

# Anomaly Detection and Severity Prediction of Air Leakage in Train Braking Pipes

Wan-Jui Lee<sup>1,2</sup>

<sup>1</sup> *Maintenance Development, Dutch Railways, Utrecht, 3500GD, the Netherlands*  
wan-jui.lee@ns.nl

<sup>2</sup> *Pattern Recognition and Bioinformatics Group, Delft University of Technology, Delft, 2628CD, the Netherlands*  
W.J.Lee@tudelft.nl

## ABSTRACT

Air leakage in braking pipes is a commonly encountered mechanical defect on trains. A severe air leakage will lead to braking issues and therefore decrease the reliability and cause train delays or stranding. However, air leakage is difficult to be detected via visual inspection and therefore most air leakage defects are run to fail. In this research we present a framework that not only can detect air leakages but also predicts the severity of air leakages so that action plans can be determined based on the severity level. The proposed contextual anomaly detection method detects air leakages based on the on/off logs of a compressor. Air leakage causes failure in the context when the compressor idle time is short than the compressor run time, that is, the speed of air consumption is faster than air generation. In our method the logistic regression classifier is adopted to model two different classes of compressor behavior for each train separately. The logistic regression classifier defines the boundary separating the two classes under normal situations and models the distribution of the compressor idle time and run time separately using logistic functions. The air leakage anomaly is further detected in the context that when a compressor idle time is erroneously classified as a compressor run time. To distinguish anomalies from outliers and detect anomalies based on the severity degree, a density-based clustering method with a dynamic density threshold is developed for anomaly detection. The results have demonstrated that most air leakages can be detected one to four weeks before the braking failure and therefore can be prevented in time. Most importantly, the contextual anomaly detection method can pre-filter anomaly candidates and therefore avoid generating false alarms. To facilitate the decision-making process, the logistic function built on the compressor run time is further used together with the duration of an air

leakage to model the severity of the air leakage. By building the prediction model on the severity, the remaining useful life of the air braking pipe until it reaches a certain level of severity can be estimated.

## 1. INTRODUCTION

Dutch Railways, the principal railway operator in the Netherlands, operates 178 VIRM (lengthened interregional rolling stock) trains which are a series of electric multiple unit (EMU) double-deck trains. These trains were built between 1994 and 2009 with on-board train management systems continuously logging particular events on the local disk or on a remote disk using wireless data communications. To ensure all trains are reliable and safe to operate at the lowest cost, Dutch Railways is continuously optimizing the maintenance schedule to plan when and what to maintain (de Vos & van Dongen, 2015). Condition-based monitoring data collected from the on-board train management systems enable us to move towards prognostics and health management (Coble, Ramuhalli, Bond, Hines, & Upadhyaya, 2015; Schenkendorf & Groos, 2015) of rolling stock. Condition monitoring (Bartam & Mahadevan, 2015; Eker, Camci, & Jennions, 2014; Prakash, Narasimhan, & Pandey, 2017) is more than detecting train failure or malfunctions. Continuous gathering of data allows for trend analysis over the entire fleet and allows for data-driven performance improvements for instance actual state-dependent maintenance (Poot-Geertman, Huisman, & van Rijn, 2015; O'Donovan, Bruton, & O'Sullivan, 2016; Qiao & Weiss, 2016). With the wealth of condition information, static and reactive preventive and corrective maintenance scheduling will be replaced by dynamic and proactive predictive maintenance in the future (Eker et al., 2014; Schenkendorf & Groos, 2015).

Like the heart of a human body which pumps blood around the body, the compressor of a train pumps air into the braking pipe. The air in the braking pipe is then consumed by op-

Wan-Jui Lee et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

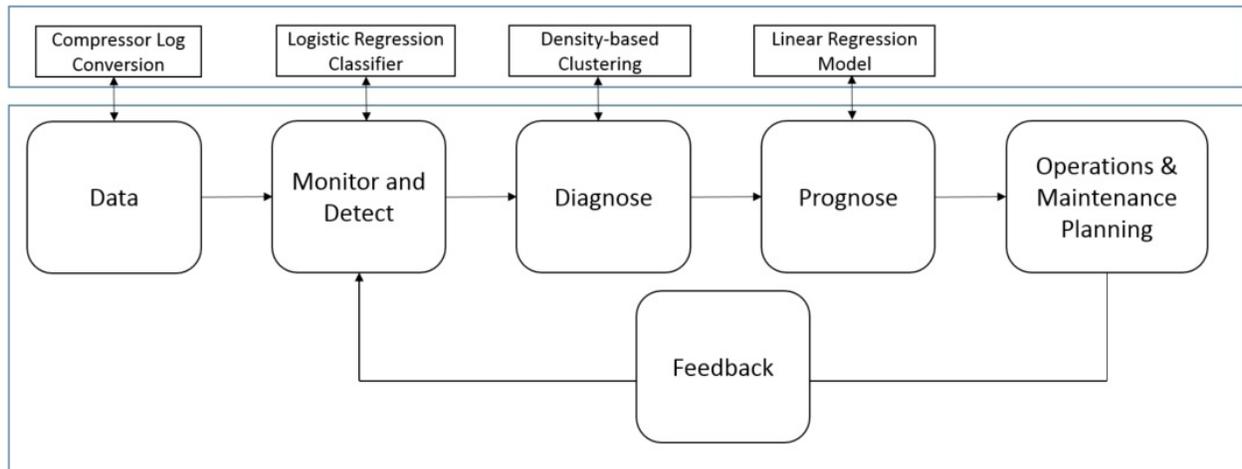


Figure 1. Proposed prognostic health management methodology for air leakage in braking pipe.

operations such as braking, door opening and closing and bio-reactor usage. Air leakage in braking pipes is a commonly encountered mechanical defect on trains which occurs roughly 1.5 times a train per year on average. A severe air leakage will lead to braking issues and cause train delays or stranding. For a dense network operating 1,000 trains in the Netherlands, it is a high risk in disturbing the passenger transportation service. In short, air leakage defects are critical both in business impact and service reliability. Therefore, checking air leakage is one of the daily maintenance tasks in the service depot carried out by technicians. Nevertheless air leakage is one of the most difficult defects to be detected by visual or audio inspection carried out in the workshops. According to historical maintenance records from Dutch Railways, only a very low percentage of air leakages can be discovered by daily inspections. Vast majority of air leakages were often reported by train drivers during operation when they were facing braking or door issues resulting in an event of train delay or stranding.

In this work, a new discovery of using switch on and off logs of a compressor to detect air leakages in braking pipes is reported. Due to the high variation and noisy nature of the compressor behavior data, clustering techniques (Upadhyaya & Singh, 2012; Behera & Rani, 2016; Khan, Awad, & Thuraisingham, 2007) are considered most applicable to our application for anomaly detection. Anomaly detection (Chandola, Banerjee, & Kumar, 2009; Behera & Rani, 2016; Khan et al., 2007; Biswas et al., 2016) is widely applied in many applications where continuous monitoring is available. The goal is to find variants that are different from normal behaviors. In applications where false positives are very expensive, post-processing or human interaction are often required to eliminate false positives. Contextual anomaly detection (Mahapatra, Srivastava, & Srivastava, 2012; Hayes & Capretz, 2015) is a newly emerging field of study that aims to detect anomalies that occur within the context of other meta-

information such as spatial or temporal information. For instance, a sensor value 0 during work hours might be normal while it is abnormal during off-work hours. In this study the logistic regression classifier (Mitchell, 1997; Ng & Jordan, 2002) is adopted for building context of "Compressor Run Time" and "Compressor Idle Time" separately for each train. The context is used for defining a threshold to filter out non-targeted regions because air leakage is most likely occur in regions where "Compressor Idle Time" is overlapped with "Compressor Run Time". This threshold differs per train due to difference in configuration, age and usage, and therefore the role of the logistic regression classifier is to model the distribution of these two classes in order to identify the decision boundary between them separately for each train and use it as the threshold.

However, during normal services, the air consumption in the braking pipes can be triggered by activities such as braking, door opening and closing and bio-reactor usage and so forth. Also the number of carriers of the train has an impact on the duration of air consumption. Therefore there might be sudden and singular occurrences of speedy air consumption due to sudden increase of air consumption demands. This kind of sudden increase of usage demand is defined as an outlier. The most intuitive way to distinguish anomaly from outlier is based on density since a leakage is a mechanical failure which will occur constantly in a certain period of time. On the other hand, a sudden increasing demand is often a single and random event. In this work, we have developed a density-based clustering approach which is inspired by DBSCAN (Ester, Kriegel, Sander, & Xu, 1996; Birant & Kut, 2007) to detect regions of high density. These regions indicate the existence of air leakage. To consider the severity degree of air leakage in anomaly detection, we have defined a dynamic density threshold based on the logistic model describing the context. The result of contextual anomaly detec-

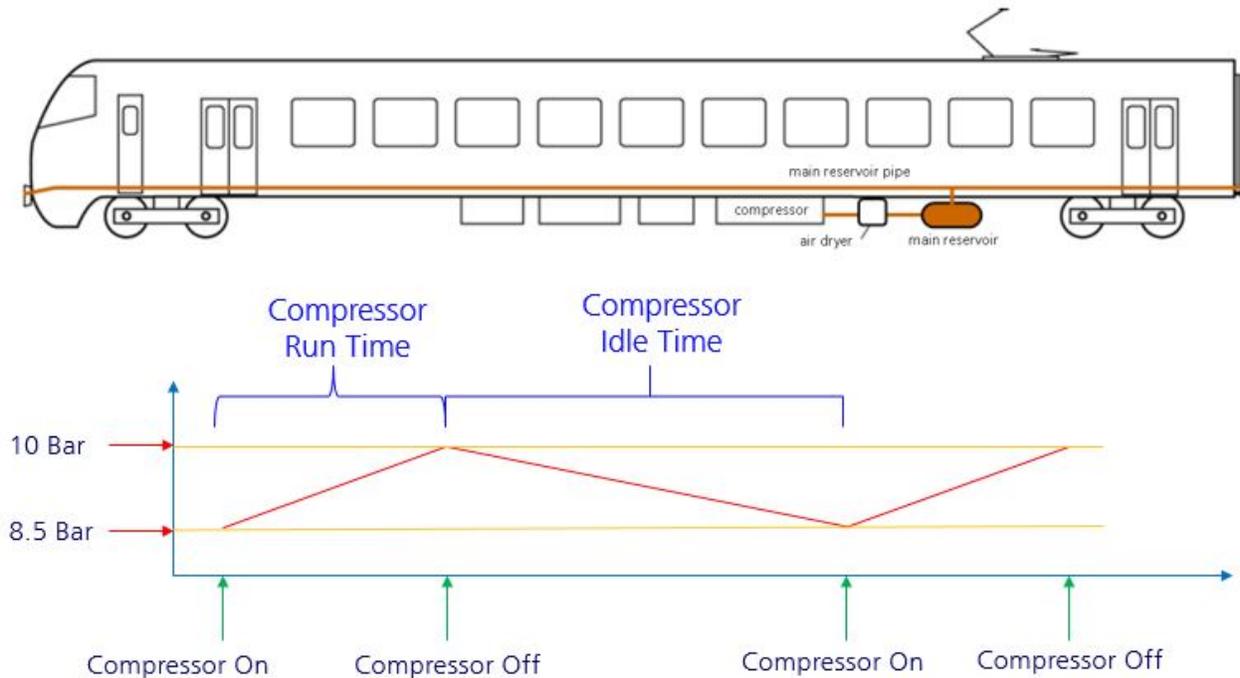


Figure 2. Illustration of air circulation system.

tion based on the density-based clustering approach suggest that air leakages in braking pipes can be detected at least one to four weeks before the braking failure.

To facilitate the decision-making process and decide the right time for operations and maintenance actions, the logistic function built on the compressor run time is further used together with the duration of an air leakage to model the severity of the air leakage. The severity level is modeled in such a method that it is always between 0 and 1, and a higher value means a higher severity degree in air leakage. By building the prediction model on the severity, the remaining useful life of the air braking pipe until it reaches a certain level of severity can be estimated.

To summarize, the main contribution of this work include:

- enabling the data-driven detection of air leakage in train braking pipes,
- development of the robust contextual anomaly detection method to detect air leakages among different data distributions,
- development of the severity index to indicate the severity level of a detected air leakage to facilitate the decision making process for maintenance planning.

Our proposed prognostic health management methodology for air leakage in braking pipe is as shown in Fig. 1. The lower part consists of the procedures and the upper part contains the proposed methods/models. In addition to the conventional procedures (Coble et al., 2015), we have added a

feedback loop to the methodology to enable both validation and continuous improvement on the models. In this paper, the application and development of logistical regression classifier in the Monitor and Detect procedure and density-based clustering in the Diagnose procedure will be described in detail in Section 2. The development of severity modeling by logistic function and severity prediction using the Linear Regression models will be introduced in Section 3. Experimental results on individual trains and over the fleet will be presented and discussed in Section 4.

## 2. CONTEXTUAL AIR LEAKAGE DETECTION IN BRAKING PIPES

In Fig. 2, the sketch of the mechanism of the air circulation system is illustrated. The air pressure of a main reservoir on VIRM trains should be kept in the level between 8.5 and 10 bar at all times. When it drops to below 8.5 bar, the compressor will be switched on to pump air into the main reservoir until the air pressure reaches 10 bar again. After 10 bar is researched, the compressor will be switched off. Air in the main reservoir will be consumed by the braking pipe during service. The time it takes for a compressor to pump air into the main reservoir is defined as the "Compressor Run Time" in this work. The time it takes for the braking pipe to consume air in the main reservoir while the compressor is switched off is defined as the "Compressor Idle Time". It is not difficult to imagine that when the speed of air generation is slower than air consumption there will be insufficient air supply to the

braking pipes and therefore braking issues will occur. One of the most possible cause of this phenomenon is air leakage in the braking pipe which is a commonly found mechanical defect on trains. To the best of our knowledge, this is the first work discovering the capability of switch on/off logs of a compressor in detecting air leakage in braking pipes. Such a discovery is extremely valuable since air leakage is one of the most difficult defect to be detected by visual or audio inspection carried out in the workshops. By converting the switch on/off logs of a compressor into duration of compressor run time and idle time, air leakage is finally possible to be continuously monitored and detected from data.

Due to the difference in configuration and operational use of each train, the range and distribution of "Compressor Idle Time" and "Compressor Run Time" differ per train. However, the physical observation that when the speed of air consumption is faster than the speed of air generation, there might exist an air leakage applies in general to all trains. To find out the region of interest for air leakage detection for each train, the logistic regression classifier is adopted for building context of "Compressor Run Time" and "Compressor Idle Time" separately for each train as a two class problem where "Compressor Run Time" is the positive class and "Compressor Idle Time" is the negative class. By building the context with logistic models, a threshold can be defined at the intersecting point where the probability of the positive class and negative class are both 0.5 to pre-filter non-targeted regions. Since air leakages occur most likely in regions where "Compressor Idle Time" is overlapped with "Compressor Run Time" and therefore only "Compressor Idle Time" with a similarity higher than 0.5 by applying the logistic model of the positive class will be considered for the clustering procedure.

## 2.1. Learning Context with Logistic Regression Classifier

Logistic regression classifier is a linear model for learning  $P(Y|X)$  in the case where  $Y$  is the class label and  $X = [x_1, x_2, \dots, x_m]$  is an input data vector. In our application we only consider the case where  $Y$  is a boolean variable (2 class problem) and  $m$  equals 1 which means the data vector is one-dimensional. The parametric model assumed by logistic regression in the 2 class setting is:

$$P(Y = 0|X) = \frac{1}{1 + \exp(w_0 + \sum_{i=1}^m w_i x_i)} \quad (1)$$

and

$$P(Y = 1|X) = \frac{\exp(w_0 + \sum_{i=1}^m w_i x_i)}{1 + \exp(w_0 + \sum_{i=1}^m w_i x_i)}. \quad (2)$$

The goal is to learn the parameters  $w_j$ , for all  $j$ 's from training data. Since the sum of the two probabilities in Eq.(1) and Eq.(2) must equal 1, Eq.(2) can be directly derived from Eq.(1).

In our application, we use the point  $\hat{X}$  where  $P(Y = 0|\hat{X})$  equals  $P(Y = 1|\hat{X})$  as the threshold. That is the point  $\hat{X}$  where

$$\frac{1}{1 + \exp(w_0 + w_1 \hat{x}_1)} = \frac{1}{2}, \quad (3)$$

and  $\hat{X} = [\hat{x}_1]$  since the data vector is one-dimensional in our application. By taking the natural logarithm, this becomes

$$w_0 + w_1 \hat{x}_1 = 0. \quad (4)$$

By transforming Eq.(4), one can derive

$$\hat{x}_1 = -\frac{w_0}{w_1}. \quad (5)$$

Therefore, after learning parameters  $w_j$ , for all  $j$ 's, the point  $\hat{X}$  can also be derived. That is, for all data points in the negative class ("Compressor Idle Time"), only those less than or equal to  $\hat{X}$  will be included in the clustering procedure for anomaly detection.

In Fig. 3, an example is given to illustrate the functionality of the logistic regression classifier in our application. From the distribution of 2 classes in Fig. 3(a), the overlapped area can be observed. The logistic regression classifier models these 2 classes with logistic functions as shown in Fig. 3(b) to find the most significant point to distinguish these 2 classes. In this example, the intersecting point of these 2 classes is at 497.4 and it is used as the threshold for filtering out any "Compressor Idle Time" with a duration value larger than 497.4 since these values are very unlikely to be generated from air leakages.

### 2.1.1. Training Logistic Regression Classifier

One common approach to train a logistic regression model is to choose parameter values that maximize the probability of the observed  $Y$  values in the training data, conditioned on their corresponding  $X$  values. That is, to choose parameters  $W$  satisfying

$$W \leftarrow \operatorname{argmax} \sum_{k=1}^n \ln P(Y^k | X^k, W), \quad (6)$$

where  $W = [w_0, w_1, \dots, w_m]$  is the vector of parameters to be estimated,  $Y^k$  denotes the observed value of  $Y$  in the  $k$ th training example, and  $X^k$  denotes the observed value of  $X$  in the  $k$ th training example. The expression to the right of the  $\operatorname{argmax}$  is the logarithm of the conditional likelihood.

By substituting with Eq.(1) and Eq.(2), the log of the conditional likelihood  $l(W)$  can be then expressed as:

$$l(W) = \sum_{k=1}^n Y^k \ln P(Y^k = 1 | X^k, W) + (1 - Y^k) \ln P(Y^k = 0 | X^k, W)$$

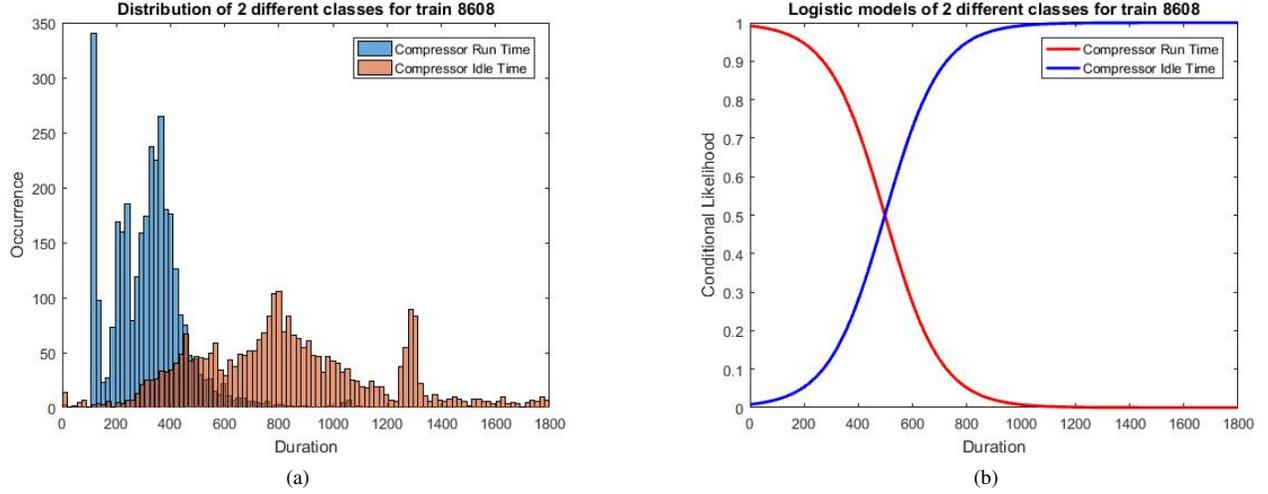


Figure 3. The (a) distribution in histograms, and (b) trained logistic models of "Compressor Run Time" and "Compressor Idle Time" data of train 8608.

$$= \sum_{k=1}^n Y^k (w_0 + \sum_{i=1}^m w_i x_i) - \ln(1 + \exp(w_0 + \sum_{i=1}^m w_i x_i)) \quad (7)$$

where  $x_i^k$  denotes the value of  $x_i$  for the  $k$ th training example.

However, there is no closed form solution to maximizing  $l(W)$  with respect to  $W$ , and one common approach is to use gradient ascent. The  $i$ th component of the vector gradient has the form

$$\frac{\partial l(W)}{\partial w_i} = \sum_{k=1}^n x_i^k (Y^k - \hat{P}(Y^k = 1 | X^k, W)) \quad (8)$$

where  $\hat{P}(Y^k = 1 | X^k, W)$  is the prediction result of the logistic regression classifier. Since the conditional log likelihood is a concave function, this gradient ascent procedure will converge to a global maximum. By beginning with initial weights of zero, the weights are iteratively updated with

$$w_i \leftarrow w_i + \eta \sum_{k=1}^n x_i^k (Y^k - \hat{P}(Y^k = 1 | X^k, W)) \quad (9)$$

where  $\eta$  is the step size which is often a small constant.

## 2.2. A Density-Based Clustering Approach for Anomaly Detection with A Dynamic Density Threshold

For clustering in a noisy dataset, density-based approaches are most commonly adopted. Among them, DBSCAN (Ester et al., 1996; Birant & Kut, 2007) is one of the most well-known approaches which requires two parameters:  $\epsilon$  and  $minPts$ . The parameter  $\epsilon$  defines the neighborhood of a considered point, and  $minPts$  is the minimum number of points

required to form a dense region. DBSCAN starts with an arbitrary starting point that has not been visited. This point's  $\epsilon$ -neighborhood is retrieved, and if it contains a sufficient number of points, a cluster is started. Otherwise, the point is considered as a noise. However this point might later be found in the  $\epsilon$ -neighborhood of a different point containing a sufficient number of points and hence be made a part of a cluster. If a point is found to be a dense part of a cluster, its  $\epsilon$ -neighborhood is also part of that cluster. Hence, all points that are found within the  $\epsilon$ -neighborhood are added, as is their own  $\epsilon$ -neighborhood when they are also dense. This procedure iterates until all points are visited.

In the case of air leakage detection, the detection capability in a severe region needs to be higher than that in a less severe region. Therefore we have defined a dynamic density threshold based on the condition of severity. The procedure of our density-based clustering approach for anomaly detection with a dynamic density threshold is described in the following:

- Step 1: Use the threshold defined in Section 2.1 to limit the search range of anomalies.
- Step 2: In the interested region, calculate the neighborhood density of each data point. The neighborhood density of a data point is the number of data points located in its  $\epsilon$ -neighborhood region. A  $\epsilon$ -neighborhood region of a data point  $x_i$  is defined by:
$$|x_i - x_j| \leq \epsilon, \forall j \quad (10)$$
 where  $\epsilon$  is an user-defined constant.
- Step 3: Classify a data point and all other points located in its  $\epsilon$ -neighborhood as anomalies if its neighborhood density is higher than the density threshold. The density

threshold should be dynamic and vary with the degree of severity. That is, a more severe air leakage (shorter idle duration) should be more easily detected by giving a lower density threshold and vice versa. By giving a user-defined density limit  $minPts$ , the dynamic density threshold  $\beta$  becomes

$$\beta = 2 * minPts * \frac{1}{1 + \exp(w_0 + \sum_{i=1}^m w_i x_i)}, \quad (11)$$

where  $\frac{1}{1 + \exp(w_0 + \sum_{i=1}^m w_i x_i)}$  is adopted from Eq.(1).

- Step 4: For a data point, if its neighborhood density is greater than or equal to its dynamic density threshold  $\beta$ , this data point and all the other data points within the neighborhood of this data points will be labeled as anomalies.
- Step 5: Data points share equivalence relations will form a connected component which is an irregular shaped cluster  $C_k$ .

In detecting mechanical defects, occurrence frequency and severity degree are two of the most commonly concerned factors. The proposed density-based clustering captures both by setting dynamic density thresholds.

### 3. SEVERITY MODELING AND PREDICTION

From the diagnostic procedures described in the above section, air leakages often can be detected one to four weeks before failure. However, for operations and maintenance planning, a exact indication for alarm triggering and action is usually desired. Therefore, in order to define the moment for action, in this work a severity index  $S$  is proposed and it is designed by combining the logistic function and the duration of a air leakage as shown in Eq.(12).

$$S_{C_k}^{t_p} = \max\left(1 - \frac{2}{1 + \exp(w_0 + \sum_{i=1}^m w_i x_i)}\right) \times \min\left(\left[\frac{t_p - t_0}{D}, 1\right], 1\right), \quad \forall x_i \in C_k \text{ at } t_p \quad (12)$$

where  $t_p$  is the current time stamp of incoming data points,  $t_0$  is the oldest time stamp of all data points  $x_i$  in  $C_k$ , and  $D$  is the pre-defined duration of an observation period. That is,  $S_{C_k}^{t_p}$  is the severity index of cluster  $C_k$  at time stamp  $t_p$ . Since  $S_{C_k}^{t_p}$  is always between 0 and 1, a threshold  $\theta$  can be easily chosen for decision making. The purpose of the observation period is to reduce the possibility of unnecessary alarms. The necessity of the observation period  $D$  may differ per application and the duration of the observation period depends on the operational and maintenance planning requirements of the train operator.

Given the current severity index, a prediction model  $f(\cdot)$  can

be designed to estimate the severity index in the upcoming hours as

$$\hat{S}_{C_k}^{t_{p+1}} = f(S_{C_k}^{t_p}, S_{C_k}^{t_{p-1}}, \dots, S_{C_k}^{t_{p-q}}), \quad (13)$$

where  $\hat{S}_{C_k}^{t_{p+1}}$  is the estimated severity index of cluster  $C_k$  for time stamp  $t_{p+1}$ . In this application, the Linear Regression model is adopted for predicting the severity indexes. The time horizon in which the severity index is predicted to reach the action threshold  $\theta$  is considered the remaining useful life of the air circulation system.

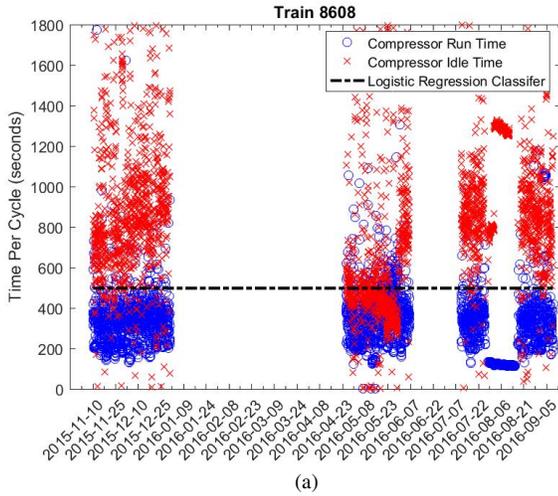
## 4. EXPERIMENTAL RESULTS

From 178 VIRM trains 632,683 data points were collected in the period from May 2015 to October 2016, in which 6,957 are labeled as "Air Leakage" and 625,726 are labeled as "Normal". To be more clear, each data point is the median "Compressor Run Time" or median "Compressor Idle Time" of a specific compressor within one hour. For different trains, the occurrences of compressor switch on/off logs differ per day (ranging from 30 to 120 counts a day) depending on the number of carriers, compressor type and bio-reactor, etc. In order to build a generic detection algorithm over the entire fleet, a more universal unit is required to construct a stable system and therefore time (in this case, hour) is used as the unit but not individual occurrence. The data labels are derived from maintenance records of air leakages in the same period. In these 178 trains, 55 trains are mounted with real-time monitoring systems and the data were sent via 4G network directly into the data center. In the rest of 123 trains, the data were read out physically with laptops in the maintenance depot. Due to the manual operations, there were sometimes long gaps in weeks or months between data records.

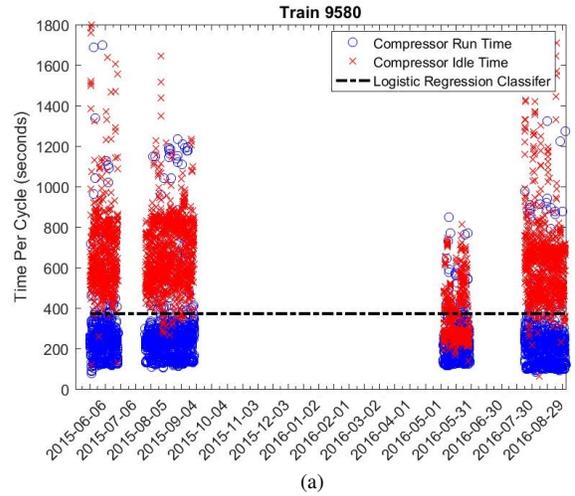
### 4.1. Diagnostics

In the experiments, the logistic regression classifier was built for each train to derive the filtering threshold on "Compressor Idle Time". For anomaly detection, our density-based clustering approach finds clusters in a two-dimensional dataset consisting two features, i.e., "Compressor Idle Time" and date converted into the number of days from January 0, 0000. The  $\epsilon$ -neighborhood of a data point for the clustering procedure is therefore two-dimensional with  $\epsilon_1$  of "Compressor Idle Time" being 0.2 times  $\hat{X}$  and  $\epsilon_2$  of the number of days being 2 days. The density limit  $minPts$  is set to 20 for computing the dynamic density threshold  $\beta$ . The selection of these  $\epsilon$  and  $minPts$  values is based on both the engineering principals and optimization against the confusion matrix (detection capability against false alarm). The considered ranges of  $\epsilon$  and  $minPts$  need to be physically sensible and then the different combination of these parameters sampled within the ranges should be tested in order to derive the optimal combination.

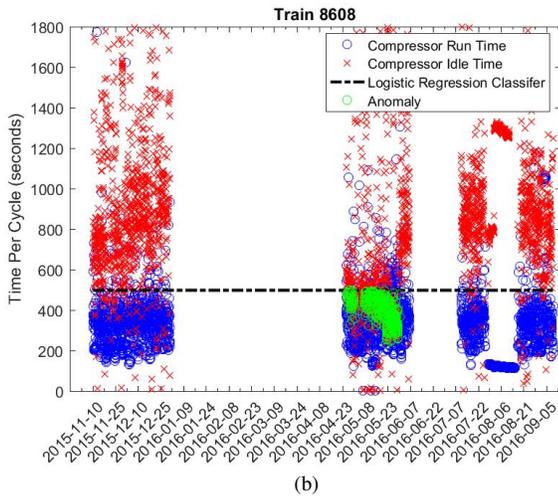
The original "Compressor Run Time" and "Compressor Idle



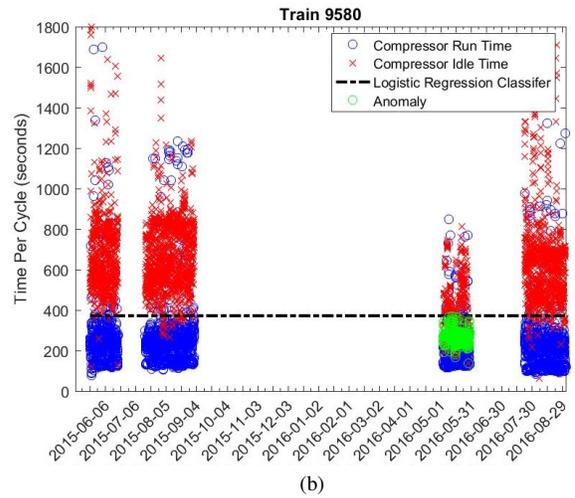
(a)



(a)



(b)



(b)

Figure 4. The (a) original compressor duration data and (b) result of the proposed air leakage detection of train 8608.

Figure 5. The (a) original compressor duration data and (b) result of the proposed air leakage detection of train 9580.

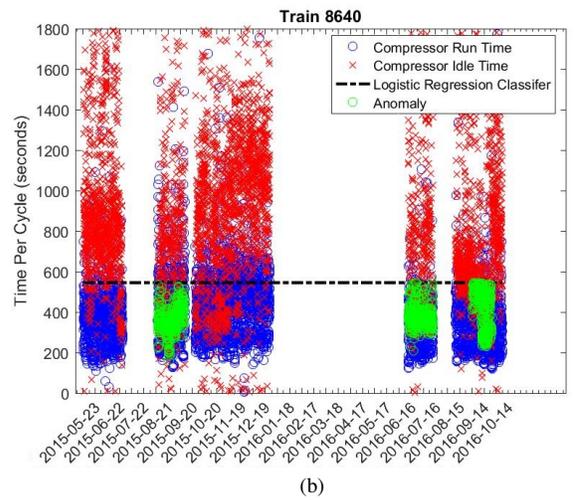
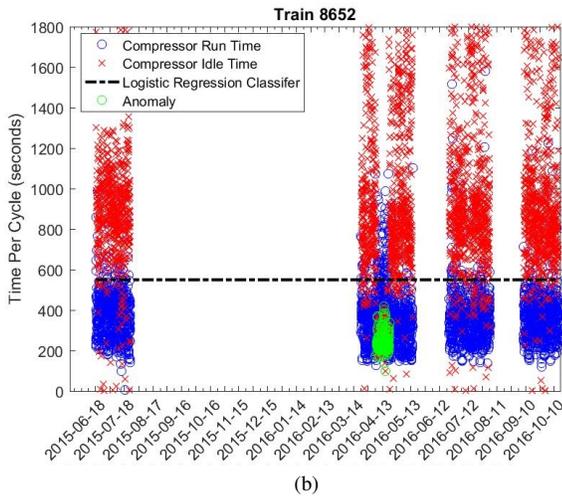
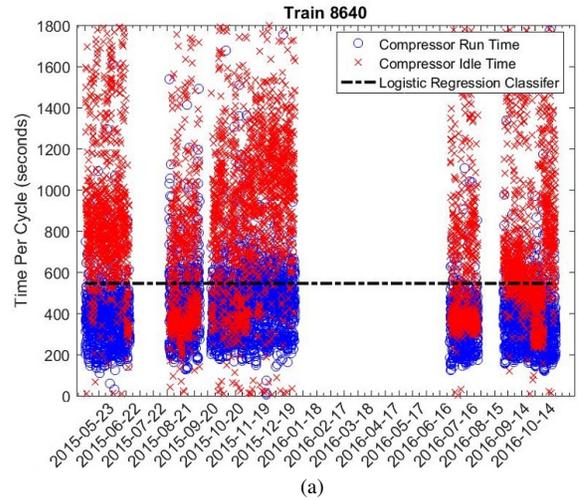
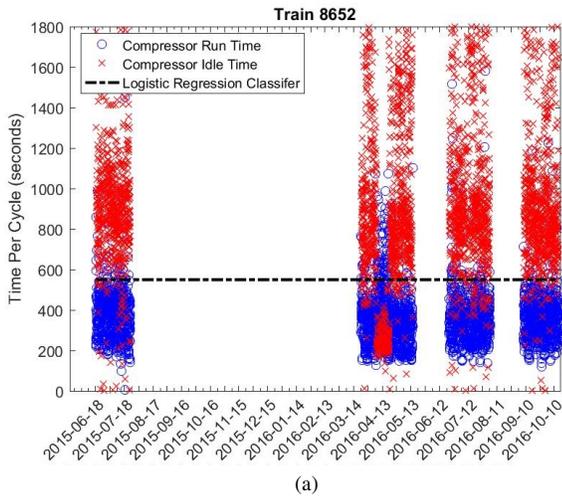
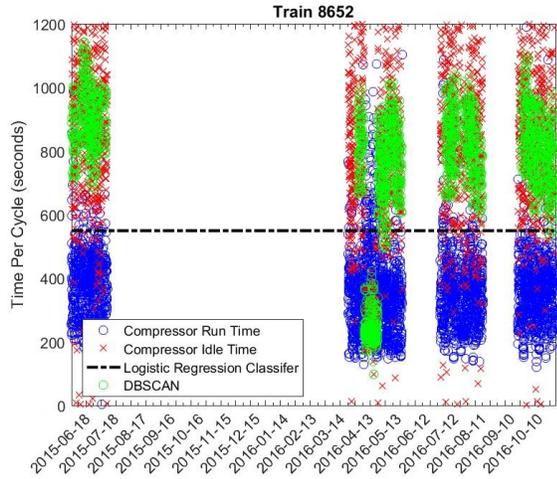
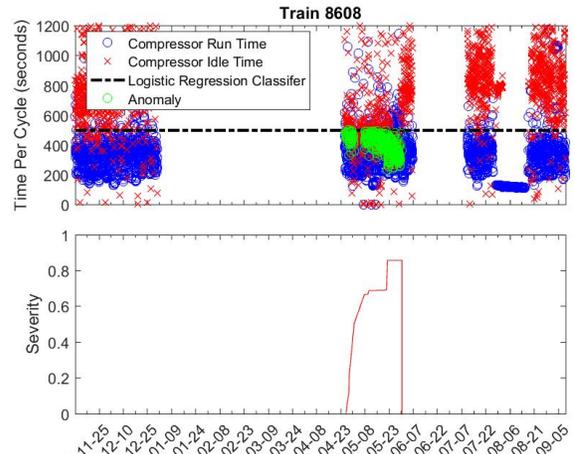


Figure 6. The (a) original compressor duration data and (b) result of the proposed air leakage detection of train 8652.

Figure 7. The (a) original compressor duration data and (b) result of the proposed air leakage detection of train 8640.



(a)



(b)

Figure 8. The results of (a) density-based clustering and (b) baseline anomaly detection for train 8652.

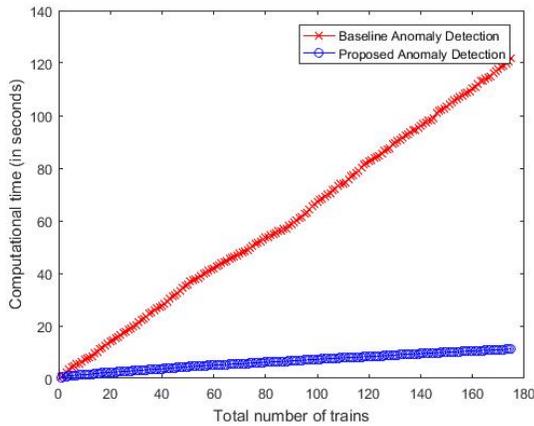
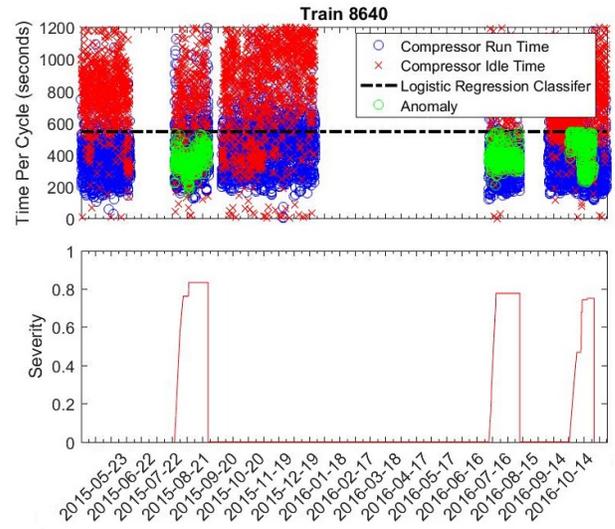
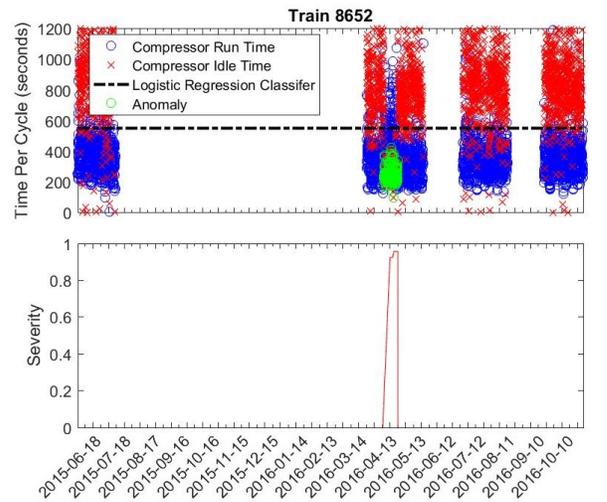


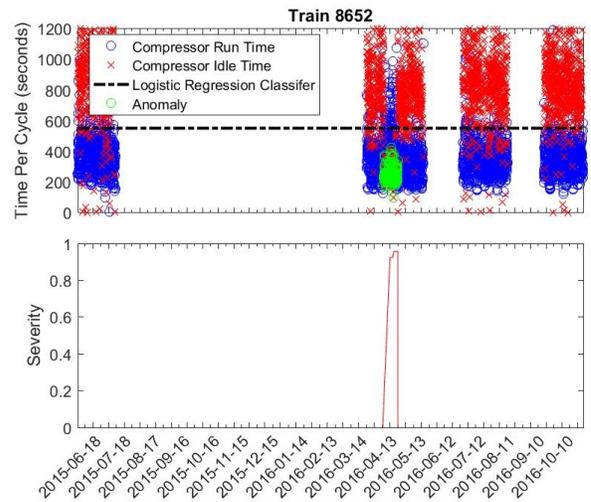
Figure 9. Comparison in computational cost between the baseline and our proposed anomaly detection method for air leakage.



(a)



(b)



(c)

Figure 10. Modeled severity index of detected air leakages in train (a) 8608, (b) 8640, and (c) 8652.

Time” data, thresholds derived from logistic regression classifiers, and the results of contextual anomaly detection for four trains with numbers 8608, 9580, 8652 and 8640, are presented in Fig. 4, Fig. 5, Fig. 6 and Fig. 7, respectively. From the figures, it can be observed that the logistic regression classifier identifies a proper boundary separating the ”Compressor Run Time” and ”Compressor Idle Time”. Please notice that the percentage of air leakage data points is relatively small and therefore it generally does not give a large impact on the logistic regression classifier.

In order to verify the effectiveness of pre-filtering using the logistic regression classifier, a baseline anomaly detection procedure is compared with the proposed procedure. The baseline anomaly detection procedure first adopted the DBSCAN clustering approach to find the dense regions on all ”Compressor Idle Time” data points as shown in Fig. 8(a). Then the logistic regression classifier is used for post-filtering to remove detected points above the threshold as given in Fig. 8(b). It can be observed in Fig. 8(b) that several points were wrongly detected as anomaly after post-filtering due to the high density in normal regions.

The confusion matrix of the results of our proposed contextual anomaly detection is given in Table 1. The values in brackets are those computed in the form of percentage. From the confusion matrix, our method for contextual air leakage detection in train braking pipes not only has a high detection capacity of 84% but also a very low false alarm ratio. Without the context modeled with the logistic regression classifier, there will be a large amount of false alarms if a density-based clustering approach is applied. Even applied with a post-filtering threshold as described in the baseline anomaly detection procedure, the amount of false alarms is also significant as shown in Table 2. Moreover, as observed in Fig. 9 the proposed anomaly detection procedure is computationally much more efficient than the baseline anomaly detection procedure since the proposed anomaly detection procedure pre-filters the ”Compressor Idle Time” which resulting in a small subset of data points considered for the density-based clustering while in the baseline anomaly detection procedure, all ”Compressor Idle Time” data points were used in density-based clustering.

#### 4.2. Prognostics

To incorporate the operational and maintenance planning requirements of Dutch Railways, the observation period  $D$  is set to 7 days in the experiments. As a result, the severity indexes derived from Eq.(12) for various air leakages cases are shown in Fig. 10.

To be able to estimate the remaining useful life, the severity indexes of all 5,844 correctly detected air leakage data points listed in Table 1 are computed. A Linear Regression prediction model  $f(\cdot)$  is then trained by 10-fold cross-validation

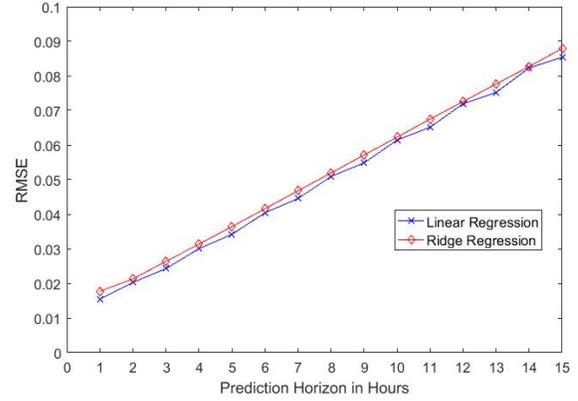


Figure 11. Comparison of RMSE in predicting time horizon from 1 to 15 hours using trained Linear Regression and Ridge Regression models.

on Root-Mean-Square Error (RMSE) by using the previous 5 hours data to estimate the severity index of the upcoming hour. That is,

$$\hat{S}_{C_k}^{t_{p+1}} = f(S_{C_k}^{t_p}, S_{C_k}^{t_{p-1}}, \dots, S_{C_k}^{t_{p-4}}). \quad (14)$$

This procedure can be iterated by using the previous estimated output as input such as

$$\hat{S}_{C_k}^{t_{p+2}} = f(\hat{S}_{C_k}^{t_{p+1}}, S_{C_k}^{t_p}, S_{C_k}^{t_{p-1}}, \dots, S_{C_k}^{t_{p-3}}) \quad (15)$$

up to a certain time horizon or until the pre-defined severity threshold  $\theta$  is reached.

To avoid collinearity issue in the input features, the experimental results of Ridge Regression is also provided in Fig. 11 in addition to Linear Regression. For severity prediction there is no collinearity issue observed since the RMSE of Linear Regression and Ridge Regression are very similar to each other. From Fig. 11 the 10-fold RMSE of the Linear Regression model in predicting severity index from the next first (operational) hour to the next fifteenth (operational) hour proportionally increases from 0.01 to 0.09. This indicates the certainty of severity prediction will decrease when the prediction time horizon increases. However, it still remains in a reasonable degree within 15 operational hours which is about 2 calendar days since a train is about 8 hours in operation each day in the Netherlands.

The methodology of remaining useful life prediction for an air leakage is illustrated in Fig. 12. After an air leakage was detected for 5 hours, the severity indexes of these 5 hours were then used to predict the upcoming 320 operational hours. The predicted severity indexes are indicated by the blue solid line while the actual severity indexes are indicated by the red solid line in the figure. Assume that the severity threshold  $\theta$  is set to 0.8, the severity index is predicted to reach 0.8 after 156 operational hours by our method. Therefore the predicted re-

Table 1. Confusion matrix of the experimental results of the proposed air leakage detection procedure

	Predicted: Air Leakage	Predicted: Normal
Actual: Air Leakage	5844 (0.840)	1113 (0.160)
Actual: Normal	665 (0.001)	625061 (0.999)

Table 2. Confusion matrix of the experimental results of the baseline anomaly detection procedure

	Predicted: Air Leakage	Predicted: Normal
Actual: Air Leakage	5892 (0.847)	1065 (0.153)
Actual: Normal	6831 (0.011)	618895 (0.989)

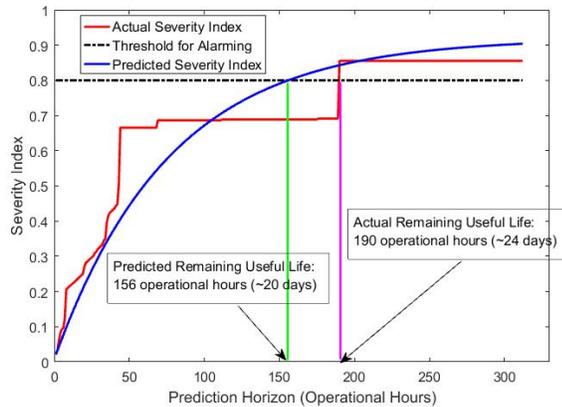


Figure 12. Illustration of remaining useful life prediction for an air leakage.

maining useful life is 156 operational hours while the actual remaining useful life is 190 hours.

## 5. DISCUSSION

In this paper, we presented a framework that not only can detect air leakages but also predicts the severity of air leakages so that action plans can be determined based on the severity level. The air leakage in train braking pipes is detected based on the compressor behavior data. In order to avoid false alarms, the logistic regression classifier is adopted to model context of "Compressor Run Time" and "Compressor Idle Time" and use the boundary separating these two classes as the threshold for pre-filtering candidate of anomalies. In order to detect anomalies according to their severity degree in a noisy dataset, a density-based clustering approach with a dynamic density threshold is developed. The experimental results have demonstrated that our method for contextual air leakage detection can detect air leakages effectively without generating false alarms. To facilitate the decision-making process and define the moment for action in operations and maintenance planning, the logistic function built on the compressor run time is further used together with the duration of an air leakage to model the severity of the air leakage. By building the prediction model on the severity, the remaining useful life of the air braking pipe until it reaches a certain

level of severity can be estimated.

## ACKNOWLEDGEMENT

The author would like to thank Mark Aalbers and Inge Kalsbeek from System Engineering Department of NS Techniek for their valuable contributions in explaining the mechanism of air circulation system in trains.

## REFERENCES

- Bart-ram, G., & Mahadevan, S. (2015). Probabilistic prognosis with dynamic bayesian networks. *International Journal of Prognostics and Health Management*, 6.
- Behera, S., & Rani, R. (2016). Comparative analysis of density based outlier detection techniques on breast cancer data using hadoop and map reduce. In *Proceedings of the second international conference on knowledge discovery and data mining*.
- Birant, D., & Kut, A. (2007). St-dbscan: An algorithm for clustering spatialtemporal data. *Data & Knowledge Engineering*, 60(1), 208-221.
- Biswas, G., Khorasgani, H., Stanje, G., Dubey, A., Deb, S., & Ghoshal, S. (2016). An approach to mode and anomaly detection with spacecraft telemetry data. *International Journal of Prognostics and Health Management*, 7.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3).
- Coble, J., Ramuhalli, P., Bond, L., Hines, J., & Upadhyaya, B. (2015). A review of prognostics and health management applications in nuclear power plants. *International Journal of Prognostics and Health Management*, 6.
- de Vos, J., & van Dongen, L. (2015). Performance centered maintenance as a core policy in strategic maintenance control. In *Procedia cirp* (Vol. 38, p. 255258).
- Eker, O., Camci, F., & Jennions, I. (2014). A similarity-based prognostics approach for remaining useful life prediction. In *Proceedings of the second european conference of the prognostics and health management society*.
- Ester, M., Kriegl, H. P., Sander, J., & Xu, X. W. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of*

*the second international conference on knowledge discovery and data mining* (p. 226-231).

- Hayes, M. A., & Capretz, M. (2015). Contextual anomaly detection framework for big sensor data. *Journal of Big Data*, 2(2).
- Khan, L., Awad, M., & Thuraisingham, B. (2007). A new intrusion detection system using support vector machines and hierarchical clustering. *The VLDB Journal*, 16(4), 507-521.
- Mahapatra, A., Srivastava, N., & Srivastava, J. (2012). Contextual anomaly detection in text data. *Algorithms*(4), 469-489.
- Mitchell, T. (1997). *Machine learning*. McGraw Hill.
- Ng, A. Y., & Jordan, M. I. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive baye. In *Advances in neural information processing systems 14* (p. 841-848).
- O'Donovan, P., Bruton, K., & O'Sullivan, D. (2016). Case study: The implementation of a data-driven industrial analytics methodology and platform for smart manufacturing. *International Journal of Prognostics and Health Management*, 7.
- Poot-Geertman, P., Huisman, B., & van Rijn, C. (2015). Application of a maintenance engineering decision method for railway operation: Managing fleet performance, cost, and risk. In *Safety and reliability of complex engineered systems: Esrel 2015* (p. 1863-1870).
- Prakash, G., Narasimhan, S., & Pandey, M. (2017). Condition based maintenance of low speed rolling element bearings using hidden markov model. *International Journal of Prognostics and Health Management*, 8.
- Qiao, G., & Weiss, B. (2016). Advancing measurement science to assess monitoring, diagnostics, and prognostics for manufacturing robotics. *International Journal of Prognostics and Health Management*, 7.
- Schenkendorf, R., & Groos, J. (2015). Global sensitivity analysis applied to model inversion problems: A contribution to rail condition monitoring. *International Journal of Prognostics and Health Management*, 6.
- Upadhyaya, S., & Singh, K. (2012). Nearest neighbour based outlier detection techniques. *International Journal of Computer Trends and Technology*, 3(2), 299-303.

## BIOGRAPHIES



**Wan-Jui Lee** received her Ph.D. degree from the Department of Electrical Engineering, National Sun Yat-Sen University, Taiwan, in 2008. From 2007 to 2011, she was a postdoctoral researcher in the Pattern Recognition Laboratory, Delft University of Technology. Currently she is a researcher of Dutch Railways focusing on predictive maintenance of train fleets on various components. Her main research interests include predictive maintenance, structural pattern recognition, (dis)similarity-based representation and machine learning. In 2016, she won the SmartRail Europe Innovation Awards - Runner up prize.