Implementation of diagnostics, prognostics and e-maintenance support under variable operating conditions

Đani Juričić[1], Pavle Boškoski[1], Janko Petrovčič[1] and Bojan Musizza[1]

[1]Department of Systems and Control, Jožef Stefan Institute
Jamova 39, Ljubljana, Slovenia
{dani.juricic, pavle.boskoski, janko.petrovcic, bojan.musizza}@ijs.si

Abstract

Efficient maintenance of critical industrial asset requires continuous condition assessment and estimation of the remaining useful life of the components in order to guarantee timely fault accommodation and avoid unanticipated breakdowns and equipment failures. The paper presents a concept of a system for on-line diagnostics and prognostics of asset condition whose aim is simple implementation, low cost and high level of portability. The key problem addressed arises from reality when operating conditions are not fully measurable or are unavailable as well as prior information concernig asset might be incomplete. The diagnostic approach, outlined in the first part of the paper, is probabilistic and relies on calculating distances between distributions of the wavelet coefficients calculated from vibrations signals. The second part of the paper presents a platform for implementation. It builds on a distributed sensor network which allows for signal acquisition from sensors, local signal processing and fusion of the diagnostic results on a central server. Key to the concept is comprehensive data storage and data manipulation framework with built-in condition monitoring taxonomy as well as allowing seamless integration with existing manufacturing execution systems. To address all these issues the potential of the MIMOSA OSA-EAI standard are exploited and a prototype is implemented on a large milling machine.

1 Introduction

Stable and predictable condition of process equipment, high process availability, product quality, and reliability are key factors that keep companies competitive in the market. However, wear, material stress and environmental influences cause equipment to fail. Unexpected failures can result in partial or total breakdown of a production line, destroyed equipment and even catastrophes.

Except of catastrophic failures, which are sudden and cause total loss of functionality, the 99% of mechanical failures go through a distinct incipient phase. This means there are some noticeable indicators, which provide advanced warning about onset of failure. The role of automated condition monitoring (CM) is to timely detect this onset and localize the root-cause. Though useful, the information about current condition is not enough. What the operators and mainteance people want to know is when to stop a machine and what are actions to take. Good estimate of the remaining useful life is indispensable.

The discipline that links the detection of failure mechanisms to system life cycle management is referred to as Prognostics & Health Management (PHM). PHM uses sensor records and prior information to allow early detection of incipient faults, remaining useful life calculations to optimize the maintenance programme.

In the last several years we have witnessed significant advances in on-line condition monitoring accompanied by emerging general purpose commercial systems. However, their applicability has been restricted mainly to the critical equipment in safety sensitive applications, power engineering and transport. Unfortunately, no massive use in industrial practice has occured so far. There are several reasons for that: (i) yet unclear the return of investment, especially when domain specific solutions have to be adopted, (ii) the approaches, mostly adequate for stationary applications, have to be supported by additional instrumentation to cope with non-stationary operating conditions, (iii) collecting of all the data relevant for determination of the operating condition, like for example work orders or maintenance reports may not be trivial to obtain.

Compared to CM, PHM is much more difficult. Limited success has been achieved in special cases like aeronautics and defence systens. The problem is notoriously demanding for many reasons: (i) data about overall

useful life from similar items of equipment are seldom available, (ii) lack of knowledge about degradation mechanisms and (iii) incomplete knowledge of operating history, disturbances and past maintenance actions.

In this paper we will address the problem of condition monitoring and prognostics in variable operating conditions. For the case of unknown mild fluctuations of the operating conditions, the use of entropy indices results in approaches, which are to a certain extent insensitive to the lack of information about variations in operating conditions. This brings significant practical advantages in the sense that diagnostics and prognostics can be done without additional sensors.

The outline of the paper is the following. The motivation for the work is briefly explained in section 2. Section 3 outlines the idea of robust diagnosis based on entropy indices. Robust prognostics using the related concepts is presented in section 4. Section 5 provides main ideas behind implementation.

2 Motivation and a statistical framework

The traditional vibration-based diagnostic approaches for rotational machines are largely designed for stationary and known operating conditions. Consequently, any change in the vibrational patterns can be directly associated with a change in the condition of the monitored machine. In case of fluctuating operating conditions, these associations become ambiguous, because changes in the operational regime usually influence the generated vibrational patterns. Therefore, reliable fault detection and isolation, irrespective of variable and presumably unknown operating conditions, is of significant practical merit.

The problem of fault detection under fluctuating load and speed has received commendable attention so far. Generally, the proposed solutions rely on precise information about current operating conditions. Usually, the information regarding operating conditions is directly used in the process of calculating feature values. Order tracking is such a method, capable of coping with speed variations by analysing signals in angular domain [22]. Also Bartelmus and Zimroz [1] successfully performed fault detection in multi-stage gearboxes by taking into account information about variations in speed and load. Some methods that exploit the statistical properties of the acquired vibrations like instantaneous power spectrum (IPS) Although the proposed solutions give satisfactory results, they heavily depend on accurate measurements of the speed and load of the monitored drives. In [4] the detection of gear deterioration under different loads is performed using instantaneous power spectrum by employing Wigner-Ville distribution (WVD). The authors successfully performed fault detection of gear faults irrespective of the variable operating conditions. Similarly, [7] combined Hilbert and wavelet packet transforms for detection of gear faults with the ability to handle non-stationary signals.

Can gear and bearing faults be reliably detected through vibration analysis in spite of unknown and variable operating conditions? Can the same question be answered in the context of prognostics?

In this paper we try to communicate (hopefully) encouraging results based on the concept of robust diagnostics and prognostics is illustrated in Figure 1. The concept builds on viewing the sensory outputs as realizations of stochastic signals. To pick up the essential aspects of the signal related to the system condition, the signal is subjected to decomposition by means of the Wavelet Packet Transform (WPT). Instead of pinpointing particular single wavelet projections, the distribution of coefficients on a given time and frequency interval is characterized by some entropy measure. Based on these measures [8] devised a fault detection system for surface bearing faults. They showed that both fault detection and fault isolation of different bearing faults can be successfully conducted by applying features extracted from the entropy measures. Despite the promising results,[8] did not provide accurate explanation of the mechanism that made these measures suitable for fault detection of rotational machines. In our approach the divergence is used to quantify the difference between actual distribution of the WT coefficients and the reference distribution captured in the fault-free case. Hence the detection can be performed.

The problems of prognostics attracted a lot of attention in the past years. Majority of the proposed approaches try to relate the relationship between the defect growth and the time evolution of some statistical characteristic of the generated vibrations like energy, peak-to-peak values, RMS, kurtosis, crest factor, etc. [12, 13]. Usually, the selected statistical characteristics are calculated on a specific frequency bands of the generated vibrations and their ratios are used as features for estimating the bearing's RUL.

To perform prognostics the concepts of entropy and divergence need to be complemented with the concept of complexity. Besides statistical dissimilarity between the reference distribution of the WPT coefficients and its current version, caused by changed inner condition of the drive, also the patterns of the waveforms change. The

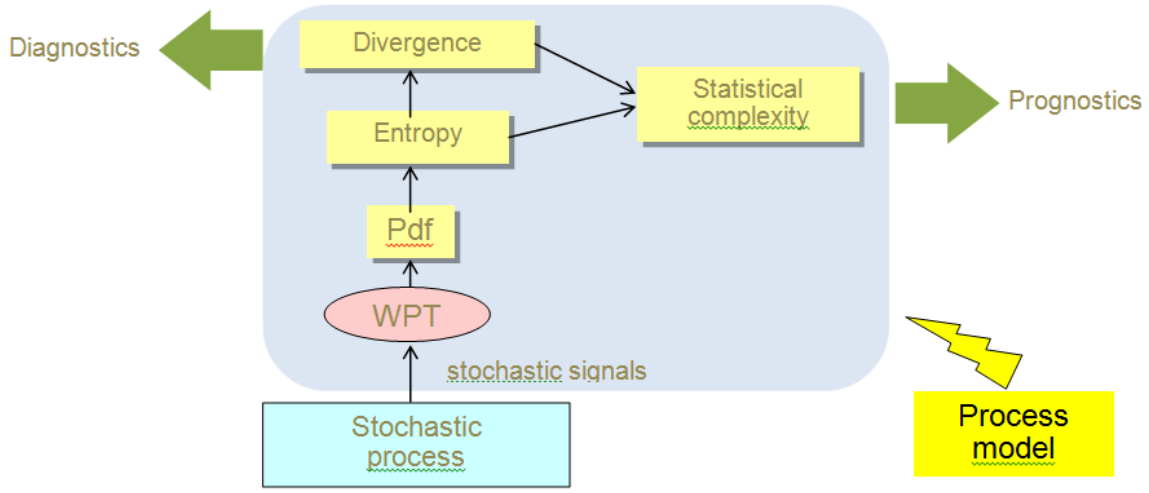

Figure 1: The concept of diagnostics and prognostics of rotational machines with time varying and partially known operating conditions

complexity of the waveform is related to its information content. In that respect one can define two extremes, i.e. periodic and purely random signals. Both bear low complexity, the former due to easily predictable repetitive pattern and the latter due to loss of any predictability, which implies no information in the waveform. The complexity of all other waveforms is somewhere in between. It is shown in [9] that statistical complexity and entropy Entropy and complexity define the system life cycle from nominal fault-free state to the failure state defining the end of life.

3 Robust diagnosis

Vibrations in rotational drives originate from complex interactions between rotating parts and excitation of the structural eigenmodes. Under constant operating conditions these interactions demonstrate almost periodic repetitive pattern. Therefore, the most natural approach to fault detection is to look for the changes in the set of spectral components that are related to such repetitive events.

However, the presence of unknown and possibly variable operating conditions makes fault detection and isolation very difficult. However, this situation is very realistic in many applications. In such a case the generated vibration signals are essentially non-stationary and the traditional methods designed for stationary processes are not applicable.

One way to alleviate these difficulties is to observe the signal's instantaneous power (IP). The instantaneous power is directly related to the signal's envelope, cf. [5], which is state-of-the art approach to fault detection of rotational machines operating under constant operating conditions. In what follows we investigate the distribution of IP of the vibration signals generated by gearbox drives operating under variable loads and speeds. Since a vibrational signal generated by running gears and damaged bearings can be regarded as the sum of sine signals with random phase and amplitude we will limit our analysis only to non-stationary signals that belong to this class. In particular, the aim is to analyse the distribution of the IP in the presence of faults and under changes in operating conditions.

3.1 Model of gears vibrations

Vibrations produced by a healthy gearbox are dominated by vibrational components due to the meshing gears. The main source of these vibrations is alteration of the mesh stiffness that occurs as a result of variation of the number of gear teeth in contact. Consequently, under constant rotational speed, the resulting vibrations are characterized by spectral components located at the mesh frequency and its higher harmonics. Additionally,

these components are accompanied by a modulation that originates from assembly errors and fluctuations in gears speed and load. To sum up, vibrations $y(t)$ produced by a pair of meshing gears rotating with constant speed can be represented as follows [25]:

$$y(t) = \sum_{m=1}^{M} A_m\big(1 + a_m(t)\big)\cos\big(2\pi m f_{gmf} t + \beta_m + b_m(t)\big), \tag{1}$$

where f_{gmf} is the gear mesh frequency, M is the highest harmonic of interest and A_m is the amplitude of the m^{th} multiple of the gear mesh frequency. The components $b_m(t)$ represent phase modulations, and β_m is the initial phase. The amplitude modulations $a_m(t)$ occur as a result of incipient localized tooth damages, i.e. surface faults that occur only on limited number of gear teeth. These modulations contain all the necessary diagnostic information.

3.2 Model of vibration signals produced by localized bearing faults

Healthy bearings produce negligible vibrations. However, in the case of surface damage, vibrations are generated by rolling elements passing across the damaged site on the surface. Each time this happens, impact between the passing ball and the damaged site triggers an impulse response of the structure $s(t)$. Under constant operating conditions impacts may be considered as truly periodic, with a period directly related to the type and location of the surface fault. However, under variable operating conditions the underlying impacts should be treated as purely random events. Furthermore, the amplitude of each impulse response also differs, due to changes in the surfaces and the microscopic variation of the way each roller element enters the load zone. Fluctuations in the amplitudes and period of occurrence of impulse response were modelled by [21] as:

$$y(t) = \sum_{i=-\infty}^{+\infty} A_i s(t - \tau_i) + n(t), \tag{2}$$

where A_i is the impulse of force that excites the entire structure and τ_i is the time of its occurrence. The final component $n(t)$ defines an additive random component that contains all non-modelled vibrations as well as environmental disturbances.

3.3 Unified signal model

The models (2) and (1)can be represented in a more general way, which is suitable for representation of nonstationary signals. The theory of evolutionary processes developed by Priestley [20] provides an elegant framework to cope with such signals resulting in the following representation

$$y(t) = \int_{-\infty}^{\infty} A(\omega, t) e^{j\omega t} dZ(\omega), \tag{3}$$

where $A(\omega,t)$ is a slow varying time and frequency dependent modulating function, and $Z(\omega)$ is a complex random process with orthogonal increments [5]. It is shown in [5] that the probability density function of the instantaneous power (IP) of the sum of random sine waves takes a closed form solution.

3.4 Wavelet Packet Transform

In reality, we have at disposal vibrational signal which is just a single realization of the sum of random sinusoids. an elegant way to estimate the pdf of the instantaneous power is by means of the coefficients of the wavelet analysis.

The wavelet analysis is based on a family of wavelets created by dilatation s and translation u of a mother wavelet $\psi(t)$ [15]

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-u}{s}\right). \tag{4}$$

Consequently, the wavelet transform $Wf(u,s)$ can be calculated using the inner product:

$$Wf(u,s) = \langle f(t), \psi_{u,s}(t)\rangle. \tag{5}$$

If we limit the values for the parameters s and u to a set of discrete values $s_j = 2^{-j}$ and $u_{j,k} = 2^{-j}k$, where $j,k \in \mathbb{Z}$, the mother wavelet (4) becomes:

$$\psi_{j,k}(t) = 2^{j/2}\psi\left(2^j t - k\right),\ j,k \in \mathbb{Z}. \tag{6}$$

The discrete wavelet analysis can be implemented using a set of discrete conjugate mirror filters able to transform an existing orthonormal basis into two orthogonal families each covering different frequency intervals. Hence we get the so-called binary wavelet packet transform (WPT), schematically shown with a wavelet packet tree in Figure 2. The nodes of the tree are marked by (d,n), where $d = \{1,\cdots,D_0\}$ represents the depth of the tree, and $n = 0,\cdots,2^d-1$ represents the number of the node at depth d. Each of the node defines a space with orthonormal basis $\psi_d^n(t-2^d k)$, $k \in \mathbb{Z}$, and contains N_d wavelet packet coefficients $\mathbf{W}_{d,n,t}$, $t = 0,\cdots,N_d-1$.

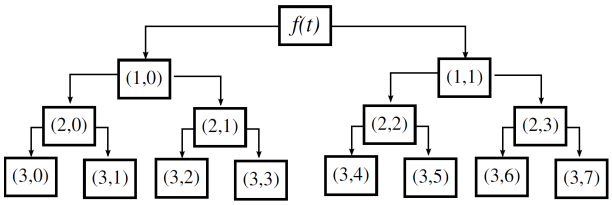

Figure 2: Example of a full WPT tree with depth $D_0 = 3$

Since the wavelet coefficients for the related node span a specific frequency band, the decomposition will give the power fluctuations contained within a specific set of frequencies [5].

3.5 The distribution of wavelet coefficients

The coefficients of each WPT node comprise information about original signal components that belong to a sufficiently narrow frequency band. By calculating the instantaneous power for each WPT node we can estimate the probability density that corresponds to the envelope of spectral components contained in that particular node.

Each of the n nodes at level d contains N_d wavelet coefficients $W_{d,n,t}$ $t = 0,\ldots,N_d-1$, $N_d = 2^{-d}N_s$, N_s is the sample length of the signal. Using these coefficients, the portion of the signal's energy $E_{d,n}$ contained within one node (d,n) reads [3]:

$$E_{d,n} = \sum_{t=0}^{N_d-1} \|W_{d,n,t}\|^2. \tag{7}$$

The total signal's energy can be obtained by summing the energy contained within the set of terminal nodes T:

$$E_{tot} = \sum_{\substack{t=0 \\ d,n\in T}}^{N_d-1} \|W_{d,n,t}\|^2 = \sum_{d,n\in T} E_{d,n}. \tag{8}$$

The set $\mathcal{P}^{d,n}$ expresses the contribution of each wavelet coefficient to the energy of the signal within the terminal node (d,n):

$$\mathcal{P}^{d,n} = \left\{ p_t^{d,n} = \frac{\|W_{d,n,t}\|^2}{E_{d,n}}, t = 0,\cdots,N_d-1 \right\}. \tag{9}$$

A similar set can be defined for the contribution of the energy of each terminal node $(d,n) \in T$ in the total energy of the signal E_{tot}:

$$\mathcal{P}^T = \left\{ p_{d,n} = \frac{E_{d,n}}{E_{tot}},\ d,n \in T \right\}. \tag{10}$$

The elements contained in both sets $\mathcal{P}^{d,n}$ and $\mathcal{P}^T$ can be treated as realisation of a random process. Based on these realisations one can estimate the corresponding probability distributions and calculate their entropies and statistical complexity.

3.6 Entropy

Entropy is a means to characterize the probability distributions. For a discrete probability distribution $\mathcal{P} = \{p_1, p_2, \ldots, p_N\}$, the simplest definition of entropy is the one according to Shannon:

$$H(\mathcal{P}) = -\sum_{p \in \mathcal{P}} p \ln(p). \tag{11}$$

For a discrete set with cardinality N Shannon entropy can reach values between 0 and $N \ln N$. A problem with the Shannon entropy is that it is relatively insensitive to the changes in the tails of the distribution. In many cases, faults in the monitored system affect the tails. An update of the Shannon entropy sensitive to the tails is Rényi entropy [2]

$$H_\alpha(\mathcal{P}) = \frac{1}{1-\alpha} \ln \sum_{p \in \mathcal{P}} p^\alpha(x),\ \alpha \geq 0\, \alpha \neq 1. \tag{12}$$

Rényi entropy contains the parameter α, which serves to tune the sensitivity of the entropy towards particular segments of the probability distribution $\mathcal{P}$.

3.7 Jensen-Rényi divergence

Divergence is a concept which is helpful in expressing the dissimilarities (or "distance") between the distribution functions. The Jensen-Rényi divergence between two distribution functions $\mathcal{P}$ and $\mathcal{Q}$ defined on the same set is [2]:

$$D_\alpha^w(\mathcal{P}, \mathcal{Q}) = H_\alpha\left(w\mathcal{P} + (1-w)\mathcal{Q}\right) - \left\{wH_\alpha(\mathcal{P}) + (1-w)H_\alpha(\mathcal{Q})\right\}, \tag{13}$$

where $w \geq 0$. The values of the exponent α governs the sensitivity of these two quantifiers to particular segments of the pdf, i.e. it specifies the relative importance of small values versus large values of the probability mass.

3.8 Fault diagnosis based on divergence

The idea of monitoring the condition of a rotational drive is illustrated in Figure 3. At the beginning of the monitoring process the reference condition should be defined by computing the values of $\mathcal{P}^{d,n}$ and $\mathcal{P}^T$ for all relevant nodes. In the course of time, the values are calculated on a short segment of signal. If the bearing's condition is normal, no significant difference between the two distributions should exist. A fault in the system can cause changes in the distribution of the particular node at hand, hence altering the corresponding values of $D_\alpha^w(\mathcal{P}, \mathcal{P}_e)$ where $\mathcal{P}_e$ refers to the nominal state. Such an approach can be used as means to detect and, in some cases, to isolate a fault.

It is important to stress that the window length is usually very short and the operating conditions within the node can therefore be assumed constant. If the speed actually varies, the spectral content will also move along the frequency axis. In spite of that, the distribution pattern associated with the WPT will not change much as the shifted harmonics are still within the specific frequency band associated to the particular node. However, if a change in the operating speed is severe enough, it might happen that the frequency content from one node moves to the adjacent node, thus fooling entirely the diagnostic reasoning. In the case of variations in the load, mild variations normally have no significant impact on the frequency distribution pattern. Furthermore, even in the case of significantly increased load, additional sideband components might occur but without any major impact on the energy distribution within a node.

3.9 An example

A two-stage gearbox operating under different speeds and loads with various combinations of embedded gear and bearing faults is used. The experimental data are taken from PHM benchmark [18], which consists of a laboratory two-stage gearbox. The test runs include 7 different fault combinations and one fault-free reference run. Each set of faults was tested under 5 different rotational speeds of the input shaft: 30, 35, 40, 45 and 50 Hz. Additionally, two different load levels were applied. The detailed list of the introduced faults is provided in Table 1. The signals were sampled with sampling frequency $f_s = 66.6$ kHz and the sampling horizon was 4 seconds long.

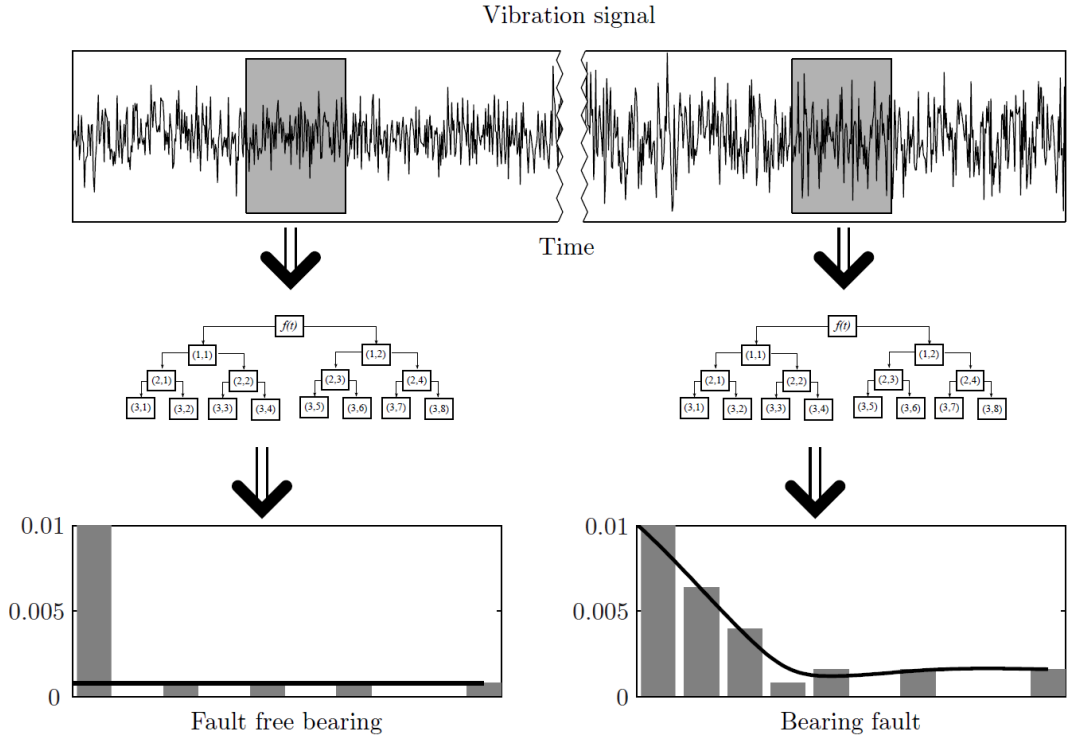

Figure 3: The diagnostic scheme.

The results of experiments with mixed gear and bearing faults will be briefly discussed. Out of 16 nodes in the WPT tree, Figure 4 shows the Jensen-Rényi divergence for the 1^{st} and the 13^{th} node since these nodes exhibit the highest sensitivity to the gear and bearing faults.

It can be seen that the Jensen-Rényi divergence for the fault-free cases #1 is negligible, regardless under which operating conditions the distributions of $p_{d,n}(t)$ were calculated. The distances calculated for the 1^{st} WPT node, and shown in Figure 4, indicate that cases #2–#5 exhibit the biggest distance from the fault-free case #1. As the 1^{st} node contains the low-frequency part of the signal, these changes can be attributed to gear faults. The cases with bearing faults #6–#8 have similar values as the fault-free case, as their influence is more expressed in the high-frequency band. This effect is visible in the 13^{th} node, shown in Figure 4. For this node the cases containing only gear faults #2–#3 show no changes, whereas all the other cases #4–#8 exhibit significant departure from the fault-free signal. It should be noticed that runs #4 and #5 show increases in both WPT nodes, since they contain both gear and bearing faults.

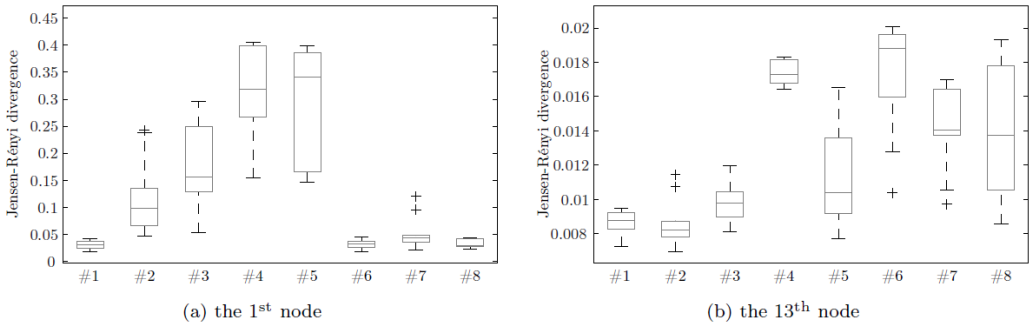

Figure 4: Jensen-Rényi divergence between the datasets taken from the fault-free machine and datasets obtained under various faults explained in Table 1. All experiments were performed under high load.

	Gear				Bearing[1]			Shaft
	1	2	3	4	1	2	3	
#1	Fault Free (FF)							
#2	Chipped	FF	Eccentric	FF	Fault Free (FF)			FF
#3	FF	FF	Eccentric	FF	Fault Free (FF)			FF
#4	FF	FF	Eccentric	Broken	Inner	Ball	FF	FF
#5	Chipped	FF	Eccentric	Broken	Inner	Ball	Outer	FF
#6	FF	FF	FF	Broken	Inner	Ball	Outer	Imbalance
#7	FF	FF	FF	FF	Inner	FF	FF	Keyway Sheared
#8	FF	FF	FF	FF	FF	Ball	Outer	Imbalance

[1] Faults were introduced only on Bearings 1–3. The other three bearings were kept fault-free during all experimental runs.

Table 1: Fault details for experimental runs.

4 Robust data-driven prognostics

Prognostics is perhaps the most crucial and in the same time the hardest ingredient of PHM systems. Everyone involved in the maintenance schedulling would highly welcome an accurate estimate of the time beyond which the machine will not operate safely or will not operate at all. The existing methods for predicting the end of life in rotating machinery can be grouped into three major concepts [24]

- approaches based on event data records,
- approaches relying on condition data and
- combination of both approaches.

The approaches belonging to the *first group* rely on accumulated data records from large number of the same items of equipment or on large set of repeated experiments in specified laboratory conditions. Time-to-fail estimates and other failure probability rates can be described by statistical failure models like Weibull or log-normal distribution [24].

The *second group* is rich with approaches that can roughly be divided into physics-based models and data-based models. The physics-based models rely on detailed physical modeling by means of finite element method, which serves to compute spatial distributions of stresses in the material [6]. The progression of local cracks in the material is computed by locally applying the Paris law [23].

A number of authors suggested use of neural networks as an efficient time series prediction tool. For example, [26] apply neural network to predict crack in rolling bearing by on-line adaptation of the network parameters. Unfortunately, only one-step ahead prediction is achieved. Problems might also occur when the training data set is too short. Some authors have also suggest the use of nonlinear state filtering and prediction based on various particle filters strategies. Their idea is to calculate the probability density function of the predicted life times, based on a dynamic model with known parameter values [10].

The *combined approaches* tend to make efficient use of reliability data in the prognostic models. Final assessment of the remaining useful life results from comprehensive treatment of the current condition and historical records.

In the above approaches it is assumed that the operating conditions are measurable. The question is what can be done if operating conditions are only partially known or unknown? Fortunately, the robust framework validated in the diagnostic context, can be extended for the prognostic purposes. The important ingredient is *signal complexity*.

4.1 Signal complexity

The statistical complexity $\mathcal{C}(\mathcal{P})$ of a signal with distribution $\mathcal{P}$ based on (12) and (13) is defined as [14]:

$$\mathcal{C}(\mathcal{P}) = Q_0 D_\alpha^w(\mathcal{P}, \mathcal{P}_e) H_\alpha(\mathcal{P}), \tag{14}$$

where $\mathcal{P}_e$ is the uniform distribution and Q_0 is a normalisation constant so that $Q_0 D_\alpha^w(\mathcal{P},\mathcal{P}_e) \in [0,1]$. The product (14) is in accordance with the initial idea that signals with perfect order $H_\alpha(\mathcal{P}) = 0$ and maximal disorder $D_\alpha^w(\mathcal{P},\mathcal{P}_e) = 0$ have the lowest complexity.

The concept of time is present in (14) indirectly through the entropy $H_\alpha(\mathcal{P})$ by using the fact that the system's entropy increases in time. An interesting perspective of system life is obtained if the statistical complexity $\mathcal{C}(\mathcal{P})$ is plotted versus the entropy $H_\alpha(\mathcal{P})$ [14]. Figure 5 shows the life cycle of measured on a real gearbox. In the beginning the random component in the vibratiton signal is strong resulting in low near-zero complexity and high entropy (note the non-monotoneous behavior). The evolution of the statistical complexity is directly related to the nature of the observed system. Therefore, by trending the evolution of the statistical complexity within the pre-defined area one can perform the prognostics task.

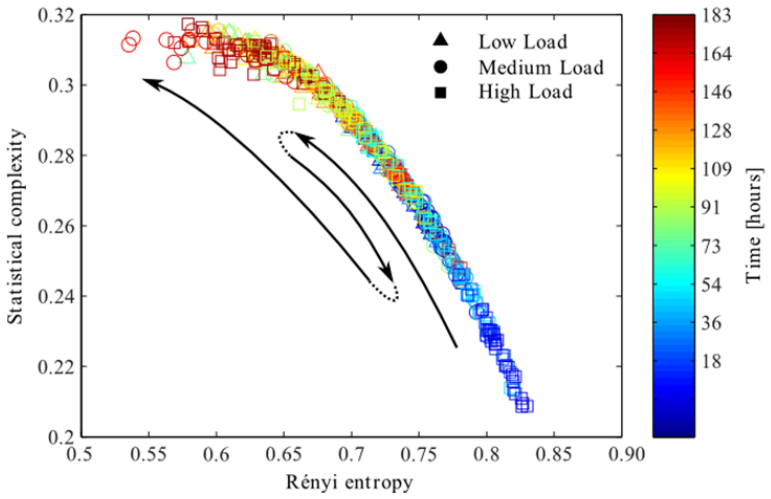

Figure 5: The life cycle of a gearbox viewed through the complexity-entropy plot.

4.2 Modelling evolution of statistical indices

The problem with the plot in Figure 5 is that time does not appear explicitly.The rationale behind is that similar components are likely to have similar life cycles accompanied by similar constellation of entropy indices. In other words, the indices $\mathcal{C}(\mathcal{P})$, $D_\alpha^w(\mathcal{P},\mathcal{P}_e)$ and $H_\alpha(\mathcal{P})$ can be viewed as a stochastic dynamic process dependent on the remaining useful life RUL. In the proposed setup the nonparametric Gaussian Process Models (GP) are applied to describe this process. A GP model [11] is a mapping from a vector of regressors into the pdf of the related model output. This pdf is Gaussian, hence the name. Final RUL prediction is done by Bayesian averaging, i.e. by fusing the partial predictions into the integral pdf.

The main idea behind the approach, originally suggested in [9], is highlighted in the sequel and illustrated on an example of bearing RUL prognostics.

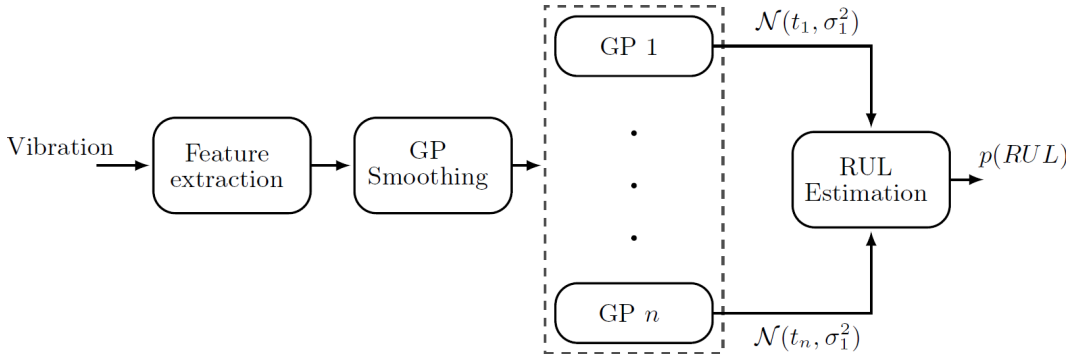

Figure 6: Schematic representation of the prognostic algorithm.

A set of repeated life-long runs on similar items of equipment has to be at disposal. For each run a sequence

of short vibrations records need to be taken repeatedly. Each record serves to calculate the statistical indices and also the time to failure. Hence each life-long run results in a time series with entropy indices as input vector and RUL as output. These time series constitute the training dataset. To cope with the fact that the duration of the training datasets varies, the actual experiment time t was normalized by the overall life-cycle time so that the relative time index τ_i is used. Hence $\tau_i = 0$ means the begining of life, while $\tau_i = 1$ denotes the end of life. An example of a normalized training set of Jensen-Rényi divergence is shown in Figure 7.

The result of the training process is a GP model which defines the evolution of the feature value for each $\tau_i \in [0,1]$. Given the input value of τ_i, the output of the GP model is a normal distribution describing the pdf of the feature value at the relative time τ_i. The training process can be described as:

$$GP : \mathcal{N}(\mu_i, \sigma_i^2 | \tau_i), \text{ where } \tau_i \in [0,1] \text{ and } i \in \mathbb{N}. \tag{15}$$

The mean values μ_i are shown with thick line in Figure 7.

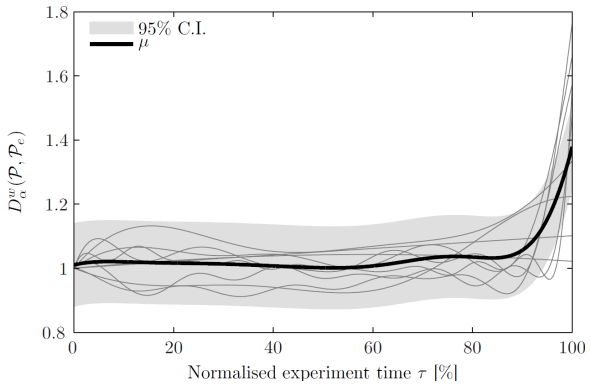

Figure 7: Time evolution of $D_\alpha^w(\mathcal{P}, \mathcal{P}_e)$ normed in the interval $[0,1]$.

4.3 RUL estimation

Having calculated the probability density functions of the particular entropy index, the evolution of RUL in dependance of the vector of indices is calculated by computing the posterior distribution of the relative time τ_i. As the training data points are normalised and lie on the interval $\tau_i \in [0,1]$, the RUL is simply $1 - \tau_i$. The posterior pdf of the distribution $p(\tau_i)$ is computed from the feature value $D_\alpha^w(\mathcal{P}_t, \mathcal{P}_e)$ at the time instant t from Bayes' rule

$$P(\tau_i | D_\alpha^w(\mathcal{P}_t, \mathcal{P}_e)) \propto P(D_\alpha^w(\mathcal{P}_t, \mathcal{P}_e) | \tau_i) P(\tau_i), \tag{16}$$

where the likelihood $P(D_\alpha^w(\mathcal{P}_t, \mathcal{P}_e)|\tau_i)$ is given by the GP model (15) and the prior $P(\tau_i)$ in (16) can either include additional knowledge related to the RUL distribution or can be set to an uninformative distribution.

If the informative prior is used in (16), the distribution $P(\tau_i)$ has to satisfy two main criteria. Firstly, it has to be conditioned on the current experiment duration t and secondly, it should be designed in a way that will give more weight to the prior at the beginning and more weight on the measurements, once they become significant. For this purpose, we propose the truncated normal distribution $T\mathcal{N}\left(\mu, \sigma^2\right)$ with pdf given as:

$$p(\tau) = \frac{1}{\sqrt{2\pi\sigma^2}} \frac{\exp\left(\dfrac{-(\tau-\mu)^2}{2\sigma^2}\right)}{\Phi\left(\dfrac{b-\tau}{\sigma}\right) - \Phi\left(\dfrac{a-\tau}{\sigma}\right)} I_{[a,b]}(\tau), \tag{17}$$

where $\Phi(\cdot)$ is the standard normal cumulative density function and $I[a,b](\tau) = 1$ if $a \leq \tau \leq b$ and zero otherwise.

The posterior distribution is interpreted as a relative time of the experiment and therefore the prior should be limited to positive values of τ_i. To achieve this, the support of (17) is set to $a = 0, b = \infty$. Furthermore, the conditioning of the prior to current experiment time t is achieved by setting its mean value to $\mu = E(\tau)$-t,

where $E(\tau)$ is the expected value of $1-RUL$ (mean time to failure). Finally, the covariance is time dependent and set to $\sigma^2 = V_0 \cdot t$, where V_0 is the inflation constant. The result of inflation is that in the initial stages of the bearing's life cycle, the prior will have a low covariance and will be the dominating part of (16). As the time progresses, the inflating covariance will effectively put more weight to the observed data and the GP model likelihood $P(D_\alpha^w(\mathcal{P}_t,\mathcal{P}_e)|\tau_i)$ will dominate.

The numerical estimation of the posterior (16) is schematically described in Figure 8. For a specific feature value $D_\alpha^w(\mathcal{P}_t,\mathcal{P}_e)$, measured at time t, the likelihood $p(D_\alpha^w(\mathcal{P}_t,\mathcal{P}_e)|\tau_i)$ is computed for each value of $\tau_i \in [0,1]$. The likelihood is then multiplied by the prior , evaluated at the same values of τ_i and normalised. The result of the computation is the posterior pdf $p(\tau_i|D_\alpha^w(\mathcal{P}_t,\mathcal{P}_e))$.

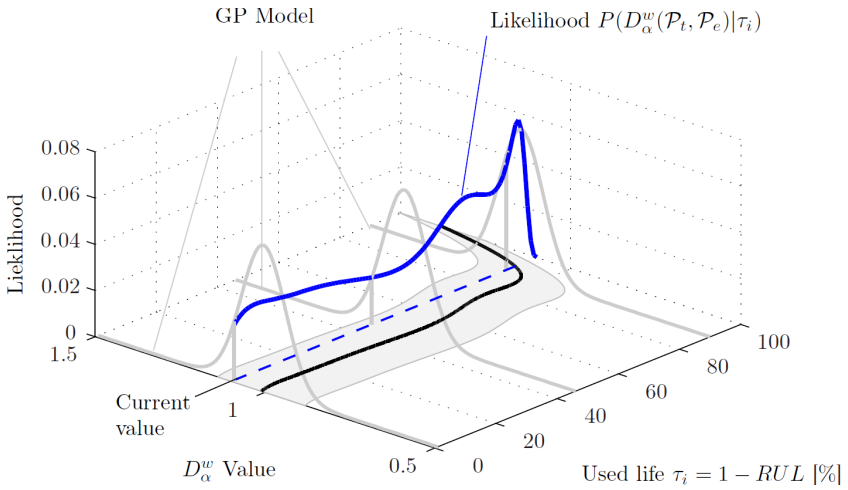

Figure 8: Calculating the probability for feature value $D_\alpha^w(\mathcal{P}_t,\mathcal{P}_e) = 1.15$.

4.4 An example

The proposed approach has been evaluated on the data set from the IEEE PHM 2012 Data Challenge [17]. The authors provided three batches of data, each one corresponding to different speed and load conditions. It is important to stress that the operating speed and load are not taken into account as independent inputs. Hence the robustness properties of the entire setup can be clearly demonstrated. The generated vibrations were sampled with $22kHz$ for duration of $100ms$, repeated every 5 minutes. The experiments were stopped when the RMS value of the generated vibrations surpassed $20ms^{-2}$. A subset of the overall set of the available experimental runs has been taken for the study since some of the runs exhibited obvious inconsistency with the rest of the data, which is ascribed to the fact the tests were accelerated, hence different mechanisms leading to failure may take action.

The procedure for RUL estimation is applied for each of the 16 WPT nodes, which results into 16 GP models. Each GP model describes the evolution of the Rényi entropy based features for each WP node. The final prediction of RUL is performed by fusing the predictions of all 16 GP models.

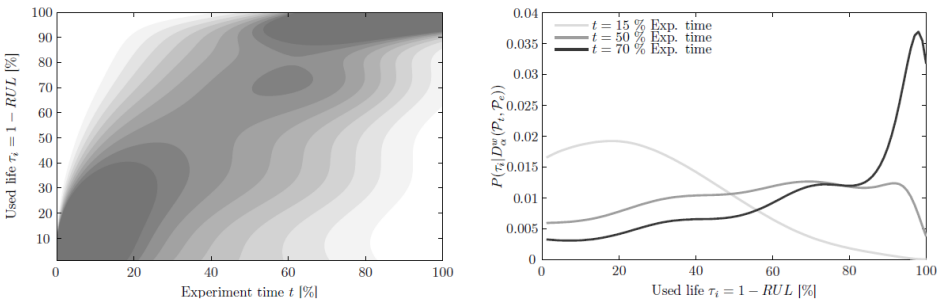

Figure 9: Evolution of $P(\tau_i|D_\alpha^w(\mathcal{P}_t,\mathcal{P}_e))$ (bearings' used life) using the 3^{rd} WP node.

Using the Bayes' rule (16) with the truncated prior bearing's RUL can be computed at any time moment.

Such an evolution of RUL is shown in Figure 9. As the duration of experiments varies, the x-axis is normalised on the interval $[0,1]$. The results exhibit almost linear relationship between the experiment time and the increase of the used life. At the very beginning, up to 20% of the experiment time, the variance of the posterior $P(\tau_i|D_\alpha^w(\mathcal{P}_t,\mathcal{P}_e))$ (16) is small. In the middle of the experiment, the uncertainty of the estimates are somewhat higher. Towards the end of the experiment, when the measured feature values become sufficiently high, the estimates become more precise.

5 Implementation

The implementation of the above algorithms is done within the context of a distributed sensor network which allows for signal acquisition from distributed sensors, local signal processing and fusion of the diagnostic results on a central server. Key to the concept is comprehensive data storage and data manipulation framework with built-in condition monitoring taxonomy as well as seamless integration capabilities with existing manufacturing execution systems.

The smart node (SN) is a low-cost, energy efficient and programmable platform capable of performing data acquisition and signal processing for condition monitoring, prognostics and health management of mechanical drives. The principle idea is illustrated in Figure 10.. SN's can be mounted on several sites on the monitored drive. They perform periodical data acquisition and execute the local feature extraction, e.g. by means of WPT. Features are transferred by various communication protocols to the central server where the system condition can be inferred and accommodation actions can be suggested to the maintenance team.

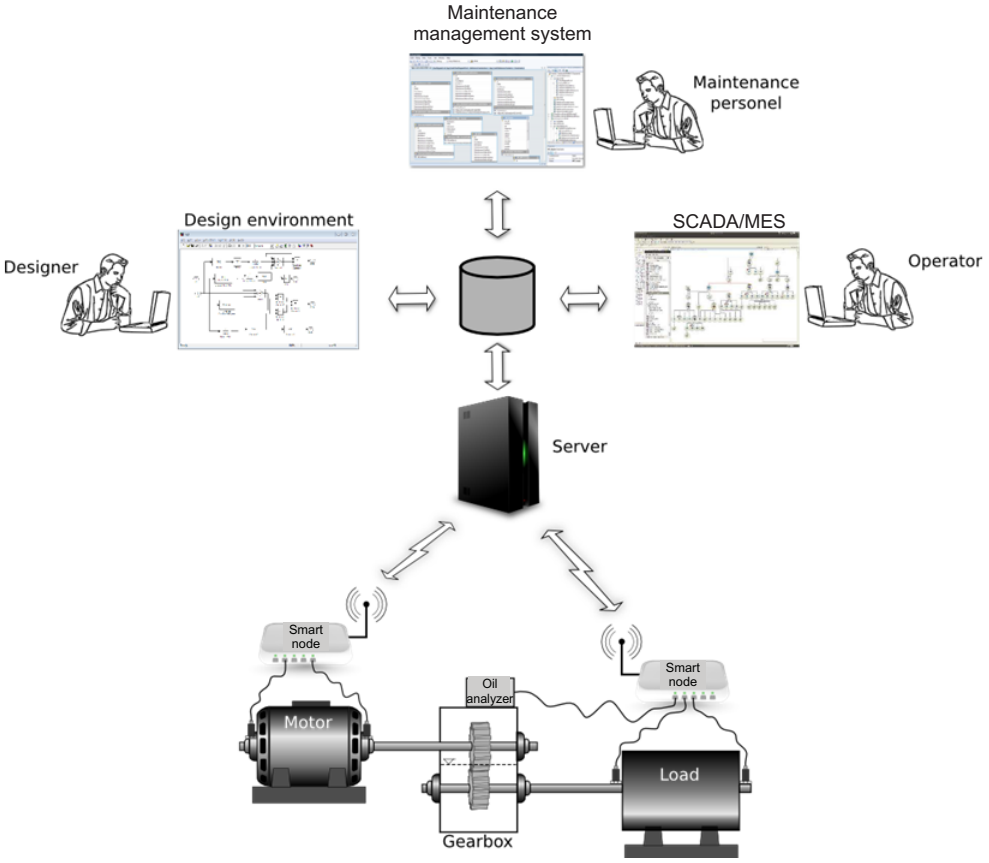

Figure 10: The condition monitoring platform based on smart nodes (SNs).

5.1 MIMOSA database

Storing data in condition monitoring applications is rarely discussed in the publications as the bulk of interest of the community is on algorithms for feature extraction, diagnostics and prognostics.

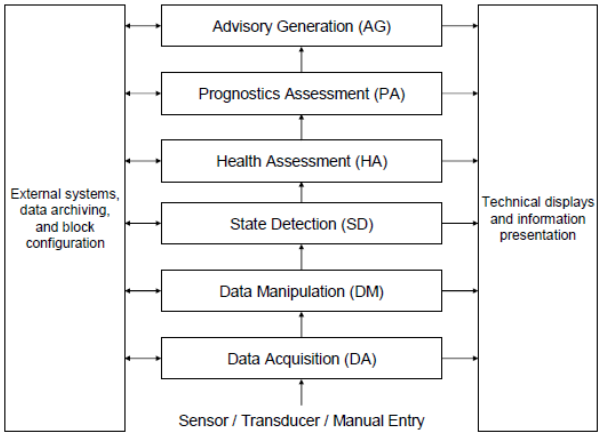

Figure 11: Data processing and information flow in MIMOSA.

MIMOSA OSA-EAI [16] standard splits the data into 4 main segments, i.e.

- the taxonomy i.e. basic data about the enterprise, machines and sensors,
- segment for machine condition monitoring,
- maintenance support and
- estimation of the machine reliability.

5.1.1 The taxonomy

The notable strength of MIMOSA comes from the completeness and consistency of the entities that constitute the system. On the topmost level there is the entity referred to as *Enterprise*. It comprises all the data relevant for the company. Hence one can handle data from different companies within the same system.

5.1.2 Condition monitoring

MIMOSA OSA-EAI delivers a model which consists of 6 modules, cf. Figure 11. The first two modules take care of correct data acquisition and evaluation of the features. The next module checks the conformity of the pattern of features with the reference pattern and an alarm is triggered if discrepancy is detected. The health assessment module provides an estimate of the current health i.e. indicating tentative faulty locations. The family of robust diagnostic algorithms above make part of this block. From the historical trends of features as well as the available degradation models the remaining useful life of the machine can be estimated. Based on the assessed current health and expected machine lifetime the last module generates guidelines for the operators and maintenance personnel.

5.1.3 Maintenance support

This segment has the role of a two-way interface between the information system on one hand and operators and maintenance people on the other. It allows the information system to deliver guidelines for actions to the operators and also receives information from the operators and maintenance staff about the realized maintenance interventions. It is very important that a concise and complete information about what exactly has been done during the maintenance actions is available in the system in order not to mislead the diagnostic and prognostic algorithms. Potential differences in the features patterns before and after the maintenance action should be taken into account by the diagnostic algorithms otherwise one can risk misinterpretation and tentative false alarms.

5.1.4 Machine reliability

In order to assess the level of reliability associated with a machine, it is essential to dispone of as much information as possible. Particularly important is the history of machine breakdowns, availability of spare parts and location of the machine. Besides reliability this segment contains also data on physical availability of the machines.

5.2 Implementation of MIMOSA database

MIMOSA OSA-EAI standard contains only technical specifications concerning the database and communication interfaces. The implementation itself is a task in the domain of the system integrator.

5.2.1 The database

The MIMOSA standard defines the database with the entity relational model as an XML scheme. The consortium that takes care of the MIMOSA standard offers also SQL script files for rapid prototyping and design of the database. Besides the basic structure, the SQL file contains also the data about look-up tables.

Each table, which corresponds to a particular entity, contains the primary key composed of three fields: (i) instance of the database, (ii) instance of the installation and (iii) unique identifier of a particular line. The tables of codings contain also an additional key, i.e. the code of the code. Such an apparently complex structure preserves the codes in the process of database upgrade.

5.2.2 Communication interfaces

MIMOSA OSA-EAI standard foresees communication interfaces realized as Web services. The standard specifies quite precisely the required commands and their parameters by means of the WSDL files. The set of commands allows for access to all the data and functions contained in the information system for condition monitoring. In addition, there are also interfaces for connections with the external information systems such as MES or ERP.

Specification by using WSDL files allows for simple implementation of the more demanding commands. A problem with communication with web services is the amount of overhead data. In the cases where reliability of transfer along with the amount of data are important, all the communication interfaces can make use of the binary data transfer.

5.3 A prototype

A prototype of the entire platform is implemented on a large milling machine. The machine serves for fine surface processing of usually large metal parts. The most critical components of the machine are (i) bearings of the milling drive, (ii) gearbox which transfers power from the motor to the platform and the cutting tool. As the machine is critical, on-line monitoring is essential to timely alert on any tentative incipient fault in the system and hence prevent unexpected breakdown.

The diagnostic system builds on 4 vibrational signals, 2 temperatures, current sensor and the encoder on the shaft of the driving motor. Periodical acquisition data by means of fast sampling serves to calculate features that reflect the current health of the drive. The calculated features are sent and stored on the server. Hence the data transfer over communication network is substantially economized. The smart node contains additional memory modules which allow for buffering raw signals from recent data acquisition sessions. In case of some event, be it departure from the normal range or simply on request, also raw data are sent to the server.

The communication interface for storing the data into the database relies on the REST Web service. In addition, the system allows for acquiring data into the database from other platforms such as PLC's and MES. MES system delivers data which provide important information about the current work order being executed on the machine, main parameters of the piece being milled and environmental data. They are used by the detection algorithms. Indeed change in operational conditions implicates likely changes in vibrations which are then appropriately normalized.

The entire MIMOSA database is realized by using open code solutions i.e. MySQL server. The additional software was realized using Java and Python. In such a way the costs related to licences are brought to zero.

6 Conclusions

In this contribution we discuss a robust data driven approaches to fault diagnostics and prognostics of rotational drives subjected to variable operating conditions (load and speed) and incomplete prior knowledge. The diagnostic method employing Jensen-Rényi divergence seems to be promising for practical applications. The diagnostic algorithm is easy to design and the results produced can be associated a plausible physical explanation. On a case study it is shown that relatively accurate detection can be achieved as well as reasonable isolation, provided the appropriate load conditions are applied.

In what concerns prognostics a data driven approach is presented based on the concept of complexity. The proposed approach has two main advantages. Firstly, the calculation of the corresponding entropy based features requires no prior knowledge about the physical characteristics of the system, only life-long tests on similar items of equipment, and no information about the operating conditions. Secondly, their wavelet based numerical estimation imposes no limits on the statistical characteristics of the analysed signals, which makes them suitable for monitoring bearings running under constant as well as fluctuating operating conditions.

An implementation framework based on MIMOSA standard is suggested for systematic and consistent data storage and tracking the maintenance actions on the asset.

Acknowledgment

We like to acknowledge the support of the Slovenian Research Agency through the grant L2-4160 and MGRT project 3211-09-000130 (Prodismon).

References

[1] W. Bartelmus, R. Zimroz, *A new feature for monitoring the condition of gearboxes in non–stationary operating conditions*, Mechanical Systems and Signal Processing, Vol. 23, No. 12, Elsevier (2009), pp. 1528–1534.

[2] M. Basseville, *Divergence measures for statistical data processing*, Tech. Rep., IRISA, 2010.

[3] S. Blanco, A. Figliola, R. Q. Quiroga, O. A. Rosso, E. Serrano, *Time-frequency analysis of electroencephalogram series. III. Wavelet packets and information cost function*, Phys. Rev. E, Vol. 57, No. 1, , (1998), pp. 932-940.

[4] N. Baydar, A. Ball, *Detection of gear deterioration under variying load conditions using the instantaneous power spectrum*, Mechanical Systems and Signal Processing, Vol. 14, No. 6, Elsevier (2000), pp. 907–921.

[5] P. Boskoski, D. Juricic, *Fault detection of mechanical drives under variable operating conditions based on wavelet packet Rényi entropy signatures*, Mechanical Systems and Signal Processing, Vol. 31, , Elsevier (2012), pp. 369-381.

[6] F.T. Chaari, M. Haddar, *Analytical modelling of spur gear tooth crack and in uence on gearmesh stiffness*. European Journal of Mechanics - A Solids, Vol. 28, No. 3,, (2009), pp. 461-468.

[7] Fan, M. J. Zuo, *Gearbox fault detection using Hilbert and wavelet packet transform*, Mechanical Systems and Signal Processing, Vol. 20, No.3, Elesevier (2006), pp. 966–982.

[8] Y. Feng, F. S. Schlindwein, *Normalized wavelet packets quantifiers for condition monitoring*, Mechanical Systems and Signal Processing, Vol. 23, No. 3, Elsevier (2009), pp. 712–723.

[9] P. Boskoski, M. Gasperin, D. Petelin, *Bearing fault prognostics based on signal complexity and Gaussian process models*, IEEE International Conference on Prognostics and Health Management,Denver, USA, 2012 June 18-21, pp. 6.

[10] M. Gasperin, D. Juricic, P. Boskoski, J. Vizintin, *Model-based prognostics of gear health using stochastic dynamical models*, Mechanical Systems and Signal Processing, Vol. 25, No. 2, Elsevier (2011), pp. 537-548.

[11] J. Kocijan, D. Petelin, *Output-error model training for Gaussian process models*. Lect. notes comput. sci., Vol. 6594, , Springer, pp. 312-320.

[12] M. N. Kotzalas, T. A. Harris, *Fatigue Failure Progression in Ball Bearings*, Transactions of ASME, Vol. 12, No. 3, (2001), pp. 238–242.

[13] R. Li, P. Sopon, D. He, *Fault features extraction for bearing prognostics*, Journal of Intelligent Manufacturing, Vol. 23, , (2012), pp. 313–321

[14] X. C. R. López-Ruiz, H.L. Mancini, *A statistical measure of complexity*, Physics Letters A, Vol. 20, No.9, (1995), pp. 321-326.

[15] S. Mallat, *A wavelet tour of signal processing*, Academic Press (2008).

[16] MIMOSA OSA-EAI v3.2.2, www.mimosa.org, (2013).

[17] P. Nectoux, R. Gouriveau, K. Medjaher, E. Ramasso, B. Morello, N. Zerhouni, C. Varnier, *PRONOSTIA: An Experimental Platform for Bearings Accelerated Life Test*, IEEE International Conference on Prognostics and Health Management, Denver, USA, 2012 June 18-21, pp. 6.

[18] PHM, Prognostics and Health Managment Society 2009 Data Challenge, http://www.phmsociety.org/competition/09, 2009.

[19] A. Poulimenos, S. Fassois, *Parametric time domain methods for nonstationary random vibration modelling and analysis: A critical survey and comparison*, Mechanical Systems and Signal Processing, Vol. 20, No. 4, Elsevier (2006), pp. 763-816.

[20] M. Priestley, *Spectral Analyses and Time Series*, Academic Press (1981).

[21] R. B. Randall, J. Antoni, S. Chobsaard, *The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals*, Mechanical Systems and Signal Processing, Vol. 15, , Elsevier (2001), pp. 945 - 962.

[22] C. Stander, P. Heyns, *Instantaneous angular speed monitoring of gearboxes under non–cyclic stationary load conditions*, Mechanical Systems and Signal Processing, Vol. 19, No. 4, Elsevier (2005), pp. 817-835.

[23] S. Suresh, *Fatigue of Materials*, Cambridge University Press (1998).

[24] G. Vachtsevanos, F.L. Lewis, M. Roemer, Hess, A., B. Wu, *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*, Wiley (2006).

[25] W. Wang, *Early Detection of gear tooth cracking using the resonance demodulation technique*, Mechanical Systems and Signal Processing, Vol. 15, , Elsevier (2001), pp. 887-903.

[26] W. Q. Wang,F. Ismail, *Prognosis of machine health condition using neuro-fuzzy systems*. Mechanical Systems and Signal Processing, Vol. 18, , Elsevier (2004), pp. 813-831.