Author(s)	Krogerus, Tomi; Hyvönen, Mika; Huhtala, Kalevi
Title	Recognition of Operating States of a Wheel Loader for Diagnostics Purposes
Citation	Krogerus, Tomi; Hyvönen, Mika; Huhtala, Kalevi 2013. Recognition of Operating States of a Wheel Loader for Diagnostics Purposes. SAE International Journal of Commercial Vehicles vol. 6, num. 2, 1-7.
Year	2013
DOI	http://dx.doi.org/10.4271/2013-01-2409
Version	Publisher's PDF
URN	http://URN.fi/URN:NBN:fi:tty-201404291150
Copyright	2013 SAE International

2013-01-2409
Published 09/24/2013

doi:10.4271/2013-01-2409
saecomveh.saejournals.org

Recognition of Operating States of a Wheel Loader for Diagnostics Purposes

Tomi Krogerus, Mika Hyvönen, and Kalevi Huhtala
Tampere University of Technology

ABSTRACT

In this paper, the operating states of a wheel loader were studied for diagnostics purposes using a real time simulation model of an articulated-frame-steered wheel loader. Test drives were carried out to obtain measurement data, which were then analyzed. The measured time series data were analyzed to find the sequences of operating states using two different data sets, namely the variables of hydrostatic transmission and working hydraulics. A time series is defined as a collection of observations made sequentially in time. In our proposed method, the time series data were first segmented to find operating states. One or more segments build up an operating state. A state is defined as a combination of the patterns of the selected variables. The segments were then clustered and classified. The operating states were further analyzed using the quantization error method to detect anomalies. The recognized operating states define the operation of the machine so the analysis can be focused on specific sections and situations in time series and to identify which kinds of operating situations generate anomalies. Simulated leakages in the main hydraulic components of the hydrostatic transmission and the working hydraulics were used as anomalies to study the changes in the recognized operating states and the magnitude of the quantization error.

CITATION: Krogerus, T., Hyvönen, M., and Huhtala, K., "Recognition of Operating States of a Wheel Loader for Diagnostics Purposes," *SAE Int. J. Commer. Veh.* 6(2):2013, doi:10.4271/2013-01-2409.

INTRODUCTION

The diagnostics of machinery require analysis of the time series of sensor and control signals. In order to analyze the time series data, different operating states which describe the transients of the measured variables need to be first detected. After that, some higher level description should be used to further analyze these operating states.

Finding operating states from the measured time series data requires segmentation of the time series data. Segmentation [1,2] is a method which allows the dividing of time series data into smaller groups of data sets which describe the patterns of the measured variables. In segmentation, the time series data are transformed into piecewise representation. A segment is a contiguous subset of a time series.

Time series segmentation is often used as a pre-processing step in time series analysis applications. Time series segmentation is exploited in a wide range of applications: for example, medical diagnostics, analysis of financial time series, speech processing, or sensor signal analysis [1]. Typically, in modern machinery there is already a lot of information available about the operation of the machine, e.g. process and control data through communication buses, which can be used in analysis. In addition to this, condition monitoring specific sensors can also be added to the system. When all this information from communication busses and additional sensors are recorded, it leads to the generation of a huge amount of data. High dimensionality complicates the processing of time series data, especially from the pattern recognition point of view [3].

After segmentation, the segments have to be organized into groups of similar members so that data can be classified against these groups. This grouping process is often referred to as clustering [4,5]. A cluster is a collection of objects which are similar between them and are dissimilar to the objects belonging to other clusters. These clusters are then interpreted to be operating states. Diagnostics can now be focused on these operating states or even a single interesting operating state. Most information in regard to the detection of anomalies is usually obtained from operating states which have the biggest changes in the analyzed variables [6].

RESEARCH PLATFORM AND SIMULATOR SYSTEM

The studied wheel loader, called GIM-machine, is a modified version of the Avant multipurpose wheel loader. It has been designed to serve as a platform for different types of research to be conducted. The frame of the machine is original, but the control system, electronics and hydraulics have been changed for research purposes. The research platform is shown in figure 1.

Figure 1. Research platform.

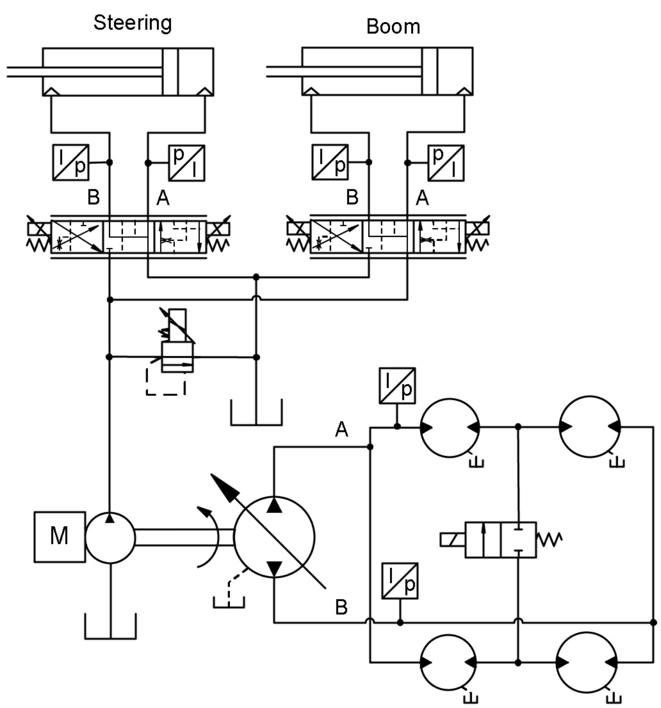

Figure 2. Simplified hydraulic circuit.

Figure 2 shows a simplified hydraulic circuit of the hydrostatic transmission (HST) and the working hydraulics of this research platform. The hydrostatic transmission is implemented with an electronically actuated variable displacement pump and fixed displacement motors. Steering is executed by a hydraulic cylinder. The stroke of the steering cylinder is controlled by means of an electronically actuated proportional valve as well as other cylinders (boom, telescope and bucket) in the working hydraulics. The telescope and the bucket have been left out off the figure for purposes of clarity. Separate fixed displacement pumps are used for the steering and for the working hydraulics.

A real time Hardware-In-the-Loop (HIL) simulation model [7] for the GIM machine has been developed utilizing Matlab/Simulink environment for the GIM research project (www.gim.tkk.fi). It is used mainly in the development of the control and the hydraulic systems of future autonomous mobile machines. The model and its sub models, such as the hydraulic component models, are verified by laboratory measurements and the model has proved its capability to even tune the control parameters of the control system. From this perspective, it is also very well suited for fault diagnostic research. In this paper, this real time HIL simulation model has been utilized.

TIME SERIES ANALYSIS

At the training phase of operating state recognition it is usually easy to make predetermined work movements and to direct the analysis and teaching measures only onto data which is collected in this way. However, after the training stage the work machine will be run in the sequences determined by its operating situations, in which case the measurement data are obtained from previously undefined driving situations. Situations which appear occasionally and correspond to the training data have to be separated from the measured time series data; in other words, certain operating states that have been defined in advance. The recognition of specific operating states is emphasized because the system, and the measurement data which is obtained from it, behave naturally in a quite different way in different driving situations, and especially when there are separate operators.

In the choice of operating states to be analyzed, a noteworthy point is the appearance of the effects of anomalies at different stages of the driving sequences in the measurement data. Earlier research results [6] have indicated that the biggest effects of anomalies will be at the transient stages of the system, where changes in the pressure and volume flows are at their biggest. After the transient stage, the effect of anomalies on the measurement results disappears, or at least decreases significantly. Thus, the analysis should be directed only to that part of the measured data where the deviations can be detected as being the clearest. Then most information in regard to the detection of anomalies of this kind is obtained from these operating states. Based on these earlier results, test drives were defined here to include fast accelerations, decelerations, and lifting and lowering the boom.

In the method used in this study to find interesting operating states from the time series data, the data are segmented to find operating states. After the segmentation, pre-processed segments are clustered into different operating states. A state is defined as a combination of the patterns of the selected variables. After the clustering process, analysis can then be focused on specific operating states. The operating states were further analyzed using the quantization error method. The method is illustrated in figure 3 and will be explained in more detail in the following sections.

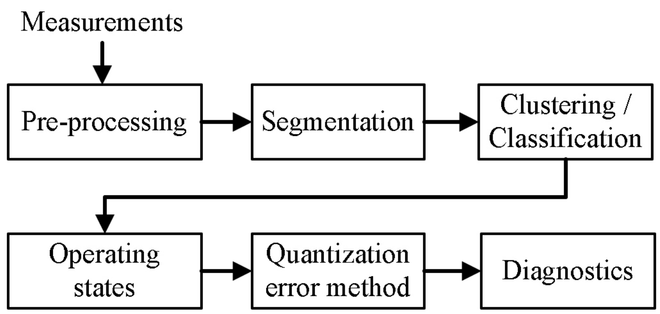

Figure 3. Time series analysis and state recognition.

Preprocessing of Time Series Data

The data have to be scaled before segmentation so that all variables have an equal effect on the segmentation result. The representative statistics of the data are needed in order to be able to scale the measurements. To do that, all the variables of the data should have enough variation. The data range and distribution used to compute the statistics should correspond to those of the whole data set.

Segmentation of Time Series Data

In segmentation, the time series data are transformed into piecewise linear representation. A segment is a contiguous subset of a time series. In this study, the time series data are segmented using the sliding window method with piecewise linear regression [2]. In this method, curves are approximated with lines. It is also possible to use other feature extraction techniques which can describe the segmented sections more precisely, for example, discrete wavelet transform [6,8].

New samples are added to the segment until cumulative squared estimation error exceeds a predefined threshold. Several measurements can be segmented together. Thus, they have common cost function and the cost is computed by summing all the measurements. The measurements together define the edges of segments.

Segmentation algorithms work either in a batch mode (offline) or online. The latter are needed for applications with harsh timing constraints (e.g. monitoring), but they are also suited for applications where a huge amount of data must be processed [1]. The sliding window segmentation method is well suited to online segmentation, because only the preceding samples are needed to define the edges. There are several modifications of the sliding window segmentation algorithm and there are other more efficient segmentation methods, but unfortunately most of them are not very suitable for online analysis. This is because the methods expect that the whole sequence will be available at the time of segmentation.

Clustering/Classification of Segment Features

After segmentation, pre-processed segments are clustered and classified. Clustering is the process of organizing objects into groups whose members are similar in some way. A cluster is therefore a collection of objects which are similar between them and dissimilar to objects belonging to other clusters. Clustering can be performed using hierarchical methods or k-means algorithm [4].

The parameters of piecewise linear regression lines are used as features. One choice is to use the slopes of lines as features. Also, the offset and the length of the segment can be used as additional features. The number of segment clusters has to be defined, or a validation index should be used to select a suitable amount of clusters [5]. It is not necessary to cluster all the available segments, as some of them can be classified using cluster prototype vectors defined using a representative set of segments. Each cluster is presented by a d-dimensional prototype vector (also known as: weight, codebook, model, reference) $\boldsymbol{m}=[m_1,\ldots,m_d]$, where d is equal to the dimension of the input vectors. It is useful to save these prototype vectors in order to classify another set of segments or to analyze classes or classification results later.

Quantization Error Method

The recognized operating states can be analyzed more thoroughly using the quantization error method [6,9]. It is based on distance calculation between the input sample vector (segmented feature vector) and the prototype vectors of operating states. Using this method it is possible to pinpoint when and where anomalies appear because each operating state corresponds to a certain operation of the machine.

MEASURED DATA FOR ANALYSIS

The test cases were defined and test drives were carried out to obtain measurement data from different operating states, which were then analyzed. The controlling functions were performed manually the first time and all the control signals were recorded. After this, the recorded control signals were used instead of manual driving. In this way it was possible to repeat the driven test drive and compare the results of the state recognition method and magnitude of the quantization error in the case of simulated faults (internal leakages) in the main hydraulic components of the hydrostatic transmission and working hydraulics, namely the HST pump and control valve of the boom cylinder [10,11]. An additional load of 300 kg was used in the test drives. Figure 4 shows an example of a drive path on a test field with the simulator.

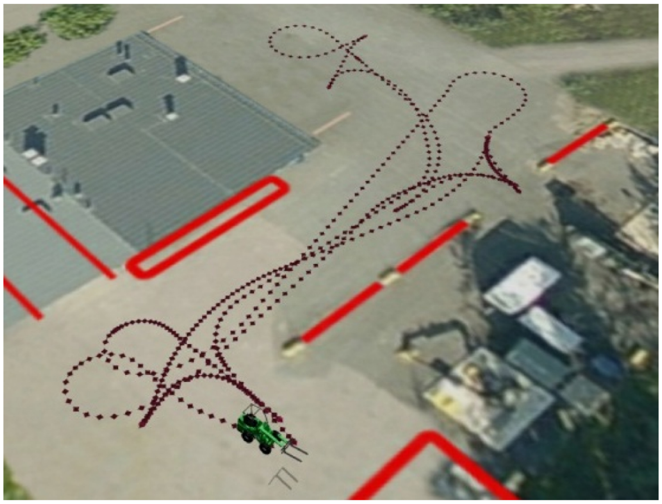

Figure 4. Example of drive path on test field with simulator.

Table 1 shows the defined test cases. The described driving sequence is repeated five times in each test drive with and without a fault (internal leakage). So altogether 20 test drives were completed. Table 2 shows the measured variables during these test drives which were then used in the analysis. The sampling frequency was 100 Hz. The measured time series data were analyzed using two different data sets, namely the variables of hydrostatic transmission and working hydraulics (boom).

Table 1. Cases of test drives.

Case id	Description
1	Hydrostatic transmission: acceleration (driving straight) – deceleration – acceleration (reversing) – deceleration
2	Working hydraulics (boom): lift up – down

Table 2. Analyzed data sets and their variables which were measured during test drives. 1st data set is hydrostatic transmission and 2nd working hydraulics

Data set	Signal	Range	Unit
1st	Diesel rotational speed reference	0…3000	rpm
	Diesel rotational speed	0…3000	rpm
	HST pump angle reference	-100…100	%
	HST pump measured angle	-100…100	%
	Pressure at port A	0…400	bar
	Pressure at port B	0…400	bar
2nd	Control valve reference	-100…100	%
	Pressure at port A	0…250	bar
	Pressure at port B	0…250	bar
	Boom angle	0…30	°

RECOGNITION OF OPERATING STATES FROM TIME SERIES DATA

Measurements were scaled to zero mean and unit variance before the analysis. Next, the pre-processed measurements were segmented using sliding window segmentation with piecewise linear regression. Cumulative squared error function with specific thresholds was used to define the edges of the segments, see table 3. Data from all the test drives were segmented similarly using two data sets. An example of segmentation of one test drive in the case of working hydraulics related measurements (data set 2, pressure A) is shown in figure 5.

Table 3. Number of test drives used in clustering and classification, number of clusters and magnitude of thresholds. N is normal, F is fault. 1st data set is hydrostatic transmission and 2nd working hydraulics.

Data set	Clustering		Classification		Clusters	Threshold
	N	F	N	F		
1st	5	0	5	10	16	0.3
2nd	5	0	5	10	25	0.15

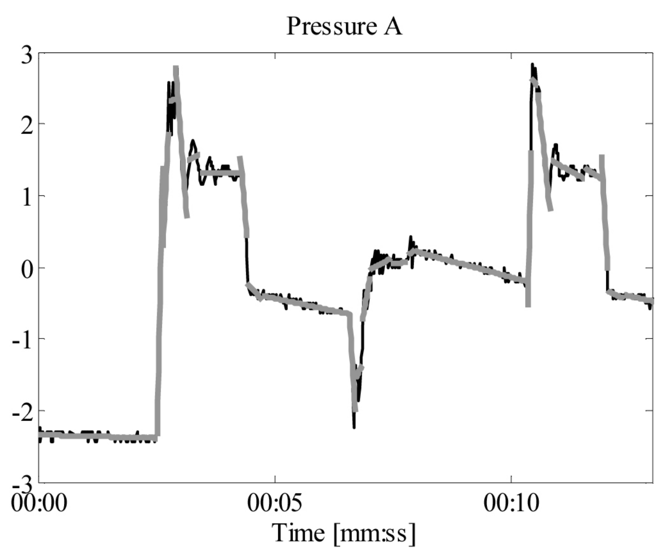

Figure 5. Example of time series segmentation of pressure A from data set 2. Measurements are normalized to zero mean and unit variance for segmentation.

The segments of ten representative test drives were clustered using k-means clustering with Davies-Bouldin validation index [5] and segment slopes as features. The grouping in the k-means algorithm is done by minimizing the sum of squares of distances between data and the corresponding cluster centers according to Eq. 1, where K is the number of groups and N is the number of sample/feature vectors, $\boldsymbol{x}_n$ is sample/feature vector and $\boldsymbol{m}_i$ is the mean of the data points in set i, i.e. prototype of cluster i, and it is calculated according to Eq. 2.

$$J = \sum_{i=1}^{K} \sum_{n=1}^{N_i} \| \boldsymbol{x}_n - \boldsymbol{m}_i \|^2 \tag{1}$$

$$\boldsymbol{m}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} \boldsymbol{x}_n \tag{2}$$

Clustering was repeated five times for every number of clusters and the result with minimum quantization error was selected. The number of clusters varied from 10 to 30 clusters and the one giving minimum validation index value was selected, see table 3. As a result of clustering, a prototype vector for each cluster was computed. These vectors were used to classify the segments of other test drives.

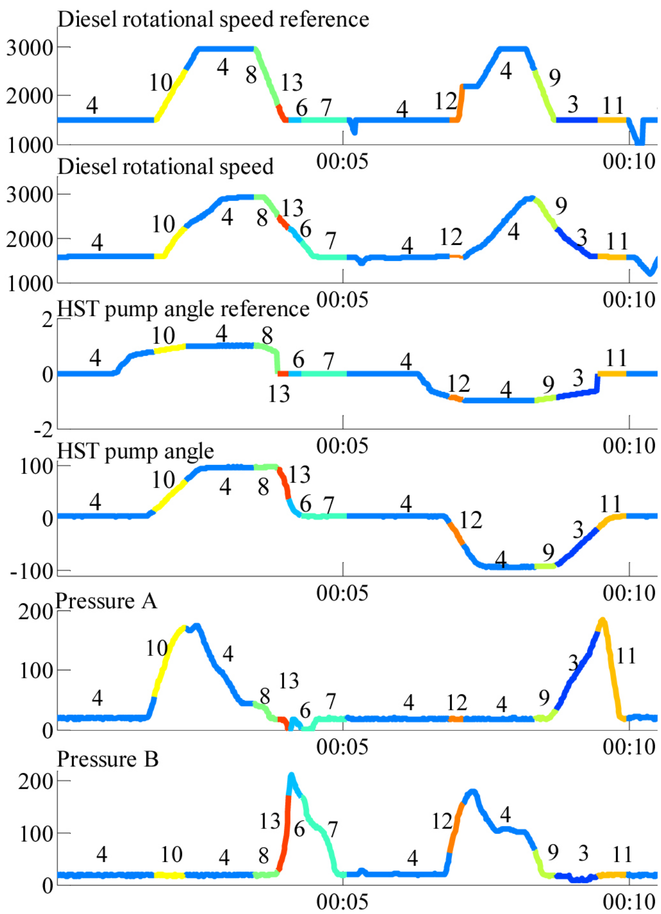

Figure 6. Example of time series of data set 1 with recognized operating states. The labels of operating states are shown with each measurement.

The segment class labels generate a sequence of operating states. Two sequences (two data sets) were built for each test drive. Operating state labels generated using working hydraulics related measurements are shifted by the number of segment classes for hydrostatic transmission measurements to obtain operating state labels which are unique over all the measurement sets. For each test drive two sequences corresponding to the measurement sets were combined. The new sequence has events with labels from 1 to 41. Figures 6 and 7 show examples of the analyzed time series of both the data sets from time of about 10 seconds. The data are from the test drives without a fault. The recognized states are plotted using different colors for each state. It can be clearly seen how changes in the measured variables define the edges of the segments.

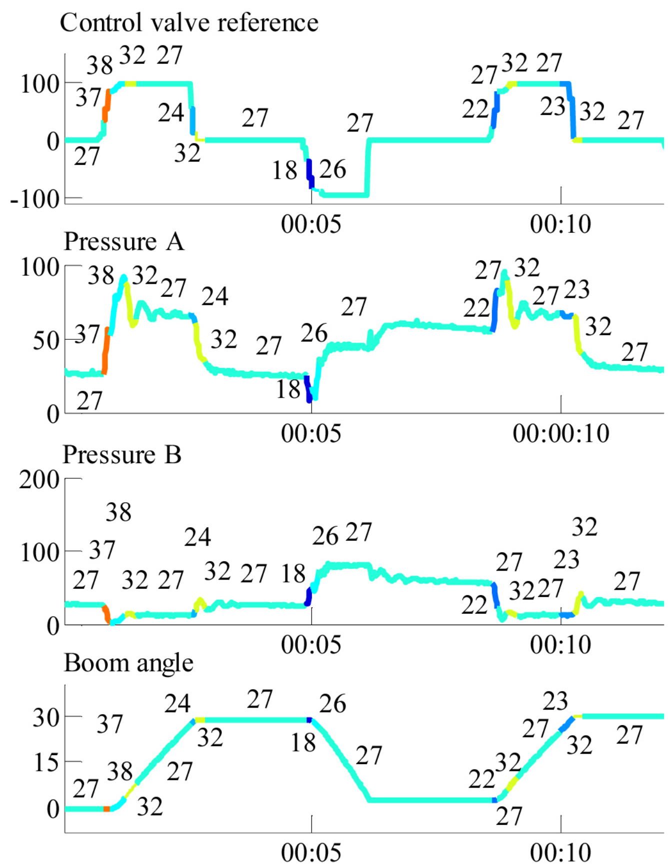

Figure 7. Example of time series of data set 2 with recognized states. The labels of operating states are shown with each measurement.

DIAGNOSTICS BASED ON RECOGNIZED OPERATING STATES

After the recognition of the operating states, these were further analyzed to detect anomalies. Simulated leakages in the main hydraulic components of the hydrostatic transmission and the working hydraulics were used as anomalies to study the changes in the recognized operating states and the magnitude of the quantization error.

Pressure dependent laminar flow was used to model the internal leakage created in the HST pump and the control

Table 4. Number of states and quantization error.

	State 12		State 37	
	No.	$\bar{e}_q$	No.	$\bar{e}_q$
N	25	0.019	5	0.032
F	50	0.048	11	0.053

valve of the boom cylinder of the working hydraulics, see Eq. 3 [10,11]. The magnitude of internal leakage was 3 l/min at 100 bar.

$$Q_i = K_l(p_A - p_B) \tag{3}$$

Quantization Error Method

Changes in the recognized operating states and the magnitude of the quantization error of the states were studied using the quantization error method. The quantization error method is based on the distance calculation, Euclidian distance, which is shown in Eq. 4. Here, the neuron whose weight vector is closest to the input sample vector $\boldsymbol{x}$ is called the BMU, denoted by c.

$$e_q = \|\boldsymbol{x} - \boldsymbol{m}_c\| = \min_i\{\|\boldsymbol{x} - \boldsymbol{m}_i\|\} \tag{4}$$

After this, a certain threshold can be set which determines the greatest distance at which recognition occurs. So when the method is tested, the distance between the sample vector and the cluster prototype vector is calculated. If the minimum distance is bigger than the threshold value set beforehand, then this sample vector is treated as an anomaly, which can be either a new operating or a fault state. After anomalies are detected and proved to be faulty operating states, data from these states can be used to detect and identify them in the future.

Changes in Recognized Operating States

One state from both the data sets was chosen here to study the changes more thoroughly in the recognized operating states using the quantization error method. The states are 12 and 37. State 12 is from the situation where the machine is reversed and state 37 where the boom is lifted up, see figures 6 and 7. The number of classified states depends on how many of them are recognized from the time series data. Both states are from the transient phases of the system. It is noted here that there are several states of this kind where anomalies clearly show. All of them are from situations when the system is in the transient stage. Figures 8 and 9 show the quantization errors of states 12 and 37. Anomalies caused by internal leakages can be clearly seen. There are some situations where normal and faulty states have almost equal quantization error. Still the mean of normal states without fault is much smaller than the mean of faulty states. Table 4 shows the number of recognized states 12 and 37, and the mean of the quantization error. The threshold can be set to 0.04 for both cases. Then most of the anomalies are detected and only a few normal states are mixed with faulty ones.

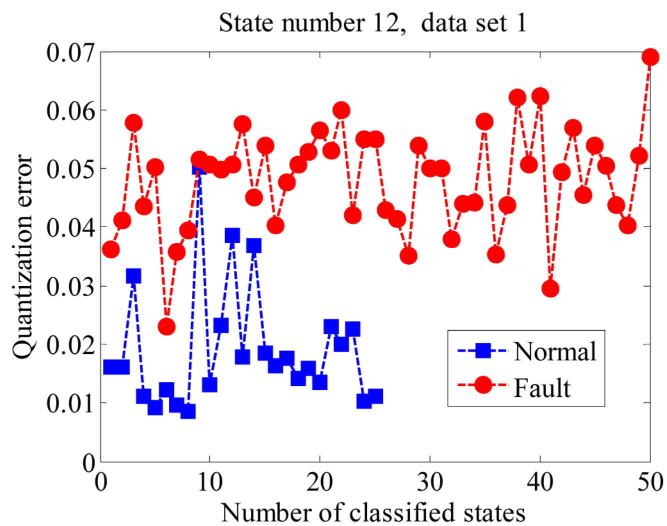

Figure 8. Quantization error of state number 12.

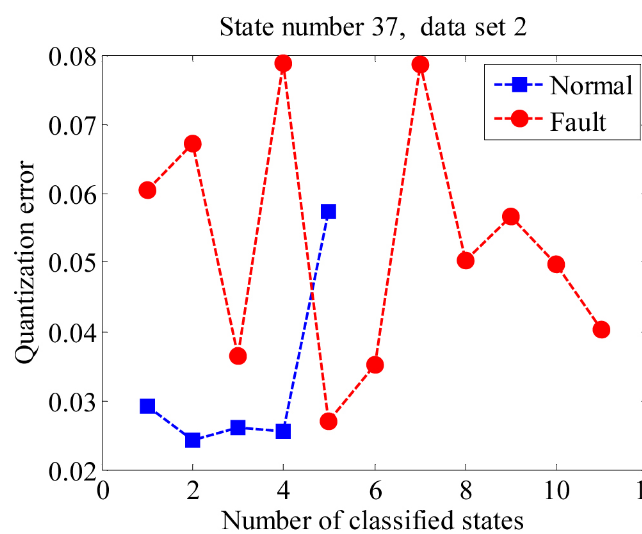

Figure 9. Quantization error of state number 37.

CONCLUSIONS

The operating states of a wheel loader were studied for diagnostics purposes using a real time simulation model of an articulated-frame-steered loader. Altogether, 20 test drives were carried out to obtain measurement data for analysis purposes. The measured data comprised two different data sets: hydrostatic transmission and working hydraulics.

41 different operating states were found from the measurement data using the sliding window method with piecewise linear regression for time series segmentation and the k-means algorithm for the clustering and classification of the pre-processed segments.

The recognized operating states were further analyzed using the quantization error method to detect anomalies. Simulated leakages in the main hydraulic components (HST pump and control valve) of the hydrostatic transmission and working hydraulics were used as anomalies. One operating state from both data sets, state 12 and 37, were selected as examples. State 12 is from the situation where the machine is reversed and state 37 where the boom is lifted up. Overall, the quantization error was higher in case of internal leakage. From the mean of the quantization error, the anomalies can be seen even more clearly.

The data-driven methods described in this study can be implemented in different kinds of machines. Insufficient sensor information may limit the number of applications, but the analysis methods in general are not restricted to a specific machine type. The thresholds in segmentation and anomaly detection need be defined based on specific applications and need specific knowledge about the operation of the system. The selection of critical measurement signals describing the operation of the machine also requires knowledge about the machine. The described operating state recognition method enables the detection of sudden critical faults as well as slowly evolving faults like internal leakages, which were studied here to demonstrate the functionality of this proposed method. The simultaneous examination of several variables and data sets enables a more generic method of detecting several different anomalies and applying it to different machine types.

In the future longer and more versatile test runs will be performed, different anomalies will be tested and the operating state recognition method will be implemented in a real machine.

REFERENCES

1. Fuchs, E., Gruber, T., Nitschke, J. and Sick, B., "Online Segmentation of Time Series Based on Polynomial Least-Squares Approximations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(12):2232-2245, 2010, doi:10.1109/TPAMI.2010.44.
2. Keogh, E., Chu, S., Hart, D. and Pazzani, M., "An Online Algorithm for Segmenting Time Series," In Proceedings of IEEE International Conference on Data Mining, 289-296, 2001, doi:10.1109/ICDM.2001.989531.
3. Ruotsalainen, M., Jylhä, J., Vihonen, J. and Visa, A., "A Novel Algorithm for Identifying Patterns from Multisensor Time Series," In Proceedings of the 2009 WRI World Congress on Computer Science and Information Engineering, CSIE 2009, Los Angeles, California, USA, 31 March - 2 April, 2009, 5:100-105, doi:10.1109/CSIE.2009.958.
4. Theoridis, S. and Koutroumbas, K., "Pattern Recognition," Academic Press, USA, 1999.
5. Davies, D. and Bouldin, D., "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2): 224-227, 1979, doi:10.1109/TPAMI.1979.4766909.
6. Krogerus, T., "Feature Extraction and Self-Organizing Maps in Condition Monitoring of Hydraulic Systems," PhD thesis, Tampere University of Technology, Tampere, Finland, Publication 949, 126, 2011.
7. Multanen, P. and Hyvönen, M., "Utilization of R&D simulators to support the maintenance of mobile work machines," In Proceedings of The Eighth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, CM 2011 and MFPT 2011, Cardiff, UK, June 20-22, 2011.
8. Krogerus, T., Rokala, M. and Koskinen, K.T, "Data Analysis Process of Working Hydraulics of Small Mobile Machine," *International Journal of Fluid Power*, 13(3):5-14, 2012.
9. Alhoniemi, E., Hollmén, J., Simula, O. and Vesanto, J., "Process Monitoring and Modeling Using the Self-Organizing Map," *Integrated Computer-Aided Engineering*, 6(1):3-14, 1999.
10. Ivantysyn, J., Ivantysynova, M., "Hydrostatic pumps and motors", Akademia Books Int., New Delhi, India, ISBN: 81-85522-16-2, 2001.
11. Merritt H.E., "Hydraulic Control Systems", J. Wiley & Sons Inc., New York, USA, ISBN: 0-471-59617-5, 1967.

CONTACT INFORMATION

Tomi Krogerus
Tampere University of Technology (TUT)
Department of Intelligent Hydraulics and Automation (IHA)
P.O. Box 589
FIN-33101 Tampere, Finland
tomi.krogerus@tut.fi

ACKNOWLEDGMENTS

The authors would like to acknowledge the Generic Intelligent Machine (GIM) project, which is a center of excellence of the Academy of Finland.

DEFINITIONS

c - Best-matching unit (BMU)

$\mathbf{e_q}$ - Quantization error

$\bar{e}_q$ - Mean of quantization error

I - Current [A]

J - Sum of squares clustering function

K - Number of groups/clusters

$\mathbf{K_l}$ - Flow gain [$m^4 \cdot s/kg$]

m - Prototype vector

$\boldsymbol{m}_\mathbf{c}$ - Prototype vector chosen as BMU

N - Number of sample vectors

$\boldsymbol{Q}_\mathbf{i}$ - Internal leakage [l/min]

p - Pressure [bar]

$\boldsymbol{p}_\mathbf{A}$ - Pressure A [bar]

$\boldsymbol{p}_\mathbf{B}$ - Pressure B [bar]

x - Data/feature vector