A Neural Network Approach for Remaining Useful Life Prediction Utilizing Both Failure and Suspension Histories

Zhigang Tian[a,*], Lorna Wong[b], Nima Safaei[b]

[a] *Concordia Institute for Information Systems Engineering, Concordia University*

1515 Ste-Catherine Street West EV-7.637, Montreal, H3G 2W1, Canada

[b] *Department of Mechanical and Industrial Engineering, University of Toronto*

5 King's College Road, Toronto, M5S 3G8, Canada

Abstract

Artificial neural network (ANN) methods have shown great promise in achieving more accurate equipment remaining useful life prediction. However, most reported ANN methods only utilize condition monitoring data from failure histories, and ignore data obtained from suspension histories in which equipments are taken out of service before they fail. Suspension history condition monitoring data contains useful information revealing the degradation of equipment, and will help to achieve more accurate remaining useful life prediction if properly used, particularly when there are very limited failure histories, which is the case in many

* Corresponding author. Tel.: 1-514-848-2424 ext. 7918; Fax: 1-514-848-3171.
Email address: tian@ciise.concordia.ca (Zhigang Tian).

applications. In this paper, we develop an ANN approach utilizing both failure and suspension condition monitoring histories. The ANN model uses age and condition monitoring data as inputs and the life percentage as output. For each suspension history, the optimal predicted life is determined which can minimize the validation mean square error in the training process using the suspension history and the failure histories. Then the ANN is trained using the failure histories and all the suspension histories with the obtained optimal predicted life values, and the trained ANN can be used for remaining useful life prediction of other equipments. The key idea behind this approach is that the underlying relationship between the inputs and output of ANN is the same for all failure and suspension histories, and thus the optimal life for a suspension history is the one resulting in the lowest ANN validation error. The proposed approach is validated using vibration monitoring data collected from pump bearings in the field.

Keywords: Remaining useful life; Prediction; Artificial Neural Networks; Suspension history; Bearing

1. Introduction

Accurate equipment remaining useful life (RUL) prediction is the key to effective implementation of condition based maintenance (CBM), which aims to prevent unexpected failures and minimizing overall maintenance costs [1]. Condition monitoring data, such as vibration data, oil analysis data and acoustic emission data, are collected and processed, and used for predicting the RUL. Existing equipment RUL prediction methods can be roughly classified into model-based (or physics-based) methods and data-driven methods [1-3]. The model-based methods predict the remaining useful life using the physical models of the components and damage propagation models based on damage mechanics [4]. A method for health condition prediction of propulsion system bearings was developed by Marble et al [5], based on a bearing spall propagation physical model and a finite element model. Methods for RUL prediction of gears were proposed by Kacprzynski et al [6] and Li and Lee [7]. If properly used, physics-based models can greatly improve the RUL prediction accuracy. However, equipment dynamic response and damage propagation processes are typically very complex, and authentic physics-based models are very difficult to build. For example, for a variety of components, structures and systems, accurate models are not available for modeling the vibration dynamic response with respect to different levels and different types of damages. In these cases, model-based methods cannot be effectively implemented without accurate physical models.

Data-driven methods utilize collected condition monitoring data for RUL prediction, and do not require physical models. The data-driven methods are based on the understanding that condition monitoring data and the extracted features will change during the damage initiation and propagation process, or the degradation process. The data-driven methods aim at modeling the relationship between equipment age, condition monitoring data, and equipment degradation and remaining useful life. Jardine et al developed the Proportional Hazards Model approach for CBM, where health condition indicators are predicted using the transition probability matrix [1, 8]. Artificial neural networks (ANNs) have been considered to be very promising tools for equipment RUL prediction because of their adaptability, nonlinearity, and arbitrary function approximation ability [9-12]. Lee et al [13] presented an Elman neural network method for health condition prediction. Gebraeel et al developed ball bearing remaining life prediction methods based on feedforward neural networks [14-15], in which the output of the ANN model was a condition monitoring measurement, such as overall vibration magnitude. A disadvantage of using such an output is that the failure threshold values are typically hard to clearly define in practical applications. Wu et al [16] proposed another ANN based RUL prediction model, where the ANN output was the life percentage, or in another word, one minus the remaining life percentage. An extended version of the model was presented by Tian [17], which can deal with data that are not equally spaced and involve multiple measurement dimensions, and was applied to the RUL prediction of field bearings.

A history of a unit refers to the period from the beginning of its life to the end of its life, failure or suspension, and the inspection data collected during this period. Thus, condition monitoring data consists of failure histories and suspension histories. In a failure history, a component ends up with a failure and it is replaced with a new component. In a suspension history, the component is taken out of service, i.e., replaced by a new component, before it fails and never used again in the equipment. Thus, a history is called a suspension history if the component is replaced with a new one during planned maintenance or inspection. If a component is replaced because other components in the system are damaged, we also call it a suspension history since the component has not failed when it is replaced. However, if a component is just temporarily stopped (or removed) and later put back into the system (after the planned maintenance or after other components are replaced), we do not call it a suspension history, and we should continue collecting condition monitoring data from the component until it fails or it is suspended and replaced with a new component. Suspension history condition monitoring data contains useful information revealing the degradation of equipment, and will help to achieve more accurate remaining useful life prediction if properly used. Due to shorter cycles of new product introduction or shorter cycles of applying a component within a specific configuration, and the fact that industry tends not to allow their equipments to fail, we typically have a small set of failure histories and a larger set of suspension histories. Thus it is particularly critical to fully utilize suspension histories if only very few failure histories are available.

In conventional reliability analysis, only failure times and suspension times are used to construct the likelihood function for estimating the lifetime distribution probability density function, typically the Weibull distribution is used in reliability analysis, and thus for predicting future reliability and expected time to failure [18]. Lu & Meeker [19] developed a two-stage statistical method for estimating the fix-effect parameters and random-effect parameters in the remaining life distribution model, using degradation data in both failure and suspension histories. The proportional hazards model can also incorporate both failure and suspension history data in estimating the model parameters. However, the time to failure distribution is relatively wide, and the state transition rates are hard to accurately predict if a large amount of condition monitoring data is not available [1, 8, 20]. Very little work has been reported on ANN methods for RUL prediction utilizing suspension histories. The only reported work we can find is a recent paper by Heng et al [2]. The ANN model in the paper uses condition monitoring measurements at several past time points as inputs, and the outputs are the reliability values, or probability of survival, at future time points. However, the outputs are difficult to interpret and convert to the predicted failure time for the unit.

In this paper, we propose an approach based on the ANN model we developed before, which takes age and condition monitoring data as inputs and the life percentage as output. The biggest challenge of using this model to deal with suspension data is to estimate the optimal failure times for each suspension history so that they can be used to train the ANN model. In the proposed approach, we address this issue based on the understanding that the underlying relationship between the inputs and output of ANN is the same for all failure and suspension histories, and thus the optimal life for a suspension history is the one resulting in the lowest ANN validation error.

The remainder of this paper is organized as follows. The ANN model and the training algorithm are presented in Section 2. The proposed ANN method is discussed in Section 3. Section 4 presents the case study, in which the proposed ANN method is validated using real-world vibration monitoring data collected in the field from pump bearings. Conclusions are given in Section 5.

Notations:

F:	Number of failure histories
S:	Number of suspension histories
NF_f:	Number of inspection points in failure history *f*.
NS_s:	Number of inspection points in suspension history *s*.
J:	The number of condition monitoring measurements
TF_f:	Failure time for failure history *f*.
$z_{f,i}^{j}$:	The measurement j at time t_i in failure history f.
TS_s:	Suspension time for suspension history *s*.
$TSD_{s,l}$:	Possible failure time value *l* for suspension history *s*
$t_{s,i}$:	Equipment age at inspection point *i* in suspension history *s*.

TS_s^*: The optimal failure time for suspension history s

2. The Artificial Neural Network Model

In this section, we present the ANN model used in this work. The structure of the ANN model was proposed by Tian [17]. But the training algorithm and the stopping criterion are different in this work, so that the trained ANN model can be more effectively used in the proposed remaining life prediction approach.

2.1 The ANN model

The ANN model proposed by Tian [17] is used in the work. The structure of the ANN model is shown in Fig. 1. The ANN model is a feedforward neural network model with an input layer, an output layer with one output neuron, and two hidden layers. We use an ANN with two hidden layers instead of one because we find it is able to produce more reliable results according to our experiments. That is, using the same ANN structure and the same data, we can get different training and prediction results when we train the ANN for different times because of the randomness in the training algorithm. According to our experiments, such differences are smaller for ANNs with two hidden layers, comparing to ANN with one hidden layer.

The inputs to the ANN include the age values and the condition monitoring measurements at the current inspection point and those at the previous inspection point. An example of the ANN structure, with two condition monitoring measurements, is shown in Fig. 1. Specifically, t_i is the age of the equipment at the current inspection point i, and t_{i-1} is the age at the previous inspection point $i-1$; z_i^1 and z_{i-1}^1 are the values of measurement 1 at the current and previous inspection points, respectively; z_i^2 and z_{i-1}^2 are the values of measurement 2 at the current and previous inspection points, respectively. Using data at the two time points allows the use of the information about the rate of change of measurement values. We only use data at the current and previous inspection points because data at these two points is able to reflect the health condition of the equipment, and more importantly, more robust ANN models with better generalization capability can be produced by effectively limiting the number of input parameters and thus the trainable weights. Wu et al [16] used equally spaced bearing vibration data collected in the lab environment, and they stated that according to their experiments, using two time points could lead to better results than using more complete monitoring data set. We also compared the option of using two time points and that using three time points in the ANN input layer, which kind of provides information on the "accelerations" of the measurement values, using our experiment data, and found that ANN using two time points is able to produce slightly more accurate prediction results. In addition, it is more computationally efficient to use ANN using data at two time points. The age values at the current and previous time points are used, so that we can handle condition monitoring data collected at discrete inspection time points that are not equally spaced, such as the vibration monitoring data for many pump bearings, and the oil analysis data for truck transmissions [8]. More measurements can be incorporated by adding more input nodes to the ANN model. For example, if we have totally I significant condition monitoring measurements to be incorporated in the ANN model, the total number of input nodes will be totally $(2 + 2I)$.

Fig. 1. Structure of the ANN model for remaining useful life prediction [17]

The output of the proposed ANN model is the life percentage, denoted by P_i. As an example, suppose the age of a bearing at the time of failure is 511 days and, at an inspection point i, the age is 400 days, then the life percentage at inspection point i would be:

$$P_i = 400/511 \times 100\% = 78.3\%. \tag{1}$$

It is true that the health condition of a piece of equipment deteriorates with time. For example, the propagation of a spall in a bearing, or the propagation of a root crack or the surface wear in a gear, is generally a monotonic process. Thus, without loss of generality, we can assume that the true inherent health condition index increases monotonically with time. Life percentage is an excellent option for indicating the inherent health condition index of a piece of equipment due to the following two reasons: (1) the mapping between the inherent health condition index and the life percentage is monotonically non-decreasing, and (2) life percentage is also able to indicate when the failure occurs, that is, the failure occurs when the life percentage reaches 100%.

The transfer function for the neurons in two hidden layers is the hyperbolic tangent sigmoid transfer function, which, for hidden neuron j, takes the form shown in Equation (2):

$$y_j = 2/(1 + e^{-2N_j}) - 1, \tag{2}$$

$$N_j = \sum_{k=1}^{K_j} w_{kj} y_k + \delta_j. \tag{3}$$

where y_j is the output value of neuron j, K_j is the number of neurons with output connections to neuron j, w_{kj} is the weight value of the connection from neuron k to neuron j, and δ_j is the bias value of neuron j. The transfer function for the neuron in the output layer is the linear transfer function, which, for output neuron j, takes the following form:

$$y_j = N_j, \tag{4}$$

where N_j takes the form given in Equation (3).

Thus, once the weight and bias values are determined during the training process, given an input vector with age and condition monitoring measurement values, the life percentage can be calculated using Equations (2)-(4).

2.2 The ANN training process

2.2.1 The training algorithm

During the training process, based on the training data set including a set of input vectors and the corresponding output values, the weights and the bias values of the ANN model are adjusted to minimize the error between the model outputs and the actual outputs. The performance function, or error function, is to be minimized during the training process. The typical performance function takes the following form:

$$F = mse = \frac{1}{N}\sum_{k=1}^{N}(e_k)^2 = \frac{1}{N}\sum_{k=1}^{N}(y_k - d_k)^2, \tag{5}$$

where N is the number training input and output pairs. d_k is the actual output. y_k is the model output, which are calculated using Equations (2)-(4), and e_k is the corresponding output error.

There are many algorithms available for training feedforward neural networks, such as the standard backpropagation algorithm, and the Levenberg-Marquardt (LM) algorithm [21]. However, as to be discussed later in this paper, the validation mechanism, or the early stopping method, will be used in the training process to improve the network generalization, and it has been reported that the resilient backpropagation (RPROP) algorithm is a method that works well with the validation mechanism [22]. The resilient backpropagation algorithm is a first-order optimization algorithm for supervised learning in ANN, and it was developed by Riedmiller and Braun in 1993 [23]. The purpose of the resilient backpropagation training algorithm is to eliminate the harmful effects of the magnitudes of the partial derivatives. Only the sign of the derivative is used to determine the direction of the weight update, and the size of the weight change is determined by a separate update value [22]. Because the importance of the validation mechanism, as to be discussed in details later, in this work, we use the resilient backpropagation algorithm to train the ANN.

2.2.2 The validation mechanism

The objective of the training process is to model the mapping between the input vector and the output without modeling the noise in the data. The problem of "overfitting" may occur during the ANN training. That is, the error on the training set is driven to a very small value, but when new data is presented to the trained network, the error is large. It is critical to ensure the generalization performance of the trained network, particularly for the application of remaining life prediction in this work.

There are basically two methods for improving network generalization [22]. One method is the use of validation data set. That is, the available data is divided into the training set and the validation set. Data in the training set is used to adjust the ANN weights and bias values, while the data in the validation set is not. During the ANN training process, the mean square error for the training set and that for the validation set will drop early in the training process, because the ANN is learning the relationship between the inputs and the outputs by modifying the trainable weights based on the training set. After a certain point, however, the mean square error for the validation set will start to increase, because the ANN starts to model the noise in the training set. The training process can be stopped at this point, and a trained ANN with good modeling and generalization capability can be achieved.

The other method for improving network generalization is called regularization. The performance function is modified in this method, so that we not only minimize the mean square error, but also minimize the sum of squares of the network trainable weights. The two minimization objectives are combined using the performance ratio γ, and the performance function is shown as follows.

$$F = \gamma \cdot mse + (1-\gamma) \cdot msw = \gamma \frac{1}{N} \sum_{k=1}^{N} (y_k - d_k)^2 + (1-\gamma) \frac{1}{J} \sum_{j=1}^{J} w_j^2, \qquad (6)$$

where msw represents the sum of squares of the network trainable weights, w_j denotes weight j, and J denotes the total number of trainable weights. The problem with this method is that the optimum value of the performance ratio is difficult to determine. Another regularization method is Bayesian regularization, in which the optimum performance ratio is determined in an automatic fashion during the training process [24].

In this work, in the training process, we need a performance measure to indicate how good the trained ANN is, given a certain set of available data. The mean square error (MSE) on the training set cannot be used as the measure, because a lower training MSE does not necessarily correspond to a better network due to the generalization issue. The Bayesian regularization cannot be used either because the optimum performance ratios are likely different if the same data is used to train the network several times, and thus the overall MSE cannot give an indication of which trained network is the best.

Thus, in this work, we use the validation data set to improve the network generalization. The available data is divided into the training set and the validation set. Based on reported results, we use 1/3 of the available data as the validation set and the other 2/3 as the training data set. The MSE on the validation data set, i.e., the validation MSE, is used as the performance measure to determine which trained ANN is the best among a number of networks trained using the same data sets. Compared to other measures, the validation MSE is the best measure for this purpose. Generally, the lower the validation MSE, the better the trained network.

3. The Proposed ANN Remaining Useful Life Prediction Approach Utilizing Failure and Suspension Histories

3.1 Overview

Most of the reported ANN remaining life prediction methods ignore suspension histories and only use failure histories. However, suspension condition monitoring histories contain valuable information reflecting the degradation of equipment, and can lead to more accurate remaining life prediction if properly utilized. From the perspective of remaining useful life prediction, there are two key pieces of valuable information associated with a suspension history:

(1) the suspension time: the equipment is taken out of service at the suspension time while it has not failed. This information can be used as a constraint when we build the ANN prediction model. That is, when the ANN prediction model is applied to the condition monitoring data in the suspension history, the predicted failure times are supposed to be beyond the suspension time.

(2) the degradation progression reflected in the condition monitoring data: a piece of

equipment degrades over time, and the age and condition monitoring data can reflect the degradation level of the equipment. The degradation progression reflected in the suspension history can help to improve the prediction accuracy of the ANN model.

The two key pieces of information mentioned above can help to establish the relationship between the collected age and condition monitoring data and the remaining useful life. Often times, we have a small number of failure histories and more than twice as many suspension histories. The failure times, as well as the age and condition monitoring data, can be quite different among different units of the equipment. By properly utilizing the suspension histories, we can have more data to model the relationship between the collected age and condition monitoring data and the remaining useful life, and thus achieve more accurate remaining useful life prediction.

As discussed in Section 2, the ANN model takes the age and condition monitoring data as inputs and the life percentage as output. The life percentage value for a certain inspection point is equal to the age value divided by the failure time of the equipment. Thus, the key challenge in utilizing data from a suspension history is that the failure time is unknown, without which we cannot determine the life percentage values for the suspension history in order to train the ANN model. In this work, we address this challenge by first determining the optimal failure time for each suspension history. The key idea is that the underlying degradation relationship between the inputs and output of the ANN is the same for all failure and suspension histories. We can specify a failure time for a suspension history, and train the ANN using the training set constructed based on that suspension history and all of the failure histories. For the suspension history, the optimal failure time is supposed to correspond to the trained ANN with the best training performance. As discussed in Section 2, the validation MSE can be used as the measure to indicate the training performance of a trained ANN. Thus, the optimal failure time for a suspension history can be obtained by finding the failure time corresponding to the lowest ANN validation MSE.

The procedure of the proposed remaining life prediction approach is shown in Fig. 2. The details of the approach are presented in the following subsections.

Fig. 2. Procedure of the remaining life prediction approach

3.2 Constructing the failure history training data set

The first step of the approach is to construct the failure history training data set, which will be combined with training data set based on the suspension histories to train the ANN. Suppose there are J condition monitoring measurements used in the ANN model. An ANN input vector based on failure history f takes the following form:

$$\boldsymbol{IN} = \left(t_{f,i-1}, t_{f,i}, z_{f,i-1}^{1}, z_{f,i}^{1}, z_{f,i-1}^{2}, z_{f,i}^{2}, \dots, z_{f,i-1}^{J}, z_{f,i}^{J}\right), \tag{7}$$

where $t_{f,i}$ denotes the equipment age at inspection point i in failure history f, and $z_{f,i}^{j}$ represents the measurement j at time $t_{f,i}$. The corresponding output value is:

$$P_{f,i} = \frac{t_{f,i}}{TF_f} \ , \tag{8}$$

where TF_f represents the failure time for failure history f. Thus, the total number of input/output pairs based on the failure histories is:

$$N_F = \sum_{f=1}^{F} \left(NF_f - 1\right) \ , \tag{9}$$

3.3 Finding the optimal failure time for a suspension history

As discussed before, the optimal failure time for a suspension history corresponds to the lowest validation MSE if we train the ANN using the training set constructed based on this suspension history and all the failure histories.

For suspension history s, we specify L discrete possible failure time values, and obtain the corresponding ANN validation MSE values. The discrete failure time values are denoted by $TSD_{s,1}$, $TSD_{s,2}$, …, $TSD_{s,L}$, respectively. These values are selected based on the suspension time for the history, TS_s. Specifically, we can have $TSD_{s,l} \geq TS_s$ $(1 \leq l \leq L)$ for most of the failure time values, and have 1-2 values smaller than TS_s, so that we can find the optimal failure time based on the validation MSE values at these discrete points.

For a certain failure time value $TSD_{s,l}$, we can obtain the ANN input/output pairs for suspension history s. The input vectors take the same form as that for failure histories, given in Equation (7). The ANN output value corresponding to the ith inspection point is given as:

$$P_{s,i,l} = \frac{t_{s,i}}{TSD_{s,l}} \ , \tag{10}$$

where $t_{s,i}$ denotes the equipment age at inspection point i in suspension history s. The ANN input/output set includes the input/output pairs based on suspension history s and the input/output data set constructed based on all the failure histories. Thus, the total number of input/output pairs is:

$$N_{Ss} = NS_s - 1 + \sum_{f=1}^{F}\left(NF_f - 1\right) \ , \tag{11}$$

where NS_s represents the total number of inspection points in suspension history s. The input/output set is further divided into the ANN training set and the ANN validation set: 2/3 of the input/output pairs for the training set and 1/3 for the validation set. Specifically, we go through the suspension history and the failure histories, and select an input/output pair in every three input/output pairs to construct the ANN validation set. The ANN is trained using the resilient backpropagation algorithm based on the training set and the validation set, and the validation MSE can be obtained. Because of the randomness in the training algorithm, typically we cannot obtain the exactly same validation MSE value each time. Thus, in this work, we train the ANN 30 times, and record the 3 lowest, or best, validation MSE values for future use, which are denoted by $ve_{l,r}^{s}$ $(r = 1, 2, 3)$.

The ANN validation MSE values $ve_{l,r}^{s}$ $(r = 1, 2, 3)$ can be obtained for all the discrete failure time values $TSD_{s,l}$ $(1 \leq l \leq L)$ for suspension history *s*. Thus, we can obtain totally $3L$ data points, each containing a validation MSE value and the corresponding failure time. In order to find the optimal failure time based on the discrete validation MSE values, we need to fit these validation MSE data points. Considering the flexibility and the ability to model simple trends, we use the third order polynomial to fit the data points:

$$y = ax^3 + bx^2 + cx + d, \tag{12}$$

where *y* represents the validation MSE, *x* represents failure time, and *a*, *b*, *c*, *d* represent the polynomial coefficients to be determined. Once the polynomial function is obtained, it is easy to find the optimal failure time corresponding to the lowest ANN validation MSE, using a simple optimization process. The optimal failure time for suspension history *s* is denoted by TS_s^*. To enforce the suspension time constraint, let $TS_s^* = TS_s$ if TS_s^* is smaller than the suspension time.

3.4 ANN Training based on the suspension histories with optimal failure times and the failure histories

Using the procedure in Section 3.3, we can find the optimal failure times for all the suspension histories: TS_s^* $(1 \leq s \leq S)$. Now we can train the ANN for remaining useful prediction based on the suspension histories with optimal failure times and the failure histories. The form of an ANN input vector is given in Equation (7). The ANN output value for an input/output pair from a failure history is given by Equation (8), and that from a suspension history is given as follows:

$$P_{s,i} = \frac{t_{s,i}}{TS_s^*} \tag{13}$$

Thus, the total number of input/output pairs is:

$$N_{\mathrm{IO}} = \sum_{s=1}^{S}(NS_s - 1) + \sum_{f=1}^{F}\left(NF_f - 1\right) \ , \tag{14}$$

The ANN training set includes 2/3 of the input/output pairs, and the ANN validation set includes the remaining 1/3 of the input/output pairs. Similarly, we train the ANN 30 times using the resilient backpropagation algorithm, and save the ANN with the smallest validation MSE.

3.5 Remaining life prediction using trained ANN

Once the ANN is trained, as discussed in the previous section, it can be used for RUL prediction for other equipments being monitored, as shown in Fig. 2. The age and condition monitoring measurements at the current and previous data points are used as inputs to the trained ANN, and the current life percentage can be obtained. The RUL is obtained by dividing the current age by the predicted life percentage. When new condition monitoring data is available, the prediction will be performed again and the RUL will be updated. The remaining useful life prediction process stops when the equipment fails or when it is preventively taken out of service.

4. Case Studies

4.1 Case study introduction

Condition monitoring data collected from the field is used to validate the proposed ANN approach for RUL prediction utilizing both failure and suspension histories, particularly when there are only few failure histories available. The condition monitoring data were collected from bearings on a group of Gould pumps at a Canadian kraft pulp mill company [25]. In total, there are 10 bearing failure histories and 14 suspension histories available. Vibration monitoring data were collected from the pump bearings using accelerometers. The collected vibration measurements include the overall vibration magnitudes in the axial, horizontal and vertical directions, and in each of these directions, the vibration magnitude values are obtained in five frequency bands. In addition, the overall acceleration values are also measured in the three directions. Not all the measurements are significantly correlated with the degradation of the bearings. Significance analysis, which is built into the software EXAKT developed by OMDEC Inc., is utilized to identify the significant condition monitoring measurements [26]. Two measurements are identified to be significant: the vibration magnitude in the horizontal direction and that in the vertical direction in the 5th vibration frequency at the problematic bearing end of the pump. We refer to these two measurements as Measurement 1 and Measurement 2. The significance analysis result is shown in Fig. 3, where “P1H_Par5” refers to Measurement 1 and “P1V_Par5” refers to Measurement 2, and both the “Sign.” column and the “p-value” column show that these two measurements are significant.

Fig. 3. Significance analysis result using EXAKT

Based on the data available, we study two cases. In the first case, we investigate the situation where there are very few failure histories: among the 10 failure histories, we use 2 of them as failure histories, and 4 of them to generate 4 suspension histories to build the ANN prediction model. In the second case, we use 5 of the 10 failure histories as failure histories and use 10 actual suspension histories to illustrate the proposed ANN prediction approach.

We compare the prediction performance of the proposed approach and the ANN method using only the failure histories. Several failure histories are used as test histories to test the prediction performance. The Average Prediction Error, denoted by $\bar{e}$, is used to quantify the prediction performance:

$$\bar{e} = \frac{1}{n} \cdot \sum_{k=1}^{n} \left| P_k - \hat{P}_k \right| \cdot 100\%, \tag{15}$$

where n is the number of inspection points for testing the prediction performance, P_k is the actual life percentage at inspection point k, and $\hat{P}_k$ is the predicted life percentage at inspection point k. Due to the randomness in the training algorithm, the resulting trained ANN model, and thus the prediction results, will not be the same each time we run the algorithm. Thus, we repeat the ANN training and prediction process 10 times, and use the average prediction errors as the as the prediction error for the inspection points in the test history. Specifically, for an inspection point at which the prediction performance is tested, the absolute value of the prediction error is calculated as:

$$\left|P_k - \hat{P}_k\right| = \frac{1}{10}\sum_{r=1}^{10}\left|P_k - \hat{P}_{kr}\right|, \tag{16}$$

where k is the inspection point index, P_k is the actual life percentage at inspection point k, and $\hat{P}_{kr}$ is the predicted life percentage when we run the ANN training and prediction process the rth time. For each of the test histories, we test the prediction performance starting from inspection point 6, since failures are not expected very early in the life.

The prediction accuracy late in the life of the equipment is more important than that early in its life, because it will more likely affect the decision on whether or not preventive replacement should be performed at the current inspection point. To investigate the prediction accuracy late in the unit life, we test the prediction performance at the last 10 inspection points for each test failure history. $\bar{e}_{\mathrm{L10}}$ is used to represent the Average Prediction Error considering only the last 10 inspection points for each test history.

4.2. Case study 1

In this case study, we investigate the situation where there are very few failure histories. It is supposed that only 2 failure histories are available, in which there are totally 37 inspection points. Among the 10 failure histories, another 4 of them are used to generate 4 suspension histories, and the remaining 4 are used as test histories to test the prediction performance. For the 4 failure histories used for generating suspension histories, the actual failure times are given in Table 1. The suspension times are randomly selected, but basically based on the principle that if the vibration magnitude measurements are too high, say over 0.045, the unit is taken out of service. The suspension times are also given in Table 1.

4.2.1 Prediction results using only the failure histories

As mentioned before, most reported ANN RUL methods only use failure histories for prediction while ignoring the suspension histories. In this section, we use data from the 2 failure histories to train the ANN, and test the prediction performance using the 4 test histories. The ANN used in the case study has 6 input neurons, since there are two significant condition monitoring measurements. The ANN has two hidden layers with 2 hidden neurons in each hidden layer, resulting in totally 23 trainable weights. The total number of inspection points in the two failure histories is 37, giving a total of 34 ANN input/output pairs, according to Equation (9). The ANN training set and the validation set are constructed from the input/output pairs. The ANN is trained using the resilient backpropagation algorithm 30 times, and the ANN with the smallest validation MSE is saved for prediction performance testing.

For the four test histories, there are a total of 127 inspection points at which the prediction performance is tested. The prediction results with the ANN method using failure histories only are shown in Table 2, where $\bar{e}_{\mathrm{All}}$ is the overall average prediction error, STD_{All} denotes the standard deviation of the prediction error, $\bar{e}_{\mathrm{L10}}$ is the average prediction error considering only the last 10 inspection points for each test history, and STD_{L10} is the corresponding standard deviation. The results show that the overall average prediction error is 22.94%, and the standard deviation of the prediction error is 10.47%. It can be seen that if we only use the two failure histories to train the ANN, the prediction error is relatively large. The relatively large standard deviation value suggests that the prediction accuracy is not very stable.

We investigated the prediction performance at the last 10 inspections points in each test history, which accounts for about 1/3 of the total test history inspection points. The remaining lives at these inspection points are between 291 days and 0 day, and the life percentage values are between 80.2% and 100%. It can be seen from Table 2 that the average prediction error, 8.18%, and the prediction error standard deviation, 6.67%, are significantly better than those considering all the inspection points. This illustrates the ability of the ANN model to achieve more accurate prediction late in the life of the equipment.

4.2.2 Prediction results using the proposed ANN approach

In this section, the prediction results using the proposed ANN approach are presented. The suspension times for the 4 suspension histories are shown in Table 1. For a certain suspension history *s*, we investigate the following 7 possible failure time values: $TS_s - 300$, $TS_s - 150$, TS_s, $TS_s + 150$, $TS_s + 300$, $TS_s + 450$, and $TS_s + 600$, that is, remaining life values -300, -150, 0, 150, 300, 450 and 600 days. We investigate a relatively small and sparse set of failure time values so as to detect the obvious changes and trends in validation MSE. The possible failure times $TS_s - 300$ and $TS_s - 150$ are also investigated for the purpose of establishing the trend in the validation MSE change, and for finding the optimal failure time for a suspension history.

Following the procedure described in Section 3, for each possible failure time, the ANN is trained 30 times using the resilient backpropagation algorithm, and the 3 lowest validation MSE values are recorded. Thus, for suspension history *s*, we can obtain 21 "validation MSE v.s. remaining life" points. The parameters for the third order polynomial function, as shown in Equation (12), can be obtained by fitting these points. Applying a simple optimization procedure to the third polynomial function, say the Newton's method, the optimal remaining life corresponding to the lowest validation MSE can be obtained. If the optimal remaining life is found to be smaller than 0, we let the optimal remaining life be equal to 0. The results for the four suspension histories are summarized in Fig. 4. The ANN validation MSE values at the 7 remaining life points are denoted by “*”. It can be observed that the validation MSE value changes with respect to the remaining life for a suspension history. Specifically, the validation MSE generally exhibits the trend of decreasing with the remaining life first and then starting to increase with it, which indicates that an optimal remaining life value that corresponds to the lowest validation MSE can be identified. For a suspension history, the fitted third order polynomial function is plotted in the figure as a solid curve, and the optimal remaining life is denoted in the figure by “o”. The optimal failure times for the four suspension histories are found to be 567 days, 495 days, 591 days and 1001 days, respectively.

Fig. 4. The validation MSE and optimal remaining life values for the suspension histories in Case 1

Using the training set and validation set constructed based on the suspension histories with the optimal failure times and the failure histories, the ANN can be trained for RUL prediction. The prediction performance of the trained ANN is tested using the same four test histories at the 127 inspection points. The prediction results are shown in Table 2 too. It can be seen that both the overall prediction performance and the prediction performance at the last 10 inspection points in each test history are apparently improved, and this shows the ability of the proposed ANN approach for producing more accurate and stable prediction results. It is worth noting that this approach achieved very good prediction accuracy at the last 10 inspection points, with an average prediction error of 5.66%. The prediction accuracy improvement is due to fact that using the proposed approach, the ANN model is able to capture useful information in the four suspension histories which is not available in the two failure histories, and use the information to achieve more accurate prediction.

4.3. Case study 2

In this section, we study the case involving actual suspension histories and more failure histories. Among the 10 available failure histories, we use 5 of them as failure histories and the remaining 5 as test histories. We use 10 actual suspension histories, which is twice the number of the failure histories used.

4.3.1 Prediction results using the failure histories only

Similarly, we first investigate the ANN method using the 5 failure histories only. The total number of inspection points in the 5 failure histories is 116, which gives a total of 111 ANN input/output pairs, according to Equation (9). Because of the complex relationship to model and the availability of more inspection data, in this case we use an ANN with 3 neurons in the first hidden layer and 2 neurons in the second hidden layer, resulting in totally 31 trainable weights. The prediction performance of the trained ANN model is tested using the 5 test histories, and there are 156 inspection points at which the prediction performance is tested. The prediction results are shown in Table 3. Compared to the results in Table 2 using only two failure histories, the four prediction performance measures in the current results are improved. The prediction accuracy improvement results from the fact that more failure history data is available for training the ANN model in this case.

4.3.2 Prediction results using the proposed ANN approach

10 actual suspension histories are used, and the suspension times range from approximately 500 days to 1400 days. Similar to Case study 1, for a certain suspension history s, we investigate the following 7 possible failure time values: $TS_s - 300$, $TS_s - 150$, TS_s, $TS_s + 150$, $TS_s + 300$, $TS_s + 450$, and $TS_s + 600$, and we can obtain the "validation MSE v.s. remaining life" points. The results for the 10 suspension histories are shown in Fig. 5, where, for each suspension history, the ANN validation MSE values at the 7 remaining life points are denoted by "*", the fitted third order polynomial function is plotted in the figure as a solid curve, and the optimal remaining life is denoted in the figure by "o".

The prediction performance of the proposed approach is tested using the 5 test histories at the 156 inspection points, and the results are given in Table 3. As can be seen Table 3, the prediction results are significantly better than those obtained with the ANN method using failure histories only, in terms of the average prediction errors and the prediction error standard deviation values. This illustrates that significantly more accurate prediction results can be achieved by properly utilizing the suspension histories. Comparing to those in Table 2, the results with the proposed ANN approach in Table 3 are better because more failure histories and more suspension histories are available for constructing the prediction model. It is particularly worth pointing out, as shown in Table 3, that the proposed approach achieves excellent prediction performance over the last 10 inspection points in the test histories: the average prediction error is 3.65%, and the standard deviation of the prediciton error is 1.09%. The prediction results on a sample test history is shown in Fig. 6. This case study illustrates the capability of the proposed ANN approach to achieve more accurate and stable remaining life preciction.

Fig. 5. The validation MSE and optimal remaining life values for the suspension histories in Case 2

Fig. 6. Prediction results for a sample test history using the proposed approach

5. Conclusions

In this paper, we develop an ANN approach utilizing both failure and suspension condition monitoring histories. The ANN model uses age and condition monitoring data as inputs and the life percentage as output. For each suspension history, the optimal predicted life is determined which can minimize the validation mean square error in the training process using the suspension history and the failure histories. Then the ANN is trained using the failure histories and all the suspension histories with the obtained optimal predicted life values, and the trained ANN can be used for remaining useful life prediction of other equipment. The key idea behind this approach is that the underlying relationship between the inputs and output of ANN is the same for all failure and suspension histories, and thus the optimal life for a suspension history is the one resulting in the lowest ANN validation error. The proposed approach is validated using real-world vibration monitoring data collected from pump bearings in the field, and the case study shows that the proposed approach can produce more accurate remaining life prediction results.

A component may fail in different failure modes. In many practical situations we only have the collected condition monitoring measurements from these failure and suspension histories, but don't have more specific information about the failure modes of different components. However, if in some applications the failure mode and degradation process information is available for the failure and suspension histories, we can adapt our approach and train an ANN using the histories with the same degradation process, and apply the trained ANN to a component subject to the same degradation process if this information is also available. It will be very challenging, though, if we don't know the degradation process for a component being monitored. This is an interesting topic and worth further investigating in future research.

Acknowledgments

This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). We appreciate very much the help from OMDEC Inc. for providing the condition monitoring data used in the case study. We also greatly appreciate the help from Dr. Andrew Jardine at the Center for Maintenance Optimization & Reliability Engineering, University of Toronto. Comments and suggestions from the editor and anonymous reviewers are very valuable in the improvement of the quality of this paper.

References

[1]. A.K.S. Jardine, D.M. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mechanical Systems and Signal Processing 20 (7) (2006) 1483-1510.

[2]. A. Heng, A.C.C. Tan, J. Mathewa, N. Montgomery, D. Banjevic, A.K.S. Jardine, Intelligent condition-based prediction of machinery reliability, Mechanical Systems and Signal Processing (2009), doi:10.1016/j.ymssp.2008.12.006.

[3]. G. Vachtsevanos, F.L. Lewis, M. Roemer, A. Hess, B. Wu, Intelligent Fault Diagnosis and Prognosis for Engineering Systems, John Wiley & Sons, 2006.

[4]. D. J. Inman, Charles R. Farrar, Vicente Lopes Junior, Damage Prognosis: for Aerospace, Civil and Mechanical Systems, John Wiley and Sons, 2005.

[5]. S. Marble, B.P. Morton, Predicting the remaining life of propulsion system bearings, In: Proceedings of the 2006 IEEE Aerospace Conference, Big Sky, MT, USA, 2006.

[6]. G.J. Kacprzynski, M.J. Roemer, G. Modgil, A. Palladino, K. Maynard, Enhancement of physics-of-failure prognostic models with system level features, In: Proceedings of the 2002 IEEE Aerospace Conference, Big Sky, MT, USA, 2002.

[7]. C.J. Li, H. Lee, Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics, Mechanical Systems and Signal Processing 19 (2005) 836–846.

[8]. D. Banjevic, A.K.S. Jardine, V. Makis V, A control-limit policy and software for condition-based maintenance optimization, INFOR 39 (2001) 32–50.

[9]. A. Heng, S. Zhang, A.C.C. Tan, J. Mathew, Rotating machinery prognostics: state of the art, challenges and opportunities, Mechanical Systems and Signal Processing 23 (2008) 724–739.

[10]. R.Q. Huang, L.F. Xi, X.L. Li, Residual Life Predictions for Ball Bearings Based On Self-Organizing Map And Back Propagation Neural Network Methods, Mechanical Systems and Signal Processing 21(1) (2007) 193-207.

[11]. Z. Tian, M.J. Zuo, Health condition prognostics of gears using a recurrent neural network approach, In: Proceedings of the Reliability and Maintainability Symposium, Fort Worth, TX, USA, 2009.

[12]. P.W. Tse, D. P. Atherton, Prediction of machine deterioration using vibration based fault trends and recurrent neural networks, Journal of Vibration and Acoustics, Transactions of ASME 121 (1999) 355-362.

[13]. J. Lee, J. Ni, D. Djurdjanovic, H. Qiu and H. Liao, Intelligent prognostics tools and e-maintenance, Computers in Industry 57 (2006) 476-489.

[14]. N. Gebraeel, M.A. Lawley, R. Liu, Residual life, predictions from vibration-based degradation signals: A neural network approach, IEEE Transactions on Industrial Electronics 51(3) (2004) 694-700.

[15]. N. Gebraeel, M.A. Lawley, A neural network degradation model for computing and updating residual life distributions, IEEE Transactions on Automation Science And Engineering 5(1) (2008) 154-163.

[16]. S.J. Wu, N. Gebraeel, M.A. Lawley, Y. Yih, A neural network integrated decision support system for condition-based optimal predictive maintenance policy, IEEE Transactions on Systems Man and Cybernetics Part A-Systems and Humans 37(2) (2007) 226-236.

[17]. Z. Tian, An artificial neural network approach for remaining useful life prediction of equipment subject to condition monitoring, In: Proceedings of the International Conference

on Reliability Maintainability and Safety, Chengdu, China, July 21-25, 2009.

[18]. A.K.S. Jardine and A.H.C. Tsang, Maintenance, Replacement, and Reliability: Theory and Applications, Boca Raton: CRC/Taylor & Francis, 2006.

[19]. C. Lu and W. Meeker, Using degradation measures to estimate a time-to-failure distribution, Technometrics 35 (2) (1993) 161–174.

[20]. D. Banjevic, A.K.S. Jardine, Calculation of reliability function and remaining useful life for a Markov failure time process, IMA Journal of Management Mathematics 17 (2006) 115−130.

[21]. R. Rojas, Neural networks: a system introduction, Berlin: Springer, 1996.

[22]. Mathworks. The MATLAB Neural Networks Toolbox, 2007.

[23]. M. Riedmiller, H. Braun, A direct adaptive method for faster backpropagation learning: The RPROP algorithm, in: IEEE Proceedings of the International Conference on NN (ICNN), San Francisco, 1993.

[24]. P.M. Williams, Bayesian regularization and pruning using a Laplace prior, Neural Computation 7(1) (1995) 117-143.

[25]. B. Stevens, EXAKT reduces failures at Canadian Kraft Mill, www.omdec.com, 2006.

[26]. D. Lin, D. Banjevic, A.K.S. Jardine, Using principal components in a proportional hazards model with applications in condition-based maintenance, Journal of the Operational Research Society 57 (2006) 910–919.

Biographies

Zhigang Tian is currently an Assistant Professor in Concordia Institute for Information Systems

Engineering at Concordia University, Canada. He received his Ph.D. degree in 2007 in Mechanical Engineering at the University of Alberta, Canada, and his M.S. degree in 2003 and B.S. degree in 2000 both in Mechanical Engineering at Dalian University of Technology, China. His research interests focus on reliability analysis and optimization, prognostics, condition monitoring, and maintenance optimization. He is a member of IIE and INFORMS.

Lorna Wong is currently a PhD student at the Center for Maintenance Optimization & Reliability Engineering, University of Toronto, Canada. She received her Honours BSc in Math and History in 2003, and her MEng in Industrial Engineering in 2007, both from the University of Toronto. Her research is focused on maintenance optimization for repairable systems.

Nima Safaei is a Postdoctoral Fellow at the Center for Maintenance Optimization & Reliability Engineering, University of Toronto, Canada. He received a BS in Applied Mathematics from University of Mazandaran, Babolsar, graduated in Industrial Engineering from Mazandaran University of Science & Technology, Babol, and received his PhD degree in Industrial Engineering from Iran University of Science & Technology, Tehran. His research focuses on combinatorial optimization, scheduling, and computational-intelligence-based problem-solving methods.

List of Figure Captions

List of Table Captions

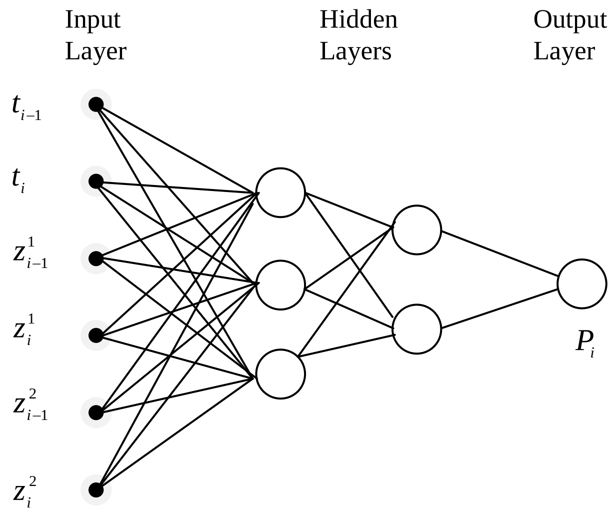


Fig. 1. Structure of the ANN model for remaining useful life prediction [17]

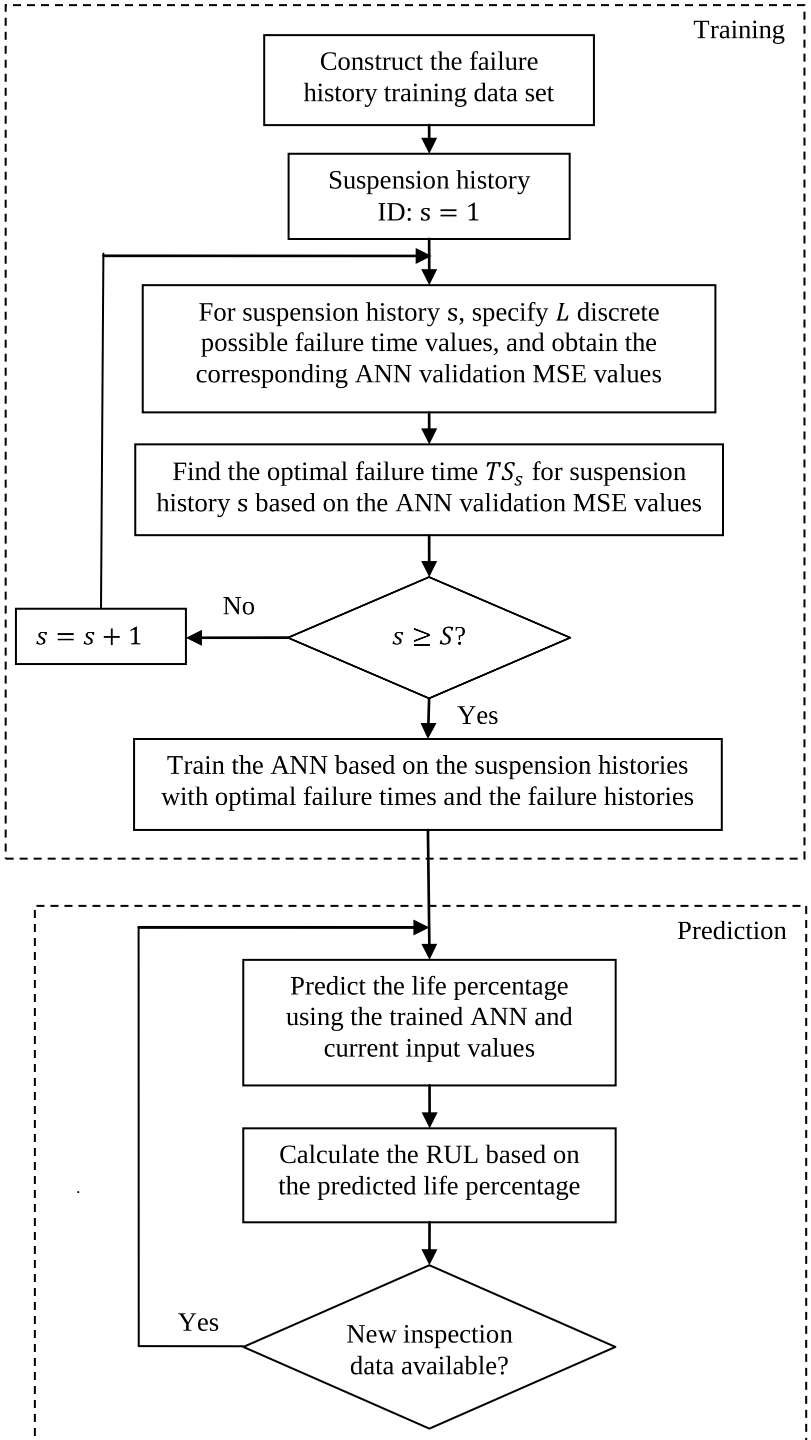


Fig. 2. Procedure of the remaining life prediction approach

Parameter	Estimate	Sign. (*)	Standard Error	Wald	DF	p - Value	Exp of Estimate	95 % CI	
								Lower	Upper
Scale	2755	-	432	-	-	-	-	1908	3602
Shape	3.398	Y	0.6761	12.58	1	0.0003893	-	2.073	4.724
P1H_Par5	20.81	Y	6.456	10.39	1	0.001265	1.095e+009	8.16	33.47
P1V_Par5	57.47	Y	14.63	15.42	1	0	9.081e+024	28.79	86.15

Fig. 3. Significance analysis result using EXAKT

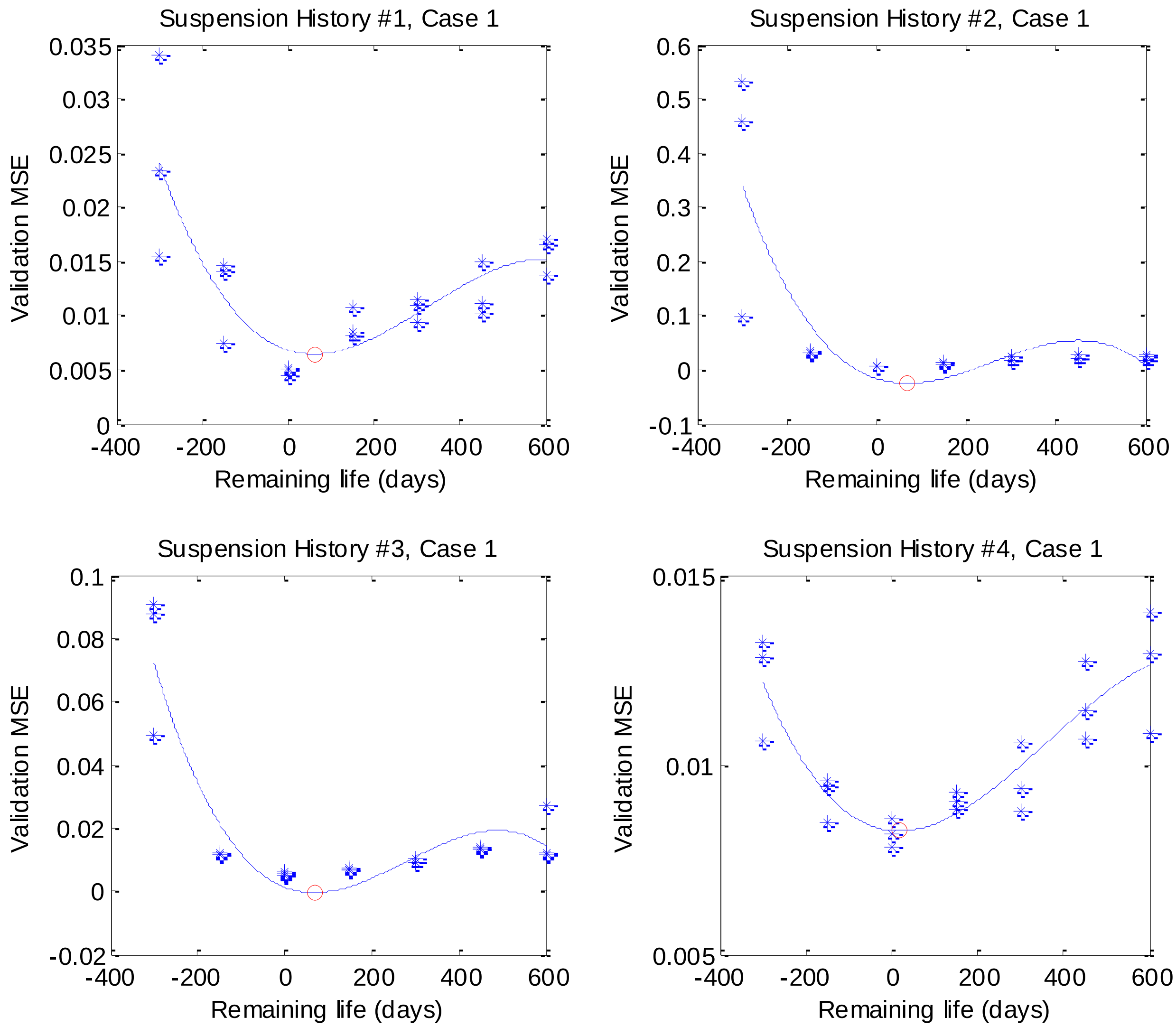
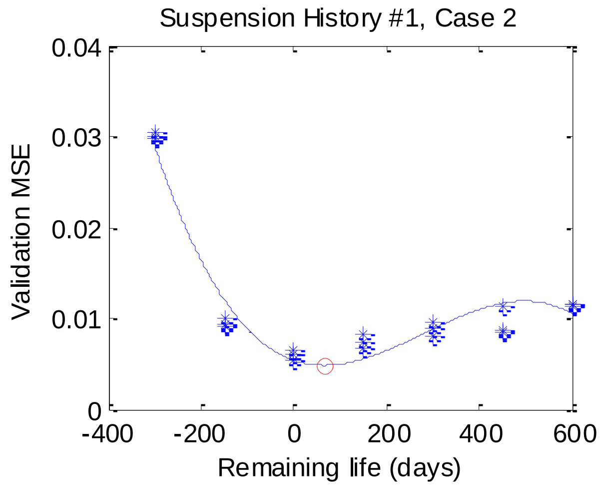
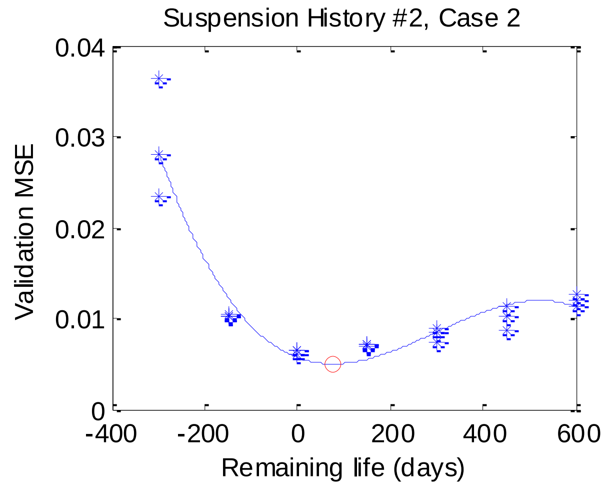
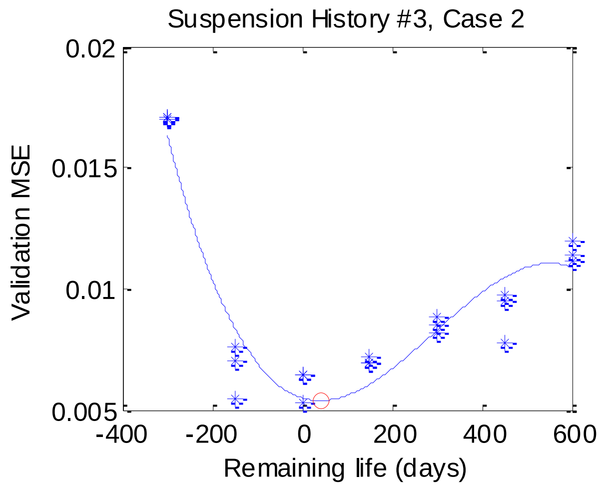
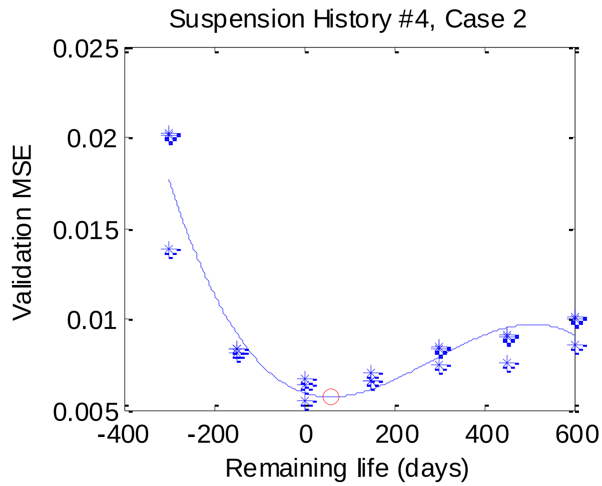


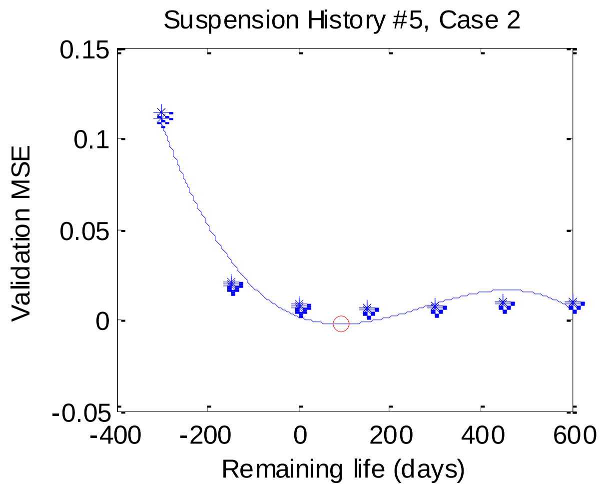
Fig. 4. The validation MSE and optimal remaining life values for the suspension histories in Case 1

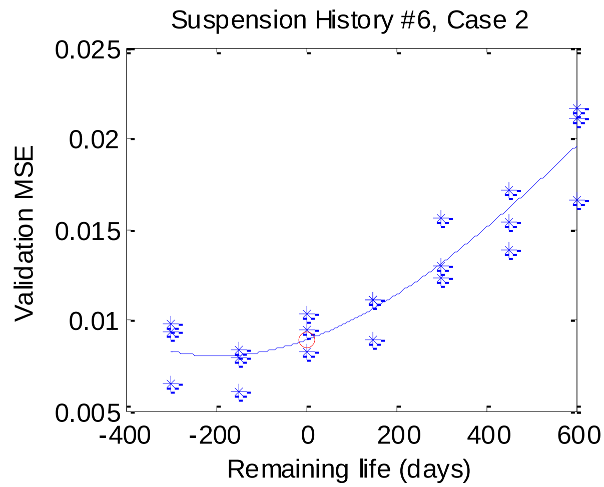

Suspension History #1, Case 2
Validation MSE
0.04
0.03
0.02
0.01
0
-400
-200
0
200
400
600
Remaining life (days)


Suspension History #2, Case 2
Validation MSE
0.04
0.03
0.02
0.01
0
-400
-200
0
200
400
600
Remaining life (days)


Suspension History #3, Case 2
Validation MSE
0.02
0.015
0.01
0.005
-400
-200
0
200
400
600
Remaining life (days)


Suspension History #4, Case 2
Validation MSE
0.025
0.02
0.015
0.01
0.005
-400
-200
0
200
400
600
Remaining life (days)


Suspension History #5, Case 2
Validation MSE
0.15
0.1
0.05
0
-0.05
-400
-200
0
200
400
600
Remaining life (days)


Suspension History #6, Case 2
Validation MSE
0.025
0.02
0.015
0.01
0.005
-400
-200
0
200
400
600
Remaining life (days)

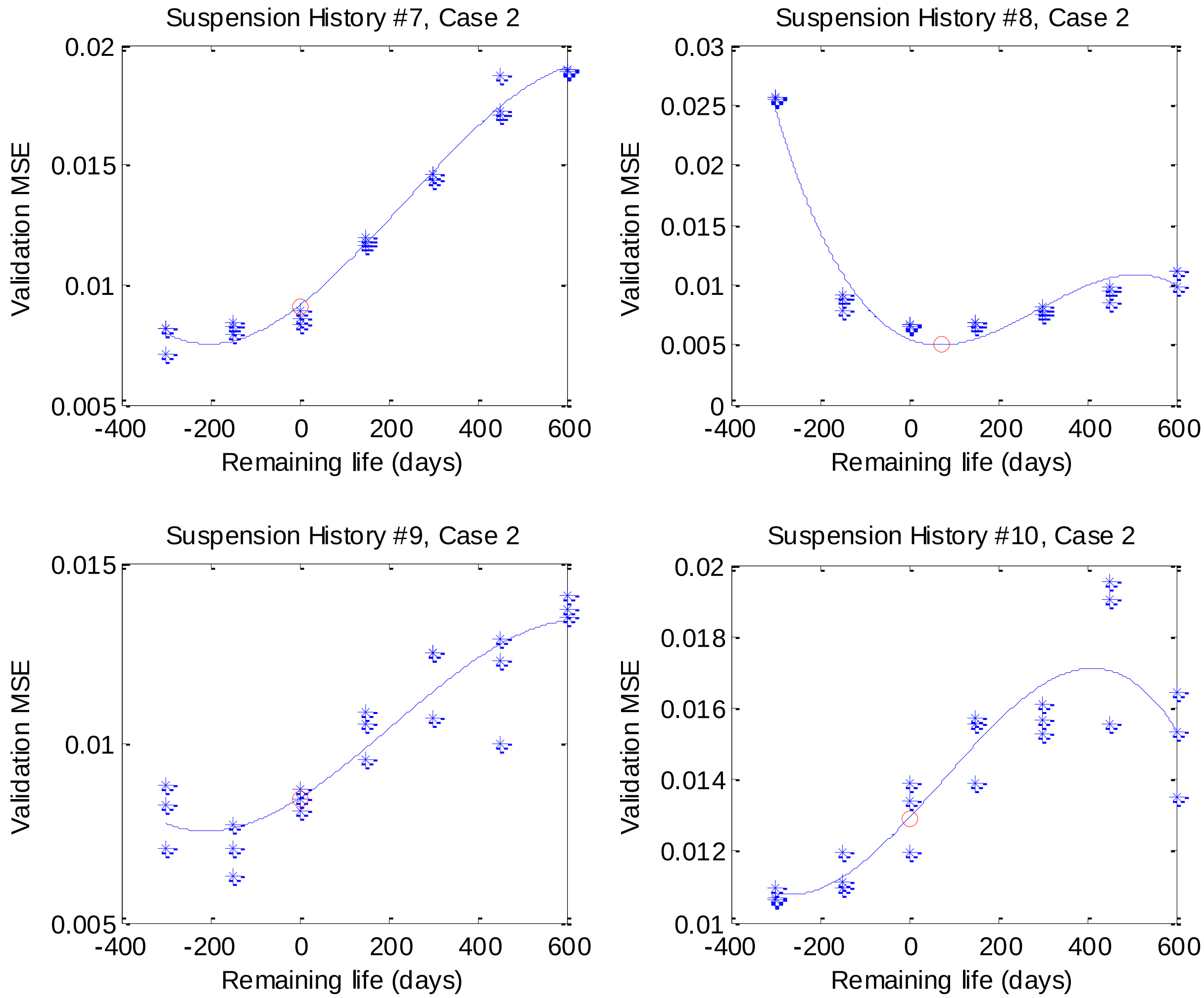


Fig. 5. The validation MSE and optimal remaining life values for the suspension histories in case 2

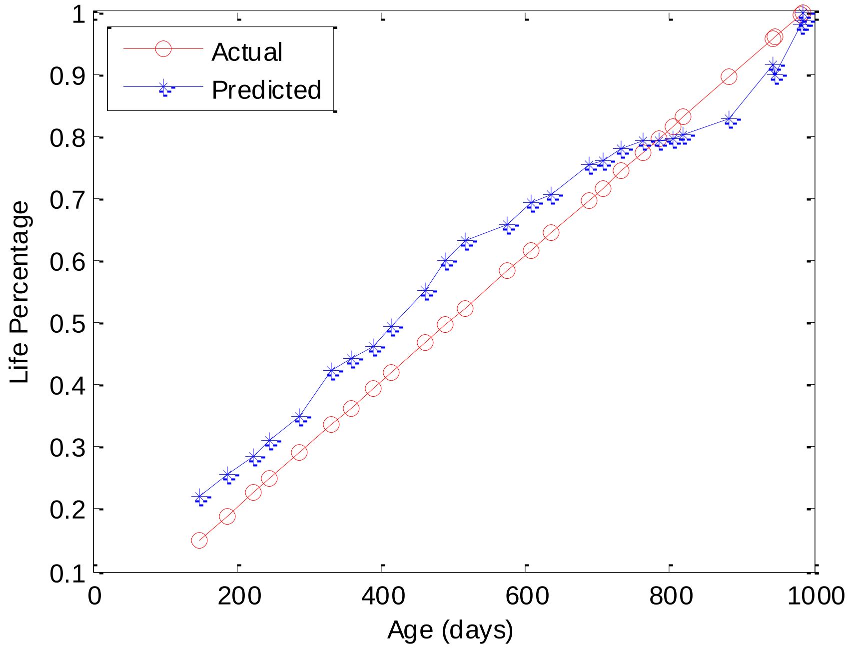


Fig. 6. Prediction results for a sample test history using the proposed approach

Table 1. the Actual failure times and the suspension times for the suspension histories

Suspension history #: s	Actual failure time (days)	Suspension time: TS_s (days)
1	601	506
2	511	425
3	692	521
4	986	982

Table 2. The RUL prediction results (Case 1)

	$\bar{e}_{\mathrm{All}}$	STD_{All}	$\bar{e}_{\mathrm{L10}}$	STD_{L10}
The ANN method using failure histories only	22.94%	10.47%	8.18%	6.67%
The proposed ANN approach	16.87%	7.05%	5.66%	2.12%

Table 3. The RUL prediction results (Case 2)

	$\bar{e}_{\text{All}}$	STD_{All}	$\bar{e}_{\text{L10}}$	STD_{L10}
The ANN method using failure histories only	19.79%	9.26%	6.79%	4.13%
The proposed ANN approach	13.82%	7.86%	3.65%	1.09%