Prognosis of gear health using stochastic dynamical models with online parameter estimation

Matej Gašperin [1], Pavle Boškoski [1], and Dani Juričič [1]

[1] *Jožef Stefan Institute, Ljubljana, Slovenia*
matej.gasperin@ijs.si
pavle.boskoski@ijs.si
dani.juricic@ijs.si

ABSTRACT

In this paper we present a statistical approach to the estimation of the time in which an operating gear will achieve the critical stage. The approach relies on measured vibration signals. From these signals features are extracted first and then their evolution over time is predicted. This is done owing to the dynamic model that relates hidden degradation phenomena with measured outputs. The Expectation-Maximization algorithm is used to estimate the parameters of the underlying state-space model on-line. Time to reach safety alarm threshold is determined by making the prediction using the estimated linear model. The results obtained on a pilot test bed are presented.

1 INTRODUCTION

On-line condition monitoring of rotational machinery has become an almost indispensable part of modern control and supervision systems. Due to almost four decades of development of the underlying methodologies the field is believed to have reached the maturity phase. On the other hand, progress in the area of prognosis of the machine condition is relatively recent so that a lot of work has to be done in the future (Howard, 1994). The driving force is the obvious importance of accurate predictions of fault propagation. This allows for alarm setting well before the machine reaches the critical stage.

In this paper we present a data driven approach to gear health monitoring and prognosis. We employ vibration signals and their processing by means of Hilbert transform (HT) (Rubini and Meneghetti, 2001; Ho and Randall, 2000). Components from Hilbert spectrum are used as fault indicators (or features). The sequence of features can be viewed as a time series i.e. a realization of the stochastic process, which is tightly related to the condition of the gear. The objective of this work is to extend feature extraction with the prediction of the feature time series in order to estimate the time of occurrence of the safety alarm (first passage time - FPT).

As feature evolution over time is a nonstationary stochastic process, we focus on an adaptive on-line tracking of the essential parameters of the dynamic model that describes the process. In the literature there are several different techniques available for time-series prediction (Wang *et al.*, 2003), among them neural networks and neuro-fuzzy systems (W. Q. Wanga and Ismailb, 2004). In this paper we assume that the time series can be viewed as an output of a second order linear, discrete time, stochastic dynamic system. Because of the progress of faults in gears, the systems parameters are expected to change in time. For on-line estimation of model parameters an Expectation-Maximization (EM) algorithm is used. The EM algorithm is a well established method, which arose in the mathematical statistics community (Dempster *et al.*, 1977), but has found wide engineering applications in different areas. One of the applications is the use of the algorithm for estimating state-space model parameters in the presence of hidden variables (unmeasured states) (Gibson and Ninness, 2005).

Fault prognosis techniques developed so far can be roughly divided into three different classes: (1) exploiting experimental models, (2) approaches based on physical models and (3) data driven techniques.

In the first case, the models of the components are typically designed by experts and validated on a large set of experimental data (Howard, 1994). The underlying models typically take the form of the probability laws, which describe the probability for the occurence of failure. Generally, such an approach is costly and applicable to a specific class of systems components.

In the second class of the approaches the damage model of the component is derived from either physical of semi-physical model in the form of state-space models. These models are further enhanced with the failure propagation entries in terms of deterministic or stochastic states. Unknown parameters can be obtained by means of the stochastic filtering approaches (Orchard and Vachtsevanos, 2009).

The third category is least demanding in terms of prior knowledge. Actually these approaches rely on a set of features, which correlate with the failure evolution over time. The time-to-failure can be estimated from the operating data provided an appropriate training process has been conducted first. The key enabler in these approaches is time series prediction which can be solved in many ways (W. Q. Wanga and Ismailb, 2004),(G. Niu, 2009). According to this segmentation, our approach can be placed in the last category.

The outline of the paper is as follows. Chapter 2 will present the experimental setup used for this study along with the experimental protocol and feature extraction. Chapter 3 will describe the stochastic time series properties and prediction along with the EM algorithm for model parameter estimation. Results of prediction using the data from the test bed are presented in Chapter 4.

2 THE EXPERIMENTAL SET-UP

The experimental test bed consists of a motor-generator pair with a single stage gearbox (Fig. 1). The motor is a standard DC motor powered through a Simoreg DC drive. A generator is being used as a break. The generated power is being fed back in the system, thus achieving the breaking force.

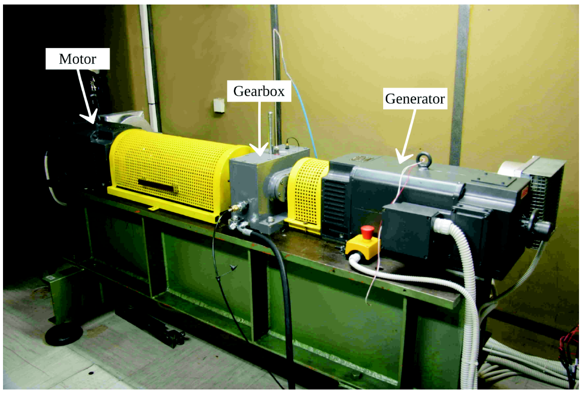


Figure 1: The test bed

The test bed is equipped with 8 accelerometers. The mounting position and sensitivity axis of each accelerometer are shown in Figure 2.

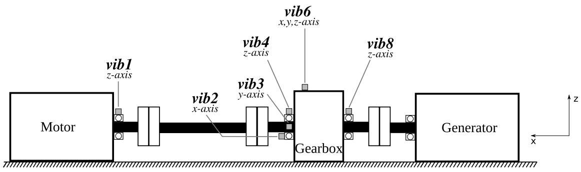


Figure 2: Vibration sensors placement scheme

The placment of the sensor determines its sensitifity for incresed vibrations caused by the degrated gear health. The highest sensitivity in our experiment has been observed in sensors labeled *vib8* and *vib3*, which measure the vibrations on gearbox output and input shafts respectively.

2.1 The experimental protocol

The test run was done with a constant torque of 82.5Nm and constant speed of 990rpm. This speed of 990rpm generates gear mesh frequency (GMF) $f_{gm} = 396$Hz, rotational speed of input shaft $f_i = 16.5$Hz, and rotational speed of output shaft $f_o = 24.75$Hz. The signals were sampled with sampling frequency $f_s = 80$kHz. Each acquisition session lasted for 5 seconds. The acquisition was repeated every 10 minutes as illustrated in Figure 3.

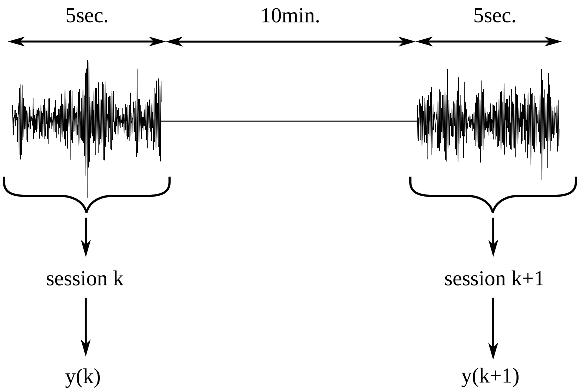


Figure 3: The concept of signal acquisition

In order to speed up the experiment, the contact surface between the gears was decreased to 1/3 of the original surface. In this manner the fault evolution horizon is made shorter. The displacement is shown in Fig. 4.

The overall experiment run took 65 hours. At the end both gears were heavily damaged i.e. on both gears spalling can be seen on all teeth, which progressed even into a plastic deformation of some of them, as shown in Fig. 4.

2.2 Feature extraction

Signals from all 8 sensors were acquired simultaneously. These signals acquired at each acquisition session, were analyzed using envelope analysis (Ho and Randall, 2000). From each sensor $s_i \in \{s_1, \cdots, s_8\}$, at each acquisition session k, a feature vector $[y_{s_i,1}(k), y_{s_i,2}(k), \cdots, y_{s_i,m}(k)]$ was derived, where m is the total number of extracted features. Each element of the feature set represents the value of the amplitude of specific spectral component from the envelope spectrum for the particular sensor s_i.

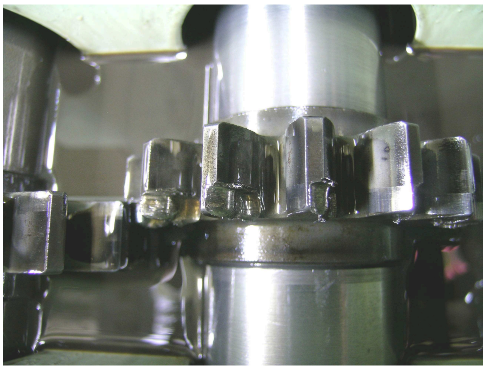
3 STOCHASTIC TIME SERIES PREDICTION

3.1 State space model of time series

State-space representation is a very general model, that can describe a whole set of different models. In our case we assume that condition of the machine is a dynamic process influenced by random tribological inputs which occur due to the impact between moving surfaces. Condition can be viewed as a random process, which can be described by a state space model:

$$\begin{aligned} \mathbf{x}_{t+1} &= f(\mathbf{x}_t, \mathbf{w}_t, \boldsymbol{\Theta}), \\ \mathbf{y}_t &= g(\mathbf{x}_t, \mathbf{e}_t, \boldsymbol{\Theta}) \end{aligned} \tag{1}$$

where $\mathbf{y}_t$ is a measured data (i.e. the output of the vibration sensor), $\mathbf{x}_t$ an unmeasured system state, $\mathbf{w}_t$ is an i.i.d. random process, $\mathbf{e}_t$ measurement noise and $\boldsymbol{\Theta}$ model parameters.



(a) Output gear

(b) Input gear

Figure 4: Output and input gear at the end of the experiment (notice heavily pitted teeth)

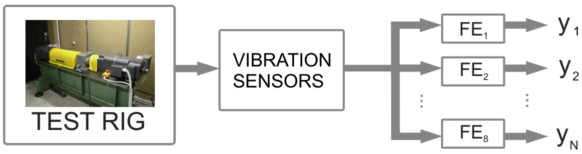


Figure 5: Feature extraction procedure

For practical use, the expression (1) can be in many cases simplified and transformed into linear form. The resulting model is described as

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}\mathbf{x}_t + \mathbf{w}_t, \\ \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t + \mathbf{e}_t, \end{aligned} \tag{2}$$

We assume that the system starts with initial vector $\mathbf{x}_0$ with mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$.

The observed system output data, indexed by time $(\mathbf{y}_t)$, which represent the time series we wish to analyze or predict.

3.2 EM algorithm for dynamic state-space system estimation

Expectation-Maximization is applied as an iterative method to estimate a vector of unknown parameters $(\boldsymbol{\Theta})$, given measurement data $(\boldsymbol{Y} = \{y_1, y_2, \ldots, y_n\})$. In other words, we wish to find the set of parameters $\boldsymbol{\Theta}$, such that $p(\mathbf{Y}|\boldsymbol{\Theta})$ is a maximum. This estimate of $\boldsymbol{\Theta}$ is known as *maximum likelihood (ML) estimate*. Usually, the log-likelihood function of the parameters is defined as,

$$L(\boldsymbol{\Theta}) = \ln p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\Theta}), \tag{3}$$

where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$. Since $\ln(x)$ is a strictly increasing function, the value of Θ that maximizes $p(\mathbf{Y}|\mathbf{X}, \boldsymbol{\Theta})$ also maximizes $L(\boldsymbol{\Theta})$.

$$\boldsymbol{\Theta}_{k+1} = \arg\max_{\boldsymbol{\Theta}} \left\{ E_{\mathbf{X}|\mathbf{Y},\boldsymbol{\Theta}_k} \left\{ \ln p(\mathbf{Y}, \mathbf{x}|\boldsymbol{\Theta}) \right\} \right\} \tag{4}$$

EM algorithm is an iterative procedure for maximizing $L(\boldsymbol{\Theta})$, meaning that after k^{th} iteration, we obtain the estimate for $\boldsymbol{\Theta}$, denoted by $\boldsymbol{\Theta}_k$.

The advantage of the algorithm is, that it can also operate when system states $(\mathbf{x})$ are not known. Because the output is dependant on the unobserved system states, direct maximization is not possible. The EM algorithm alternates between two steps, first maximizing the likelihood function with respect to the system states (E-step) and than with respect to the parameters (M-step).

E-step

Given an estimete of the parameter values $(\boldsymbol{\Theta}_k = \{\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{C}, \boldsymbol{R}, \boldsymbol{x}_0, \boldsymbol{Q}_0\})$, the Rauch-Tung-Striebel (RTS) smoother provides an optimum estimate of the unobserved state sequence $(\boldsymbol{x}_t)$ of a state space model (2). In dynamical systems with hidden states the E-step corresponds directly to solving the smoothing problem.

In other words, for any time t we would like to compute

$$p(\mathbf{x}_t|\mathbf{y}_{1:T}) \tag{5}$$

where $\mathbf{y}_{1:T} = \{y_1, y_2, \ldots, y_T\}$, $T > t$. The RTS smoother procedure is as follows:

- First, Kalman filter is applied to the observable data in a forward manner ($t = 1, 2, \ldots, n$), starting with initial estimate of the state mean and variance $(\mathbf{x}_0, \boldsymbol{\Sigma}_0)$.
- Recursive smoother is applied in a backward manner ($t = n, n-1, \ldots, 1$), taking the filtered state estimate at time n as initial condition.

Summary of the algorithm is given in Table 1.

M-step

As stated above, the vector of unknown system parameters in the case of state space model is given as

$$\boldsymbol{\Theta} = \{\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{C}, \boldsymbol{R}, \boldsymbol{x}_0, \boldsymbol{Q}_0\} \tag{6}$$

If the system states can be observed in addition to system outputs, the joint pdf can be written as

$$p(\boldsymbol{Y}|\boldsymbol{\Theta}) \quad = \quad p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\Theta}) p(\boldsymbol{X}|\boldsymbol{\Theta})$$

Assuming Gaussian distributions and ignoring the constants, the complete data log-likelihood can be written as

Table 1: RTS smoother algorithm

Forward filter

Initialization

$\boldsymbol{x}_0 = \boldsymbol{x}_0$

$\boldsymbol{P}_0 = \boldsymbol{P}_0$

Computation: For $t = 1, 2, \ldots, n$

$\boldsymbol{x}_{t+1|t} = \boldsymbol{A}\boldsymbol{x}_{t|t}$

$\boldsymbol{P}_{t+1|t} = \boldsymbol{A}\boldsymbol{P}_{t|t}\boldsymbol{A}^T + \boldsymbol{Q}$

$\boldsymbol{K}_t = \boldsymbol{P}_{t+1|t}\boldsymbol{C}^T(\boldsymbol{C}\boldsymbol{P}_{t+1|t}\boldsymbol{C}^T + \boldsymbol{R})^{-1}$

$\boldsymbol{x}_{t+1|t+1} = \boldsymbol{x}_{t+1|t} + \boldsymbol{K}_t(y_{t+1} - \boldsymbol{C}\boldsymbol{x}_{t+1|t})$

$\boldsymbol{P}_{t+1|t+1} = \boldsymbol{P}_{t+1|t} - \boldsymbol{K}_t\boldsymbol{C}\boldsymbol{P}_{t+1|t}$

Backward filter

Computation: For $T = n, n-1, \ldots, 1$

$\boldsymbol{J}_t = \boldsymbol{P}_{t|t}\boldsymbol{A}^T\boldsymbol{P}^{-1}_{t+1|t}$

$\boldsymbol{x}_{t|T} = \boldsymbol{x}_{t|t} + \boldsymbol{J}_t(\boldsymbol{x}_{t+1|T} - \boldsymbol{x}_{t+1|t})$

$\boldsymbol{P}_{t|T} = \boldsymbol{P}_{t|t} - \boldsymbol{J}_t(\boldsymbol{P}_{t+1|T} - \boldsymbol{P}_{t+1|t})\boldsymbol{J}_t^T$

$$\begin{aligned} -2\ln L(\mathbf{\Theta}) &= \ln|\mathbf{\Sigma}_0| + (\mathbf{x_0} - \mu_0)'\mathbf{\Sigma}_0^{-1}(\mathbf{x_0} - \mu_0) \\ &+ n\ln|\mathbf{Q}| \\ &+ \sum_{t=1}^{n}(\mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1})'Q^{-1}(\mathbf{x}_t - \mathbf{A}\mathbf{x}_{t-1}) \\ &+ n\ln|\mathbf{R}| \\ &+ \sum_{t=1}^{n}(\mathbf{y}_t - \mathbf{C}\mathbf{x}_t)'R^{-1}(\mathbf{y}_t - \mathbf{C}\mathbf{x}_t) \end{aligned} \quad (7)$$

where $\mathbf{x}_0 \sim N(\mu_0, \mathbf{\Sigma}_0)$. Taking the expected value of the expression with respect to the current parameter estimate $(\mathbf{\Theta}_k)$ and complete observed data $\mathbf{Y}_n$

$$l(\mathbf{\Theta}|\mathbf{\Theta}_k) = E\left\{-2\ln L(\mathbf{\Theta})|\mathbf{Y}_n, \mathbf{\Theta}_k\right\} \quad (8)$$

Using the results from RTS smoother equations:

$$E_{\mathbf{X}|\mathbf{Y_n},\mathbf{\Theta}_k}(\mathbf{x}_t\mathbf{x}_t') = \mathbf{x}_t^n\mathbf{x}_t^{n\prime} + \mathbf{P}_t^n \quad (9)$$

$$E_{\mathbf{X}|\mathbf{Y_n},\mathbf{\Theta}_k}(\mathbf{x}_t\mathbf{x}_{t-1}') = \mathbf{x}_t^n\mathbf{x}_{t-1}^{n}{}' + \mathbf{P}_{t,t-1}^n \quad (10)$$

$$E_{\mathbf{X}|\mathbf{Y_n},\mathbf{\Theta}_k}(\mathbf{x}_{t-1}\mathbf{x}_{t-1}') = \mathbf{x}_{t-1}^n\mathbf{x}_{t-1}^{n}{}' + \mathbf{P}_{t-1}^n \quad (11)$$

$$E_{\mathbf{X}|\mathbf{Y_n},\mathbf{\Theta}_k}(\mathbf{x}_t) = \mathbf{x}_t^n \quad (12)$$

and taking expectation over the expression (7) yields

$$\begin{aligned} & l(\mathbf{\Theta}|\mathbf{\Theta}_k) = \\ = \; & \ln|\mathbf{\Sigma}_0| \\ + \; & \mathrm{tr}\left[\mathbf{\Sigma}_0^{-1}\left(\mathbf{P_0^n} + (\mathbf{x_0}^n - \mu_0)(\mathbf{x_0}^n - \mu_0)'\right)\right] \\ + \; & n\ln|\mathbf{Q}| \\ + \; & \mathrm{tr}\left[\mathbf{Q}^{-1}\left\{\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{A}' - \mathbf{A}\mathbf{S}_{10}' + \mathbf{A}\mathbf{S}_{00}\mathbf{A}'\right\}\right] \\ + \; & n\ln|\mathbf{R}| \\ + \; & \mathrm{tr}\left[\mathbf{R}^{-1}\left\{\mathbf{S}_{22} - \mathbf{S}_{20}\mathbf{C}' - \mathbf{C}\mathbf{S}_{20}' + \mathbf{C}\mathbf{S}_{11}\mathbf{C}'\right\}\right] \end{aligned} \quad (13)$$

where

$$\mathbf{S}_{11} = \sum_{t=1}^{n}\mathbf{x}_t^n\mathbf{x}_t^{n\prime} + \mathbf{P}_t^n \quad (14)$$

$$\mathbf{S}_{10} = \sum_{t=1}^{n}\mathbf{x}_t^n\mathbf{x}_{t-1}^{n}{}' + \mathbf{P}_{t,t-1}^n \quad (15)$$

$$\mathbf{S}_{00} = \sum_{t=1}^{n}\mathbf{x}_{t-1}^n\mathbf{x}_{t-1}^{n}{}' + \mathbf{P}_{t-1}^n \quad (16)$$

$$\mathbf{S}_{22} = \sum_{t=1}^{n}E(\mathbf{y}_t\mathbf{y}_t') = \sum_{t=1}^{n}(\mathbf{y}_t\mathbf{y}_t') \quad (17)$$

$$\mathbf{S}_{20} = \sum_{t=1}^{n}\mathbf{y}_t\mathbf{x}_t^{n\prime} \quad (18)$$

Our goal is to find the maximum of the function $l(\mathbf{\Theta}|\mathbf{\Theta}_n)$ with respect to parameters $\mathbf{A}, \mathbf{C}, \mathbf{Q}, \mathbf{R}, \mathbf{x_0}, \mathbf{\Sigma_0}$. Calculating derivatives of the Eq. (13) with respect to all parameters we obtain the following result

$$\mathbf{A} = \mathbf{S}_{10}\mathbf{S}_{00}^{-1} \quad (19)$$

$$\mathbf{Q} = n^{-1}\left(\mathbf{S}_{11} - \mathbf{S}_{10}\mathbf{S}_{00}^{-1}\mathbf{S}_{10}'\right) \quad (20)$$

$$\mathbf{C} = \mathbf{S}_{20}\mathbf{S}_{11}^{-1} \quad (21)$$

$$\mathbf{R} = n^{-1}\left(\mathbf{S}_{22} - \mathbf{S}_{20}\mathbf{S}_{11}^{-1}\mathbf{S}_{20}'\right) \quad (22)$$

$$\mathbf{x}_0 = \mathbf{x}_0^n \quad (23)$$

$$\mathbf{Q}_0 = \mathbf{P}_0^n \quad (24)$$

Both E and M steps are iterated until the increase in the likelihood at the current time step, compared to the previous one, is greater than the selected threshold.

Time series prediction

With known model parameters, predicting the future values of the time series is straightforward. We start at the time of the last measure, with the filter estimate of the state vector distribution ($\boldsymbol{x}_{n|n}$, $\boldsymbol{P}_{n|n}$), where n is the time index of the last measured sample.

At every future time step $t > n$, the mean and variance of the output vector $\mathbf{y}_t$ can be calculated analytically.

4 EXPERIMENTAL RESULTS

The algorithm has been used to predict first passage time in the setup as described in section 2. Each feature is represented by a time series of 390 samples. The goal is to use the algorithm for online prognosis, which is done in the following way.

At time t, we estimate the parameters of the underlying stochastic model (Eq. 2) based on data window $y_{(t-Nw+1):t}$, where Nw is window length. As the condition of the machine will change over time, the values of the model parameters will change as well. Model parameters will determine the trend in the feature values, while noise covariance parameters comprise the influence of the varying noise component (which increases as the damage progresses).

The estimated model, obtained from each time window is used to predict future behavior, that is time T

when $y(T)$ crosses the alarm value y^* i.e. $y(T) \geq y^*$ for the first time.

The alarm value y^* has been set to values 7000 and 1100 for feature values of *vib8* and *vib3* respectively, which corresponds to the time of 60 hours after the start of the experiment.

4.1 Model structure

The underlying model is assumed to be of the following form:

$$\begin{aligned} x_1(t+1) &= a_{11}x_1(t) + a_{21}x_2(t) + w_1(t) \\ x_2(t+1) &= a_{22}x_2(t) + w_2(t) \\ y(t) &= c_1x_1(t) + e(t) \end{aligned} \quad (25)$$

or in matrix state-space form

$$\begin{aligned} \mathbf{x}_{t+1} &= \mathbf{A}\mathbf{x}_t + \mathbf{w}_t, \\ \mathbf{y}_t &= \mathbf{C}\mathbf{x}_t + \mathbf{e}_t, \end{aligned} \quad (26)$$

It is important to note, that the system states (x) in this model do not directly correspond to state of machine or gear health or have any physical meaning. The system states serve only to describe the dynamical behaviour of the feature values.

4.2 Online tracking of model parameters

Whenever a new measurement is obtained, the algorithm estimates model parameters using the last 100 samples. Initial parameters of the model are set to the following values:

$$\begin{aligned} \mathbf{A} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \\ \mathbf{Q} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \\ \mathbf{C} &= \begin{bmatrix} 1 & 0 \end{bmatrix} \\ \mathbf{R} &= \begin{bmatrix} 1 \end{bmatrix} \\ \mathbf{x}_0 &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ \mathbf{\Sigma}_0 &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \end{aligned} \quad (27)$$

Convergence of the EM procedure is achieved when the relative increase of the log likelihood function is less than 10^{-4}.

The algorithm has been first tested on the time series corresponding to the 8th vibration sensor measurements (c.f. *vib8* in Figure 2), which measures the vibrations in z direction on the output shaft. Because of the gear placement, the impacts between gear teeth are causing the strongest vibrations in this direction, therefore the feature values from this particular sensor are expected to be the most informative about the gear health.

Figure 6 shows the measured time series, the first eigenvalue of the estimated matrix $\mathbf{A}$ and diagonal elements of the estimated noise covariance matrix $\mathbf{Q}$.

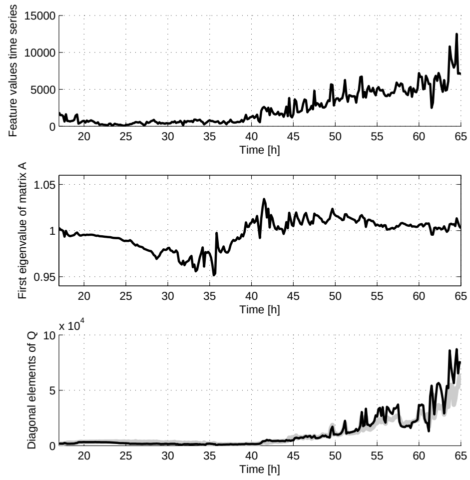


Figure 6: Values of the estimated parameters

Up to 40 hours of operation, it can be seen that the feature values are relatively small and there is no trend present. Consequently, the estimated system turns to be stable (eigenvalue less than 1). However, when the feature values start to increase, the system eigenvalues become greater than one. The increase in system noise parameters corresponds to stronger noise component present in the signal, which can be clearly seen from time series (between 40 and 65 hours).

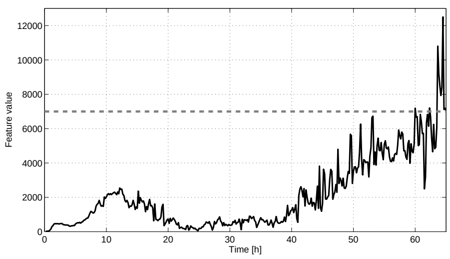
4.3 Prediction of the first passage time

The goal of our prediction is to determine the time, at which the feature will exceed the critical value. Our approach goes the following way. At a given data acquisition session (t), model parameters are estimated using windowed data ($y_{(t-Nw+1):t}$). Based on the estimated model, a Monte-Carlo simulation of the future feature trend is repeated 1000 times. Each simulation run results in a realization of the FPT. Based on all realizations a probability density function for the FPT is calculated, in particular its first and second moment. The results obtained for the feature that corresponds to the gear-frequency of the envelope spectrum of the vibration sensor (Figure 7).

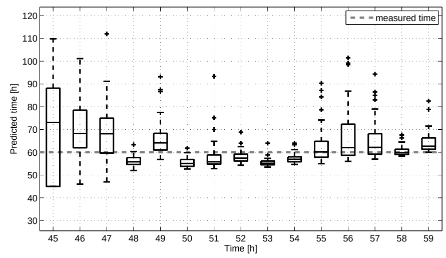
It can be seen, that the first estimates of the FPT's are made around 20 hours before the feature actually achieves the critical value. The estimation variance gradually decreases and 15 hours before the actual FPT the uncertainty of the predicted FPT's falls in the range ± 5 hours.

The validation of the concept has been performed on a different time series, in this case 3rd vibration sensor (c.f. *vib3* in Figure 2), the results of which are shown in Figure 8.

Again, the mean predicted time is approximately the same as the true time as early as 20 hours in advance. Due to higher noise variance in this time series, the variance of the predicted FPT is greater and is approximately ± 10 hours.

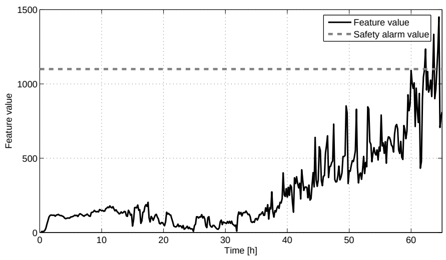


(a) Feature

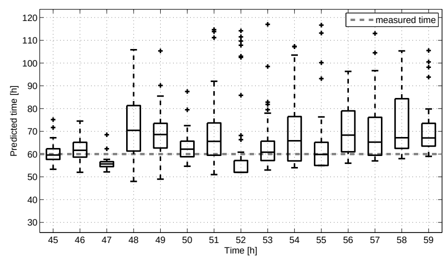


(b) Prediction

Figure 7: Results of MC analysis using *vib8* sensor data



(a) Feature



(b) Prediction

Figure 8: Results of MC analysis using *vib3* sensor data

5 CONCLUSION

It has been shown that the dynamic behavior of the vibration feature value can be approximated as the output of a second order dynamic linear stochastic state space model. Model parameters have been estimated using Expectation-Maximization algorithm, which iteratively maximizes the model likelihood function with respect to hidden states and to unknown parameters. The results show, that the model obtained in this way can effectively predict the future behavior of the feature and can therefore be used to predict the time of safety alarm. From current experiments, we estimate that using this method, the accurate prediction can be made 15 to 20 hours in advance. This offers the machine operators or maintenance a reasonable amount of time to replace the gear system without causing unnecessary production downtime.

In this stage, our algorithm has only been tested in the case of constant load. In real application, it is common that the load as well as rounds per minute are changing in time. Next step in the development of the procedure is to modify the algorithm so that it will include these as a measured system input. There are also several other issues that will have to be addressed for this approach to be used in industrial applications. One is the selection of the reference feature values, where the algorithm signals the alarm. Because there is no direct relation between our system states and the state of gear health, the alarm values for each setup will have to be determined experimentaly and in cooperation with system operators.

REFERENCES

(Dempster *et al.*, 1977) A. Dempster, N. Lard, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39:1–38, 1977.

(G. Niu, 2009) Bo-Suk Yang G. Niu. Dempstershafer regression for multi-step-ahead time-series prediction towards data-driven machinery prognosis. *Mechanical Systems and Signal Processing*, 23:740–751, 2009.

(Gibson and Ninness, 2005) S. Gibson and B. Ninness. Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica*, 41:1667–1682, 2005.

(Ho and Randall, 2000) D. Ho and R. B. Randall. Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mechanical Systems and Signal Processing*, 14:763–788, 2000.

(Howard, 1994) I. Howard. *A review of rolling element bearing vibration "Detection, Diagnosis and Prognosis"*. DSTO Aeronautical and Maritime Research Laboratory, 1994.

(Orchard and Vachtsevanos, 2009) M. E. Orchard and G. J. Vachtsevanos. A particle-filtering approachfor on-line fault diagnosis and failure prognosis. *Transactions of the Institute of Measurement and Control*, 31:221246, 2009.

(Rubini and Meneghetti, 2001) R. Rubini and U. Meneghetti. Application of the envelope and wavelet transform and analyses for the diagnosis of incipient faults in ball bearings. *Mechanical Systems and Signal Processing*, 15:287–302, 2001.

(W. Q. Wanga and Ismailb, 2004) M. F. Golnaraghib W. Q. Wanga and F. Ismailb. Prognosis of machine health condition using neuro-fuzzy systems. *Mechanical Systems and Signal Processing*, 18:813–831, 2004.

(Wang *et al.*, 2003) W. Q. Wang, M. F. Golnaragh, and F. Ismail. Prognosis of machine health condition using neuro-fuzzy systems. *Mechanical Systems and Signal Processing*, 18:813–831, 2003.