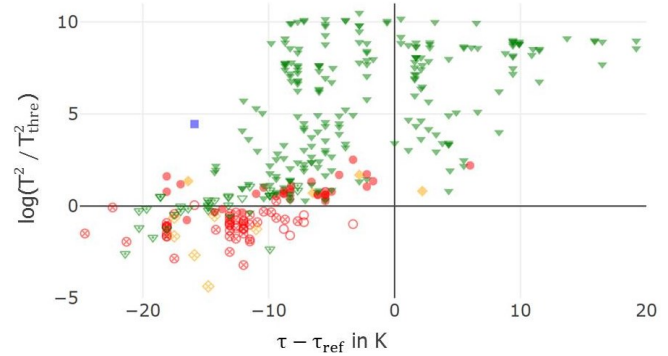

Figure 5. See caption of Figure 2c. POSS® alerts of seven switches in direction 1: "Time too long" (triangles), "power too high" (circles), "current too high" (diamonds), "power too low" (squares). Maintenance expert assessment of alerts: true (solid symbols), undefined (open symbols), false (symbols with cross or dot inside).

The model potential to identify systematic abnormal behavior on an early stage of emerging failures is exemplified in Figure 6. Starting around December 15th 2013 a systematic increase in the T^2 parameter is detected (see data points inside the drawn ellipse). The subsequent failure reported on December 26th was originated by a rusting gear box. The increasing and sustained abnormal behavior of the switch is a premise for failure forecast.
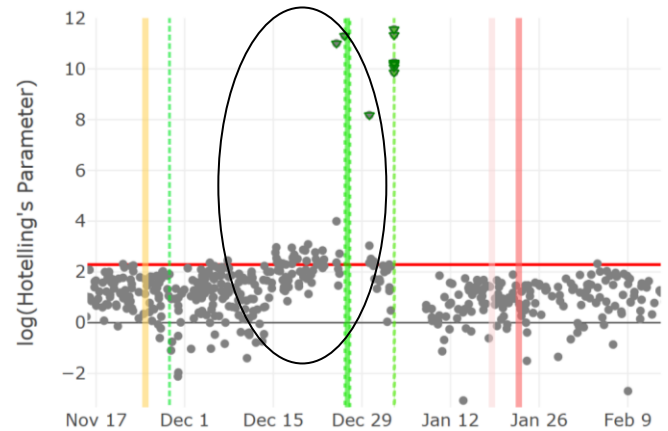

Figure 6. Switch 3076 in direction 1. Logarithm of T^2 as a function of time (Nov. 2013 – Feb. 2014). Horizontal (red) solid line: T^2_{thre}. Vertical bars: reported incidents (dashed) and maintenance (solid); the different colors have no

meaning. Ellipse highlights data points with systematic increase.

5. DISCUSSION AND OUTLOOK

The four most common alert types triggered by (the current release of) POSS® can be differentiated between false and true (indicative of switch anomalies) based on their $\tau - \tau_{ref}$ value and alert type. This argumentation is in line with the maintenance expert findings about the alerts. "Power too low", "time too long" and "power too high" alerts in the upper-left quadrant, as well as "current too high" alerts in the upper right quadrant of Figure 5 certainly reflect compromised switch functionality and a real problem. The switch failure detection model detected them all as anomalies. Furthermore true and false alerts classified according to an expert's knowledge and qualitative arguments showed a good agreement with the model results: of all true alerts 94% were detected as abnormal switchblades movements and of all false alerts 84% were detected as normal ones by the model. In spite of the assumptions regarding normal behavior, the model is found to be temperature robust and capable of detecting the majority of alerts without the need to manually select reference curves and corresponding thresholds for each switch and in each direction.

Furthermore the model is designed in such a way that it can detect anomalies that are not necessarily reflected in deviations from expected total power (i.e. area under the curve), and that would be missed by state-of-the-art condition monitoring systems even if the reference curve is valid for the measurement. For example, if one considers a sinusoidal curve, its area under the curve over one period is equal to zero. If a current curve would show fluctuations described by a sinusoidal curve, this abnormal behavior would not be reflected in the total power and thus no alert would be triggered in systems like POSS®. However since the model considers current standard deviation during switchblades movement, this example current curve would in principle be detected by the model as abnormal.

The pattern recognized for switch 3083 in Figure 2a for summer and winter T^2 values could be due to differences between the air temperature at the relay house and the asset temperature, and also due to seasonal variations of weather variables other than temperature. For example precipitation evaporates faster in a warm sunny day than in a cold and cloudy one. Thus even when the model is temperature robust, this does not imply that other factors, which are temperature-correlated and that affect the switch behavior, are being accounted for. Additionally the model is trained with one-year data and applied over the next 4 years. The way a switch reacts to temperature (and other weather conditions) or load might change with time. Additionally maintenance actions performed on the switch can modify the normal relation between features (e.g. maintenance

actions can cause step-changes in the median value of one or more features). Changes on the switch functioning induced by time or maintenance imply a modified normal behavior, and require re-training the model in order to keep it up-to-date. Model accuracy depends on its range of validity/applicability, which is a topic of major importance. Current research is dedicated to automatically identify this validity range considering the factors of influence.

The method of training set selection requires more sophisticated methods in order to train the model with current curves of not only normal, but representative of well-functioning switch behavior. The current selection method has no way of differentiating between a functional switch with abnormal behavior from a functional switch with normal/healthy behavior. A good alternative for training set selection is clustering; with this approach one could e.g. consider the current curves belonging to the cluster, which according to maintenance experts shows the most normal /functional behavior.

Results for switch 2604 (Figure 3a) are a good example to show the importance of normal switch behavior and its impact on the model output. In the training period T^2 values have large deviations and there is much structure in the data points contained in it. This is also reflected in the T^2-distribution of the training set and thus on the thresholds of normal switch behavior, ultimately affecting anomaly detection and leading to the detection of too many anomalies. The way to overcome this issue is by finding more and more adequate features representing switch behavior. Ideally, all parameters influencing the system or common causes of variation are accounted for in the model, such that normal switch behavior (represented by the training set) is a stable process. In this ideal situation the training set output parameters should not present structure.

T^2 and SPE values indicating an anomaly are of concern but further investigation is needed to categorize them and provide degrees of abnormal behavior and criticality of the anomaly. Moreover, for condition-based predictive maintenance support, detection anomaly needs to be accompanied by a diagnosis. The link between a switch functional model, which relates switch sub-functions (see Figure 1) to switch components, and the data-based model output are the features. Domain knowledge will further provide features that are directly linked to switch components. In this way, when an anomaly is detected, the features can be traced back to identify the components that are compromised.

6. CONCLUSION

The data-based switch failure detection model is verified against POSS® alerts from seven switches over more than five years; it identifies true alerts triggered by abnormal switch behavior as anomalies. The model does not rely on

manual reference and threshold selection, while it produces reliable detections. The implementation of the model for anomaly detection could improve the reliability of switch condition monitoring systems, such as POSS® current released version. Furthermore the model detects evolving abnormal behavior, setting the path towards failure forecast.

Further research on feature engineering is necessary to enable more accurate modelling of switch behavior. This will increase model accuracy and reliability when anomalies are detected, and provide diagnostic information for more efficient switch interventions. Additionally, other weather variables and actions (e. g. scheduled preventive maintenance) performed on the switches which modify their normal behavior need to be accounted for.

ACKNOWLEDGEMENTS

NOMENCLATURE

T^2	Hotelling's parameter
T^2_{thre}	Threshold of normal switch behavior T^2 range
SPE	Squared prediction error
τ	Temperature at the time current curve is measured
τ_{ref}	Temperature at the time of reference curve

REFERENCES

Böhm, T.: Remaining useful life prediction for railway switch engines using classification techniques, International Journal of Prognostics and Health Management, 8, 2017.

Böhm, T. and Doegen, C.: Diagnosis without Sensors - Integration of External Data for Condition Monitoring of Railway Switches, in: COMADEM 2010 - Advances in maintenance and condition diagnosis technologies towards sustainable society, Okumura, S., Kawai, T., Chen, P., and Rao, R. (Eds.), Condition Monitoring and Diagnostic Engineering Management, Sunrise Publishing, Japan, Hikone, 2010.

Böhm, T., Schenkendorf, R., and Lemmer, K.: Statistical Process Control for Modern Switch Failure Detection, in: Proceedings of the 11th World Congress on Railway Research, 11th World Congress on Railway Research, Mailand, 29.05.-02.06.2016, 2016.

Camci, F., Eker, O. F., Baskan, S., and Konur, S.: Comparison of sensors and methodologies for effective prognostics on railway turnout systems, Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 230, 24–42, doi:10.1177/0954409714525145, 2016.

Dutschk, B., Groos, J. C., Hadzic, E., and Böhm, T.: Deeper insight in railway switch condition nowcasting, in: Proceedings of WCCM 2017 congress, The British Institute of Non-Destructive Testing (Ed.), World Congress on Condition Monitoring, London, 13.-16.06.2017, 2017.

Eker, O. F., Camci, F., and Kumar, U.: Failure Diagnostics on Railway Turnout Systems Using Support Vector Machines, in: EMaintenance 2010: Proceedings of the 1st International Congress on eMaintenance Lule°a, Sweden, June 22-24, Kumar, U. (Ed.), Univ, Lulea, 248–251, 2010.

Jackson, J. E. and Mudholkar, G. S.: Control Procedures for Residuals Associated with Principal Component Analysis, Technometrics, 21, 341, doi:10.2307/1267757, 1979.

Kuhn, M., & Johnson, K., 2016. *Applied Predictive Modeling*. New York: Springer Science+Business Media.

Letot, C., Pierre, D., Pugnaloni, M., Dehombreux, P., Guillaume, F., Cyril, D., and Piero, L.-C.: A data driven degradation-based model for the maintenance of turnouts: a case study, 48, 2015.

Sotiris, V. A. and Pecht, M.: Support Vector Prognostics Analysis of Electronic Products and Systems, AAAI, Artificial Intelligence for Prognostics Fall Symposia, Arlington, 9.11.2017, 2017.

Stoll, H. and Bollrath, B.: Weichendiagnosesystem SIDIS W, last access: 1 June 2016.

BIOGRAPHIES

Daniela Narezo Guzman received her physics Licenciatura degree from the Universidad Nacional Autónoma de México in 2009. In 2011 she obtained her M.Sc. degree in applied physics from the University of Twente (UT). For her doctorate she joined a collaboration project between the UT and the Dept. of Physics from the University of California in Santa Barbara (USCB), USA. In 2015 she received her Ph.D. title from the UT for her experimental work done at the UCSB. In 2016 she joined the EXPACT project (impacts of extreme weather events) at the Potsdam Institute for Climate Impact Research, in Germany. In July 2017 she started working as data scientist in the Asset and System Monitoring group at the Institute of Transportation Systems from the German Aerospace Center, in Germany.

Edin Hadžić was born in 1974. In 2002 he received his Bachelor of Engineering degree from the department of Electrical Engineering and Computer Sciences, Rotterdam University of Applied Sciences in the Netherlands. He started his career in 1997 working for an industrial engineering company in Rotterdam. In 2003 he joined

Strukton Rail as Maintenance & Reliability engineer working on improvements of maintenance concepts and processes for railway infrastructure. For this challenge he has performed a variety of different activities. Lately focusing on data and currently leading several projects on using machine learning tools and making data driven models applicable and usable in real life operations.

Robert Schuil is reliability engineer and responsible for the optimization of maintenance concepts, designing new inspection methods using the latest techniques and setting up risk analyses this with a good balance between risks and cost-effectiveness. Known experience in FMECA, RAMS, RCM and RCA. Robert was born in 1982 in Leeuwarden, the Netherlands. In 2008 he completed the Bachelor of Business Studies and Engineering at the NHL University of Applied Sciences Leeuwarden. 2017 was the year in which he completed maintenance management at the HU University of Applied Sciences Utrecht.

Eric Baars is reliability engineer railway specialized in signaling and monitoring. Eric was born in Amsterdam, the Netherlands, in 1965. In 1987 he completed the vocational training for maritime officer in Haarlem. In 2015 Eric received his MEng degree for maintenance and asset management from the University of Applied Science in Utrecht. He started his professional carrier in 1987 as a cost engineer and draftsman for a small contractor for electrical, HV/AC, pluming and roofing installations. In 2002 Eric started to work for Strukton Rail as general foreman signaling. A few years later he became maintenance engineer. From the beginning Eric was involved in the development of the POSS wayside monitoring of Strukton Rail and currently he is the product architect for POSS detections. Eric is professional competences assessor for the railAlert, a non-profit organization to ensure that no fatal accidents occur in the Netherlands by working on the rail infrastructure.

Jörn Christoffer Groos received his diploma in geophysics from the University of Karlsruhe, Germany, in 2007 and his Dr. rer. nat. from the faculty of physics of the Karlsruhe Institute of Technology in 2010. Since 2014 he is with the German Aerospace Center DLR at the Institute of Transportation Systems in Braunschweig, Germany. Jörn is leading the Asset and System Monitoring group within the Department Data Management and Knowledge Discovery. His current research interests include data acquisition, data analysis, and machine learning for condition monitoring of railway assets.