

On-board Clutch Slippage Detection and Diagnosis in Heavy Duty Machine

Elisabeth Källström^{1,5}, Tomas Olsson², John Lindström³, Lars Håkansson⁴ and Jonas Larsson⁵

¹ *Department of Engineering Sciences and Mathematics, Luleå University of Technology, Luleå, Sweden*

² *RISE SICS Västerås, Sweden*
tomas.olsson@ri.se

³ *ProcessIT Innovations, Luleå University of Technology, Luleå, Sweden*
john.lindstrom@ltu.se

⁴ *Department of Applied Signal Processing, Blekinge Institute of Technology, Karlskrona, Sweden*
Lars.Hakansson@bth.se

⁵ *Volvo Construction Equipment, Eskilstuna, Sweden*
elisabeth.kallstrom@volvo.com
jonas.jl.larsson@volvo.com

ABSTRACT

In order to reduce unnecessary stops and expensive downtime originating from clutch failure of construction equipment machines; adequate real time sensor data measured on the machine in combination with feature extraction and classification methods may be utilized.

This paper presents a framework with feature extraction methods and an anomaly detection module combined with Case-Based Reasoning (CBR) for on-board clutch slippage detection and diagnosis in heavy duty equipment. The feature extraction methods used are Moving Average Square Value Filtering (MASVF) and a measure of the fourth order statistical properties of the signals implemented as continuous queries over data streams. The anomaly detection module has two components, the Gaussian Mixture Model (GMM) and the Logistics Regression classifier. CBR is a learning approach that classifies faults by creating a new solution for a new fault case from the solution of the previous fault cases. Through use of a data stream management system and continuous queries (CQs), the anomaly detection module continuously waits for a clutch slippage

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

event detected by the feature extraction methods, the query returns a set of features, which activates the anomaly detection module. The first component of the anomaly detection module trains a GMM to extracted features while the second component uses a Logistic Regression classifier for classifying normal and anomalous data. When an anomaly is detected, the Case-Based diagnosis module is activated for fault severity estimation.

1. INTRODUCTION

Being present in a highly competitive business area, the heavy duty construction equipment industry strives to compete effectively with the market challenges by continuously providing better features/systems to meet customer needs and requirements. The customer needs and requirements include, e.g., improved availability and avoided unplanned stops, predictable/proactive maintenance instead of reactive maintenance, as well as highly accurate work planning.

With increasing complexity in the machines, more and more research is directed towards developing intelligent machines where it is possible to automatically (remotely) monitor the health of sub-systems and major components in the machine (Setu et al., 2006). Such a component is the automatic transmission clutches, which may be considered

as a crucial component of its driveline. To reduce service cost and to improve uptime, an on-board data driven detection and diagnosis technique based on real time sensor data from the machine is considered. In this way, the health of the clutch material may be continuously monitored, and if the clutch health starts to degrade a service and/or repair may be scheduled well in advance of a potential clutch failure. The feature extraction and anomaly detection module combined with case-based reasoning (CBR) for on-board clutch slippage detection and diagnosis are implemented via a data stream management system (DSMS) and continuous queries (CQs), which allows numerical analysis packages to be plugged in (Xu, Wedlund, Helgason, & Risch, 2013; Zeitler & Risch, 2011).

In (Olsson, Källström, et al., 2014), we proposed the diagnostic approach shown in Fig. 1. The system consists of the three parts. The first part (1) is the feature extraction module presented in the previous section. The second part (2) is the on-board anomaly detection part. The third part (3) is the off-board case-based fault diagnosis part. The anomaly detection was done by a probabilistic classifier (we used logistic regression) that was trained to recognize normal cases and anomalous cases based on a small sample of known anomalies (faults), while the anomalies were diagnosed by a CBR approach off-line. In this paper, we give a more detailed analysis of each component and we also assess the performance of the parts as a whole while in the previous paper we only tested them individually.

2. AUTOMATIC TRANSMISSION CLUTCHES

A clutch enables the connection and transfer of torque between two rotating shafts when engaged (Lingesten, 2012). Heavy duty equipment generally has an automatic transmission with a multiple disc wet clutch (Lingesten, 2012). A wet clutch is simply a clutch that operates while submerged in a lubricant, and this lowers the friction between the discs as compared to a dry clutch (Mäki, 2005). The multiple disc wet clutches allow engagement while there is a large difference between the rotation speeds of the two shafts (Lingesten, 2012). A multiple disc wet clutch pack consists of separator discs, friction discs, lubricant, piston and two shafts (Lingesten, 2012). Furthermore, multiple disc wet clutches are illustrated in Fig. 2 and explanations are available below:

1. Gear /Output shaft
2. Hub (output shaft side)
3. End plate
4. Friction disc
5. Returning spring
6. Separator disc
7. Drum (input shaft side)

8. Piston
9. Input shaft
10. Lubrication line
11. Bearing

The clutch plates are arranged in such a way that one of the discs is driven by a hub and the other by a drum, see Fig. 1 (Ompusunggu, Papy, Vandenplas, Sas, & Brussel, 2012). The drum and hub are driven by a joint that allows axial movement such as splines and lugs (Ompusunggu et al., 2012). In the clutch pack along its axial direction every other disc is a separator disc and in between the separator discs are the friction disc. The friction material in the friction discs are coated with either paper, asbestos or sintered bronze, while the separator discs are basically steel plates (Lingesten, 2012). The asbestos material is not used anymore due to its high toxicity (Lingesten, 2012).

To engage the clutch, a hydraulic induced normal force is applied to the clutch piston thereby clamping together the friction disc and the separator disc, which allows torque transfer between the two shafts (Mäki, 2005). Clutch discs in the multiple disc wet clutch pack are designed to slip for a defined period of time (slip time) in order not to burn the clutch material due to excessive friction (Berglund, 2013). The friction characteristics of the wet clutches are crucial for the ultimate performance of the automatic transmission because they define how long time the clutches slip during an engagement (Fatima, Marklund, & Larsson, 2013).

Furthermore, a clutch is considered to have failed when it can no longer transmit the desired torque. The level of torque transfer in wet clutches is controlled by the generated friction in it, and a good and stable friction coefficient which keeps output torque at a required level is important (Fatima et al., 2013; Fatima, Marklund, & Larsson, 2012). Thus, clutch slippage is a result of diminishing frictional characteristics of the clutch system (Fatima et al., 2013). The friction characteristics of the clutch material are influenced by different factors such as the clutch material structure, porosity, lubricant and permeability (Berglund, 2013; Marklund, 2010). Furthermore, the coefficient of friction may be affected by sliding speed, varying load, boundary friction, contact temperature of clutch plates and friction due to fluid flow through the friction material (thin-film friction) (Fatima et al., 2012; Devlin et al., 2004).

Thus, degradation of the wet clutch results in a continual drop in the coefficient of friction throughout the clutch service life (Fatima et al., 2013).

To sum up, many factors influence the service life of the multiple disc wet clutches and most of these factors are difficult to isolate and accurately measure. This makes it almost impossible to match the service life condition of the transmission in an actual machine with corresponding test

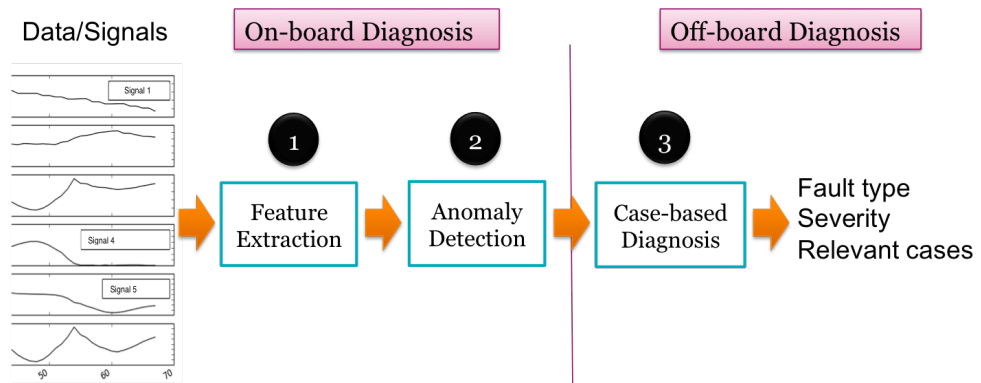


Figure 1. The proposed on-board and off-board diagnosis framework.

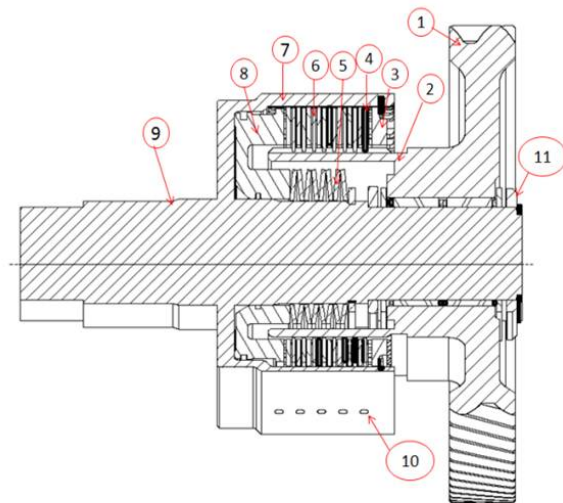


Figure 2. Multiple disc wet Clutch pack 2-D view

rig computation (Kazunari, Akihiko, & Takeshi, 2009). This concerns factors such as temperature of clutch plates, coefficient of friction, torque transfer, drag torque, normal force, oil viscosity, oil quality, oil temperature in the clutch pack, absorbed energy, absorbed energy rate, etc. Kazunari et al. 2009 focused on the degradation level of wet clutches due to temperature and they developed the T-N curve (i.e. temperature vs. frequency of occurrence) for the life calculation of multiple wet clutches (Kazunari et al., 2009). However, the method presented by Kazunari et al. required knowledge of the inner and outer temperatures of the multiple wet clutch pack as well as the S-N curve (i.e. fatigue strength vs. frequency of occurrence) of the metal thermal deformation, which only is possible to measure in a test rig.

Since many of the factors that influence the frictional characteristic of the clutch are only measurable in a test rig but not measurable in today's actual heavy duty machine, this paper addresses the gap in condition monitoring of automatic transmission clutches in an actual heavy duty machine by monitoring the health of the clutch material on-board the machine using the available controller area networks (CAN-bus) signals in the machine together with the feature extraction and anomaly detection module combined with case-based reasoning (CBR) to prevent clutch failure.

3. RESEARCH APPROACH FOR INDUSTRIAL CASE STUDY

The research approach was based on experimentation in an industrial setting using a Volvo L90F wheel loader. The experimental set-up, data collection and extraction, data analysis and feature extraction, and CBR are further described below in this section.

3.1. Experimental Set-up and Data Collection/Extraction

Real time sensor data measurements were logged on the machine with the CAN-bus and broadcasted to the on-board DSMS via a CAN-bus wrapper. The CAN-bus (controllers area network) is the standard message-based protocol which allows different electronic components (e.g. electronic control units, sensors, micro-controllers, actuators, devices, etc.) to communicate (Marx, Luck, Pitla, & Hoy, 2016). Furthermore, the CAN-bus allows data logging from different sensors (Marx et al., 2016).

The signals logged from the machine CAN-bus are the transmission oil temperature, turbine torque, clutch 1 and 2 differential speeds, out-going speed, input speed, turbine speed, off-going slip, on-going slip, engaged gear, gear direction, shifting from 1 to 2 and 2 to 1. The data was logged with a 32-bit CAN-bus at a baud rate of 250 kBaud corresponding to 7.995 Mbits/s. To read the digital data from machine CAN-bus signals, a sampling frequency of 500Hz was used.

Due to too much heat generated in the Forward 2 and 1

clutches of the L90F wheel loader machine, only gear shifts from gear one to two and vice versa were logged for this experiment. To simulate leakage in the clutches, two manual needle valves were installed on the pressure out-takes on the clutch 1 and clutch 2. This enables the adjustment of the oil pressure going to the piston in the clutches. Each of the needle valves can be opened in seven steps (each step corresponding to 360°) simulating different severity of the fault. The system was set up as Fig. 4 shows.

3.2. DATA ANALYSIS/FEATURE EXTRACTION

3.2.1. Higher Order Statistical Properties

Commonly when statistical properties of stochastic processes are studied the mean, autocorrelation, autocovariance, etc. of a process are considered (Bendat & Piersol, 2010). The autocorrelation, autocovariance, etc. are so-called second-order statistical properties, higher-order statistical properties or non-Gaussian properties are of third-order or higher (Manolakis, Ingle, & Kogon, 2000). The moments m_k of a random process $X(n), n \in \mathbb{Z}$ are given by (Papoulis, 1991):

$$m_k = E[x^k(n)] \quad k = 1, 2, \dots$$

where $E[\cdot]$ is the expectation operator and the central moments of random process are defined as:

$$m_k^c = E[(x(n) - E[x(n)])^k] \quad k = 1, 2, \dots$$

The first central moment is always zero, the second central moment is the variance. However, the skewness $\gamma_{x(n)}^3$ is defined as a normalized third central moment according to (Manolakis et al., 2000):

$$\gamma_{x(n)}^3 = E \left[\frac{(x(n) - E[x(n)])^3}{\sigma_{x(n)}^3} \right] = \frac{m_3^c}{\sigma_{x(n)}^3} \quad (1)$$

where $E[\cdot]$ is the expectation operator. The skewness provides a measure on the asymmetry of a probability density function around its mean. Furthermore, the kurtosis $\gamma_{x(n)}^4$ is defined as the normalized fourth central moment subtracted with three and is given by (Manolakis et al., 2000):

$$\gamma_{x(n)}^4 = E \left[\frac{(x(n) - E[x(n)])^4}{\sigma_{x(n)}^4} \right] - 3 = \frac{m_4^c}{\sigma_{x(n)}^4} - 3 \quad (2)$$

The kurtosis gives an indication of the "peakedness" and the "tailedness" of a probability density function (DeCarlo, 1997; Manolakis et al., 2000). For a Gaussian distributed



Figure 3. L90F Wheel Loader



Figure 4. Newly Installed needle valves on the pressure outlets on the Transmission of the L90F complete machine

random process the kurtosis value is 0 (DeCarlo, 1997; Manolakis et al., 2000).

3.2.2. Mean Square Value and Moving Average Square Value Filtering (MASVF)

Usually mean and mean square values of random signals may be estimated with the aid of time averages and/or ensemble averages depending on the underlying physical phenomenon a signal originates from (Bendat & Piersol, 2010). For instance, if the underlying physical phenomenon from which a random signal originates enables time averaging for the estimation of an unbiased and consistent mean square value of the signal, the signal may be considered to be weakly ergodic. Weakly ergodic stochastic processes constitute a subset of weakly stationary stochastic processes (Bendat & Piersol, 2010). For instance, unbiased and consistent estimates of the mean value, the auto correlation and auto covariance of a weakly ergodic stochastic process may be estimated with the aid of time averages (Bendat & Piersol, 2010). The mean value $E[x(n)]$ of a weakly ergodic signal $x(n)$, $n = 1, 2, \dots, N$ may be estimated using a time average, according to:

$$\widehat{E}[x(n)] = \frac{1}{N} \sum_{n=0}^{N-1} x(n) \quad (3)$$

Where n is the discrete time and N is the number of samples included in the time average. The estimate $\widehat{E}[x(n)]$ is an unbiased estimate of the true mean value, $E[x(n)]$ (Bendat & Piersol, 2010). In the same way, an estimate of the mean square value $E[x^2(n)]$ of a weakly ergodic process may be produced as (Bendat & Piersol, 2010):

$$\widehat{E}[x^2(n)] = \frac{1}{N} \sum_{n=0}^{N-1} x^2(n) \quad (4)$$

The variance or the second central moment, $\sigma_{x(n)}^2$, of weakly ergodic stochastic process may now conveniently be estimated as (Bendat & Piersol, 2010):

$$\widehat{\sigma}_{x(n)}^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x(n) - \widehat{E}[x(n)])^2 \quad (5)$$

If a stochastic process $X(n)$, $n \in \mathbb{Z}$ is fourth-order ergodic the kurtosis $\gamma_{x(n)}^4$ may be consistently estimated as (Manolakis et al., 2000):

$$\widehat{\gamma}_{x(n)}^4 = \frac{\frac{1}{N} \sum_{n=0}^{N-1} (x(n) - \widehat{E}[x(n)])^4}{(\frac{1}{N} \sum_{n=0}^{N-1} (x(n) - \widehat{E}[x(n)])^2)^2} - 3 \quad (6)$$

To estimate the mean value, mean square value, etc. for a

non-stationary stochastic process so-called moving time averaging may be utilized (Bendat & Piersol, 2010; Andren, Håkansson, Brandt, & Claesson, 2004). An estimate of a time varying mean value for a signal $x(n)$ with the aid of moving averaging may be produced as:

$$\widehat{E}[x(n)] = \frac{1}{N} \sum_{j=0}^{N-1} x(n-j), \quad n \in \{N-1, N, \dots\} \quad (7)$$

For the selection of the length N of the moving time average the time constant of the non-stationary behavior of the stochastic process and the variance of the estimates have to be considered. Consequently, the moving averaging procedure for the estimation of a time varying mean square value may, e.g., be expressed as:

$$\widehat{E}[x^2(n)] = \frac{1}{N} \sum_{j=0}^{N-1} x^2(n-j), \quad n \in \{N-1, N, \dots\} \quad (8)$$

The moving averaging procedure may for instance be carried out with the aid of a FIR filter having the impulse response:

$$h(n) = \begin{cases} \frac{1}{N} & , 0 \leq n \leq N-1 \\ 0 & , \text{otherwise} \end{cases} \quad (9)$$

The MASVF is realized by filtering squared samples of a signal with an adequate filter (Andren et al., 2004). Thus, an estimate of the time varying mean square value of a signal $x(n)$ may for instance be produced according to the convolution sum:

$$\widehat{E}[x^2(n)] = \frac{1}{N} \sum_{j=0}^{N-1} x^2(n-j)h(j), \quad n \in \{N-1, N, \dots\} \quad (10)$$

The moving average square value filtering not only smoothens random variations of the signal but also gives an indication of the mean square properties of a signal (Andren et al., 2004). The mean square value estimates may also provide information about the stationarity of a signal (Bendat & Piersol, 2010). The moving average filter acts as a low-pass filter over the squared magnitude of the signal, the part of the squared signal that is within the bandwidth of the filter is not attenuated while the part of the squared signal that is outside the bandwidth of the filter is attenuated (Andren et al., 2004). The averaging time defines the length of the filter (Andren et al., 2004). If different time varying properties of the mean square value of a signal are desired, filters with different lengths may be used instead of filters with fixed length (Andren et al., 2004).

3.2.3. The Gaussian Mixture Model and Logistic Regression

A common statistical model for modelling continuously valued data is the multivariate Gaussian mixture model (GMM) (Murphy, 2012). A GMM assumes that cases are generated by a set of clusters of Gaussian distributions. Thus, a GMM is the weighted sum of the set of Gaussian distributions:

$$p(\vec{x}) = \sum_{z=1}^Z p(\vec{x}|z)p(z) \quad (11)$$

$$p(\vec{x}|z) = \frac{1}{\sqrt[2]{2\pi^K |\Sigma_z|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu}_z)^T \Sigma_z^{-1} (\vec{x}-\vec{\mu}_z)} \quad (12)$$

where \vec{x} is a case represented as a numerical vector with length K , Z is the number of clusters, z denotes a specific cluster, $p(z)$ is the probability of the cluster and $p(\vec{x}|z)$ is the likelihood of case \vec{x} conditioned on cluster z , while $\vec{\mu}_z$ is a vector of mean values and Σ_z is the correlation matrix for cluster z , and $|\Sigma_z|$ is the determinant of Σ_z . The parameters Σ_z , $\vec{\mu}_z$, and $p(z)$ are estimated using the Expectation-Maximization algorithm (Dempster, Laird, & Rubin, 1977). A Gaussian mixture model can in principle model any type of distribution with a large enough number of cluster components.

A commonly used algorithm for classifying data is the logistics regression classifier (LRC). LRC is a binary classifier that can separate between two classes (Murphy, 2012). LRC is, as the name tells, a linear classifier, which can be considered a discrete version of linear regression. The LRC probability distributions for two classes $c \in [0, 1]$ given a feature vector \vec{x} is

$$p(c = 1|\vec{x}) = \frac{1}{1 + \exp(-\vec{\omega}^T \vec{x})} \quad (13)$$

$$p(c = 0|\vec{x}) = \frac{\exp(-\vec{\omega}^T \vec{x})}{1 + \exp(-\vec{\omega}^T \vec{x})} \quad (14)$$

where $\vec{\omega}$ is a weight vector with $K + 1$ weights assuming that x has $K + 1$ features including an extra feature that is 1 for all cases. A case is then classified as $c = 1$ if $\vec{\omega}^T \vec{x} \geq 0$ and $c = 0$ otherwise. The logistic regression is well suited to use on board a machine since it is a simple algorithm with a small number of parameters. Also, since it is a discriminative classifier, it makes few assumptions of the distribution on of the independent features of \vec{x} in contrast to generative classifiers where the distribution of \vec{x} is also modeled.

3.3. ANOMALY DETECTION

Anomaly detection is about finding patterns that deviate substantially from what is considered normal (Chandola, Banerjee, & Kumar, 2009). Typically, it is assumed that

the normal cases are much more common than the abnormal and faulty cases, so that what constitutes the normal pattern can be learned. Since not all fault classes need to be known in advance, anomaly detection has become a popular approach for fault detection. There are two approaches to anomaly detection, unsupervised and supervised anomaly detection. In the unsupervised approach, it is assumed that there are no known anomalous cases in which case a model is created based on all data and the cases not fitting the model with respect to a specified criterion are considered anomalous. In the supervised approach there is a small set of known anomalous cases that can be used for training a machine learning model that can be adjusted to an imbalanced data set.

In this work, we assume that there is a large set of cases known to be normal and a relatively small set of cases known to be anomalous. In addition, not all fault classes are known beforehand, so new faults should also be detected. Thus, given these assumptions, an ordinary classifier is not sufficient, and therefore, we use a supervised anomaly detection approach instead.

In the proposed approach, the anomaly detection component is continuously monitoring the vehicle by classifying the extracted signal features into normal or anomalous using a continuous query running on the machine. If a case is considered anomalous, the signals are sent off-board for further analysis. In addition, the anomaly detection method should be fast and light-weight, since it should be able to handle continuous streams of data on-board a machine. However, it is not required that the anomaly detection model is created on-board, so that is done off-board in the current setup.

A common way of doing anomaly detection is to fit a statistical model to the non-anomalous cases and then, by choosing a suitable threshold, classify cases above the threshold as normal and below the threshold as anomalous since they are unlikely (Chandola et al., 2009). So, in this simple statistical approach \vec{x} is defined to be anomalous if $p(\vec{x}) < \alpha$ where α is a small threshold that is selected using the anomalous cases and $p(\vec{x})$ is the probability of \vec{x} given the statistical model. This can be formulated as a probability distribution as follows ($c = 1$ means \vec{x} is anomalous):

$$p(c = 1|\vec{x}) = 1 \text{ if } p(\vec{x}) < \alpha \text{ else } 0 \quad (15)$$

Cases below the threshold are unlikely to have been generated by the statistical model and therefore considered anomalous. However, this can be seen as a binary classification problem with two outcomes and where the probability $p(\vec{x})$ is the single input to the classifier.

Generalizing from the above approach, we can transform the problem into a statistical classification problem, where instead a soft threshold is learned by training a probabilistic

classifier using the output of the statistical model as input. Thus, the statistical model is used to generate feature to the classifier. By fitting the non-anomalous data to a statistical model and then fit a probabilistic classifier to the probability distribution of the classifier, we get two advantages: (1) the degree of anomaly is now measured as a probability (not only yes or no), and (2) the threshold is part of the statistical model and can be automatically learned. In addition, by using the statistical model of non-anomalous cases for feature generation, we also hypothesize that it will be easier to detect faults from unknown fault classes compared to training a classifier directly on the original features. As statistical model of the normal data, we use a set of GMMs and as a probabilistic classifier, we use logistic regression. So, the GMMs are only trained on normal data while the logistic regression is fitted on both anomalous and normal data and thereby, automatically learns the thresholds for deciding when a case is anomalous or not.

Normally, in case of anomaly detection, GMMs are fitted to all features, while in our approach, the GMMs are fitted to each feature independently of the other features. Thus, each GMM can measure the anomalousness of each feature independently of the others. Output is then a new feature vector with the log-likelihood of each pair of feature value and cluster. For instance, 5 signals where 5 features are extracted from each signal will result in 25 features in total. In addition, fitting a GMM with 5 cluster components to each extracted feature will result in a total of 125 log-likelihoods. Thus, the GMMs can also be seen as a way of discretizing the data into a vector of the same length as the number of clusters with a value for each cluster. However, if there are any dependencies between features, the logistic regression will at least partially take that into account. Thus, the logistic regression uses the log-likelihood features to learn to separate between normal and anomalous signals.

3.4. CASE-BASED REASONING (CBR)

Given that the anomaly detection module detects an anomaly, we use CBR to make a diagnosis of the anomaly. CBR makes the assumption that similar problems tend to have similar solutions (Aamodt & Plaza, 1994). Thus, CBR assumes that new solutions to a new problem can be constructed by retrieving a set of previous solution to similar problems. The most commonly used CBR algorithm is the k-nearest neighbor algorithm (Aha, Kibler, & Albeit, 1991). Similar to our previous work, we use CBR, not only for fault diagnosis, but also to support manual decisions by presenting the set of most relevant cases (Leake & McSherry, 2005).

CBR, in contrast to model-based approaches, does not generalize into a model, but makes predictions directly from the cases. Therefore, an advantage of using CBR compared

to model-based approaches is that, if the classification algorithm does not propose a good solution, a CBR-based approach can nevertheless support experts in finding a diagnosis by being able to retrieve and present the set of the most relevant cases. Thereby, CBR can support manual decision making in addition to automatic fault classification.

A CBR approach requires a measure of similarity between cases, and as in our previous papers, we define the similarity between two cases as how similarly they deviate from the normal cases with respect to a statistical model. For measuring the similarity between cases, we, as before, use the symmetric Kullback-Leibler divergence (J-divergence) (Kullback & Leibler, 1951). The J-divergence is a statistical measure for comparing the similarity between two probability distributions. For statistically modeling the normal cases, we also as before fit a GMM to each feature independent of the other features (Olsson, Gillblad, Funk, & Xiong, 2014).

Subsequently, we compare two cases as the difference between the probabilities of the clusters given the cases. Therefore, let \vec{z} be a vector with one cluster for each feature $k = 1, \dots, K$ then the J-divergence between two cases \vec{x}_i, \vec{x}_j with respect to the distribution of \vec{z} is as follows:

$$\begin{aligned} J(\vec{x}_i, \vec{x}_j) &= \sum_{\vec{z}} \log \frac{p(\vec{z}|\vec{x}_i)}{p(\vec{z}|\vec{x}_j)} \left(p(\vec{z}|\vec{x}_i) - p(\vec{z}|\vec{x}_j) \right) \\ &= \sum_{k=1}^K \sum_{z^k=1}^Z \log \frac{p(z^k|x_i^k)}{p(z^k|x_j^k)} \left(p(z^k|x_i^k) - p(z^k|x_j^k) \right) \\ &\leq \sum_{k=1}^K \sum_{z^k=1}^Z \left| \log(p(x_i^k|z^k)) - \log(p(x_j^k|z^k)) \right| \end{aligned}$$

where $p(\vec{z}|\vec{x}_i) = p(z^1|x_i^1) \cdot p(z^2|x_i^2) \cdot \dots \cdot p(z^K|x_i^K)$ (assuming independence of cluster z^k given feature k and value x_i^k) and $p(z^k|x^k) = \frac{p(x^k|z^k)p_{z^k}}{p(x^k)}$ and $\log \frac{p(\vec{z}|\vec{x}_i)}{p(\vec{z}|\vec{x}_j)} = \log p(\vec{z}|\vec{x}_i) - \log p(\vec{z}|\vec{x}_j)$. The less-than-equal is valid since $\log p(z^k|x_i^k) - \log p(z^k|x_j^k) = \log p(x_i^k|z^k) - \log p(x_j^k|z^k) + \log p(x_j^k) - \log p(x_i^k)$ (last terms are independent of z^k and thereby canceled out in the sum) and $\max(|p(z^k|x_i^k) - p(z^k|x_j^k)|) = 1$. Thus, from the J-divergence, we can derive, as an upper bound, the Manhattan distance with respect to the log-likelihood of each cluster. As final metric for comparing cases, we will use the Manhattan distance with normalized and weighted log-likelihood features. The weights were estimated using the average of the maximum information coefficient (MIC) between the classes or severity and the normalized features (Murphy, 2012). Then the resulting metrics is as follows:

$$d(\vec{x}_i, \vec{x}_j) = \sum_{k=1}^K \sum_{z^k=1}^Z \omega_k \left| \log p(x_i^k|z^k) - \log p(x_j^k|z^k) \right|$$

Table 1. Kurtosis values when there is no clutch slippage

Diff. Speed1	Diff. Speed2	Output Speed	Turbine Speed	Turbine Torque
5.634	4.9055	5.6837	3.5152	3.5172
6.5279	5.8958	4.1985	4.3244	4.1297
7.1882	6.6851	4.4596	4.7167	5.1261
2.4357	2.2671	3.8482	1.8557	1.8729

where normalization and MIC are included in ω_k for each feature.

For making off-board fault classification, we would use the k -nearest neighbor algorithm that makes a majority vote to classify the anomalous cases, but where an anomalous case is sent to manual investigation if the percentage of nearest neighbors voting for the class is less than a threshold of 90%. Thus, we require at least 9 out of 10 neighbors to be of the same class to accept it as the final classification. For severity estimation, we use the average severity of the retrieved set of cases.

4. RESULTS

4.1. Mean Square Value and Sliding Mean Square Value filtering

With needle valve 2 alone fully opened the clutch hydraulic system compensated for the leakage and no slippage was observed. So, needle valve 2 was kept fully opened and valve 1 was opened gradually to give different degrees of valve opening indicating clutch slippage. In order to detect slippage, the engagement part of the clutch 1 differential speed signal for each gear shifts were passed through the moving average square value filter. The averaging length of 10 samples was used. After filtering, the absolute mean square value of each differential speed signal was subtracted from the filtered signal to clearly give an indication of slippage as presented in Fig. 5.

4.2. Higher order Statistics:Kurtosis

The kurtosis of the engagement part of each signal is estimated and the results given in tables 1 and 2.

The results from the kurtosis values in tables 1 and 2 shows a variation for the different signals for clutch slippage and non-clutch slippage.

4.3. Detecting Anomalies

For evaluation of the anomaly detection, we have collected 389 cases of which 110 are fault cases with valve opening 0-7 where 0 indicates a normal case and 7 indicates the fault with the highest severity. In case of anomaly detec-

Table 2. Kurtosis values when there is clutch slippage

Diff. Speed1	Diff. Speed2	Output Speed	Turbine Speed	Turbine Torque
1.811	2.0128	3.2147	1.6865	1.7863
1.9542	2.0576	4.6225	1.8431	2.2753
1.6627	1.8836	3.8121	1.6731	1.8809
1.6541	1.7516	2.2406	1.6332	1.7420
1.7767	1.6713	2.8705	1.5620	1.7888

Table 3. The average precision (AP) for the clutch slip diagnosis.

Feature Type	AP (3 features)	AP (5 features)
Original features	0.934	0.926
BIC cluster	0.947	0.898
AP cluster (2 and 4)	0.925	0.923

tion, we assumed the two classes: normal (0 valve openings) and anomalous (1-7 valve openings) cases. We fitted the GMMs and the logistic regression 10 times on randomly split data where the training data constituted 80% of the normal data and 20% of the anomalous data and for testing, we used the remaining data.¹ When fitting the logistic regression, we also used $l1$ norm regularization with 5-fold cross-validation on the training data, and for managing the imbalance between the classes in the data since there are many more normal data points than anomalous, we trained the logistic regression using a cost function where the classes are weighted proportional to the size of each class. The performance was measured using the Average Precision score (AP) and the Precision-Recall curves (PRC) (Saito & Rehmsmeier, 2015). The AP corresponds to the area under the PRC.

In Table 3, the result from applying the anomaly detection algorithm to a test set is shown. The signal features used in the 3 feature set-up consists of the signal length, its mean value and standard deviation. The signal features used in the 5 features set up also included the kurtosis and the maximum of the sliding mean square value filtering of the clutch 1 differential speed signal. The meaning of each row is as follows: *Original features* approach means that no clustering is done, and only the original 3 or 5 features were used. *BIC cluster* approach refers to automatically selecting the number of used clusters per features by testing the model fit of the GMM to the data using the Bayesian Information Criterion (BIC) measure (Schwarz et al., 1978). In contrast, the

¹Recall from previous section that the GMM is fitted only to the normal data, while logistic regression to both normal and anomalous data.

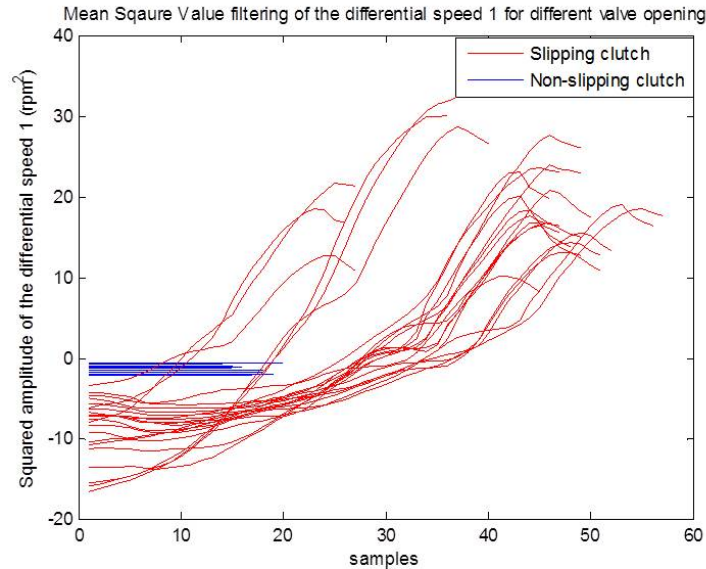


Figure 5. The Sliding Mean Square value filtering of clutch 1 differential speed showing clutch slippage and non-clutch slippage

AP cluster approach selects the number of clusters with the largest AP for a validation set when using the same number of clusters for every signal feature (the BIC cluster selects a unique number of clusters per feature).

The results are shown Fig. 6 in form of PRC curves for each approach to clustering.² The further up to the right the curves are, the better is the showed performance. The curves look quite similar, but apparently, the best AP score is 0.947 for the BIC cluster with 3 features. However, what is striking is that the BIC cluster for 5 features has the worst AP score of 0.898, while the difference between 3 or 5 features in the other cases are small. Adding the two extra features was bad for the BIC cluster, but the 5 features was also worse in the two other cases, although only a little worse. Regardless, the result shows that we are indeed able to detect the oil leakage using the proposed approach. In the next section, we will investigate the performance when we also add diagnosis after the anomaly detection.

4.4. Case-based Prediction

For evaluation of the CBR diagnosis component, we have used the same data set as for anomaly detection. However, since the data set only have a single type of fault but with varying severity the evaluation can only be restricted

²Notice that the Recall is only approximately equal to zero when the Precision is approximately equal to one.

to severity estimation. In this evaluation, we diagnose the output from the anomaly detection component, and thus, false positives must be managed. In this case, since we are testing the fault diagnosis and the anomaly detection in combination, we increase the training data to include 50% anomalous data but still with 80% normal data as when the anomaly detection was evaluated. When training the CBR, we have selected the number of neighbors using 10-fold cross validation. In this case, the k-nearest neighbor algorithm has to be trained using the same data as was used for training the anomaly detection algorithm. However, it is not straightforward whether to add all training cases or a subset since there are more normal cases than the anomalous cases. Thus, we tried three different ways of managing the normal cases in the training set: (1) add all normal cases, (2) add no normal cases, and (3) add only normal cases that are misclassified by the anomaly detection module. Table 4 shows the average MSE of the test sets for the different clustering approach. We only show for the 3 features set up. As can be seen, the MSE is lowest for the second way of managing the normal cases in the training set. Thus, in this case, it is better to add no normal cases. However, this will probably change as we get more cases. In Table 5, we show the performance for the different valve openings and the valve opening 1 had the least while 7 has the highest error. We see a similar pattern for the mean absolute error. However, we can only conclude that it is really hard to

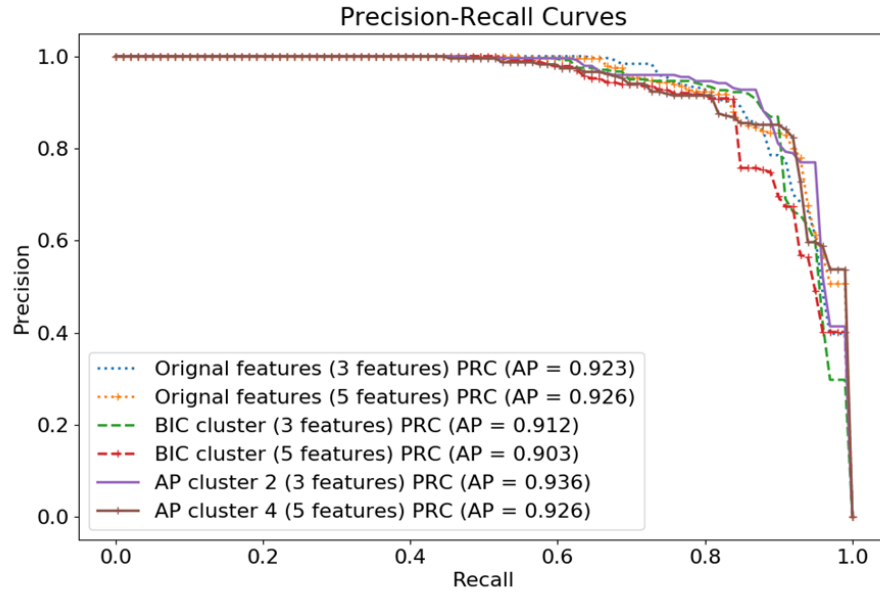


Figure 6. Precision-Recall curves (PRC) for anomaly detection with Average Precision scores (AP) as point values.

Table 4. Result from running both anomaly detection and CBR diagnosis measured in MSE, (1) all normal cases, (2) no normal cases, and (3) only misclassified normal cases.

	All (1)	No (2)	Miss. (3)
Original features	8.42	7.60	8.11
BIC cluster	8.54	7.56	8.12
AP cluster (6, 10, and 10)	8.32	7.52	7.75

predict the larger valve openings from the data.

To summarize the most important results, we find the use of combining feature extraction, anomaly detection and CBR as a useful framework for clutch slippage detection and diagnosis. Further, we have verified and extended previous research into a framework usable for other sub-systems or major components in heavy duty machines or construction equipment.

5. DISCUSSIONS AND CONCLUSIONS

The main results of the study are a framework with feature extraction and anomaly detection combined with CBR for clutch slippage detection and diagnosis. The above results, which are enabled by the use of on-board sensor technology and distributed analytics through a DSMS and CQs, demonstrates that clutch slippage patterns of the automatic

Table 5. Mean Absolute Error and Mean Square Error for each valve opening

Valve Opening	Mean Absolute Error	Mean Squared Error
0	0.044	0.044
1	0.0	0.0
2	1.0	1.0
3	2.11	4.55
4	3.0	9.0
5	4.21	17.9
6	5.19	27.1
7	6.03	36.4

transmission clutches in an actual heavy duty construction equipment can be detected using a Moving Average Square Value filter combined with a measure of the higher order statistics, Kurtosis. It has also been established that the clutch slippage patterns can further be diagnosed into different severity of fault cases, i.e. levels of valve openings, using the anomaly detection module and case-based reasoning. However, we showed that detecting the presence of valve openings is easy but to estimate the severity is not easy. We also showed that fitting a GMM on each individual

feature improved the anomaly detection but did not largely affect the CBR performance. However, more data is needed to draw any certain conclusions. The detection and diagnosis framework is of a general nature, and can be transferred to other settings and machines - however, the specific data collection and analysis methods applied may need to be exchanged for adequate ones fitting the context and specific needs.

Regarding related future work, it would be interesting to continue with monitoring the health of Automatic Transmission Fluids (ATF) on-board via additional sensors attached to the machine, since the oil has a lot of information concerning the health of the automatic transmission clutches (Fatima et al., 2012). In practical settings, to offer customers a possibility to monitor a fleet of machines with diagnostics for each machine may save both money and time, as for instance potential clutch failures may be predicted well in advance before they occur. Being able to predict potential problems to a large extent allows for acting in a proactive manner and planning the maintenance instead of doing reactive maintenance when something has already occurred or broken down. Other components, which are critical for the availability of the machine and its function, could further be monitored as well in order to improve the customers' productivity and the availability level of the construction equipment. The data-driven approach, which has been developed, is generic and can thus be applied to other components other than the clutches. In this way, several critical components on a machine can be continuously monitored.

The result from this work may also be used to support new and emerging business models, which for instance require fleet management and monitoring to be able to predict problems and act proactively related to maintenance and long-term management of operations. Offers based on these emerging business models may also be sold with availability, result or productivity clauses, requiring the provider to take additional costs and manage additional responsibility and risk - and consequently being compensated for that. Examples of such emerging business models are: Product-Service Systems/Industrial Product-Service Systems (Meier, Roy, & Seliger, 2008) and Functional Products (Lindström et al., 2013). In order to stay ahead in the global competition and meet the customers' needs, it is required that corporations develop their core competences and customer offers. To successfully develop core competences, technology and customer offers, it is necessary to learn more on what is offered, how it is used in the customer applications and how to keep it operating exceeding the customers expectations.

ACKNOWLEDGEMENTS

This work has been partially supported by the FP7 EU Large Scale Integrating Project SMART VORTEX (Scalable Semantic Product Data Stream Management for Collaboration and Decision Making in Engineering) co-financed by the European Union.

For more details, visit <http://www.smartvortex.eu/>. Further, the Swedish Knowledge Foundation (KK-stiftelsen) through ITS-EASY Research School and Swedish Governmental Agency for Innovation Systems (VINNOVA) grant no 10020 and JU grant no 100266 making this research possible.

REFERENCES

- Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues. In *Ai communications*.
- Aha, D., Kibler, D., & Albeit, M. (1991). Instance-based learning algorithms, machine learning. In *Machine learning*.
- Andren, L., Håkansson, L., Brandt, A., & Claesson, I. (2004). Identification of dynamic properties of boring bar vibrations in a continuous boring operation. In *Mechanical systems and signal processing*.
- Bendat, J., & Piersol, A. (Eds.). (2010). *Random data; analysis and measurement procedures*. Wiley.
- Berglund, K. (2013). *Predicting wet clutch service life performance*, (Unpublished doctoral dissertation). Luleå University of Technology.
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 79-86.
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. In *Psychological methods*.
- Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the em algorithm. In *Journal of the royal statistical society*.
- Devlin, M., Tersigni, S., Senn, J., Turner, T., Jao, T., & Yatsunami, K. (2004). Effect of friction material on the relative contribution of thin-film friction to overall friction in clutches. In *Sae international*.
- Fatima, N., Marklund, P., & Larsson, R. (2012). Water contamination effect in wet clutch system. In *Proceedings of the institute of mechanical engineers, part d, journal of automobile engineering*.
- Fatima, N., Marklund, P., & Larsson, R. (2013). Influence of clutch output shaft inertia and stiffness on the performance of the wet clutch. *Tribology Transactions*, 56(2), 310-319.
- Kazunari, O., Akihiko, F., & Takeshi, H. (2009). Proposal of field life design method for wet multiple plate clutches of automatic transmission on forklift-trucks. In *Sae international*.
- Kullback, S., & Leibler, R. (1951). On information and suf-

- iciency. *The Annals of Mathematics Statistics*, 22(1), 79-86.
- Leake, D., & McSherry, D. (2005). *Introduction to the special issue on explanation in case-based reasoning* (Vol. 24; Tech. Rep. No. 2). Artificial Intelligence Review,.
- Lindström, J., Plankina, D., Nilsson, K., Parida, V., Ylinenää, H., & Karlsson, L. (2013). Functional products: Business model elements. In *Proceedings of 5th cirp international conference on industrial product-service systems*.
- Lingesten, N. (2012). *Wear behavior of wet clutches* (Unpublished doctoral dissertation). Luleå University of Technology.
- Mäki, R. (2005). *Wet clutch tribology; friction characteristics in limited slip differentials* (Unpublished doctoral dissertation). Luleå University of Technology.
- Manolakis, D., Ingle, V., & Kogon, S. (Eds.). (2000). *Statistical and adaptive signal processing*. McGraw-Hill.
- Marklund, P. (2010). Permeability measurements of sintered and paper based friction materials for wet clutches and brakes. In *Sae*.
- Marx, S., Luck, J., Pitla, S., & Hoy, R. (2016). Comparing various hardware/software solutions and conversion methods for controller area network (can) bus data collection. *Computers and Electronics in Agriculture*, 128(3), 141-148.
- Meier, H., Roy, R., & Seliger, G. (2008). Industrial product-service systems - ips2. *CIRP Annals Manufacturing Technology*, 1-24.
- Murphy, K. (2012). *Machine learning: a probabilistic perspective* (Tech. Rep.). MIT Press.
- Olsson, T., Gillblad, D., Funk, P., & Xiong, N. (2014). Explaining probabilistic fault diagnosis and classification using case-based reasoning. In *Proceedings of case-based reasoning research and development international conference on case-based reasoning*.
- Olsson, T., Källström, E., Gillblad, D., Funk, P., Lindström, J., Håkansson, L., ... Larsson, J. (2014). Fault diagnosis of heavy duty machines: Automatic transmission clutches. In *Proceedings of workshop on synergies between cbr and data mining at 22nd international conference on case-based reasoning*.
- Ompusunggu, A., Papy, J., Vandenplas, S., Sas, P., & Brusel, H. (2012). Condition monitoring method for automatic transmission clutches. In *International journal of prognosis and health management*.
- Papoulis, A. (Ed.). (1991). *Probability, random variables, and stochastic processes*. McGraw-Hill.
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLoS one*, 10(3), e0118432.
- Schwarz, G., et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2), 461-464.
- Setu, M., Wilcutts, M., Chigusa, S., Qiao, L., Choi, K., & Pattipati, K. (2006). Systematic data-driven approach to real-time fault detection and diagnosis in automotive engines. In *Proceedings of IEEE*.
- Xu, C., Wedlund, D., Helgesson, M., & Risch, T. (2013). Model-based validation of streaming data. In *Proceedings of the 7th ACM international conference on distributed event-based systems, DEBS*.
- Zeitler, E., & Risch, T. (2011). Model-based validation of streaming data. In *Proceedings of the VLDB Endowment*.

BIOGRAPHIES

Elisabeth Källström received the M.Sc. degree in Electrical Engineering with emphasis in Signal Processing from Blekinge Institute of Technology, Karlskrona, Sweden, in 2012. She also has a Licentiate degree from Luleå University of Technology, Luleå, Sweden. She is currently pursuing the PhD degree in the division of Product and Production Development at Luleå University of Technology, Luleå, Sweden. Her research interest include, on-board condition monitoring of driveline parts, data mining of big data, engineering vibrations and diagnostics. She is a member of the International Journal of Acoustics and Vibration (IJAV). She is currently employed at Volvo CE as a Diagnostic Engineer where she works with on-board diagnostics.

Tomas Olsson obtained a PhD degree from Mälardalen University and he is a senior researcher at RISE SICS Västerås, with a licentiate degree from Uppsala University (2006) and M.Sc. from KTH (1998). His research interests are in applied AI, statistical machine learning, Deep learning, Bayesian statistics and Case-Based Reasoning. He has been working at SICS as researcher in various projects since 1998 where he has done AI-related research in domains like multi-agent systems, recommender systems, service composition, SLA management, security analysis, and fault analysis.

John Lindström associate professor, received his PhD in information systems science at Luleå University of Technology, Sweden. He manages a Research and Development centre, ProcessIT Innovations, at Luleå University of Technology and his research interests include development processes, availability matters on system and organizational level, as well as modeling and simulation applied in industrial development processes. One of his main research interests lies in the area of product development for function provision i.e. functional product development. He has published research papers at international journals and conference proceedings. Prior to joining academy, he worked for 15 years in different industries in both management and

specialist positions with product, service, process, and business development.

Lars Håkansson received the M.Sc. degree in Electrical Engineering from Lund University of Technology, Lund, Sweden, and the Ph.D. degree, in Mechanical Engineering from Lund University of Technology, Lund, Sweden, in 1989 and 1999, respectively. He joined Blekinge Institute of Technology (BTH) and was appointed Senior Lecturer in Electrical Engineering and continued to expand his research within the area of noise and vibration control. In 2005 he was appointed associate professor and received the responsibility, as principal researcher and advisor, for the active control group at the Department of Signal Processing, BTH. Dr. Håkansson is a professor at the department of Electrical Engineering (former Department of Signal Processing), at BTH. Currently, he is the research director for the signals systems sensors and remote engineering group, S3 R, at BTH. His current research interests are in Signal Analysis, Signal Processing, Adaptive Signal Processing,

Active Noise and Vibration Control, Automatic Control, Remotely Controlled Laboratories, Analytical and Experimental Modeling of Mechanical and Acoustic Systems. He has a keen interest on developing new technology and his research is generally in collaboration with industry, which led to several patents. Lars Håkansson is a member of the Scandinavian vibration association (SVIB) and a member of the Editorial Board of the Journal of Advances in Acoustics and Vibration.

Jonas Larsson reached his PhD degree in Mechanical Engineering at Linköping University in Sweden 2003. He has since then worked at Volvo Construction Equipment as among others Chief Project Manager for research projects, which is also his current position. As part of that role he has set up the Volvo part of the Smart Vortex project in which the major part of the research of this paper has been performed.