INDEPENDENT COMPONENT ANALYSIS FOR DETECTION OF CONDITION CHANGES IN LARGE DIESEL ENGINES

Niels Henrik Pontoppidan[1], Jan Larsen[1] and Torben Fog[2]

[1] Informatics and Mathematical Modelling, Technical University of Denmark,
Richard Petersens Plads 321, 2800 Kgs. Lyngby, Denmark
{nhp,jl}@imm.dtu.dk, Tel +45 4525 3351, Fax +45 4587 2599
[2] MAN B&W Diesel A/S
Teglholmsgade 41, 2450 København S, Denmark
Torben.Fog@manbw.dk, Tel +45 3385 1759, Fax +45 2285 1030

ABSTRACT

Automatic detection and classification of operation conditions in large diesel engines is of significant importance. This paper investigates an independent component analysis (ICA) framework for unsupervised detection of changes in and possibly classification of operation conditions such as lubrication changes and increased wear based on acoustical emission (AE) sensor signals.
The probabilistic formulation of ICA enables a statistical detection of novel events which do not conform to the current ICA model, thus indicating significant changes in operation conditions. Novelty of an observation is measured through the likelihood that the model has produced that observation. Evaluation of likelihood ratios allows the framework to also handle multiple models, thus enabling classification of operation conditions; furthermore the likelihood also serves as a link to traditional change detection.
The framework is evaluated on measured AE signals in an experiment where the operational condition varies. In particular, we compare the performance of mean field ICA, information-maximization ICA, and Principal Component Analysis. For detection of changes the performance is also compared to standard methods, e.g. mean value step detection.

INTRODUCTION

Identification of engine conditions and faults is important for automatic monitoring of critical failures in large marine diesel engines and stationary power plants. Early detection of small defects prior to evolving into serious breakdowns often reduces the costs for repair significantly.
The literature suggests that monitoring based on acoustical emission (AE) offer advantages over sensor techniques such as pressure and vibration [1, 2]. The signal-to-noise ratio is typically better for AE sensor signals, and further a system based on AE is more feasible from an operational point of view. Previous work on adaptive signal processing and machine learning [3, 4, 5, 6, 7, 8, 9] has mainly been focusing on supervised learning from sensor data and known faults. This paper focuses on unsupervised learning for significant detection of changes in measured AE signals, that is, modelling the probability density of the AE signal. Since we measure many samples of the AE signal we suggest a model, which also offers a compact data representation, such as the Independent Component Analysis (ICA) and Principal

Component Analysis (PCA) models. The probability density associated with the trained ICA and PCA models [10, 11, 12, 13] can be used to identify events which do not conform to the model assumptions [14, 15] and thus represent a significant change in engine condition.
In section 2 the data acquisition and experimental setup is described. Section 3 presents the modelling framework and a novel change detection algorithm based on ICA or PCA models. A comparative analysis and discussion using the suggested method is provided in section 4, and finally, section 5 state the conclusions.

Throughout vectors and matrices are identified by lowercase bold and uppercase bold letters respectively, i.e. the vector **x** and matrix **X**.

EXPERIMENTAL SETUP

The data set consist of two acoustic emission (AE) energy (RMS) signals $y_1(t)$, $y_2(t)$ acquired at 20 kHz with two very sensitive Physical Acoustics Corporation sensors placed on the cylinder liner and cover, respectively. The signals are resampled into the crank angle domain to provide 2048 samples per engine revolution. Further the two signals are stacked into the d=4096 dimensional feature (row) vector **x**.

$$\mathbf{x} = \left[y_1(1), \quad \ldots, \quad y_1(2048), \quad y_2(1), \quad \ldots, \quad y_2(2048) \right]^{\mathrm{T}} \tag{1}$$

In addition, 21 other channels, including top and crank-pulses were acquired from the cylinder at MAN B&W Diesel's Research Engine[1] in Copenhagen.
For each experiment, we consider three data sets:

- A Training set containing stationary examples under the current engine operation condition H_0.
- Test set 1 containing examples under the same condition as in the training set, H_0, which is used for model validation.
- Test set 2 containing examples that are investigated for changed in engine condition, H_1.

Thus we are able to check against false rejection of H_0 and to some extent also false accept of H_1.
During the experiment, the engine load was changed from 25% to 75%. In the middle of the 25% load period the cylinder lubrication was turned off, and in the middle of the 75% load period this system was turned on again.

Experiment 1: Shutting Off Lubrication

Initially the engine is stabilized at 25% load. After a while the lubrication to the cylinder is turned off. The objective is to detect this change in operation condition shortly after it occurred.

Experiment 2: Unstable Revolution Speed

The engine is running at 50% load and the lubrication system is turned off. Inspection of the revolution speed obtained from timing signal indicated that the engine undergo some sudden changes in the middle of this period. This is possibly caused by engine load fluctuations. We aim to detect the start and end of this period.

Experiment 3: Increased wear and re-establishment of lubrication

[1]Test bed, 4 cylinders, 500 mm. bore, 10.000 BHP.

The engine is running at 75 % load without lubrication. After 30 minutes lubrication is re-established, lowering the wear rate. We aim to detect this change of AE activity.

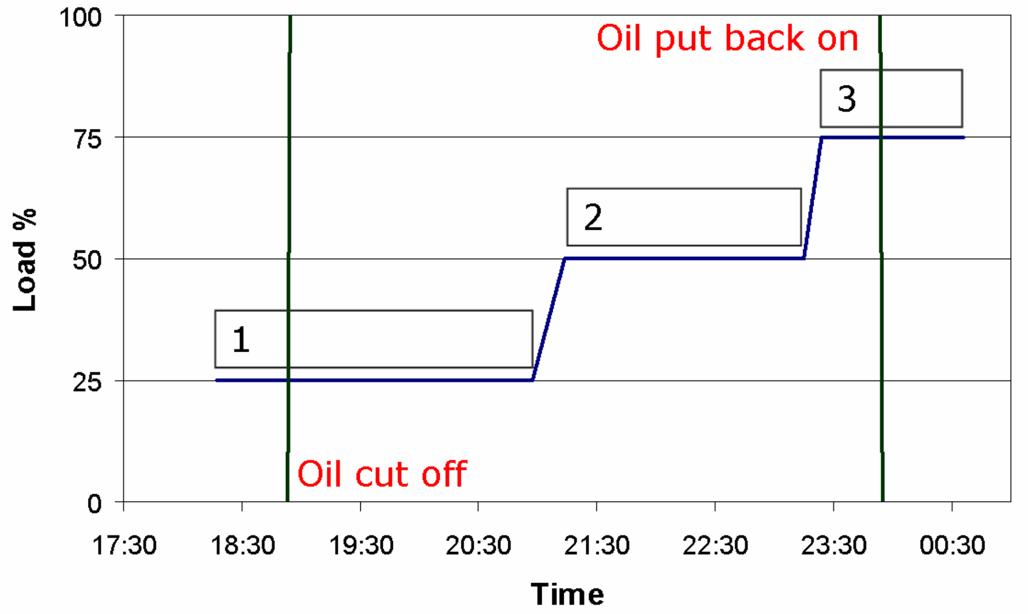

Figure 1, Time line of experiment. The stair like curve shows the increasing load as function of time. The numbered boxes refer to the three experiments described in the previous sections. The two vertical lines indicate when the lubricating system was turned off and on.

MODELING FRAMEWORK

Novelty detection

A general treatment of change detection is presented in e.g., [16, 17] here we deploy the novelty detection method proposed in [14, 15] which makes it possible to evaluate whether new examples conform to the model trained on the training set $\mathcal{T}$. The novelty detection is based on input density $p(\mathbf{x}|\mathcal{T})$ of the trained model. Consider the cumulative distribution of density values over the training set for all thresholds t. By selecting a low threshold $Q_{\min}$ identifying the corresponding $t_{\min} = \arg\min_t Q(t) \geq Q_{\min}$, novel events are detected as those where Q(t) is less than $t_{\min}$, see further figure 2.

$$Q(t) = P(\mathbf{x} \in R), R = \{\mathbf{x} : p(\mathbf{x}|\mathcal{T}) < t\} \tag{2}$$

So Q(t) is the probability that the example **x** is under the same condition H_0 as examples in the training set. The presented method assumes that examples, **x**, in the training set, i.e. drawn from the normal condition model H_0, share underlying hidden sources, and that we are able to identify those (or linear combinations) correctly. As usual, we are faced with the problem of over fitting, where too many sources allow the model to adapt to the noise in the training examples, and too few sources prohibits the model in learning the different variations. With test set 1 we are able to detect over fitting, as it contains examples that should be accepted as H_0, like the training set.
For an example **x** in the training set with mixing matrix **A** and corresponding source vector **s** the log-likelihood[2] is given by

$$\log p(\mathbf{x}|\mathcal{T}) \log p(\mathbf{x}|\mathbf{A},\mathbf{\Sigma}_\varepsilon,\Sigma_\mathbf{s}) = \log \int p(\mathbf{x}|\mathbf{A},\mathbf{\Sigma}_\varepsilon,\Sigma_\mathbf{s},\mathbf{s})p(\mathbf{s})d\mathbf{s} \tag{3}$$

---
[2] The probability density of **x** given the estimated model parameters.

Where is the covariance of the residuals from the training set, and s is the covariance of the sources estimated from the training set.
Define $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ as the set of N examples[3] and the number of used sources/components, K.

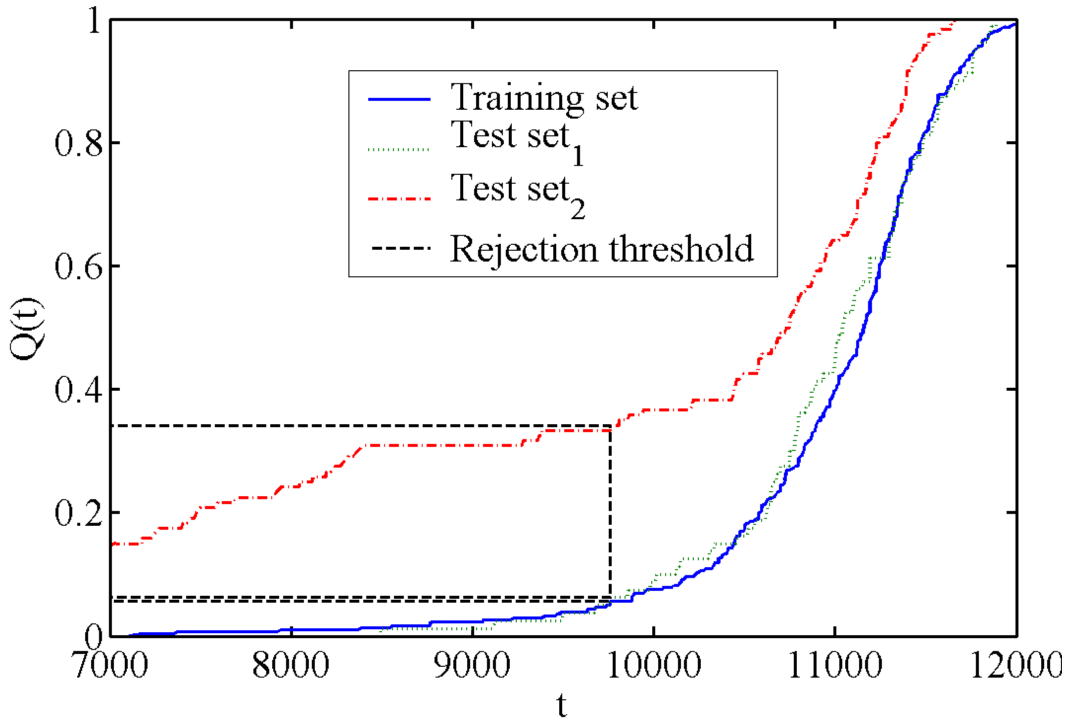

Figure 2, Cumulated log-likelihood densities, Q(t), from experiment 2 using ICA with 2 components. The solid (and smooth) line shows the cumulated density for the training examples. The dotted line show the cumulated density for test set 1 and is close to the training set. The dash-dotted line show the cumulated density for test set 2 and is above the training set curve, showing that many of these examples are rejected. The vertical (dashed) line show the threshold t_{min} together with the corresponding (horizontal) lines at the different rejection levels.

3.2. PCA

$$\mathbf{X} = \mathbf{UDV} \tag{4}$$

Where **X** is d×N, **U** is d×d, **D** is d×N, and **V** is N×N. We identify the mixing matrix as the K first columns of **U**, $\mathbf{A} = \mathbf{U}_K$, and the source matrix $\mathbf{S}_K$ as the first K principal components $\mathbf{D}_K\mathbf{V}_K$. Given a new example **x** we get the corresponding source $\mathbf{s} = \mathbf{Ax}$, and the residual $\boldsymbol{\varepsilon} = \mathbf{x} - \mathbf{As}$. We assume that the residual is Gaussian with diagonal covariance, and that the source distribution p(**s**) can be approximated by a Gaussian with zero mean and known diagonal covariance given by $\mathbf{D}_K$. Under these assumptions **Eqn. (3)** is analytically tractable [10, 11] and is given by

$$\log p(\mathbf{x} \mid \mathbf{A}, \boldsymbol{\Sigma}_\varepsilon, \boldsymbol{\Sigma}_\mathbf{s}) = \frac{1}{2}(\log|\boldsymbol{\Sigma}| - \log|\boldsymbol{\Sigma}_\varepsilon| - \log|\boldsymbol{\Sigma}_\mathbf{s}|) - \frac{1}{2}\mathbf{x}^\mathrm{T}\left(\boldsymbol{\Sigma}_\varepsilon^{-1} + \boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}_\varepsilon^{-1}\right)\mathbf{x} - \frac{1}{2}\log 2\pi \tag{5}$$

Where

$$\boldsymbol{\Sigma} = \left(\mathbf{A}^\mathrm{T}\boldsymbol{\Sigma}_\varepsilon^{-1}\mathbf{A} + \boldsymbol{\Sigma}_\mathbf{s}^{-1}\right)^{-1} \tag{6}$$

Information-maximization ICA (IM ICA)

[3] Each example corresponds to one revolution of the engine, for which two AE signature waveforms of 2048 samples are acquired.

Using PCA as a pre-processing dimensionality reduction step onto K dimensions, we can the apply Infomax ICA with square mixing matrix [18, 12].

$$\mathbf{X} = \mathbf{U}_K \mathbf{A}\mathbf{S} \tag{7}$$

Where **X** is d×N, **U** is d×K, **A** is K×K and **S** is K×N. We use the ICA-ML DTU:toolbox [19] for training

Mean Field ICA with Positive Constraints on Source and Mixing Matrices (MF ICA)

An advanced Bayesian ICA using mean field training [13] enables the possibility of avoiding PCA as a pre-processing step as well as imposing priors on the sources and mixing matrix.
The AE signals is the observable result of an additive process combining energy from various sources in the cylinder. If we want to model this, both source- and mixer matrix must only contain non-negative elements. We estimate a positive source matrix **S**, where the elements of each column are exponential distributed, and a mixer matrix **A** having non-negative elements using the ICA-ADATAP DTU:toolbox [19].

$$\mathbf{X} = \mathbf{A}\mathbf{S} \tag{8}$$

X is d×N, **A** is d×K and **S** is K×N. Given a new example and the trained model the code provides estimates of sources and the associated log-likelihood. The returned log-likelihood (as well as the sources) is a mean field approximation to **Eqn. (3)** obtained by minimizing a Kullback-Leibler divergence [13].

RESULTS

We have selected the number of components that reject the expected number of examples from test set 2 while still accepting examples from test set 1. The following tables show these results and the obtained performance. Figures 3-6 show Q(t) for individual examples in different experiments using the algorithms. Looking at these figures, the condition changes are easily spotted.
Generally mean field ICA and PCA works best, which is due to the fact that their log-likelihood also depends on the noise. These algorithms are both able to detect that the sources are changing and/or the yielding those examples is evaluated using both the sources and the residuals.

	PCA	IM ICA	MF ICA
Test set 1	10 %	13 %	10 %
Test set 2	93 %	80 %	89 %
No. of components	3	27	3

Table 1, Experiment 1: detecting oil off. The expected rejection rate of test set 2 is 93%.

	PCA	IM ICA	MF ICA
Test set 1	6 %	4 %	5 %
Test set 2	35 %	31 %	33 %
No. of components	5	12	2

Table 2, Experiment 2: Detecting temporary external condition change. The expected rejection rate of test set 2 is 34 %.

	PCA	IM ICA	MF ICA
Test set 1	13 %	5 %	10 %
Test set 2	98 %	94 %	98 %
No. of components	97	12	1

Table 3, Experiment 3: Detecting oil on. The expected rejection rate of test set 2 is 100%

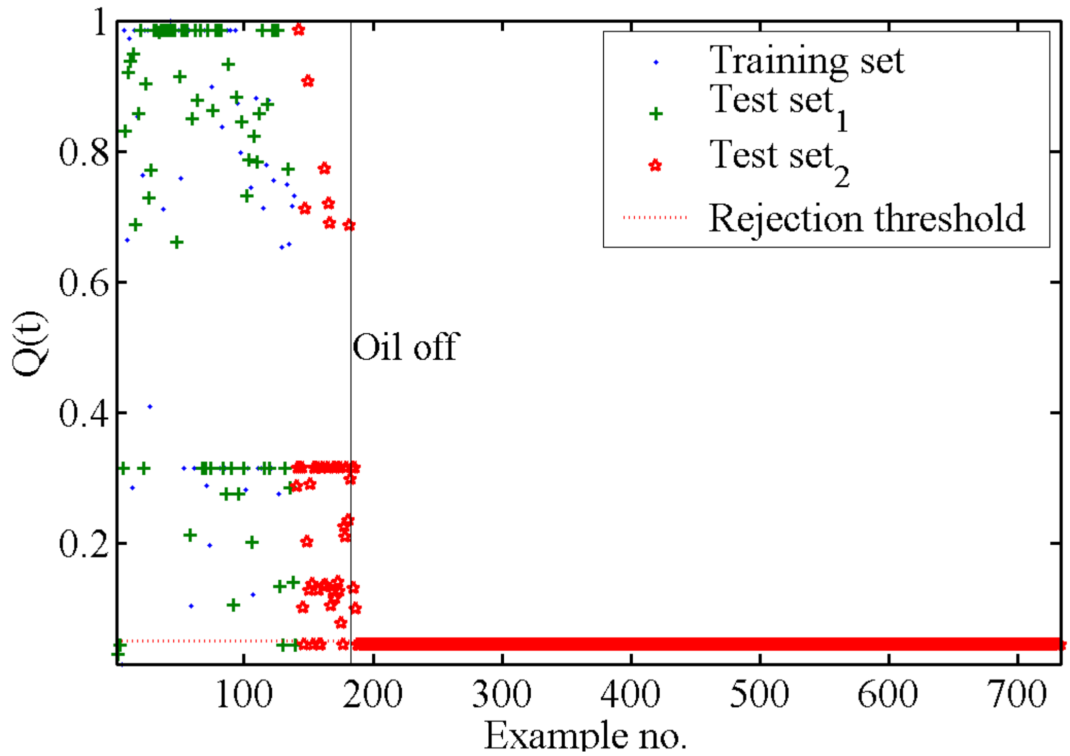

Figure 3, Q(t) for each example using PCA in experiment 1. The probability of coming from the normal condition clearly drops after the oil was cut off.

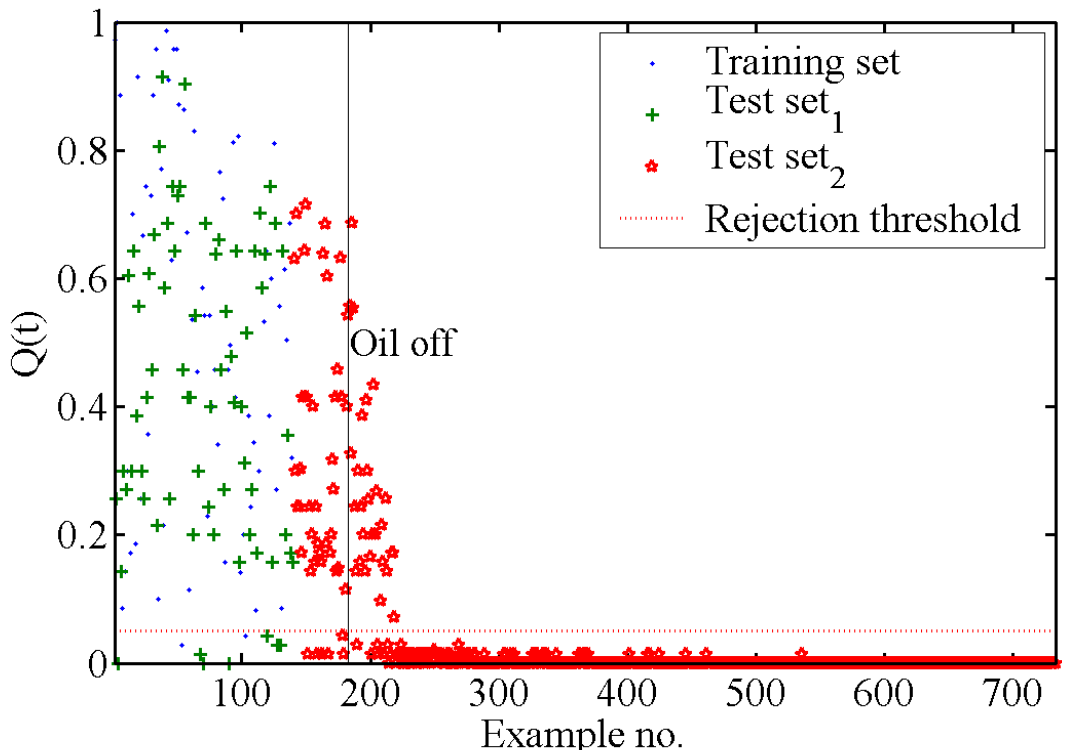

Figure 4, Q(t) for each example using mean field ICA in experiment 1. The probability of coming from the normal condition clearly drops after the oil was cut off.

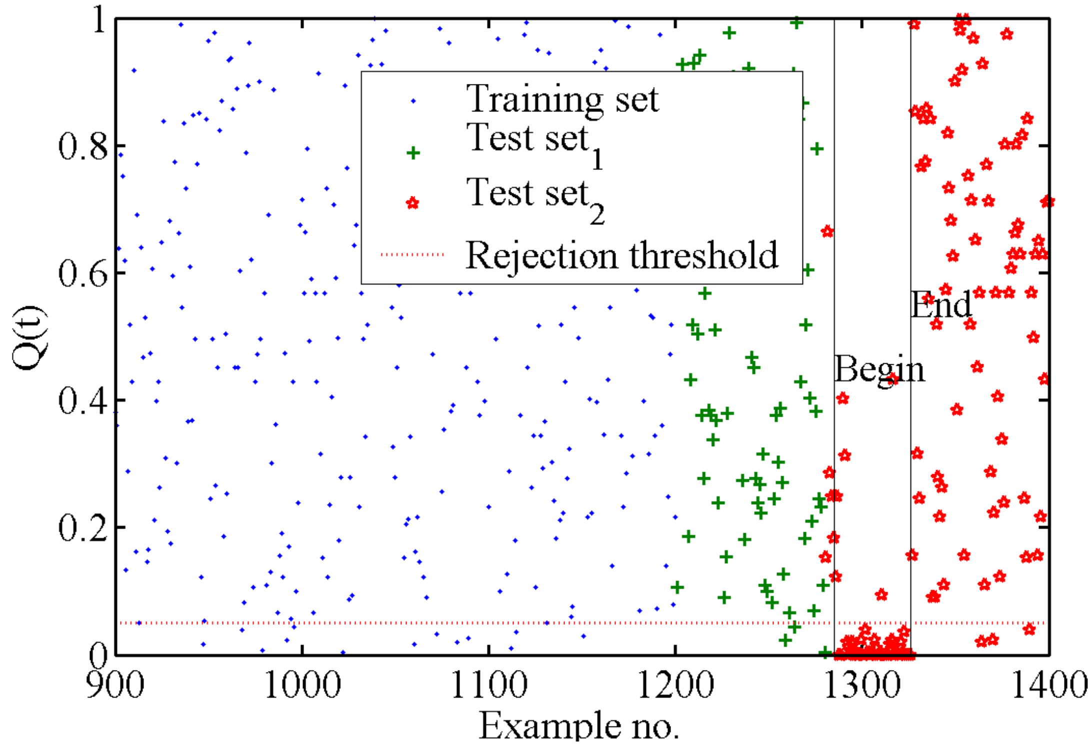

Figure 5, Q(t) for each example using Infomax ICA in experiment 2. The probability of coming from the normal condition clearly drops and returns thus indicating that the engine return to the previous condition.

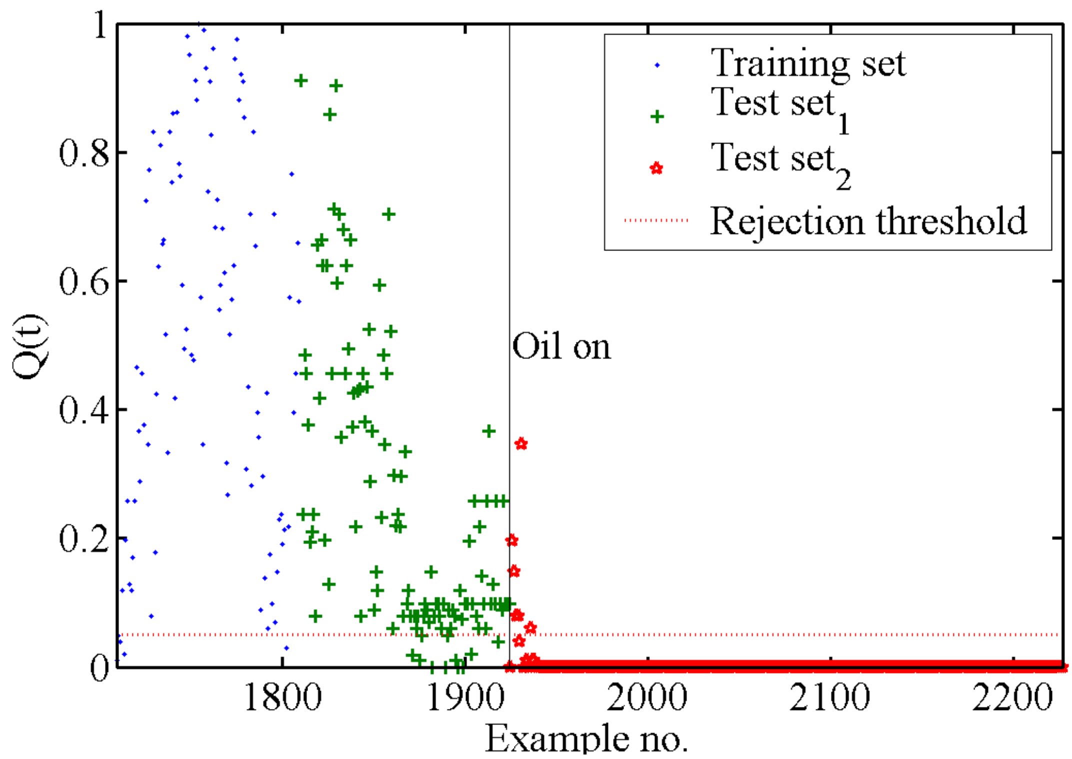

Figure 6, Q(t) for each example using mean field ICA in experiment 3. The probability of coming from the normal condition clearly drops after the oil was put back.

Simpler schemes

With ICA and PCA we are able to detect the changes. Experiments with the CUSUM algorithm [16] and Bayesian step detection [20] using means of the revolutions[4], show that simpler schemes also detect some of these condition changes. In settings where the amount of data is too large, these simples schemes can be used to pre-select time-windows for further analysis.

CONCLUSION

We have demonstrated the ability to detect changes in the operating parameters, including some parameters that where not monitored, for instance an external parameter. Furthermore we detect transitory condition changes, where the engine quickly returns to the previous condition.
In future we'll exploit this fact and extend the method aiming for classification, based on likelihood ratios. This should fix the apparent problems with only detecting changes namely, indication of causes as well as verification of fixes – when the condition returns after repair. Given vast amounts of data, segmented by simpler schemes our extended method should also be able to classify and group the segments.

Acknowledgements

This work is supported by EU Competitive and Sustainable Growth Programme GRD2-2001-50014 – acronym AE-WATT (http://isp.imm.dtu.dk/aewatt).

REFERENCES

1 R. L. Reuben, "The role of acoustic emission in industrial condition monitoring," International Journal of COMA-DEM, vol. 1, no. 4, pp. 35–46, Oct. 1998, ISSN 1363-7681.
2 G. D. Neill, S. Benzie, J. D. Gill, P. M. Sandford, E. R. Brown, J. A. Steel, and R. L. Reuben, The relative merits of acoustic emission and acceleration monitoring for detection of bearing faults, COMADEM, 1998.
3 G. Chandroth, A. J. C. Sharkey, and N. E. Sharkey, "Cylinder pressures and vibration in internal combustion engine condition monitoring," in Comadem 99, July 1999.
4 G. Chandroth, A. J. C. Sharkey, and N. E. Sharkey, "Vibration signatures, wavelets and principal components analysis in diesel engine diagnostics," in Marine Technology ODRA 99, October 1999.
5 T. L. Fog, L. K. Hansen, J. Larsen, H. S. Hansen, L. B. Madsen, P. Sørensen, E. R. Hansen, and P. S. Pedersen, "On condition monitoring of exhaust valves in marine diesel engines," in Proceedings of the IEEE Workshop on Neural Networks for Signal Processing IX, Y. H. Hu, J. Larsen, E. Wilson, and S. Douglas, Eds., Piscataway, New Jersey, 1999, IEEE, pp. 225–234.
6 Torben Fog, Condition Monitoring And Fault Diagnosis in Marine Diesel Engines, Ph.D. thesis, Technical University of Denmark, 1998, IMM-PHD-1998-52.
7 Mos Kaveh, Ahmed H. Tewfik, Kevin M. Buckley, et al., Integrated Predictive Diagnostic Tools, chapter 4.b: Signal processing for the detection and classification of acoustic emissions in integrated diagnostics, MultiUniversity Center for Integrated Diagnostics, 2000.
8 L.M. Rogers, "Use of acoustic emission methods for crack growth detection in offshore and other structures," Trans IMarE, vol. 110, no. part 3, pp. 171–180, 1998.
9 Alexander Ypma, Learning methods for machine vibration analysis and health monitoring, Ph.D. thesis, Dept. of Applied Physics, Delft University of Technology, Nov. 2001.

---

[4] $\frac{1}{2048}\sum_{n=1}^{2048} y_1(n)$

10 L.K. Hansen and J. Larsen, “Unsupervised learning and generalization,” in Proceedings of the IEEE Int. Conf. on Neural Networks 1996, 1996, vol. 1, pp. 25–30.
11 Thomas P. Minka, “Automatic choice of dimensionality for pca,” in Advances in Neural Information Processing Systems 13 , T.K. Leen, T.G. Dietterich, and V. Tresp, Eds. 2001, pp. 598–604, MIT Press.
12 Te-Won Lee, Independent Component Analysis: Theory and Applications, Kluwer Academic Publishers, September 1998, ISBN: 0 7923 8261 7.
13 Pedro Højen-Sørensen, Ole Winther, and Lars Kai Hansen, “Mean field approaches to independent component analysis,” Neural Computation, vol. 14, pp. 889–918, 2002.
14 Jan Larsen, Lars Kai Hansen, Anna Szymkowiak, Torben Christiansen, and Thomas Kolenda, “Webmining: Learning from the World Wide Web,” Computational Statistics and Data Analysis, vol. 38, pp. 517–532, 2002.
15 C.M. Bishop, “Novelty detection and neural network validation,” IEE Proceedings - Vision Image and Signal Processing, vol. 141, no. 4, pp. 217–222, 1994.
16 M. Basseville and I.V. Nikiforov, Detection of Abrupt Changes: Theory and Application, Prentice-Hall, 1993.
17 Fredrik Gustafsson, Adaptive Filtering and Change Detection, John Wiley and Sons LTD, 2000, ISBN: 0 471 49287 6.
18 Bell and T.J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” Neural Computation, vol. 7, pp. 1129–1159, 1995.
19 T. Kolenda, S. Sigurdsson, O. Winther, L.K. Hansen, and J. Larsen, “DTU:toolbox,” Internet, 2002, http://isp. imm.dtu.dk/toolbox/.
20 J. J. K. O. Ruanaidh and W.J. Fitzgerald, “Numerical Bayesian Methods Applied to Signal Processing”, Springer Verlag, 1996.