# Accuracy Improvement of Condition Diagnosis of Railway Switches via External Data Integration

T. BÖHM

**ABSTRACT**

A highly available infrastructure is a premise for capable railway operation of high quality. Therefore maintenance is necessary to keep railway infrastructure elements available. Especially switches are critical because they connect different tracks and allow a train to change its moving direction without stopping. Their inspection, maintenance and repair have been identified as a cost.

The Institute of Transportation Systems in cooperation with the German Railways (DB AG) is exploring ways to apply a diagnostic and prognostic health management by monitoring the condition of switches and their degeneration process to reduce failures and thus maintenance costs. Due to the fact that switches are exposed to strong forces and sometimes extreme weather conditions, any sensor applied in the field has to be very reliable and robust. But such sensors are expensive. There are only a few monitoring systems on the market that fulfil these requirements, but none of them provides a satisfying accuracy in terms of failure diagnosis.

This contribution compares the failures indicated by the system with the actual failures that have occurred using ROC graphs as a measurement. These inaccuracies result from several external parameters influencing the switch condition, hence producing noise in the measurement. These parameters and how they are measured without additional sensors are explained. It is shown how external data sources are integrated and used to reduce the noise. This involves a combination of data mining methods like K-Modes clustering and artificial Neural Networks. The resulting improvement of the diagnostic accuracy is then expressed using false positive and true positive rate as a primary measure.

---

Thomas Böhm, German Aerospace Center, Institute of Transportation Systems, Lilienthalplatz 7, 38108 Braunschweig, Germany

## INTRODUCTION

A highly available infrastructure is a premise for capable railway operation of high quality. Therefore maintenance is necessary to keep railway infrastructure elements available. Especially switches are critical because they connect different tracks and allow a train to change its moving direction without stopping. Besides, switches are responsible for 19% of all minutes of delay in the network of German Railways (DB AG) [1]. In 2010 the DB AG registered 147.5 million minutes of delay [2]. Thus 28.025 million minutes respectively 53.32 years of delay are caused by faults or failure of switches. This makes switches crucial for the quality of operation and the attractiveness of rail transport. But their inspection, maintenance and repair have been identified as a cost driver for infrastructure managers.

The Institute of Transportation Systems in cooperation with the DB AG is exploring ways to apply a diagnostic and prognostic health management by monitoring the condition of switches and their degeneration process to reduce failures and maintenance costs. Due to the fact that switches are exposed to strong forces and sometimes extreme weather conditions, any sensor applied in the field has to be very reliable and robust. But such sensors are expensive. The investment for equipping all of the 71674 switches and crossings would far exceed the available budget.

Additionally the railway operator has to prove a reactionless functionality to ensure that no safety issues arise from the monitoring and the corresponding data transmission (e.g. accidently repositioning of the switch leading to derailment or crash). There are only a few monitoring systems on the market that fulfil these requirements. Field experience has shown that none of them provides a satisfying accuracy in terms of failure diagnosis [3].

The research goal of the Institute of Transportation Systems is to reduce delays for rail passengers and the maintenance costs for the infrastructure by improving the switch condition diagnosis and prediction. Therefore this contribution compares the failures indicated by a diagnostic system with the actual failures that have occurred using ROC graphs as a measurement. These inaccuracies result from several external parameters influencing the switch condition, hence producing noise in the measurement. These parameters and how they are measured without additional sensors are explained in the following. It is also shown how external data sources are integrated and used to reduce the noise. The resulting improvement of the diagnostic accuracy is then expressed using the false positive and the true positive rate as a primary measure.

## SWITCH DIAGNOSIS SYSTEM AS FUNDAMENTAL CONDITION SENSOR

The switch engine (also referred to as point machine) moves the switch tongues (also referred to as blades), which make contact with one or another rail and thus enabling a passing train to take one or the other direction. Switch diagnosis systems have been invented to monitor the function of the switch engine respectively the switch by monitoring the electrical power consumption during tongue repositioning.

From the few systems available on the market in Europe, e.g. Roadmaster 2000 from VAE, POSS from Strukton, and SIDIS W from Siemens, this research is based on the latter. SIDIS W only uses the switch motor as “sensor”. Relevant data, like voltage, amperage, and effective power, are directly measured at the engine and then processed in a remote diagnostic component. During the repositioning of tongues the power timeline graph shows a typical development (see Figure 1 left side). Based on that idealistic, general pattern each individual switch is recorded with its characteristic level of power at the installation. Also four threshold values are defined. A yellow and a red alert, each for an upper and a lower boundary, are set (see Figure 1 right side). In case the measured values fall below or exceed their thresholds respectively, first a yellow alert is given. If the measured values decrease or increase further respectively, a red alert is given. SIDIS W presents eight indicators, each displaying the condition state during a different phase or of a different parameter. Should the measured power level during operation, respectively the corresponding indicator flag, reach the yellow or even the red alert, a failure might occur or might have occurred. At this time at the latest the responsible operation staff should maintain or at least inspect the corresponding switch. Theoretically, SIDIS W is able to detect tongue deformation, exceptional slackness of the locking, engine failures, and hardly moving tongues caused by a bad switch condition [4].

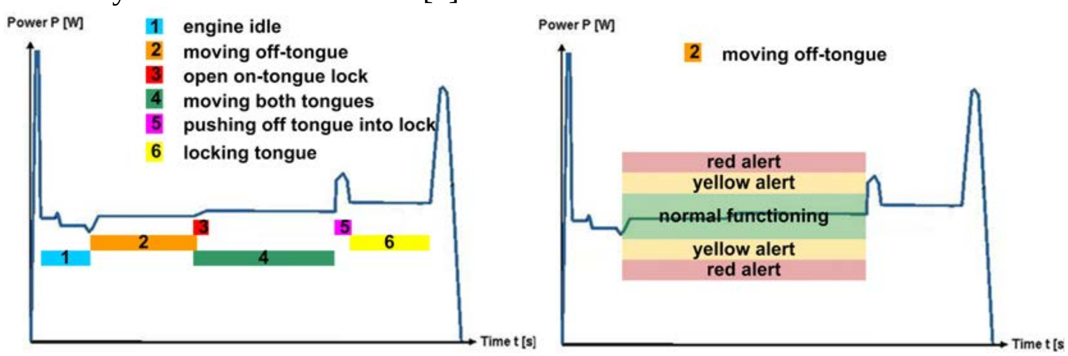


*Figure 1. Pattern of effective power during repositioning of tongues within SIDIS W [4].*

## ACCURACY OF THE ORIGINAL DIAGNOSIS

The question is how accurate is SIDIS W detecting failures? The yellow or even the red alerts are supposed to indicate a failure at the switch. Hence the maintenance staff should react to an alert to avoid any disturbance in the railway operation. On the one hand this would eventually reduce the down time of switches and the delays. On the other hand every inspection or repair needs budget spending and requires the time of the maintenance staff. Therefore a balance between detecting failures and avoiding false alerts has to be found, as with every other diagnosis system.

One way to look at the accuracy is to simply count the number of measures taken during a given time period. Then the number of alerts and the number of actual failures of a switch is determined. Table 1 shows the results. For the 11 switches in the table, which are all at the same railway station, this simple counting gives the impression that most of the time an alert can be ignored because only a few of the many alerts are accompanied by a recorded failure.

Table 1. Switches and the results from the diagnosis system during a given time period.

| Switch | Period [d] | Measures | Yellow Alerts (no Red) | % of Measures | Red Alerts | % of Measures | Failures |
|---|---|---|---|---|---|---|---|
| 1 | 712 | 2429 | 819 | 33,72% | 12 | 0,49% | 2 |
| 2 | 712 | 1376 | 576 | 41,86% | 124 | 9,01% | 1 |
| 3 | 713 | 754 | 393 | 52,12% | 20 | 2,65% | 0 |
| 4 | 596 | 760 | 45 | 5,92% | 9 | 1,18% | 0 |
| 5 | 126 | 10992 | 7523 | 68,44% | 2510 | 22,83% | 4 |
| 6 | 323 | 31229 | 4134 | 13,24% | 358 | 1,15% | 13 |
| 7 | 322 | 10911 | 1889 | 17,31% | 757 | 6,94% | 11 |
| 8 | 126 | 432 | 134 | 31,02% | 24 | 5,56% | 0 |
| 9 | 713 | 2224 | 485 | 21,81% | 220 | 9,89% | 1 |
| 10 | 709 | 2059 | 930 | 45,17% | 681 | 33,07% | 0 |
| 11 | 709 | 1800 | 69 | 3,83% | 9 | 0,50% | 0 |

Although the analysis of the measurement data showed that SIDIS W is able to reveal the degradation of the switch over months [5], alerts seem to occur randomly in smaller periods of weeks or days. Former research showed that the effective power consumption of the engine is influenced by various parameters. They cause a fluctuation in the level of power consumption, which every now and then strikes the thresholds for alerts without an abnormal behaviour and hence gives false alerts [6]. Beside the constructional characteristics of the switch the climate and especially the temperature has been identified as the main influence on the measures. Table 2 shows the average correlation between some of the measured attributes of SIDIS W and some climate parameters. The temperature has a high correlation to the power consumption during the idle engine at the beginning (Phase 1) and the maximum power at the end of the repositioning (P Maximum). The last row of the table gives the standard deviation of the correlation to the temperature. It shows how wide the range can be, meaning that the influence of the temperature on some switches is bigger than on others, depending on the constructional characteristics of the switch.

Table 2. Average power and climate correlation matrix of 29 switches.

| | 1 Phase | 2 Phase | 4 Phase | 6 Phase | Time of | P Maxin | Engine \ | Peak of | Humidity | Tempera | Atmospl | Precipita |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1 Phase** | 1,00 | 0,24 | 0,31 | 0,16 | 0,06 | -0,47 | 0,35 | 0,16 | -0,42 | 0,60 | -0,01 | 0,01 |
| **2 Phase** | 0,24 | 1,00 | 0,57 | 0,51 | 0,32 | 0,18 | 0,27 | 0,42 | -0,03 | -0,14 | 0,07 | -0,02 |
| **4 Phase** | 0,31 | 0,57 | 1,00 | 0,53 | 0,23 | 0,00 | 0,33 | 0,57 | -0,09 | 0,05 | 0,02 | -0,01 |
| **6 Phase** | 0,16 | 0,51 | 0,53 | 1,00 | 0,40 | 0,13 | 0,18 | 0,71 | -0,06 | -0,06 | 0,07 | -0,03 |
| **Time of Repositioning** | 0,06 | 0,32 | 0,23 | 0,40 | 1,00 | 0,05 | -0,01 | 0,31 | -0,05 | -0,07 | 0,01 | -0,01 |
| **P Maximum** | -0,47 | 0,18 | 0,00 | 0,13 | 0,05 | 1,00 | 0,02 | 0,06 | 0,43 | -0,77 | 0,09 | -0,01 |
| **Engine Voltage** | 0,35 | 0,27 | 0,33 | 0,18 | -0,01 | 0,02 | 1,00 | 0,16 | 0,03 | 0,01 | -0,01 | 0,01 |
| **Peak of 5 Phase** | 0,16 | 0,42 | 0,57 | 0,71 | 0,31 | 0,06 | 0,16 | 1,00 | -0,06 | -0,02 | 0,04 | -0,02 |
| **Humidity** | -0,42 | -0,03 | -0,09 | -0,06 | -0,05 | 0,43 | 0,03 | -0,06 | 1,00 | -0,50 | -0,12 | 0,08 |
| **Temperature** | 0,60 | -0,14 | 0,05 | -0,06 | -0,07 | -0,77 | 0,01 | -0,02 | -0,50 | 1,00 | -0,11 | 0,03 |
| **Atmospheric Pressure** | -0,01 | 0,07 | 0,02 | 0,07 | 0,01 | 0,09 | -0,01 | 0,04 | -0,12 | -0,11 | 1,00 | -0,08 |
| **Precipitation** | 0,01 | -0,02 | -0,01 | -0,03 | -0,01 | -0,01 | 0,01 | -0,02 | 0,08 | 0,03 | -0,08 | 1,00 |
| **Std. Deviation to Temp.** | 0,36 | 0,36 | 0,31 | 0,35 | 0,39 | 0,29 | 0,15 | 0,30 | 0,21 | 0,00 | 0,09 | 0,03 |

The conclusion drawn from the influence of the temperature on the measured attributes or on the alerts respectively is to exclude this noise to reduce the false alerts. Beforehand, the temperature has to be known. One way to receive the information would be to install a temperature sensor at every switch. But such sensors must withstand the rough conditions in the field, like strong forces and dirt. They also should be reliable in order not to become a source of failure themselves. Such sensors would be expensive. Therefore, it is the aim to get the temperature data without the installation of new sensors.

## INTEGRATION OF EXTERNAL DATA SOURCES TO REDUCE THE NOISE IN THE DIAGNOSIS

As presented in [6] the temperature data needed for the integration in the diagnosis can be retrieved from an external source. Weather stations in the proximity of the switches provide the information openly. With the help of climate data from a weather station less than ten kilometers away, data adjustment of the original measures from the diagnosis system can be reached by the following steps:

### Clustering Switches According to their Constructional Characteristics

First, a clustering approach is used. With the K-Modes-algorithm introduced by [7] the switches are grouped to treat those equally which have the same constructional characteristics, e.g. the switch type, the curve radius, the sleepers, etc.

### Learning the Alert Thresholds with an Artificial Neural Network (aNN)

Secondly, an artificial Neural Network (aNN) is used to learn the thresholds of the measured attributes at which the yellow and red alerts are given. This is done individually for each switch in each cluster. Since only five out of eight indicators are influenced by the temperature, only their corresponding thresholds are learned. This step is necessary because neither the exact figure of the thresholds was known nor the precise calculation in relevance to the depending attributes. For example, the indicator for failures of the switch tongue lock at the end of the repositioning depends on several measured attributes. The measured power level during tongue locking (Phase 6) is set in relation to the power level during the movement of both tongues (Phase 4) and the maximum power. So the corresponding alerts are the result of an equation taking into account these tree attributes with unknown factors. In this experiment the learning algorithm introduced by [8] was used. No aNN predicted the alerts with accuracy less than 99.6%. However, this step is redundant in case the thresholds are known precisely.

### Adjusting the Original Diagnosis Measures

Third, the original measured values are adjusted to the temperature. For each switch in each cluster the correlation between diagnosis attributes and temperature are calculated. Whenever the correlation is below -0.6 or above 0.6 its coefficient of determination is calculated. The coefficient of determination can be interpreted as a metric of how much the change in one value can be explained with the change of another. Additionally, the average change of an attribute per °C is calculated and then multiplied by the coefficient of determination. The result is a correction value per °C for each relevant measured attribute of each switch. This correction value is applied to the measures while the temperature is set to its overall average during the time period, in this case 11.3 °C. This way the measures of the diagnosis system are smoothed and the fluctuations caused by the temperature are reduced.

### Resetting the Alerts of the Diagnosis System

As fourth and last step, the trained aNN is fed with the adjusted measures. The result contains diagnosis measures of each switch, but with the red and yellow alerts for five of the eight indicators reset.

## IMPROVED ACCURACY OF THE TEMPERATURE ADJUSTED DIAGNOSIS

Once the indicators are reset with the integration of external data source (the temperature retrieved from a weather station) it has to be evaluated if this really improved the accuracy. The simple numbers seem to indicate this. The yellow alerts are reduced to 14134 from 16997 (83.1%) while the red alerts have decreased to 386 from 4724 (8.1%). Regarding the balance between detecting failures and avoiding false alerts, these numbers give little information about the performance of either the original diagnosis or the temperature adjusted diagnosis. A better way to look at the diagnostic performance of the system is a Receiver Operating Characteristics (ROC) graph. As introduced by [9], a ROC curve is a visualisation technique used to analyse classifiers according to their performance. It is a well established method to analyse diagnostic systems [10]. It is based on a confusion matrix which contains the predicted and the actual class. The metrics displayed in an ROC graph are calculated from the matrix. The false positive (also referred to as false alert rate) is the relation between wrongly predicted faults and all non-faults in a data set. The true positive rate is the relation between the correctly predicted faults and all faults. In the ROC graph the best predictor is the one closest to the left upper corner. There every fault is predicted while no false alert is given. Figure 2 shows the basic concept of ROC graphs.

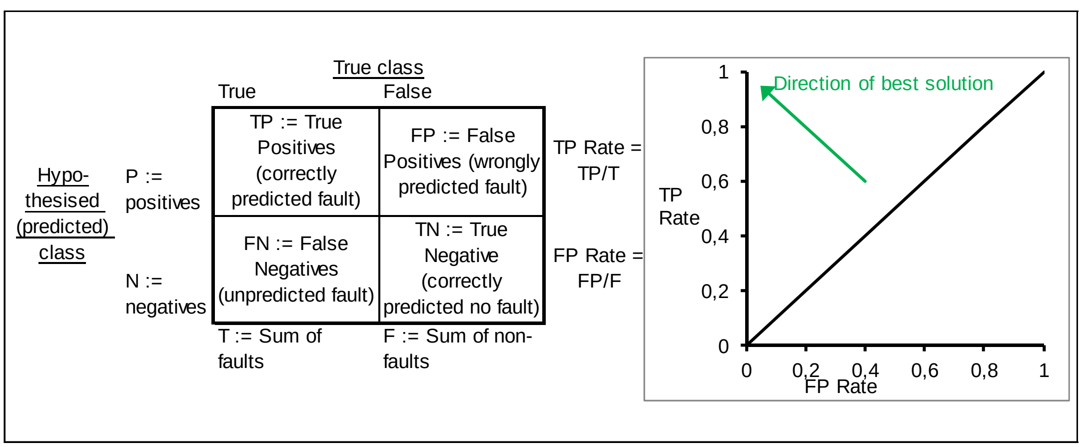


*Figure 2. Confusion matrix of predicted and actual class and basic ROC graph.*

Taking only the plain number of alerts and failures would provide only two single points for the systems. In order to compare the performance of the original diagnosis with the one after the integration of the external data more points are necessary. Therefore, the following metrics was used to determine a failure from the indicators and their alerts:

- Any of the indicators gave a red alert.
- Any of the indicators gave a red or yellow alert.
- All indicators gave a red alert.
- All indicators gave a red or yellow alert.
- Any indicator gave x consecutive red alerts, in which x is 5, 10, 20 or 50.
- Any indicator gave x consecutive red or yellow alerts, in which x is 5, 10, 20 or 50.
- In y hours any indicator gave z percent red alerts, in which y is 48, 96 or 144 and z is either 67%, 75%, 80% or 90%.
- In y hours any indicator gave z percent red or yellow alerts, in which y is 48, 96 or 144 and z is 67%, 75%, 80% or 90%.

Due to the comprehensiveness, not all of the results are displayed in this paper. Only those which are significant for the comparison are provided in Table 3. It lists some metrics and their corresponding FP Rate und TP Rate, while Figure 3 shows a sample of the analysis in the ROC graph (note that the graph is pruned above 0.5 for space saving). In the ROC graph one line connects one pair of the original and adjusted diagnosis results for a given metric. The dot at the line represents the FP/TP rate of the original diagnosis system. The arrowhead represents the FP/TP rate of temperature adjusted diagnosis.

Table 3. False and true positive rates from the original and the temperature adjusted diagnosis.

| **Metric** | | Origninal Diagnosis | | Temperatrue Adjusted Diag | |
|---|---|---|---|---|---|
| | | **FP Rate** | **TP Rate** | **FP Rate** | **TP Rate** |
| 1. | 5 Consecutive Red Alerts before Failure | 0,0107 | 0 | 0,0007 | 0 |
| 2. | 67% Red Alerts in the 48 h before Failure | 0,0015 | 0,0313 | 0 | 0 |
| 3. | 80% Red or Yellow Alerts in the 48 h before Failure | 0,0057 | 0,1563 | 0,0044 | 0,1250 |
| 4. | 75% Red or Yellow Alerts in the 48 h before Failure | 0,0058 | 0,1563 | 0,0046 | 0,1250 |
| 5. | 67% Red or Yellow Alerts in the 48 h before Failure | 0,0072 | 0,1563 | 0,0062 | 0,1250 |
| 6. | 75% Red or Yellow Alerts in the 144 h before Failure | 0,0225 | 0,1563 | 0,0127 | 0,0938 |
| 7. | 67% Red or Yellow Alerts in the 144 h before Failure | 0,0237 | 0,1563 | 0,0151 | 0,1563 |
| 8. | Any Red Alert before Failiure | 0,0727 | 0,1875 | 0,0059 | 0,0313 |
| 9. | 10 Consecutive Red or Yellow Alerts before Failure | 0,1279 | 0,1875 | 0,0756 | 0,1563 |
| 10. | 5 Consecutive Red or Yellow Alerts before Failure | 0,1641 | 0,1875 | 0,0923 | 0,1875 |
| 11. | Any Alert before Failure | 0,3343 | 0,5000 | 0,2234 | 0,4375 |

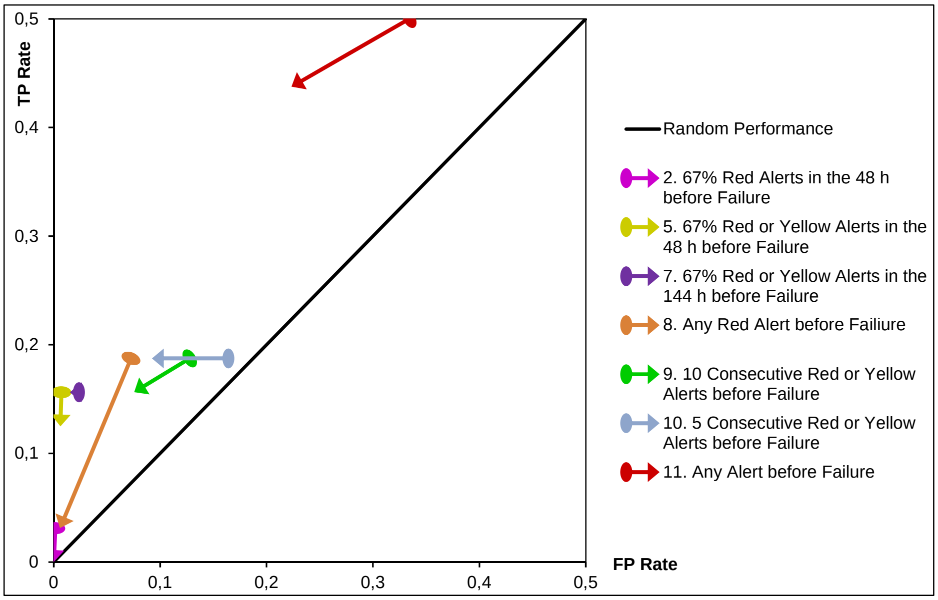


*Figure 3. ROC graph comparing the original diagnosis to the temperature adjusted diagnosis.*

In most cases the diagnosis adjusted to the temperature performs better. Anyhow, there are metrics at which the original diagnosis performs better, like any red alert before failure or a high percentage of alerts in a short period of time. The high number of alerts of the original diagnosis makes it more likely to alert before an actual failure. But those metrics were outperformed by others producing results in the upper left area of the graph. What Table 1 and Figure 3 do not show is that the integration of external temperature data did not detect any failure undetected before. Since most of the measured values of effective power have been reduced and most alerts came from exceeding the thresholds, only a very small number of new red alerts had been set. In general, the main improvement is the decrease of false alerts. Hence, this will make the diagnosis more reliable, even if not to a level at which the maintenance staff reacts on every hypothetical failure indicated by the diagnosis system.

## CONCLUSION AND FURTHER RESEARCH

The research aims to reduce the delays caused by failures of railway switches and also to reduce maintenance costs. Therefore the focus is on finding ways to apply a diagnostic and prognostic health management by monitoring the condition of switches and their degeneration process. The basis is an existing diagnosis system. But its accuracy is not sufficient enough, mainly because of its high number of false alerts. Other parameters interfere with the diagnosis and produce noise responsible for false alerts. After quantifying the temperature as one parameter of significance, a way has been presented to adjust the measures of the original diagnosis. The temperature has been retrieved and integrated using a weather station as external data source, available at almost no additional costs. Hence the equipping of additional sensor is not necessary. The measures of the diagnosis system have been recalculated excluding the influence of the temperature and the alerts have been reset. Both, the original and the adjusted diagnosis have been compared according to their false positive and true positive rates. In regard to the ROC graph of both systems, the adjusted diagnosis has shown some advantages compared to the original system. Though the adjusted diagnosis did not detect any new failures, the main improvement is the decrease of false alerts and thus the improvement of reliability.

However, the results also show that the true positive rate needs to be improved. Therefore other parameters which influence the switch condition need to be quantified and included in the diagnosis. Additionally, the metrics by which a failure is indicated will be subject to further research. Thus the results at hand are only an early stage on the way to predict switch failures in the railway network.

## REFERENCES


[1] Rausch, Wolfgang: Diagnosesysteme für Weichen als Grundlage für eine optimierte Instandhaltungsplanung. In: VDEI (Hrsg.): *Symposium zum Thema: "Moderne Instandhaltungsverfahren für Weichen – Qualitätsansprüche – Wirtschaftlichkeit"*. Brandenburg-Kirchmöser, 2010

[2] DB Netz AG: *Infrastrukturzustands- und -entwicklungsbericht 2010 : Internet Version.* Frankfurt am Main, April 2011

[3] Holst, Söhnke: *Lösung zur Instandhaltung signaltechnischer Anlagen*. In: Deine Bahn 40 (2012), Nr. 2, pp. 19–23

[4] Stoll, Horst; Bollrath, Bernhard: *Weichendiagnosesystem SIDIS W*. In: *Signal + Draht* 94 (2002), Nr. 4, pp. 26–29

[5] Gutsche, Katja; Böhm, Thomas: *e-Maintenance of Railway Assets Based on a Reliable Condition Prediction.* In: *International Journal of Performability Engineering* Vol. 7 (2011), Nr. 6, pp. 573–582

[6] Böhm, Thomas; Doegen, Christian: *Diagnosis without Sensors - Integration of External Data for Condition Monitoring of Railway Switches*. In: Okumura, Susumu; Kawai, Tadao; Chen, Peng; Rao, R. BKN (Hrsg.): *COMADEM 2010 - Advances in maintenance and condition diagnosis technologies towards sustainable society : Proceedings of the 23rd International Congress on Condition Monitoring and Diagnostic Engineering Management,* Hikone, Japan : Sunrise Publishing, 2010, pp. 619–622

[7] Huang, Zhexue: *Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values.* In: *Data Mining and Knowledge Discovery* 2 (1998), Nr. 3, pp. 283–304.

[8] Riedmiller, M.; Braun, H.: *A direct adaptive method for faster backpropagation learning: theRPROP algorithm*. In: *Proceedings of the IEEE International Conference on Neural Networks*. Piscataway, NJ, USA, 1993, pp. 586–591.

[9] Egan, James P.: *Signal Detection Theory and ROC Analysis.* New York : Academic Press, 1975 (Series in Cognitition and Perception)

[10] Swets, John A.: *Measuring the accuracy of diagnostic systems.* In: *Science* 240 (1988), pp. 1285–1293.