

# Virtual sensing for gearbox condition monitoring based on kernel factor analysis

Jin-Jiang Wang<sup>1</sup> · Ying-Hao Zheng<sup>1</sup> · Lai-Bin Zhang<sup>1</sup> · Li-Xiang Duan<sup>1</sup> · Rui Zhao<sup>2</sup>

Handling editor: Jian Shuai

Received: 12 December 2016 / Published online: 22 May 2017  
© The Author(s) 2017. This article is an open access publication

**Abstract** Vibration and oil debris analysis are widely used in gearbox condition monitoring as the typical indirect and direct sensing techniques. However, they have their own advantages and disadvantages. To better utilize the sensing information and overcome its shortcomings, this paper presents a virtual sensing technique based on artificial intelligence by fusing low-cost online vibration measurements to derive a gearbox condition indicator, and its performance is comparable to the costly offline oil debris measurements. Firstly, the representative features are extracted from the noisy vibration measurements to characterize the gearbox degradation conditions. However, the extracted features of high dimensionality present nonlinearity and uncertainty in the machinery degradation process. A new nonlinear feature selection and fusion method, named kernel factor analysis, is proposed to mitigate the aforementioned challenge. Then the virtual sensing model is constructed by incorporating the fused vibration features and offline oil debris measurements based on support vector regression. The developed virtual sensing technique is experimentally evaluated in spiral bevel gear wear tests, and the results show that the developed kernel factor analysis method outperforms the state-of-the-art feature

selection techniques in terms of virtual sensing model accuracy.

**Keywords** Gearbox condition monitoring · Virtual sensing · Feature selection and fusion

## 1 Introduction

In the oil and gas industry, the gearbox is a very important element and its health and safety are critical to the smooth operation and efficiency of relevant facilities. However, gearboxes generally work under complex conditions, which may accelerate degradation and further induce different defects such as fatigue crack, pitting. These gearbox defects may even cause the breakdown of the whole system, leading to significant economic losses, costly downtime, and catastrophic damage. Therefore, gearbox condition monitoring is of great significance to its safe operation and maintenance schedule.

Increasing need for gearbox reliability has accelerated the integration of sensing techniques for condition monitoring. These sensing techniques could be roughly categorized into indirect sensing and direct sensing techniques, of which vibration and oil debris analysis are two typical sensing techniques, respectively. Various vibration analysis techniques have been widely investigated in gearbox fault diagnosis and prognosis. For instance, a fault characteristic order (FCO) analysis method was proposed to extract the vibration signal components related to rotational speed from the time–frequency representation (TFR) for gear fault detection under time-varying rotational speed (Wang et al. 2014). A modified cantilever beam model was investigated to analytically evaluate the time-varying mesh

---

✉ Jin-Jiang Wang  
jiw09005@foxmail.com; jwang@cup.edu.cn

✉ Lai-Bin Zhang  
zhanglb@cup.edu.cn

<sup>1</sup> School of Mechanical and Transportation Engineering, China University of Petroleum, Beijing 102249, China

<sup>2</sup> School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore

stiffness of a planetary gear set for the detection of crack severity and location via vibration analysis (Liang et al. 2014). The vibration signal properties of a planetary gearbox were investigated to differentiate healthy and cracked tooth conditions (Liang et al. 2015). A neuro-fuzzy approach was investigated by Samanta and Nataraj (2008) for modeling and prediction of gearbox dynamics utilizing various health indices. Defect prognostics were performed to estimate the temporal evolution of features (Wang 2007). The degradation condition prediction of a planetary gearbox was also investigated by Hussain and Gabbar (2013) based on acoustic phenomena along with neural networks and neuro-fuzzy approaches.

Vibration sensing is the most commonly used online monitoring technique because of its rugged cost-effective, but the sensing quantity is an indirect indicator of gearbox condition due to the low signal-to-noise ratio (SNR) of the sensing measurement. On the other hand, oil debris analysis, as a direct sensing method, provides an alternative solution by offline inspection of the gearbox condition. The oil debris measurements can directly indicate the machinery condition with high accuracy. An oil debris analysis was performed by Bolander et al. (2009) to estimate the spall size as the damage progressed in aircraft engine prognosis. It can be found that oil debris monitoring is suitable to provide an early indication and quantification of internal damage of a gearbox, but it is far from convenient or cost-effective because of high cost and human intervention requirements during normal operations of gearboxes.

To bridge the gap between direct sensing and indirect sensing, virtual sensing has emerged as a viable, noninvasive, and cost-effective method to infer difficult-to-measure or expensive-to-measure parameters based on computational models (Tham et al. 1991). It has been investigated for active noise and vibration control (Petersen et al. 2008), industrial process control (Cheng et al. 2004), building operation optimization (Ploennigs et al. 2011), lead-through robot programming (Ragaglia et al. 2016), product quality of hydrodesulfurization (HDS) (Shokri et al. 2015), and tool condition monitoring (Bustillo et al. 2011; Li and Tzeng 2000). Data-driven virtual sensing techniques are favorable by fusing the extracted features from noisy online measurements to infer the difficult-to-measure parameters based on artificial intelligence models (Gelman et al. 2013). A good feature representation method should be able to remove the irrelevant and redundant features, while preserving important (geometric or statistical) properties of the original data (Bolón et al. 2015). Thus, it is critical to devise a systematic feature selection and representation scheme to extract and select the most representative features. Moreover, it may reduce the computational complexity and storage, improve the

efficiency of virtual sensing models, and provide insight for knowledge discovery.

Different feature selection and fusion techniques have been developed including principal component analysis (PCA) and its kernel version (He et al. 2007; Schölkopf et al. 1998), factor analysis (FA) (Bishop 2006), dominant feature identification method (Zhou et al. 2011), minimum redundancy maximum relevance technique (Peng et al. 2005), and locally linear embedding (LLE) (Roweis and Saul 2000). In the above methods, PCA, FA, and kernel PCA are widely used in machine learning and data mining. Kernel PCA, as a nonlinear extension of PCA, was developed to explore the nonlinear relationship among variables by the use of a kernel function. Owing to high computational efficiency and nonlinear projection ability of using kernel functions, kernel methods are applied to extend the traditional linear model for feature selection and fusion. FA, as a linear model based on second-order statistics, generally has difficulty processing real-world data which has non-Gaussian distributions. The kernel methods provide the inspiration to develop a new feature selection and fusion method based on FA for feature fusion in gearbox condition monitoring.

To better utilize the sensing measurements for gearbox condition monitoring, this paper presents a virtual sensing technique based on artificial intelligence by fusing the low-cost online vibration measurements to infer the gearbox condition, and its performance can be comparable to the costly offline oil debris measurements. Firstly, the representative features are extracted from the noisy vibration measurements to characterize the gearbox degradation conditions. Next, a new nonlinear feature selection and fusion method, named kernel factor analysis (KFA), is proposed to reduce the feature dimensionality. Then the virtual sensing model is constructed by incorporating the fused vibration features and offline oil debris measurements based on support vector regression. The developed virtual sensing technique is experimentally evaluated in the spiral bevel gear wear tests, and the results show that the developed kernel factor analysis method outperforms the state-of-the-art feature selection techniques in terms of virtual sensing model accuracy.

The main contributions of this study rest on: (1) A virtual gearbox condition sensing framework is proposed to bridge the gap between vibration analysis and oil debris monitoring methods, and (2) a new feature selection and representation method (KFA) is presented to exploit the nonlinear representative features with non-Gaussian distributions, and the effectiveness of the KFA method is experimentally validated by the gear wear study. The rest of this paper is constructed as follows. After introducing the theoretical background of conventional feature representation techniques in Sect. 2, the details of the kernel

factor analysis-based virtual sensing model are then discussed in Sect. 3. The effectiveness of the presented technique is experimentally demonstrated in Sect. 4 based on direct and indirect sensing data acquired from a spiral bevel gear case study. Finally, conclusions are drawn in Sect. 5.

## 2 Theoretical framework

### 2.1 Principal component analysis and its kernel variant

Principal component analysis (PCA) has been widely investigated for dimensionality reduction of feature space. It transforms a set of observations of possible correlated variables into a set of uncorrelated variables called principal components, where the first principal component has the largest variance, and each succeeding principal component has comparative lower variance orthogonal to the preceding principal components. However, if the sample data has more complicated structures which cannot be well represented in a linear subspace, PCA may be not applicable. Kernel principal component analysis (KPCA) generalizes the traditional PCA to the nonlinear dimensionality reduction method by incorporating kernel techniques. The key idea of KPCA is to define a nonlinear transformation  $\phi(\bullet)$  which transforms the sample data into a high-dimensional data space, where each data point  $X_i$  is projected to a point  $\phi(X_i)$ . Then, the traditional PCA is performed in the new feature space (He et al. 2007). The first several principal components can well represent the original data with minimal mean squared approximation error, and thus KPCA has been widely used in the dimensionality reduction applications (He et al. 2007).

### 2.2 ISOMAP method

The ISOMAP algorithm extends the metric multidimensional scaling (MDS) method by integrating the geodesic distances instead of pairwise Euclidean distances to compute the graph shortest path distances (Tenenbaum et al. 2000). The key idea of ISOMAP is to find a low-dimensional embedding of data points, which is characterized as a nonlinear and global optimal method since only one free parameter (e.g.,  $\varepsilon$  or  $K$ ) needs to be optimized. The implementation of this algorithm mainly includes the following steps. Firstly, a neighborhood graph  $G$  is constructed by determining which points are neighbors on the manifold  $M$  based on the distances  $d_X(i, j)$  between pairs of points  $i, j$  in the input space  $X$ , where  $\varepsilon$ -ISOMAP and  $K$ -ISOMAP methods can be used to determine the neighborhood points. Secondly, the geodesic distances  $d_M(i, j)$

between all pairs of points on the manifold  $M$  are estimated by computing the shortest path distances  $d_G(i, j)$  in the graph  $G$ . Finally, the classical MDS is applied to the matrix  $D_G$ , constructing an embedding of the data in a  $d$ -dimensional space  $Y$  which best preserves the manifold's estimated intrinsic geometry. However, a typical shortcoming of ISOMAP method is the high computational complexity, characterized by the full matrix eigenvector decomposition (Tenenbaum et al. 2000).

### 2.3 Locally linear embedding algorithm

Locally linear embedding (LLE), as a representative manifold learning technique, is a nonlinear dimension reduction method by mapping the high-dimensional data to a lower dimensional space while preserving the essential properties of the raw data. It attempts to discover the underlying nonlinear structure (nonlinear manifold) in high-dimensional data by exploiting the local symmetries of linear reconstructions (Roweis and Saul 2000). Like the ISOMAP algorithm, the implementation of the LLE method also requires several steps. First of all, the neighbors of each data point  $x_i$  are obtained by calculating the Euclidean distances between neighbor points and the data point of interest. Next, the weights matrix  $W$  is computed by minimizing the reconstruction error of the data point from its neighbors. Finally, each high-dimensional observation  $X$  is mapped to a low-dimensional vector  $Y$  representing the global internal coordinates on the manifold. When implementing the LLE algorithm, only one free parameter needs to be optimized, which is quite straightforward. Therefore, once the number of neighbors per data point  $K$  is chosen, the optimal weights  $W_{ij}$  and coordinates  $Y_i$  are computed by standard methods in linear algebra. However, this method is sensitive to noise and prone to ill-conditioned eigen issues, which may lead to unsatisfactory performance of feature selection and fusion.

### 2.4 Factor analysis

Factor analysis (FA), as a typical variance-based feature selection and representation technique, is different from traditional PCA which is formulated on matrix decomposition. FA is a linear Gaussian latent variable model and releases constraint by forming a diagonal covariance (Bishop 2006). The constructed factor model represents the original variable by a linear combination of latent and measured variables. Thus FA is able to deal with the uncertainty of extracted features in gearbox condition monitoring, since the gearbox degradation process under complex operating conditions (such as accumulation of

fatigue, crack propagation, wear) may be subject to uncertainty and changeable operations. Moreover, FA is invariant to the component-wise rescaling of the feature space for input data by preserving the intrinsic data structures (Bishop 2006). Unfortunately, FA is a linear model based on the second-order statistics, which means that the processed data need to obey Gaussian distributions. However, the vibration signals in practice contain much noise and a variety of frequency components obeying non-Gaussian distributions. By taking into account the high computational efficiency and nonlinear projection ability of kernel functions, a kernel factor analysis is investigated for feature selection and representation in gearbox condition monitoring.

### 3 Proposed virtual sensing framework

During the normal operation process, online sensing techniques such as accelerometer and tachometer signals are continuously recorded to reflect gearbox conditions, but they are indirect indicators of gearbox conditions. On the other hand, oil debris is usually measured offline by experienced engineers to inspect the gearbox conditions, but it can directly reflect the gearbox condition. The proposed virtual sensing model for gearbox condition monitoring takes advantage of online measurements to estimate the gearbox conditions which are comparable to the oil debris measurements based on artificial intelligence as illustrated in Fig. 1. The virtual sensing framework mainly consists of four modules: (1) a data acquisition system capable of measuring vibration measurements during gearbox operation, (2) a feature extraction module to extract the representative gearbox condition indicators (CIs) by preprocessing the raw noisy measurements, (3) a kernel factor analysis-based feature fusion module to select

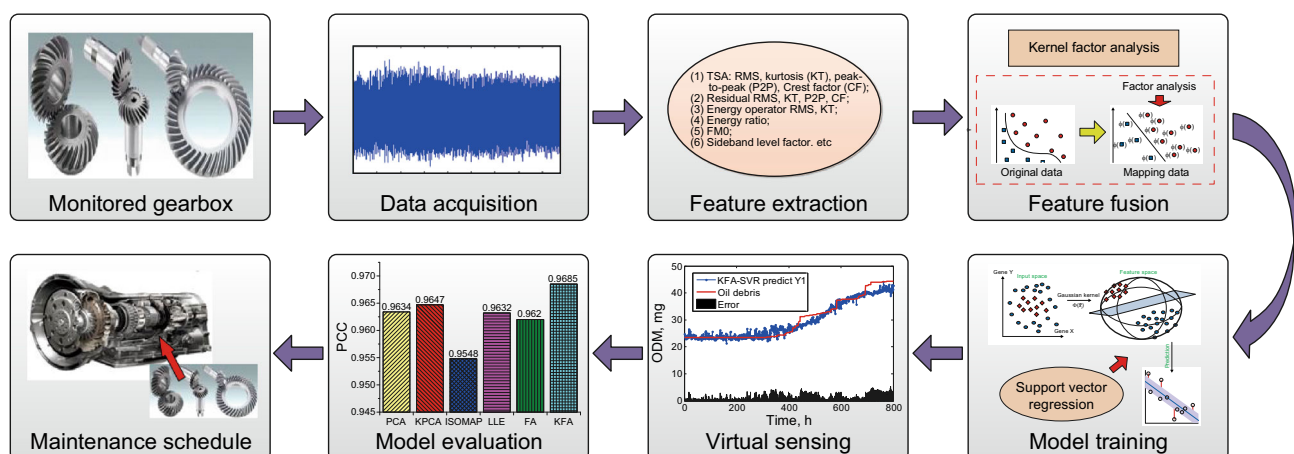
and fuse the extracted features for dimension reduction, and (4) a support vector regression-based artificial intelligence model to infer gearbox conditions from the fused features. The developed virtual sensing method is a complement to direct sensing or indirect sensing and provides a more effective tool for gearbox condition monitoring. The details of each module are discussed below.

#### 3.1 Data acquisition and feature extraction

The vibration signal measurements are usually collected continuously to characterize the gearbox condition. Due to the poor signal-to-noise ratio (SNR) and multi-component interaction in a gearbox, vibration signal processing is required to de-noise the signal and extract defective signatures. In this study, a total of 21 features or condition indicators (CIs) from time, frequency, and time–frequency domains are investigated including (1) time synchronous averages (TSA): root mean square (RMS), kurtosis (KT), peak-to-peak (P2P), crest factor (CF); (2) residual RMS, KT, P2P, CF; (3) energy operator RMS, KT; (4) energy ratio; (5) FM0; (6) sideband level factor; (7) narrowband (NB) RMS, KT, CF; (8) amplitude modulation (AM) RMS, KT; (9) derivative AM KT; and (10) frequency modulation (FM) RMS, KT. The detailed formulation of these condition indicators have been published (Zakrajsek et al. 1993; Wemhoff et al. 2007).

#### 3.2 Feature selection and fusion

The extracted features are formulated as feature vectors and further constructed as feature space of high dimensionality. To remove irrelevant and redundant features, and to improve model computational efficiency, a proper feature selection and fusion strategy is needed to lower the dimension of feature space. In the factor analysis, the



**Fig. 1** Diagram of the developed virtual sensing model

feature set  $X$  is defined as a linear combination of latent variable set  $Z$  plus a noise term as follows:

$$X = WZ + \mu + \varepsilon \tag{1}$$

where  $W$  is a  $D \times M$  factor loading matrix capturing the correlations behind the extracted feature variables;  $\mu \in \mathbb{R}^D$  is the mean vector for feature set  $X$ ; and  $\varepsilon$  denotes a  $D$ -dimensional Gaussian noise with zero mean and a diagonal covariance, i.e.,  $\varepsilon \sim \mathcal{N}(0, \Psi)$ , where  $\Psi$  is a  $D \times D$  diagonal matrix modeling the independent noise variance for each original dimension. For conciseness, the mean vector  $\mu$  is ignored in the following derivation, since the data are easily assumed to be zero-centered after preprocessing.

With the proper transformation, the original features could be well represented by a low-dimensional latent variable space. However, an underlying constraint in factor analysis is that the variables follow Gaussian distributions which are difficult to meet in real-world gearbox condition monitoring. Thus, a kernel version of the factor analysis method is formulated to tackle this issue. The original features are projected into a new feature space  $\mathcal{B}$  with a mapping function  $\phi$ , and the new feature matrix is generated and written by:

$$F = \begin{pmatrix} \Phi(x_1) \\ \Phi(x_2) \\ \vdots \\ \Phi(x_n) \end{pmatrix} \tag{2}$$

Then, the FA is introduced in the new feature space  $\mathcal{B}$ , which can be treated as performing a nonlinear FA in the original space. Similar to the FA method, the data in the new space  $\mathcal{B}$  can also be represented as follows:

$$F = WT + E \tag{3}$$

where  $W \in \mathbb{R}^{N \times M}$  that is the projected latent data matrix,  $T \in \mathbb{R}^{M \times P}$ , and  $E \sim \mathcal{N}_{N,P}(0, \Psi \otimes I_M)$  that is the noise variance matrix following the independent and identical distribution. To estimate model parameter set  $\theta$  containing  $\{W, \Psi\}$ , the expectation–maximization (EM) algorithm is used, which is an iterative method proposed for maximum likelihood of a latent probabilistic model (Dempster et al. 1977, Moon 1996). Considering that the new data matrix  $F$  is centered, the parameter estimation in  $E$ -steps and  $M$ -steps can be obtained as (Wang et al. 2016):

$$W_{q+1} = K_{\text{norm}} \Psi_q^{-1} W_q E [T]^T (I + G_q W_q^T \Psi_q^{-1} K_{\text{norm}} \Psi_q^{-1} W_q)^{-1} \tag{4}$$

$$\Psi_{q+1} = \frac{1}{N} \text{diag}\{K_{\text{norm}} - W_{q+1} G_q W_q^T \Psi_q^{-1} K_{\text{norm}}\} \tag{5}$$

$$G_q = (I + W_q^T \Psi_q^{-1} W_q)^{-1} \tag{6}$$

where  $q$  and  $q + 1$  represent two successive iteration steps. The kernel factor analysis only needs to address the kernel

matrix  $K$ , which is different from the traditional FA and has a more efficient learning process.

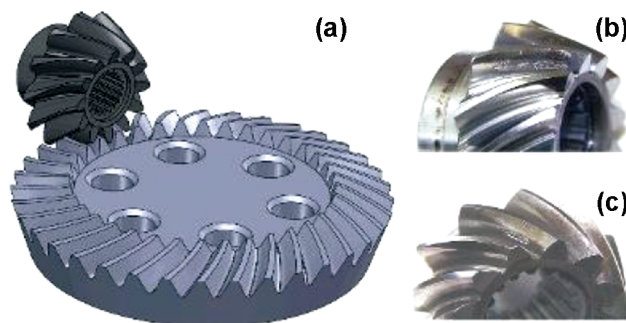
### 3.3 Virtual sensing model construction

The selected features are fed into the artificial intelligence model to construct the virtual sensing technique. Different artificial intelligence techniques could fit the purpose including artificial neural network (Dong et al. 2010), support vector regression (Widodo and Yang 2007), and fuzzy logic (Gokulachandran and Mohandas 2015). The artificial neural network technique has been widely investigated, but it requires a large amount of historical data for model training and suffers from local optima and overfitting issues. Support vector regression raises much attention because of high generalization capability and lower training sample requirements (Widodo and Yang 2007). Considering only a limited number of labeled experimental data sets are available, the support vector regression is selected to build the virtual sensing model in this study. During the model construction process, the selected features from vibration measurements are taken as the inputs while the oil debris measurements are treated as the outputs. The selection of parameters and kernel functions in the support vector regression model is determined using a grid search algorithm following a leave-one-out cross-validation method. Then the built support vector regression model fuses the selected features from vibration measurements to infer the gearbox condition indicator which is comparable with the oil debris measurements for gearbox condition monitoring.

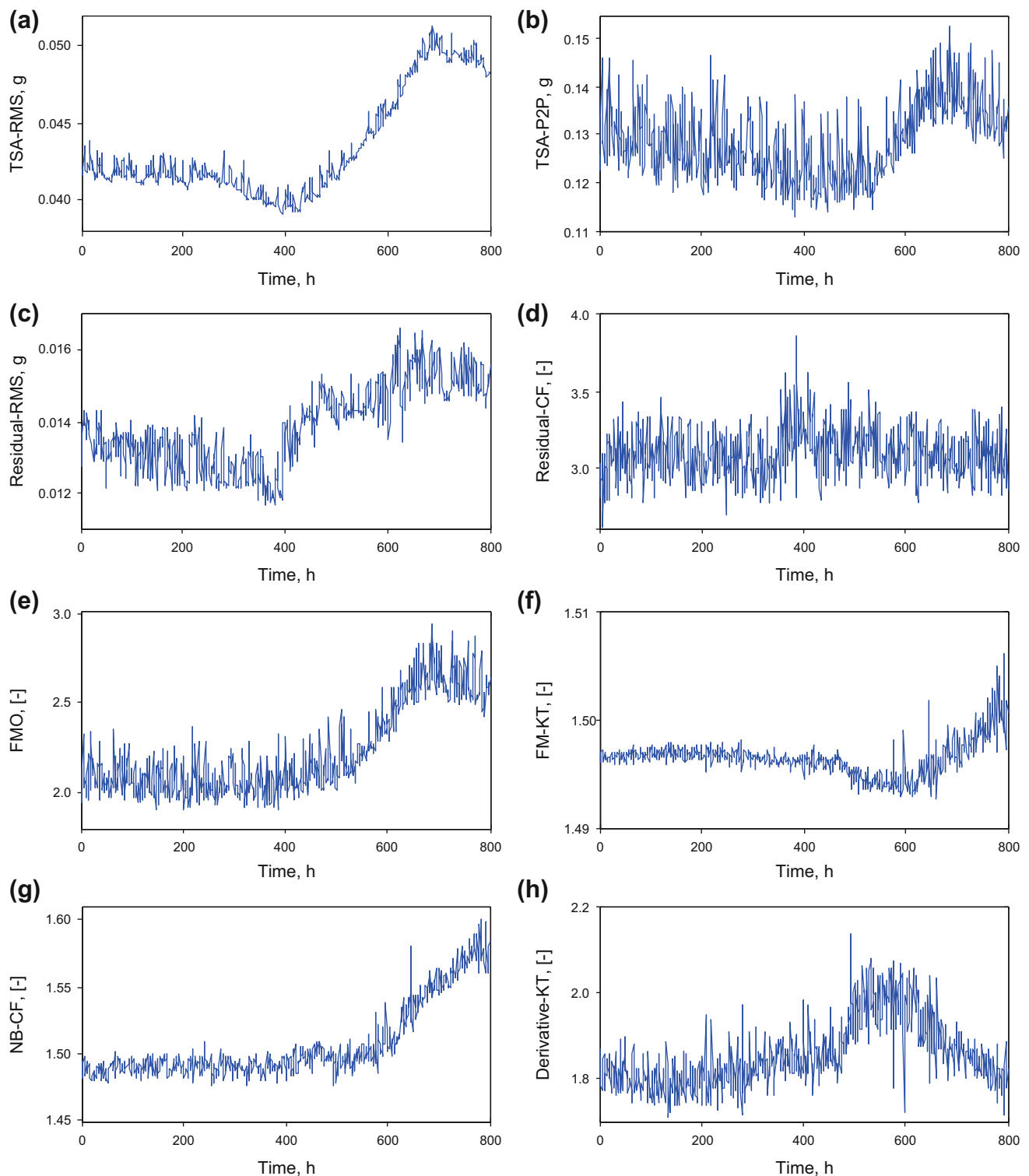
## 4 Experimental studies

### 4.1 Data preparation

Experimental data obtained from a spiral bevel gear case study (Dempsey et al. 2002) is used to evaluate the



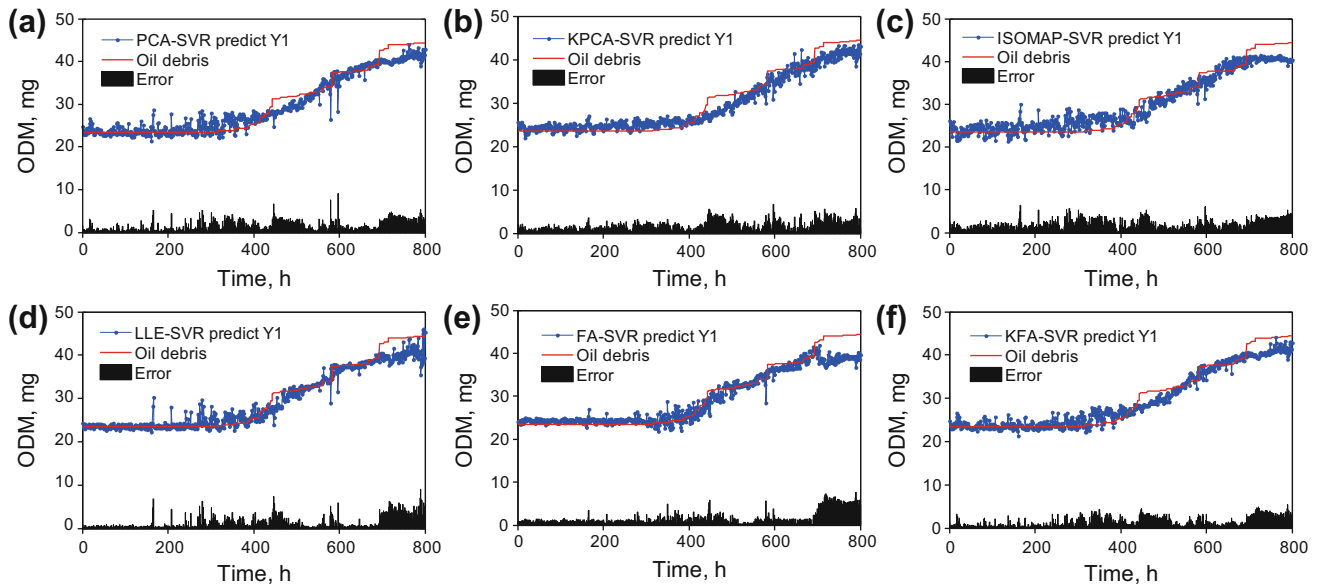
**Fig. 2** Illustration of gearbox test. **a** Bevel gear test rig. **b** Damaged spiral bevel gear in experiment Y1. **c** Damaged spiral bevel gear in experiment Y3 (Dempsey et al. 2002)



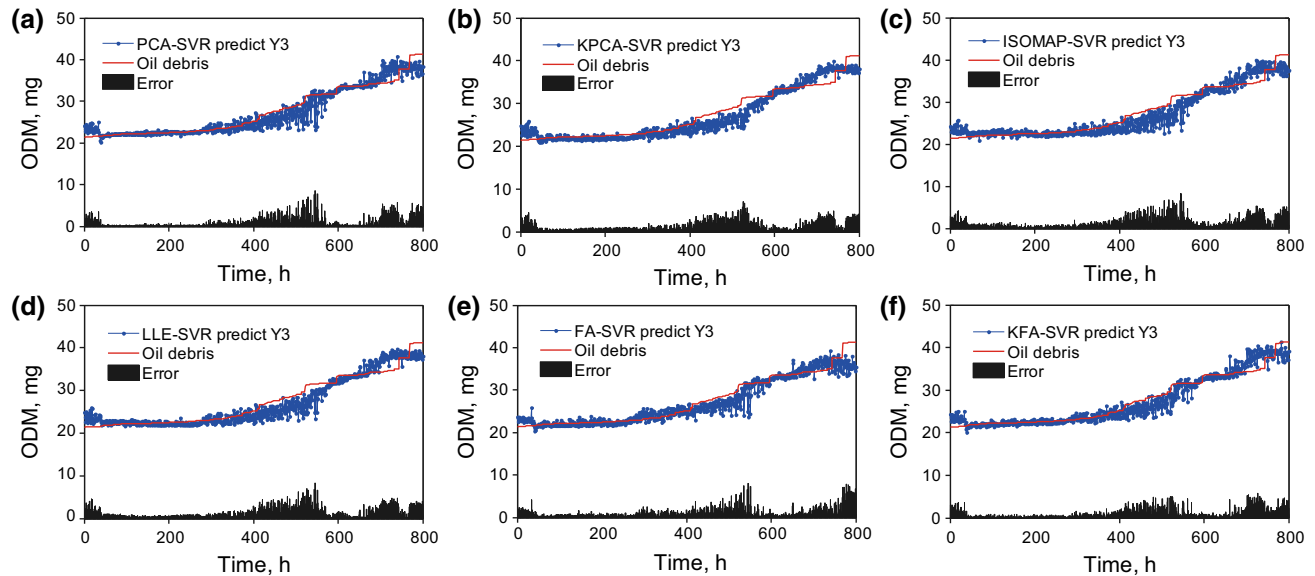
**Fig. 3** Exemplified feature sets extracted from vibration measurements

presented virtual sensing method. The schematic diagram of the bevel gear test rig is shown in Fig. 2. A number of gear wear tests were performed until surface fatigue occurred, during which the vibration and oil debris

measurements were collected to characterize the gearbox conditions. Vibration data was measured by two accelerometers located on the left and right pinion shaft bearing housing. They were collected once per minute



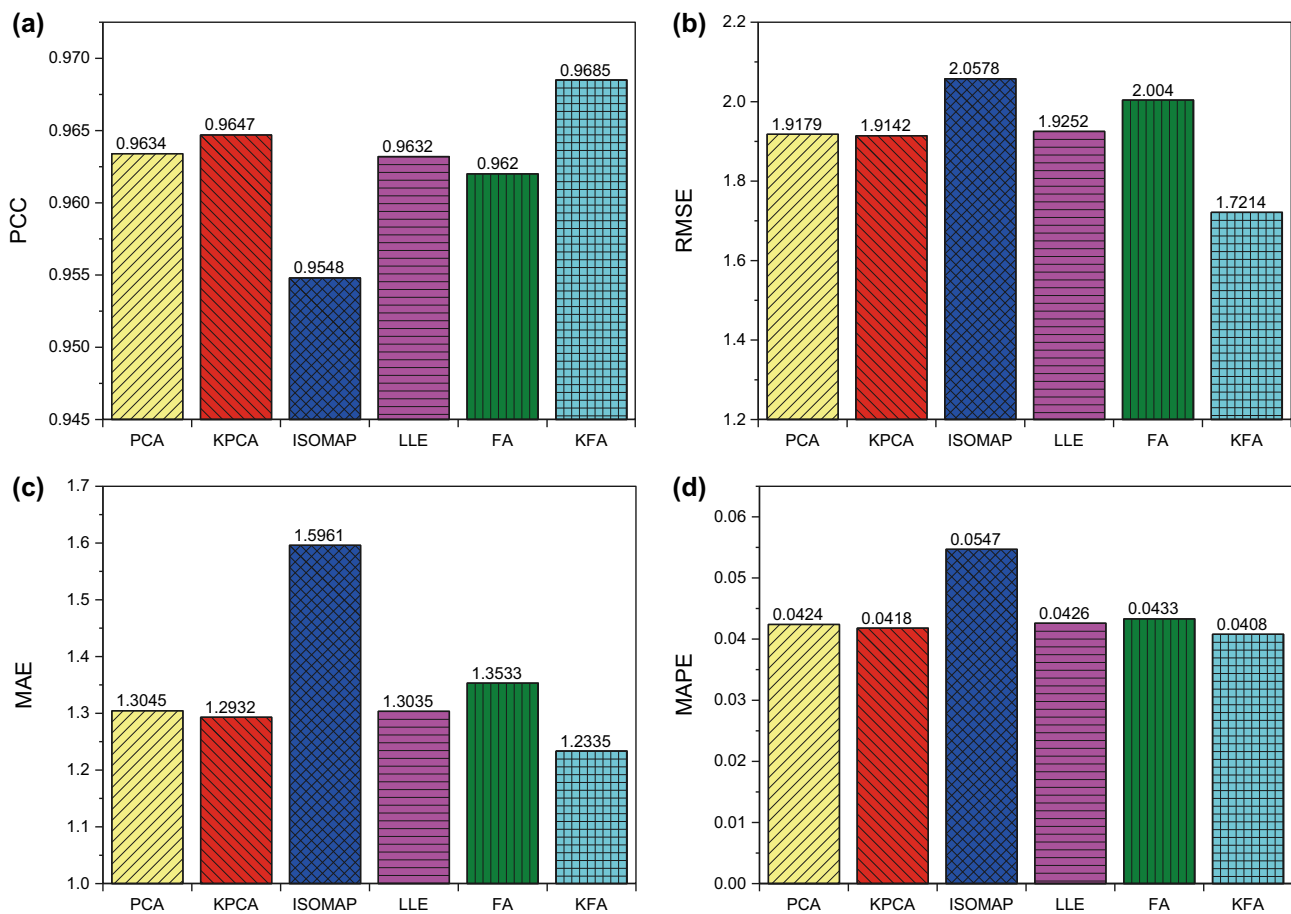
**Fig. 4** Performance comparison of different virtual sensing schemes using dataset Y1



**Fig. 5** Performance comparison of different virtual sensing schemes using dataset Y3

**Table 1** Quantitative evaluation criteria for performance comparison

Metrics	Mathematic formulation
Pearson correlation coefficient (PCC)	$PCC = \frac{\sum_i (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_i (y_i - \bar{y})^2 \sum_i (\hat{y}_i - \bar{\hat{y}})^2}}$
Root-mean-square error (RMSE)	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$
Mean absolute error (MAE)	$MAE = \frac{1}{N} \sum_{i=1}^N  y_i - \hat{y}_i $
Mean absolute percentage error (MAPE)	$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{ y_i - \hat{y}_i }{y_i}$



**Fig. 6** Performance comparison of different virtual sensing schemes using different criteria. **a** PCC. **b** RMSE. **c** MAE. **d** MAPE

**Table 2** Performance comparison of different virtual sensing schemes

Methods	PCC	RMSE	MAE	MAPE
PCA based	0.9634 ± 0.0089	1.9179 ± 0.0240	1.3045 ± 0.1370	0.0424 ± 0.0042
KPCA based	0.9647 ± 0.0121	1.9142 ± 0.0513	1.2932 ± 0.0318	0.0418 ± 0.0024
ISOMAP based	0.9548 ± 0.0116	2.0578 ± 0.0995	1.5961 ± 0.1806	0.0547 ± 0.0055
LLE based	0.9632 ± 0.0115	1.9252 ± 0.0095	1.3035 ± 0.0984	0.0426 ± 0.0054
FA based	0.962 ± 0.0103	2.004 ± 0.2077	1.3533 ± 0.1632	0.0433 ± 0.0028
KFA based	0.9685 ± 0.0081	1.7214 ± 0.0850	1.2335 ± 0.1246	0.0408 ± 0.0027

using a sampling rate of 100 kHz for 2 s duration. The shaft speed was measured by an optical sensor once per each gear shaft revolution, generating time synchronous averages (TSA). Oil debris data were collected using a commercially available oil debris sensor to detect the pitting damage on spiral bevel gears (Howe and Muir 1998).

A total of 21 representative features (as discussed in Sect. 3.1) are extracted from time, frequency and time–frequency domains by preprocessing the time synchronous averaging signal. The exemplified features are shown in Fig. 3. Whitening and eigenvalue decomposition (EVD) are firstly performed to select six dominant features by

preserving almost 95% of the cumulative variances. Next, kernel factor analysis is performed for dimension reduction to remove the irrelevant and redundant features.

## 4.2 Performance evaluation

The presented virtual sensing model is used to exploit the complex relationship between the vibration and oil debris analysis methods. A total of three sets of gearbox life test data (e.g., Y1, Y2, and Y3, etc.) are available. The leave-one-out strategy is followed to cross validate the performance of the virtual sensing model. More specifically, two



data sets are chosen for model training, and the remaining one is for model testing. Firstly, the SVR model is built by optimizing the cost parameter  $C$  and Gaussian kernel parameter  $\gamma$  using the grid search method to prevent overfitting. Next, the selected features obtained by KFA are fed into the constructed SVR model to infer gearbox conditions. To compare the performance of KFA, several state-of-the-art dimension reduction techniques are also investigated including PCA, KPCA, LLE, ISOMAP, and FA. The virtual sensing results of these different feature selection schemes are shown in Figs. 4 and 5 using different sets of experimental data. It is found that the predicted gear conditions by these virtual sensing models generally follow the trend of the actual oil debris measurement.

To quantitatively compare the performance of different virtual sensing schemes, different criteria are investigated including the Pearson correlation coefficient (PCC), root-mean-square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) as illustrated in Table 1. In the evaluation indexes,  $y$  represents the actual oil debris measurement and  $\hat{y}$  is the estimated oil debris measurement using the virtual sensing model. Generally, the larger the PCC value, the better the model performance, while the lower the RMSE/MAE/MAPE value, the better the model performance.

According to the above evaluation criteria, the performance of these virtual sensing schemes is compared and the results are shown in Fig. 6 and Table 2. It can be found that different feature selection techniques play important roles in the performance of virtual sensing models. The KFA-based virtual sensing model outperforms the conventional feature selection based virtual sensing models. By incorporating the kernel techniques into factor analysis, the superiority of KFA method is demonstrated to tackle the uncertainty and non-Gaussian features in the vibration measurements for feature selection and fusion.

## 5 Conclusions

Virtual sensing, as a complement to direct sensing or indirect sensing, provides a new perspective for machinery condition monitoring. According to the results obtained in this study, the conclusions can be drawn as follows.

- (1) A new virtual sensing technique is presented for gearbox condition monitoring by taking the merits of vibration analysis and oil debris analysis methods.
- (2) By incorporating the kernel technique into the factor analysis, a new feature selection method (KFA) is presented to exploit the nonlinear representative features with non-Gaussian distributions.

- (3) The effectiveness of the presented virtual sensing method is validated in the experimental studies of gear wear, and the comparison results show that the presented KFA scheme outperforms the conventional feature selection techniques in terms of virtual sensing model accuracy.

A variety of experimental tests will be performed to evaluate the robustness of the proposed method in our next-step research.

**Acknowledgements** This research acknowledges the financial support from the National Science Foundation of China (No. 51504274 and No. 51674277), the National Key Research and Development Program of China (No. 2016YFC0802103), and the Science Foundation of China University of Petroleum, Beijing (No. 2462014YJRC039 and 2462015YQ0403).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Bishop CM. Pattern recognition and machine learning. New York: Springer Science; 2006.
- Bolander N, Qiu H, Eklund N, et al. Physics-based remaining useful life prediction for aircraft engine bearing prognosis. In: Annual conference of the prognostics and health management society. 2009.
- Bolón-Canedo V, Sánchez-Marroño N, Alonso-Betanzos A. Recent advances and emerging challenges of feature selection in the context of big data. *Knowl Based Syst.* 2015;86:33–45. doi:10.1016/j.knsys.2015.05.014.
- Bustillo A, Correa M, Renones A. A virtual sensor for online fault detection of multitooth-tools. *Sensors.* 2011;11(3):2773–95. doi:10.3390/s110302773.
- Cheng JW-J, Chao TC, Chang LH, et al. A model-based virtual sensing approach for the injection molding process. *Polym Eng Sci.* 2004;44(9):1605–14. doi:10.1002/pen.20158.
- Dempsey PJ, Handschuh RF, Afjeh AA. Spiral bevel gear damage detection using decision fusion analysis. In: The 5th international conference on information fusion, Cleveland, OH, USA; July 8–11, 2002. doi:10.1109/icif.2002.1021136.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B Methodol.* 1977;39(1):1–38. doi:10.2307/2984875.
- Dong X, Wang S, Sun R, et al. Design of artificial neural networks using a genetic algorithm to predict saturates of vacuum gas oil. *Pet Sci.* 2010;7(1):118–22. doi:10.1007/s12182-010-0015-y.
- Gelman L, Petrunin I, Jennions IK, et al. Diagnostics of local tooth damage in gears by the wavelet technology. *Int J Progn Health Manag.* 2013;3:52–8.
- Gokulachandran J, Mohandas K. Comparative study of two soft computing techniques for the prediction of remaining useful life of cutting tools. *J Intell Manuf.* 2015;26(2):255–68. doi:10.1007/s10845-013-0778-2.

- He Q, Kong F, Yan R. Subspace-based gearbox condition monitoring by kernel principal component analysis. *Mech Syst Signal Process.* 2007;21(4):1755–72. doi:[10.1016/j.ymsp.2006.07.014](https://doi.org/10.1016/j.ymsp.2006.07.014).
- Howe B, Muir D. In-line oil debris monitor (ODM) for helicopter gearbox condition assessment. In: Defense Technical Information Center, VA; 1998.
- Hussain S, Gabbar HA. Vibration analysis and time series prediction for wind turbine gearbox prognostics. *IJPHM Special Issue on Wind Turbine PHM (Color)*. 2013:69.
- Li CJ, Tzeng TC. Multimilling-insert wear assessment using nonlinear virtual sensor, time-frequency distribution and neural networks. *Mech Syst Signal Process.* 2000;14(6):945–57. doi:[10.1006/mssp.1999.1282](https://doi.org/10.1006/mssp.1999.1282).
- Liang X, Zuo MJ, Hoseini MR. Vibration signal modeling of a planetary gear set for tooth crack detection. *Eng Fail Anal.* 2015;48:185–200. doi:[10.1016/j.engfailanal.2014.11.015](https://doi.org/10.1016/j.engfailanal.2014.11.015).
- Liang X, Zuo MJ, Pandey M. Analytically evaluating the influence of crack on the mesh stiffness of a planetary gear set. *Mech Mach Theory.* 2014;76:20–38. doi:[10.1016/j.mechmachtheory.2014.02.001](https://doi.org/10.1016/j.mechmachtheory.2014.02.001).
- Moon TK. The expectation-maximization algorithm. *IEEE Signal Process Mag.* 1996;13(6):47–60.
- Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell.* 2005;27(8):1226–38. doi:[10.1109/tpami.2005.159](https://doi.org/10.1109/tpami.2005.159).
- Petersen CD, Fraanje R, Cazzolato BS, et al. A Kalman filter approach to virtual sensing for active noise control. *Mech Syst Signal Process.* 2008;22(2):490–508. doi:[10.1016/j.ymsp.2007.06.007](https://doi.org/10.1016/j.ymsp.2007.06.007).
- Ploennigs J, Ahmed A, Hensel B, et al. Virtual sensors for estimation of energy consumption and thermal comfort in buildings with underfloor heating. *Adv Eng Inform.* 2011;25(4):688–98. doi:[10.1016/j.aei.2011.07.004](https://doi.org/10.1016/j.aei.2011.07.004).
- Ragaglia M, Zanchettin AM, Bascetta L, et al. Accurate sensorless lead-through programming for lightweight robots in structured environments. *Robot Comput Integr Manuf.* 2016;39:9–21. doi:[10.1016/j.rcim.2015.11.002](https://doi.org/10.1016/j.rcim.2015.11.002).
- Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. *Science.* 2000;290(5500):2323–6. doi:[10.1126/science.290.5500.2323](https://doi.org/10.1126/science.290.5500.2323).
- Samanta B, Nataraj C. Prognostics of machine condition using soft computing. *Robot Comput Integr Manuf.* 2008;24(6):816–23. doi:[10.1016/j.rcim.2008.03.011](https://doi.org/10.1016/j.rcim.2008.03.011).
- Schölkopf B, Smola A, Müller KR. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* 1998;10(5):1299–319. doi:[10.1162/089976698300017467](https://doi.org/10.1162/089976698300017467).
- Shokri S, Sadeghi MT, Marvast MA, et al. Improvement of the prediction performance of a soft sensor model based on support vector regression for production of ultra-low sulfur diesel. *Pet Sci.* 2015;12(1):177–88. doi:[10.1007/s12182-014-0010-9](https://doi.org/10.1007/s12182-014-0010-9).
- Tenenbaum JB, De Silva V, Langford JC. A global geometric framework for nonlinear dimensionality reduction. *Science.* 2000;290(5500):2319–23. doi:[10.1126/science.290.5500.2319](https://doi.org/10.1126/science.290.5500.2319).
- Tham MT, Montague GA, Morris AJ, et al. Soft-sensors for process estimation and inferential control. *J Process Control.* 1991;1(1):3–14. doi:[10.1016/0959-1524\(91\)87002-F](https://doi.org/10.1016/0959-1524(91)87002-F).
- Wang J, Xie J, Zhao R, et al. A new probabilistic kernel factor analysis for multisensory data fusion: application to tool condition monitoring. *IEEE Trans Instrum Meas.* 2016;65(11):2527. doi:[10.1109/tim.2016.2584238](https://doi.org/10.1109/tim.2016.2584238).
- Wang T, Liang M, Li J, et al. Rolling element bearing fault diagnosis via fault characteristic order (FCO) analysis. *Mech Syst Signal Process.* 2014;45(1):139–53. doi:[10.1016/j.ymsp.2013.11.011](https://doi.org/10.1016/j.ymsp.2013.11.011).
- Wang W. An adaptive predictor for dynamic system forecasting. *Mech Syst Signal Process.* 2007;21(2):809–23. doi:[10.1016/j.ymsp.2005.12.008](https://doi.org/10.1016/j.ymsp.2005.12.008).
- Wemhoff E, Chin H, Begin M. Gearbox diagnostics development using dynamic modeling. In: AHS 63rd annual forum, Virginia Beach, VA; 2007.
- Widodo A, Yang BS. Support vector machine in machine condition monitoring and fault diagnosis. *Mech Syst Signal Process.* 2007;21(6):2560–74. doi:[10.1016/j.ymsp.2006.12.007](https://doi.org/10.1016/j.ymsp.2006.12.007).
- Zakrajsek JJ, Townsend DP, Decker HJ. An analysis of gear fault detection methods as applied to pitting fatigue failure data. In: 47th Mechanical Failure Prevention Group Virginia Beach, Virginia; April 13–15, 1993.
- Zhou JH, Pang CK, Zhong ZW, et al. Tool wear monitoring using acoustic emissions by dominant-feature identification. *IEEE Trans Instrum Meas.* 2011;60(2):547–59. doi:[10.1109/tim.2010.2050974](https://doi.org/10.1109/tim.2010.2050974).