Anomaly Detection Using Dynamical Linear Models and Sequential Testing on a Marine Engine System

Erik Vanem[1, 2], Geir Olve Storvik[2]

[1] *DNV GL, Group Technology and Research, Høvik, Norway*
Erik.Vanem@dnvgl.com

[2] *University of Oslo, Department of Mathematics, Oslo, Norway*
erikvan@math.uio.no
geirs@math.uio.no

ABSTRACT

This paper presents a study on the use of Dynamical Linear Models for anomaly detection and condition monitoring of a marine engine system. Various sensors are installed at different places within the engine system and records essential parameters such as power output from the engine, engine speed, bearing temperatures and various other temperatures, speeds and pressures for selected engine components. The idea is to utilize the information in these sensor signals in order to monitor the condition of the engine. Such a condition monitoring system should include means of fault detection, diagnosis and prognostics, where robust anomaly detection is a prerequisite for reliable management of the system. Dynamical Linear Models (DLM) constitute a flexible framework for modelling of sensor signals, where the sensor signals are modelled conditional on some latent states, and the model provides forecasts of the signals that can be compared to new sensor readings. Statistical sequential model testing will then be performed on the forecast errors and model breakdown can be an indication of deviation from normal conditions and possible impending failures of the engine system. This will then call for further diagnostics and prognostics tasks to interpret the nature of the deviations. The Dynamical Linear Model framework can accommodate a range of candidate models. However, very complicated models in high dimensions may be computationally expensive to estimate and apply, so various pre-processing techniques are investigated in this paper to improve model performance, including simple regression models, cluster analysis and principal component transformation.

1. INTRODUCTION

Shipping accidents can have major consequences in terms of lives lost, environmental pollution and economic losses, and the safety and reliability of ships and ship systems is a main concern within the maritime industries. There exist international regulations for controlling the risk of maritime transport. For example, ships operating in international waters are subject to a number of IMO regulations such as SOLAS (Safety of Life At Sea (IMO, 2014b)) and the ISM Code (International Safety Management Code (IMO, 2014a)), and there are additional requirements from flag and port states. Moreover, ships are required to undergo regular surveys and inspections in a classification regime to ensure that the technical systems onboard the ships are maintained and operated in a safe and reliable manner (DNV GL, 2017a, 2017b).

Notwithstanding a strong and continued focus on safety and reliability of shipping, maritime accidents continue to occur and recent studies have indicated that the frequency of occurrence of major maritime accidents may even have increased in the last decade (Eleftheria, Papanikolaou, & Voulgarellis, 2016). Even though many of the accidents are ascribed to human errors (Psarros, 2015), the reliability of the ship and its various technical subsystems greatly influence the overall safety and risk of shipping. Hence, it is of paramount importance for the maritime industries to implement good maintenance strategies and to continuously monitor the condition of the ship and different ship systems and components.

Ship machinery systems provide safety critical functions to the ship related to propulsion and manoeuvrability as well as electrical power production for other ship systems. Thus, any failure in the ship machinery system may potentially be very critical to the safety of the ship. Loss of propulsion and manoeuvrability may lead to drift grounding and collision accidents with potential loss of stability and sinking of the vessel

as ultimate consequences (Vanem, Rusås, Skjong, & Olufsen, 2007). Loss of electric power production may lead to blackout and loss of other safety-critical functions from other ship systems (Mindykowski & Tarasiuk, 2015) and failures in ship machinery system may cause engine room fires which is one of the most frequent category of fire incidents for merchant ships (Vanem & Skjong, 2004). The condition and performance of the machinery system is also important for the overall fuel efficiency of the vessel and may have a large impact on the economics and environmental footprint of ship operations (Vanem, Brandsæter, & Gramstad, 2016). Hence, the condition of the ship machinery system is critical to the safety and economics of the vessel and needs to be carefully monitored.

The shipping industry is currently experiencing a shift towards digital solutions for monitoring of equipment and systems, and more and more sensors are being installed on components and systems onboard ships (Niculita, Nwora, & Skaf, 2017; Zymaris, Alnes, Knutsen, & Kakalis, 2016; DNV GL, 2015). These provide real-time information about the performance and condition of the monitored systems and may be utilized for fault detection, diagnosis and prognosis. A critical first step in such a setup is to automatically detect deviations from normal operational conditions, which may indicate failure of the system or even an impending failure that has not yet occured. It is assumed that data-driven, statistical methods can be exploited in order to detect anomalies in the, often high-dimensional, sensor data streams as early as possible.

This paper presents an application of one such statistical approach for anomaly detection, based on Dynamical Linear Models (DLMs) for describing and predicting the sensor time series. One of the advantages of DLM compared to several other methods is that it implicitly takes dynamical behaviour and dependence in time into account. Multivariate sensor data for a main generator engine onboard a ship in actual operations will be used for anomaly detection and condition monitoring. These data contains simultaneous measurements of 23 parameters related to the main generator engine, which is one out of four main generator engines onboard the ship. It is noted that these data does not contain any know failures or faults of the system.

A previous application of DLM for condition monitoring of ship machinery systems is presented in (Vanem & Storvik, 2016), but then on a smaller, lower-dimensional dataset for an overall ship system. In this paper, a more comprehensive dataset has been used with more detailed sensor measurements of specific components of a marine engine subsystem. The autoassociative kernel regression method (AAKR) (Hines & Garvey, 2006; Baraldi, Di Maio, Pappaglione, Zio, & Seraoui, 2012) has recently been used with a similar aim, as presented in (Brandsæter, Manno, Vanem, & Glad, 2016).

2. METHODOLOGY

2.1. Dynamical Linear Models

A thorough introduction to the theory and application of dynamical linear models is given in (West & Harrison, 1997), and only a brief introduction will be given below. Letting $\boldsymbol{Y}_t$ denote the r-dimensional vector of observed signals at time t, and $\boldsymbol{\theta_t}$ be the corresponding m-dimensional state-vector of an unobserved, latent process, the DLM model equations are on the following form:

$$\begin{aligned} \boldsymbol{Y}_t &= \boldsymbol{F}_t^{'}\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, & \boldsymbol{\nu}_t &\sim N(\boldsymbol{0}, \boldsymbol{V_t}) \\ \boldsymbol{\theta}_t &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t &\sim N(\boldsymbol{0}, \boldsymbol{W_t}). \end{aligned} \tag{1}$$

The first line of eq. (1) are often referred to as the observation equation and the second line is called the evolution, state or system equation. Such a model is fully specified by the quadruple $\{\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t\}$, as follows

- $\boldsymbol{F}_t$ is an $(m \times r)$ matrix, sometimes referred to as the regression matrix
- $\boldsymbol{G}_t$ is an $(m \times m)$ matrix, sometimes referred to as the state evolution matrix
- $\boldsymbol{V}_t$ is an $(r \times r)$ variance matrix for the observational variances
- $\boldsymbol{W}_t$ is an $(m \times m)$ variance matrix for the evolution variances

The error sequences $\boldsymbol{\nu_t}$ and $\boldsymbol{\omega_t}$ are assumed to be independent and mutually independent, and will be further assumed to be multivariate normally distributed with zero mean. The model further assumes a conditional independence structure so that, conditioned on the state-vector $\boldsymbol{\theta_t}$, the observations at time t, $\boldsymbol{Y_t}$ are independent of the history of observations $D_{t-1} = \{\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots, \boldsymbol{y}_1, D_0\} = \{\boldsymbol{y}_{t-1}, D_{t-2}\}$. Moreover, given the whole history until time t, all model information of the future is contained in $\boldsymbol{\theta}_t | D_t$. Such a conditional independence structure is illustrated in figure 1.

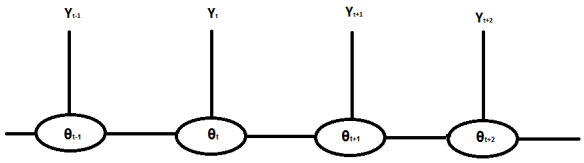

Figure 1. The conditional independence structure of the dynamical linear models

In order to fully specify such a model, one also needs an initial prior for the state vector at time $t = 0$, i.e. assuming a multivariate normal distribution with a mean vector $\boldsymbol{m}_0$ and an initial variance matrix $\boldsymbol{C}_0$:

$$(\boldsymbol{\theta}_0 | D_0) \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0). \tag{2}$$

2.1.1. Model Updating and Forecasting

Having established such a dynamical linear model, one may obtain, at each time t, posterior distributions for the state vectors and forecast distributions for future observations within a Bayesian setting, i.e. by applying Bayes' theorem, as follows.

a) Posterior at time $t-1$:
For some mean vector $\boldsymbol{m}_{t-1}$ and variance matrix $\boldsymbol{C}_{t-1}$,

$$(\boldsymbol{\theta_{t-1}}|\boldsymbol{D_{t-1}}) \sim N\left(\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}\right). \tag{3}$$

b) Prior at time t:

$$(\boldsymbol{\theta_t}|D_{t-1}) \sim N(\boldsymbol{a}_t, \boldsymbol{R}_t), \tag{4}$$

where $\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1}$ and $\boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t^{'} + \boldsymbol{W}_t$.

c) One-step ahead forecast for time t:

$$(\boldsymbol{Y}_t|D_{t-1}) \sim N(\boldsymbol{f}_t, \boldsymbol{Q}_t), \tag{5}$$

where $\boldsymbol{f}_t = \boldsymbol{F}_t^{'}\boldsymbol{a}_t$ and $\boldsymbol{Q}_t = \boldsymbol{F}_t^{'}\boldsymbol{R}_t\boldsymbol{F}_t + \boldsymbol{V}_t$.

d) Posterior at time t:

$$(\boldsymbol{\theta_t}|\boldsymbol{D_t}) \sim N\left(\boldsymbol{m}_t, \boldsymbol{C}_t\right) \tag{6}$$

with $\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t\boldsymbol{e}_t$ and $\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{Q}_t\boldsymbol{A}_t^{'}$, and where $\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{Q}_t^{-1}$ and $\boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t$

The vector $\boldsymbol{e}_t$ is the one-step ahead forecast errors and the matrix $\boldsymbol{A}_t$ is a matrix of adaptive coefficients. Looking further ahead from time t, for $k \geq 0$, the k-step ahead forecast distributions for $\boldsymbol{\theta}_{t+k}$ and $\boldsymbol{Y}_{t+k}$ given current information D_t are given by, for the state and forecast distributions, respectively:

$$(\boldsymbol{\theta}_{t+k}|D_t) \sim N\left(\boldsymbol{a}_t(k), \boldsymbol{R}_t(k)\right) \tag{7}$$

$$(\boldsymbol{Y}_{t+k}|D_t) \sim N\left(\boldsymbol{f}_t(k), \boldsymbol{Q}_t(k)\right) \tag{8}$$

with the following moments defined recursively, from starting values $\boldsymbol{a}_t(0) = \boldsymbol{m}_t$ and $\boldsymbol{R}_t(0) = \boldsymbol{C}_t$:

$$\boldsymbol{f}_t(k) = \boldsymbol{F}_t^{'}\boldsymbol{a}_t(k)$$

where

$$\boldsymbol{a}_t(k) = \boldsymbol{G}_{t+k}\boldsymbol{a}_t(k-1)$$

and

$$\boldsymbol{Q}_t(k) = \boldsymbol{F}_t^{'}\boldsymbol{R}_t(k)\boldsymbol{F}_t + \boldsymbol{V}_{t+k}$$

where

$$\boldsymbol{R}_t(k) = \boldsymbol{G}_{t+k}\boldsymbol{R}_t(k-1)\boldsymbol{G}_{t+k}^{'} + \boldsymbol{W}_{t+k}.$$

There are many versions of the dynamical linear model, and in general the model parameters are allowed to vary in time. In these investigations, however, a constant DLM is assumed, i.e. a model where the model parameters are constant in time $\{\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t\} = \{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{V}, \boldsymbol{W}\}$. Moreover, the parameters are estimated from a training data set and consequently assumed as known.

2.1.2. Model Intervention

The updating and forecast distributions outlined above assumes that the model does not receive any external information, i.e. that the information available at each time t is simply $D_t = \{\boldsymbol{Y}_t, D_{t-1}\}$. However, the modelling framework allows for intervention and incorporation of external information at any time t. This information could be based on other external observations, control of some of the parameters, knowledge about missing data or observations that should be ignored or could be a subjective intervention based on expert judgement. Such interventions can be included in the model at any time, and letting I_t denote the intervention at time t, it can be represented by an update of the available information prior to observing $\boldsymbol{Y}_t$, i.e. $D_{t-1} \rightarrow \{I_t, D_{t-1}\}$. For more details on how to include such effects in the models, reference is made to (West & Harrison, 1997).

2.1.3. Model Monitoring

One important feature of any statistical model used for condition monitoring is the ability to monitor the model and to detect, as early as possible, any departure from the model. With a DLM this can be done by Bayes' factors, assuming a null model M_0 and an alternative model M_1 for the signal time series. Both models provide a predictive distribution for $\boldsymbol{Y}_t$ conditioned on the available information D_{t-1}, i.e. for $i = 0, 1$

$$p_i(\boldsymbol{Y}_t|D_{t-1}) = p(\boldsymbol{Y}_t|D_{t-1}, M_i). \tag{9}$$

The Bayes' factor for M_0 versus M_1 based on the observed value of $\boldsymbol{Y}_t$ can then be defined as

$$H_t = \frac{p_0(\boldsymbol{Y}_t|D_{t-1})}{p_1(\boldsymbol{Y}_t|D_{t-1})}, \tag{10}$$

where p_i refers to the probability of observing $\boldsymbol{Y}_t$ conditioned on the model i, $i = 0, 1$ according to the predictive densities at time t; $p_i(\boldsymbol{Y}_t|D_{t-1}) = p(\boldsymbol{Y}_t|D_{t-1}, M_i)$. For k consecutive observations, $\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \ldots, \boldsymbol{Y}_{t-k+1}$, the cumulative Bayes' factor for M_0 versus M_1 can be defined as

$$H_t(k) = \prod_{r=t-k+1}^{t} H_r = \frac{p_0(\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \ldots, \boldsymbol{Y}_{t-k+1}|D_{t-k})}{p_1(\boldsymbol{Y}_t, \boldsymbol{Y}_{t-1}, \ldots, \boldsymbol{Y}_{t-k+1}|D_{t-k})}. \tag{11}$$

These Bayes' factors measure the predictive performances of model M_0 relative to model M_1 and evidence for or against the null model M_0 accumulates multiplicatively as data are observed and processed; for each $t > 1$, $H_t(k) = H_t H_{t-1}(k-1)$. On the log-scale the evidence is additive

and a value of 0 (1 on the original scale) indicate no preference for either model but a positive value of the Bayes' factor indicate evidence in favour of model M_0. In practice, one may monitor the forecast errors rather than the forecasts themselves. This use of Bayes' factors is similar to the sequential probability ratio test (SPRT) often applied in process control (Wald, 1945), and which will be applied in the study presented herein.

2.2. Sequential Testing

The sequential probability ratio test is based on the SPRT index and is similar to the Bayes' factor. The index is defined as the natural logarithm of the likelihood ratio for M_1 versus M_0 (the reciprocal of eq. 10). The SPRT is additive and is updated as subsequent observations arrive. To calculated these indices, there is a need to specify the alternative model M_1, in order to know the predictive density $p_1(\boldsymbol{Y}_t|D_{t-1})$. However, by monitoring the forecast errors this is greatly simplified.

According to the model assumptions, the one-step forecast errors, $\boldsymbol{e}_t = \boldsymbol{Y}_t - \boldsymbol{f}_t$. should be normally distributed with zero mean and variance matrix $\boldsymbol{Q}_t = \boldsymbol{F}_t^{'}\boldsymbol{R}_t\boldsymbol{F}_t + \boldsymbol{V}_t$ (see eqs. 3-6). Accordingly, the standardized one-step forecast errors, $\tilde{\boldsymbol{e}}_t = \boldsymbol{e}_t/\boldsymbol{Q}_t^{1/2}$, should be uncorrelated and distributed according to the standard normal distribution. Hence, rather than specifying an alternative model for $\boldsymbol{Y}_t$, one may specify an alternative model for the standardized forecast errors. For a sequence of errors, $e_1, \ldots, e_n$, the SPRT index can be calculated as

$$l_n = \ln L_n = \ln \prod_{t=1}^{n} \frac{p_1(e_t|D_{t-1})}{p_0(e_t|D_{t-1})}. \tag{12}$$

The test builds evidence for or against the null model and as soon as enough evidence is obtained to make a decision (accept or reject the alternative model), the SPRT index is reset to zero. Two threshold values, A and B need to be specified. If the SPRT index reaches the upper threshold an alarm is flagged for possible model breakdown. If the lower threshold is reached, the null model is accepted as adequate. In either case, the SPRT index is reset to zero and the monitoring continues until any of the thresholds are reached again. The limits can be related to the acceptable probabilities of false alarms (α) and missed alarms (β) as follows

$$A = \ln\left(\frac{\beta}{1-\alpha}\right) \quad \text{and} \quad B = \ln\left(\frac{1-\beta}{\alpha}\right). \tag{13}$$

2.2.1. SPRT for Mean Level Change

Under the null model, M_0, the standardized forecast errors, $\tilde{\boldsymbol{e}}_t$ are uncorrelated standard normally distributed variables and various alternative models can easily be specified, e.g. as a change in level (mean), in spread (variance) or in covariance between the errors. In such situations the predictive densities for the errors become $p_0(\tilde{\boldsymbol{e}}_t|D_{t-1}, M_0) \sim N(\boldsymbol{0}, \boldsymbol{I})$ and $p_1(\tilde{\boldsymbol{e}}_t|D_{t-1}, M_1) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for some values $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_{ij}$. Having established one or more alternative models, sequential probability ratio tests can be performed by monitoring the desired SPRT indices.

In this study, alternative hypotheses of a mean level change will be applied and the SPRT index for the standardized forecast error for each sensor signal, *i*, will be monitored. Two tests will be performed, i.e. for a positive or negative level change, $\pm\mu_i$ for each of the signals. The alternative levels μ_i must be specified and in this study these are assumed to be the same value for each signal, with $\mu = \pm 2$. Moreover, under both the null and the alternative hypotheses the errors are assumed uncorrelated across the signals. Hence, the SPRT indices for the two tests take the following form (assuming now unit variance for the standardized forecast errors)

$$\begin{aligned} l_{1,n} &= \mu \sum_{i=1}^{n} \left(\tilde{e}_i - \frac{\mu}{2}\right) \\ l_{2,n} &= \mu \sum_{i=1}^{n} \left(-\tilde{e}_i - \frac{\mu}{2}\right). \end{aligned} \tag{14}$$

2.3. Model selection

The Akaike Information Criterion (AIC) (Akaike, 1974) is a measure based on the likelihood that also penalizes excessive number of parameters. The AIC is on the form of eq. 15, where L is the likelihood of the model and k is the number of free parameters in the model. The preferred model according to this criterion is the one with the lowest AIC score. The AIC is calculated for all the candidate models for both the training data and the test data in this study. Alternatively, for the test data, the Root Mean Square Error (RMSE) is calculated based on the one-step ahead forecast errors from the models. The RMSE combines the RMSE for each sensor signal without any weighting, and is calculated as shown in eq. 16, where T is the number of time-points in the test data, r is the number of sensor signals and $\hat{y}_{i,t}$ and $y_{i,t}$ are the predicted and observed value of sensor signal i at time t, respectively.

$$AIC = 2k - 2\ln(L) \tag{15}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}\sum_{i=1}^{r}\left(\hat{y}_{i,t} - y_{i,t}\right)^2}{rT}} \tag{16}$$

It should be noted that these metrics are not entirely appropriate in this case, since the DLMs have not been fitted to the same data. Since different pre-processing of the raw data have

been performed prior to fitting the different models, comparing the models by terms of likelihood-based metrics or RMSE may not be fair. For example, the RMSE measures the square distance between the forecasted principal components of the residuals and not the forecast errors of the actual raw signals. Notwithstanding, the AIC and the RMSE for each model are computed, and they give an indication of how the models perform relatively to the others. However, model selection is perhaps best carried out by looking at the performance in terms of ratio of alarms.

3. SENSOR DATA FROM A MARINE ENGINE SYSTEM

The dataset that is explored in this study contains several sensor signals that can be related to the main bearing condition of one of four separate diesel engines on a ship. It is noted that the collected data do not contain any known faults or failures of the system, and the data are not compared to maintenance logs of the system. The list of signals are included in table 1, and engine 1 contains 24 sensor signals. By inspecting the data, it is observed that the signals for MG1TE702 contain only zero-values, presumably due to malfunctioning sensor or loss of connectivity and these signals are excluded from the subsequent analysis.

Table 1. Sensor signals in the dataset

MAIN GENERATOR ENGINE 1	
MG019	MGE1 ENGINE SPEED [rpm]
MG1PT201	MGE1 LO PRESS ENGINE INLET [bar]
MG1PT401	MGE1 HT WATER JACKET INLET PRESS [bar]
MG1PT601	MGE1 CHARGE AIR PRESS AT ENGINE INLET [bar]
MG1SE518	MG1 TC A SPEED [rpm]
MG1SE528	MG1 TC B SPEED [rpm]
MG1TE201	MGE1 LO TEMP ENGINE INLET [C]
MG1TE272	MGE1 LO TEMP TC OUTLET A [C]
MG1TE282	MGE1 LO TEMP TC OUTLET B [C]
MG1TE511	MGE1 EXHAUST GAS TEMP TC A INLET [C]
MG1TE517	MGE1 EXHAUST GAS TEMP TC A OUTLET [C]
MG1TE521	MGE1 EXHAUST GAS TEMP TC B INLET [C]
MG1TE527	MGE1 EXHAUST GAS TEMP TC B OUTLET [C]
MG1TE600	MGE1 AIR TEMP TC INLET [C]
MG1TE601	MGE1 CHARGE AIR TEMP AT ENGINE INLET [C]
MG1TE700	MAIN BEARING NO 0 TEMP MGE1 [C]
MG1TE701	MAIN BEARING NO 1 TEMP MGE1 [C]
MG1TE702	MAIN BEARING NO 2 TEMP MGE1 [C]
MG1TE703	MAIN BEARING NO 3 TEMP MGE1 [C]
MG1TE704	MAIN BEARING NO 4 TEMP MGE1 [C]
MG1TE705	MAIN BEARING NO 5 TEMP MGE1 [C]
MG1TE706	MAIN BEARING NO 6 TEMP MGE1 [C]
MG1TE707	MAIN BEARING NO 7 TEMP MGE1 [C]
PM100.07	MG1 POWER [kW]

The sensor signals covers a period of about 10 months starting from December 2014 with a sampling frequency of one minute. It is observed that many of the signals are highly correlated. For example, the various temperature measurements for the main bearings are all very strongly correlated, see Figure 2. A couple of interesting features that can be observed from this plot is a) that all the temperature sensors seems to fall out at a certain point (all drops to zero) and b) that the sensor for main bearing No 2 seems to stop working and reports the same value over a period of time. Traceplot of the engine speed is also shown in the figure.

The time series data are divided into a training dataset (75%) and a test dataset (25%), containing 249858 and 83286 samples, respectively. Correlation plots of the data are shown in Figure 3. These illustrate the the overall correlation structure remains relatively unchanged across the subsets of data, perhaps with the exception of one signal that goes from being weakly negatively correlated in the training data to being more strongly negatively correlated in the test data.

Density plots of the sensor signals reveal another interesting feature of the data. There seem to be two dominating modes of operation, corresponding to ”on” and ”off”. This is most evident in the density plot of the engine speed and is shown in Figure 4, where the two peaks of the density is clearly distinguished. Note that even in ”off” mode, the engine speed is never exactly zero, apart from a few times when the sensor signal is presumably lost. These modes are clearly reflected in most of the other signals as well, and Figure 4 includes the densities of the various pressure signals and for some of the temperature measurements. The exception is the air temperatures at the inlets that are not very dependent on the operational mode. The trace plots in Figure 2 demonstrates that the different levels of the main bearing temperatures coincide with the main levels of the engine speed.

Separating the data into two subsets depending on whether the engine speed is above or below 100 rpm, one can easily see that the characteristics are different for the two subsets of data for most of the signals. In Figure 5, the densities for such conditional data are shown for each signals; the black density curves correspond to operational mode ”on” (i.e. engine speed ≥ 100) and the red density curves correspond to operational mode ”off”. It is observed that there are distinct differences between most of the signals in different modes.

3.1. Data Preprocessing

Dynamical Linear Models will be applied to the raw sensor data, but will also be applied to different versions of preprocessed data. In particular, two alternative regression models will be applied in order to try to account for the main modes of the data. Then, DLM will be applied on these residuals. Moreover, principal component analysis is applied for dimension reduction.

3.1.1. Regression on Engine Speed

It is observed that the data essentially falls in two groups depending on the value of engine speed, see Figures 2, 4 and 5. This feature will be difficult for a statistical model to predict, and can be construed to coincide with decisions made on the bridge. Rather than trying to model these shifts in the data, for example by some conditional model or a Markov chain

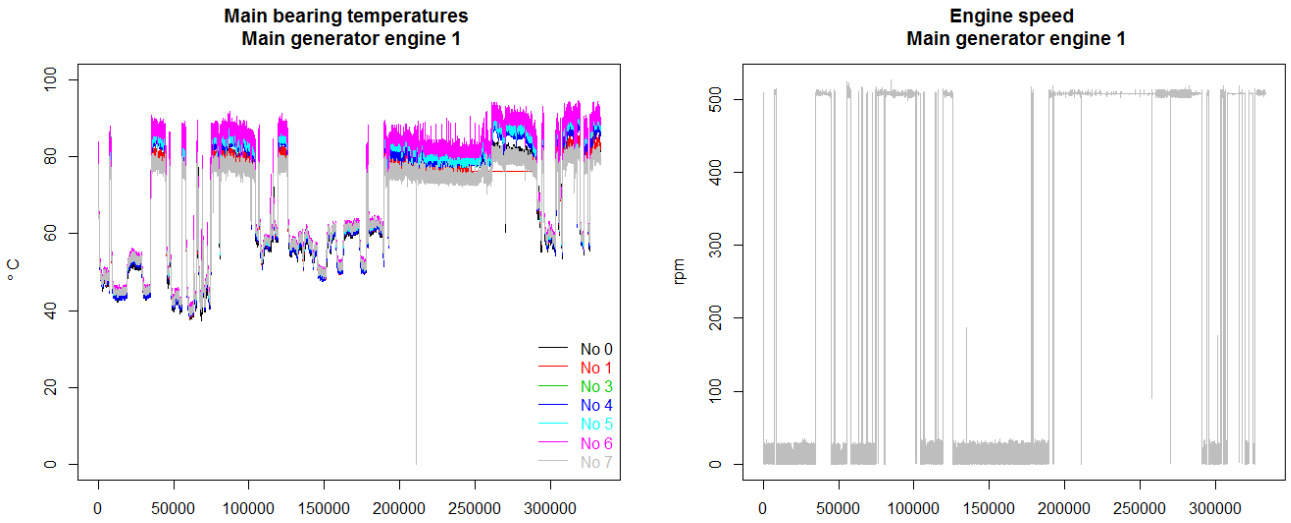

Figure 2. Example of traceplots illustrating strong correlation between some of the signals (engine 1); the various temperature readings for main bearing temperatures (left) and the engine speed (right)

model, the engine speed will be used as covariate in linear regression models for the other sensor signals. Then, dynamical linear models will be applied to the residuals of this regression model. Presumably, the residuals will not contain the same degree of jumpiness and can be better modelled with a dynamical linear model. Moreover, the strong correlation between engine speed and the other sensor signals will mean that all the other signals will also be strongly correlated, and it is believed that this correlation will be reduced when the effect of engine speed is removed. The residuals of such a regression model on the value of engine speed, for each of the main bearing temperatures, are shown in Figure 6. Compared to the trace plots in Figure 2, it is seen that the residuals vary around the approximate same level for both operational modes of the engine, and that the correlation between the signals remains high. However, there are some spikes corresponding to the transient conditions between the two main operating modes. A correlation plot of the residuals is also shown in Figure 6, and this can be compared to the correlation plots in Figure 3. Note that for engine speed (*MG019*), there are no residuals, so the actual values have been used for this sensor in the correlation plot.

Regression models with different lags of the engine speed were also tried out, in order to better captures transient behaviour, but adding the lags did not improve the model noteworthy. Another model, with categorical variables corresponding to e.g. engine speed larger or smaller than 100 rpm was also tried out, but did not change the results notably either. Hence, a simple linear regression model with an intercept and the value of engine speed was fitted to all the other sensor signals, and the further modelling was performed on the residuals.

3.1.2. Principal Components of Regression Residuals

In order to reduce dimensions one more pre-processing step will be performed before fitting dynamical linear models. That is, a principal component analysis is applied to extract the principal components of the residuals. For the data from main engine 1, it is found that 99% of the variance in the residual signals can be explained by the first 9 principal components, as shown in Figure 7. The horizontal green line indicates a cumulative proportion of variance of 0.99, which corresponds to the first 9 principal components (the first 9 PCs explain 99.246 of the total variance). Hence, the first 9 principal components of the residuals will be used to fit a dynamical linear model for the data. This corresponds to a significant dimension reduction without significant loss of information. Trace plots of the transformed data (residuals of regression model on engine speed) along the first 9 principal components are shown to the right in Figure 7. It can be seen that there are still significant spikes in the data, presumably associated with the transient states between main operational modes of the engine.

Another effect of the principal component decomposition is that the principal components will be uncorrelated, and hence the transformed data will have zero linear correlation. Even though zero correlation does not necessarily mean that the signals are independent, one may assume that the uncorrelated, transformed data can be modelled independently and that the effect of any remaining dependence can be ne-

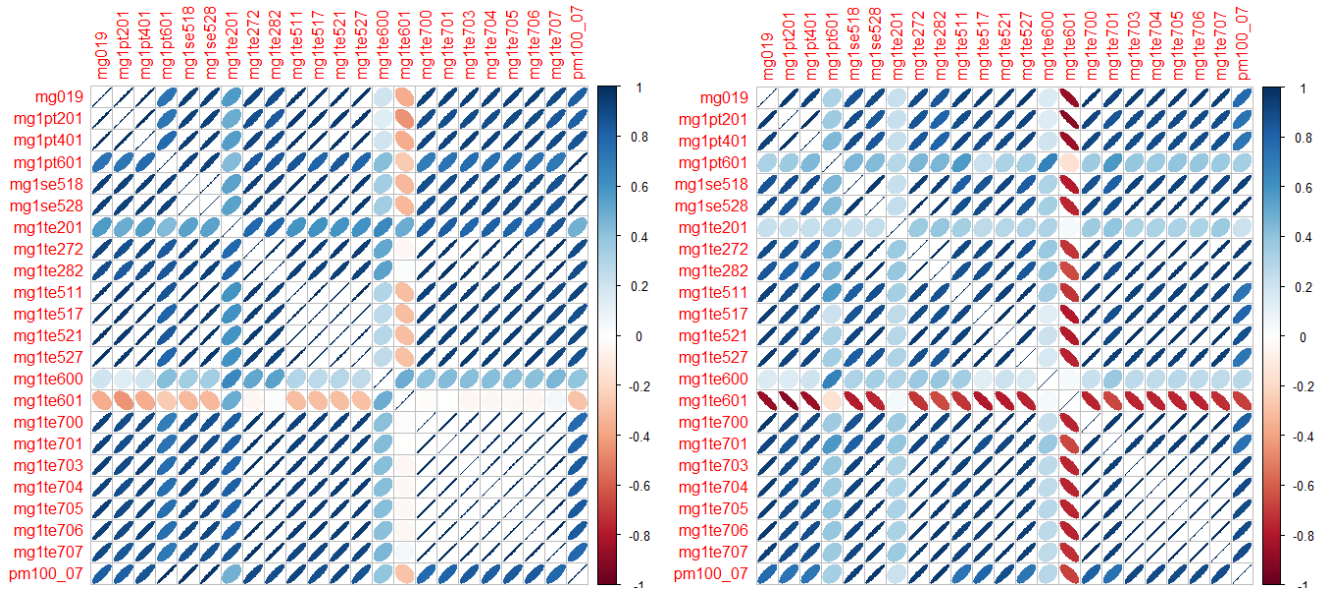

Figure 3. Correlation plots for the training data (left) and the test data (right) for main generator engine 1

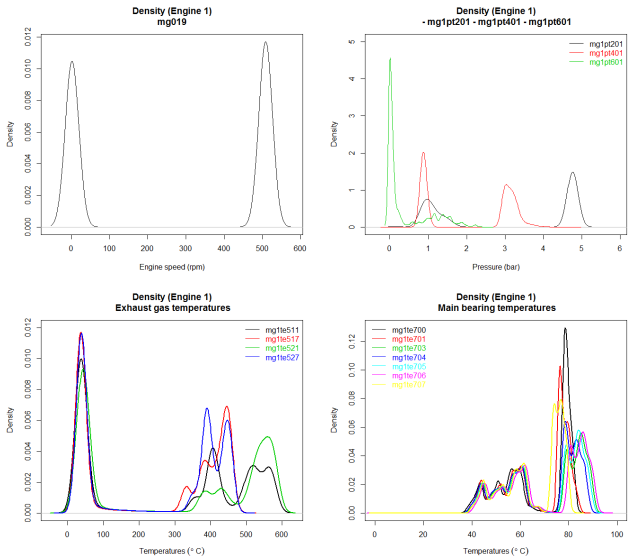

Figure 4. Density plots for some of the sensor signals for main generator engine 1; top: engine speed (left) and various pressures (right), bottom: Exhaust gas temperatures (left) and main bearing temperatures (right)

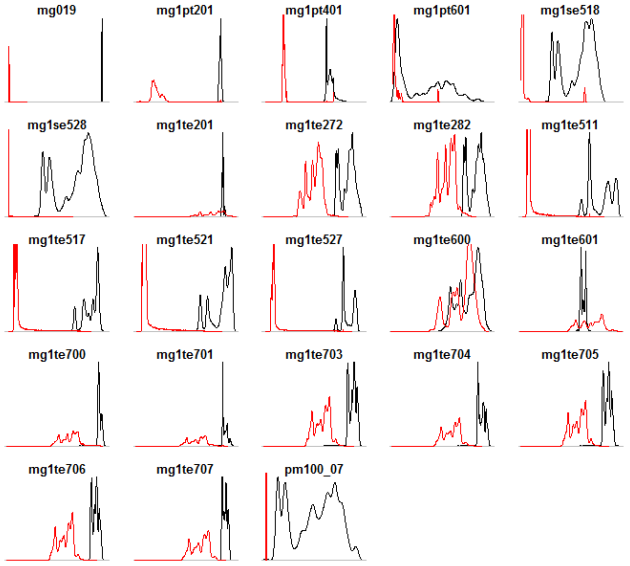

Figure 5. Density plots for the various signals from main generator engine 1 in "on" and "off" modes, respectively; Black curves correspond to "on" and red curves correspond to "off", where the mode is determined by the engine speed.

glected. Hence, the transformed sensor signals will be modelled independently, with a dynamical linear model for each transformed signal independently in both the evolution- and observation-level. This significantly reduces the complexity of the dynamical linear model.

3.1.3. Regression based on Cluster Analysis

Even though the data seems to correspond to two main operational modes of the engine, it may be investigated whether this is reasonable by carrying out a cluster analysis on the data. A simple cluster analysis using k-means clustering for different values of k suggests that there might in fact be 3 or 4 clusters in the data, but also indicates that much of the within-clusters total sum of squares are significantly reduced

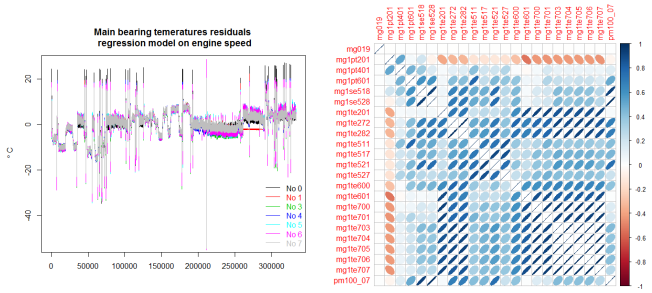

Figure 6. Traceplots of residuals of the main bearing temperatures after being regressed on engine speed (left) and correlation plot of the residuals (right)

by $k = 2$. A plot of the total within-cluster sum of squares as a function of k in the k-means cluster algorithm is shown in Figure 8.

Interestingly, for $k = 2$, not all data points with high engine speed are assigned to the same cluster, and for $k = 3$ and $k = 4$, there will be 2 and 3, respectively, clusters with very similar engine speed values, around 500 rpm. Possibly, this can be due to vector-autocorrelation in the data, i.e. due to inertia in the systems causing a delayed effect of the turning on or off of the engine on the other sensor signals. In order to account for this, an extended regression model on the assigned cluster will be fitted to the data, and anomaly detection will be performed on the residuals as above. Assuming 4 clusters, and letting the first cluster correspond to the reference level, the regression model becomes, for each signal independently,

$$y_i = \alpha + \beta_1 C_{2,i} + \beta_2 C_{3,i} + \beta_3 C_{4,i} + \varepsilon_i, \qquad (17)$$

where $C_{2,i}, C_{3,i}, C_{4,i}$ are dummy variables taking values 0 or 1, respectively, depending on whether observation i belongs to cluster 2, 3, or 4 or not. Observations assigned to cluster 1 will have all dummy variables equal to 0, which is the reference level. The correlation plot of the residuals from this regression model is shown in Figure 9 and trace plots of the residual engine speed and main bearing temperatures are shown in Figure 10.

Principal component analysis of the residuals are performed, and it is found that more than 99 % of the total variance in the data is explained by the first 11 principal components. Hence, DLM models will be fitted to the residuals projected onto the 11 first principal components, where these components are uncorrelated and modelled independently.

4. APPLICATION OF DLM FOR ANOMALY DETECTION ON A MARINE ENGINE SYSTEM

Various dynamical linear models are fitted to the raw data as well as the different datasets obtained by different preprocessing of the raw data. The results are outlined below.

4.1. DLM on the Raw Sensor Signals

Different dynamical linear models were applied to the initial sensor data for main engine 1. However, only the trivial model where the 23 sensor signals are modelled independently with a full-dimensional state vector is somewhat successful. Other models, where different state vectors selected based on e.g. the correlation structure in the data, by assuming 7 latent states representing each main bearing or assuming one latent state failed to succeed due to lack of convergence of the maximum likelihood optimiser. The number of model parameters in each of these model alternatives were between 30 and 50, and it is obviously a challenging optimization problem to find the maximum likelihood estimates of such high-dimensional likelihoods. The model that was successfully fitted to the data, with full-dimensional state vector, contained 46 parameters associated with the observational and system variances for each signal.

This model was fitted to the training data and subsequently applied to the test data. Some examples of the signals together with the one-step predictions as well as the filtered and predicted hidden states are shown in Figure 11.

4.2. DLM on Principal Components of Residuals from Regression on Engine Speed

The principal component decomposition of the residuals from regression models yields principal components that are uncorrelated, which simplifies the modelling with dynamical linear models. It means that the regression and state evolution matrices can be modelled as unit diagonals and the two variance matrices will be diagonal with 9 free parameters each. The estimated model parameters, as estimated from fitting such a model to the test data, are presented in Table 2, where $\sigma^2_{\nu,i}$ denotes the ith diagonal element in the estimated V-matrix and $\sigma^2_{\omega,i}$ are the diagonal elements of the W-matrix.

Having estimated the model on the training data, it may be applied to the test data in order to derive filtering and smoothing distributions as well as forecast distributions. The expected values of the 1-step ahead forecast distributions are shown in Figure 12 together with the actual signals. It can be seen that the model performs quite well in predicting the principal components one step ahead. However, some problems can be observed for e.g. principal component # 7.

The k-step ahead forecasts for the principal components of the residuals from the regression model on engine speed are shown in Figure 13 for k = 100, starting from the last observation in the test data. Also the last 250 observations of the test data are included in the plot. The forecasts are shown as the mean forecast and the 95% confidence interval for 100 future observations. It is observed that the confidence intervals grow wider further into the future, as expected, and that the confidence intervals seems to correctly capture at least the

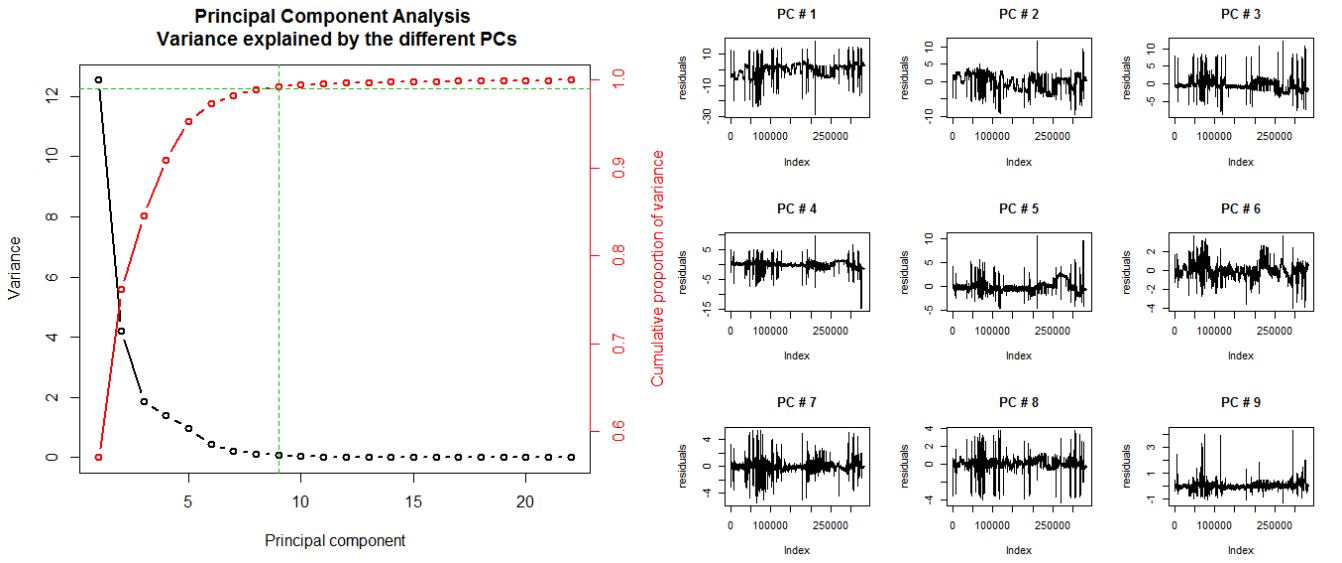

Figure 7. Results of a Principal Component Analysis on the residuals; Variance explained by each principal component and the cumulative proportion of variance by the first principal components (left). Transformed data along the first 9 principal components (right).

Table 2. Parameter estimates for the model of the principle components of residuals from a regression model on engine speed

$\sigma^2_{\nu,1}$	$\sigma^2_{\nu,2}$	$\sigma^2_{\nu,3}$	$\sigma^2_{\nu,4}$	$\sigma^2_{\nu,5}$	$\sigma^2_{\nu,6}$	$\sigma^2_{\nu,7}$	$\sigma^2_{\nu,8}$	$\sigma^2_{\nu,9}$
1.62×10^{-7}	0.00369	0.00823	0.0394	0.00513	0.00778	0.0438	0.00164	0.00189

$\sigma^2_{\omega,1}$	$\sigma^2_{\omega,2}$	$\sigma^2_{\omega,3}$	$\sigma^2_{\omega,4}$	$\sigma^2_{\omega,5}$	$\sigma^2_{\omega,6}$	$\sigma^2_{\omega,7}$	$\sigma^2_{\omega,8}$	$\sigma^2_{\omega,9}$
0.0787	0.0165	0.0136	0.00954	0.00327	0.00132	0.00287	0.00330	0.000415

range of values in the preceding 250 measurements. For principal component # 7, which the model had some difficulties in predicting, it is observed that there are a relatively wider confidence interval also for the 1-step ahead forecasts.

4.3. DLM on Principal Components of Residuals from Regression on Cluster Membership

Residuals from another regression model on cluster membership were also used for anomaly detection with DLMs. In this case, the 11 first principal component explain more than 99% of the variation in the data, so DLMs on these variables have been fitted. Again, the principal components are uncorrelated, so a simple, independent DLM is fitted to these signals, where both the regression and the state evolution matrices will be 11-dimensional unit diagonal matrices. The variance matrices V and W will be diagonal matrices with elements estimated from the test data, as presented in Table 3. Again, $\sigma^2_{\nu,i}$ denotes the ith diagonal element in the estimated V-matrix and $\sigma^2_{\omega,i}$ are the diagonal elements of the W-matrix.

The 1-step ahead point forecasts for the 11 first principal components of the residuals after regression on cluster membership are shown in Figure 14 together with the actual principal component measurements. The k-step ahead forecasts for $k = 250$ future measurements after the end of the test data are shown in Figure 15. It is observed that the 95 % confidence bands seems reasonable and seems to capture at least most of the variability in the preceding 250 measurements. However, for principal components # 9 and 11 the variability in the signals is relatively larger compared to the forecast variance than for the other signals. Figure 14 confirms that these signals seems to be most challenging for the model.

4.4. Model Monitoring and Predictive Performance

In order to use the models for anomaly detection and condition monitoring, it is assumed that the model will describe the signals well in normal operating conditions, and that model breakdown indicates deviation from normal behaviour of the system, which may need to be investigated with more scrutiny. According to the model assumptions, the standardized one-step ahead forecast errors should be uncorrelated

Table 3. Parameter estimates for the model fitted to the principle components of the residuals from a regression model on cluster membership with 4 clusters

$\sigma^2_{\nu,1}$	$\sigma^2_{\nu,2}$	$\sigma^2_{\nu,3}$	$\sigma^2_{\nu,4}$	$\sigma^2_{\nu,5}$	$\sigma^2_{\nu,6}$	$\sigma^2_{\nu,7}$	$\sigma^2_{\nu,8}$	$\sigma^2_{\nu,9}$	$\sigma^2_{\nu,10}$	$\sigma^2_{\nu,11}$
0.0308	0.0533	0.121	0.104	0.00630	0.110	0.0640	0.0416	0.0190	0.00660	0.00889

$\sigma^2_{\omega,1}$	$\sigma^2_{\omega,2}$	$\sigma^2_{\omega,3}$	$\sigma^2_{\omega,4}$	$\sigma^2_{\omega,5}$	$\sigma^2_{\omega,6}$	$\sigma^2_{\omega,7}$	$\sigma^2_{\omega,8}$	$\sigma^2_{\omega,9}$	$\sigma^2_{\omega,10}$	$\sigma^2_{\omega,11}$
0.0562	0.0147	0.0134	0.0169	0.0117	0.00360	0.00671	0.00597	0.000657	0.00208	0.000299

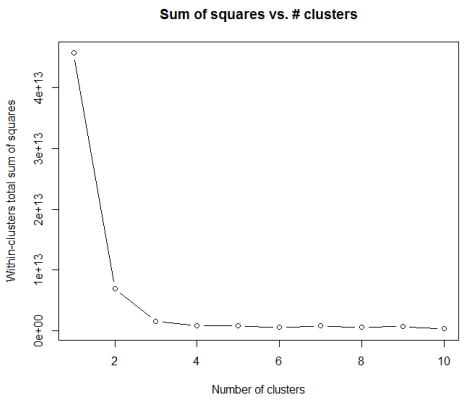

Figure 8. Total within-clusters sum of squares as a function of number of clusters

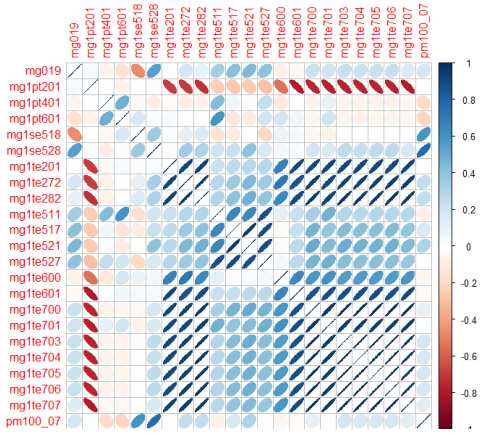

Figure 9. Correlation plot of the residuals after applying a linear regression model on the categorical variables corresponding to cluster assignment with 4 clusters

and standard normally distributed. Hence, one may look at the one-step ahead forecast errors for the test data and compare to this assumption in order to check if the model perform well and to get indications of model breakdown. Traceplots and densities of the one-step ahead forecast errors for the model on residuals after regression on engine speed are shown in Figure 16 (Similar plots for the other models are not shown, but display similar characteristics).

It is observed that most of the forecast errors are centered around zero and that the variances are lower than for the standard normal except for some large spikes. These coincide with large spikes in the data (principal components of the residuals from regression models) which again coincides with the transient states between ”on” and ”off” modes in the raw data (see discussion of the data above). These jumps in the data are difficult for the model to predict, but once the jumps have been made, the models seems to be able to tune in to the new behaviour quite fast.

In order to use the DLMs for condition monitoring and anomaly detection, sequential model testing will be applied, as outlined above. Tests for positive and negative mean changes of $\mu = \pm 2$ will be applied, with $\alpha = 0.005$ and $\beta = 0.001$, and an example of the results of sequential probability ratio tests (SPRT) with these settings on one of the models is presented in Figure 17. The tests are performed

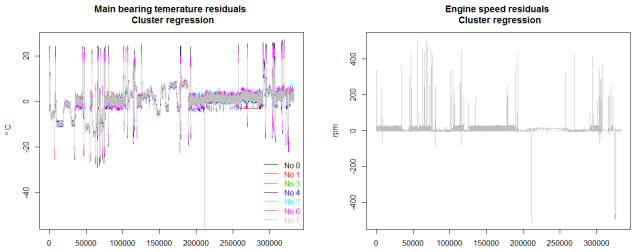

Figure 10. Traceplots of the residuals after applying a regression model on cluster groups; main bearing temperatures (left) and engine speed (right)

separately for each signal making it possible to see in which signal there is a deviation from normal behaviour. However, for the cases where the signals are principal components, it is not straightforward to interpret how an alarm in one signal relates to the physical system, and any alarms need to be investigated closely.

In the plots of the SPRT indices, alarms are marked in red for every time the SPRT crosses the upper level, B. It is observed that the SPRT raises quite many alarms on these data and many red flags are shown in the SPRT-plots. However, compared to the number of measurements, the alarm rate still seems reasonable. A close-up of an SPRT-plot is shown in

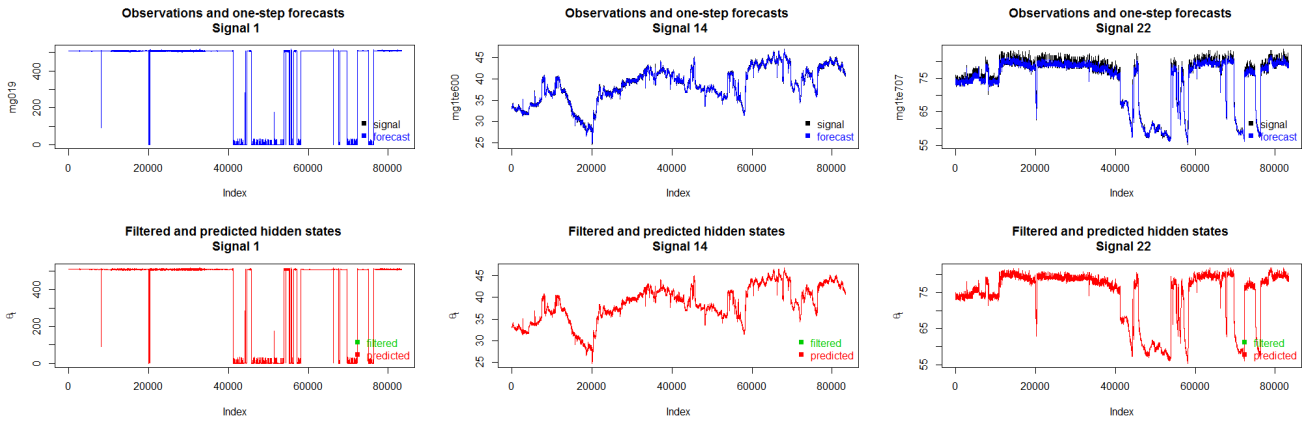

Figure 11. Predictions and filtering on the test data for the independent model; selected signals for engine 1

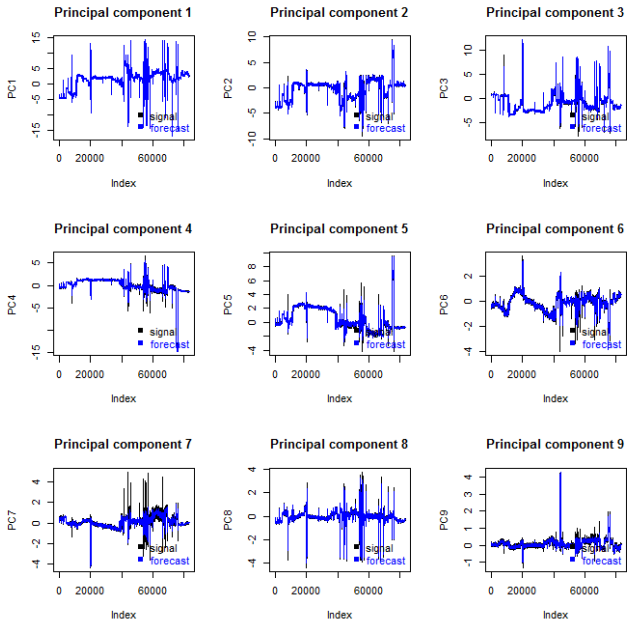

Figure 12. Expected 1-step ahead forecasts of the signals from the Dynamical Linear Model on the principal components of the residuals from regression on engine speed

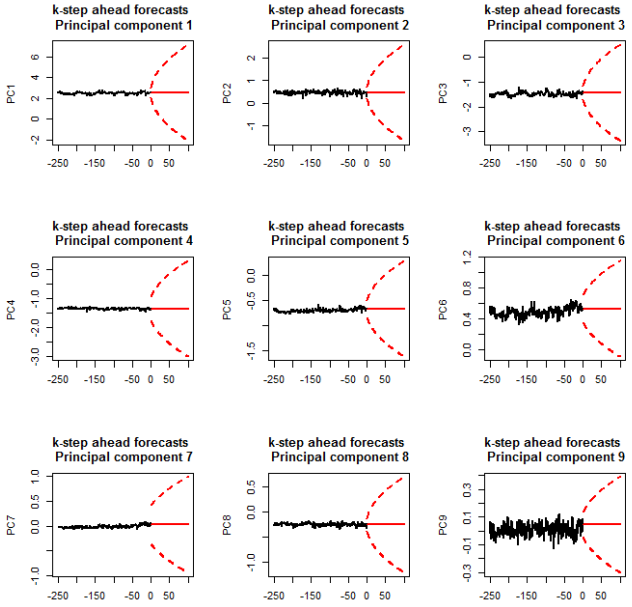

Figure 13. Expected 1-step ahead forecasts of the signals from the Dynamical Linear Model on the principal components of the residuals from regression on engine speed

Figure 18. In this example, there are two alarms and it can be seen how the SPRT builds evidence for or against the null hypothesis (normal state) in approximately four measurements. The sensitivity of the SPRT can be fine-tuned by tuning the parameters α, β and μ.

The number of alarms for positive and negative mean changes are shown in Table 4. For simplicity, the independent model on the raw data are referred to as Ind Raw, the model on the residuals from regression on Engine speed is referred to as Reg ES, the model on the principal components of the same residuals are referred to as PC Reg ES and the model on the principal components of the residuals from regression on cluster membership are referred to as PC Reg Cluster. The table reports the total number of alarms on any signal (total alarms) and also the total number of unique alarms, meaning that only unique time-steps where there is an alarm is counted. Obviously, on some time-steps there will be simultaneous alarms from different signals. The total alarm ratio is then the ratio of time-steps where there is an alarm.

From Table 4 it can be seen that the number of alarms are larger for the model based on the raw data. Comparing the two models based on principal components of residuals, the model on residuals from the regression on cluster memberships give more alarms than the one based on the residu-

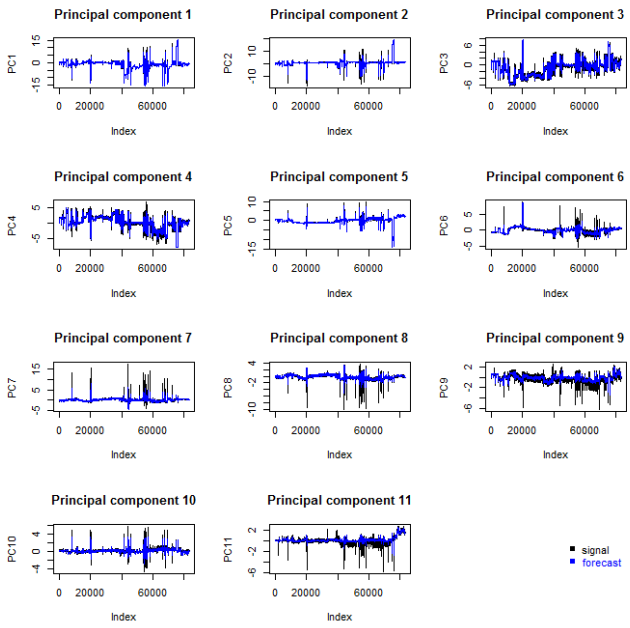

Figure 14. Expected 1-step ahead forecasts of the signals from the Dynamical Linear Model on the principal components of the residuals from regression on cluster membership

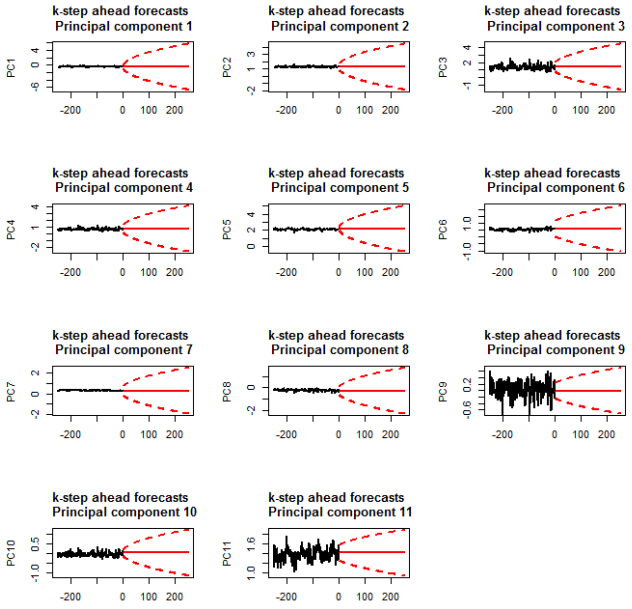

Figure 15. Expected k-step ahead forecasts of the signals from the Dynamical Linear Model on the principal components of the residuals from regression on cluster membership

Table 4. Number of anomalies detected with each modelling alternative (total number of anomalies for all signals)

	Ind Raw	PC Reg ES	PC Reg Cluster
N	83 286	83 286	83 286
Positive mean change	6307	1852	5344
Negative mean change	5036	1586	5470
Total alarms	11343	3438	10814
Total unique alarms	5885	1431	4554
Total alarm ratio	0.0707	0.0172	0.0547

als from regression on engine speed. The total alarm ratios vary from 0.07, meaning that an alarm is raised every 14 time step, to 0.017, corresponding to an alarm about every 60th time step. Since there are no known faults in this data, there should presumably not be any anomalies and most of the alarms could be interpreted as false alarms. Hence, the model with the lowest number of alarms could be preferred. However, the large spikes in the pre-processed data are difficult to predict for the dynamical linear models, and it is not unreasonable for them to flag an anomaly each time such a spike is encountered. Hence, these false alarms are not due to failure of the DLMs to model the data, but rather the failure of the pre-processing steps in removing these spikes. One may also see this as an alarm whenever the engine suddenly shifts from one steady state to another. More study on how to account for the transient states in the original data could improve this situation if pre-processed data without these spikes could be fed into the DLM framework. On the other hand, for data displaying transient states from one steady state to another, it may not be unreasonable with an alarm when entering into the transient states. At any rate, the alarm ratios for the individual signals vary from 0.0011 to 0.024 (independent DLM on raw signals), 0.0026 to 0.0053 (regression model based on engine speed) and from 0.0032 to 0.028 (regression model on clusters) and this is not unreasonable given that the allowed probability of false alarms were set to 0.005 for both positive and negative mean level change separately.

4.5. Model Selection

The number of alarms raised over the dataset can be used as an indication of which model performs best on the sensor data. However, more formal statistical metrics may also be applied, and the Akaike Information Criterion (AIC) and the root mean square error (RMSE) have been calculated for the different models. The AIC is based on the likelihood and favours the models with largest likelihood, but also includes a penalty for large number of parameters. The model with smallest AIC are normally selected. The RMSE is calculated based on the one-step ahead forecast errors and is calculated for the test data only. It should be noted that these metrics are not entirely appropriate in this case, since the DLMs have not been fitted to the same data. Since different pre-processing of the raw data have been performed prior to fitting the different models, comparing the models by terms of likelihood-based metrics or RMSE may not be fair. For a more fair comparison, one could try to use the forecasted principal components to transform back to the residuals of the regression model and

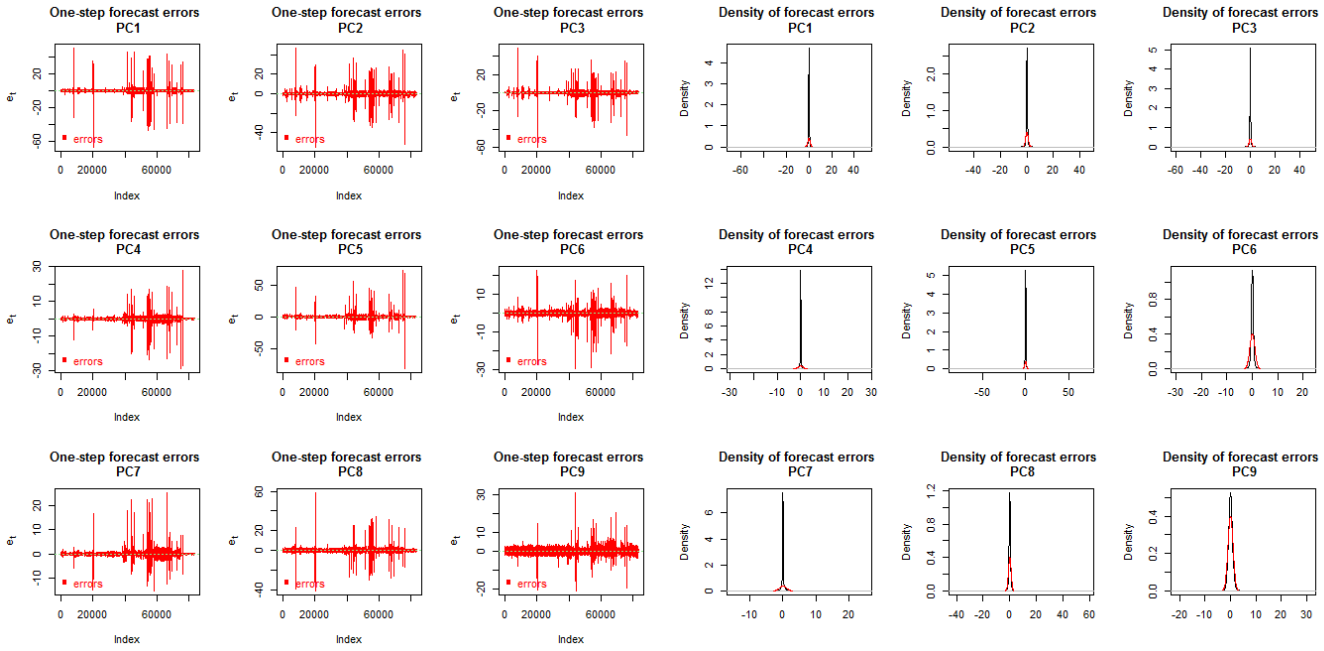

Figure 16. One-step forecast errors (left) and density of 1-step forecast errors (right) for DLM fitted to principal components of residuals from a regression model on engine speed

then compare this with predictions from the regression models, but this has not been pursued in this study. Notwithstanding, the AIC and the RMSE for each model is reported in Table 5, and it gives an indication of how the models perform. However, model selection is perhaps best carried out by looking at the performance in terms of ratio of alarms.

Table 5. AIC and RMSE for the different models

	Ind Raw	PC Reg ES	PC Reg Cluster
k	46	18	22
AIC_{train}	6 215 338	−6 584 128	−5 274 406
AIC_{test}	2 021 041	−2 283 094	−1 501 852
RMSE_{test}	84.47	0.1578	0.3158

Even though the model selection criteria may not be fully adequate in this setting, it indicates that the dynamical linear model performs better for anomaly detection on the residuals on a sensible regression analysis and that it is worthwhile to perform some preprocessing of the data. Comparing the alternatives based on principal components of residuals, the dynamical linear models on principal components of residuals from a regression model on engine speed seems to be preferable to the one based on regression on cluster memberships. This model has lower RMSE on the test data and also lower AIC on both the training and the test data. Moreover, it is a simpler model with fewer parameters. This modelling approach also resulted in fewer alarms, which essentially means that there were fewer spikes in the pre-processed data. Hence, this modelling may be preferable, among the ones investigated in this study.

5. DISCUSSION

The approach presented in this paper utilizes dynamical linear models for monitoring of multivariate sensor signals. Based on the forecast distributions of future signals and comparing these with the sensor measurements as they arrive, online model monitoring can be performed by sequential tests. However, a crucial prerequisite for this approach to work well is the ability to establish a model that describes the sensor data well. This study has revealed that this can be challenging but feasible.

In this study, different DLMs have been tried out on a set of sensor signals measuring various variables of a marine generator engine. As it turns out, it is difficult to establish good models on the raw sensor signals. Two reasons for this is the sudden shifts in the data between different operational modes and the somewhat high dimensionality of the data. The sudden shifts between main modes of operation is of course difficult for a statistical model to predict, and full DLMs for the 23-dimensional dataset would include quite many model parameters that would be hard to estimate accurately. In order to come around these problems, some pre-processing steps were carried out before fitting the DLMs, i.e. regression models to try to filter out the sudden shifts and principal component analysis to reduce the dimensionality and also the cross-correlations in the data. Even though the models on

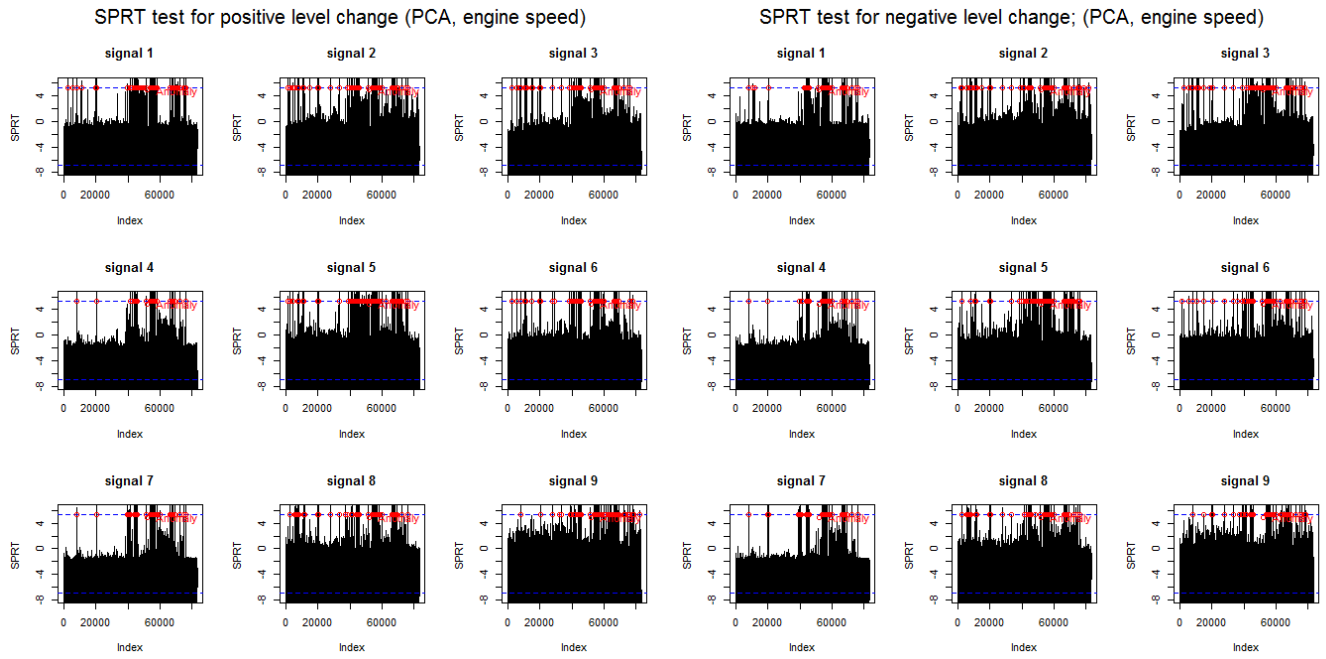

Figure 17. Sequential probability ratio tests (SPRT) on the one step ahead forecast errors from DLMs fitted to the principal components of the residuals from a regression model on engine speed; positive mean level change (left) and negative mean level change (right)

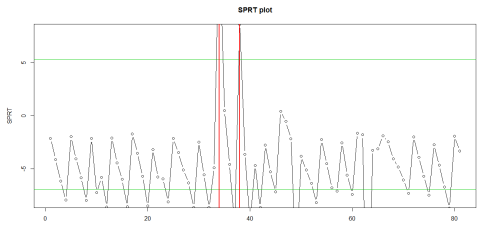

Figure 18. A close-up of the SPRT plot for a short time-window

the pre-processed data appear to be performing better than on the raw data, there are still some problems that needs to be solved. For example, even though the pre-processing steps diminish the effect of the different operational modes in the data, there are still spikes in the residual time series, mostly coinciding with the transient states between modes. Other pre-processing techniques may be applied in order to reduce these further.

Other extraneous factors that may influence the sensor signals, and that is perhaps particular to ship systems, are the loads and responses from the ship environment. It is possible that the sensor signals from a ship in perfect condition may look very different in calm conditions compared to stormy seas. Such operational conditions will then introduce larger variability in the recorded sensor signals and need to be captured by the statistical models used for anomaly detection. Clearly, one wants to distinguish between anomalous behaviour of the ship system and anomalous environmental conditions. Possibly, such environmental conditions could be used as covariates in a regression-type component of the model. Previously, it has been shown that up to 50% of the variation in the efficiency of a ship machinery system can be explained by such covariates, suggesting that such effects should be taken into account (Vanem et al., 2016). If such operational conditions are found to be very influential on the monitoring signals and, consequently, on the sensor based anomaly detection, methods for conditional or contextual anomaly detection might be required (Song, Wu, Jermaine, & Ranka, 2007; Hayes & Capretz, 2015).

It should be possible to construct a hierarchical model where the condition of the ship or the system being monitored is directly modelled by some unobserved state. For example, the hidden layer could be made up of binary or multinary variables indicating whether the system being monitored is in a fully functional state, some intermediate degraded state or a system failure state. Condition monitoring could then be performed by analysing the predictive, filtered and smoothed distribution on the system level directly, and alarms could be raised when the probability of being in a degraded or failed state is higher than some acceptable value. Possibly, DLM could be extended to include diagnostics in such a setting. However, additional information about the actual state of the system would presumably be needed in order to train such a model to be able to recognize such underlying states from the

sensor signals. Further investigation of this kind of usage of DLM for condition monitoring is left for further research.

5.1. Composite Sequential Probability Ratio Tests

The performance of the model is monitored by way of Bayes' factors or sequential probability ratio tests. In standard SPRT, the established model, the null model M_0, is compared to a specific alternative model, M_1, in simple statistical tests. In the study presented herein, two alternative models were used, with a positive and negative expected forecast error, respectively.

One way to improve such tests is to formulate composite one-sided SPRTs. Thus, rather than testing a simple null hypothesis $\mu = 0$ against a specific simple alternative hypothesis $\mu = m$, one can test the simple null hypothesis against the composite alternative hypothesis $\mu \geq m$. This will give more weight to extreme deviations larger than m. A composite SPRT for the Poisson model was proposed in (Kulldorff et al., 2011). A similar extension for the normal case could be the following modification of eq. 14 (for positive mean change; assuming still unit variance).

$$l_n = \sum_{i=1}^{n} \left[\mu \left(e_i - \frac{\mu}{2} \right) I(e_i \leq \mu) + \frac{e_i^2}{2} I(e_i > \mu) \right]. \quad (18)$$

This index would treat each new observation independently of the previous ones and the indicator function ensures that each new term in the sum depends on whether the value of the new residual is above or below the hypothesized value μ. For values of the residuals smaller than or equal to the μ-value, the added term will be identical to the standard simple test, but for values greater than μ, the modified SPRT index will pick up a larger contribution proportional to e_i^2 rather than e_i. Hence, this modified test will more quickly pick up alarms for large deviations from the null hypothesis (large compared to the alternative hypothesis). Hence, faults are detected more quickly for large residuals.

The above modification treats each new observation (residual) independently and is a true sequential test in that regard. However, if one rather look at the average residual since the SPRT index was last reset one can formulate a test based on all observations within the interval of the test. This will then not be a truly sequential test, since it updates the complete sum rather than merely adding a term at each step, but it might be more appropriate. Such a SPRIT index would be on the following form, where $\bar{e}$ refers to the average residual since the SPRT was last reset to zero, i.e. $\bar{e} = \frac{1}{n} \sum_{i=1}^{n} e_i$,

$$l_n = \sum_{i=1}^{n} \left[\mu \left(e_i - \frac{\mu}{2} \right) \right] I(e_i \leq \mu) + \sum_{i=1}^{n} \bar{e} \left[e_i - \frac{\bar{e}}{2} \right] I(e_i > \mu). \quad (19)$$

For large values of e_i compared to μ, such a modified SPRT would also increase more quickly towards the threshold for raising an alarm.

One could also formulate two-sided tests, rather than two one-sided test (positive and negative mean change), e.g. testing $\mu = 0$ towards $\mu = |m|$ or two-sided composite tests against $\mu \geq |m|$. However, such tests might not work as well in a sequential setting, where tests are updated sequentially as new observations arrive. With a two sided test, one would not be able to distinguish between successive deviations to one side (e.g. indicating a drift of the signal) and alternating deviations on either side. If this is important information, it might be advisable to rather formulate several tests and perform separate tests for positive and negative drifts, respectively. For situations where there are alternating deviations to either side, additional tests for increase (or decrease) in the variance of the forecast errors could be formulated. Other tests could be formulated with e.g. composite H_0 against one- or two-sided H_1, one-sided H_0 against one-sided H_1, etc. However, it should be kept in mind that multiple testing will give more frequent false alarms just by chance (multiple testing problem), and this might have to be compensated for.

6. CONCLUSIONS

This paper applies the framework of dynamical linear models on a set of sensor signals from a marine generator engine system in order to detect anomalies and monitor the condition of the system. Different pre-processing of the data are applied in order to lighten the burden of the DLM and make it easier to fit the models to the signals. These include various regression models, principal component analysis and clustering techniques. This study demonstrates that this improves the performance of the dynamical linear models, even though there are still challenges in handling spikes in the pre-processed data which remains from the transient parts of the original signals. These spikes can be construed as anomalies resulting in too many warnings being issued by the anomaly detection scheme.

It may not be unreasonable for a condition monitoring system based on dynamical linear models to flag an alarm when the signals indicate a sudden departure from one steady state to another, but it is believed that the frequency of such false alarms can be significantly reduced if more precise pre-processing steps are applied. This is an area for further research. Moreover, it is acknowledged that DLM is a framework that will work better for slowly varying systems, and it may not be ideally suited for processes where there are frequent and abrupt changes in the signal values.

Nevertheless, the framework presented in this paper, based on some pre-processing steps and utilizing dynamical linear models and sequential testing, are able to detect sudden jumps in the data and to issue alarms whenever this is detected, and thereafter swiftly recover to the new level of the signals to

continue to model the normal behaviour of the system. This indicates that a system such as the one presented in this paper may be useful for anomaly detection.

ACKNOWLEDGMENT

The study presented in this paper is partly carried out within the centre for research-based innovation, BigInsight.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716-723.

Baraldi, P., Di Maio, F., Pappaglione, L., Zio, E., & Seraoui, R. (2012). Conditional monitoring of electrical power plant components during operational transients. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *226*, 568-583.

Brandsæter, A., Manno, G., Vanem, E., & Glad, I. K. (2016, June). An application of sensor based anomaly detection in the maritime industry. In *Proc. IEEE PHM2016.*

DNV GL. (2015). *Ship connectivity* (Tech. Rep. No. Strategic Research and Innovation Position Paper 4-2015). DNV GL.

DNV GL. (2017a). *Rules for classification - general regulations.* (DNVGL-RU-0050)

DNV GL. (2017b). *Rules for classification: Ships.* (DNVGL-RU-SHIP)

Eleftheria, E., Papanikolaou, A., & Voulgarellis, M. (2016). Statistical analysis of ship accidents and review of safety level. *Safety Science*, *85*, 282-292.

Hayes, M. A., & Capretz, M. A. (2015). Contextual anomaly detection framework for big sensor data. *Journal of Big Data*, *2*, 1-22.

Hines, J. W., & Garvey, D. R. (2006). Development and application of fault detectability performance metrics for instrument calibration verification and anomaly detection. *Journal of Pattern Recognition Research*, *1*, 2-15.

IMO. (2014a). *International Safety Management Code (ISM Code) with guidelines for its implementation (2014 edition).* International Maritime Organization.

IMO. (2014b). *SOLAS, consolidated edition 2014.* International Maritime Organization.

Kulldorff, M., Davis, R. L., Kolczak, Lewis, E., Lieu, T., & Platt, R. (2011). A maximized sequential probability ratio test for drug and vaccine safety surveillance. *Sequential Analysis*, *30*, 58-78.

Mindykowski, J., & Tarasiuk, T. (2015). Problems of power quality in the wake of ship technology development. *Ocean Engineering*, *107*, 108-117.

Niculita, O., Nwora, O., & Skaf, Z. (2017). Towards design of prognostics and health management solutions for maritime assets. *Procedia CIRP*, *59*, 122-132.

Psarros, G. A. (2015, May-June). Comparing the navigator's response time in collision and grounding accidents. In *Proc. 34th international conference on ocean, offshore and arctic engineering (omae 2015).*

Song, X., Wu, M., Jermaine, C., & Ranka, S. (2007). Conditional anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*, *19*, 631-645.

Vanem, E., Brandsæter, A., & Gramstad, O. (2016, September). Regression models for the effect of environmental conditions on the efficiency of ship machinery systems. In *Proc. esrel 2016.*

Vanem, E., Rusås, S., Skjong, R., & Olufsen, O. (2007). Collision damage stability of passenger ships: Holistic and risk-based approach. *International Shipbuilding Progress*, *54*, 323-337.

Vanem, E., & Skjong, R. (2004, July). Fire and evacuation risk assessment for passenger ships. In *Proc. 10th international fire science and engineering conference (interflam) 2004* (Vol. 1, p. 365-374).

Vanem, E., & Storvik, G. O. (2016, August). Dynamical linear models for condition monitoring with multivariate sensor data. In *Proc. comadem 2016.*

Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, *16*, 117-186.

West, M., & Harrison, J. (1997). *Bayesian forecasting and dynamic models* (Second ed.). Springer-Verlag.

Zymaris, A. S., Alnes, Ø. Å., Knutsen, K. E., & Kakalis, N. M. P. (2016, July). Towards a model-based condition assessment of complex marine machinery systems using systems engineering. In *Proc. Third European Conference of the Prognostics and Health Management Society 2016.*