

Optimally Parameterized Wavelet Packet Transform for Machine Residual Life Prediction

M. F. Yaqub, I. Gondal and J. Kamruzzaman

Gippsland School of Information Technology, Monash University, Melbourne, Australia

ABSTRACT

One of the prevalent issues in condition based maintenance (CBM) is to predict the residual life of the equipment. This paper proposes a novel framework to predict the remnant life of the equipment, called Residual life prediction based on optimally parameterized Wavelet transform and Mute-step Support vector regression (RWMS). In optimally parameterized wavelet transform, a generalized criterion is proposed to select the wavelet decomposition level which works for all the applications and decomposition nodes are selected by characterizing their dominance level based upon relative fault signature-signal energy contents. The prediction model is based on multi-step support vector regression (MSVR) and prediction accuracy is improved in comparison with the techniques based on support vector regression (SVR). Performance of RWMS is evaluated in terms of Root Means Square Error (RMSE), studies show that proposed algorithm predicts the residual life of the equipment accurately.

INTRODUCTION

Machine health monitoring (MHM) is crucial in all industrial processes to achieve high reliability, reduced man power and scheduled maintenance. MHM specifically deals with abnormality diagnosis and prognosis. Many works have already been done in the field of fault diagnosis, also proposed by the authors (Yaqub et al., 2011a, Yaqub et al., 2011c, Yaqub et al., 2011e, Yaqub et al., 2011d, Yaqub et al., 2011b, Hu et al., 2007, Teotrakool et al., 2009). One of the challenging problems in CBM is to predict the residual life of the equipment. Figure 1 gives the flowchart for condition based monitoring of the rotary system.

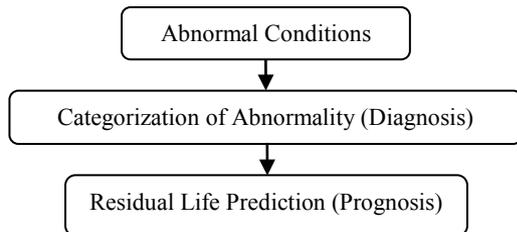


Figure 1. Condition based maintenance system.

In Figure 1, the first phase is to categorize the type of the fault. Once the type of the fault is categorized, it is crucial to predict the residual life of the equipment as it facilitates the maintenance staff to schedule the repair by optimizing demand-supply relationships rather than sudden and unplanned break down of the equipment. The conventional techniques are based upon scheduled plant maintenance after a specific predefined period of time, which is prone to sudden break down of the equipment as well as disassembling and reassembling may initiate problems in already perfectly running process. The proposed model can schedule on-demand and intelligent maintenance by predicting the residual life of the equipment and repair is only advised when necessary.

In order to predict residual life of the equipment, it is important to record certain physical parameters such as 'vibration' which varies according to the variation in the machine dynamics. In case of rotary machinery, malfunctioning in the operation of the bearing is the most common fault. It has been investigated that 40% of the total machine faults are

because of bearing (Morel, 2002). These bearing faults change the machine dynamics and generate certain vibration patterns which depend upon bearing characteristics frequency. The vibration characteristics frequencies (El Hachemi Benbouzid, 2000) for inner race (f_{ID}), outer race (f_{OD}) and ball (f_{BD}) defects can be represented by (1) – (3):

$$f_{ID} = \frac{n}{2} f_{rm} \left(1 + \frac{d_{ball}}{d_{pitch}} \cos \phi \right), \quad (1)$$

$$f_{OD} = \frac{n}{2} f_{rm} \left(1 - \frac{d_{ball}}{d_{pitch}} \cos \phi \right), \quad (2)$$

$$f_{BD} = \frac{d_{pitch}}{2d_{ball}} f_{rm} \left(1 - \left(\frac{d_{ball}}{d_{pitch}} \right)^2 \cos^2 \phi \right), \quad (3)$$

where f_{rm} , d_{pitch} , d_{ball} , n and ϕ represent frequency of rotation, pitch diameter, ball diameter, number of balls and the contact angle respectively as highlighted in Figure 2.

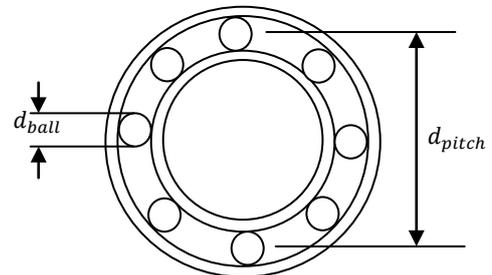


Figure 2. Rolling element ball bearing geometry.

Vibration signal is non-stationary in nature, i.e., its spectral contents vary with respect to time. Wavelet transform (WT) is effectively used in order to extract the time-frequency domain contents of the vibration signal (Peng and Chu, 2004). In particular, wavelet packet transform (WPT) (Peng and Chu, 2004, Eren and Devaney, 2004, Teotrakool et al., 2009, Lau and Ngan, 2010, Yen and Kuo-Chung, 1999, Zhao et al., 2009, Eren et al., 2010, Jianhua et al., 2010) decomposes the signal into multiple frequency nodes and provides multi-resolution analysis. The fault diagnostic schemes use fixed predefined decomposition levels for WPT (Eren and Devaney, 2004, Teotrakool et al., 2009, Lau and Ngan, 2010, Yen and Kuo-Chung, 1999, Zhao et al., 2009, Eren et al., 2010, Jianhua et al., 2010) and work only for a particular application, which is very limiting. The first contribution of the proposed technique, i.e., Residual life prediction based on optimally pa-

parameterized Wavelet transform and Multi-step Support vector regression (RWMS) lies in developing a generalized criterion for optimal selection of the decomposition level for accurate profile of fault-signatures.

In case of vibration signal, not the entire frequency band contains the true information about the fault vibration signal, only sub-portion (Yaqub et al., 2011c) of the whole frequency spectrum contains true fault signatures. In wavelet packet decomposition, the vibration data is decomposed into different frequency sub-bands and feature vector is extracted from these frequency sub-bands. All the features do not contain true fault signatures, so feature selection is vital along with optimal decomposition level selection. The second contribution of RWMS is to characterize the dominance level of the features and find an optimal number of features.

In CBM, a few of the techniques deal with prognosis (Heng et al., 2009). Some of these techniques treat prognosis as the classification problem, i.e., datasets with multiple severity levels of the faults is used to build the fault prognostic model and same datasets is used to validate the performance of the prediction model (Hu et al., 2007, Zhang et al., 2010, Zhang, 2010, Kwan et al., 2003). These prognostic models only predict the fault severity levels which were used in building the model and cannot predict the state when severity level of the fault is different than training levels for the prediction model. In (Tran et al., 2008, Samanta and Nataraj, 2009, Shao and Nezu, 2000, Wang and Vachtsevanos, 2002, Berenji and Yan, 2006), the predictor uses previous observations to predict the future values. These techniques can predict the future state of the equipment based upon current observations, but cannot determine the residual life in terms of how long the equipment is going to last. In these techniques, if number of steps to be predicted is increased, their reliability is decreased (Tran et al., 2008). Moreover, the cracks in the rotary machinery components such bearings propagate in a non-linear way with respect to time, from inception to breakdown as described in Paris's fatigue model (Pugno et al., 2006). In order to incorporate the impact of non-linearity in defect propagation, RWMS is based upon multi-step SVR rather than simple SVR as in the existing literature (Sotiris and Pecht, 2007, Kim et al., 2009). The third contribution of RWMS is to enhance the prediction accuracy of the residual life of the equipment based on multi-step SVR.

The paper is organized as follows: Section II presents the framework for the proposed technique; Section III illustrates the data acquisition and parameter optimization; Section IV contains the results and validates the performance of the proposed predictive maintenance model in comparison with the existing scheme and Section V presents the concluding remarks.

FRAMEWORK OF RWMS

Figure 3 gives hierarchical paradigm for RWMS. In order to extract the time-frequency information in the non-stationary vibration signal, digitized data are decomposed using WPT (Peng and Chu, 2004, Eren and Devaney, 2004, Teotrakool et al., 2009, Lau and Ngan, 2010, Yen and Kuo-Chung, 1999, Zhao et al., 2009, Eren et al., 2010, Jianhua et al., 2010). WPT helps in investigating the frequency contents of the vibration data in different frequency ranges, i.e., nodes. A generalized criterion is proposed for optimal selection of wavelet decomposition level. Feature vectors are computed by evaluating RMS values of wavelet decomposition nodes at the selected decomposition level (Eren and Devaney, 2004, Teotrakool et al., 2009, Lau and Ngan, 2010). Once the optimal feature vector is finalized, the predictive maintenance model is built based upon multi-step support vector regression.

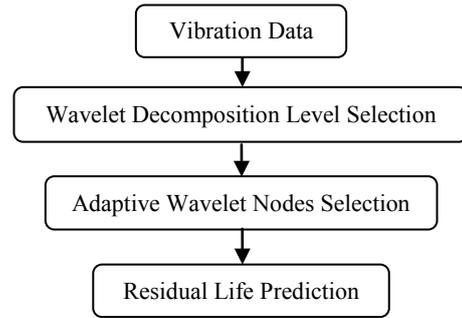


Figure 3. Framework of RWMS.

Wavelet Packet Decomposition and Feature Extraction:

Among the time-frequency domain signal processing techniques, e.g., discrete Fourier transform (DFT), short time Fourier transform (STFT) and wavelet packet transform (WPT), WPT can be used for comprehensive analysis of non-stationary vibration signal to reliably extract its time and frequency domain contents (Peng and Chu, 2004). DFT of a non-stationary signal $x[n]$ (4) does not exploit the variation in frequency contents with respect to time. Rather it averages out the frequency content over the whole signal range (Peng and Chu, 2004).

$$X(k) = \sum_{n=0}^{N-1} x[n] e^{-j\left(\frac{2\pi}{N}\right)kn}, \quad k = 0, 1, \dots (N-1). \quad (4)$$

The short comings of DFT can be overcome by STFT (5), but STFT suffers from the problem that it yields same time and frequency resolution for low and high frequencies. The time and frequency resolution remains same because window size $\omega[n]$ remains constant throughout the analysis (Peng and Chu, 2004).

$$X(m, k) = \sum_{n=0}^{N-1} x[n] \omega[m-n] e^{-j\left(\frac{2\pi}{N}\right)kn}, \quad k = 0, 1, \dots (N-1). \quad (5)$$

In order to overcome the drawback of fixed time-frequency resolution in STFT, wavelet transformation can be used which has the tendency to perform multi-resolution analysis. Wavelet packet transform provides multi-resolution analysis, i.e., time and frequency resolution can be adjusted (Eren and Devaney, 2004, Teotrakool et al., 2009, Lau and Ngan, 2010, Yen and Kuo-Chung, 1999, Zhao et al., 2009, Eren et al., 2010, Jianhua et al., 2010). Figure 4 gives the decomposition tree for WPT. Digitized vibration data are passed through high pass $h[n]$ and low pass $g[n]$ Quadrature Mirror Filters (QMFs) (6)-(7) and then down sampled. QMFs are finite impulse response (FIR) filters or infinite impulse response (IIR) filters (Walker, 1999). Filter selection is a very crucial part in case of analysis using WT. In the proposed scheme, Daubechies (Db5) filter (Yen and Kuo-Chung, 1999) is used which is an FIR filter. Figure 4 also shows that the total number of nodes at any decomposition level is given by (8):

$$y_{approx}[n] = x[n] * g[n],$$

$$\text{or } y_{approx}[n] = \sum_{k=-\infty}^{k=\infty} x[k] \times g[n-k]. \quad (6)$$

$$y_{detailed}[n] = x[n] * h[n],$$

$$\text{or } y_{detailed}[n] = \sum_{k=-\infty}^{k=\infty} x[k] \times h[n-k]. \quad (7)$$

$$N_j = 2^{j_{decomp}}. \quad (8)$$

RMS value of wavelet decomposition nodes is computed as in (9) which is extensively used as feature extraction metric in the literature (Eren and Devaney, 2004, Teotrakool et al., 2009, Lau and Ngan, 2010).

$$X_{rms} = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (9)$$

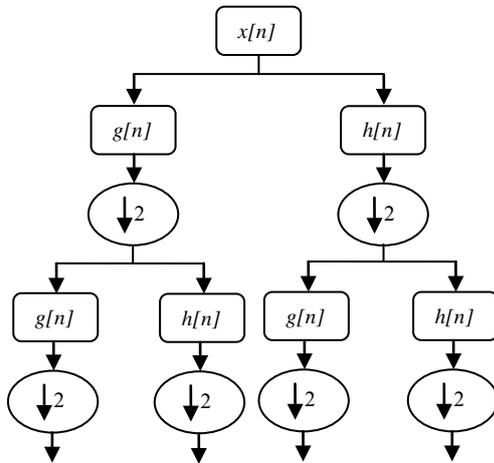


Figure 4. Wavelet packet transform (WPT).

Table I lists the RMS values for the first node (lowest frequency node) in each of the six decomposition levels. It shows that the RMS value of the first decomposition node decreases by increasing the decomposition level, which justifies that the level of distribution of the signal energy among decomposition nodes increases by increasing the decomposition level. Moreover, Figure 5 plots the first (lowest) frequency node for each of the six decomposition levels (vibration data is for outer race fault). The transients in the first node decrease by increasing the decomposition level, which justify that by increasing the decomposition level, the distribution of signal transients is improved among wavelet decomposition nodes.

Table 1. RMS value of the lowest frequency nodes at levels 1-6

1	2	3	4	5	6
0.16	0.15	0.13	0.11	0.10	0.08

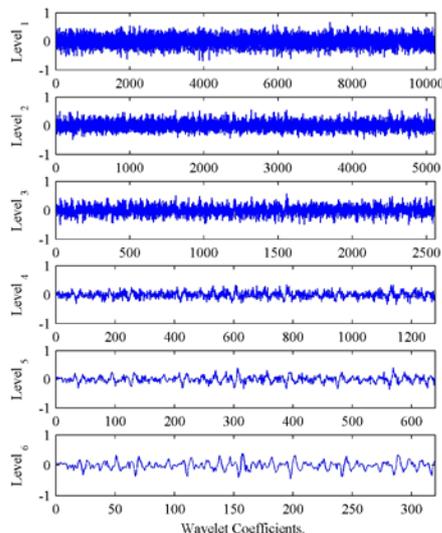


Figure 5. Coefficients for the lowest frequency node at levels 1-6.

Table 1 and Figure 5 validate that the distribution of the signal transients is improved by increasing the decomposition level, i.e., the increment in the decomposition level provides more detailed distribution of the signal contents in different frequency bands. At the same time, increase in the decomposition level gives rise in the number of decomposition nodes (features) so it is very important to devise a generalised criterion, which defines an upper limit for the optimal decomposition level selection. Apparently, the decomposition level should be as high as possible, but there should be an upper limit to optimally select the decomposition level, after which the performance of the prognostic system is not improved further.

Wavelet Decomposition Level Selection:

RWMS proposes a novel criterion for optimal selection of the decomposition level which works for all the application contrary to existing schemes (Peng and Chu, 2004, Eren and Devaney, 2004, Teotrakool et al., 2009, Lau and Ngan, 2010, Yen and Kuo-Chung, 1999, Zhao et al., 2009, Eren et al., 2010, Jianhua et al., 2010) which are application specific. The decomposition level selection is such that the high-energy frequency bands in the vibration signal are split so that maximum fault signature-signal energy is achieved. This is because only a particular band of the overall vibration spectrum contains true fault signature, called resonant frequency bands with maximum signal to noise ratio (SNR) (Yaqub et al., 2011c). In the given vibration signal, splitting of high energy frequency bands into multiple lower level decomposition nodes helps in identifying the nodes which are true representative of the fault. The optimal decomposition level can be determined based on the criterion defined in proposition 1:

Proposition 1: The high energy frequency bands in the vibration signal should be decomposed into lower level wavelet nodes such that the maximum fault signature-signal energy is achieved in a node.

To study the analytical relations and expressions for the proposed scheme, let us assume that R_i^j represents the RMS value of the i -th node at the j -th decomposition level and R^j is the vector with RMS values for the nodes at j -th decomposition level as represented in (9). It is shown in (8) that wavelet decomposition at level j divides the overall vibration spectrum into 2^j nodes. R_i^j gives a measure of signal energy at wavelet decomposition nodes, i.e., greater RMS value represents greater signal energy.

$$R^j = \{R_1^j, R_2^j, R_3^j, \dots, R_{2^j}^j\} \quad (10)$$

In order to achieve high energy frequency bands for the fault signature-signal, nodes are reordered in the descending order of their energy values at their respective level. Let \hat{R}_i^j represents the i -th maximum value at the j -th decomposition level. Then the number of nodes N_L which contain τ_Q signal energy can be determined from the relation in (11):

$$\gamma^j = \frac{\sum_{i=1}^{N_L} \hat{R}_i^j}{\sum_{i=1}^{2^j} R_i^j} \geq \tau_Q \quad (11)$$

where τ_Q represents the threshold for quantifying the node ratio containing certain portion of the signal energy, i.e., N_L is the number of nodes which contain $\tau_Q\%$ of the signal contents in terms of RMS value. The node-ratio (11) is computed for every decomposition level. By increasing the decomposi-

tion level, the ratio decreases as the number of sub-bands is increased and the probability that lesser number of nodes would contain signal energy greater than τ_Q . Once the maximum signature-signal energy nodes are determined, then the node-ratio (11) does not change even if decomposition levels are further increased, because least possible number of nodes has been achieved which contain signal energy less than τ_Q . Analytically, the rate of change of γ^j with respect to j should be zero, resulting in maximum fault signature-signal energy in the nodes at that level (12):

$$\frac{d\gamma^j}{dj} \approx 0. \quad (12)$$

Studies were conducted for bearings faults: inner race, ball and outer race faults to determine the dependency of τ_Q on node-ratio (11), by varying τ_Q from 10% to 90% and it was observed (Section IV) that the horizontalness of the node-ratio (11) is always achieved at the same decomposition level irrespective of the value of threshold, i.e., decomposition level selection is independent of the threshold value, τ_Q . The optimal decomposition level is found to be '6', which results in 64 feature values (8).

Adaptive Wavelet Nodes Selection

It has been investigated (Yaqub et al., 2011c) that only a small portion of the overall vibration spectrum contains true information regarding the fault induced vibrations and the rest is noise. The useful information lies in the resonant frequency band which contains the highest SNR (Yaqub et al., 2011c). All the features do not contain true fault signatures, so feature selection is vital along with optimal decomposition level selection. A criterion is proposed to select the nodes containing relatively larger signal energy. The rest of nodes corresponding to poor SNR values are discarded and not used in the prediction model which ensures the robustness in the proposed scheme under poor SNR.

Let us assume that R_i^j represents the RMS value of the i -th node at the j -th decomposition level. R_i^j gives a measure of the signal energy at i -th node, i.e., greater RMS value represents greater signal energy.

$$\mathbf{R}^j = \{R_1^j, R_2^j, \dots, R_{2^j}^j\}. \quad (13)$$

\mathbf{R}^j represents the RMS values for all the nodes at j -th decomposition level. If \hat{R}_i^j represents the nodes order in the ascending order of their RMS values, ratio γ^n is defined such that N_D nodes are selected out of the total 2^j nodes as in (14). The optimal value for γ^n is determined using parameter optimization techniques given in Section III. Irrespective of the machine dynamics, the optimal value for γ^n gives the number of nodes which contain relatively higher value for signal energy, and the corresponding node indices give feature vector for the residual life prediction model.

$$\frac{\sum_{i=1}^{N_D} \hat{R}_i^j}{\sum_{i=1}^{2^j} R_i^j} \geq \gamma^n. \quad (14)$$

Residual Life Prediction:

RWMS develops a novel multi-step predictive maintenance system based upon support vector regression. Vibration data from multiple severity levels is used and feature vectors are extracted as described in subsection 'adaptive wavelet nodes selection'. The multi-step SVR (MSVR) model works in two

steps. In the first step, the two severity levels (upper and lower) of the training data are selected which surround the test query (test datapoint) with unknown severity level of the fault. In the second step, the SVR based prediction model is built with the selected severity levels. Instead of building the prediction model with all the severity levels of the training data (Sotiris and Pecht, 2007, Kim et al., 2009), MSVR uses only two severity levels of the training data within which test query lies. It enhances the overall prediction accuracy because defect propagation or RMS based feature values follow a non-linear curve with respect to time as described by Paris's law (Pugno et al., 2006). It explains the rate of change of defect length with respect to speed of rotation as in (15):

$$\frac{da}{dN} = C_o(\Delta K)^n \quad (15)$$

where a and N represent instantaneous defect length and speed respectively, and C_o and n are material dependent constants. ΔK represents range of strength intensity. It can also be written as in (16):

$$D' = \frac{dD}{dt} = C_o D^n \quad (16)$$

It shows that the rate of change of instantaneous defect area D is exponential function of instantaneous defect area. Figure 6 plots the average feature value (extracted by computing RMS value) with respect to the variation in the residual life of the equipment. It shows that the feature values used to determine the residual life of the equipment vary exponentially with defect propagation.

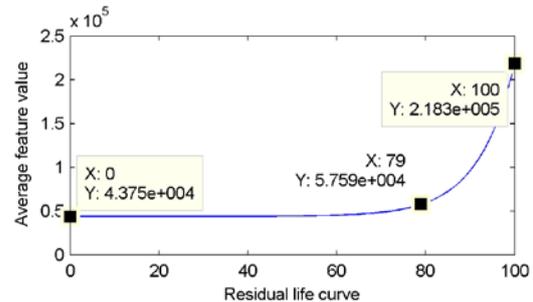


Figure 6. Average feature value vs. residual life.

Figure 7 gives the orientation of the data points with the three severity levels, i.e., sev1, sev2 and sev3. Based on the position of the test query, represented as 'star', the training severity levels are selected, i.e., 'sev1 = ●', 'sev2 = ■'. In the result section, it is validated that the selection of the training datapoints which are in the vicinity of test point, enhances the prediction accuracy contrary to the techniques (Sotiris and Pecht, 2007, Kim et al., 2009) based on building the predict

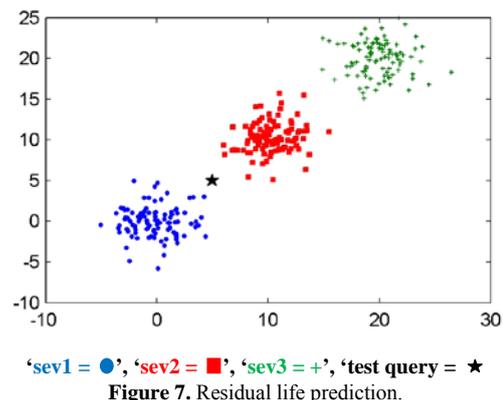


Figure 7. Residual life prediction.

tion model by using the training data points from all the severity levels.

Support vector machine (SVM) is extensively used technique for classification and estimation (Sotiris and Pecht, 2007). Figure 8 gives the orientation of the optimal hyper-plane in the case when two severity levels are used to build the fault estimation model. The data points which are nearest to the decision boundary are called support vectors and help in calibrating the fault prediction model, represented as sv_1 , sv_2 , sv_3 , and sv_4 in Figure 8.

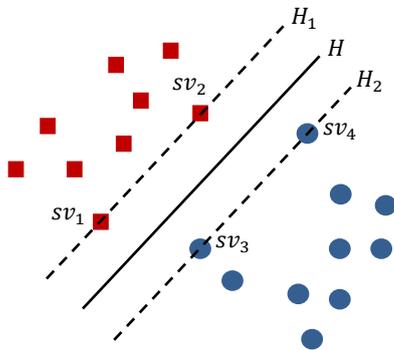


Figure 8. Optimal hyper-plane in SVM.

In order to build the residual life prediction model, consider the training dataset $D = \{(\mathbf{x}_i, y_i)\}$ with an input feature vector as $\mathbf{x}_i \in R^{N_D}$ and the corresponding severity levels are assigned the labels as $y_i \in \{-1, 1\}$. To find an optimal separating hyper plane, each input \mathbf{x}_i is mapped to higher dimensional space via a non-linear mapping, such that $\mathbf{z}_i = \varphi(\mathbf{x}_i)$. There exists a vector \mathbf{w} and scalar b that define the separating hyperplane as $\mathbf{w} \cdot \mathbf{z}_i + b = 0$ as in (17):

$$y_i(\mathbf{w} \cdot \mathbf{z}_i + b) \geq 1 - \xi_i \text{ where } (\mathbf{w} \in R^n, b \in R), \quad (17)$$

where $\xi_i \geq 0$ are slack variables and only misclassified training samples generate non-zero ξ_i . As in Figure 8, the optimal hyper-plane with maximal margin is equivalent to minimizing the value of $\|\mathbf{w}\|$ which may be defined as quadratic optimization problem as in (18):

$$\min \frac{1}{2} \langle \mathbf{w}, \mathbf{w} \rangle + C \sum_{i=1}^L \xi_i, \quad (18)$$

where C is a constant parameter, called regularization parameter, determines the trade-off between maximum margin and minimum classification error. The optimization problem in (18) is quadratic programming problem and can be formulated in terms of Lagrangian multipliers (Vapnik et al., 1994) as in (19):

$$\begin{aligned} \max L(\boldsymbol{\alpha}) &= \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,j=1}^m \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \\ \text{subject to } \sum_{i=1}^L \alpha_i y_i &= 0 \text{ and } 0 \leq \alpha_i \leq C, \end{aligned} \quad (19)$$

where α_i is non-negative Lagrangian multipliers and K is the Kernel function which is equivalent to transforming the input feature vector \mathbf{x}_i to higher dimensional feature space as in (20):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle, \quad (20)$$

There are many choices of Kernel's such as RBF (21), Polynomial (22) and Hyperbolic tangent (23):

$$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma^k \|\mathbf{x}_i - \mathbf{x}_j\|^2}, \quad (21)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i \cdot \mathbf{x}_j + 1)^d, \quad (22)$$

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\beta \mathbf{x}_i \cdot \mathbf{x}_j + b). \quad (23)$$

In fault severity estimation model, labels are only defined to determine the orientation of the optimal hyper-planes otherwise SVM also gives the probabilistic measure for the nearness and fairness of the test data point with unknown severity level from the training data points with known severity levels (Platt, 1999), which gives the residual life of the equipment. In support vector regression (SVR), the optimal regression hyperplanes (17)-(19) are determined that best fit the training datasets belonging to different severity levels. In this paper, the performance of RWMS is measured by using linear and RBF kernels which are the most commonly used kernel functions (Hu et al., 2007, Hsu et al., 2003).

DATA ACQUISITION AND PARAMETERS OPTIMIZATION

Data Acquisition:

The data were generated by NSF I/UCR Center on Intelligent Maintenance Systems (IMS) with support from Rexnord Corp. in Milwaukee, WI (Qiu et al., 2003). Four bearings were installed on one shaft. The speed of rotation was kept constant at 2000RPM with a radial load of 6000lb. On each bearing two PCB 353B33 High Sensitivity Quartz ICP accelerometers were installed for a total of 8 accelerometers (one vertical Y and one horizontal X on each). Figure 9 shows the test rig and illustrates sensor placement. All failures occurred after exceeding designed life time of the bearings that is more than 100 million revolutions. Vibration data were recorded every 10 minutes for the period of 02/12/2004 10:32:39 – 02/19/2004 06:22:39 using a NI DAQ Card 6062E. At the end of the test-to-failure experiment an outer race failure occurred on bearing 1.

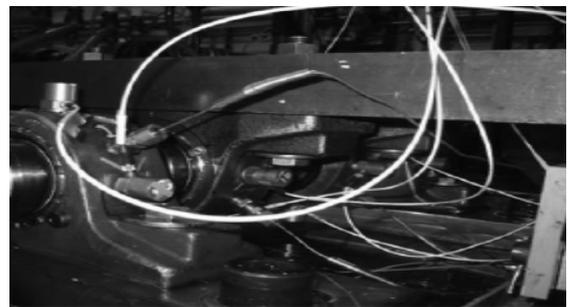


Figure 9. Experimental setup (Qiu et al., 2003).

Parameters Optimization:

There are five parameters to be optimized; decomposition level (n_{decomp}), number of nodes (N_D), regularization parameter C , RBF kernel γ^k and epsilon ϵ . Two of the parameters: n_{decomp} and N_D are related to wavelet decomposition, and rest of the parameters i.e., C , γ^k and ϵ belong to SVR. Results for optimizing n_{decomp} and N_D are presented in detail in Section IV. The applicability of the proposed generalized criterion for the optimal selection of the decomposition level is also presented. In order to optimize the SVR parameters, i.e., C , γ^k and ϵ , vibration data are divided into two subsets, training dataset (67%) and test dataset (33%). The training dataset is used for optimizing the parameters using 5-fold cross validation (Hsu et al., 2003). Root mean square

error (RMSE) (24) is minimized as quantification metric in parameters optimization. It exploits the difference between the actual and the estimated value. Greater the value of RMSE, poorer is the performance of the predictor.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Actual_i - Estimated_i)^2}{N}} \quad (24)$$

EXPERIMENTAL RESULTS

This section presents the result to validate the performance of RWMS. The performance is measured in terms of RMSE value. The actual value is taken from the curve presented in Figure 6. The first part of this section contains the performance evaluation of RWMS. The second part illustrates optimal selection of the decomposition level. The last part presents the results which emphasize the need of dominant decomposition nodes selection.

Performance Evaluation of RWMS:

Table 2 presents the results for the residual life prediction in RWMS. The experimental datasets is used for 13 levels, level-1 correspond to the data when remnant life of the equipment is 100% and level-13 correspond to the point, when remaining life of the equipment is 0%. The regressor parameters are optimized as described in subsection ‘parameters optimization’ using the training data given in Table 2 and performance is validated using the test data. Performance is measured using (24). Table 2 shows that the residual life prediction model is divided into three steps, i.e., [1→5], [5→9] and [9→13]. The performance of the proposed model is measured using linear as well as RBF kernel. The prediction accuracy is better in case of RBF kernel as compared to linear kernel as it has tendency to define boundaries for non-linearly separable datasets (Hsu et al., 2003).

Table 2. RMSE vs. Node Ratio

Train	Test	Kernel	RMSE	Avg. RMSE
[1 5]	[2 3 4]	Linear	3.7635	
[5 9]	[6 7 8]	Linear	1.3358	2.4195
[9 13]	[10 11 12]	Linear	2.1591	
[1 5]	[2 3 4]	RBF	2.5261	
[5 9]	[6 7 8]	RBF	1.3349	1.5288
[9 13]	[10 11 12]	RBF	0.7254	

Table 2 gives the performance evaluation with 3-step prediction model. Table 3 presents the performance evaluation by varying the number of steps in the prediction model. In case of ‘conventional SVR’, the prediction model is built by using training levels-1, 5, 9 & 13. Table 3 shows that the prediction accuracy of MSVR (2-step, 3-steps, 4-steps) is better as compared to the prediction model based on SVR. In case of MSVR, since the defect propagation is non-linear as described in Figure 6, by increasing the number of steps, accuracy is improved. Intuitively, the prediction accuracy is increased as it requires larger information in terms of historical (training) data to build the fault prediction model.

Table 3. Multi-step Prediction

Experimental Setup	RMSE (Linear)	RMSE (RBF)
Conventional SVR (Sotiris and Pecht, 2007, Kim et al., 2009)	5.5190	3.6636
2-Steps Prediction	5.3909	2.4422
3-Steps Prediction (RWMS)	2.5067	1.9557
4-Steps Prediction	1.9190	1.4398

Optimal Decomposition Level Selection:

Figure 9a-d present the results for optimal selection of decomposition level. The node ratio γ^j in (11) is plotted against decomposition level for different value of the thresholds τ_Q (30%, 60% and 90%). The ratio corresponds to the number of nodes containing $\tau_Q\%$ of the signal RMS value to the total number of nodes at a particular decomposition level. The results are presented for the training vibration datasets in Table 2: Figure 10a (Level-1), Figure 10b (Level-5), Figure 10c (Level-9) and Figure 10d (Level-13). Figures 10a-d show that if the number of decomposition level is increased then the node-ratio decreases to a certain extent and then horizontalized. These figures show that the optimal decomposition level is ‘6’, after which there is no significant variation in the ratio. Moreover, Figures 10a-d manifest that the optimal decomposition level selection is independent of the node-ratio threshold in (11), i.e., τ_Q .

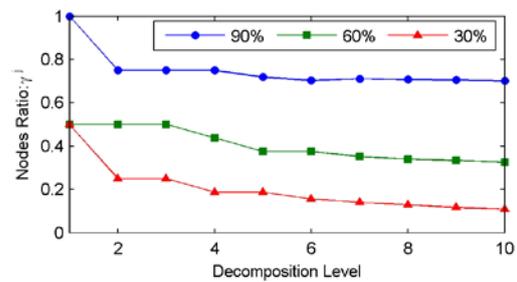


Figure 10a. Optimal decomposition level selection, Level-1

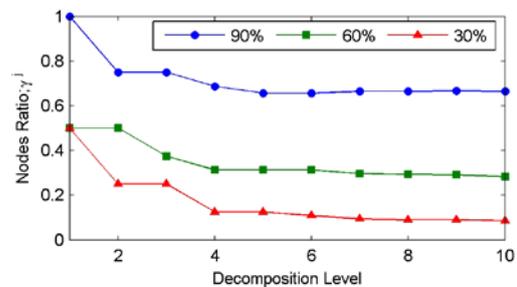


Figure 10b. Optimal decomposition level selection, Level-5.

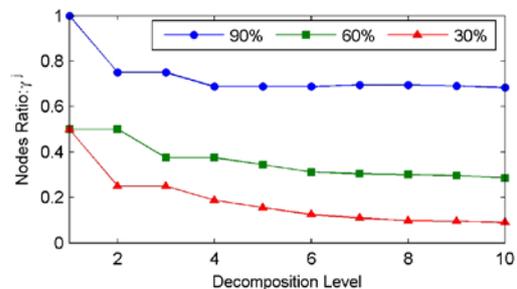


Figure 10c. Optimal decomposition level selection, Level-9.

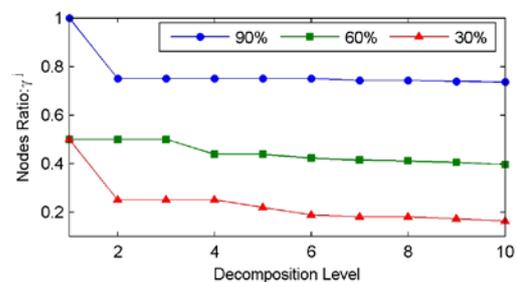


Figure 10d. Optimal decomposition level selection, Level-13.

In order to further elaborate the working of the proposed scheme, Table 4a-b list the prediction accuracy by varying the decomposition level. It shows that by increasing the decomposition level, prediction accuracy is improved. Moreover, if the decomposition level is increased beyond '6' (optimal decomposition level), the prediction accuracy almost remains same, i.e., further increment in the decomposition level is not enhancing the prediction accuracy.

Table 4a. RMSE vs. Decomposition Level (Linear Kernel)

Level	RMSE	Level	RMSE
1	3.5705	5	2.7748
2	3.4558	6	2.5067
3	3.1832	7	2.6412
4	2.8479	8	2.5367

Table 4b. RMSE vs. Decomposition Level (RBF Kernel)

Level	RMSE	Level	RMSE
1	3.9407	5	2.2350
2	2.9724	6	1.9557
3	2.7693	7	1.9727
4	2.1900	8	1.9131

Optimal Decomposition Nodes Selection:

RWMS emphasises the need for the optimal selection of the decomposition nodes rather than building the prediction model with all the nodes. Figure 11 presents the results both for 'linear' and 'RBF' kernel to substantiate the need for the selection of the dominant decomposition nodes. It shows that the prediction accuracy is maximum (RMSE is minimum) when number of decomposition nodes is '9', although at sixth decomposition level, total number of nodes is '64'. The optimal selection of the decomposition nodes not only enhances the prediction accuracy but also reduces the complexity of the prediction model by decreasing the number of decomposition nodes.

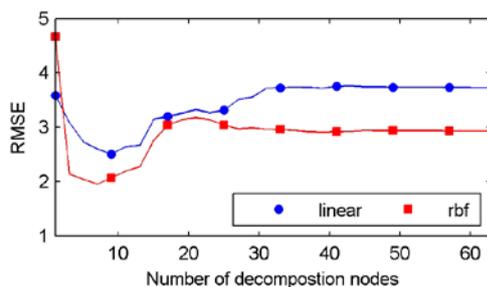


Figure 11. RMSE vs. decomposition nodes.

CONCLUSION

This paper proposes a novel technique for residual life prediction of the rotary equipment based on optimally parameterized wavelet transform and multi-step support vector regression. In optimally parameterized wavelet transform, a generalized criterion is proposed for optimal selection of the decomposition level as well as decomposition nodes. The criterion for decomposition level selection defines an upper limit on the decomposition level. Optimal decomposition nodes selection does not only enhances the prediction accuracy, but also reduces the complexity. The 'multi-step' regressor further enhances the prediction accuracy as compared to the conventional SVR based prediction techniques. The over-

all performance of the RWMS in terms of prediction accuracy justifies the efficacy of the proposed scheme in industrial processes to predict the residual life of the equipment.

REFERENCES

Berenji, H. R. & Yan, W. 2006. Wavelet Neural Networks for Fault Diagnosis and Prognosis. *IEEE International Conference on Fuzzy Systems*, 1334-1339.

El Hachemi Benbouzid, M. 2000. A review of induction motors signature analysis as a medium for faults detection. *IEEE Transactions on Industrial Electronics*, 47, 984-993.

Eren, L., Cekic, Y. & Devaney, M. J. 2010. Enhanced feature selection from wavelet packet coefficients in fault diagnosis of induction motors with artificial neural networks. *IEEE Instrumentation and Measurement Technology Conference (I2MTC)*, 960-963.

Eren, L. & Devaney, M. J. 2004. Bearing damage detection via wavelet packet decomposition of the stator current. *IEEE Transactions on Instrumentation and Measurement*, 53, 431-436.

Heng, A., Zhang, S., Tan, A. C. C. & Mathew, J. 2009. Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*, 23, 724-739.

Hsu, C. W., Chang, C. C. & Lin, C. J. 2003. A practical guide to Support Vector Classification, *Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei*.

Hu, Q., He, Z., Zhang, Z. & Zi, Y. 2007. Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble. *Mechanical Systems and Signal Processing*, 21, 688-705.

Jianhua, Z., Zhixin, Y. & Wong, S. F. 2010. Machine condition monitoring and fault diagnosis based on support vector machine. *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2228-2233.

Kim, H., Tan, A. C. C., Mathew, J., Kim, E. Y. H. & Choi, B. 2009. Prognosis of bearing failure based on health state estimation. *Proceedings of the 4th World Congress on Engineering Asset Management*.

Kwan, C., Zhang, X., Xu, R. & Haynes, L. 2003. A novel approach to fault diagnostics and prognostics. *IEEE International Conference on Robotics and Automation*, 1, 604-609.

Lau, E. C. C. & Ngan, H. W. 2010. Detection of Motor Bearing Outer Raceway Defect by Wavelet Packet Transformed Motor Current Signature Analysis. *IEEE Transactions on Instrumentation and Measurement*, 59, 2683-2690.

Morel, J. 2002. Vibratory monitoring and predictive maintenance. *Techniques de l'Ingénieur, Measurement and Control*, RD.

Peng, Z. K. & Chu, F. L. 2004. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mechanical Systems and Signal Processing*, 18, 199-221.

Platt, J. C. 1999. *Probabilities for SV Machines*, MIT Press.

Pugno, N., Ciavarella, M., Cornetti, P. & Carpinteri, A. 2006. A generalized Paris' law for fatigue crack growth. *Journal of the Mechanics and Physics of Solids*, 54, 1333-1349.

- Qiu, H., Lee, J., Lin, J. & Yu, G. 2003. Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. *Advanced Engineering Informatics*, 17, 127-140.
- Samanta, B. & Nataraj, C. 2009. Prognostics of Machine Condition Using Energy Based Monitoring Index and Computational Intelligence. *Journal of Computing and Information Science in Engineering*, 9, 044502-6.
- Shao, Y. & Nezu, K. 2000. Prognosis of remaining bearing life using neural networks. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 214, 217-230.
- Sotiris, V. & Pecht, M. 2007. Support Vector Prognostics Analysis of Electronic Products and Systems. *AAAI Fall Symposium on Artificial Intelligence for Prognostics.*, 120-127.
- Teotrakool, K., Devaney, M. J. & Eren, L. 2009. Adjustable-Speed Drive Bearing-Fault Detection Via Wavelet Packet Decomposition. *IEEE Transactions on Instrumentation and Measurement*, , 58, 2747-2754.
- Tran, V. T., Yang, B.-S., Oh, M.-S. & Tan, A. C. C. 2008. Machine condition prognosis based on regression trees and one-step-ahead prediction. *Mechanical Systems and Signal Processing*, 22, 1179-1193.
- Vapnik, V., Levin, E. & Cun, Y. L. 1994. Measuring the VC-dimension of a learning machine. *Neural Comput.*, 6, 851-876.
- Walker, J. S. 1999. *A Primer on Wavelets and their Scientific Applications*, New York, Chapman & Hall/CRC.
- Wang, P. & Vachtsevanos, G. 2002. Prognosis of remaining bearing life using neural networks. *AI EDAM*, 15, 349-365.
- Yaqub, M. F., Gondal, I. & Kamruzzaman, J. 2011. Machine Fault Severity Estimation Based on Adaptive Wavelet Nodes Selection and SVM (Accepted for publication). *IEEE International Conference on Mechatronics and Automation*.
- Yaqub, M. F., Gondal, I. & Kamruzzaman, J. 2011b. Machine Health Monitoring Based on Stationary Wavelet Transform and 4th Order Cumulants (Accepted for publication). *Australian Journal of Electrical & Electronics Engineering*.
- Yaqub, M. F., Gondal, I. & Kamruzzaman, J. 2011c. Resonant Frequency Band Estimation using Adaptive Wavelet Decomposition Level Selection (Accepted for publication). *IEEE International Conference on Mechatronics and Automation*.
- Yaqub, M. F., Gondal, I. & Kamruzzaman, J. 2011d. Severity Invariant Feature Selection for Machine Health Monitoring. *International Review of Electrical Engineering*, 6, 238-248.
- Yaqub, M. F., Gondal, I. & Kamruzzaman, J. 2011e. Severity Invariant Machine Fault Diagnosis (Accepted for publication). *IEEE International Conference on Industrial Electronics and Application*.
- Yen, G. Y. & Kuo-Chung, L. 2009. Wavelet packet feature extraction for vibration monitoring. *IEEE International Conference on Control Applications*, 2, 1573-1578.
- Zhang, L., Xiong, G., Liu, H., Zou, H. & Guo, W. 2010. Applying improved multi-scale entropy and support vector machines for bearing health condition identification. *Proceedings of the Institution of Mechanical Engineers Part C- Journal of Mechanical Engineering Science*, 224, 1795-1795.
- Zhang, L., Xiong, G., Liu, H., Zou, H. 2010. Bearing condition identification using improved multiscale entropy and SVMs. *Key Engineering Materials*, 817-820.
- Zhao, F., Chen, J. & Xu, W. 2009. Condition prediction based on wavelet packet transform and least squares support vector machine methods. *Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering*, 223, 71-79.