Fault detection method for railway point systems

Marius Vileiniskis, Rasa Remenyte-Prescott and Dovile Rama

Abstract

Failures of railway point systems (RPS) often lead to service delays or hazardous situations. A condition monitoring system can be used by railway infrastructure operators to detect the early signs of the deteriorated condition of RPS and prevent failures. This paper presents a methodology for early detection of the changes in the measurements of current drawn by the motor of the point operating equipment (POE) of RPS, which can be used to warn about a possible failure in the system. The proposed methodology uses the One Class Support Vector Machines (OCSVM) classification method with the similarity measure of Edit distance with Real Penalties (ERP). The technique has been developed taking into account specific features of the data of in-field RPS and therefore is able to detect the changes in the measurements of current of the POE with greater accuracy

Nottingham Transportation Engineering Centre, University of Nottingham, UK

Corresponding author

Rasa Remenyte-Prescott, Nottingham Transportation Engineering Centre, University of Nottingham, Nottingham, NG7 2RD, UK.

Email: R.Remenyte-Prescott@nottingham.ac.uk

compared to the commonly used threshold-based technique. The data from in-field RPS, which relate to incipient failures of RPS, were used after the deficiencies in the data labelling were removed using expert knowledge. In addition, the possible improvements in the proposed methodology were identified in order for it to be used as an automatic online condition monitoring system.

Keywords

Railway point systems, fault detection, point operating equipment, One Class Support Vector Machines, Edit distance with Real Penalties, fault classification

Introduction

An increasing demand for high capacity and cost-effective railway transportation requires railway systems to be fault tolerant and safe for both passengers and cargo transport. For example, in 2012/2013, Network Rail, UK railway infrastructure operator, was responsible for

8.7 million minutes of train delays; almost 7.4 million of delay minutes were recorded on passenger trains, while the remaining delay minutes were assigned to freight transport.[1] One of the main causes of train delays was failures of railway point systems (RPS), resulting in 575,679 minutes of total delays. If the target for the delays set out by the regulators is not met, fines are imposed to the infrastructure operator.

Most problems with RPS are associated with wear or misalignment of their components. Therefore, railway points require regular adjustments to compensate for wear or to correct any misalignments.[2] Even though these processes require a lot of time and are very costly, it is essential to keep RPS in a good condition. Poor maintenance and missed faults can lead to serious accidents. For example, 7 people were killed and 76 injured in the train derailment in Potters Bar in 2002, when the points failed while the train was passing through them.[3] Therefore, advanced detection and diagnoses of RPS failures can increase railway safety and improve operation through better maintenance planning.

RPS condition monitoring systems are increasingly used by infrastructure operators.[4-8] The main railway infrastructure operator in the UK reported that the RPS failures per year decreased from 8048 to 5053 failures between the years 2008/2009 and 2012/2013, after remote condition monitoring has been implemented and a better RPS maintenance regime has been applied.[1] The condition monitoring systems often contain an alarm system based on a threshold technique. The threshold of the observed parameter that is associated with the presence of a fault has to be predefined and it is usually chosen based on expert knowledge. Moreover, it requires adjustments due to seasonal changes and wear-out of the equipment. Commonly, the threshold technique cannot detect a fault at its earliest stage, i.e. before a failure actually occurs. The potential benefits of being able to detect incipient faults include fewer service disruptions due to better maintenance planning, increased safety and extended asset lifetime due to preventive maintenance.

A large variety of methods for fault detection of RPS is available in literature, such as: statistical analysis methods,[9-14] classification methods,[15-20] model based methods[21, 22] and other methods.[2, 23, 24]

Classification methods seem to show promising results in detecting RPS failures and have the potential to be a general fault detection method for different types of RPS and POE.

For example, Chamroukhi et al.[15] performed Mixture Discriminant Analysis on the extracted features from the electrical power signal to classify the signals as without and with defects with 95% accuracy. Asada and Roberts[16] proposed to extract features of the electrical active power of AC point machines using Discrete Wavelet Transform and use them as inputs to the SVM, which classified 86 operations of RPS with a 100% classification accuracy. More testing has been performed in a later work by Asada et al.[17], by considering failure modes with different severity, and achieving the perfect 100% classification accuracy. The same methodology was tested with the DC point machines by Asada and Roberts[18] giving similar results to those in the previous work. Eker et al.[19] performed fault detection with a combination of Principal Components Analysis and SVM based on the measurements of linear ruler to classify 20 operations of RPS as fault free or misalignment of drive rod with 100% classification accuracy. Thomas Böhm[20] suggested using temperature data to

reduce the number of false alarms, produced by an industrial fault detection system, based on a threshold technique, SIDIS W.[5] The amount of alerts was reduced by 16.84% and 91.83% for lower and higher level alerts respectively by applying correction values per each °C and then using a Neural Network to classify the movements.

Despite the fact that the approaches, described above, reported very high fault detection accuracy (almost 100%), small data sets (e.g. 86 operations of RPS in a study presented by Asada and Roberts[16]) used during the testing phase was one of their main deficiencies. Moreover, the majority of the approaches were tested using data obtained from the laboratory based RPS where failures were introduced manually, rather than occurring as a result of the natural wear out process. Manual fault simulation in the RPS led to the fact that the laboratory data did not contain some of the important features observed in data from the in-field RPS (see the section Main features of in-field data). Furthermore, the mentioned approaches needed a similar amount of data for different states (fault-free or with a certain type of fault) of RPS to be trained to detect a fault, which in practice is not always available. For example, data of 30

operations for representing each category of fault-free condition, left-hand underdriving and right-hand underdriving and 26 operations for representing each category of left-hand overdriving and right-hand overdriving were used in a study presented by Asada and Roberts[18].

This paper proposes a novel fault detection methodology developed and tested on the in-field RPS data, which is based on the classification method of One Class Support Vector Machines (OCSVM) and uses the elastic similarity measure of Edit distance with Real Penalties (ERP) to compare two time series. The main features of the proposed methodology are:

- Development and testing of the proposed methodology using in-field RPS data;
- Ability to detect the changes in the measurements of current earlier than the commonly used alarm threshold technique;
- Suitability for any type of RPS and POE;
- Ability to deal with the lack of historical data of failures.

Railway point systems

An RPS consist of electrical and mechanical installations, enabling railway trains to be guided from one track to another. An example of an RPS with a clamp lock can be seen in Figure 1.
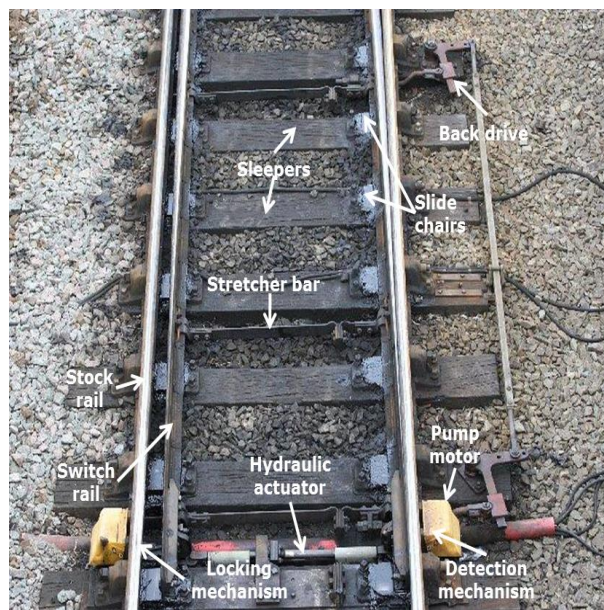

Figure 1. RPS fitted with a clamp lock[25] as POE

Various types of POE can be used within an RPS. In this study, RPS with clamplocks are considered, but the same fault detection methodology can be applied to point systems with other types of POE (e.g. HW 1000/2000 point machine, Westinghouse style 63 point

machine). Depending on the type of POE, some system components may differ, e.g. supplementary drives instead of a back drive, but they all have a motor to provide the drive of an actuator.

The operation of RPS can be described as follows:

1. A signal is received by the RPS to change the direction of track from normal to reverse or reverse to normal. The oil pump motor starts up (motor start-up phase in Figure 2).
2. The switch rails are unlocked (unlocking phase in Figure 2).
3. The corresponding pump motor pumps oil to the hydraulic actuator, which together with the back drive moves the switch rails (switch rail movement phase in Figure 2).
4. When the switch rails are moved to the desired position, they are locked and a signal is sent to the signaller that the railway point system has successfully completed the operation (locking phase in Figure 2). Due to a slight delay of the detection of locking the motor keeps running after the points have completed the movement. This is common to clamplock POE

and thus a 'shark fin' at the end of the operation is observed, as shown in Figure 2.

A typical trend of electric current drawn by the pump motor of a clamplock, demonstrating the four operation phases of the RPS, is shown in Figure 2.
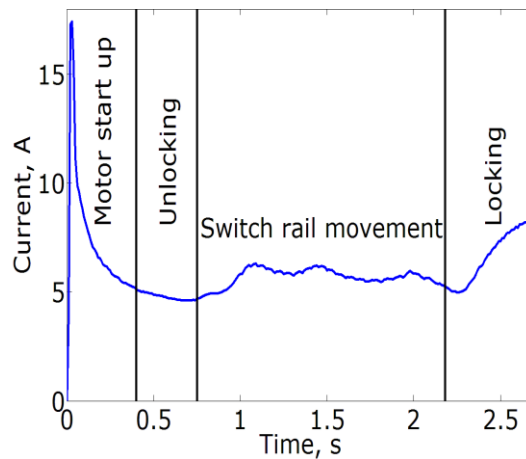

Figure 2. Operating phases of the clamplock

Main features of in-field data

The proposed method for fault detection has been developed using in-field data of RPS, such as shown in Figure 2. A number of specific

features have been observed in the in-field data of RPS. They have been of great importance while developing the proposed method for the analysis.

Diversity of operational behaviour within a group of the same type POE. Grouping RPS with the same type of POE for fault detection would be useful, since large data samples for each type of fault could be obtained. However, during this study it has been observed that each POE seems to have its own specific current trend, as illustrated in Figure 3.
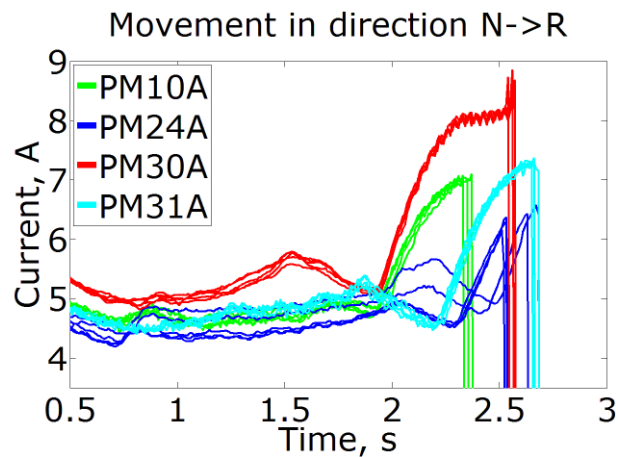

Figure 3. Measurements of current of different RPS with clamplock

Note that, the first 0.5 seconds, where the shapes are very similar, are ignored in this study, as shown in Figure 3. The issue of unique operational behaviour of each POE has also been raised by Tunley.[26]

Due to this diversity it is, therefore, possible that the current trends of fault-free movements of some point systems are similar to the current trends of movements with the faults of other point systems, and therefore the fault detection algorithm might identify them incorrectly. Thus, data sets of individual point systems, not their groups, had to be analysed in this study. Furthermore, there were far more data records for fault-free movements than those for movements with faults. Thus fault detection techniques that require a similar size of data samples for fault-free and each faulty state could not be used in this study, and an alternative approach was needed.

Variation in the duration of POE movements. Variation in the duration of the movement of the same POE is another feature observed in the in-field data. When solving a classification problem, the time series of measurements of current can be compared using a similarity measure. However, one of the standard similarity measure, known as the Euclidean distance,[27] cannot be used to compare two time series of different length. Thus, pre-processing of data has to be performed before the Euclidean distance can be applied or alternative similarity measures and techniques need to be used, such as Discrete Wavelet Transform (DWT)[16-18], Time Warp Edit Distance (TWED)[28] and etc. The use of ERP distance was proposed in this study since it mimics the way a human would intuitively compare the shapes of two time series, i.e. trying to match the corresponding peaks and valleys in the time series.

Variation in movements of fault-free RPS. The last feature of the in-field RPS data, which usually is not observed in the data obtained in the laboratory environment, is the variation in current trends of movements of fault-free RPS. Figure 4 illustrates a number of movements for one RPS when no faults have been recorded. Such variability might be explained by the deterioration of the point system, i.e. when a failure is due to occur in the future, influence of weather conditions or preventive maintenance actions, i.e. before a failure occurs.
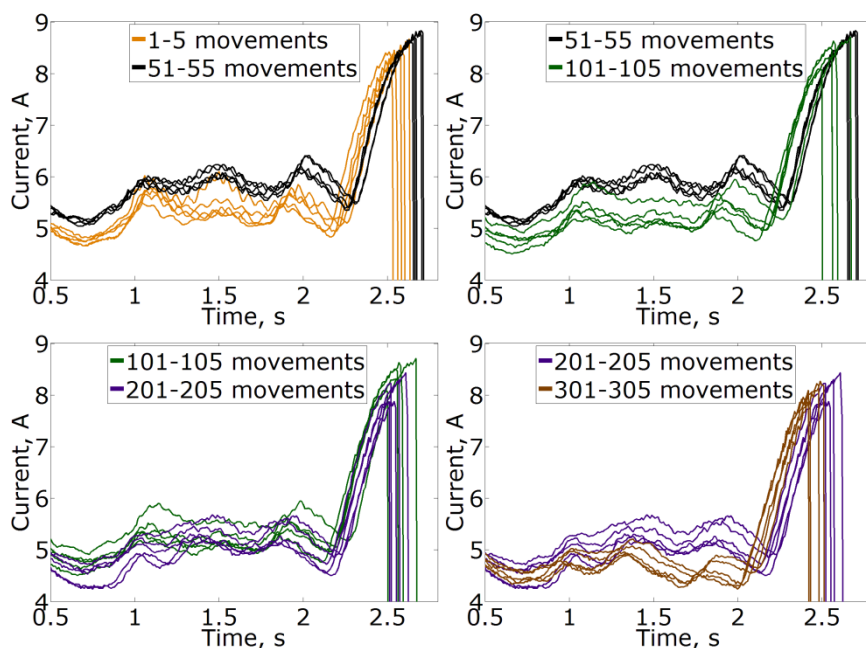

Figure 4. Variation in movements of fault-free point system

This variation needs to be captured appropriately, especially if current trends of fault-free movements and movements with faults are similar making it difficult to distinguish between them. Therefore, in this study the dataset was divided into four seasons (winter, spring, summer and autumn) and a similar amount of movements was randomly chosen for each season to be used as input data for model development. The methodology to detect failures of RPS is presented in the next section.

Proposed methodology

Problem definition and proposed approach

The problem, considered in this study, is to perform fault detection of in-field RPS given the measurements of current (also referred as current trends in the paper) of each POE operation. The idea of the proposed methodology is to consider fault detection of RPS as a classification task, known as one class classification or outlier/novelty

detection. Such a classification task needs certain features, on which the classification decision can be made. In this study, the similarity between two time series of measurements of current of POE is used to differentiate between the fault-free and faulty operations of RPS. A changing shape and an increase in the absolute values of current trends might indicate a failure of the RPS, thus these parameters are considered when determining the similarity of time series. The similarity between two time series is compared by the Euclidean and ERP metrics. The ERP metric tries to align the current trends in time, thus it also represents the difference of shapes of current trends, as well as difference in absolute values. Historical data of current trends of in-field POE is used to train the OCSVM method to form a decision boundary based on the similarities of the most representative fault-free operations from historical data. This boundary is formed by Support Vectors (SV) in the OCSVM. The OCSVM method has several parameters, which are optimised using a grid search approach based on the classification accuracy (for fault-free and faulty operations individually) of OCSVM, which is further described in section "Determining the best performing model". The SV of the optimised

OCSVM are then used to make a decision between fault-free and faulty operations. That is, if the similarity between the measurements of current of new POE operation and the SV does not fall within the formed decision boundary, that operation of POE is considered to be faulty.

Steps of the methodology

The proposed methodology can be summarised in the following steps:

1. Label the current trends of RPS movements as fault-free or with faults.

2. Pre-process the current trends with uniform scaling (optional for ERP, necessary for the Euclidean distance), exponential smoothing (optional) or down sampling (optional) techniques.

3. Calculate the pairwise similarity measures for all of the current trends.

4. Select a grid of OCSVM parameters and find the ones, which give the best classification accuracy of OCSVM. Use the best performing model for the fault detection of RPS.

Each of these steps is explained in more detail in the following sections by presenting the data labelling approaches first, then introducing to data pre-processing techniques, OCSVM theory, ERP metric, and finally giving the details of an algorithm to determine the best performing OCSVM model.

Labelling of fault-free movements and movements with faults

Correct labelling of data used in a classification algorithm is of critical importance, since mislabelled data could lead to poor classification accuracy. The labelling of data was carried in two stages. First, an automatic labelling of data was carried out in the preliminary analysis:

1. Label the data according to records of failures, i.e. the movements carried out between the date of fault occurrence and the date of fault rectification were labelled faulty, other movements were labelled good.

2. Label the data according to records of alarms (based on a threshold technique), i.e. the movements that raised an alarm were labelled faulty and other movements were labelled good.

Inconsistences resulting from the automatic labelling, such as inconsistent labelling of current time series with similar shapes, were then eliminated by relabeling the time series were necessary using expert knowledge supported by both failure and alarm dates.

Data pre-processing techniques

Several data pre-processing techniques have been used to speed up the classification process and also to achieve a better alignment of current time series for when comparing movements (due to the variation in durations of the movements). The techniques explored are: uniform scaling,[29] exponential smoothing[30] and down-sampling.

Uniform scaling. In order to compare results obtained using the ERP and the Euclidean distance, the time series have to be rescaled to become of the same length. In this study, the rescaling was done by uniformly scaling[29] the time series to an average length of time series of all movements of the RPS, according to the formula, given in equation (1).

$$a_i^R = a_{[i \cdot N/M]}, i = 1, 2, \ldots, M, \tag{1}$$

where a_i^R is the i-th element of the rescaled time series, N is the length of the original time series, M is the length of the rescaled time series, $\lceil i \rceil$ is the ceiling function of i, $a_{\lceil i \cdot N/M \rceil}$ is the $\left\lceil i \cdot \frac{N}{M} \right\rceil$ element of original time series.

Exponential smoothing. There are various methods used to remove the noise from the measurements provided by sensors, e.g. moving average filter,[30] exponential smoothing[30] and Kalman filter.[31] In this study, the exponential smoothing was used due to its simplicity, as shown in equation (2).

$$a_j^S = a_{j-1}^S + \alpha\big(a_j - a_{j-1}^S\big), j = 2,3,\dots,N,\ a_1^S = a_1, 0 < \alpha < 1, \tag{2}$$

where a_j is the jth data point of the original time series, a_j^S is the jth data point of the smoothed time series, α is the smoothing factor.

Down-sampling. One of the simple ways to speed up the calculations of similarity measures is to down-sample the original time series. In this study the linear down-sampling was used when every k-th data point from the original time series was removed. Here $k = 1, 2, \ldots, 6$ is the down-sampling factor. The speed of the calculations can be improved, however, an analyst should be aware of the potential loss of the accuracy of the classification.

One class support vector machines

A lack of historical data of failures for RPS was observed in the data available in this study, as mentioned in the section "Main features of in-field data". One class support vector machines were chosen due to the ability to classify such highly imbalanced data.

Scholkopf et al.[32] suggested this modification to SVM in order to make it suitable to be trained on one class data when the outliers from a given data set are separated. The idea of the method is to form a decision function $f(x)$, as shown in equation (3), that has a

value of +1 for the majority of the data and a value of -1 for the outliers.

$$f(x) = sgn\left(\sum_i \alpha_i k(x, x_i) - \rho\right),$$ (3)

where sgn is a sign function, defined as $sgn(x) = \begin{cases} -1 \ if \ x < 0 \\ 1 \ if \ x \geq 0 \end{cases}$, ρ is the bias term of the hyperplane, α_i is the weight of ith support vector, $k(x, x_i)$ is the kernel function between a new point x and ith support vector x_i.

The decision function forms an optimal separating hyper-plane with the values of parameters v and γ, carefully chosen for optimum performance (as discussed in the next section), when a case study is considered.

A radial basis function (RBF) kernel is commonly used in literature when SVM are used for fault detection.[16, 33, 34] The RBF kernel has also been used in this study, since it allows to choose a distance function used in the kernel:[35]

$$k(x, y) = \exp(-\gamma D^2(x, y)),$$ (4)

where γ is the width parameter, $D(x,y)$ is the selected distance function, x and y are the feature vectors.

Any function can be used as $D(x,y)$, as long as it is a metric.[36, 37] A commonly used distance function is the Euclidean distance.[32] In this study, an elastic metric ERP was proposed to be used as an alternative to the Euclidean distance. This way, no rescaling of the data would be necessary (rescaling was still performed in the case when the performance of ERP was compared with the Euclidean distance) and, if the phases of the current time series were not well aligned in time, ERP would provide a better matching of the time series. Effects on the classification accuracy were analysed when both distance functions were used and the results are provided in the Case study section.

Edit distance with Real Penalties

This similarity measure, proposed by Chen and Ng,[38] can support a local time shifting observed in the in-field data. Thus, ERP is more robust in comparing similarity of shapes of two time series, even if

they are not perfectly aligned in time. This measure is also a metric, which is the necessary requirement for a function to be used in the kernel function of the OCSVM.

Consider two time series $A = (a_1, a_2, \ldots, a_N)$ and $B = (b_1, b_2, \ldots, b_M)$, where $N \neq M$. Before calculating the ERP distance the normalisation of the time series needs to be carried out. Denote a subsequence of A as $A_i^p = (a_i, a_{i+1}, \ldots, a_p)$ and a subsequence of B as $B_i^q = (b_i, b_{i+1}, \ldots, b_q)$. Then the ERP distance can be defined as follows:

$$\delta_{ERP}(A,B) = \begin{cases} \displaystyle\sum_{i=1}^{N} d_{ERP}(a_i, gap), & if\ M = 0, \\ \displaystyle\sum_{i=1}^{M} d_{ERP}(b_i, gap), & if\ N = 0, \\ \min \begin{cases} \delta_{ERP}(A_1^{N-1}, B_1^M) + d_{ERP}(a_N, gap), \\ \delta_{ERP}(A_1^{N-1}, B_1^{M-1}) + d_{ERP}(a_N, b_N), \\ \delta_{ERP}(A_1^N, B_1^{M-1}) + d_{ERP}(b_M, gap) \end{cases}, & otherwise\ . \end{cases} \tag{5}$$

where

$$d_{ERP}(a,b) = \begin{cases} |a - b|\ if\ a, b\ is\ not\ a\ gap, \\ |a - g| & if\ b\ is\ a\ gap, \\ |b - g| & if\ a\ is\ a\ gap, \end{cases} \tag{6}$$

and g is a gap penalty.

This metric uses a real penalty between two non-gap elements and a constant penalty value for the gap. Chen and Ng proved that for any constant value of g ERP is a metric, thus they suggest using $g = 0$.

Determining the best performing model

Modified 5-fold cross-validation technique. A modification of a common technique to assess the accuracy of classification models, known as 5-fold cross-validation procedure was used in this study.[18, 38] The modification was introduced due to the imbalance of data in each class. Only the data of movements identified as fault-free was divided into 5 folds (all data is divided into 5 folds in the original algorithm). 4 folds (or parts) of data of fault-free movements were used to train the model and 1 fold, followed by all the movements with faults, was used for testing.

If the data imbalance between two classes used for testing is large, the total classification accuracy of the OCSVM method might be misleading, since the classification accuracy of fault-free movements might overwhelm the classification accuracy of movements with faults. The classification accuracy in each ith fold of fault-free and

faulty movements, $P_G(i)$ and $P_F(i)$ respectively, was evaluated in this study by equations (7) and (8).

$$P_G(i) = \frac{1}{N_{Gi}} \sum_{j=1}^{N_{Gi}} I(L_j^c) \cdot 100\%, \qquad (7)$$

$$P_F(i) = \frac{1}{N_F} \sum_{j=1}^{N_F} I(L_j^c) \cdot 100\%, \qquad (8)$$

where $P_G(i)$ is the classification accuracy of fault-free movements in ith fold, N_{Gi} is the total number of fault-free movements in ith fold, $P_F(i)$ is the classification accuracy of movements with faults, when ith fold of fault-free movements is used for testing, N_F is the total number of movements with faults, I is the indicator function defined by equation (9).

$$I(L^c) = \begin{cases} 0 \ if \ L^c \neq L^t \\ 1 \ if \ L^c = L^t, \end{cases} \qquad (9)$$

where L^c is the label given by the classification algorithm and L^t is the true label.

The 5-fold cross-validation accuracies of fault-free and faulty movements, P_G and P_F respectively, were obtained using equations (10) and (11).

$$P_G = \frac{1}{5} \sum_{i=1}^{5} P_G(i), \tag{10}$$

where P_G is the 5-fold cross-validation accuracy of fault-free movements, $P_G(i)$ is as defined by equation (7).

$$P_F = \frac{1}{5} \sum_{i=1}^{5} P_F(i), \tag{11}$$

where P_F is the 5-fold cross-validation accuracy of faulty movements, $P_F(i)$ is as defined by equation (8).

Since the division into 5 folds is done randomly, the accuracy of the classification algorithm given by the cross-validation estimate (equations (10) and (11)) depends on this random division. The best estimate of the 5-fold cross-validation would be obtained by going through all of the possible random divisions into folds, however this is computationally too expensive to perform. Repeating the division into folds multiple times should provide a good estimate of the accuracy, as suggested by Kohavi[39]. Thus, repetitive random divisions were performed to obtain the 5-fold cross-validation estimate of the OCSVM accuracy.

Determination of optimal model parameters. The classification accuracy of the OCSVM model depends on the OCSVM parameters γ and ν. A pair of OCSVM parameters (ν, γ), that gives the best classification accuracy was obtained using a grid search technique.[34] An array of values of each OCSVM parameter ν and γ were chosen. All the possible pairs (ν, γ) were used to obtain the 5-fold cross-validation estimates of the OCSVM accuracy. An algorithm to find the pair (ν, γ) that led to the best classification accuracy of the OCSVM model can be summarised in Figure 5.
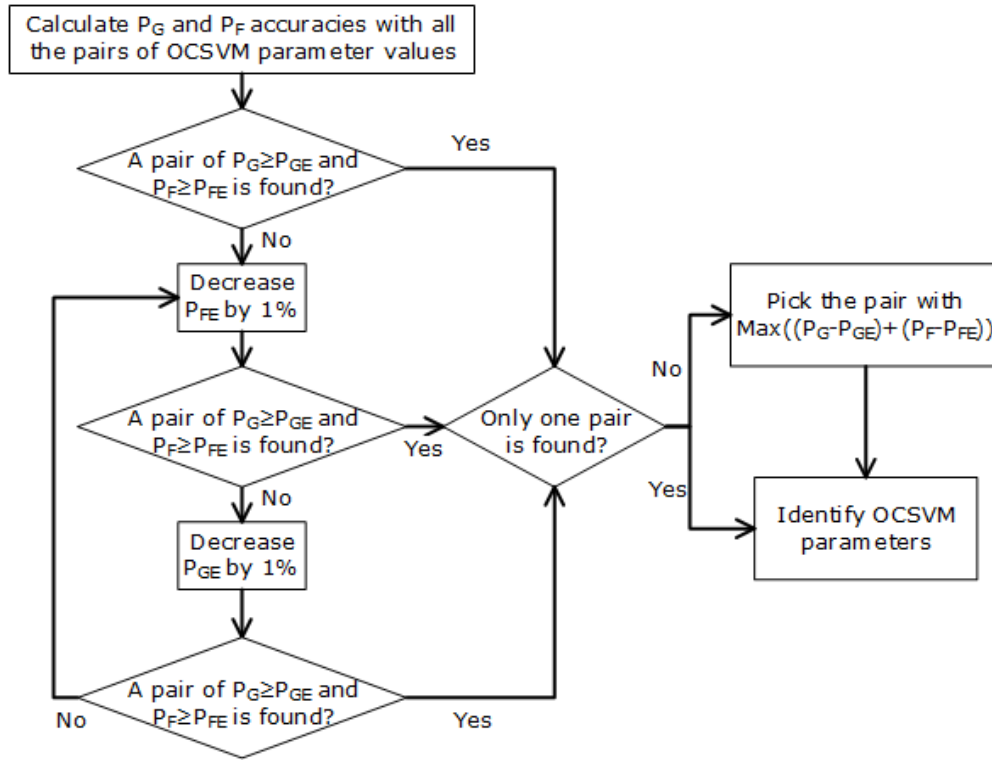
Figure 5. Algorithm to find optimal values of OCSVM parameters

Using the algorithm, the expected classification accuracies of fault-free and faulty movements, P_{GE} and P_{FE} respectively, are set and a search for a model, which would satisfy a system of inequalities (12), is performed.

$$\begin{cases} P_G \geq P_{GE}, \\ P_F \geq P_{FE}. \end{cases} \tag{12}$$

If no model satisfying the inequalities (12) is found, the P_{FE} and P_{GE} are decreased by 1% and the search is repeated until at least one model is found. The decrease in accuracy of 1% is chosen, since a smaller change is considered insignificant due to performing the cross-validation technique with multiple random divisions, as explained in previous section. If several models are found, the one with the highest difference between the expected and achieved accuracies (see equation (13)) is chosen.

$$\max\big((P_G - P_{GE}) + (P_F - P_{FE})\big) \tag{13}$$

This algorithm favours the values of OCSVM parameters, which have similar classification accuracy for both classes. Depending on the requirements of the fault detection process, this algorithm could be modified to place more emphasis on improving the accuracy of one class, for example, better accuracy of classification of movements with faults.

Case study

For the development of the proposed methodology, data from several RPS have been analysed. Here, the results of the analysis of a single RPS will be presented to illustrate the proposed methodology. A calendar year for 2012 was selected, with data from 589 movements in normal to reverse operation. 5 faults (3 cases of obstructions, 1 case of contamination and 1 case of crushed cable) were recorded, which corresponded to 58 movements with faults and 47 movements that raised an alarm. Following the labelling process discussed in the section "Labelling of fault-free movements and movements with faults", 482 movements were labelled as fault-free and 107 movements were labelled as faulty.

Application of the methodology

The grid of the values of the parameters was chosen as follows: $v = [0.01, 0.05, 0.1, 0.15, 0.2]$ and $\gamma = 10^n$, where $n = [-3, -2, -1, 0, 1, 2, 3]$. 100 random divisions of data into folds for the 5-fold cross-validation were performed for each pair of parameter values (v, γ) and the classification accuracies were averaged over the number of runs.

Different data pre-processing techniques were applied, such as the exponential smoothing, with the values $\alpha = [0.2, 0.3, 0.4, 0.5, 1]$, where 1 stands for no smoothing and the down-sampling with the values $k = [1, 2, 3, 4, 5, 6]$, where 1 stands for no down-sampling. For every pair of (α, k), which gives 30 combinations in total, the best performing OCSVM model was found using the algorithm summarised in Figure 5. Performance of the methodology using different distance function was also compared.

Results of the methodology

The performance of OCSVM models with different data pre-processing techniques was summarised by 4 levels of accuracy, expressed in colours, as given in Table 1. 95% level of accuracy has been chosen subjectively following an assumption that 5% misclassification rate represents good performance of the method.

Note, that uniform scaling is needed in order to be able to compare the results of Euclidean and ERP similarity measures.

Table 1. Colour-coding of classification accuracy

Colour	Description	Values
	desired classification accuracy	$P_F \geq 95$ and $P_G \geq 95$
	good classification accuracy	at least $P_F \geq 90$ and $P_G \geq 90$
	average classification accuracy	at least $P_F \geq 85$ and $P_G \geq 85$
	unacceptable classification accuracy	$P_F < 85$ or $P_G < 85$

Comparison of Euclidean distance and ERP metric. The Euclidean distance outperformed the ERP metric (both on non-scaled (see Table 3) and rescaled data (see Table 4) in terms of average classification accuracy and the number of models achieving better classification rates. When Euclidean distance was used, 94% and 86% classification accuracies for P_F and P_G respectively were obtained, and the standard deviation was less than 2%. Note that, the way the best model is found and P_G and P_F is selected will influence the results of the comparison between the two methods.

Table 2. Classification rates using Euclidean distance on rescaled data for PM9A

δ_E^R	P$_G$≥95	90≤P$_G$<95	85≤P$_G$<90	P$_G$<85
P$_F$≥95	0	0	0	0
90≤P$_F$<95	0	0	25	4
85≤P$_F$<90	0	0	0	0
P$_F$<85	0	0	0	1
$E(P_F) \pm \sigma(P_F)$	93.72%±1.89%			
$E(P_G) \pm \sigma(P_G)$	86.42%±1.06%			

Table 3. Classification rates with ERP distance (non-scaled data) for

PM9A

δ_{ERP}	P$_G$≥95	90≤P$_G$<95	85≤P$_G$<90	P$_G$<85
P$_F$≥95	0	0	0	0
90≤P$_F$<95	0	0	0	0
85≤P$_F$<90	0	0	22	1
P$_F$<85	0	0	7	0
$E(P_F) \pm \sigma(P_F)$	86.07%±1.20%			
$E(P_G) \pm \sigma(P_G)$	88.74%±0.89%			

Table 4. Classification rates with ERP distance (rescaled data) for PM9A

δ_{ERP}^{R}	$P_G \geq 95$	$90 \leq P_G < 95$	$85 \leq P_G < 90$	$P_G < 85$
$P_F \geq 95$	0	0	0	3
$90 \leq P_F < 95$	0	0	9	3
$85 \leq P_F < 90$	0	0	0	7
$P_F < 85$	0	0	0	8
$E(P_F) \pm \sigma(P_F)$	89.71%±4.87%			
$E(P_G) \pm \sigma(P_G)$	84.94%±0.85%			

However, the good performance of Euclidean distance can be explained by looking at the data of the RPS used in the case study (see Figure 6). One can observe that peaks and valleys in the time series (on the left hand side of Figure 6) before the rescaling were matched very well. Furthermore, after the uniform rescaling was applied (on the right hand side of Figure 6), these peaks and valleys were even better aligned.
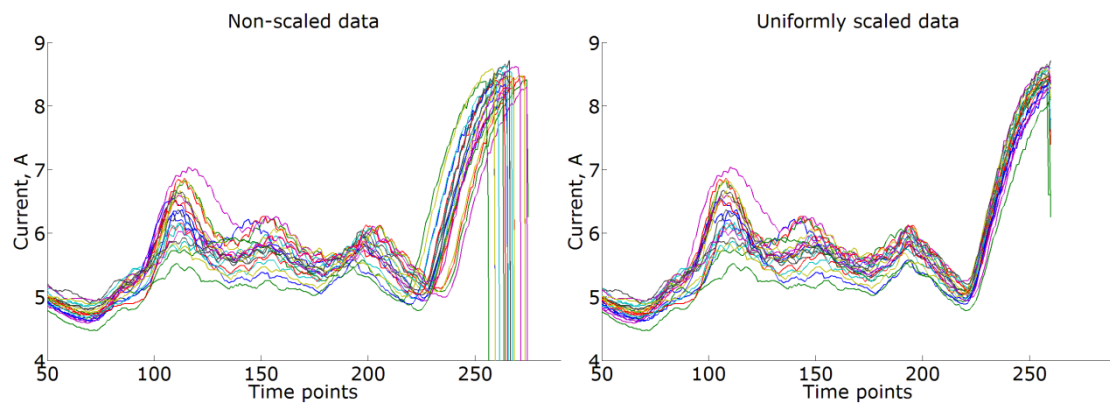
Figure 6. Non-scaled (on the left hand side) and uniformly scaled (on the right hand side) data

Thus, in this case the Euclidean distance gives the desired alignment of two time series. It performs better than the elastic metric, which can introduce unnecessary warping in the time series and make the alignment poor. However, if peaks and valleys did not match well in the first place, it is likely that poor alignment would be achieved using the Euclidean distance and the ERP metric would perform better.

It is important to note that the classification accuracies P_G and P_F of different similarity measures need to be carefully interpreted. The main reason for this is that the classification accuracies, P_G and P_F, depend on the way the data is labelled. Thus, one has to fully trust

that the data is labelled well, which might not always be the case. If the data is poorly labelled, the similarity measure might produce correct results of similarity between two time series, but still give poor classification rates when used in OCSVM.

Effects of data pre-processing techniques. Effects of the data pre-processing techniques were further examined in terms of the difference in the classification accuracies when no or one of the data pre-processing techniques were used.

For example, the biggest effect of exponential smoothing (maximum of 9%) was noticed when the ERP was used on rescaled data. In this case a considerable increase in both P_G and P_F (see Table 5) was observed, while there was a slight increase (maximum 2.5%) in the classification accuracy (P_G and P_F) for the Euclidean distance. For the ERP used on non-scaled data the smoothing of the time series increased the accuracy of P_G (maximum 0.16%) negligibly and decreased the accuracy of P_F (maximum 1.97%) slightly.

Table 5. Change in accuracy (%) for the values of smoothing factor α

Similarity measure	$\alpha = 0.2$ vs $\alpha = 1$		$\alpha = 0.3$ vs $\alpha = 1$		$\alpha = 0.4$ vs $\alpha = 1$		$\alpha = 0.5$ vs $\alpha = 1$	
	P_G	P_F	P_G	P_F	P_G	P_F	P_G	P_F
δ_E^R	+2.20	+0.48	+2.50	+0.40	+2.53	+0.23	+2.01	+0.50
δ_{ERP}	+0.16	-1.97	+0.13	-1.17	+0.08	-0.82	+0.06	-0.57
δ_{ERP}^R	+2.86	+8.08	+2.46	+7.52	+1.93	+8.23	+1.54	+9.04

Since in many cases the classification accuracy increased, the exponential smoothing on the time series should be performed to reduce the noise level in raw data.

The analysis of influence of down-sampling factor k on the classification accuracy (P_G and P_F) showed that in the most of the cases the classification accuracy dropped, when the data was down-sampled as can be seen from Table 6.

Table 6. Change in accuracy (%) for the values of down-sampling factor k

	$k = 2$ vs $k = 1$		$k = 3$ vs $k = 1$		$k = 4$ vs $k = 1$		$k = 5$ vs $k = 1$		$k = 6$ vs $k = 1$	
	P$_G$	P$_F$	P$_G$	P$_F$	P$_G$	P$_F$	P$_G$	P$_F$	P$_G$	P$_F$
δ_E	+0.09	+0.94	+0.06	-9.56	-0.47	+1.55	+2.26	+0.78	-0.34	+0.67
δ_{ERP}	+0.09	-0.04	-0.20	-0.89	-0.18	-0.44	-0.38	-2.78	-0.29	-0.48
δ_{ERP}^{R}	0	-0.41	+0.05	-0.37	+0.08	-1.33	+0.11	+1.99	-0.08	-3.94

However, this was expected as fewer data points were used in the computation of similarity measures. The down-sampling led to the vastly reduced time needed to calculate the ERP metric. For example, the calculation time, needed to calculate similarity measures for all the movements used in the presented example, decreased when the down-sampling factor increased, as shown in Figure 7.
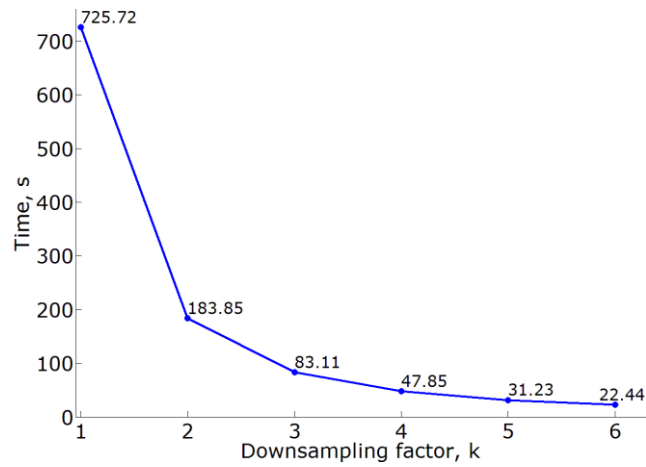
Figure 7. Computing time of ERP metric for different down-sampling factors

As shown in Figure 7, the experimental computing time decreased from 725 seconds when $k = 1$ (no down-sampling) to 183 seconds when $k = 2$, ($\frac{725}{183} \approx 3.96 \approx k^2$) which is similar to a theoretical decrease in computing time (k^2 times). Thus, data down-sampling should be considered if the proposed method was to be used as an online fault detection algorithm, however the trade-off analysis between the accuracy and speed of the proposed method should be carried out.

Comparison of alarm threshold technique and OCSVM model. A direct comparison of the accuracy of the commonly used alarm system based on a threshold technique and the proposed method is not straightforward, since the former method identifies movements that are very dissimilar (and usually warn about a failure), whereas the latter method identifies more subtle changes in movements which might indicate a deteriorated state (a failure that has not yet occurred).

For example, consider a situation when an alarm led to a failure recorded as crushed cable, and the alarm was raised when an increase of current and exceedance in the duration was observed, as red line shown in Figure 8.
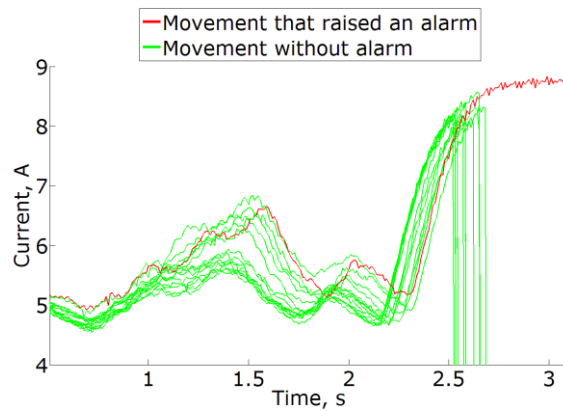
Figure 8. Movement that raised an alarm with the alarm system

Using the proposed methodology, a few earlier movements are also classified as signs of conditions for failure, since a similar increase in current values is observed, whereas an alarm has not been raised for these movements due to the duration staying within the set limits.

On the left-hand side of Figure 9 it can be seen that a group of movements is identified as with faults (before an alarm is raised for movement #527). On the right-hand side a more detailed illustration is given, demonstrating that the first warning of the fault occurrence would be given for movement #517, i.e. well before the failure occurs.
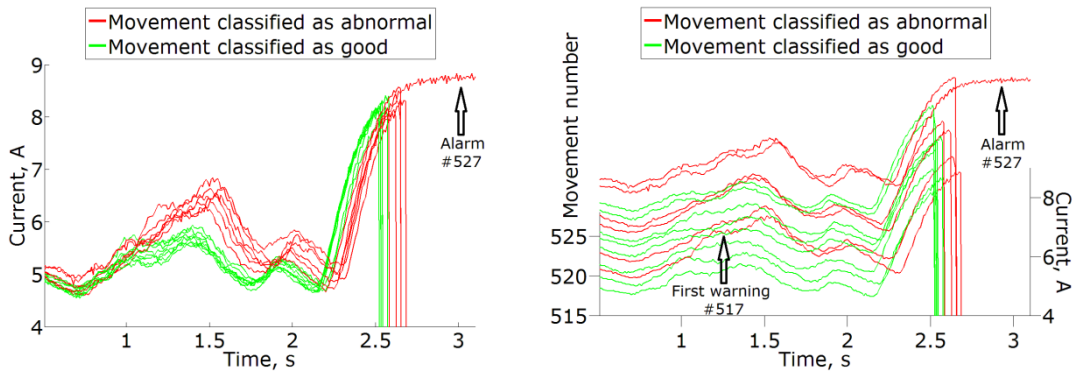
Figure 9. Movement classification with the proposed methodology

The decision values of the OCSVM method are plotted in Figure 10. They indicate that the movements that were classified as abnormal (red dots in the Figure 10) were different from the ones that were classified as good (green triangles in the Figure 10), since the decision values were in different sides of the decision boundary (blue line in the Figure 10).
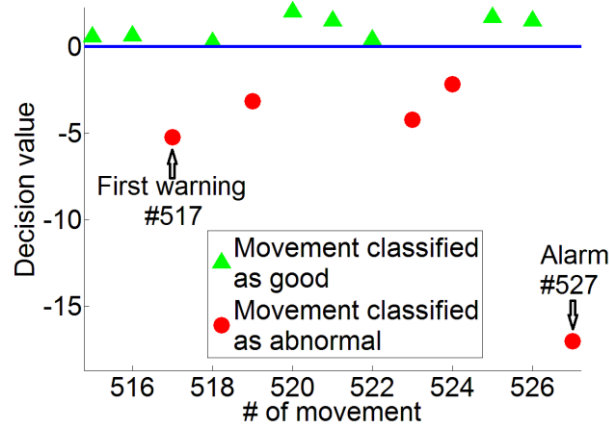
Figure 10. OCSVM decision values for a selected example

Similar situations were observed for all the faults in this dataset. It can be concluded that using the proposed methodology the changes in the measurements of current can be detected before an alarm is raised by the threshold technique. Such early warnings can give a good opportunity to plan maintenance actions in advance and act before a failure occurs. However, more data analysis of situations when changes in the current measurements lead to a failure and are not caused by seasonal or operational changes, is needed, so that the number of unnecessary warnings can be minimised.

Conclusions and future work

A novel fault detection methodology, when the fault detection was considered as a classification task, is presented in this paper. It was developed taking into account the features of in-field RPS data and was tested using the data obtained throughout a year period from an in-field RPS. The results are promising, but further development of the proposed methodology is necessary in order to use it in practice.

The main contributions of this proposed methodology in the area of fault detection methods of RPS consist of:

- Development and testing of the proposed methodology using data of in-field RPS instead of commonly researched lab-based RPS.

- The novel application of the OCSVM model in this area in order to deal with the lack of historical data of failures.

- The novel use of the ERP elastic metric in the OCSVM in this area, as an alternative to the Euclidean distance, for the comparison of two time series of measurements of current, which differ in length.

In order to be able to use the proposed methodology in practice, several aspects should be considered further.

1. The process of training and testing the OCSVM model cannot be fully automated in its current form. This issue could be resolved by combining expert knowledge and an automated labelling system of movements before the model is used in practice.

2. As demonstrated by this study, it might be unnecessary to use the elastic metrics for comparison of current trends and the commonly used approach of the Euclidean distance can be applied instead, especially, when the time series are aligned in time well. The choice of the similarity measure can be based on an initial inspection of the rescaled data, in order to achieve the required accuracy.

3. If the ERP elastic metric is used, the data should be down-sampled to speed up the calculations, given small decreases in classification accuracy rates are acceptable.

4. The proposed methodology should be used in combination with expert knowledge, since the changes in the measurements of current might not always indicate a failure of RPS in practice.

A number of additional features have been considered which could be introduced to advance the methodology further. First of all, alternative inputs to the OCSVM model could be chosen, for example, similarity measures with selected recent movements of RPS. This approach would make the method faster, because there would be no need to calculate the similarity measures with all of the movements, but rather with the several selected ones. Moreover, it might help to deal with the variation in the movements of fault-free RPS, since only some previous movements would be compared and they should be representing similar seasonal or operational conditions. This could lead to more accurate results of fault detection. Secondly, duration of RPS inactivity could also be considered as an additional input to the OCSVM model, since this would possibly reduce the number of falsely detected failures due to the rise of absolute values of current trends for POE which are not operated frequently. Thirdly, the movement could be analysed in terms of time to closure than to the time from the start in order to achieve better accuracy. Finally, individual phases of the movement could be compared separately. This would

potentially help to develop a fault diagnostics method for RPS, so that causes of failures can be localised.

Acknowledgments

References

1. Network Rail. Annual Return 2013. http://www.networkrail.co.uk/publications/Annual-return/2013.

2. Atamuradov V, Camci F, Baskan S and Sevkli M. Failure Diagnostics for Railway Point Machines Using Expert Systems. *2009 Ieee International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*. 2009: 335-9.

3. Bagwell P. The sad state of British railways: The rise and fall of Railtrack, 1992–2002. *Journal of Transport History*. 2004; 25: 111-24.

4. Strukton Rail. POSS online monitoring. 2013.

5. Siemens AG. Point Diagnostic System SIDIS W. 2011.

6. SPX Rail Systems. Modular Point Operating Machine SPX Clamplock. 2013.

7. CDS Rail. Points Condition Monitoring. 2013.

8. Thinking Forward. Integrated Remote System for Monitoring Point Machines. 2013.

9. Márquez FPG, Schmid F and Collado JC. A reliability centered approach to remote condition monitoring. A railway points case study. *Reliability Engineering & System Safety*. 2003; 80: 33-40.

10. McHutchon MA, Staszewski WJ and Schmid F. Signal processing for remote condition monitoring of railway points. *Strain*. 2005; 41: 71-85.

11. Márquez FPG and Chacón Muñoz JM. A pattern recognition and data analysis method for maintenance management. *International Journal of Systems Science*. 2010: 1-15.

12. Márquez FPG and Garcia-Pardo IP. Principal Component Analysis Applied to Filtered Signals for Maintenance Management. *Qual Reliab Eng Int*. 2010; 26: 523-7.

13. Ardakani HD, Lucas C, Siegel D, et al. PHM for railway system - A case study on the health assessment of the point machines. *Prognostics and Health Management (PHM), 2012 IEEE Conference on*. 2012, p. 1-5.

14.   Bolbolamiri N, Sanai MS and Mirabadi A. Time-Domain Stator Current Condition Monitoring: Analyzing Point Failures Detection by Kolmogorov-Smirnov (K-S) Test. *World Academy of Science, Engineering and Technology*. 2012: 1049-55.

15.   Chamroukhi F, Same A, Aknin P and Antoni M. Switch mechanism diagnosis using a pattern recognition approach. *4th IET International Conference on Railway Condition Monitoring*. 2008, p. 1-4.

16.   Asada T and Roberts C. Development of an effective condition monitoring system for AC point machines. *Railway Condition Monitoring and Non-Destructive Testing (RCM 2011), 5th IET Conference on*. 2011, p. 1-6.

17.   Asada T, Roberts C and Koseki T. An algorithm for improved performance of railway condition monitoring equipment: Alternating-current point machine case study. *Transportation Research Part C: Emerging Technologies*. 2013; 30: 81-92.

18.   Asada T and Roberts C. Improving the dependability of DC point machines with a novel condition monitoring system. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*. 2013.

19.   Eker OF, Camci F and Kumar U. SVM Based Diagnostics on Railway Turnouts. *International Journal of Performability Engineering*. 2012; 8: 289-98.

20.   Böhm T. Accuracy Improvement of Condition Diagnosis of Railway Switches via External Data Integration. In: Boller C, (ed.). *Proceedings of the Sixth European Workshop on Structural Health Monitoring*. Dresden, Deutschland: Deutsche Gesellschaft für Zerstörungsfreie Prüfung, 2012, p. 1550-8.

21.   Pedregal DJ, Garcia FP and Roberts C. An algorithmic approach for maintenance management based on advanced state space systems and harmonic regressions. *Ann Oper Res*. 2009; 166: 109 - 24.

22.   Márquez FPG, Pedregal DJ and Roberts C. Time series methods applied to failure prediction and detection. *Reliability Engineering & System Safety*. 2010; 95: 698-703.

23.   Márquez FPG and Schmid F. A digital filter-based approach to the remote condition monitoring of railway turnouts. *Reliability Engineering & System Safety*. 2007; 92: 830-40.

24.   Silmon JA and Roberts C. Improving railway switch system reliability with innovative condition monitoring algorithms. *Proceedings of the Institution of Mechanical Engineers Part F-Journal of Rail and Rapid Transit*. 2010; 224: 293-302.

25.   Skipsey   D.   Clamp   lock. http://www.rmweb.co.uk/community/index.php?/topic/17767-clamp-locks/2010.

26.   Tunley M. Points to Failure. *IRSE news*. Institution of Railway Signal Engineers, 2010, p. 5-8.

27.   Deza E and Deza M. *Encyclopedia of Distances*. Springer, 2009.

28.   Marteau PF. Time Warp Edit Distance with Stiffness Adjustment for Time Series Matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2009; 31: 306-18.

29.   Fu AWC, Keogh E, Lau LYH, Ratanamahatana CA and Wong RCW. Scaling and time warping in time series querying. *Vldb J*. 2008; 17: 899-921.

30.   Lorenz D and Köhler T. *A Comparison of Denoising Methods for One Dimensional Time Series*. Zentrum für Technomathematik, 2005.

31.   Kalman RE. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME – Journal of Basic Engineering*. 1960: 35-45.

32.   Scholkopf B, Platt JC, Shawe-Taylor J, Smola AJ and Williamson RC. Estimating the support of a high-dimensional distribution. *Neural Comput*. 2001; 13: 1443-71.

33.   Ganyun LV, Cheng HZ, Zhai HB and Dong LX. Fault diagnosis of power transformer based on multi-layer SVM classifier. *Electr Pow Syst Res*. 2005; 74: 1-7.

34.   Li Y, Fang Q, Wang X, Zhang Z and Xie C. Research on air-conditioning fault diagnosis method based on SVM. *Sixth International Conference on Natural Computation (ICNC)*. 2010, p. 3397-400.

35.   Dongyu Z, Wangmeng Z, Zhang D and Hongzhi Z. Time Series Classification Using Support Vector Machine with Gaussian Elastic Metric Kernel. *20th International Conference on Pattern Recognition (ICPR)*. 2010, p. 29-32.

36.   Vanschoenwinkel B and Manderick B. Appropriate Kernel Functions for Support Vector Machine Learning with Sequences of Symbolic Data. In: Winkler J, Niranjan M and Lawrence N, (eds.).

Deterministic and Statistical Methods in Machine Learning. Springer Berlin Heidelberg, 2005, p. 256-80.

37. Hansheng L and Bingyu S. A Study on the Dynamic Time Warping in Kernel Machines. *Signal-Image Technologies and Internet-Based System, 2007 SITIS '07 Third International IEEE Conference on*. 2007, p. 839-45.

38. Chen L and Ng R. On the marriage of Lp-norms and edit distance. *Proceedings of the Thirtieth international conference on Very large data bases - Volume 30*. Toronto, Canada: VLDB Endowment, 2004, p. 792-803.

39. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*. Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc., 1995, p. 1137-43.