

Saari, Juhamatti; Odelius, Johan

Article

Detecting operation regimes using unsupervised clustering with infected group labelling to improve machine diagnostics and prognostics

Operations Research Perspectives

Provided in Cooperation with:

Elsevier

Suggested Citation: Saari, Juhamatti; Odelius, Johan (2018) : Detecting operation regimes using unsupervised clustering with infected group labelling to improve machine diagnostics and prognostics, Operations Research Perspectives, ISSN 2214-7160, Elsevier, Amsterdam, Vol. 5, pp. 232-244,
<https://doi.org/10.1016/j.orp.2018.08.002>

This Version is available at:

<https://hdl.handle.net/10419/246352>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by-nc-nd/4.0/>



Detecting operation regimes using unsupervised clustering with infected group labelling to improve machine diagnostics and prognostics

Juhamatti Saari^{*,a,b}, Johan Odelius^b

^a SKF-LTU University Technology Centre, Luleå University of Technology, Luleå, SE-97187, Sweden

^b Division of Operation, Maintenance and Acoustics, Luleå University of Technology, Luleå, SE-97187, Sweden



ARTICLE INFO

Keywords:

Maintenance
Operation regime
Clustering
Data mining
LHD

ABSTRACT

Estimating the stress level of components while operation modes are varying is a key issue for many prognostic models in condition monitoring. The identification of operation profiles during production is therefore important. Clustering condition monitoring data with regard to operation regimes will provide more detailed information about the variation of stress levels during production. The distribution of the operation regimes can then support prognostics by revealing the cause-and-effect relationship between the operation regimes and the wear level of components.

In this study unsupervised clustering technique was used for detecting operation regimes for an underground LHD (load-haul-dump machine) by using features extracted from vibration signals measured on the front axle and the speed of the Cardan axle. The clusters were also infected with a small portion of the data to obtain the corresponding labels for each cluster. Promising results were obtained where each sought-for operation regime was detected in a sensible manner using vibration RMS values together with speed.

1. Introduction

Prognostic and health management (PHM) of a system is a discipline that link studies of failure mechanism to system lifecycle management [25]. One of the challenges of PHM is to estimate the stress level of the components of a system when operation modes vary. For many systems, it is either impossible or impractical to measure component stress accurately, so the next best thing may be to detect operation profiles during production. However, for complex systems, even the operation profile can be unknown and may change on daily basis. There is a need for methods which can use pre-existing data (condition monitoring or process data), often collected for other purposes, to detect operation regimes. Results can be used to predict different life scenarios in case of incipient faults or to determine the correct time and place to apply diagnostic techniques.

Machine learning and pattern recognition techniques for data mining have been improving dramatically recently, with many more areas of application, including PHM. They have been adapted for and are used in the PHM of machines in the automotive industry [6], defense and space programs [30] and heavy industries [33]. Machine learning techniques used in PHM can be divided into three rough categories: classification, regression and clustering techniques.

Classification algorithms are used to classify two or more categories

by using data to distinguish, for example, a faulty system from a healthy one [19,27]. Regression models are mainly used for prognosis where the time to failure is estimated using existing historical data (see for instance [28]). Regression analysis involves the use of such techniques as neural networks, fuzzy logic systems and simpler univariate regression models; these techniques are not strictly reserved for regression analysis and can also be used for data mining. Recently Hanafizadeh et al. [16] used supervised neural networks to identify flow regimes in a pipe to determine when the flow type was changing during operation. This technique aims to improve the control of the process by determining when it is not optimal. However, it is not the most practical one for identifying operation regimes of complex machines; the data need to be labelled while training the model, and this is seldom done in a varying operating environment, as, for instance, with mobile machines. Suarez et al. [26] tracked real-time onboard damage accumulation using a model called PHM/ALPS. The goal was to evaluate the current mission profile (operating conditions) using past mission profiles (historical data) to demonstrate independent life prediction capability. It is difficult to adapt this type of technique for operation regime detection, however, unless several mission profiles are pre-recorded or simulated. Unsupervised clustering techniques, may be more practical than supervised ones in some cases since they do not require historical data from several different operating conditions. The benefit of

* Corresponding author at: SKF-LTU University Technology Centre, Luleå University of Technology, Luleå, SE-97187, Sweden

E-mail address: juhamatti.saari@ltu.se (J. Saari).

<https://doi.org/10.1016/j.orp.2018.08.002>

Received 7 February 2018; Received in revised form 6 July 2018; Accepted 3 August 2018

Available online 04 August 2018

2214-7160/© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

unsupervised techniques is the possibility of finding natural groups and patterns in the data by optimizing the boundaries and the clusters in the data. Mostly these techniques are used for anomaly detection, where several clusters are formed to characterize typical system behaviour and alarm is sent when data vector is outside of clusters [17]. Perhaps the most common unsupervised clustering method is the k-means algorithm [18]. This algorithm is initialized by picking k initial cluster points and allocating all data points to the closest one. Another popular cluster algorithm proven to be successful in many situations is the expectation-maximization [9,11]. When detecting operation regimes, there are limitations to using these algorithms (see for instance [13,31]). Perhaps the biggest problem when trying to implement these techniques is the need to set the number of clusters in advance, as this is rarely known for complex machines operating under unknown conditions or in a changing environment. For instance, load change in one time and position during production might create three separate clusters which cannot be treated as one mode.

To overcome this problem, Corduneanu and Bishop [10] have developed a variational Bayesian Gaussian mixture (VBGM) model. With this algorithm, it is not necessary to know the exact number of clusters (k) in the beginning, since only the maximum number of clusters needs to be set. Similar techniques have been applied to defining operation regimes in the process industry using control parameters, such as valve openings or temperature [34]. but in these techniques, the k value is defined using another algorithm [12].

Although unsupervised techniques have advantages when compared to supervised ones, there are some practical limitations. One is the validation of cluster labels, i.e., what each cluster actually represents. To overcome this problem, we propose a method where the VBGM algorithm is first used to separate a large set of condition monitoring data into groups (clusters) which are later infected with a smaller set of data with labels. We apply the method to the analysis of vibration data collected from a complex machine operating under harsh conditions (underground mining loader, LHD). Aim is to see how the unsupervised algorithm, together with infection data, can be applied for separating operation modes using only condition monitoring data. We use vibration measurement data collected for diagnosis purposes and consisting of noise from many natural sources. Work is novel in that it applies the VBGM clustering algorithm to real data and explains how it can be used generically with infection data to predict labelled clusters.

2. Background and labelling operation regimes

Clustering technique (VBGM) used in this study for separating data for different clusters is based on the work by Corduneanu and Bishop [10], which can be also found in the book written by Bishop [7]. When the VBGM algorithm is used for mining condition monitoring data, more specifically, to separate data into meaningful operation regimes, it is not necessary to know the exact number of clusters, since components whose expected mixing coefficients are numerically indistinguishable from zero are not plotted [7]. The method is also more practical (generalizable) since it can rely on data when the training set is large and on the prior distribution assumption when the data set is small.

In Gaussian mixture model for each observation x_n we have a corresponding latent variable z_n comprising a 1-of-K binary vector with elements z_{nk} for $k=1, \dots, K$. Denotation for observed data set is $\mathbf{X} = x_1, \dots, x_N$, similarly latent variables are denoted as $\mathbf{Z} = z_1, \dots, z_N$.

Conditional distribution of \mathbf{Z} , given the mixing coefficients $\boldsymbol{\pi}$, is defined as follows [7]

$$p(\mathbf{Z} | \boldsymbol{\pi}) = \prod_{n=1}^N \prod_{k=1}^K \pi_k^{z_{nk}}. \tag{1}$$

For the observed data, the conditional distribution, given the latent variables and the component parameters, is as follows [7]

$$p(\mathbf{X} | \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \prod_{n=1}^N \prod_{k=1}^K \mathcal{N}(x_n | \mu_k, \Lambda_k^{-1})^{z_{nk}}, \tag{2}$$

where $\boldsymbol{\mu} = \{\mu_k\}$ is mean and $\boldsymbol{\Lambda} = \{\Lambda_k\}$ is precision.

Using conjugate prior distribution and choosing a Dirichlet distribution over the mixing coefficient $\boldsymbol{\pi}$, which is defined as [7]

$$p(\boldsymbol{\pi}) = \text{Dir}(\boldsymbol{\pi} | \boldsymbol{\alpha}_0) = C(\boldsymbol{\alpha}_0) \prod_{k=1}^K \pi_k^{\alpha_0 - 1}, \tag{3}$$

where $C(\boldsymbol{\alpha}_0)$ is the normalization constant for the Dirichlet distribution. Hyperparameter $\boldsymbol{\alpha}_0$ can be interpreted as the effective number of observations associated with each component of a mixture. If $\boldsymbol{\alpha}_0$ is small, the posterior distribution will be influenced primarily by the data rather than the prior.

By introducing independent Gaussian-Wishart prior governing the mean and precision of each Gaussian component, the distribution can be written as [7]

$$p(\boldsymbol{\mu}, \boldsymbol{\Lambda}) = p(\boldsymbol{\mu} | \boldsymbol{\Lambda})p(\boldsymbol{\Lambda}) = \prod_{k=1}^K \mathcal{N}(\mu_k | m_0, (\beta_0 \Lambda_k)^{-1}) \mathcal{W}(\Lambda_k | W_0, \nu_0). \tag{4}$$

Joint distribution of all of the random variables, is given by the equation [7]

$$p(\mathbf{X}, \mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = p(\mathbf{X} | \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Lambda})p(\mathbf{Z} | \boldsymbol{\mu}, \boldsymbol{\Lambda})p(\boldsymbol{\pi})p(\boldsymbol{\pi} | \boldsymbol{\Lambda})p(\boldsymbol{\Lambda}). \tag{5}$$

In the Eq. 3 only the variables \mathbf{X} are observed.

Considering a variational distribution which factorizes between the latent variables and the parameters, so that [7]

$$q(\mathbf{Z}, \boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = q(\mathbf{Z})q(\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}). \tag{6}$$

With this assumption it is possible to obtain a traceable practical solution to the Bayesian mixture model. The optimal solution is found by seeking a distribution for which the lower bound is largest.

A toolbox for the algorithm is publicly available at Mathworks [22]. In this study, we kept the parameter settings at default each time the algorithm was run. These parameters, $\boldsymbol{\alpha}_0$, was 1 and β_0 , which affects to the initial precision value ($\boldsymbol{\Lambda}$), was 1.

To overcome the problem of not knowing what each cluster represent, we propose method to collect another set of data which is much smaller than the training set (See Fig. 1). This smaller set of data can be used to infect some or all of the found clusters in order to know what they represent by predicting their clusters using already trained models. Benefit of the technique is that the training can be carried out for a much larger data set and rare patterns which may occur during production in some situations, will be included in the model. However disadvantage may be the difficulties of interpreting cluster labels, if data is distributed evenly among clusters. In these cases, parameters needs to re-selected or use different initial parameter values to achieve better results. Infection data should be collected in such manner that one complete cycle of the operation is present.

With this technique, once the computationally demanding training phase is over (although it is the same as compared to traditional maximum likelihood ones), real time or near real time cluster prediction for new data set is achievable for several system/components by using on-site feature extraction and wireless communication together with centralized computing.

The ideal way to collect infection data set would be to let the operator determine when to acquire data during operation (first-hand knowledge) or to automatize data collection and use RFID tags or other similar techniques. These are used in many industries to keep track of mobile machines (for instance, in mining industry). Time period for data collection should cover the whole operation mode in the beginning and only later, if the operation mode is distributed evenly into many clusters, a deeper analysis and better selection should be done.

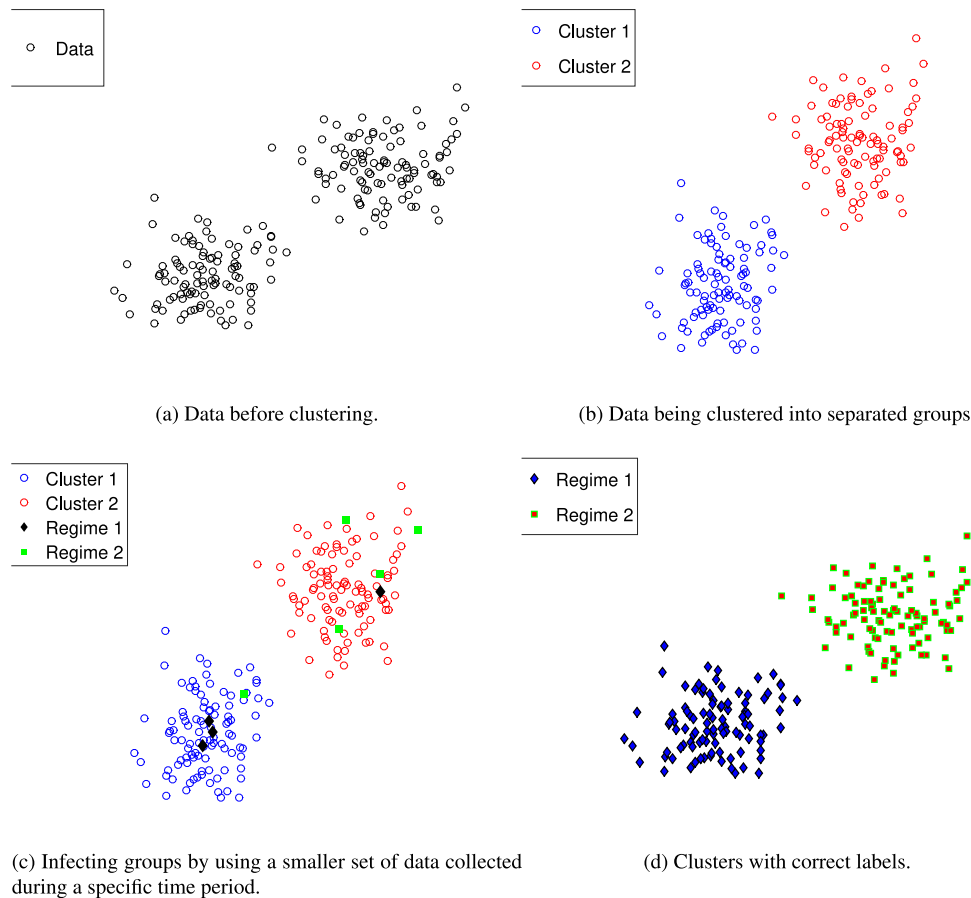


Fig. 1. Proposed method how to determine operation groups by using unsupervised clustering.

2.1. Parameter selection

The number of input parameters for detecting operation regime can concern two or more dimensions. Optimal parameter selection would be to choose the ones which are connected to the stress induced in the monitored component. For instance, the power of an electric motor can be a good parameter when the assumption is that an increase in the torsional load might be causing problems for the system being monitored, since it is known that $P = M\omega$, where P is the power, M is the moment and ω is the angular velocity.

However, in practice sensor network is always optimized based on to the price-benefit ratio and therefore many of the measured parameters are only indirectly related to the stress of a certain component. Therefore it is rational to study those parameters which are most commonly collected for other purposes (e.g. diagnosis purposes). Note, that this may create another problem where used feature values start to change once the monitored component is degrading and thus the trained clusters are not valid anymore. For instance vibration based features have a tendency to increase near the end of its technical life. One solution how the problem can be avoided is to use vibration sources which are not directly related to the diagnosis of the given component, but are located nearby (monitoring other component faults). For instance as in the given case study (see Section 4), sensors located on the left side of the front axle can be used for monitoring operation changes happening at the right side and vice versa. This way, (once detection algorithms are available), both can be studied in parallel without using the same features for diagnosis purposes and for detecting the operation regimes.

Generic unsupervised feature selection is the focus of many studies [8,15,35]. If these selection filters can be successfully used as a pre-processing step, it would make the operation regime clustering easily

adaptable for alternate systems. However authors do believe that comprehensive, generic method for selecting input parameters probably cannot be achieved in the near future for solving the explained problem. Not at least before some of the solution are in generic use and practical shortcomings are encountered. Therefore the best approach is to use input features, which are tailored to each system separately. However first priority should be to find features which has a potential to function with many types of systems. Therefore as a first step, we have chosen to study how common vibration features, without any pre-processing, are able to separate operation regimes in our case study.

3. Relation to diagnostic and prognostic techniques

In generic, the formal definition of the RUL can be expressed as the stochastic degradation process of the system [36]. Stochastic process can be modelled as a first hitting time (FHT model [21], denoted by $X(t), t \in \mathcal{T}, x \in \mathcal{X}$ with initial value $X(0) = x_0$, where \mathcal{T} is the time space and \mathcal{X} is the state space of the process. Boundary set when the failure occurs can be denoted as a boundary set \mathcal{B} , where $\mathcal{B} \subset \mathcal{X}$. FHT model denotes that RUL has ended when the process lie outside the boundary set \mathcal{B} . Two important aspects for defining the RUL correctly is to define the Boundary set \mathcal{B} correctly and to estimate the time when the process $X(t)$ lies outside the Boundary set [21].

General description estimating the RUL cannot consider any operational changes and assumption is made that the system will only have one homogeneous operation mode. However in most of the real world systems, it is reasonable to assume that some of the working state will cause a higher degradation rate than others [24]. Therefore the state space \mathcal{X} can consist of possible states during the operation $\mathcal{X}_r, r = 0, 1, 2, \dots$. With the help of proposed clustering technique it is possible to separate the process variable data into several clusters so

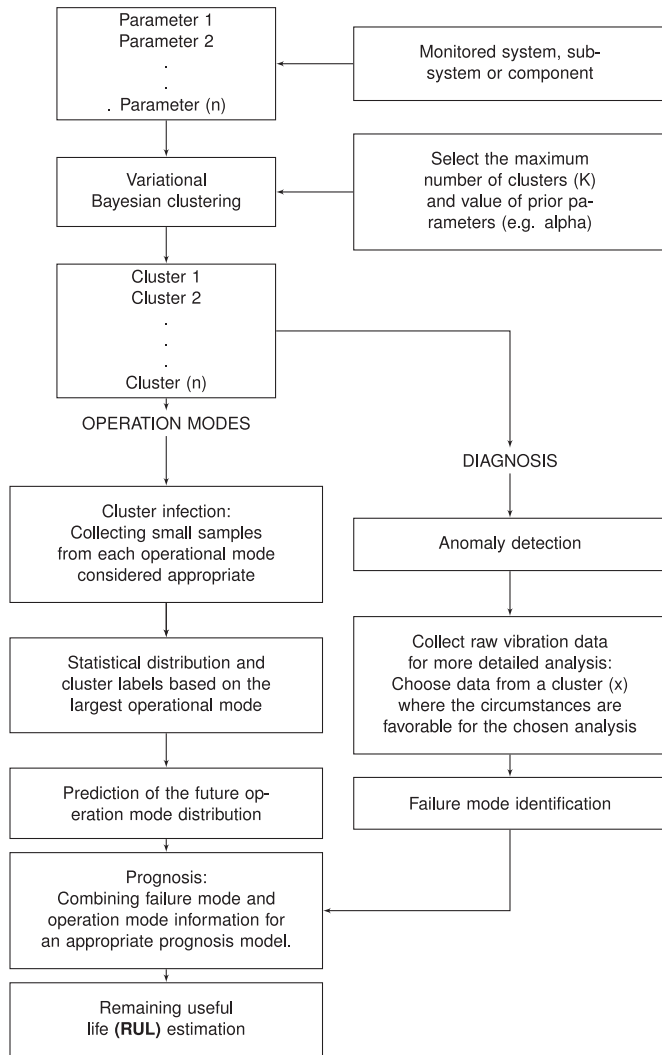


Fig. 2. Flowchart of the proposed method: monitoring the condition of a complex system by combining operational regimes with fault diagnosis and prognosis techniques.

that the number of states $r \leq K$, where K is the number of found clusters.

Once the separation of the process variable data is done, there are several approaches that can be performed: First, the degradation path can be recorded during independent failures in order to estimate what was the degradation rate in each of them. Later this information can be used by recalculating the RUL by varying the estimated distribution of each operation state. Second, for each operation mode it is possible to define individual degradation model that can be used instead one universal degradation model. For instance in fracture mechanics, crack propagation can have three modes (opening sliding and tearing) and once the relation between operation and the stress type is known, correct model can be used in each defined operation states. Third, it can also be beneficial for diagnostic to acquire the training data only when the system is in a particular state (i.e. cluster (x) in Fig. 2).

A universal flowchart showing how defined operation regimes can be used for condition monitoring is provided in Fig. 2. In Fig. 2 steps related to diagnosis are for defining the correct failure mode, which can help to define the correct degradation model or to assist to select an appropriate boundary set \mathcal{B} . Eventually the estimation of the RUL can be achieved using either knowledge-based reasoning [2], physical models [3] or data-driven approaches [3]. More detailed explanation of each approaches was recently reviewed by Zhang et al. [36].

Proposed model is built by completing the following steps.

- Collect a set of data using the chosen parameters with sufficient time segment.
- Data normalization (e.g. z-score method).
- Choose the maximum number of clusters (K).
- Collect small samples from each operational mode to infect and label the clusters, see Fig. 2c.
- Once the model has been trained, study the distribution of new data instances to determine how the operation modes vary during the process and re-evaluate the need of merging or separating found clusters.

4. Case study and selected input parameters

To test the proposed method in a real environment, data were collected from a LHD (load-haul-dump machine) working in an underground mine (Fig. 3a). Such LHD's are subjected daily to several different operating conditions where the environment, boulder sizes, road condition and even the operator are changing. Loaders belong to a class of vehicles for which it can be very challenging to use traditional condition monitoring techniques, since the rotation speed changes and the loads and type of load vary. This combination of factors is usually so demanding that there are no good methods for estimating when some of the critical components are going to fail, since even the performance of diagnoses can sometimes be challenging. Nowadays this type of machine mostly relies on preventive maintenance and weekly inspections [1]. However, the development of these machines is heading towards full automation [14] and therefore the demand to find methods for the creation of a condition-based maintenance protocol is increasing.

The LHD model in question is made by Sandvik and is an LH621. Vibration measurements were performed using a National Instruments CompactRIO 9024 data logger where four SKF Copperhead CMPT 2310 accelerometer sensors were used. Originally data was collected for diagnosis purposes and therefore were installed to be near the most critical component (see [20]). Sensors were installed on the front axle, two on the left side of the axle and two on the right side, as seen in Fig. 3. The vibration measurements were synchronized with the cardan axle speed, which was obtained using the tachometer pulse from the drive shaft. The vibration measurements were continuous, which means that every operation regime of the LHD was recorded with a precise time stamp. The sample rate was 12.8 kHz.

In these machines there is an in-built condition monitoring tool that can record several parameters from the machine, such as the RPM of the engine, the machine speed, the RPM of the cardan axle, the temperature at several positions, the driven gear and the hydraulic pressure. All this information can be used for detecting the operational regimes in the future with the proposed technique once the data is synchronized and shared online with condition monitoring systems (e.g. using machine to machine communication). Some feasibility analysis using the given tool is already performed in other studies [37] and using Kalman filter technique [32].

4.1. Feature extraction

Since raw vibration measurements can rarely be used, a common method is to extract features which will indicate certain attributes (qualitative or quantitative). Some features are sensitive to overall vibration levels, for instance, the RMS value. Other features can give good results when used to detect impacts, such as the peak value or the peak-to-peak value. We selected five commonly used features that would distinguish how the shape and form of the vibration signal changed: the RMS, peak, peak-to-peak, kurtosis and skewness values. All these parameters should vary while the machine is in operation. Although many studies (related to fault identification) consider pre-processing



(a) A LHD machine.



(b) Sensor on a horizontal direction.



(c) Sensor on a vertical direction.

Fig. 3. A typical underground loader and two sensors mounted on to the front axle.

vibration measurements using frequency filtering or time-frequency filtering [4,5,23,29], we have only studied time domain features for two reasons. First, because the proposed method should be generic, features which do not require extensive manual pre-processing should be prioritized. Second (based on our literature study), similar studies using vibration measurements for detecting operation regimes have not been done using time domain features.

4.2. Operation regimes of the LHD

Roughly the operation can be divided into five different phases: driving between the working face and the maintenance hall (located underground), transit to the loading position, loading, hauling and idling. How each of these phases are wearing a monitored component(s) depends on the failure type (the malfunctioning component and the type of failure). One can estimate the harmfulness of each stage only when each group has been analysed separately after failures have occurred and have been documented correctly.

Transit from the maintenance hall to the working face

In this phase the speed is usually higher than that in the other stages and at the same time the LHD has the lowest static load, since there is no ore inside the shovel. During this phase the LHD is almost always driven using manual operation and some speed decreases might occur, depending on the traffic inside the caverns and how tight the corners are in the tunnel. During this phase, an uneven road or holes in the road can have an effect on the health of a monitored component.

Transit to the loading position

This phase is similar to the first phase, since there is no load which needs to be carried. However, the difference is that, during this phase, the LHD is operated automatically and the maximum speed is limited to a lower RPM than in manual driving. Moreover, there might be some differences in the road condition, since the smaller tunnels used for this operation may not be in as good a condition as the main tunnels. Therefore, there might be a slight increase in the vibration levels compared to those encountered in the first phase.

Loading

Loading is the event when the ore is picked up after it has been

drilled and blasted from the working face. This operation can be performed using a manned LHD or an LHD remotely controlled by an operator sitting in a van. During this phase the LHD is normally operated at lower speeds, but is subjected to heavy impacts, since some of the rocks can be rather big. According to previous studies, this might be the most harmful phase for the e.g. gears in the front axle [14]. This is especially true of LHD's driven using remote control, since in this case the operator loses their intuitive sense of the machine's handling and cannot perceive the subtle differences between smooth and rough handling.

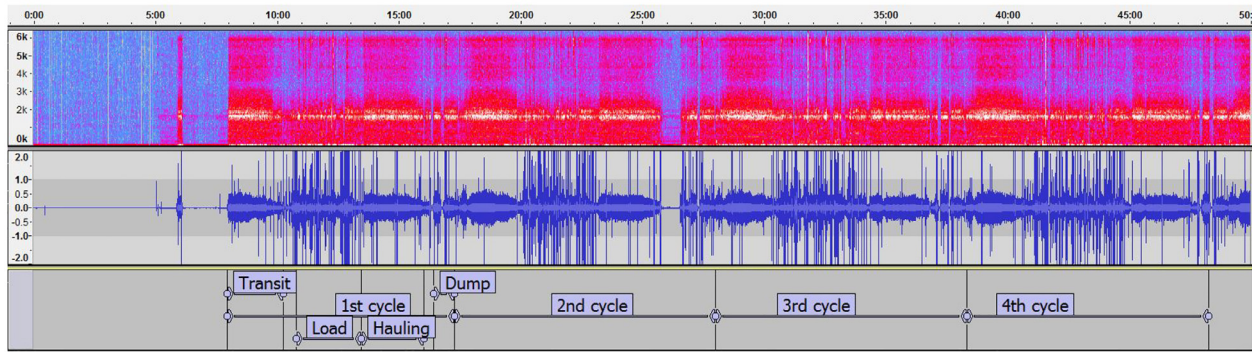
Hauling and dumping

During the hauling phase, the ore is carried inside the LHD's shovel and is transported to the dumping point, where it is crushed or prepared for crushing into smaller pieces. The difference between the hauling operation and the two transit stages is that, during hauling, the total weight of the LHD is much greater and the average speed is less than during the transit stages. Moreover, the hauling operation is performed using an autopilot and the operator only monitors the event. During hauling, impact forces do not usually occur, since sudden movements are avoided and collisions caused by human error do not happen, since the operation is automated. However, higher static loads might be harmful to some of the components, especially the front axle, since most of the weight is carried by the front axle because of the shovel's location at the front.

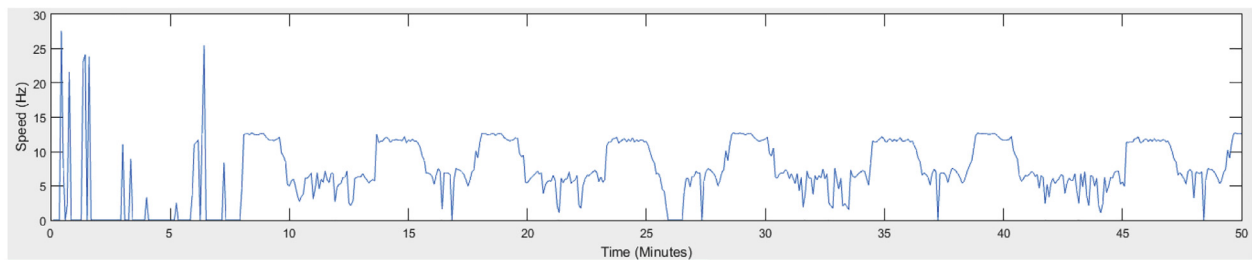
Dumping is a very quick operation and has little effect on mechanical components. Moreover, when concentrating on monitoring the condition of the other components, for example the hydraulic cylinders of the shovel, it would be useful to differentiate between this operation and hauling. However, when defining operation modes affecting the lifetime of a e.g. front axle, this would be futile and, therefore, hauling and dumping can be considered as two parts of a single phase (which is done in this study).

4.3. Extraction of the infection data

In order to collect the infection data, we extracted the WAV signal of one of the acceleration signals and its spectrogram (0–6 kHz) to characterize and reveal different regimes. Infection data was collected on Tuesday for defining operation regimes manual transit and idle and



(a) Waveform of the acceleration signal, spectrogram and estimated stages during several cycles (Wednesday 1st shift, see Figure 6).



(b) LHD cycles during operation and the variation of the rotation speed. Each speed value is the average of a 5 second segment.

Fig. 4. LHD cycles during operation (50 min. time periods).

Wednesday for defining other operation modes. As shown in Fig. 4 the LHD works for periods of approximately 50 min.. During each period, there are three cycles where the LHD will first transit to the loading position (transit with automatic operation), collect the ore (loading) and return to the dumping position (hauling + dumping). As can be observed, there is a similar pattern when each mode is changing, except that the separation between a transit and a hauling operation cannot be seen. However, these two operations can be distinguished based on the fact that hauling always follows loading and transit always follows dumping.

Fig. 5 shows the time period when the LHD is in transit from the maintenance hall to the cavern where the ore is being removed from the face. The duration of this phase is approximately one hour and 45 min.. The difference between this transit and that visualized in Fig. 4 is that in the former transit, the LHD is operated manually and the speed can be much higher. The reason why this mode of operation should be separated from automated transit is that, in some cases, the speed increase might be crucial for the health of the components, since e.g. the front axle is lubricated using oil and sump starvation may occur at higher speeds and thus increases the wear process in some circumstances. Moreover, this phase may also be a good opportunity to acquire vibration signals which can be used later for diagnosis purposes.

Sometimes there are periods when the LHD is stopped for some unknown reason and must wait before it can become productive again. This can be seen clearly in Fig. 5 during the period from 39 to 44 min.. It is also advantageous to separate this mode from the other regimes, since it should not damage the machine at all. If this mode increases during production, this can be taken into account when calculating the RUL.

As can be observed in Figs. 4b and 5 b the rotation speed of the Cardan axle seems to be a very promising indicator for distinguishing most operation regimes from each other. Therefore, in further tests, speed was one of the parameters used in all the models.

Based on this analysis, we manually selected (by using audiovisual

tools) the isolated data segments which belonged to these five different groups. Later these data were used when infecting clusters in the way mentioned in Section 2. Note, however, that this type of analysis is unnecessary when using the proposed method in real life, since infection data can be separated from the other data during production.

4.4. Typical operation during one week

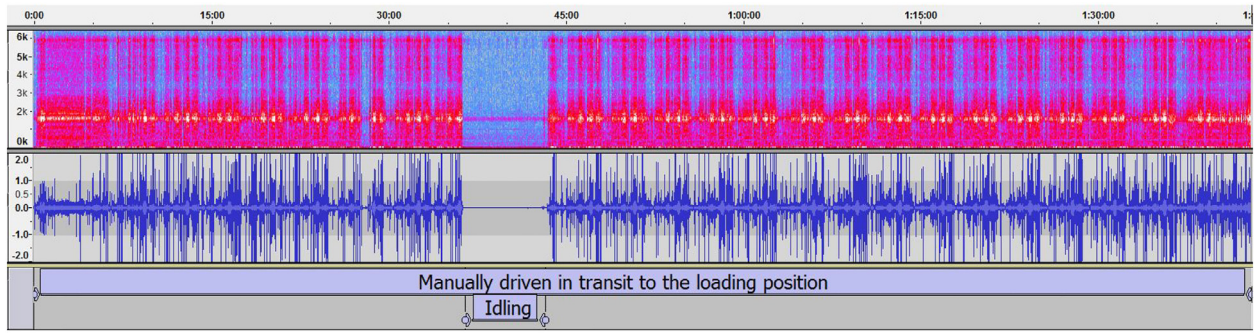
In Fig. 6 it can be observed that the overall pattern for most days in the week seems to be rather similar. There are two shifts during each day which are operated quite punctually. However, on Monday and Friday, the daily operation time is quite shorter for some reason, possibly due to weekly predictive maintenance tasks and other maintenance issues. Therefore, the further tests were performed comparing the data collected during the three days from Tuesday to Thursday.

5. Results and discussion

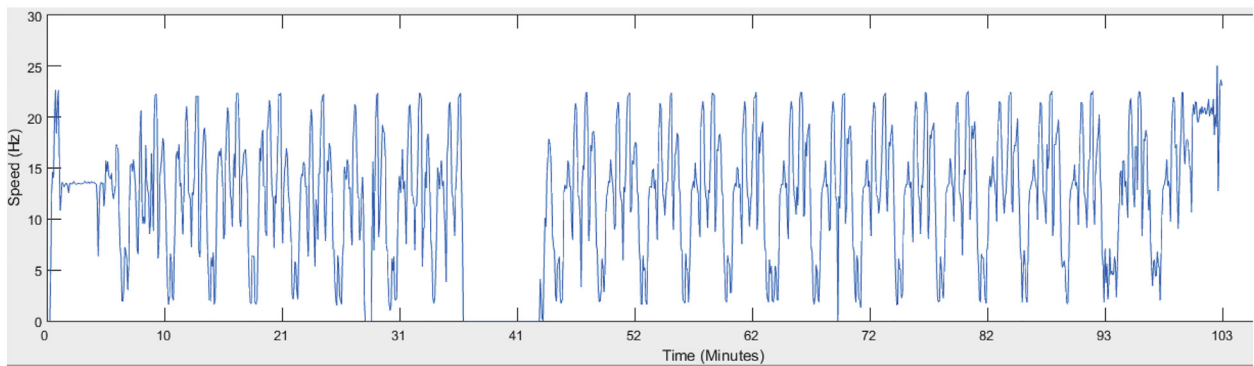
The results were processed in three steps. Firstly the distribution and values of the selected input parameters were visualized in order to see how noise and other factors might have influenced the measurements.

Secondly the models, after clustering, was evaluated by estimating how the trained models converged when the training data were collected on different days. If the parameters are suitable and can be used for data separation, each model should have approximately the same amount of clusters in similar positions when data are taken at a random time during production.

Thirdly the models trained using data collected on Tuesday were evaluated using the infection data, to determine how well each operation regime was represented by each cluster.



(a) Waveform of the acceleration signal, spectrogram and estimated stages during several cycles (Tuesday 1st shift, see Figure 6).



(b) LHD cycles during operation and the variation of the rotation speed. Each speed value is the average of 5 second segment.

Fig. 5. LHD transit to the working face (1h. 43 min. time period).

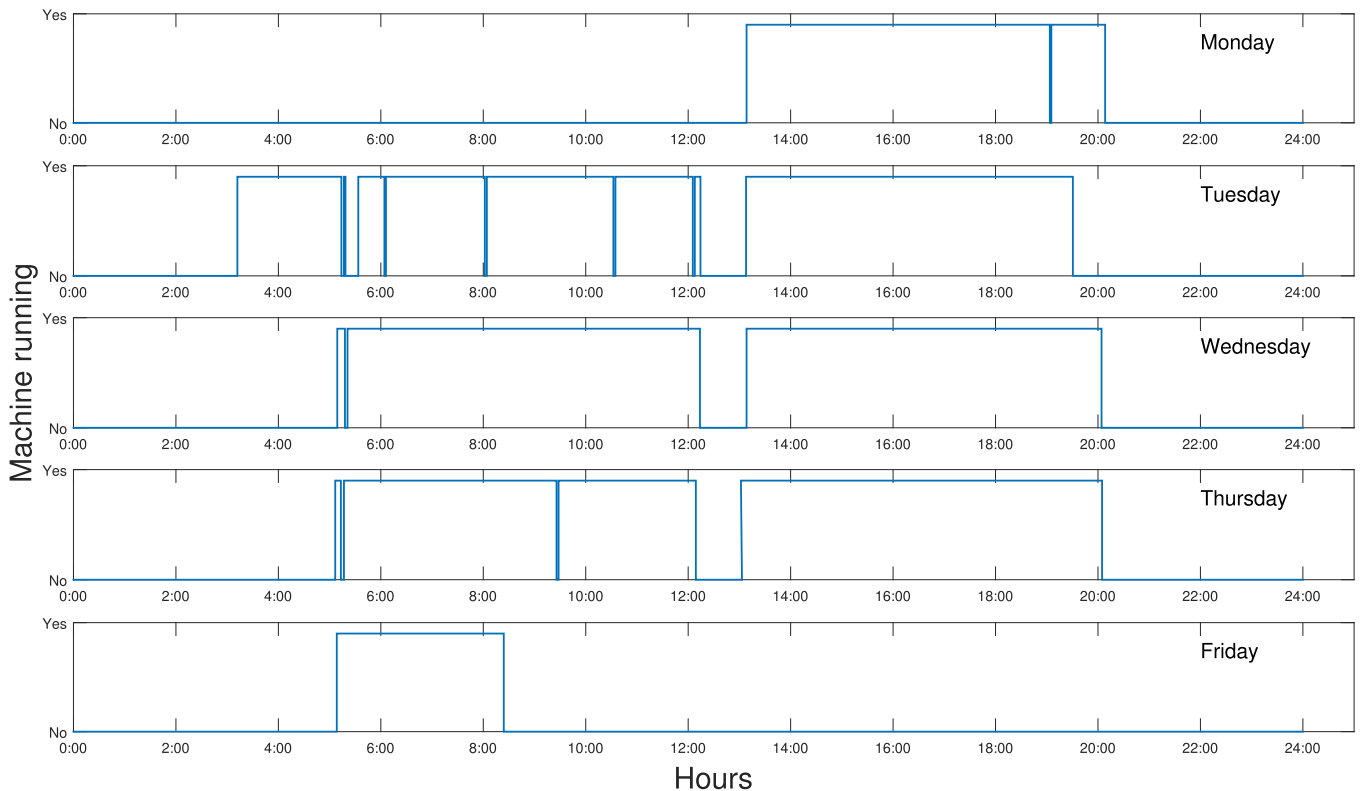


Fig. 6. Time periods when the loader was running during one week.

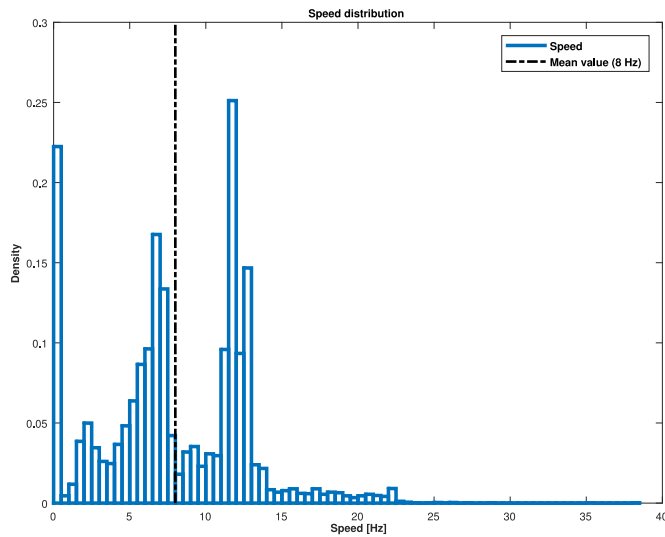


Fig. 7. Distribution of the rotation speed of the Cardan axle on a typical day (Tuesday). Each speed value is the average of a 5 sec, segment.

Parameter evaluation

The nine parameters selected were as follows: the rotation speed of the Cardan axle,

RMS_V (vertical), RMS_H (horizontal), $Peak_V$, $Peak_H$, $Kurtosis_H$, $Kurtosis_V$, $Skewness_H$ and $Skewness_V$. To test how these parameters could be used together to determine the operational behaviour of the LHD, each parameter was statistically evaluated in order to ensure that mixture of Gaussian distribution is meaningful. The distribution of the speed for each chosen day can be seen in Fig. 7. When comparing 3 full working days (Tuesday, Wednesday and Thursday) it was clear that each day had a similar speed distribution where 3 almost Gaussian modes were present which were centred around the speeds 2.5 Hz, 7 Hz and 13 Hz. Moreover, there was a narrow spike at 0 Hz, that indicates that the machine is idling.

Take the result first and then you may discuss it and draw conclusions: As shown in Fig. 7 the idling state (bin where the speed is zeros) is very dominant. Using scheduled maintenance based on operation time is therefore not suitable if idle time is not considered. This finding is well aligned with the study done by [21] who stated that the time scale for the process is not the same as the calendar or clock time. Further, using mean speed is not a good indicator for the wear of bearing and gears.

As shown in Fig. 7 the idling state (bin where the speed is zeros) is very dominant. Using scheduled maintenance based on operation time is therefore not suitable if idle time is not considered. Further, using mean speed is not a good indicator for the wear of bearing and gears. Better approximation would be to have separate analysis into three Gaussian distribution centred into 2 separated frequency areas (bins 2.5 Hz, 7 Hz and 13 Hz). Naturally this is only how the speed will affect into the RUL and several other aspects are also present such as the level of vibrations.

Other features extracted from the vibration signals can be found in Table 1. Unlike the speed, each of the vibration features has a shape which is more like a unimodal shape and is quite close to the shape of a normal distribution, except that there is a high end tail. Reason for the long tail might be due rocks and boulders, which are randomly hitting the bottom of the loader and causing transient increases in the acceleration levels. This can have a huge effect, especially on features like kurtosis and range. However, when examining the difference between the median and mean values of the kurtosis, one can observe that these peaks happen quite seldom, since the median value is much lower.

By comparing the range and peak values seen in Table 1 one can

observe that the range values are almost the same as the peak values scaled by 2. After normalization, which is almost always performed in many machine learning applications, it would be redundant to use both parameters as inputs for the cluster algorithm and, therefore, models are trained only by using the peak as an input parameter without considering the range value.

Separating the data using the clustering technique

Each model was trained using the speed of the Cardan axle as a single parameter together with each vibration feature (RMS, kurtosis, peak values and skewness) individually, to see how well the algorithm would produce similar results when the training data were collected on different days (Tuesday, Wednesday and Thursday). Only the shape and size could be evaluated, since the numbering varied because the initial starting point was chosen randomly inside the feature space. Model 5 is the exception, since it includes all the nine parameters and cannot be evaluated visually. Therefore, model 5 was only evaluated using the infection data. Before using the chosen parameters as inputs for the model, normalization was performed using the z-score method. After normalization, those feature values which were four times larger than the standard deviation were excluded from the final models in order to avoid anomalies caused by noise. As can be observed in Fig. 8 the algorithm was not able to reduce the number of clusters to five (the number of analysed regimes). Instead all the models found ten clusters, which was the initial K value. This indicates that the data are not dense and focused exclusively on the operation regimes.

Results indicate that, it would be advisable to assume the K value to be higher than the assumed number of operational regimes; later, if the cluster infection explained in Fig. 1 works properly, all the clusters without a label can be neglected as being noise. Moreover, if one operation mode is dominant in two clusters, the clusters can either be combined into a single cluster or kept separate and treated differently. For instance, in the loading mode, crashing into a boulder and gently lifting ore can be seen in different clusters and later be defined as separate operations.

Model 1 (RMS and speed)

In Fig. 8 model 1 represent clusters which were trained using the RMS and speed parameters. When comparing results when training data was collected on different days, one can observe that, cluster positions and sizes are rather similar. This indicates that the RMS and speed as an input features are able to produce similar results and do not depend heavily on the collection day.

When the values are low, this model has some difficulty in finding similar clusters. For example, when comparing clusters 1 and 4 (Tuesday), cluster 10 (Wednesday) and clusters 4 and 2 to each other, one can observe that the data in this area are sometimes divided into two clusters and are sometimes concentrated to only one. Nevertheless, the results seem to be promising, since these small variations are to be expected when using these types of approximation schemes.

Model 2 (kurtosis and speed)

Model 2 represent clusters which were trained using the kurtosis and speed parameters. One can observe that, when comparing the cluster positions and sizes, most of them are not similar. These results indicate that using kurtosis values in a real industrial environment can give poor results. This might be due to random noise peaks, which can alter the kurtosis value quite dramatically, as seen in Table 1. Perhaps in the future, if kurtosis is used as an input feature, one should use pre-processing filter where frequency band is taken on a more narrower spectra, which should reduce impacts coming from noise sources.

Model 3 (peak values and speed)

Model 3 represent clusters which were trained using the peak and speed parameters. By comparing the cluster positions and sizes, most of

Table 1

Statistical information on each feature before using the VBGM algorithm. V stands for the vertically and H stands for the horizontally mounted accelerometer values. The signal length was five sec. for each individual value, and during each day around 10,000 parameter values were calculated during different operation regimes.

Parameter	Tuesday					Wednesday					Thursday				
	Median	Mean	Std	Max	Min	Median	Mean	Std	Max	Min	Median	Mean	Std	Max	Min
RMS V (g)	0.12	0.13	0.10	4.86	0.00	0.12	0.12	0.08	3.42	0.00	0.11	0.11	0.09	5.10	0.00
RMS H (g)	0.09	0.10	0.09	4.98	0.00	0.09	0.10	0.08	3.49	0.00	0.08	0.09	0.08	5.20	0.00
Kurtosis V	4.13	78.82	424.76	24921.61	1.70	4.25	91.98	419.14	28050.70	2.12	4.50	90.59	473.01	27711.59	2.12
Kurtosis H	4.45	138.68	452.19	16851.51	1.63	4.83	162.74	507.36	27201.94	1.64	5.25	143.30	566.03	27989.68	1.85
Peak V (g)	0.68	1.91	2.82	31.60	0.00	0.69	2.07	3.07	31.60	0.00	0.68	1.91	2.92	31.60	0.00
Peak H (g)	0.58	2.08	3.87	32.19	0.00	0.60	2.22	3.96	32.19	0.00	0.60	1.92	3.42	32.19	0.00
Skewness V	-0.01	-0.09	2.55	104.76	-32.87	-0.01	-0.12	2.42	147.07	-28.46	-0.01	-0.12	2.81	146.42	-109.79
Skewness H	0.01	0.15	2.41	93.24	-34.30	0.01	0.05	2.50	144.15	-47.75	0.01	0.08	3.25	147.10	-110.64
Range V (g)	1.38	3.91	5.76	58.61	0.01	1.39	4.20	6.23	63.19	0.01	1.37	3.87	5.85	63.19	0.01
Range H (g)	1.16	4.13	7.59	64.41	0.01	1.20	4.42	7.84	64.41	0.01	1.21	3.82	6.70	64.41	0.01

them are not similar.

The initial assumption was that the peak values might be a good indicator when trying, for example, to separate the loading regime from the other regimes, since the vibration peaks increase during loading, which can be seen when comparing each stage manually, as in Figs. 4 and 5. The increase in the peak values can be seen especially at the horizontal level. The poorly converging results obtained when using the VBGM method in this case might be due to the fact that such peak values also occur during production in other stages, and therefore are masked and cannot form clear operation regimes.

Model 4 (skewness and speed)

Model 4 represent clusters which were trained using the skewness and speed parameters. The data behaviour for each day is quite similar and, when the speed increases, the skewness dispersion is very small. When the speed is low, the skewness values are greatly dispersed and many clusters can be found within this region. These results indicate that skewness might be a good parameter for separating operational regimes. At least the VBGM method can give promising results when cluster are not labelled.

5.1. Results after defining operation regimes using data infection

For the label infection, the models trained using the data collected on Tuesday were chosen. The results can be found in Table 2 for model 1–5, which were all infected using the same data. If the proposed method worked, all five different modes in the 'largest mode' row should be found. For simpler cases, the percentage proportion of each mode should be close to 100%. However, for complex systems, it is not reasonable to assume a perfect data separation and, therefore, the percentage proportion of each mode can be somewhere between 30% and 100%, depending on how many operation regimes we are trying to separate. Minimum threshold limit is cannot be defined since it depends also how many labelled clusters are aimed to find. For instance, each operation mode should have a single cluster where its percentage proportion is dominant when comparing the percentages column-wise. For example, in Table 2, the largest mode in cluster 9 in model 1 is loading, which can be regarded as a good result since 49% is a dominant proportion. To calculate how large the percentage proportion is with regard to the column-wise percentage distribution, we can add all the values in the column in question and then perform the following calculation:

$$LargestMode = 100 * 49 / (8 + 49 + 6) = 77\%, \text{ which means that loading is a very dominant operation mode}$$

The idling mode was detected accurately with most of the models, especially with model 1 (RMS and speed), where it was 100% concentrated to one cluster. For prognostic purposes, this can be used when estimating the total production time. For instance, maintenance tasks based on the total operating h. can be postponed if the idling time

increases dramatically.

One interesting finding is that inside any given operation mode, for example loading, there are data instances that are actually labelled as idling. In model 1, for example, 23% of the infection data of the loading cluster belongs to the idling cluster. Actually, this makes sense, since the LHD occasionally stops during loading before choosing the next portion of ore to be loaded.

When comparing the results of each model seen in Table 2, model 1 was the only model which was able to separate each sought-for operation regime in such a way that each operation mode is seen as the largest mode in at least one of the clusters. Although hauling, transit with manual operation ('transit (man)') and transit with automatic operation ('transit (auto)') were rather mixed, this can be considered a good result. Even the fact that loading is separated into four different clusters can be considered a reasonable result, since there are so many different phases included in loading that it would perhaps be wise to divide loading into two or more operation regimes. Fig. 9, model 1 (data collected on Tuesday) clusters are labelled and similar cluster are merged together by using the collected infection data (Table 2). The final model using RMS together with speed was able to detect the different operation regimes. The result indicate that the final model using RMS together with speed can be used to detect operation regimes.

Model 2 and model 4 do not work for the given case study because one of the clusters include a majority of the data points. Furthermore.

The results for model 3 (peak values and speed) show that the separation of regimes was unsuccessful. However, cluster 6 was identified correctly, where the was a large proportion of loading. Therefore, when combining the peak parameter with other parameters, one could obtain some extra information which could be useful.

Model 5, where all the parameters are combined, found only three clusters where all the operation regimes were located, and most of them were in two data clusters. One can conclude from this that combining good parameters (like RMS in this case) and bad parameters (like kurtosis in this case) will lead to poorer results than just using good parameters.

6. Conclusions

There is a need for techniques that can use existing information to estimate external factors such as operation regimes. The proposed method is one such method. It employs an unsupervised clustering technique in condition monitoring data and then infects these data with a smaller data set to label each cluster. It is suggested as a technique useful for the industry, as a larger amount of training data can be collected without needing to know the correct labels for all operation modes beforehand. Using speed and vibration RMS values (model 1) gave reasonable result for the distribution of the operation regimes during production.

The use of common statistical features, such as kurtosis, skewness

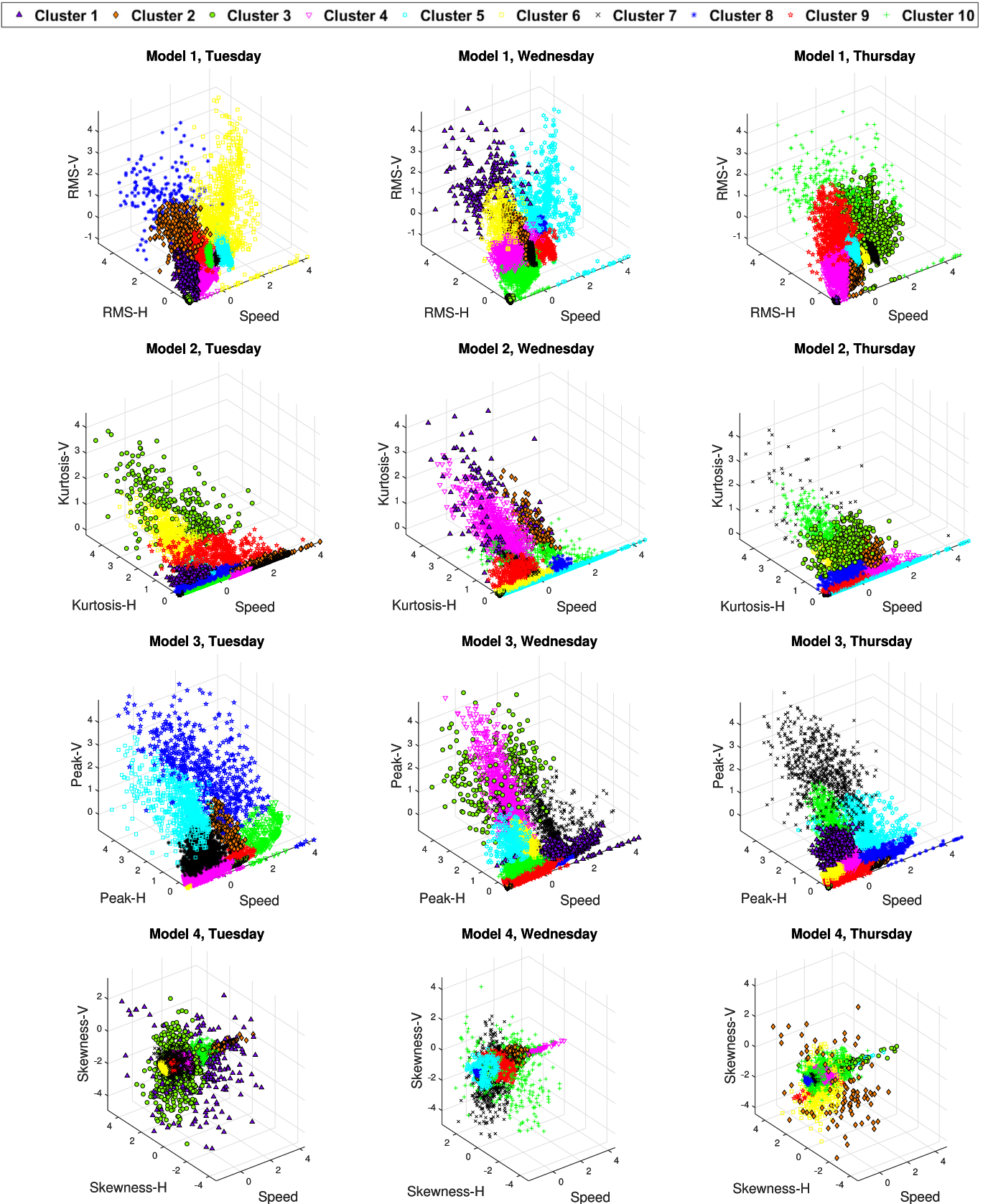


Fig. 8. Results obtained using the VBGM method for data collected on separate days. The speed is the rotation speed of the Cardan axle and features were obtained from acceleration sensors mounted vertically and horizontally on the front axle. The initial K value was 10.

Table 2
Distribution of the validation data (Tuesday).

Operation regime	Model 1 (RMS and speed)										no. of points
	Distribution of data points %										
	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6	cluster 7	cluster 8	cluster 9	cluster 10	
Idling	0	0	0	100	0	0	0	0	0	0	70
Transit (Man)	1	2	0	15	14	47	5	7	8	2	1125
Transit (Auto)	0	0	0	5	39	15	36	1	0	3	97
Loading	7	10	0	23	0	0	0	5	49	5	154
Hauling	0	0	0	16	10	11	56	0	6	0	98
Largest mode	Loading	Loading	Empty	Idling	Transit (Auto)	Transit (Man)	Hauling	Transit (Man)	Loading	Loading	
Operation regime	Model 2 (Kurtosis and speed)										no. of points
	Distribution of data points %										
	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6	cluster 7	cluster 8	cluster 9	cluster 10	
Idling	0	0	100	0	0	0	0	0	0	0	70
Transit (Man)	0	0	73	0	1	19	0	1	6	0	1125
Transit (Auto)	0	0	93	0	0	7	0	0	0	3	97
Loading	0	0	82	0	3	15	0	0	0	0	154
Hauling	0	0	82	0	3	15	0	0	0	0	98
Largest mode	Empty	Empty	Idling	Empty	Loading + Hauling	Transit (Man)	Empty	Transit (Man)	Transit (Man)	Transit (Auto)	
Operation regime	Model 3 (Peak and speed)										no. of points
	Distribution of data points %										
	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6	cluster 7	cluster 8	cluster 9	cluster 10	
Idling	0	0	0	0	0	0	0	100	0	0	70
Transit (Man)	15	0	1	5	10	8	0	23	32	5	1125
Transit (Auto)	22	0	2	0	23	2	0	43	4	4	97
Loading	0	0	0	0	8	36	0	42	14	0	154
Hauling	1	0	0	0	5	2	0	83	7	2	98
Largest mode	Transit (Auto)	Empty	Transit (Auto)	Transit (Man)	Transit (Auto)	Loading	Empty	Idling	Transit (Man)	Transit (Man)	
Operation regime	Model 4 (Skewness and speed)										no. of points
	Distribution of data points %										
	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6	cluster 7	cluster 8	cluster 9	cluster 10	
Idling	49	0	0	0	0	0	51	0	0	0	70
Transit (Man)	85	0	7	1	0	0	3	0	0	4	1125
Transit (Auto)	93	0	4	0	0	2	0	0	1	4	97
Loading	71	0	24	1	0	0	3	0	0	0	154
Hauling	86	0	10	4	0	0	0	0	0	2	98
Largest mode	Transit (Auto)	Empty	Loading	Hauling	Empty	Transit (Auto)	Idling	Empty	Transit (Auto)	Transit (Auto + Man)	
Operation regime	Model 5 (All combined)										no. of points
	Distribution of data points										
	cluster 1	cluster 2	cluster 3	cluster 4	cluster 5	cluster 6	cluster 7	cluster 8	cluster 9	cluster 10	
Idling	0	0	0	100	0	0	0	0	0	0	70
Transit (Man)	0	0	0	33	0	0	0	18	0	50	1125
Transit (Auto)	0	0	0	12	0	0	0	1	0	87	97
Loading	0	0	0	69	0	0	0	1	0	31	154
Hauling	0	0	0	33	0	0	0	0	1	66	98
Largest mode	Empty	Empty	Empty	Idling	Empty	Empty	Empty	Transit (Man)	Hauling	Transit (Auto)	

and peak values, which were extracted from the vibration signals, failed to detect operation regimes using the proposed method. Possibly environmental noise affected the measurements too much. If time domain features are to be used as inputs, some pre-processing filters are needed to reduce the noise and improve the results.

In the proposed approach, the chosen number of clusters (k) should be larger than the number of sought-for operation regimes to avoid having a cluster where only noise is included and two operation modes become mixed as noise is mixed with one of the operation groups. Using the infection method, it is possible to take an operation mode which is divided into two clusters and either merge the two clusters into one

operation regime or treat them as two individual regimes and further analyze why they should be treated as two separate groups.

Future work

Based on the identified operation regimes, future work can investigate the relation of a component wear together with detected operation modes using the studied approach. Furthermore simulations can be done where RUL is estimated using several different operational distribution and can help to decide is it possible to continue the production using fail safe operation modes or should the system be

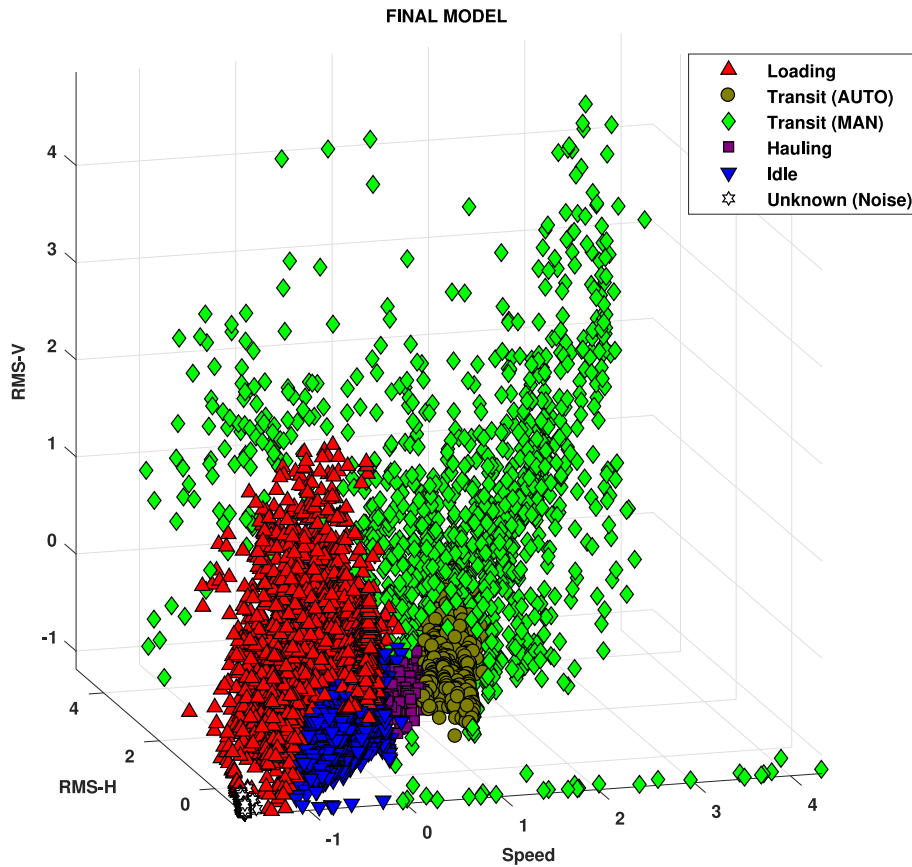


Fig. 9. Final model where clusters (Model 1 Tuesday) are merged based on the collected infection data.

maintained promptly. Therefore more work is needed where these models are used while collecting several degradation data sets in order to truly validate their effectiveness.

In the future proposed method should be tested using process parameters (together with or without vibration features). More emphasis should be placed on feature selection and on the pre-processing step where filtering techniques are also applied without losing the focus of finding input parameters which are generic and do not require extensive pre-processing before they can be applied to many different systems. The infection step should be carried out with more precise data (let the operator do the selection) with a view to gaining a better understanding of the relations between the parameters and the operation regimes and possibly the harmful effects of the various operation regimes on machine health. Furthermore, studies where investigated method is compared with other similar methods, should be done.

Acknowledgements

The authors would like to thank SKF AB, National Instruments, Pyhäsalmi Mine and Sandvik for their contributions and support.

References

[1] Al-Chalabi H, Lundberg J, Ahmadi A, Jonsson A. Case study: model for economic lifetime of drilling machines in the swedish mining industry. *Eng Econ* 2015;60(2):138–54.
 [2] Alamaniotis M, Grelle A, Tsoukalas LH. Regression to fuzziness method for estimation of remaining useful life in power plant components. *Mech Syst Signal Process* 2014;48(1–2):188–98.
 [3] An D, Kim NH, Choi J-H. Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliab Eng Syst Safety* 2015;133:223–36.
 [4] Antoni J, Randall R. The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mech Syst Signal Process* 2006;20(2):308–31.
 [5] Bartelmus W, Zimroz R. A new feature for monitoring the condition of gearboxes in non-stationary operating conditions. *Mech Syst Signal Process*.

2009;23(5):1528–34.
 [6] Benedettini O, Baines TS, Lightfoot H, Greenough R. State-of-the-art in integrated vehicle health management. *Proceedings of the Institution of Mechanical Engineers* 2009;223(2):157–70.
 [7] Bishop CM. *Pattern recognition and machine learning*. springer; 2006.
 [8] Boutemedjet S, Bouguila N, Ziou D. A hybrid feature extraction selection approach for high-dimensional non-gaussian data clustering. *IEEE Trans Pattern Anal Mach Intell* 2009;31(8):1429–43.
 [9] Chamroukhi F, Samé A, Aknin P, Govaert G. Model-based clustering with hidden markov model regression for time series with regime changes. *Neural Networks (IJCNN), The 2011 International Joint Conference on. IEEE*; 2011. p. 2814–21.
 [10] Corduneanu A, Bishop CM. *Variational bayesian model selection for mixture distributions*. Artificial intelligence and Statistics. 2001. Morgan Kaufmann Waltham, MA; 2001. p. 27–34.
 [11] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the em algorithm. *J R Stat Soc Series B (methodological)* 1977:1–38.
 [12] Figueiredo MA, Jain AK. Unsupervised learning of finite mixture models. *Pattern Anal Mach Intell, IEEE Trans* 2002;24(3):381–96.
 [13] Fraley C, Raftery AE. How many clusters? which clustering method? answers via model-based cluster analysis. *Comput J* 1998;41(8):578–88.
 [14] Gustafson A, Schunnesson H, Galar D, Kumar U. The influence of the operating environment on manual and automated load-haul-dump machines: a fault tree analysis. *Int J Min Reclam Environ* 2013;27(2):75–87.
 [15] Guyon I, Elisseeff A. An introduction to variable and feature selection. *J Mach Learn Res* 2003;3(Mar):1157–82.
 [16] Hanafizadeh P, Eshraghi J, Taklifi A, Ghanbarzadeh S. Experimental identification of flow regimes in gas-liquid two phase flow in a vertical pipe. *Meccanica* 2015:1–12. <https://doi.org/10.1007/s11012-015-0344-4>.
 [17] Iverson D.L. *Inductive system health monitoring*2004.
 [18] Jain AK. Data clustering: 50 years beyond k-means. *Pattern Recognit Lett* 2010;31(8):651–66.
 [19] Kotsiantis SB, Zaharakis ID, Pintelas PE. Machine learning: a review of classification and combining techniques. *Artif Intell Rev* 2006;26(3):159–90.
 [20] Laukka A, Saari J, Ruuska J, Juuso E, Lahdelma S. Condition-based monitoring for underground mobile machines. *Int J Ind Syst Eng* 2016;23(1):74–89.
 [21] Lee M-LT, Whitmore GA. Threshold regression for survival analysis: modeling event times by a stochastic process reaching a boundary. *Stat Sci* 2006:501–13.
 [22] *Pattern recognition and machine learning*. *Toolbox*. Mathworks; 2016.
 [23] Peng Z, Peter WT, Chu F. A comparison study of improved hilbert–huang transform and wavelet transform: application to fault diagnosis for rolling bearing. *Mech Syst Signal Process* 2005;19(5):974–88.
 [24] Si X-S, Hu C-H, Kong X, Zhou D-H. A residual storage life prediction approach for

- systems with operation state switches. *IEEE Trans Ind Electron* 2014;61(11):6304–15.
- [25] Si X-S, Wang W, Hu C-H, Zhou D-H, Pecht MG. Remaining useful life estimation based on a nonlinear diffusion degradation process. *IEEE Trans Reliab* 2012;61(1):50–67.
- [26] Suarez EL, Duffy MJ, Gamache RN, Morris R, Hess AJ. Jet engine life prediction systems integrated with prognostics health management. *Aerospace Conference, 2004. Proceedings. 2004 IEEE*. 6. IEEE; 2004. p. 3596–602.
- [27] Timusk M, Lipsett M, Mechefske CK. Fault detection using transient machine signals. *Mech Syst Signal Process* 2008;22(7):1724–49.
- [28] Tse P, Atherton D. Prediction of machine deterioration using vibration based fault trends and recurrent neural networks. *J Vib Acoust* 1999;121(3):355–62.
- [29] Urbaneck J, Barszcz T, Zimroz R, Antoni J. Application of averaged instantaneous power spectrum for diagnostics of machinery operating under non-stationary operational conditions. *Measurement* 2012;45(7):1782–91.
- [30] Walker M, Figueroa F, Toro-Medina J. Phm enabled autonomous propellant loading operations. *Aerospace Conference, 2017 IEEE*. IEEE; 2017. p. 1–11.
- [31] Wu X, Kumar V, Ross Quinlan J, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. *Knowl Inf Syst* 2008;14(1):1–37. <https://doi.org/10.1007/s10115-007-0114-2>.
- [32] Wylomańska A.P.D.A., Śliwiński M.P.. Identification of loading process based on hydraulic pressure signal 2016;:459–466.
- [33] Yang Z-X, Wang X-B, Zhong J-H. Representational learning for fault diagnosis of wind turbine equipment: a multi-layered extreme learning machines approach. *Energies* 2016;9(6):379.
- [34] Yu J, Qin SJ. Multimode process monitoring with bayesian inference-based finite gaussian mixture models. *AIChE J* 2008;54(7):1811–29.
- [35] Yuwono M, Guo Y, Wall J, Li J, West S, Platt G, et al. Unsupervised feature selection using swarm intelligence and consensus clustering for automatic fault detection and diagnosis in heating ventilation and air conditioning systems. *Appl Soft Comput* 2015;34:402–25.
- [36] Zhang Z, Si X, Hu C, Lei Y. Degradation data analysis and remaining useful life estimation: a review on wiener-process-based methods. *Eur J Oper Res* 2018.
- [37] Zimroz R, Wodecki J, Król R, Andrzejewski M, Sliwinski P, Stefaniak P. Self-propelled mining machine monitoring system—data validation, processing and analysis. *Mine planning and equipment selection*. Springer; 2014. p. 1285–94.