A hybrid prognostic model for multistep ahead prediction of machine condition

D Roulias¹, T H Loutas¹,² and V Kostopoulos¹
¹Applied Mechanics Laboratory, Department of Mechanical and Aeronautics engineering, Patras university campus, 265 04, Rio, Greece

E-mail: loutas@mech.upatras.gr

Abstract. Prognostics are the future trend in condition based maintenance. In the current framework a data driven prognostic model is developed. The typical procedure of developing such a model comprises a) the selection of features which correlate well with the gradual degradation of the machine and b) the training of a mathematical tool. In this work the data are taken from a laboratory scale single stage gearbox under multi-sensor monitoring. Tests monitoring the condition of the gear pair from healthy state until total brake down following several days of continuous operation were conducted. After basic pre-processing of the derived data, an indicator that correlated well with the gearbox condition was obtained. Consecutively the time series is split in few distinguishable time regions via an intelligent data clustering scheme. Each operating region is modelled with a feed-forward artificial neural network (FFANN) scheme. The performance of the proposed model is tested by applying the system to predict the machine degradation level on unseen data. The results show the plausibility and effectiveness of the model in following the trend of the timeseries even in the case that a sudden change occurs. Moreover the model shows ability to generalise for application in similar mechanical assets.

1. Introduction

Maintenance action for mechanical asset and more particularly Condition Based Maintenance (CBM) is an issue of high importance. CBM is steadily gaining ground in scientific literature [1]. In practice, sectors such as Department of Defence (DoD), air-transport and industry sectors are reluctant towards the application of CBM methods. Nonetheless CBM can substantially lower the maintenance cost in comparison to traditional methods as stated in [2]. One of the most interesting issues that arise in the context of CBM is prognostics. It is essential for the maintenance engineer, not only to pinpoint the current structural condition of the operating machine but also to know with some certainty the condition of the machine several time instants in the future. Rotating machinery prognostics in particular have gained the attention of researchers and engineers the last few years. There have been many efforts in addressing the latter problem. All, however, suffer from certain weaknesses. The basic issues relate to i) lack of accuracy in multi-step ahead prediction and ii) in tracking abrupt changes in the machine operating condition.

The current work is dedicated in developing a model that better adapts to the form of a prognostic feature timeseries and it is divided into six sections. In section 2, feature extraction takes place, along with certain preprocessing steps. Section 3 describes the prognostic system that was developed. The
system that was developed comprises of two steps, the definition of two very distinctive regions for the operating condition of the machine and then the training of a prognostics model for each of the two regions. In section 4, the experimental procedure is briefly described. The results are depicted in section 5. Finally, several conclusions are drawn and are summarised in section 6.

2. Feature extraction and preprocessing

2.1. Feature extraction
The first step towards the development of a reliable prognostics model is an appropriate feature selection. A good CBM candidate feature should have monotonic trend and low stochasticity. A highly stochastic parameter would yield very uncertain results.

In the current framework, offline visual inspection of a variety of sensor recordings took place. A plethora of extracted features from each sensor was evaluated in the current framework, among them rms, kurtosis and crest factor. Moreover, frequency related features were also extracted. More particularly the energy of certain frequency bands was extracted and the time evolution of the latter was evaluated.

2.2. Feature preprocessing
The aim of feature preprocessing was to “clean” the selected CBM feature from unwanted characteristics. This would result in an even more robust prediction model development for the particular operating asset. All preprocessing was implemented in such a way that can also be adapted for on-line data “cleaning”.

2.2.1. Run-in period disposal. In rotating machinery, vibration recordings are highly correlated with contextual parameters and most notably the machine temperature. The effect of temperature rise was neglected with the disposition of a certain time interval of the beginning of the experiment.

2.2.2. Outlier detection. Outlier detection is an issue for any robust time series model. There is no explicit definition of when a time series sample is an outlier or not. Generally, such a recording stems neither from the normal physical condition of the system under consideration nor contextual noise. It can be attributed to sensor problems, DAQ malfunctioning or some other event that has nothing to do with the operation of the asset. The outlier detection and cancelling that took place is based on [4] and is completely causal.

2.2.2.1. The decision threshold. The outlier decision threshold would be of statistical nature. Most often, the standard deviation of a moving window is defined as a good means to distinguish between normal and abnormal samples. This is not very robust though. In [4] a threshold based on the median absolute deviation is applied (equation (2.1)) normalised with 0.6745.

\[
MAD = \frac{\sum_{k=0}^{N-1} |x_k - d_N|}{N} 
\]  

(2.1)

Referring to equation (2.1), \(x_k\) is the \(k\)th sample of the timeseries window of width \(N\) and \(d_N\) is the median of the respective timeseries. If the next sample is larger than three or four times the normalised MAD then it is replaced by the median of the windowed timeseries else it is left untouched.

2.2.2.2. The window length. The length \(N\) of the time series window in equation (2.1) should also be defined. Too long a window wouldn’t take into account the local dynamics of the phenomenon. On the other hand, too short a window could allow outliers be accepted as valid recordings, especially in
cases that patches of outliers exist in the timeseries. The window should have length longer than the larger outlier patch.

2.2.3. Noise reduction. The timeseries can incorporate noise that may have an impact on timeseries model training. A simple causal low pass filter can be applied.

3. Prognostic model development
Prognostics is an important notion in CBM. Derived from the greek pro (prior) and gnosis (knowledge), it means the knowledge of the asset health condition after a certain operating time. Several issues should be resolved prior to adapting a mathematical model to the time series for prediction purposes. These are going to be briefly discussed in the following subsections.

3.1. Definition of operating regions on the timeseries data
The majority of scientific literature assumes that an operating machine can be characterised by a single prognostics model throughout its operating condition. Such an assumption narrows the potential of a condition monitoring system assuming the asset degradation mechanism uniform throughout the experiment. In most cases an unexpected event can accelerate the degradation procedure. The simplest way to distinguish between operating conditions is by defining fixed thresholds on the extracted feature value based on some evidence. More generic decision boundaries are possible to be defined via clustering methods.

3.2. Mixture of Gaussians
Gaussian mixture model (GMM) is an interesting clustering method. A brief description of GMM clustering follows. Most frequently, the experienced user defines the number of classes a priori. The cluster centres correspond to the mean of each distribution. The distribution covariance matrix is a measure of each cluster radius. Frequently an implementation of K-means algorithm can provide a first estimation for initial centre and covariance matrices. Afterwards, the expectation maximisation (EM) algorithm (presented in [5]) is implemented to adjust these parameters. At the end of the training procedure, a set of variances and means are defined. These will be linked to the various operating regions on the CBM feature timeseries.

3.3. Embedding dimension
Another parameter that needs to be defined for timeseries prognosis is the embedding dimension of the timeseries ([6]) that is the minimum time window that contains all the intrinsic information of a timeseries. This number will define the minimum number of past samples that are going to be fed in the Artificial Neural Network (ANN) for multistep prediction.

3.4. Feedforward Artificial Neural Network(FFANN) for prognosis
Finally, feed forward artificial neural network([8]) are applied. ANN have been shown ([9]) to outperform other timeseries modelling methodologies. An issue to be solved is pruning the ANN that is to define the number of hidden units. Too few hidden units would result in poor capture of the dynamics of the system. Too many units would risk fitting the random jitter of the time series. This would result in poor generalisation of the model.

4. Experimental procedure
The experiments were conducted in the context of [10]. Figure 1 depicts the experimental setup used for the gears testing. The test rig consists of two gears made from 045M15 steel with a module of 3 mm, pressure angle 20°, which have 53 and 25 teeth with 10 mm face width. The axes of the gears are supported by two ball bearings each. All the above are settled in an oil basin in order to ensure proper lubrication. The gear box is powered by a motor and consumes its power on a generator. Their characteristics are as follows:
- 1 stage gearbox with two gears (25 and 53 teeth)
- 3-phase 5 hp motor (220V, 9A, 50Hz, 1400 rpm) controlled by inverter
- Single phase generator with continuous power consumption control (4.2 KVA, 3000 rpm, 50Hz) with option for load fluctuation,
- The oil pump is of wet type without oil recirculation
- The shafts are ball bearing supported.

Figure 1. Figure 1a depicts the test rig under consideration. Figure 1b is a zoom of the gear box. The sensor positions are also shown.

The narrow teeth width combined with the relatively high applied load simulated accelerated fatigue conditions. This experiment was particularly designed for monitoring of the gradual gear teeth degradation. The experiment started with a healthy gear configuration. The experiment ran for about 160hrs. In the final visual inspection, both spur gears had suffered complete destruction with multiple teeth having been chopped off or completely worn out. No continuous visual inspection was possible and the monitoring of the machine was largely based on the sensor recordings. Periodic visual inspection was avoided due to the impact that start stop events have on the recordings. Three Bruel & Kjaer accelerometers were used for the vibration monitoring both mounted upon the gearbox case, one in each side-axis. The sampling frequency used was 50 kHz and recordings of 1 sec duration were obtained. The recording of all the above data is realized by a National Instruments NI-6070 1MS/SEC FIREWIRE data acquisition card and is assisted by special software in-house developed in Labview.

5. Results

5.1. Feature extraction and preprocessing
After the procedure described in section 2.1, a particular feature was kept for timeseries prognosis. This was derived from vibration sensor 1 (figure 1) recordings. This feature was the energy of 12.5Khz-18.75Khz vibration frequency band and will comprise the degradation feature. This degradation feature seemed of particular interest. First of all, the general increasing monotonic trend correlates well with the steady gradual degradation of the gear pair under consideration. Moreover, a jump of the timeseries value (after time sample 1050, figure 2b) is in fact an increase in the energy of the particular band and is attributed to a sudden change in the degradation processes, probably a tooth being worn out, that inserts high frequency resonances in vibration spectrum. Then a preprocessing based on section 2.2 was applied. The first 1.5 hour of operation was disposed off (see section 2.2.1). An outlier cleaning algorithm was also applied according to section 2.2.2. The outlier detection moving window length was defined as N=3 (according to section 2.2.2.2). The outlier detection threshold was defined as 4 times the normalised MAD (see section 2.2.2.1). The noise in the timeseries was reduced with the use of a simple low pass MA filter, more specifically the
{1/4, 1/4, 1/4, 1/4} filter. The result of the latter preprocessing is depicted in figure 2. The latter timeseries corresponds to continuous operation of the machine (no start-stop events in between). The full timeseries under consideration has a length of 1218 samples and corresponds to 101hrs of continuous operation (no start-stop event in between).

![Figure 2](image)

**Figure 2.** Figure 2a depicts the original extracted feature and 2b the feature after the preprocessing according to section 2.2.

5.2. Prognostic model

Referring to figure 2b, two operating regions can be distinguished. One that spans from the beginning of the experiment to at about time sample 1050 (“operating region 1/normal operation”) and another corresponding to the rest of the experiment (“operating region 2/substantial gear degradation”). The data were divided in half for model training and testing purposes respectively. Samples with even indices (N=0,2,4…) are kept for training and samples with odd indices (N=1,3,5…) are kept for testing the system resulting in 609 samples for training and an equal number of samples for testing the system.

5.2.1. Gaussian Mixture clustering. A gaussian mixture model was applied on the training data set. The number of clusters was defined equal to the number of the user defined operating regions. The training procedure returned a mean and a variance for each operating region. The mean feature amplitude for region 1 is 0.273 and the variance 0.0101. The mean feature amplitude for region 2 is 1.463 and the variance 0.1171. The mapping of the training data set in two regions according to the clustering model is depicted on figure 3.

![Figure 3](image)

**Figure 3.** The clustering results on training data set. Red crosses correspond to samples belonging to first class and green ‘x’ correspond to samples belonging to second class.
5.2.2. Defining embedding dimension. The embedding dimension is defined for each operating region (see figure 3) on the training data set. For this purpose the false nearest neighbour algorithm is applied ([6]). It was calculated by means of [7] as equal to 3 for both regions. It is reasonable to assume that 3 consequent samples \((x[n], x[n-1], x[n-2])\) is the minimum input vector for the data driven prognostics model that is to follow.

5.2.3. Feed Forward Artificial Neural Network (FFANN). Finally, a FFANN is adapted to each operating region. The configuration of each of the two FFANN is depicted in table 1. The number of hidden units was defined with few trial-and-errors. Two FFANNs were trained, one for each operating region. The input vectors were formed by triads of consecutive time samples. The input vectors for “FFANN region 1” comprised of triads of samples that were classified as region 1 samples, according to 5.2.1. The input vectors for “FFANN region 2” comprised of samples that were classified as region 2 samples except for the point where a change in operating condition took place. In the latter instance a few input vectors were allowed to be a mixture of “region 1” and “region 2” samples. The output vectors were of dimension 1 and were defined as being a certain number of time samples ahead of the respective input vector. Two cases were considered, 5 and 10 steps ahead prediction.

The input vectors and output vectors for training each FFANN were formed from the training data set. On the other hand, the testing data set was used for the testing of the FFANN model. The testing results from both FFANN were aggregated and a full outlook of the prognostic model was derived. An indicative efficiency metric for the FFANN training-testing procedure was opted to be the mean square error (MSE) between the actual and predicted timeseries. Table 1 depicts this MSE measured of course on the testing data set.

5.3. Discussion
In figure 4a,b the full results for the predicted timeseries are plotted along with the respective actual timeseries. The predictions were made on unseen data (input vectors from the testing set). Figure 5a depicts the results for 5 steps ahead prediction (50 minutes) and figure 4b depicts the results for 10 steps ahead prediction (1 hour and 40 minutes). The results are deemed satisfactory. The absolute error is also plotted for each case. This error is not negligible, especially for operating region “2” (figure 4a,b from sample 460 to sample 560). However, it seems that it is of the scale of the local stochasticity or local random jitter of the tested timeseries. This is evidence that the proposed model can generalise well. Moreover, the scale of the error term is not substantially changed when going from “5 steps” to “10 steps” prediction case, except for the region of change of operating mode. Figure 4c and 4d depicts a detail of figures 4a,b respectively. The latter gives a detailed view of the point where the operating region changes. The prediction model follows quite well the jump from operating condition 1 to operating condition 2 for both cases under consideration. Figure 5 summarises the proposed prediction scheme for on line application.
### Table 1. FFANN configuration.

| ANN parameter                                      | “5 steps ahead”                          | “10 steps ahead”       |
|----------------------------------------------------|------------------------------------------|------------------------|
| Inputs (according to section 5.2.2)                | 3 consequent time samples                | -α-                    |
| Number of outputs                                  | 1                                        | -α-                    |
| Number of hidden units                             | 3, for region “1”                         | -α-                    |
|                                                | 1, for region “2”                         | -α-                    |
| Hidden layer transfer function                     | tanh                                     | -α-                    |
| Output layer transfer function                     | linear                                   | -α-                    |
| ANN Initialisation algorithm                       | Nguyen Widrow                            | -α-                    |
| ANN Optimisation algorithm                         | Levenburg-Marquadt                       | -α-                    |
| Mean square error for ANN                          | 4*10^-4                                  | 7*10^-4                |
| of operating region “1”                            | (~200 iterations)                        | (~150 iterations)      |
| Mean square error for ANN                          | 0.0157                                   | 0.017                  |
| of operating region “2”                            | (~200 iterations)                        | (~300 iterations)      |

**Figure 4.** Testing data set (green line), predicted data set (red line), absolute error between the latter (blue line). Figure 4a depicts results from the 5 step (50 minutes) ahead prediction and 4b from the 10 step (100 minutes) ahead prediction case. Figure 4c depicts a detail from figure 4a at the operating region change point. Figure 4d depicts a detail from figure 4b at the operating region change point.
6. Conclusions
In the current work a hybrid, on-line, prognostics model was designed (figure 5). GMMs along with FFANN are applied in order to derive a robust multistep ahead prediction model for condition monitoring. Feature pre-processing was deemed useful for even more robust prognostics and it was designed in such a way that it could be applied to on-line CBM applications. The model was tested on unseen data. The results are deemed satisfactory. The model could very well adapt to the indicator’s steady trend as well as sudden changes in the operating condition. At the same time, overfitting seems to have been avoided and may generalise well to similar mechanical assets.

Acknowledgments
This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

References
[1] Jardine A K S, Lin D and Banjevic D 2006, Mechanical Systems and Signal Processing, 20 1483-510
[2] Byington C S and Garga A K 2001 Handbook of multisensory data fusion, CRC Press, Boca Raton, FL, D. L. Hall and Linas, eds, pp 23-1-23-31
[3] Menold P H, Pearson R K and Allgower F. 1999 Proceedings of the 7th Mediterranean Conference on Control and Automation (MED99) Haifa, Israel - June 28-30
[4] Dempster AP, Laird NM and Rubin D B 1977 Journal of the Royal Statistical Society Series B (Methodological) 39 1–38
[5] Sprott J C 2003 Oxford University Press
[6] Matlab code from http://ideas.repec.org/c/boc/bocode/t741510.html
[7] Bishop C.M. (1995), Oxford University Press
[8] Tse PW and Atherton DP 1999 Journal of vibration and acoustics 121, 355-62
[9] Nguyen D and Widrow B 1990 Proceedings of the International Joint Conference on Neural Networks, 3, 21-26
[10] Loutas T H, Roulias D, Pauly E. and Kostopoulos V 2011, Mechanical Systems and Signal Processing, 25 1339-52