Health index extraction for power-shift steering transmission using selected oil field data

Jianhua Chen¹,², Biao Ma¹, Shufa Yan¹,³ and Changsong Zheng¹

Abstract
The concentration of wear debris in used lubricating oil is generally observed from spectral oil analysis. The element concentration data are one of the commonly used oil field data for degradation evaluation of power-shift steering transmission. However, in practical applications, the underlying degradation degree of the power-shift steering transmission is difficult to evaluate due to the complexity of element concentration data. To solve this problem, we propose a health index extraction methodology using a weighted average method to better characterize the underlying degradation, which leads to an accurate estimation of the residual life before the power-shift steering transmission needs to be maintained. The extracted health index not only maximizes the monotonic trend of underlying degradation but also minimizes the failure threshold variance between different power-shift steering transmissions. The method includes element concentration data modification, data selection, and data fusion steps that result in a reasonable power-shift steering transmission degradation model. Finally, a case study is provided to illustrate the proposed method. The results show that the extracted health index outperforms each selected element concentration data.

Keywords
Power-shift steering transmission, health index, oil field data, wear debris, residual life

Date received: 3 October 2019; accepted: 17 March 2020

Handling Editor: James Baldwin

Introduction
Power-shift steering transmission (PSST) is one of the vital components that widely used in many heavy tracked vehicles. With PSST operating, metal wear debris spalling from each mechanical pair evenly mixed in lubricating oil and its concentration increase, which is a slow degradation process. Metal wear debris mixed in lubricating oil accelerates wear and tear of every mechanical components, consequently leading to the degradation of the PSST system.¹,² After a certain degree of degradation, the PSST system will fail to perform normal operations. For mechanical powertrains, faults caused by oil contamination are the primary failure modes, with more than 50% associated with metal wear debris.¹ Failure of the PSST will result in production downtime and delays, which often leads to economic losses and even safety problems that, in turn, may lead to catastrophic consequences. Thus, the PSST system should be regularly monitored, and the system soft failure occurrence time should be accurately estimated to perform predictive maintenance (PM). To address this issue, lubricant condition monitoring

¹School of Mechanical Engineering, Beijing Institute of Technology, Beijing, China
²Jianglu Machinery and Electronics Group Company, NORINCO Group, Xiangtan, China
³Department of Mechanical & Industrial Engineering, University of Toronto, Toronto, ON, Canada

Corresponding author:
Shufa Yan, School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China.
Email: yanshufa1990@163.com

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
(LCM) technology has been used to observe oil field data for oil-lubricated machines.\textsuperscript{3,4} The purpose of this article is to address a health index (HI) extraction problem for PSST with the observed oil field data to determine the expected residual life (RL) before the PSST needs to be maintained.

In the literature of LCM,\textsuperscript{5–7} the amount of wear debris is one of the most common types of degradation data to assess the severity of system degradation. The amount of wear debris has been sampled from oil spectral analysis using MOA II (atomic emission spectroscopy) during the oil sampling period.\textsuperscript{8} Using this method, 15 groups of element concentration data sample can be obtained. With the sampled element concentration data, prognostic analysis can then be performed to estimate the RL of an operating PSST.\textsuperscript{5,10} In the existing literature, many techniques and methods have been used to model the relationship between system degradation of oil-lubricated machines and its association with element concentration data.\textsuperscript{11–16} For example, Wang et al. described the analysis procedure of oil field data in the study of Wang and Burairah\textsuperscript{17} and developed a machine degradation model using stochastic diffusion process in their study\textsuperscript{18} by assuming the oil field data are a valuable condition indicator for characterizing the underlying degradation. In this article, the rationality of using element concentration data for degradation modeling was investigated. Recently, by modeling the increasing trend of wear particles in oil, Vališ and Žák\textsuperscript{19} built a system degradation model for an internal combustion engine to determine the expected moment when a soft failure occurred. It is noted that the expected moment can be used as the system RL and further be set as the time to perform planned PM. In the most recent, Yan et al.\textsuperscript{20} presented a condition-based maintenance (CBM) problem with selected oil field data to determine the optimal time of machine maintenance. In addition, the work of Yan et al.\textsuperscript{20} presents a framework of using sampled oil field data for maintenance optimization. A review of the application of oil field data for LCM and CBM can be found in Wakiru et al.\textsuperscript{21} and the references therein.

Most of the existing literature simply use single-element data (e.g., Fe, Cu, and Mo\textsuperscript{9–12}) in wear debris concentrations to establish degradation models. In practice, however, many studies have shown that using a single-element concentration, data-based degradation model is problematic,\textsuperscript{22,23} which may cause inaccuracy in system RL estimation.

To our knowledge, no HI extraction method has developed in the literature that can be utilized for the fusion of oil field data. Therefore, this article seeks to fulfill this gap by fusing multiple element concentration data to extract an HI to characterize system degradation process that can be used for RL estimation. To address this issue, a unified degradation model method has been proposed in our previous research\textsuperscript{24} in which a copula function is utilized to deal with multiple oil field data. However, it is often challenging to determine an appropriate copula function, especially when dealing with more oil field datasets. Unlike the existing works, this article proposes an HI extraction methodology by developing a quadratic programming problem that can simultaneously maximize the HI monotonicity and minimize the failure threshold variance of different units. With the proposed efforts, we expected to attain more accurate RL estimation.

The remaining parts are organized as follows. The “Overview of the element concentration data” section describes the PSST system and the oil field data used in this article. The “Development of an HI extracting method” section describes the key steps associated with the preprocessing of selected oil field data, including element concentration modification for lubricating oil supplement and element concentration data selection, and develops explicit formulas for extracting an HI based on quadratic programming for degradation modeling and RL estimation. The “Case study” section provides an illustrative case study. Finally, the “Discussion and conclusion” section provides the conclusion and future research.

**Overview of the element concentration data**

**System model description**

This article considers a PSST system monitored using regular oil spectral analysis to implement LCM and predict its soft failure occurrence. The concerned PSST system combines a multi-speed shifting system with an infinite steering system, and is widely used in heavy industrial applications such as heavy-duty tracked vehicles. Figure 1 shows a sketch diagram of a PSST, and

![Figure 1. Sketch of the PSST.](attachment:image.png)

1: hydraulic torque convertor; 2: CV clutch; 3: CH clutch; 4: first shaft; 5: steering pump; 6: second shaft; 7: C1C2 clutch; 8: third shaft; 9: steering motor; 10: C3 clutch; 11: CLCR clutch.
Figure 2 represents the configuration of the test-bed. The PSST was inspected every 5 motor-hour (Mh) during the operational life. The sample collection criteria are listed in Table 1, and detailed sampling principles can be found in Yan et al.7,8

Table 1. Criterion of sample acquiring.

| Position                  | Operating condition | Interval                   |
|---------------------------|---------------------|----------------------------|
| Entrance of fine-filter   | Different gear stages | Load of each gear stages   |
|                           | Variable load       | 50%, 75%, 100%             |

Figure 2. Life cycle test-bed of PSST. 1: diesel engine; 2, 4, 5: torque and speed sensor; 3: PSST; 6, 7: moment of inertia; 8, 9: loading pump.

Table 2. Element concentration data of one PSST (unit: µg/g).

| Sample | Time/Mh | Ca  | P   | Zn  | Fe  | Mg  | Na  | Cu  | Pb  | Si  | Al  | Ni  | Cr  | Mn  | Mo  | Sn  |
|--------|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1      | 5       | 2368| 1046| 1020| 12.8| 18.1| 14.7|10.5 | 5.8 | 3.2 | 0.5 | 0.2 | 0.0 | 0.0 | 0.2 |
| 2      | 10      | 2536| 1020| 1090| 19.1| 21.8| 16.5|10.3 | 6.5 | 5.3 | 3.1 | 0.5 | 0.6 | 0.8 | 0.5 |
| 3      | 15      | 2543| 1031| 1080| 19.7| 20.9| 15.3|11.1 | 7.0 | 4.9 | 3.1 | 0.4 | 0.4 | 0.9 | 0.6 |
| ...    | ...     | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 35     | 175     | 2380| 943 | 969 | 604 | 16.6| 11.2|17.3 | 6.1 | 4.3 | 3.2 | 9.9 | 5.3 | 4.3 | 0.5 |
| 36     | 180     | 2468| 1014| 1039| 648 | 18  | 10.6|18   | 7.2 | 4.5 | 3.5 | 10.3| 5.5 | 4.6 | 0.9 |

PSST: power-shift steering transmission.

Dataset description

The element concentration dataset used herein includes 20 training units and 5 test units. Each unit was run to failure due to severe wear under the same operating conditions. Time was measured in inspection periods, in this case, 5 Mh. All of the collected samples were analyzed immediately. The concentration data in parts per thousand of 15 types of element were obtained using MOA II, which are related to the degradation process of the operating PSST.7,8 The element concentration data of one unit are shown in Table 2.

As mentioned above, modeling the degradation mechanism and representing the degradation state of the operating PSST are difficult due to the information redundancy of multiple element concentration data.20 Thus, to characterize the degradation level of the operating PSST, in the next section, a composite HI will be extracted based on the fusion of the element concentration data.

Development of an HI extracting method

In this section, several key steps associated with the preprocessing of the element concentration data and constructing of a composite HI are discussed.

Dataset preprocessing

Modification of element concentration. Since a certain volume of used lubricating oil is sampled for LCM, an equal amount of unpolluted oil is then injected. The concentrations of wear particle in the subsequent inspection periods are not ideal; specifically, it is less than the actual concentration. Thus, the first step to enable an accurate degradation condition characterization is to modify the original element concentration data during preprocessing. Thus, a function was proposed to modify the element concentration data

\[
x_{i,j} = \frac{x_{i,j}V_0 + \sum_{k=1}^{j} [x_{i,k} - x_0] V_s}{V_0}
\]

where \(x_{i,j}\) and \(x_{i,j}'\) are the measured value of the wear particle concentrations and the actual concentrations for unit \(i\) in the inspection period \(j\); \(x_0\) represents the element concentration data of the unpolluted oil; \(V_0\) represents the initial volume of the lubricating oil; and \(V_s\) is
the sampling size. To eliminate the influence of initial wear particles, the concentration value can be obtained by subtracting \(x_0\) from \(x'_{i,j}\).

**Selection of element concentration data.** Recalled that 15 types of major element concentration data are obtained from oil spectral analysis using MOA II. Therefore, it is necessary to determine the element concentration data used as the input of our HI extraction method. In engineering practice, time-series data that exhibit a significant increasing or decreasing trend during operation are of interest. Based on this criterion, if the last sample is larger or smaller than the initial sample, the element concentration data are selected.25

**Methodology development**

Researchers usually use weighted average functions to fuse multiple time-series data and extract indicators that can characterize the implicit information.22,26 Therefore, to use the selected element concentration data, we use the weighted average function to formulate the HI extracting method, represented as

\[
d_i = \mathbf{X}_{i, \omega}
\]

where \(\omega \in \mathbb{R}^{n \times 1}\) represents the vector of weight coefficients that is used for fusing multiple element concentration data at each inspection period of each unit, and \(s\) represents the number of selected element concentration data; \(d_{i,j}\) and \(\mathbf{X}_{i,j}\) are the values of the extracted HI and the vector of the selected element concentration data for unit \(i\) in inspection period \(j\); \(\omega' \mathbf{M}^{(1)} = 1\), where \(\mathbf{M} \in \mathbb{R}^{n \times s}\) is a diagonal matrix used for representing the degradation trend information, of which the diagonal element is \(1\) \((-1)\) when the corresponding selected element concentration data have an increasing (decreasing) trend.

**Desirable properties for extracting the HI.** To guarantee the obtained HI is more suitable for LCM and RL estimation than the element concentration data, a few desirable properties for the HI are defined to enhance its effectiveness for reasonable degradation modeling and successful RL estimation.

The element concentration data are always nonmonotonic due to measurement errors; thus, an HI that has a monotonic trend should be extracted. Furthermore, for given element concentration data, the failure thresholds between different units are always different, so the obtained HI should have little variation in the failure threshold. Considering these two requirements, the two basic properties, which were proposed by Saxena et al.21 to make the condition monitoring data have a better application in degradation modeling and RL prediction, are adopted for the extracted HI to have a better LCM application.

**Property 1.** The degradation trend of the extracted HI should be monotonic with the PSST system degradation.

**Property 2.** The variance in the failure threshold of different PSST systems should be minimal under the same unit and operating condition.

**Methodology optimization.** Since each PSST system is assumed to fail due to wear debris accumulation under the same operating condition, we expect to extract an HI that has a similar degradation pattern for all systems. Thus, the expectation can be achieved by jointly programming these two properties when extracting the HI. Therefore, the objective function should consist of two parts: the weighted amount of the nonmonotonicity of Property “1” and the variance of the failure threshold of Property “2.”

For Property “1,” a slack variable is used to compensate for monotonic conflicts. In the case where the value of HI is less than the previous inspection period, the difference between HI and the reasonable value is measured using a slack variable. Specifically, \(\bar{e}_{i,j} = \max(d_{i,j-1} - d_{i,j}, 0)\). Then, the total weighted amount of confliction in monotonicity is obtained by

\[
\sum_{i=1}^{m} \sum_{j=1}^{n-1} c_{i,j} e_{i,j}
\]

where \(e_{i,j}\) represents a slack variable used to quantify the degree of confliction in monotonicity of unit \(i\) in inspection period \(j\); \(c_{i,j}\) represents the weight coefficient of the slack variable \(e_{i,j} \); \(m\) represents the number of training units; and \(n_i\) represents the total number of inspection periods of unit \(i\).

For Property “2,” the variance of the failure threshold of the HI is measured using an unbiased inspection variance, which can be formulated as

\[
\frac{(Y\omega)'(Y\omega) - m((Y')\omega)/m^2)}{(m - 1)} = \omega Y'DY\omega
\]

where \(Y \in \mathbb{R}^{m \times s}\) is a matrix in which rows represent each unit and columns represent each selected element concentration data, and \(D \in \mathbb{R}^{m\times s}\) represents a symmetric matrix of the form: \(D = (I - O)/m(m - 1)\), where \(O\) is a matrix of full values of 1, and \(I\) is an identity matrix.

Above all, by maximizing the monotonic trend and minimizing the threshold variance between different units, the method optimization model is formulated as a quadratic programming problem, represented as

\[
\text{obj} = \min_{\omega, e_{i,j}} \sum_{i=1}^{m} \sum_{j=1}^{n-1} c_{i,j} e_{i,j} + (1 - r)\omega Y'DY\omega
\]

s.t. \(M\omega \geq 0, E_i\omega_i \geq 0\)
where \( r \) is a tuning parameter used to measure the relative importance of these two properties in the optimization model, and \( \mathbf{E}_i \in \mathbb{R}^{(n_i-1) \times (s + n_i - 1)} \) is a matrix used to maximize the monotony for unit \( i \), which can be formulated as

\[
\mathbf{E}_i = \begin{bmatrix}
  x_{i,1,2} - x_{i,1,1} & \ldots & x_{i,s,2} - x_{i,s,1} \\
  \vdots & \ddots & \vdots \\
  x_{i,1,n_i} - x_{i,1,n_i-1} & \ldots & x_{i,s,n_i} - x_{i,s,n_i-1}
\end{bmatrix}
\]

where \( x_{i,k,j} \) is the \( k \)th type of element concentration data of unit \( i \) in inspection period \( j \), and \( \mathbf{\omega} \in \mathbb{R}^{(s + n_i - 1) \times 1} \) is a vector that needs to be determined to ensure that the extracted HI is monotonic after introducing the slack variable \( e_{i,j} \). The vector \( \mathbf{\omega} \) can be formulated as

\[
\mathbf{\omega} = \left[ \omega', e_{i,1}, \ldots, e_{i,n_i-1} \right]^T
\]

**Parameter setting.** In this section, how the crucial parameters in our optimization model are set is discussed such that the obtained HI is reasonable for the operating conditions and degradation mechanism. Specifically, we focus on setting weight coefficients \( c_{i,j} \) and tuning parameter \( r \).

1. **Weight coefficient setting:** Recalling the each slack variable \( e_{i,j} \), \( c_{i,j} \) is the corresponding weight coefficient as unit \( i \) degrades and \( c_{i,j}e_{i,j} \) measures the weighted amount of confliction in monotonicity for unit \( i \) in inspection period \( j \). As a unit degrades, the accuracy of degradation modeling and RL prediction becomes increasingly sensitive to the number of monotonic conflicts of HI. Therefore, as the inspection period \( j \) increases, the slack variable \( e_{i,j} \) is assigned higher weights. To be specific, the weight coefficient \( c_{i,j} \) must satisfy the following condition

\[
c_{i,j+1} \geq c_{i,j} \geq 0, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n_i - 1
\]

The value of weight coefficient \( \{c_{i,j}\} \) is further assumed to follow an arithmetic series \( (2c_{i,j} = c_{i,j+1} + c_{i,j-1}) \) and the weight coefficients have the following form

\[
c_{i,j} = c_{i,1} + (j - 1) \frac{2 - 2c_{i,1}n_i}{(n_i - 1)n_i}, \quad j = 1, \ldots, n_i
\]

Since more weight is assigned to \( c_{i,j} \) as the inspection period \( j \) increases, \( \mathbf{\omega} \), the estimated weight coefficient vector, becomes increasingly dominated by the element concentration data at inspection period close to failure time. Note that the arithmetic series assumption may not be suitable in some applications; thus, the geometric series \( (c_{i,j} = c_{i,1}c_{i,j-1}) \) and other assumptions may be used to initialize the optimization model based on the emphasis on the monotonic confliction.

2. **Tuning parameter setting:** The relative importance of the two terms in the optimization model in equation (5), that is, confliction in monotonicity and variance in the failure threshold, is controlled by the tuning parameter \( r \). We can use cross-validation to obtain the optimal value of \( r \). Details of the calculation process can be found in Kohavi.\(^{28}\)

Increasing \( r \) puts more emphasis on reducing the confliction of the monotonicity property of the extracted HI at the expense of increasing the variance of the failure threshold. This multi-objective optimization problem is often solved by plotting the efficient frontier with regard to these two items of the optimization function in equation (5). In practice, we set the optimal value of tuning parameter \( r \) based on the importance assigned to the two items of the optimization function.

**Flowchart of the proposed method**

Figure 3 shows the flowchart of the proposed HI extraction method. Using this method, the optimal values of \( \mathbf{\omega} \) and \( r \), represented by \( \mathbf{\omega}^* \) and \( r^* \), can be obtained. Then, the HI for LCM and RL prediction can further be extracted based on \( \mathbf{\omega}^* \) and the selected element concentration data.

**Case study**

In this section, we provide a numerical study using element concentration data collected for inspections from
each PSST system in the “Overview of the element concentration data” section to illustrate the HI extraction procedure. To challenge the developed method, we further investigated the performance of our proposed method and compared it with the existing single-element concentration data-based degradation modeling method. Specifically, we compare the accuracy of the RL prediction by using the constructed HI and using each selected element concentration data based on the same Wiener process (WP)–based degradation model.

**Dataset preprocessing and HI extraction**

Concentration modification and data selection. After the wear particle concentration is modified using equation (1), the element concentration data used for extracting HI are then selected. The selection is based on the modified element concentration data showing a consistent decreasing or increasing trend for all units. Among the 15 types of element concentration data shown in Table 2, six (i.e. \( s = 6 \)) element concentrations are selected for HI extraction and degradation modeling, namely, Fe, Cr, Cu, Ni, Mn, and Mo. As a result, the corresponding diagonal elements of \( \mathbf{M} \) are set as \([1, 1, 1, 1, 1, 1]\).

Weight coefficient and tuning parameter setting

1. Weight coefficient setting: Since we assume that the weight coefficient increases linearly in the “Development of an HI extracting method” section, we adopt an arithmetic series for \( \{c_{i,j}\} \). These weight coefficients represent the confliction in monotonicity for Property “1.” Furthermore, we select the weighted average function to demonstrate the HI extraction method.

2. Tuning parameter setting: In this case study, the tuning parameter was set as \( r = 0.5 \) to illustrate the proposed method. The optimal model in equation (5) is solved based on the selected element concentration data. Table 3 shows the optimal weight \( \mathbf{\omega}^* \). Note that the value of \( r \) can be selected based on the importance assigned to each property when extracting the HI.

**WP-based degradation modeling**

Degradation model development. The PSST system initially operates in a healthy state with unpolluted oil and is subject to wear particles obtained from regular inspections. The selected six element concentration data are shown in Figure 4.

With the plots of element concentration data in Figure 4, the element concentration data show random diffusion forms. Therefore, the WP-based degradation model proposed in Si et al.\(^ {29} \) is utilized for evaluating the performance of the extracted HI when used for RL prediction. The degradation model is given by

\[
X(t) = X(0) + \sigma B(t) + \theta t
\]

where the degradation process \( \{X(t), t \geq 0\} \) is driven by a standard Brown movement \( \{B(t), t \geq 0\} \); \( \sigma \) is the diffusion parameter; \( \theta \) is an unknown drift coefficient characterizing the degradation rate; \( \sigma B(t) \sim N(0, \sigma^2 t) \) denotes the randomness and time-varying uncertainty of the degradation process.

It is assumed that a PSST system is identified as failed and that needs to be replaced when \( \{X(t), t \geq 0\} \) exceeds the failure threshold. Therefore, the RL of PSST system is defined based on the first hitting time (FHT) of the degradation process. Given the failure threshold \( \omega \), RL \( L_k \) at time \( t_k \) can be formulated as

\[
L_k = \inf \{l_k > 0 : X(l_k + t_k) \geq \omega\}
\]

Based on the Markov property of the WP, the degradation track of PSST system after time \( t_k \) is given the following

\[
X(t) = x_k + \theta l_k + \sigma B(t - t_k), t \geq t_k
\]

According to the above description of the degradation model, if time \( t \) is the FHT of the degradation process \( \{X(t), t \geq 0\} \), then the RL of PSST system at time \( t_k \) can be represented as \( t - t_k \). Assume that

**Table 3. Optimal value of \( \mathbf{\omega}^* \)**

| Element | Fe     | Cr     | Cu     |
|---------|--------|--------|--------|
| Value   | 0.2913 | 0.1785 | 0.2571 |
| Element | Ni     | Mn     | Mo     |
| Value   | 0.1087 | 0.0875 | 0.0867 |

**Figure 4.** Curves for element concentrations in a training unit.
\( l_k = t - t_k, \bar{X}(l_k) = X(t) - x_k \) and \( w_k = \omega - x_k \), then the RL of PSST system at time \( t_k \) is equal to the FHT of the degradation process \( \{X(l_k), l_k \geq 0\} \)

\[
\bar{X}(l_k) = \theta l_k + \sigma(B(l_k + t_k) - B(t_k))
\]

Under the above construction, the FHT of the WP conforms to an inverse Gauss distribution. Thus, the RL of PSST system at time \( t_k \) is denoted as \( x_k \).

Using the maximum likelihood estimation (MLE) based parameter estimation. To initialize the model defined in equation (8), we estimate the model parameters \( \sigma^2, \theta \) using the maximum likelihood estimation (MLE) based on the element concentration data of the training unit. The element concentration data of the \( i \)th training unit at time \( t_j \) are denoted as \( x_{i,j} \), and the entire dataset is \( \{X(t_j) = x_{i,j}, i = 1, \ldots, N, j = 1, \ldots, M\} \). The degradation model parameter vector is denoted as \( \Theta = (\sigma^2, \theta)^T \). Then, the likelihood function \( \xi(\Theta | X) \) of all element concentration data histories is expressed as

\[
\xi(\Theta | X) = \frac{-NM}{2} \ln(2\pi) - \frac{N}{2} \ln|\Omega| - \frac{1}{2} \sum_{i=1}^{N} (x_i - \theta t_i)^T \Omega^{-1} (x_i - \theta t_i)
\]

where \( x_i \sim N(\mu, \Omega) \sim N(\theta t, \sigma^2 Q) \), \( Q = \min\{t_i, t_j\}, 1 \leq i,j \leq M \). The MLE of \( \sigma^2, \theta \) can be easily obtained by maximizing \( \xi(\Theta | X) \). A detailed description about the MLE method can be found in Pecht and Jain.\(^{30}\)

Table 4 shows the estimated value of \( \sigma^2 \) for all selected element concentration data and the extracted HI, from which we can see that the extracted HI fits the degradation process better than each original element concentration data.

The dataset of the last inspection period before failure in all training units for element concentration data \( k \) is denoted as \( X_{m,k,n} = [X_{1,k,n}, \ldots, X_{M,k,n}] \). The variance of the failure threshold for element concentration data \( k \) is denoted as \( \nu^f_k \), as shown in Table 5. We can see that the \( \nu^f_k \) in the extracted HI is smaller than that in any other selected element concentration data.

\( \sigma^2 \) and \( \nu^f_k \) into consideration, the extracted HI conforms to the two desirable properties.

**Prediction results**

Using the estimated parameters, the degradation model can be initialized to predict the RL of the testing units. Figure 5 shows the degradation state of a random testing unit. The FHT of the extracted HI is 180 Mh, which represents the failure period of PSST system degradation that provides a reference for the maintenance policy formulation. In addition, Figure 6 shows the predicted conditional distribution of the degradation state at several inspection periods, which characterizes the uncertainty degradation of the PSST system.

The relative error between the RL prediction and actual RL is calculated and the comparison considers two cases: (1) the RL prediction based on the extracted HI and (2) the RL prediction based on each selected element concentration data. Specifically, the relative error, \( \text{err}_{r,k} \), is defined as the relative difference between

| Table 4. Estimated value of \( \sigma^2 \). |
|-----------------|-------------|-------------|-------------|-------------|
| Element | Fe | Cr | Cu | Ni |
| Value/10\(^{-3}\) | 2.258 | 2.775 | 2.487 | 3.879 |
| Element | Mn | Mo | HI | – |
| Value/10\(^{-3}\) | 3.236 | 4.246 | 0.6983 | – |

| Table 5. Calculated value of \( \nu^f_k \). |
|-----------------|-------------|-------------|-------------|-------------|
| Element | Fe | Cr | Cu | Ni |
| Value/10\(^{-3}\) | 0.867 | 1.258 | 1.025 | 2.153 |
| Element | Mn | Mo | HI | – |
| Value/10\(^{-3}\) | 1.856 | 1.958 | 0.3358 | – |
the RL prediction values and the actual RLs for unit \( i \) and element concentration data \( k \), and is given by

\[
err_{i,k} = \frac{\hat{T}_{i,k} - T_i}{n_{i}t_s + T_i}
\]

where \( T_i \) represents the actual RL for testing unit \( i \), \( \hat{T}_{i,k} \) represents the RL prediction for unit \( i \) and element concentration data \( k \), \( n_i \) represents the number of inspection periods of unit \( i \), and \( t_s \) represents the inspection interval (i.e. \( t_s = 5 \text{Mh} \)).

Figure 7 shows the absolute value of the relative error using each selected element concentration data and the extracted HI at different quantiles of the machine operating time.

Figure 7 shows that (1) compared with each selected element concentration data, the extracted HI provides the best prognostic result due to the control of the two desirable properties. In other words, the monotonic property is maximized and the variance in the failure threshold is minimized when extracting the HI. (2) The RL prediction using the extracted HI becomes increasingly accurate with the unit operating due to the control of the weight coefficient \( c_{i,j} \). Specifically, given more penalties to the slack variables, \( \{ \kappa_{i,j} \} \) in the inspection periods are closer to failure using the arithmetic series for \( \{ c_{i,j} \} \). These useful characteristics have practical applications when dealing with equipment that has high reliability requirements.

In addition to the relative errors shown in Figure 7, researchers are often interested in comparing the root mean square error (RMSE)\(^3\) of the RL predictions and actual RLs. A small RMSE indicates a better RL prediction with less absolute error. The RMSE value for all selected element concentration data and the HI is shown in Table 6. Based on Table 6, the HI extracted using our proposed method has the smallest RMSE compared with using each selected element concentration data. The extracted HI results in a more accurate RL prediction results, which provides a useful foundation for an optimal PM strategy for oil-lubricated systems.

### Discussion and conclusion

This article provides a systematic methodology that includes concentration modification, data selection, data generalization, and data fusion procedures combining element concentration data (e.g. Fe, Cu) obtained for inspections from a PSST system to extract an HI that accurately characterizes the system degradation condition. The novelty of this method is integrating multi-dimensional element concentration data in a unified HI. The constructed method is advanced in that it can maximize the monotonicity of the indicator and minimize the variance in the failure threshold simultaneously. The developed method was tested and validated using element concentration data from several PSST systems. The WP-based degradation model was utilized to evaluate the validity of the extracted HI by estimating the RL of each oil-lubricated system in time. The results show the improved performance of the extracted HI compared with each selected original element concentration data.
There are several important directions deserving further studies. First, more oil field data (e.g., ferrography, viscosity, and acidity) tailored to LCM are necessary. Second, kernel methods that can fuse nonlinear time-series data should be investigated. Third, an effective maintenance strategy optimization method based on the extracted HI should be developed.

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 is partially supported by the National Natural Science Foundation of China (NSFC) under grant numbers 51475044 and 51975047, and partially supported by the China Scholarship Council under grant number 201806030083.

ORCID iD
Shufa Yan https://orcid.org/0000-0002-6122-146X

References
1. Sheng S. Monitoring of wind turbine gearbox condition through oil and wear debris analysis: a full-scale testing perspective. Tribol T 2016; 59: 149–162.
2. Zheng C, Liu P, Liu Y, et al. Oil-based maintenance interval optimization for power-shift steering transmission. Adv Mech Eng 2018; 10: 1687814018760921.
3. Fan B, Li B, Feng S, et al. Modeling and experimental investigations on the relationship between wear debris concentration and wear rate in lubrication systems. Tribol Int 2017; 109: 114–123.
4. Salameh JP, Cuet S, Etien E, et al. Gearbox condition monitoring in wind turbines: a review. Mech Syst Signal Pr 2018; 111: 251–264.
5. Kumar P, Hirani H and Agrawal AK. Online condition monitoring of misaligned meshing gears using wear debris and oil quality sensors. Ind Lubr Tribol 2018; 70: 645–655.
6. Hong W, Wang S, Tomovic MM, et al. A novel indicator for mechanical failure and life prediction based on debris monitoring. IEEE T Reliab 2017; 66: 161–169.
7. Yan S, Ma B and Zheng C. Health index extracting methodology for degradation modelling and prognosis of mechanical transmissions. Eksploaat Niezawodn 2019; 21: 137–144.
8. Yan S, Ma B, Zheng CS, et al. Weighted evidential fusion method for fault diagnosis of mechanical transmission based on oil analysis data. Int J Auto Tech Kor 2019; 20: 989–996.
9. Du Y, Wu T and Makis V. Parameter estimation and remaining useful life prediction of lubricating oil with HMM. Wear 2017; 376: 1227–1233.
10. Wang D, Tsui KL and Miao Q. Prognostics and health management: a review of vibration based bearing and gear health indicators. IEEE Access 2018; 6: 665–676.
11. Loutas TH, Roulias D, Pauly E, et al. The combined use of vibration, acoustic emission and oil debris on-line monitoring towards a more effective condition monitoring of rotating machinery. Mech Syst Signal Pr 2011; 25: 1339–1352.
12. Liu Y, Ma B, Zheng CS, et al. Research and evaluation on wear in power-shift steering transmission through oil spectral analysis with RKPCA method. Spectrosc Spect Anal 2015; 35: 1370–1375.
13. Kattelus J, Miettinen J and Lehtovaara A. Detection of gear pitting failure progression with on-line particle monitoring. Tribol Int 2018; 118: 458–464.
14. Cao W, Dong G, Xie YB, et al. Prediction of wear trend of engines via on-line wear debris monitoring. Tribol Int 2018; 120: 510–519.
15. Yan SF, Ma B and Zheng CS. Remaining useful life prediction for power-shift steering transmission based on fusion of multiple oil spectra. Adv Mech Eng 2018; 10: 1687814018784201.
16. Yan SF, Ma B and Zheng CS. Condition-based maintenance for power-shift steering transmission based on oil spectral analysis. Spectrosc Spect Anal 2019; 39: 3470–3474.
17. Wang W and Burairah H. Plant residual time modelling based on observed variables in oil samples. J Oper Res Soc 2009; 60: 789–796.
18. Wang W, Hussin B and Jefferis T. A case study of condition based maintenance modelling based upon the oil analysis data of marine diesel engines using stochastic filtering. Int J Prod Econ 2012; 136: 84–92.
19. Vališ D and Žilk L. Contribution to prediction of soft and hard failure occurrence in combustion engine using oil tribo data. Eng Fail Anal 2017; 82: 583–598.
20. Yan S, Ma B, Zheng CS, et al. An optimal lubricating oil replacement method based on selected oil field data. IEEE Access 2019; 7: 92110–92118.
21. Wakiru JM, Pintelon L, Muchiri PN, et al. A review on lubricant condition monitoring information analysis for maintenance decision support. Mech Syst Signal Pr 2019; 118: 108–132.
22. Liu K, Gebrael NZ and Shi J. A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis. IEEE T Autom Sci Eng 2013; 10: 652–664.
23. Yang F, Habibullah MS, Zhang T, et al. Health index-based prognostics for remaining useful life predictions in electrical machines. IEEE T Ind Electron 2016; 63: 2633–2644.
24. Yan S, Ma B and Zheng C. A unified system residual life prediction method based on selected tribodiagnostic data. IEEE Access 2019; 7: 44087–44096.
25. Liu L, Wang S, Liu D, et al. Quantitative selection of sensor data based on improved permutation entropy for system remaining useful life prediction. Microelectron Reliab 2017; 75: 264–270.
26. Chen Y, Fung RYK and Tang J. Rating technical attributes in fuzzy QFD by integrating fuzzy weighted average
method and fuzzy expected value operator. *Eur J Oper Res* 2006; 174: 1553–1566.

27. Saxena A, Celaya J, Balaban E, et al. Metrics for evaluating performance of prognostic techniques. In: *Proceedings of international conference on prognostics and health management*, Denver, CO, 6–9 October 2008. New York: IEEE.

28. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of international joint conference on artificial intelligence*, Montreal, QC, Canada, 20–25 August 1995. Burlington, MA: Morgan Kaufmann Publishers, Inc.

29. Si X, Wang W and Zhou D. Remaining useful life estimation—a review on the statistical data driven approaches. *Eur J Oper Res* 2011; 213: 1–14.

30. Pecht M and Jaai R. A prognostics and health management roadmap for information and electronics-rich systems. *Microelectron Reliab* 2010; 50: 317–323.

31. Willmott CJ. Some comments on the evaluation of model performance. *B Am Meteorol Soc* 1982; 63: 1309–1313.