IMA Journal of Management Mathematics (2002) **13**, 3–16

A model to predict the residual life of rolling element bearings given monitored condition information to date

WENBIN WANG†
Centre for OR and Applied Statistics, University of Salford, UK

[Received on 30 January 2001; revised on 30 November 2001; accepted on 17 December 2001]

In condition monitoring practice, one of the primary concerns of maintenance managers is how long the item monitored can survive given condition information obtained to date. This relates to the concept of the condition residual time where the survival time is not only dependent upon the age of the item monitored, but also upon the condition information obtained. Once such a probability density function of the condition residual time is available, a consequencial decision model can be readily established to recommend a 'best' maintenance policy based upon all information available to date. This paper reports on a study using the monitored vibration signals to predict the residual life of a set of rolling element bearings on the basis of a chosen distribution. A set of complete life data of six identical bearings along with the history of their monitored vibration signals is available to us. The data were obtained from a laboratory fatigue experiment which was conducted under an identical condition. We use stochastic filtering to predict the residual life distribution given the monitored condition monitoring history to date. As the life data are available, we can compare them with the prediction. The predicted results are satisfactory and provide a basis for further studies. It should be pointed out that although the model itself is developed for the bearings concerned, it can be generalized to modelling general condition-based maintenance decison making provided similar conditions are met.

Keywords: residual life; condition monitoring; prediction; bearing.

1. Introduction

The use of condition monitoring techniques within industry to direct maintenance actions has increased rapidly over recent years to the extent that it has marked the beginning of what is likely to prove a new generation in production and maintenance management practice. There are both economic and technological needs for this development driven by tight profit margins, and high outage costs as plant has increased in cost, complexity and automation. Technical advances in condition monitoring techniques have provided a means to achieve high availability and to reduce unscheduled production shutdowns. In all cases, the measured condition information, in addition to potentially improving decision making, has an insurance role for a manager in that there is now a more objective means of explaining actions if challenged.

Today there exists a large and growing variety of forms of condition monitoring techniques for machine health monitoring and fault diagnosis. A particularly popular

† Also at the Management School of Harbin Institute of Technology, People's Republic of China

one for rotating and reciprocal machinery is vibration analysis. However, irrespective of the particular condition monitoring techniques used, the working principle of condition monitoring is the same: namely, condition data become available which need to be interpreted and appropriate actions taken accordingly. There are generally two stages in condition-based maintenance. The first stage is related to condition monitoring data acquisition and their interpretation. There have been numerous papers contributing to this stage, as evidenced by the proceedings of COMADEM over recent years (Starr & Rao, 2001). This first stage is characterized by engineering skill, knowledge and experience. For detailed technical aspects of condition monitoring and fault diagnosis, see Collacott (1977). The second stage is maintenance decision making, namely what to do now given that condition information and its interpretation are available. The decision at this stage can be complicated and entails consideration of cost, downtime, production demand, preventive maintenance shutdown windows, and most importantly, the likely survival time of the item monitored. Compared with the extensive literature on condition monitoring techniques and their applications, relatively little attention has been paid to the important problem of modelling the decision making stage in condition-based maintenance. Only a few studies have appeared. Christer & Wang (1992, 1995) addressed maintenance decision problems of directly monitored systems in which the actual condition of the system can be observed by condition monitoring. Christer *et al.* (1997) presented a case study of furnace erosion prediction and replacement using the state space model and Kalman filtering. In Wang *et al.* (1995, 1996, 1997) various models were explored and discussed without actual applications. Coolen & Dekker (1995) discussed the sensitivity analysis of a two-stage failure process. There have been attempts at modelling condition-based maintenance decision making using the proportional hazard model (PHM) (Kumar & Westberg, 1997; Makis & Jardine, 1991). However, there is a fundamental problem in the PHM models as only the current condition information or functions of current information are used to predict the future development of the item monitored, rather than the whole monitoring history. Christer & Wang (1995) showed that good maintenance decision making can be dependent upon the rate at which plant reached a particular maintained state, that is the past history profile of maintenance observations. Aven (1996) and Heinrich & Jensen (1992) developed stochastic process models of degradation, but again models were established for the directly observed systems. Scarf (1997) surveyed the papers on modelling condition-based maintenance and highlighted the need for more theoretical and case studies in the area.

In this paper we present a case study of predicting the residual life of items monitored based upon the condition information obtained in the form of a distribution. Once such a prediction is available, a model for condition-based maintenance decision making can be readily established (Wang & Christer, 2000). The model used in this paper is an extension of the general model developed by Wang & Christer (2000) using stochastic filtering theory. The items concerned in this study are six rolling element bearings set up in a laboratory fatigue experiment where the overall vibration signal of the bearings is monitored on an irregular basis. This is a typical case of indirect monitoring since the true condition of the item monitored is unknown but correlated with the measured vibration signals. The bearings were under accelerated life testing, and therefore complete life data are available to us over a relatively short period of time. This is a rare case where the time to failure is available since the bearing with abnormal vibration signals will not be allowed

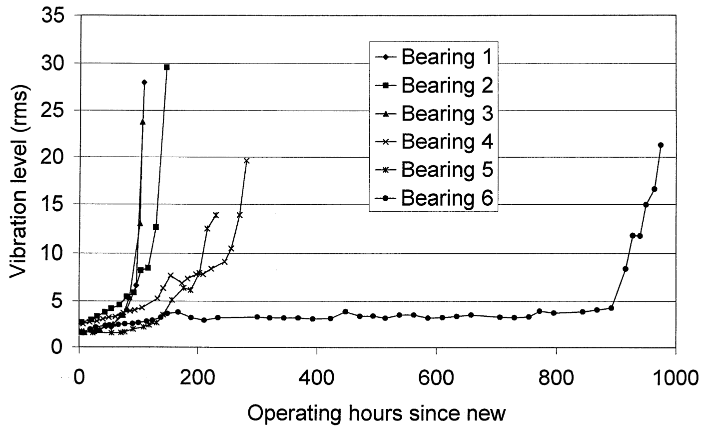

FIG. 1. Vibration data of six bearings.

to run to failure in practice. The delay time concept (Christer & Waller, 1984) was used to classify the failure process into two stages where the abnormal vibration signal will usually only appear over the second stage, namely the failure delay time period, because of the hidden defect. We discuss the ways to identify the initiation of a defect, and use this data set to fit a model to it. As the life data are available, we can compare the prediction with the actual data. The fitted results are satisfactory since the failure times of the six bearings are within the range of the predicted residual distributions. It is noted that although this is a special case study, the model developed here can be generalized into other condition monitoring practice if similar conditions are satisfied.

2. The data

Figure 1 shows the data of vibration in rms of six bearings. It can be seen from Fig. 1 that the bearing lives vary from around 100 hours to over 1000 hours, which shows a typical stochastic nature of the life distribution. The monitored vibration signals also indicate an increasing trend with bearing age in all cases, but with different paths. The important observation from the data set is the pattern of vibration signals which stays relatively flat in the early stage of the bearing life and then increases rapidly (a defect may have been initiated). This indicates the existence of the two-stage failure process as defined by Christer & Waller (1984). It also demonstrates that more frequent condition monitoring checks might be needed once an increasing trend is picked up.

The information in Fig. 1 is the basis for subsequent model formulation and analysis. It is noted that the bearings concerned in this study are maintenance free since the only action is replacement. This considerably simplifies the modelling process. It will be shown later that, as expected, the key element in the model is to establish and estimate the distribution of the residual life at time t since a defect may have initiated given condition information to date. We first discuss in the next section the identification of such an initial point of a hidden defect, and then the establishment of the conditional probability density function of residual life will be discussed in detail.

3. Identification of the initiation point of a defect

The observed condition information is assumed in Wang & Christer (2000) to be a function of the residual life from time zero. This may be the case in oil-based monitoring where the amount of the oil debris in the lubricating oil could be positively correlated with wear, and indeed influences the residual life of the item from time zero. This may, however, not be the case in vibration monitoring. Figure 1 demonstrates that there should be no concern over the time period where the vibration signal stays stable unless a consistent increase has been observed. We know that the residual life decreases monotonically as the bearing ages in the absence of corrective maintenance intervention, but the vibration signal may stay relatively flat until a defect has been initiated, which will usually trigger a rapid increase in vibration. In this case we can assume that the relationship between the observed vibration information and the residual life exists over the time period since a defect may have been initiated. This relates to the concept of the delay time defined by Christer & Waller (1984).

The failure delay time is defined in Christer & Waller (1984) as the time period from when a defect is first observable to when a repair (replace) would be essential (breakdown) in the absence of corrective action. It has been generally accepted in the case of vibration monitoring that a consistent and rapid increase in the vibration signal is a good indicator of the presence of a hidden defect. Our interest is to identify the starting point of such a delay time and estimate its residual length using the available vibration signal as opposed to the usual 0–1 type of information used in delay time modelling (Baker & Wang, 1991).

There are two approaches to deal with the identification of a potential defective state of the monitored item. The first is a subjective approach using engineers' experience and expertise, that is, it uses engineers' judgement to specify whether the bearing is in a defective state or not based on available vibration and other data. This approach is practical and commonly used in practice. However, it is subject to personal experience that varies from person to person, and is also difficult to automate unless an appropriate expert system is established. The second approach is to use a control chart to detect the vibration signal deviation from its normal position. This approach is statistically based and easy to implement without prior engineering knowledge of the item concerned. Since the data in Fig. 1 were collected several years ago and the associated information about the size and type of bearing, running speed and loading have been lost, it is difficult for an expert to judge the bearing state purely from the vibration data, and therefore, we choose to use the control chart approach.

Since the data are taken one point at a time, the range chart is not appropriate, and we choose a Shewhart average level chart in this case. Table 1 shows the result of the action limits of the six bearings. The action limit is the threshold level that indicates an abnormal vibration signal, which corresponds to the initial point of a potential defect. Note that the exact location of the initial point of a defect is not significant for subsequent modelling since false alarms should be allowed, and it does not matter too much if we start to model the residual life sooner or later in the earlier stage of the delay time period. In this case we choose an action limit of 5 for all bearings.

4. Establishment of the conditional residual time distribution

The time lapse from any point that the operating system is checked by condition monitoring to the time that it may fail is defined in Wang *et al.* (1995) as the conditional residual

TABLE 1 *Action limits of the vibration signals*

Bearing	μ	σ	Action limit
1	2.577	0.829	5.064
2	3.616	0.669	5.623
3	2.273	0.627	4.156
4	3.394	0.538	5.144
5	2.095	0.609	3.921
6	3.176	0.575	4.901
Average	2.855	0.649	4.802

time (CRT). In the case considered here, only the signals over the potential failure delay time will be used, and therefore the conditional residual time refers to the time period from a check within the delay time to the actual failure conditional on the information collected since the initiation of a defect to date. Referring to the delay time concept, we call this the conditional residual delay time, or the residual delay time for short. Two pieces of information are present at the time of checking, namely the age of the item and the condition monitoring information history to date. The residual delay time at the time of checking is dependent upon the age or time that the item has survived, but also upon the condition information obtained to date. It is the availability of such condition information which informs us more about the likely survival probability of the item monitored over a period of time. If at time t we are interested in assessing the probability that the item monitored can survive over a period of time given the age and condition information obtained, the problem reduces to determining the residual delay time distribution conditional upon time t, and available condition information. This is, in general, of prime significance to condition monitoring practice since one of the purposes of monitoring is to assess the survival probability using available condition information, and thereby to initiate appropriate maintenance actions. We establish in this section the conditional residual time distribution based upon stochastic filtering theory (Wang & Christer, 2000).

4.1 *Modelling assumptions and notation*

According to the nature of the data and condition monitoring practice, we propose the following modelling assumptions:

1. Plant items are monitored irregularly at discrete time points.
2. There is no maintenance to the bearings.
3. The residual delay time is a random variable, which might be described by a probability distribution, and may be conditional on available condition information.
4. The condition information obtained at time t_i, say the current time, is also a random variable which may be described by a distribution function whose mean is assumed to be a function of the current residual delay time.

Assumptions 1 and 2 are the condition monitoring and maintenance practice of the case described here. Assumption 3 is self-evident. Assumption 4 extends the assumption

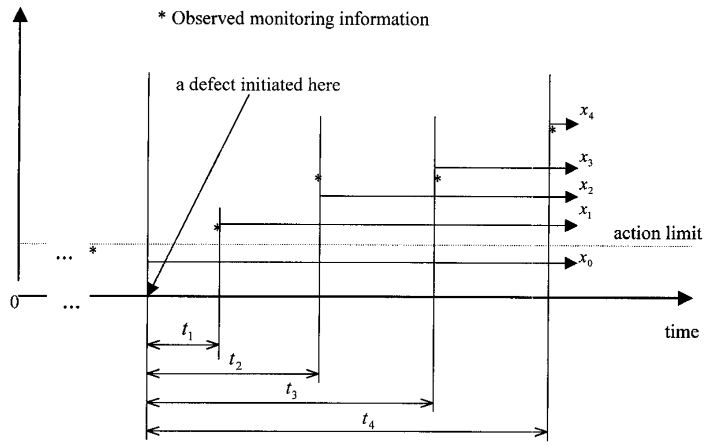

FIG. 2. Relationship between times of checking and residual times since a defect has initiated.

established in Wang & Christer (2000) using the residual delay time concept, which specifies the stochastic relationship between the residual time and condition monitoring signals.

The notation used in this paper is as follows;

- t_i denotes the ith and the current monitoring time since the bearing was initially detected to be defective but still operating.
- x_i denotes the residual delay time at time t_i.
- y_i denotes the condition information obtained at time t_i.
- $Y_i = \{y_i, y_{i-1}, y_{i-2}, \ldots, y_1\}$ denotes the history of condition monitoring information obtained to t_i.
- $p_i(x_i|Y_i)$ denotes the PDF of x_i conditional upon Y_i.
- $p(y_i|x_i)$ denotes the PDF of y_i conditional upon x_i.

Note that our interest is in the establishment of $p_i(x_i|Y_i)$ over the delay time period so t_i is measured since the initiation point of a defect, see Fig. 2. As the actual initiation point of a defect is unknown to us, we used, in subsequent model fitting, the mid-point between the first check where the reading is above the action limit and the check point immediately before it as our approximation to the initial point of the defect.

4.2 *The residual delay time distribution*

Since the residual delay time at t_i is the residual delay time at t_{i-1} minus the interval between t_i and t_{i-1} provided the item has survived to t_i and no maintenance action has been taken since then, it follows that

$$x_i = \begin{cases} x_{i-1} - (t_i - t_{i-1}) & \text{if } x_{i-1} > t_i - t_{i-1}\text{, and no maintenance intervention,} \\ \text{not defined} & \text{otherwise.} \end{cases} \tag{1}$$

The relationship between y_i and x_i is yet to be identified. From assumption 4 we know that it can be described by a distribution, say, $p(y_i|x_i)$. We shall discuss this later when fitting the model to the data.

We seek to establish the expression for $p_i(x_i|Y_i)$ so that a consequent decision model can be constructed on the basis of such a conditional probability. It can be shown that (Wang & Christer, 2000) $p_i(x_i|Y_i)$ is given by

$$p_i(x_i|Y_i) = p(x_i|y_i, Y_{i-1}) = \frac{p(x_i, y_i|Y_{i-1})}{p(y_i|Y_{i-1})}, \tag{2}$$

where using the chain rule, the joint distribution, $p(x_i, y_i|Y_{i-1})$, and $p(y_i|Y_{i-1})$ are given respectively by

$$p(x_i, y_i|Y_{i-1}) = p(y_i|x_i, Y_{i-1})p(x_i|Y_{i-1}) = p(y_i|x_i)p(x_i|Y_{i-1}), \tag{3}$$

$$p(y_i|Y_{i-1}) = \int_0^\infty p(y_i|x_i, Y_{i-1})p(x_i|Y_{i-1})\mathrm{d}x_i = \int_0^\infty p(y_i|x_i)p(x_i|Y_{i-1})\mathrm{d}x_i, \tag{4}$$

where, using (1),

$$p(x_i|Y_{i-1}) = \frac{p_{i-1}(x_i + t_i - t_{i-1}|Y_{i-1})}{\int_{t_i - t_{i-1}}^\infty p_{i-1}(z|Y_{i-1})\mathrm{d}z}. \tag{5}$$

Equations 3 and 4 are established based on assumption 4, that is $p(y_i|x_i, Y_{i-1}) = p(y_i|x_i)$,which requires more explanation. If condition monitoring is useful, then there must be some form of correlation between y_i and x_i. It is known that the root cause of an abnormal signal y_i is usually a hidden defect, and not the previously observed Y_{i-1}. There is normally a negative correlation between the amplitude of the vibration signal and the severity of the defect. If we use the generic word of residual delay time to represent the severity of a defect, that is the more severe the defect, the shorter the residual delay time, then the negative correlation between y_i and x_i is established. This implies that x_i is equivalent to the root cause of the defect, and if x_i is known, then y_i is uniquely determined by $p(y_i|x_i)$.

It can be seen from (3)–(5) that since $p(x_i|Y_{i-1})$ can be determined from $p_{i-1}(x_{i-1}|Y_{i-1})$, then if $p_0(x_0|Y_0)$ and $p(y_i|x_i)$ are known, (2) may be determined recursively. Equations (1)–(5) constitute a stochastic filtering process, which is a special case of the model in Wang & Christer (2000).

5. Fitting the model to the data

In this section we fit the model introduced earlier to the vibration monitoring data of six bearings. The first question is obviously what distributional forms to take for $p_0(x_0|Y_0)$ and $p(y_i|x_i)$. As there is no condition monitoring reading available when the bearings are first installed, $p(x_0|Y_0) = p(x_0)$ which is just the conventional PDF of the bearing life. For the bearing delay time distribution, an appropriate choice is the Weibull distribution which has been widely used in bearing fatigue experiments (Chen, 1993). It is noted, however, that the choice of the distribution of y_i conditional upon x_i could be Weibull, gamma, lognormal or normal distributions. The conditional relationship between y_i and x_i is established by

TABLE 2 *Distribution and the associated set-up for* $E(y_i) \propto A + B\mathrm{e}^{-Cx_i}$

Distribution	PDF of y_i	$E(y_i)$	Setup
Weibull	$\rho\eta(\rho y_i)^{\eta-1}\mathrm{e}^{-(\rho y_i)^\eta}$	$\frac{1}{\rho}\Gamma(1+\frac{1}{\eta})$	$\frac{1}{\rho} = A + B\mathrm{e}^{-Cx_i}$
Gamma	$\frac{\mathrm{e}^{-\rho y_i}\rho(\rho y_i)^{\eta-1}}{\Gamma(\eta)}$	$\frac{\eta}{\rho}$	$\frac{1}{\rho} = A + B\mathrm{e}^{-Cx_i}$
Normal	$\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{1}{2\sigma^2}(y_i-\mu)^2}$	μ	$\mu = A + B\mathrm{e}^{-Cx_i}$
Lognormal	$\frac{1}{\sqrt{2\pi}\sigma y_i}\mathrm{e}^{-\frac{1}{2\sigma^2}(\log y_i-\mu)^2}$	$\mathrm{e}^{\mu+\frac{1}{2}\sigma^2}$	$\mu = \log(A + B\mathrm{e}^{-Cx_i})$

letting $E(y_i) \propto A + B\mathrm{e}^{-Cx_i}$, where A, B and C are parameters to be estimated from the data. This will create a negative correlation between y_i and x_i as expected.

The actual setup for the relationship between y_i and x_i depends on the distribution used. Table 2 shows the setup for each of the candidate distribution functions used. In all cases the relationship of $E(y_i) \propto A + B\mathrm{e}^{-Cx_i}$ holds. The function form of $A + B\mathrm{e}^{-Cx_i}$ is chosen based on the data in Fig. 1 as it gives a closer fit to the data.

The model parameters are estimated in two steps. The first step is to estimate the parameters in $p(x_0)$, that is the scale and shape parameters of the Weibull distribution as the failure delay time distribution. The actual starting point of the delay time for each bearing is unknown, but we know that it is between the interval of the point just before $y_i \geqslant 5$ and the first point after $y_i \geqslant 5$. For simplicity, we assumed that it is at the midpoint between these two points. We used the maximum likelihood method based upon the six bearings data, which gives $\alpha = 0{\cdot}011$ and $\beta = 1{\cdot}8729$.

We wish to estimate A, B, C and the other parameter within $p(y_i|x_i)$ based upon the available vibration monitoring data collected at discrete points and the survival time information at each checking point using the maximum likelihood method. However, the classical theory of maximum likelihood is based upon the situation in which observations are independently and identically distributed. It should be noted in our case that the observations are neither independent or identical. Here the definition of a condition probability density function is used to write the likelihood function as

$$L = \prod_{j=1}^{m} p(z_{j1}) \times p(z_{j2}|z_{j1}) \times p(z_{j3}|z_{j2}, z_{j1}) \times \cdots \times p(z_{jn_j}|z_{j,n_j-1}, \ldots, z_{j1}) \quad (6)$$

where $p(\cdot|\circ)$ denotes the PDF of observing $\cdot$ given that $\circ$ has occurred, m is the number of the rolling element bearings tested, z_{ji} is the observed information of the jth bearing at t_i and t_{n_j} is the last monitoring check time for bearing j. This is called the multiplication law of likelihood where L is the likelihood of $\sum_{j=1}^{m} n_j$ events.

Taking a bearing life as an example, at t_1 we have two pieces of information: the first is that we have observed condition monitoring reading y_1 and the second is that the residual delay time is longer than t_1. If the joint PDF for having such information is denoted by $p(y_1, x_0 > t_1)$, then using Bayes' theorem, we have

$$p(y_1, x_0 > t_1) = p(y_1|x_0 > t_1)P_0(x_0 > t_1).$$

Since $x_0 = x_1 + t_1$ from (1), we have $p(y_1|x_0 > t_1) = p(y_1|x_1 > 0) = p(y_1)$. Note that throughout the text a lower-case p is used to denote PDF and an upper-case P is used for probability.

At t_2, again we have two pieces of information: the first is that we have observed condition monitoring reading y_2 and the second is that the bearing has survived over t_2, both are conditional upon our having observed y_1 at t_1 and $x_0 > t_1$. The corresponding PDF for events at t_2 is given by

$$p(y_2, x_1 > t_2 - t_1 | y_1, x_0 > t_1) = p(y_2 | y_1, x_1 > t_2 - t_1, x_0 > t_1) P(x_1 > t_2 - t_1 | y_1, x_0 > t_1). \tag{7}$$

As $x_1 > t_2 - t_1$ implies that we must have $x_0 > t_1$ and $x_1 = x_2 + t_2 - t_1$, we have

$$p(y_2 | y_1, x_1 > t_2 - t_1, x_0 > t_1) = p(y_2 | x_2 > 0, y_1) = p(y_2 | y_1)$$

and

$$P(x_1 > t_2 - t_1 | y_1, x_0 > t_1) = \int_{t_2 - t_1}^{\infty} p_1(z | y_1) \mathrm{d}z = P_1(x_1 > t_2 - t_1 | y_1).$$

Then (7) becomes

$$p(y_2, x_1 > t_2 - t_1 | y_1, x_0 > t_1) = p(y_2 | y_1) P_1(x_1 > t_2 - t_1 | y_1).$$

Similarly at t_i, we have

$$\begin{aligned} &p(y_i, x_{i-1} > t_i - t_{i-1} | Y_{i-1}, x_{i-2} > t_{i-1} - t_{i-2}, \dots, x_0 > t_1) \\ &\quad = p(y_i | Y_{i-1}) P_{i-1}(x_{i-1} > t_i - t_{i-1} | Y_{i-1}). \end{aligned}$$

If the last observation is a failure at time t_f, $t_f > t_n$, where t_n is the time of the last monitoring check, its contribution to the likelihood function is

$$p_n(x_n = t_f - t_n | Y_n).$$

The likelihood function for this bearing is then

$$L = \left(\prod_{i=1}^{n} p(y_i | Y_{i-1}) P_{i-1}(x_{i-1} > t_i - t_{i-1} | Y_{i-1}) \right) p_n(x_n = t_f - t_n | Y_n).$$

The likelihood for m bearings is given by

$$\begin{aligned} L = \prod_{j=1}^{m} &\left(\prod_{i=1}^{n_j} p(y_{ji} | Y_{j,i-1}) P_{j,i-1}(x_{j,i-1} > t_{ji} - t_{ji-1} | Y_{j,i-1}) \right) \\ &\times p_{jn_j}(x_{jn_j} = t_{jf} - t_{jn_j} | Y_{jn_j}), \end{aligned} \tag{8}$$

where n_j is the number of condition monitoring checks for the jth bearing, y_{ji} is the vibration reading of the jth bearing at t_{ji}, x_{ji} is the residual delay time of the jth bearing at t_{ji}, and $Y_{j,i-1}$ is the monitoring history of the jth bearing to the $(i-1)$th check.

In the case that both $p(y_i|x_i)$ and $p_0(x_0)$ follow a Weibull distribution,

$$p_0(x_0) = \alpha\beta(\alpha x_0)^{\beta-1}\mathrm{e}^{-(\alpha x_0)^\beta} \tag{9}$$

and

$$p(y_i|x_i) = \rho\eta(\rho y_i)^{\eta-1}\mathrm{e}^{-(\rho y_i)^\eta} \tag{10}$$

where $\rho = 1/(A + B\mathrm{e}^{-Cx_i})$.

Using (1)–(5) and (9)–(10), and letting $\psi_k(z, t_{ji-1}) = \frac{\mathrm{e}^{(-y_{jk}(A+Be^{-C(z+t_{ji-1}-t_{jk})})^{-1})^\eta}}{A+Be^{-C(z+t_{ji-1}-t_{jk})}}$ after some manipulation, we have

$$p(y_{ji}|Y_{ji-1}) = \begin{cases} \dfrac{\int_0^\infty \eta y_{ji}^{\eta-1}(z+t_{ji-1})^{\beta-1}\mathrm{e}^{-(\alpha(z+t_{ji-1}))^\beta}\prod_{k=1}^{i}\psi_k(z,t_{ji-1})\mathrm{d}z}{\int_{t_{ji}-t_{ji-1}}^\infty (z+t_{ji-1})^{\beta-1}\mathrm{e}^{-(\alpha(z+t_{ji-1}))^\beta}\prod_{k=1}^{i-1}\psi_k(z,t_{ji-1})\mathrm{d}z}, & i \geqslant 2 \\ \dfrac{\int_0^\infty \eta y_{j1}^{\eta-1}(z+t_{j1})^{\beta-1}\mathrm{e}^{-(\alpha(z+t_{j1}))^\beta}\psi_1(z,t_{j1})\mathrm{d}z}{\int_{t_{j1}}^\infty z^{\beta-1}\mathrm{e}^{-(\alpha z)^\beta}\mathrm{d}z}, & i = 1. \end{cases} \tag{11}$$

The PDF of $(x_{j,i-1}|Y_{j,i-1})$ is

$$p_{ji-1}(x_{ji-1}|Y_{ji-1}) = \begin{cases} \dfrac{(x_{ji-1}+t_{ji-1})^{\beta-1}\mathrm{e}^{-(\alpha(x_{ji-1}+t_{ji-1}))^\beta}\prod_{k=1}^{i-1}\psi_k(x_{ji-1},t_{ji-1})}{\int_0^\infty (z+t_{ji-1})^{\beta-1}\mathrm{e}^{-(\alpha(z+t_{ji-1}))^\beta}\prod_{k=1}^{i-1}\psi_k(z,t_{ji-1})\mathrm{d}z}, & i \geqslant 2 \\ \alpha\beta(\alpha x_{j0})^{\beta-1}\mathrm{e}^{-(\alpha x_{j0})^\beta}, & i = 1. \end{cases} \tag{12}$$

Substituting (11) and (12) into (8), after some manipulation the likelihood function of (8) becomes

$$L = \prod_{j=1}^{m}\left(\alpha^\beta\beta t_{jf}^{\beta-1}\prod_{i=1}^{n_j}\left(\eta y_{ji}^{\eta-1}\rho_k(0)^\eta\mathrm{e}^{-\sum_{k=1}^{n_j}(y_{jk}\rho_k(0))^\eta-(\alpha t_{jf})^\beta}\right)\right). \tag{13}$$

Taking the log on both sides of (13), we have

$$\log L = \sum_{j=1}^{m}\left(\log(\alpha^\beta\beta) + \log(t_{jf}^{\beta-1}) + \sum_{i=1}^{n_j}\log\left(\eta y_{ji}^{\eta-1}\rho_k(0)^\eta\mathrm{e}^{-\sum_{k=1}^{n_j}(y_{jk}\rho_k(0))^\eta-(\alpha t_{jf})^\beta}\right)\right). \tag{14}$$

The final problem to be solved is the failure time of bearing. The actual life recorded is the time period from a brand new bearing being installed to the point of a high-vibration signal being detected, when the bearing is considered to be unacceptable, and therefore is replaced. There might be some residual time left since the last reading on the bearings although it may be small. This information is not available to us and we estimate that it is around 10 hours based upon the consultation with the researcher who conducted the

TABLE 3 *Estimated parameter values*

Distribution used for $p(y_i\|x_i)$	A	B	C	Other parameter	Log likelihood value
Weibull	7·0688	27·089	0·0528	$\eta = 4{\cdot}5593$	−82·147
Gamma	0·8442	3·267	0·0643	$\eta = 8{\cdot}2481$	−85·131
Normal	7·4126	34·637	0·0862	$\sigma = 3{\cdot}2042$	−90·419
Lognormal	6·9516	27·701	0·0714	$\sigma = 0{\cdot}2639$	−83·468

experiment: that is, the time to failure of a bearing, say the jth bearing, is $t_{n_j}+10$. Applying the above procedures we have estimated the values of A, B, C, and other parameters as shown in Table 3.

As the Weibull model produces the highest likelihood value of the four candidate distributions for $p(y_i|x_i)$, we use it as our choice of $p(y_i|x_i)$.

We seek to plot the predicted residual times of these six bearings at various monitoring points based upon estimated model parameters. The functional form of $p_i(x_i|Y_i)$ can be obtained recursively from (1)–(5), (12) and (9) as

$$p_i(x_i|Y_i) = \frac{(x_i + t_i)^{\beta-1}\mathrm{e}^{-(\alpha(x_i+t_i))^\beta}\prod_{k=1}^{i}\psi_k(x_i, t_i)}{\int_0^\infty (z + t_i)^{\beta-1}\mathrm{e}^{-(\alpha(z+t_i))^\beta}\prod_{k=1}^{i}\psi_k(z, t_i)\mathrm{d}z} \tag{15}$$

where

$$\psi_k(z, t_i) = \frac{\mathrm{e}^{-(y_k(A+Be^{-C(z+t_i-t_k)})^{-1})^\eta}}{A + Be^{-C(z+t_i-t_k)}}.$$

Note the role played by y_k where all the historical information of y has been taken into account. The predicted residual time at some monitoring points given the history information based upon bearing 6 data in Fig. 1 is plotted in Fig. 3.

In Fig. 3, the actual residual lives at each checking point are also plotted with symbol *. It can be seen that the actual residual lives are well within the predicted residual time distributions as expected, which gives us confidence of the model established. The important point here is the ability of the model to pick up individual remaining lives given both the condition monitoring information history and the current age. At the beginning all bearings are assumed to follow an identical Weibull delay time distribution, namely $p_0(x_0)$. As time passes and more condition monitoring information becomes available we can see the completely different paths of the bearing lives which have been indicated by our model.

6. The decision model

The fundamental decision to make at various monitoring points is whether we should replace the bearing or not given all information available. If the answer is yes then what is the best time for such a replacement, and should we wait until a suitable production window such as a scheduled shutdown arrives? Suppose that the mean cost of a failure is C_f and C_p denotes a mean preventive replacement cost, t is the current monitoring check point measured since new and T is the planned replacement time. If monitoring cost

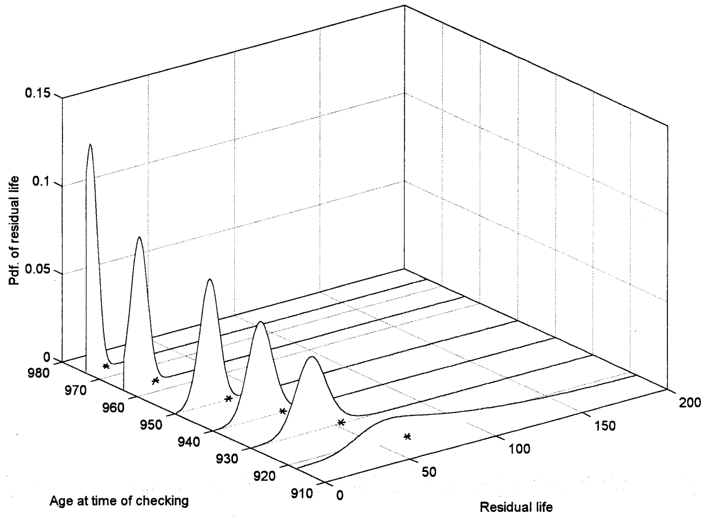

FIG. 3. Predicted conditional residual time of bearing 6.

is negligible and subscript i denotes the order of checks since $y \geqslant 5$, we have the total expected cost per unit time (Wang & Christer, 2000) given by

$$
\begin{aligned}
C(t) &= \frac{C_f P_i(x_i < T - t|Y_i) + C_p(1 - P_i(x_i < T - t|Y_i))}{t + (T - t)(1 - P_i(x_i < T - t|Y_i)) + \int_0^{T-t} z p_i(z|Y_i)\,\mathrm{d}z} \\
&= \frac{(C_f - C_p)P_i(x_i < T - t|Y_i) + C_p}{t + (T - t)(1 - P_i(x_i < T - t|Y_i)) + \int_0^{T-t} z p_i(z|Y_i)\,\mathrm{d}z}.
\end{aligned}
\tag{16}
$$

Equation (16) utilizes the renewal reward theory in that the numerator is the expected cycle cost and the denominator is the expected cycle length given that T is the planned replacement time. If the minimum of (16) is outside of a specified planning period, there is no need to perform a preventive replacement within the period, otherwise a planned replacement should be recommended at the time point where the minimum is located. Suppose $C_f = £6000$ and $C_p = £2000$, which are based upon our experience in previous case studies, and we are at monitoring check point $t = 916$, 930, 940, 950, 964, and 974 hours respectively for bearing number 6; using (16) and the model parameters obtained in the last section, we have the graphical output of our decision model as illustrated in Fig. 4.

The bearing eventually failed shortly after $t = 974$ hours, and we can see from our model that a preventive replacement was recommended as early as at $t = 916$ hours.

We can also see from this example that the model is dynamic and updated at each monitoring point once new information becomes available.

7. Conclusion

This paper reports on a case study of the application of a model to the prognosis of bearing residual delay time given available condition monitoring information. The model

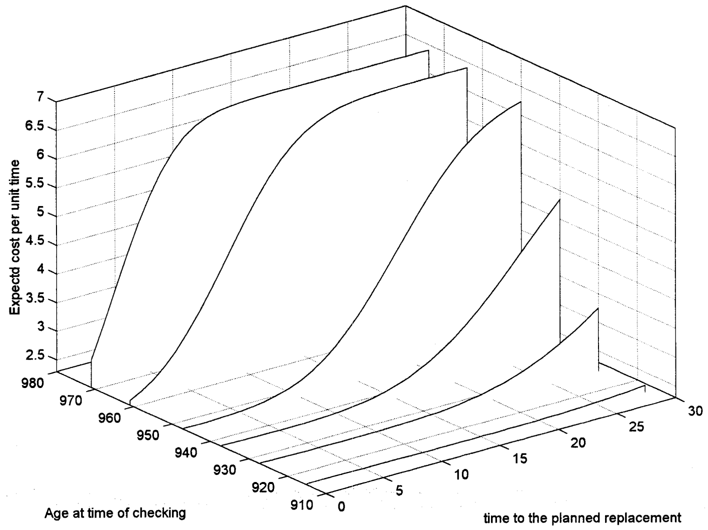

FIG. 4. Expected cost per hour in terms of the planned replacement at time t given that the current monitoring check time is t.

is established under a special case where life information is available. The fundamental concept behind the model is the conditional residual delay time which differs from the conventional concept of the residual time in that it not only depends upon the current age of the item but also on the condition information available to date. The condition information we used were taken from vibration readings of six bearings used in a lab. fatigue experiment.

Since vibration analysis is the most popular condition monitoring means used in industry, the model developed has a potential to be applied to the much wider situation where similar condition monitoring processes exist. It should be pointed out that this model is a condition-based dynamic model which requires the on-line input of condition information, and therefore must be installed in the user's system. For this reason we are in the process of developing a software package based upon the model presented earlier. This software will be tested and installed in a company system if resources are available.

8. Acknowledgement

The research reported here is partly supported by EPSRC under grant number GR|M96582.

REFERENCES

AVEN, T. (1996) Condition based replacement policies—a counting process approach. *Reliab. Eng. & System Safety*, **51**, 275–281.

BAKER, R. D & WANG, W. (1991) Estimating the delay-time distribution of faults in repairable machinery from failure data. *IMA J. Maths. Appl. Bus. Ind.*, **3**, 259–281.

CHEN, J. (1993) Survival analysis and reliability (in Chinese). Anhui Education Press.

CHRISTER, A. H. & WALLER, (1984) Delay time models of industrial inspection maintenance problems. *J. Opl. Res. Soc.*, **35**, 401–406.

CHRISTER, A. H. & WANG, W. (1995) A simple condition monitoring model for a direct monitoring process. *Euro. J. Opl. Res.*, **82**, 258–269.

CHRISTER, A. H. & WANG, W. (1992) A model of condition monitoring inspection of production plant. *I. J. Prod. Res.*, **30**, 2199–2211.

CHRISTER, A. H., WANG, W. & SHARP, J. M. (1997) A state space condition monitoring model for furnace erosion prediction and replacement. *Euro. J. Opl. Res.*, **101**, 1–14.

COLLACOTT, R. A. (1977) *Mechanical Fault Diagnosis and Condition Monitoring*. London: Chapman and Hall.

COOLEN, F. R. A. & DEKKER, R. (1995) Analysis of a 2-phase model for optimisation of condition monitoring. *IEEE Trans. Reliab.*, **44**, 505–511.

HEINRICH, G. & JENSEN, U. (1992) Optimal replacement rules based upon different information levels. *Naval Research Logistics*, **39**, 937–955.

KUMAR, D. & WESTBERG, U. (1997) Maintenance scheduling under age replacement policy using proportional hazard modelling and total-time-on-test plotting. *Euro. J. Opl. Res.*, **99**, 507–515.

MAKIS, V. & JARDINE, A. K. S. (1991) Computation of optimal policies in replacement models. *IMA J. Maths. Appl. Bus. Ind.*, **3**, 169–176.

SCARF, P. (1997) On the modelling of condition based maintenance. *Proceedings of ESREL97, June*. (C. Greder Soars, ed.). Lisbon, Portugal: Pergamon.

STARR, G. & RAO, B. K. N. (eds) (2001) *Proceedings of COMADEM international congresses and exhibitions*. London: Elsevier.

WANG, W. & CHRISTER, A. H. (2000) A general of condition based maintenance model for a stochastic dynamic system. *J. Opl. Res. Soc.*, **51**, 145–155.

WANG, W., CHRISTER, A. H. & SHARP, J. M. (1996) Stochastic decision modelling of condition based maintenance. *Proceedings of COMADEM96, 6-8 July, 1996, Sheffield*. (B. K. N. Rao, R. A. Smith & J. L. Wearing, eds). Sheffield: Sheffield Academic Press, pp. 1175–1184.

WANG, W., CHRISTER, A. H. & SHARP, J. M. (1995) A model of condition monitoring by stochastic filtering. *Proceedings of COMADEM95, July, 1995, Kingston, Canada*. (B. K. N. Rao, T. N. Moore & J. Jeswiet. eds). pp. 329–336.

WANG, W., SCARF, P. & SHARP, J. M. (1997) Modelling condition based maintenance of production plant. *Proceedings of COMADEM97, 9–11 June, 1997, Espoo, Finland*. (Jantunen Erkki. ed.). pp. 75–84.