Case Study

Institution of MECHANICAL ENGINEERS

Journal of ENGINEERING TRIBOLOGY

Proc IMechE Part J:
J Engineering Tribology
0(0) 1–8

DOI: 10.1177/1350650119874560
journals.sagepub.com/home/pij
SAGE

Remaining useful life prediction of lubricating oil with dynamic principal component analysis and proportional hazards model

Ying Du[1], Tonghai Wu[2], Shengxi Zhou[1] and Viliam Makis[3]

Abstract

Lubricating oil contains a lot of tribological information of the machine and plays an important role in machine health. Oil degrades with serving time and causes severe wear afterwards, which is a complex dynamic process, and difficult to be accurately described by a single property. Therefore, the main purpose of deterioration prediction is to estimate the remaining useful life that the oil can still fulfill its functions by analyzing oil condition monitoring data. With a large amount of oil condition monitoring data collected, a vector autoregressive model is applied to the original oil data to describe the dynamic deterioration process. Then dynamic principal component analysis, an effective dimensionality reduction method, is employed to obtain the principal components capturing the most information of the oil data. The proportional hazards model is then built to calculate the failure risk of the lubricating oil based on the condition monitoring information, where its baseline function represents the aging process assuming to follow the Weibull distribution and its positive link function represents the influence of covariates (the principal components) on the failure risk. Finally, the remaining useful life prediction of lubricating oil can be obtained by explicit formulas of the characteristics such as the conditional reliability function and the mean residual life function. This work provides an approach to assess the health of lubricating oil, and a guidance for oil maintenance strategy.

Keywords

Deterioration modeling, time series model, dynamic principal component analysis, proportional hazards model, remaining useful life prediction

Date received: 18 January 2019; accepted: 7 August 2019

Introduction

Lubricating oil is one of the key components of a machine, which contains a lot of tribological information. Lubricating oil characteristics can describe the deterioration of the oil and the operation conditions of the machine.[1,2] With serving time and the influences of many factors, lubricating oil degrades and produces acidic substances, moisture, and insoluble deposits,[3,4] such as carbon deposits, sludges, etc. Under such circumstances, lubricating oil deteriorates, and its performance characteristics are reduced,[5] which subsequently leads to machine failure caused by friction and wear problems.[6,7] By oil monitoring and analysis technologies, the physical and chemical properties of lubricating oil obtained from condition monitoring (CM) have been used to assess oil deterioration and evaluate the current status of the machine.[8,9] Due to the fact that lubricating oil performance does not depend only on one or several indicators, different kinds of monitoring methods are required to obtain as much oil information as possible. However, to our knowledge, there may be a strong correlation between variables of the CM data with both cross-correlation and autocorrelation because they are related to the same deterioration process of lubricating oil. Moreover, the large amount of data makes it difficult to assess oil deterioration and evaluate machine condition. Therefore, it is first necessary to reduce the

[1]School of Aeronautics, Northwestern Polytechnical University, Xi'an, China
[2]Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an, China
[3]Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada

Corresponding author:
Tonghai Wu, Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an, Shaanxi, China.
Email: wt-h@163.com

dimensionality of the oil data. Dimensionality reduction method is applied, which maps multi-dimensional data to a low-dimensional space using fewer irrelevant variables to represent the original oil data. Principal component analysis is one of the most popularly used methods for dimensionality reduction to overcome "dimensionality disaster" and improve analysis efficiency.[10] Dynamic principal component analysis (DPCA) is an extension of principal component analysis, which is very suitable for multivariate continuous data that are both cross-correlated and autocorrelated.[11,12]

Lubricating oil will fail when the deterioration degree reaches the pre-specified fault threshold. In industry, the threshold of physical and chemical properties of lubricating oil is utilized to determine the oil change time, but the value of the threshold relies heavily on empirical values, and the control limit is different depending on the characteristics of lubricating oil performance and the working condition in actual engineering applications.[13] Therefore, at present, there is still no criterion for determining the failure of lubricating oil. In the literature, several statistical models have been applied to describe the deterioration process and obtain the failure time of a system or a machine, such as the filtering-based model,[14] the hidden Markov model (HMM),[15,16] the proportional hazards model (PHM),[17,18] the regression model,[19] etc. These models are data-driven methods which utilize the lifetime information obtained from inspections. PHM, which was proposed by Cox in 1972, has been widely used to calculate the failure risk (hazard rate) with CM information and predict the remaining useful life (RUL) of the deterioration system first in the field of biomedical sciences.[20] Recently, this kind of method has also been employed in the field of reliability and maintenance of machines.[21–23] Furthermore, the method of PHM does not require the prior knowledge of a given threshold.

In this paper, a multivariate time series analysis of the original lubricating oil data based on physical and chemical properties is firstly applied to fit a vector autoregressive (VAR) model. A dimensionality reduction methodology, DPCA, is then employed to reduce the model dimensionality. The principal components (PCs) are then obtained, which carry most information of the oil data without cross-correlation and autocorrelation. Subsequently, a PHM is built to calculate the failure risk of lubricating oil, where the PCs are regarded as the covariates of the model. Finally, the explicit formulas of the conditional reliability function (CRF) and the mean residual life (MRL) function are computed to predict the RUL of lubricating oil.[24] The CRF represents the probability that the lubricating oil can fulfill its functions during the period of time and has not failed yet, and the MRL function is employed to obtain the remaining time that the oil can still survive.[15] It is a very new approach to apply PHM in the field of Tribology to model the oil deterioration process and estimate the RUL of lubricating oil with the CM data based on physical and chemical properties. The complete procedure is shown in Figure 1.

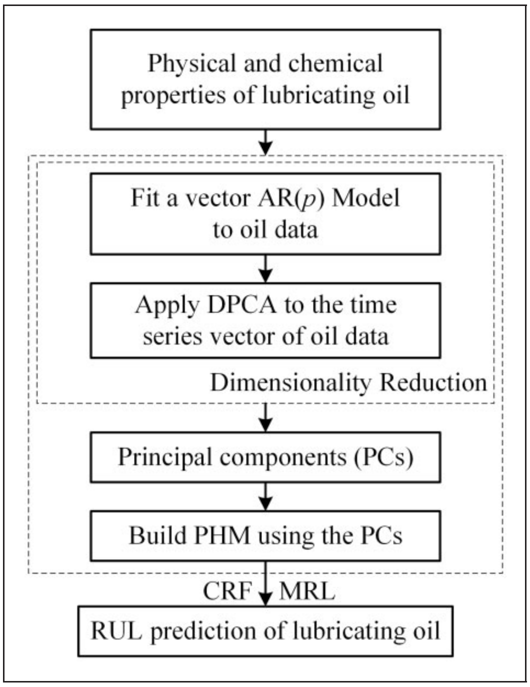

Figure 1. Procedure for RUL prediction of lubricating oil.

Lubricating oil data description

The real experimental CM oil data based on physical and chemical properties were obtained. The lubricating oil, which was used for large machines, was collected from the engine of a loader. The CM data of lubricating oil represent the oxidation of the oil, the loss of additives and the deterioration of the oil. There are several monitoring methods to inspect the physical and chemical properties of lubricating oil, and dozens of performance parameters are used to characterize oil deterioration. In industry, the selected monitoring indicators are oxidation, kinematic viscosity (40 °C), TAN, TBN, Ca, Zn, P, etc. In this paper, we use the seven-dimensional monitoring data to assess the lubricating oil, where a typical data sheet is given in Table 1. The oil data were collected every $\Delta = 100$ h by off-line monitoring of lubricating oil. TAN and TBN are the total acid number and the total base number of the lubricating oil, respectively. The measurements of TAN and TBN represent the oxidation degree of the lubricating oil, where the increase of TAN and the decrease of TBN indicate the production of the acid.[2] The measurements of the metal elements phosphorus (P), zinc (Zn), and calcium (Ca) in ppm represent the content of additives in lubricating oil,[25] so the loss of additives can be obtained by monitoring the contents of these elements. The total number of oil data histories is 6, which consist of $N = 2$ failure histories that end with a failure, and $M = 4$ suspension

Table 1. Physical and chemical properties of lubricating oil.

No.	Characteristic	Unit	0 h	100 h	200 h	300 h	400 h	500 h
1	Oxidation	A/cm	11.67	12.39	12.19	11.65	11.58	11.62
2	Viscosity	cSt	103.3	95.80	95.52	94.89	94.69	96.1
3	TAN	mgKOH/g	1.84	1.03	1.65	2.43	2.66	2.2
4	TBN	mgKOH/g	6.89	6.17	5.54	4.24	4.49	3.04
5	Ca	ppm	1332	1446	1396	1321	1290	1167
6	Zn	ppm	773	809	788	737	740	670
7	P	ppm	724	709	687	634	643	564

histories that end when the lubricating oil is still in operation and has not lost its functions.

Vector autoregressive modeling of the oil data

As described by Makis et al.,[17,26] the oil data histories are considered as a multivariate time sequence and follow a VAR process[27] given by

$$Z_n = \mu + \sum_{r=1}^{p} \Phi_r Z_{n-r} + \varepsilon_n, \quad n \in \mathbf{Z} \tag{1}$$

where $Z_n = (z_{1n}, z_{2n}, \ldots, z_{7n})'$, $n \in \mathbf{Z}$ denotes the seven-dimensional time series vector of random variables (CM data of lubricating oil), ε_n is the seven-dimensional vector representing the white noise process of lubricating oil with covariance matrix Σ following i.i.d. $N_7(\mathbf{0}, \mathbf{\Sigma})$, $p \in \mathbf{N}$ is the model order, $\Phi_r \in \mathbf{R}^{7\times 7}$, $r = 1, \ldots, p$ are coefficient matrices, $\mu_0 \in \mathbf{R}^7$ and $\Sigma \in \mathbf{R}^{7\times 7}$ are the mean vector and covariance matrices of Z_n, respectively. All the model parameters are unknown and need to be estimated.

The regression representation is $\mathbf{W} = \mathbf{VA} + \mathbf{E}$.[15,26] $\mathbf{W}$ is defined as the $a \times 7$ data matrix $\mathbf{W} = \left[Z_{p+1}, Z_{p+2}, \ldots Z_T\right]'$, where T is the time series length of lubricating oil, $T = \sum_{i=1}^{N+M} (t_i - p)$, and $a = T - p$. $\mathbf{V}$ is defined as the $a \times (7p+1)$ matrix, whose typical row is $\mathbf{V}_n = \left[1, Z_{n-1}, \ldots, Z_{n-p}\right]$, $n = p+1, \ldots, T$. $\mathbf{A}$ is defined as $\mathbf{A} = [\mu, \Phi_1, \ldots, \Phi_p]'$, and $\mathbf{E}$ is defined as the error matrix $\mathbf{E} = \left[\varepsilon_{p+1}, \varepsilon_{p+2}, \ldots, \varepsilon_T\right]'$.

Therefore, we have[15]

$$\mathbf{W} = \left[z^1_{p+1}, \ldots, z^1_{t_1}, \ldots, z^{N+M}_{p+1}, \ldots, z^{N+M}_{t_{N+M}}\right]' \tag{2}$$

$$\mathbf{A} = \left[\mu, \Phi_1, \ldots, \Phi_p\right]' \tag{3}$$

$$\mathbf{E} = \left[\varepsilon^1_{p+1}, \ldots, \varepsilon^1_{t_1}, \ldots, \varepsilon^{N+M}_{p+1}, \ldots, \varepsilon^{N+M}_{t_{N+M}}\right]' \tag{4}$$

$$\mathbf{V} = \begin{bmatrix} 1 & z^1_p & \cdots & z^1_1 \\ 1 & \vdots & \vdots & \vdots \\ 1 & z^1_{t_1-1} & \cdots & z^1_{t_1-p} \\ 1 & \vdots & \vdots & \vdots \\ 1 & z^{N+M}_p & \cdots & z^{N+M}_1 \\ 1 & \vdots & \vdots & \vdots \\ 1 & z^{N+M}_{t_{N+M}-1} & \cdots & z^{N+M}_{t_{N+M}-p} \end{bmatrix} \tag{5}$$

Using the method of least squares (LS)[27] to estimate the vector AR(p) model parameters of lubricating oil, the LS estimator of $\mathbf{A}$ minimizes the following objective function[15,26]

$$\mathrm{tr}(S_p) = \mathrm{tr}\left[(\mathbf{W} - \mathbf{VA})'(\mathbf{W} - \mathbf{VA})\right] \tag{6}$$

where "tr" is the trace of the square matrix, which is the sum of diagonal elements in the matrix, and S_p is the residual sum of the matrix obtained from the vector AR model of order p.

Then, the LS estimators of $\mathbf{A}$ and $\mathbf{\Sigma}$ can be computed by the following formulas

$$\hat{\mathbf{A}} = \left[\hat{\mu}, \hat{\Phi}_1, \ldots, \hat{\Phi}_p\right]' = (\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'\mathbf{W} \tag{7}$$

$$\begin{aligned} \hat{\mathbf{\Sigma}} &= \frac{1}{T - (7p+1)} (\mathbf{W} - \mathbf{VA})'(\mathbf{W} - \mathbf{VA}) \\ &= \frac{S_p}{T - (7p+1)} \end{aligned} \tag{8}$$

The estimation of the order of the VAR model $p \in \mathbf{N}$ is obtained by testing $H_0 : \Phi_p = 0$ against $H_a : \Phi_p \neq 0$. The likelihood ratio statistic is given by

$$M_p = -\left(T - 7p - 1 - \frac{1}{2}\right) \ln\left(\frac{\det(S_p)}{\det(S_{p-1})}\right) \tag{9}$$

where the likelihood ratio statistic M_p is approximately distributed as $\chi^2_{k^2}$, k is the dimension of the CM oil data. We reject $H_0 : \Phi_p = 0$, when $M_p > \chi^2_{\alpha,k^2}$. And α

is the selected significance level of the likelihood ratio test.

For our real seven-dimensional oil data, $k=7$. For $p=2$ and $p=3$, we obtain the LR statistic $M_2 = 81.9735$ and $M_3 = -168.23$. For the significance level $\alpha = 0.05$ and $k^2 = 49$ degrees of freedom, the critical value is $\chi^2_{0.05,49} = 65.2352$. Since $M_2 > \chi^2_{0.05,49}$ and $M_3 < \chi^2_{0.05,49}$, we reject $H_0 : \Phi_2 = 0$ and fail to reject $H_0 : \Phi_3 = 0$. Therefore, $\hat{p} = 2$ is the adequate model order for the oil data. For the fitted vector AR(p) model, the parameter estimates $\{\hat{\mu}, \hat{\Phi}_{12}, \hat{\Phi}_{22}, \hat{\Sigma}\}$ are given by equations (10) to (13)

$$\hat{\mu} = [11.8184, \quad 98.1250, \quad 1.9756, \quad 5.3547, \quad 1293.9167, \quad 727.5556, \quad 669.3611]' \tag{10}$$

$$\hat{\Sigma} = \begin{bmatrix} 0.1149 & -0.0389 & 0.0364 & -0.0274 & -4.0151 & -1.8773 & 2.1394 \\ -0.0389 & 2.7879 & -0.2121 & 0.5263 & -20.8087 & 10.6991 & 32.7385 \\ 0.0364 & -0.2121 & 0.1220 & -0.0709 & -5.3152 & -3.5211 & -1.2782 \\ -0.0274 & 0.5263 & -0.0709 & 0.6189 & 15.7660 & 2.3264 & 20.9676 \\ -4.0151 & -20.8087 & -5.3152 & 15.7660 & 1950.7451 & 134.3497 & 322.1918 \\ -1.8773 & 10.6991 & -3.5211 & 2.3264 & 134.3497 & 425.8500 & 158.7244 \\ 2.1394 & 32.7385 & -1.2782 & 20.9676 & 322.1918 & 158.7244 & 1123.1992 \end{bmatrix} \tag{11}$$

$$\hat{\Phi}_{12} = \begin{bmatrix} 0.2232 & -2.6817 & 0.3663 & -0.1406 & -42.2457 & 7.5913 & 8.5408 \\ 0.0030 & -0.1545 & 0.0552 & -0.0502 & -5.9917 & 3.7117 & -1.5628 \\ 0.2382 & -4.4831 & 1.0524 & -0.1513 & -87.9208 & -34.9342 & 17.8352 \\ 0.0253 & 0.2606 & 0.0537 & 0.4421 & -44.4522 & -13.5743 & -20.6324 \\ -0.0011 & 0.0079 & -0.0028 & -0.0012 & 1.0136 & 0.0527 & 0.0507 \\ -0.00001 & 0.0020 & 0.0013 & 0.0047 & -1.3197 & 0.4610 & 0.2472 \\ 0.0032 & -0.0002 & -0.0059 & 0.0001 & 0.8816 & -0.1745 & 0.8429 \end{bmatrix} \tag{12}$$

$$\hat{\Phi}_{22} = \begin{bmatrix} -0.4685 & 0.2687 & 0.0885 & 0.1630 & -28.5648 & 15.0614 & -17.2206 \\ 0.0047 & 0.2210 & -0.0185 & 0.0503 & -5.0669 & -0.0650 & -1.9170 \\ 0.0018 & -3.6632 & 0.2316 & -0.8936 & -60.9622 & 43.1238 & -16.2975 \\ 0.1986 & -0.5004 & 0.2145 & -0.1676 & 43.2904 & 0.4863 & 43.7250 \\ -0.0006 & -0.0422 & 0.0060 & -0.0017 & -0.1838 & 0.2209 & 0.0210 \\ 0.0015 & 0.0169 & -0.0089 & -0.0015 & 0.6542 & 0.3032 & -0.19750 \\ -0.0034 & -0.0784 & 0.0113 & 0.0064 & -0.8847 & 0.4499 & -0.3031 \end{bmatrix} \tag{13}$$

Dimensionality reduction of oil data using DPCA

Principal component analysis is a method of multivariate statistical analysis, which is mainly used for dimensionality reduction, feature extraction, and fault diagnosis. This kind of method applies the linear transformation to the CM data. DPCA is an extension of the original principal component analysis method, which can be applied to the matrix composed of the time-shifted data vectors.[16] By using DPCA, the multivariate CM data of lubricating oil can be reduced to a data set of variables (the PCs) that accounts for the most information of the oil data. The PCs are uncorrelated and independent of each other,[28] and the value of the covariance is 0. Therefore, we adopt DPCA method to reduce the dimensionality of the oil data, and the vectors of the oil data consist of the current data vector Z_n and the time-shifted vectors Z_{n-1}, Z_{n-2}.

Firstly, the correlation matrix R of the oil data should be obtained, $R(i) = D^{-1}\Gamma(i)D^{-1}$, $i = 0, 1, 2$, where i is the time lag.[17] $\Gamma(0)$ is the cross-covariance matrix, $\Gamma(i)$, $i = 1, 2$ are the auto-covariance matrices, and $D = \Gamma(0) - \Gamma^{*'}_{(p-1)}\Gamma^{-1}_{(p-1)}\Gamma^{*}_{(p-1)}$ is the diagonal matrix of standard deviations of the oil data, $\Gamma^{*'}_{(p-1)} = [\Gamma(p-1), \ldots, \Gamma(1)]$, $\Gamma_p = \begin{bmatrix} \Gamma_{p-1} & \Gamma^{*}_{(p-1)} \\ \Gamma^{*'}_{(p-1)} & \Gamma(0) \end{bmatrix}$. $R(i-j)$ is the (i, j)th block of the correlation matrix, and $R(i-j) = R(j-i)'$, if $i-j < 0$, $i, j = 1, 2, 3$.

Secondly, the score vector of the dynamic PCs of the oil data can be obtained as[11]

$$S_n = U'O_n = (u_1, u_2, \ldots, u_k)' \cdot O_n \tag{14}$$

where $U = u_1,\ u_2,\ \ldots,\ u_k$ are the eigenvectors of the correlation matrix R of the oil data, the vector $O_n = \left(O'_n,\ O'_{n-1},\ O'_{n-2}\right)'$ is the standardized vector obtained from the oil data set Z_n, and $O'_n = \left(\frac{Z_{n,1}-\overline{Z_1}}{s_1},\ \frac{Z_{n,2}-\overline{Z_2}}{s_2},\ \ldots,\ \frac{Z_{n,k}-\overline{Z_k}}{s_k}\right)$, $\overline{Z_i}$ and s_i are the mean and the standard deviation of variables Z_i, $i = 1, \ldots, k$, respectively.

Cattell's scree test[29] is adopted to determine the number of dynamic PCs to retain, and the result of the test is shown in Figure 2. The eigenvalues are drawn in continuous descending order of their extraction, and then an elbow in the curve is identified so that the bottom of the eigenvalues after the elbow forms an approximate straight line. The retained dynamic PCs are the eigenvalues that are above the line.

From Figure 2, we can conclude that there is a break between the first 4 and the remaining 17 eigenvalues, which indicates that we need to retain the first four dynamic PCs for the subsequent fitting of PHM.

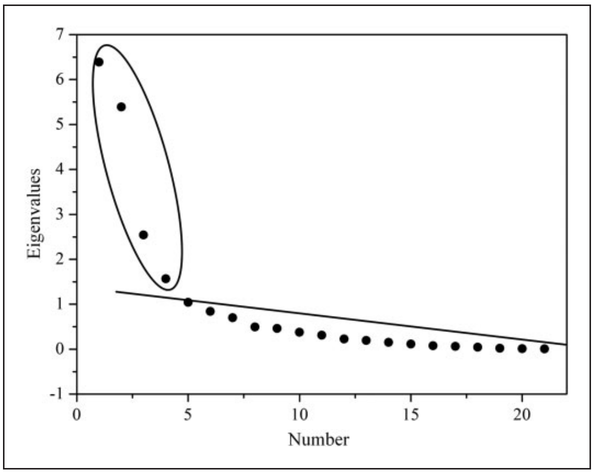

Figure 2. Scree test for the lubricating oil data.

A more detailed description of the eigenvalues and their contribution rates obtained after applying DPCA is shown in Table 2, where the eigenvalues l_i are in the successive descending order. It can be seen that the first four eigenvalues l_i, $i = 1, 2, 3, 4$ are $\{6.3883, 5.3891, 2.5405, 1.5635\}$, and the contribution rates C_i, $i = 1, 2, 3, 4$ of the first four components are $\{30.42\ \%, 25.66\ \%, 12.10\ \%, 7.45\ \%\}$. It is indicated that the selected four dynamic PCs contain the most information of the lubricating oil performance from the original oil data. Therefore, it is very reasonable and feasible to choose the four dynamic PCs.

Proportional hazards modeling and RUL prediction of lubricating oil

Proportional hazards modeling

PHM is utilized to estimate the failure time of the system according to the CM variables and lifespan data of lubricating oil. The relationship between the CM data and the hazard rate is described by using PHM shown in Figure 3,[22] which is considered to be very successful to model the lifetime data.

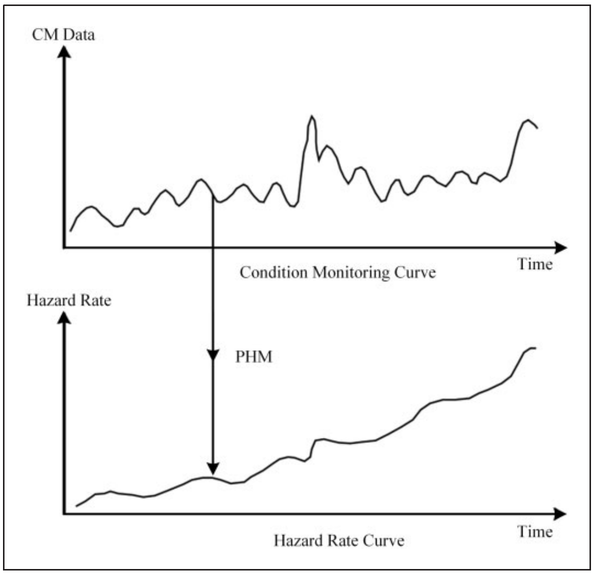

Figure 3. Hazard rate calculation by CM data using PHM.

Table 2. Eigenvalues l_i and contribution rates C_i of the oil data vector in the successive descending order.

	$i = 1$	2	3	4	5	6	7	8	9	10	11
l_i	6.3883	5.3891	2.5405	1.5635	1.0414	0.8416	0.6981	0.4927	0.4587	0.3762	0.3102
$l_i - l_{i+1}$	0.9992	2.8486	0.9769	0.5221	0.1998	0.1435	0.2054	0.0339	0.0826	0.0659	0.0836
C_i %	30.42	25.66	12.10	7.45	4.96	4.01	3.32	2.35	2.18	1.79	1.48
	12	13	14	15	16	17	18	19	20	21	
l_i	0.2266	0.1912	0.1496	0.1154	0.0768	0.0599	0.0437	0.0200	0.0116	0.0050	
$l_i - l_{i+1}$	0.0355	0.0416	0.0342	0.0386	0.0169	0.0162	0.0237	0.0084	0.0066	–	
C_i %	1.08	0.91	0.71	0.55	0.37	0.29	0.21	0.10	0.06	0.02	

PHM is one of the statistical regression models to calculate the hazard rate of the system for the purpose of life analysis,[18] which provides a mapping relationship between the CM data and the hazard rate of the system. Therefore, by monitoring the CM data of lubricating oil, the hazard rate of the oil can be obtained by using PHM.

The composition of the PHM is the failure rate, which is the product of a baseline hazard rate $h_0(t)$, and the positive link function $\psi(\gamma \cdot z_t)$, which represents the effect of the operational environment on the deterioration process. Thus, the hazard rate function for the proportional hazards modeling has the common form[17]

$$h(t, z_t) = h_0(t) \cdot \psi(\gamma \cdot z_t) \tag{15}$$

where $h(t,\ z_t)$ is the hazard rate under the condition of the covariates $\{z_t\}$ of lubricating oil, which is assumed to be a continuous time Markov process; $h_0(t)$ is the baseline hazard rate commonly using Weibull distribution, which has the form $h_0(t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}$, and β, η are the parameters of Weibull distribution; $\psi(\gamma \cdot z_t)$ is the positive link function commonly having the form $\psi(\gamma \cdot z_t) = \exp(\gamma \cdot Z(t))$, and γ is the vector of regression coefficient indicating the effect of the PCs of oil data on the hazard rate function. Therefore, the PHM using the Weibull proportional hazard function can be computed by equation (16) [17]

$$h(t, Z_t) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \cdot \exp(\gamma_1 \cdot z_1(t) + \ldots + \gamma_k \cdot z_k(t)) \tag{16}$$

Equation (16) can also be written as

$$\log(h(t, Z(t))) = \log\left(\frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1}\right) + \gamma \cdot Z(t) \tag{17}$$

For the oil data histories, dynamic PCs computed previously are regarded as the covariates for the PHM, $Z(t) = (PC1_t,\ PC2_t,\ \ PC3_t,\ PC4_t,\ PC1_{t-1},\ PC2_{t-1},\ PC3_{t-1},\ PC4_{t-1})'$. The columns of matrix $U = [u_1,\ u_2,\ u_3,\ u_4]$ are the eigenvectors corresponding to the first four largest eigenvectors of the correlation matrix R of the oil data, which is presented in equation (18).

The hazard rate at each sampling epoch of the lubricating oil is plotted in Figure 4. It can be seen that the hazard rate increases gradually. Specifically, the hazard rate increases slowly at the early stage of the oil deterioration, and with the deterioration of the lubricating oil, the hazard rate gets higher in the later period.

$$U = [u_1,\ u_2,\ u_3,\ u_4] = \begin{bmatrix} -0.1845 & 0.0665 & 0.1419 & -0.0560 \\ 0.2729 & -0.1131 & -0.1486 & 0.1577 \\ -0.1696 & 0.2910 & 0.1947 & 0.1544 \\ 0.0941 & -0.3273 & -0.2273 & 0.2630 \\ 0.1607 & -0.2664 & -0.2329 & -0.2776 \\ 0.1519 & -0.3565 & -0.1243 & -0.0248 \\ 0.0897 & -0.3331 & -0.0972 & 0.3073 \\ -0.1767 & -0.0663 & 0.0171 & 0.2254 \\ 0.2393 & 0.1935 & 0.1642 & -0.1775 \\ -0.3517 & 0.1100 & -0.0615 & 0.0051 \\ 0.2507 & -0.0207 & 0.4365 & 0.0733 \\ 0.3119 & -0.0795 & 0.0887 & -0.2972 \\ 0.3408 & -0.0613 & 0.1979 & -0.0018 \\ 0.2356 & 0.0037 & 0.4251 & 0.2702 \\ -0.0467 & -0.2102 & 0.0240 & 0.3835 \\ 0.0604 & 0.2118 & -0.4411 & -0.1322 \\ -0.2846 & -0.2530 & 0.0314 & -0.0176 \\ 0.1514 & 0.2682 & -0.2613 & 0.3211 \\ 0.2607 & 0.1885 & -0.1667 & -0.1264 \\ 0.2534 & 0.2535 & -0.1641 & 0.1063 \\ 0.1055 & 0.3044 & -0.1437 & 0.4032 \end{bmatrix} \tag{18}$$

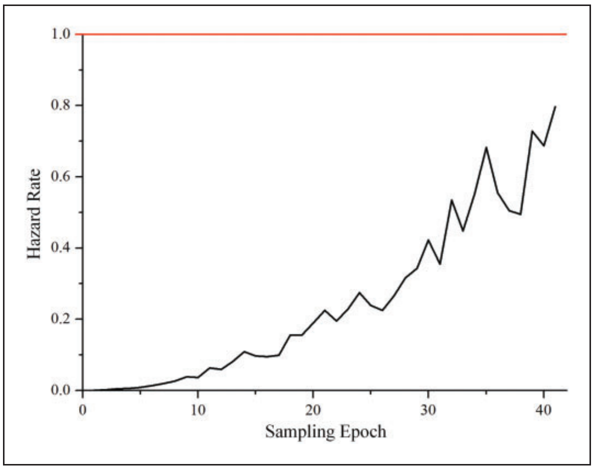

Figure 4. Hazard rate of the oil data history.

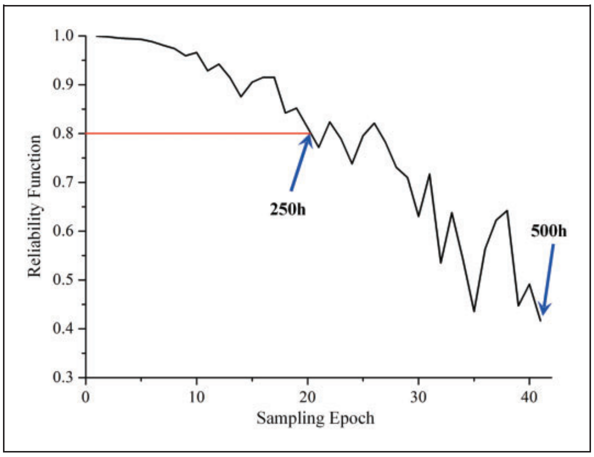

Figure 5. Conditional reliability function of the oil data histories.

RUL prediction of the lubricating oil

The CRF and MRL functions have been widely used for RUL prediction.[15,30] Therefore, CRF and MRL functions are employed in this paper to predict the RUL of lubricating oil. The CRF based on equation (16), which denotes the probability that the lubricating oil will not fail at time t, can be computed by

$$\begin{aligned} R(t, Z(t)) &= P(T > t | 0 \leqslant s \leqslant t) \\ &= \exp\left(-\int_0^t \exp(\gamma \cdot Z(s)) \mathrm{d}\left(\frac{s}{\eta}\right)^\beta \right) \end{aligned} \tag{19}$$

Then, the MRL function can be calculated by

$$MRL_t = \int_0^\infty R(t, Z(t)) \mathrm{d}t \tag{20}$$

The computed CRF for the selected failure history is illustrated in Figure 5.

It can be seen that the CRF of the PHM decreases gradually as the lubricating oil gets degraded. Meanwhile, the reliability decreases slowly at the early period of the oil deterioration, but drops faster in the late stage of deterioration. For the estimated results of the oil data history, the reliability remains high at time $t = 250$ h, and the oil can still fulfill its functions at time $t = 500$ h but with a low performance. The results indicate that the oil can still fulfill its functions at the end of the detection time as well. Therefore, the predicted results agree with the actual results very well.

Conclusions and future research

In this paper, we have applied a vector AR(2) model to represent the dynamic deterioration process of the lubricating oil by a seven-dimensional oil monitoring data obtained from the loader. The method of DPCA has been employed to reduce the data dimensionality. By doing this, four dynamic PCs have been selected using the scree test. Then, the four PCs have been regarded as the covariates for a Weibull PHM. RUL prediction of lubricating oil has finally been developed by deriving the explicit formulas of CRF and MRL functions. It is verified by the agreement between the predicted results and the actual results that the proposed approach can assess the deterioration of lubricating oil and provide a guidance to oil maintenance strategy as well.

However, the Weibull PHM in this paper assumes the values of the model parameters of Weibull distribution based on experience, and additionally, the number of oil data histories used in this paper is very limited. Due to the influencing factors, the prediction accuracy of the proposed model will be reduced. In future research, the pseudo-likelihood function, which can be used to estimate the unknown model parameters of Weibull distribution by maximizing the function iteratively, will be considered. As the precision of the Weibull PHM depends on the CM data, enough oil data histories will be collected in the subsequent research. Moreover, the PCs could be investigated and interpreted with the corresponding physical mechanisms, which is another research focus in the future work.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work has been supported by the National Science Foundation of China (Grant Nos. 11802237, 51975455 and 51675403), the Fundamental Research Funds for the Central Universities (Grant No.

G2018KY0306), NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1709215), and through the international collaborative plan of Shaanxi province (No. 2017kw-034). The authors also gratefully acknowledge the support of K. C. Wang Education Foundation.

ORCID iD

Shengxi Zhou https://orcid.org/0000-0002-4034-193X

References

1. Márquez FPG, Tobias AM, Pérez JMP, et al. Condition monitoring of wind turbines: techniques and methods. *Renewable Energy* 2012; 46: 169–178.
2. Vähäoja P and Pikkarainen H. Trends in industrial oil analysis – a review. *Int J Condition Monit* 2011; 1: 4–10.
3. Mortier RM, Orszulik ST and Fox MF. *Chemistry and technology of lubricants*, volume 107115. Berlin: Springer, 2010.
4. Plumley MJ, Wong V and Martins T. Oil degradation in a diesel engine with dual-loop lubricating system. *Tribol Trans* 2018; 61: 596–603.
5. De Feo M, Minfray C, Bouchet MDB, et al. Ageing impact on tribological properties of modtc-containing base oil. *Tribol Int* 2015; 92: 126–135.
6. Du Y, Wu T and Cheng J. Age detection of lubricating oil with on-line sensors. In: *IEEE SENSORS, 2015*, 2015, pp.1–4. USA: IEEE.
7. Yang YF, Wu QY, Wang YL, et al. Dynamic characteristics of cracked uncertain hollow-shaft. *Mech Syst Signal Process* 2019; 124: 36–48.
8. Vališ D, Žák L and Pokora O. Perspective approach in using anti-oxidation and anti-wear particles from oil to estimate residual technical life of a system. *Tribol Int* 2018; 118: 46–59.
9. Zhu X, Zhong C and Zhe J. Lubricating oil conditioning sensors for online machine health monitoring – a review. *Tribol Int* 2017; 109: 473–484.
10. Gewers FL, Ferreira GR, de Arruda HF, et al. Principal component analysis: a natural approach to data exploration [J]. *arXiv preprint arXiv:1804.02502*, 2018.
11. Makis V and Wu J. Effective fault detection and cbm based on oil data modeling and dpca. In: *Handbook of performability engineering*. Berlin: Springer, 2008, pp.825–841.
12. Wang W, Sun L, Lu X, et al. Application of dpca based stochastic filtering model and comparison of optimal cbm policies. In: *International conference on quality, reliability, risk, maintenance, and safety engineering (QR2MSE)*, 2013, pp.700–706. USA: IEEE.
13. Liu B, Liang Z, Parlikad AK, et al. Condition-based maintenance for systems with aging and cumulative damage based on proportional hazards model. *Reliab Eng Syst Saf* 2017; 168: 200–209.
14. Tang L, DeCastro J, Kacprzynski G, et al. Filtering and prediction techniques for model-based prognosis and uncertainty management. In: *Prognostics and health management conference, 2010, PHM'10,* 2010, pp.1–10. USA: IEEE.
15. Du Y, Wu T and Makis V. Parameter estimation and remaining useful life prediction of lubricating oil with HMM. *Wear* 2017; 376: 1227–1233.
16. Camci F and Chinnam RB. Health-state estimation and prognostics in machining processes. *IEEE Trans Automat Sci Eng* 2010; 7: 581–597.
17. Makis V, Wu J and Gao Y. An application of dpca to oil data for cbm modeling. *Eur J Oper Res* 2006; 174: 112–123.
18. Ghasemi A, Yacout S and Ouali MS. Evaluating the reliability function and the mean residual life for equipment with unobservable states. *IEEE Trans Reliab* 2010; 59: 45–54.
19. Li Y and Nilkitsaranont P. Gas turbine performance prognostic for condition-based maintenance. *Appl Energy* 2009; 86: 2152–2161.
20. Si XS, Wang W, Hu CH, et al. Remaining useful life estimation–a review on the statistical data driven approaches. *Eur Jo Oper Res* 2011; 213: 1–14.
21. Lin D, Banjevic D and Jardine AK. Using principal components in a proportional hazards model with applications in condition-based maintenance. *J Oper Res Soc* 2006; 57: 910–919.
22. Jardine AK and Tsang AH. *Maintenance, replacement, and reliability: theory and applications*. Boca Raton: CRC Press, 2005.
23. Pham HT, Yang BS, Nguyen TT, et al. Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine. *Mech Syst Signal Process* 2012; 32: 320–330.
24. Zhou S and Zuo L. Nonlinear dynamic analysis of asymmetric tristable energy harvesters for enhanced energy harvesting. *Commun Nonlinear Sci Numer Simul* 2018; 61: 271–284.
25. Zzeyani S, Mikou M, Naja J, et al. Spectroscopic analysis of synthetic lubricating oil. *Tribol Int* 2017; 114: 27–32.
26. Kim MJ, Makis V and Jiang R. Parameter estimation for partially observable systems subject to random failure. *Appl Stochastic Models Bus Ind* 2013; 29: 279–294.
27. Reinsel GC. *Elements of multivariate time series analysis*. Berlin: Springer Science & Business Media, 2003.
28. Purushotham V, Narayanan S and Prasad SA. Multifault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. *NDT E Int* 2005; 38: 654–664.
29. Cattell RB. The scree test for the number of factors. *Multivariate Behav Res* 1966; 1: 245–276.
30. Khaleghei A and Makis V. Model parameter estimation and residual life prediction for a partially observable failing system. *Nav Res Log* 2015; 62: 190–205.