A Hybrid Framework Combining Data-level Fusion and Model-based Models for Remaining Useful Life Prediction

Xiaochuan Li
Faculty of Computing, Engineering and Media
De Montfort University
Leicester, UK
xiaochuan.li@dmu.ac.uk

David Mba
[1]Faculty of Computing, Engineering and Media
De Montfort University
Leicester, UK
[2]Department of Mechanical Engineering
University of Nigeria
Nsukka, Nigeria

Tianran Lin
School of Mechanical and Automotive Engineering
Qingdao University of Technology
Qingdao, China

***Abstract*—In this work, a hybrid prognostic framework is put forward to predict remaining useful life (RUL) through the fusion of multiple degradation-based sensor data. This approach is tested on measurements obtained from an industrial centrifugal pump. The prognostic framework consists of determination of predication start point, sensor selection and fusion, and prognostics steps that lead to accurate RUL prediction. This approach first applies the canonical variate analysis (CVA) method to the pump for determining the prediction start time. We also present an approach for constructing a single-valued prognostic health indicator through the fusion of fault root-case variables. Moreover, the particle filter (PF) algorithm is deployed to improve the metabolism grey forecasting model (MGFM) to realize the prediction of both RUL and uncertainties.**

Keywords-canonical variate analysis; grey model; particle filter; Pearson correlation analysis

I. INTRODUCTION

Unexpected downtime caused by machinery failures is costly and can incur large economic losses and security threats. In light of this industrial challenge, numerous predictive maintenance programs/approaches have been put forward in recent years to predict equipment failures and to ensure maintenance is carried out only when necessary. However, it is difficult and costly to carry out RUL prediction when equipment is under normal conditions since little information about the degradation trend can be found during this stage. To make a prognostic framework suitable for online monitoring, it is essential to include a module which can automatically determine prediction start time such that the RUL prediction is implemented only after certain failures are detected. Widely used data-driven process monitoring tools, known in general as multivariate statistical process monitoring (MSPM) techniques, have been widely used for detecting process abnormalities. Principal component analysis (PCA) [1], independent component analysis (ICA) [2] and their improved versions [3]–[6] are representative methods in early studies for industrial process monitoring. CVA [7] is also a MSPM technique, and it was shown to be superior to other dimension reduction techniques for monitoring dynamic process under varying operational conditions [5]. In this study, a CVA based HI is adopted together with kernel density estimation (KDE) to automatically determine the prediction start time. Additionally, CVA is employed in this work to generate degradation features based on which system failure time can be estimated.

From a prognostic perspective, variables that show little correlation with the fault evolution can be eliminated for RUL prediction [8]. To this end, we propose an EWMA-Pearson method to differentiate faulty variables from normal variables in that the contributions from each process variable to the fault are quantified. Those variables with small contributions are eliminated and variables with large contributions are retained. EWMA is a popular smoothing technique for time series analysis. It has been proven to be effective in detecting small shifts in data [9] and hence has aroused attention for detecting incipient faults [10], [11]. Little research related to the use of EWMA for fault identification has been conducted. Therefore, we employ EWMA in this work to enhance the performance of the correlation analysis for the purpose of improving the fault identification ability of the latter.

Numerous data-driven prognostic approaches and their applications to the engineering field can be found in the

978-1-7281-0860-5/19/$31.00 ©2019 IEEE

literature. These are support vector regression, feed-forward neural network and deep learning algorithms, to name a few. However, these methods normally require large amounts of failure data to ensure precise predictions. Failure data are rather scarce in practice especially for large scale rotating machines. Therefore, we deploy a predictor, MGFM [12], which is particularly suitable for small sample and scarce failure data problems to realize RUL predictions. The inherent predictive errors would accumulate over time when it comes to long-term prediction. As a result, large gain of uncertainties is inevitable and should be estimated for characterizing the precision of the prediction. To this end, we use PF together with MGFM to quantify both the RUL and its uncertainties.

II. METHODOLOGY

A. CVA for the determination of prediction start time

Fault detection in real rotating machinery using CVA was introduced in [7]. Although fault detection is not the main purpose of this study, it provides the starting point of RUL prediction in the sense that the prognostic analysis can be triggered once anomalies have been detected. A brief introduction of CVA-based diagnostic method is given in the following subsections.

Given two data sets $y_{1,t} \in \mathcal{R}^n$ and $y_{2,t} \in \mathcal{R}^n$, CVA aims to find the projections $z_{1,t} = K \cdot y_{1,t}$ and $z_{2,t} = G \cdot y_{2,t}$ such that the correlation between $z_{1,t}$ and $z_{2,t}$ is maximized. $z_{1,t}$ and $z_{2,t}$ are also known as canonical variates (CVs). The past and future column vectors $y_{a,t} \in \mathcal{R}^{na}$ and $y_{b,t} \in \mathcal{R}^{nb}$ can be formed as:

$$y_{a,t} = \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-a} \end{bmatrix} \in \mathcal{R}^{na} \tag{1}$$

$$y_{b,t} = \begin{bmatrix} y_{t} \\ y_{t+1} \\ \vdots \\ y_{t+b-1} \end{bmatrix} \in \mathcal{R}^{nb} \tag{2}$$

where a and b are the size of the past and future windows of data, respectively. Then the past and future Hankel matrices $\hat{\mathrm{Y}}_a$ and $\hat{\mathrm{Y}}_b$ can be formed as:

$$\hat{Y}_a = [\,\hat{y}_{a,t+1}, \hat{y}_{a,t+2}, \dots, \hat{y}_{a,t+N}] \in \mathcal{R}^{na \times N} \tag{3}$$

$$\hat{Y}_b = [\hat{y}_{b,t+1}, \hat{y}_{b,t+2}, \dots, \hat{y}_{b,t+N}] \in \mathcal{R}^{nb \times N} \tag{4}$$

where $\hat{y}_{a,t}$ and $\hat{y}_{b,t}$ are the normalized past and future vectors, respectively. $N = M - a - b + 1$ represents the number of $y_{a,t}$ contained in the Hankel matrices $\hat{\mathrm{Y}}_a$. The sample covariance matrices $\sum_{a,a}$, $\sum_{b,b}$ and $\sum_{a,b}$ can be obtained as

$$\sum_{a,a} = \hat{Y}_a \hat{Y}_a{}^T/(N-1)\ ; \quad \sum_{b,b} = \hat{Y}_b \hat{Y}_b{}^T/(N-1)\ ; \quad \sum_{b,a} = \hat{Y}_b \hat{Y}_a{}^T/(N-1) \tag{5}$$

CVA aims to maximize the correlation between the linear combinations of the projections $z_{1,t} = K \cdot y_{1,t}$ and $z_{2,t} = G \cdot y_{2,t}$. The algebraic solution is obtained by performing singular value decomposition (SVD) on the scaled matrix $\mathcal{H}$:

$$\mathcal{H} = \sum_{b,b}{}^{-1/2} \sum_{b,a} \sum_{a,a}{}^{-1/2} = U \sum V^T \tag{6}$$

The canonical residuals are given by the difference between the first q projections in the past and the future space:

$$r_t = L_q^T \hat{y}_{b,t} - \sum_q J_q^T \hat{y}_{a,t} \tag{7}$$

where $L_q^T \hat{y}_{b,t}$ is the first q future projections, and $L^T = \sum_{b,b}^{-1/2} U_q^T$. Similarly, $J_q^T \hat{y}_{a,t}$ is the first q past projections, and $J^T = \sum_{a,a}^{-1/2} V_q^T$. $\sum_q = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_q)$ is the diagonal matrix of the descending first q singular values.

In this paper, the determination of prediction start time is carried out by comparing a CVA distinction-based index T_c with a statistically determined fault threshold.

$$T_c = \frac{T^2}{\sigma^{T^2}} + \frac{Q}{\sigma^Q} + \frac{T_d}{\sigma^{T_d}} \tag{8}$$

$$T^2 = z_t^T z_t, z_t = V_q^T \Sigma_{a,a}^{-1/2} \hat{y}_{a,t} \tag{9}$$

$$Q = e_t^T e_t, e_t = V_{na-q}^T \Sigma_{a,a}^{-1/2} \hat{y}_{a,t} \tag{10}$$

$$T_d = f\big(c(r_t - 0)^T \mathrm{S}^{-1}(r_t - 0)\big) = \frac{|c(r_t-0)^T \mathrm{S}^{-1}(r_t-0)|}{|c|[(r_t-0)^T \mathrm{S}^{-1} S \mathrm{S}^{-1}(r_t-0)]} = [(r_t)^T \mathrm{S}^{-1}(r_t)]^{1/2} = [r_t^T (I - \sum\sum^T) r_t]^{1/2} \tag{11}$$

where σ^{T^2}, σ^Q and σ^{T_d} are the threshold of Hotelling's T^2 and Q statistics [12], and, T_d index, respectively. c demotes a normalizing constant, and $\mathrm{S} = I - \sum\sum^T$ denotes the covariance matrix of the training data. σ^{T^2}, σ^Q and σ^{T_d} are calculated using the kernel density estimation method as proposed in [13].

B. Enhancement of Pearson correlation analysis using EWMA

Pearson correlation analysis. Pearson correlation analysis is utilized to calculate a numerical value (i.e. Pearson correlation coefficient (PCC)) in expression degrees of how well two variables correlates. PCC is one of the most commonly used statistical metrics in statistics that measures the direction and strength of a linear relationship between two random variables [17], [18].

Given two sets of zero-mean random variables x and y, the PCC is defined as [19]:

$$\rho(x, y) = \frac{E[x,y]}{\sigma(x)\sigma(y)} \tag{12}$$

where $E[x, y]$ is the cross-correlation between the random variables x and y; $\sigma(x)$ and $\sigma(y)$ denote the standard deviation of x and y, respectively. Standard PCCs were determined for each variable versus the fault detection index using data collected from early stages of degradation. The PCCs give an indication of the amount of contributions from different variables during the early degradation. The higher the PCC of a performance variable, the larger the contribution of the specific variable to the detected fault.

EWMA-Pearson. EWMA smoothing technique is sensitive to small shifts in data [9], [10], which allows fault incurred variations to present more consistently in the monitoring index. Therefore, EWMA is applied the contributions calculated from

the data collected from early degradation states to enhance the fault identification performance of Pearson correlation analysis. Compare to Pearson correlation analysis [19], EWMA-Pearson improves the fault identifiability of Pearson correlation analysis by enhancing the contributions from influential variables. The contribution of a variable based on the EWMA-enhanced approach can be obtained as:

$$c_t = (1-\delta)c_{de,t} + \delta c_{t-1} \quad (13)$$

$$c_{t-1} = \frac{\sum_{k=t-W}^{t-1} c_{de,t}}{W} \quad (14)$$

where δ is the forgetting factor and W is the width of the moving window. The EWMA-enhanced contributions can provide information regarding the most strongly affected variables when a fault occurs. The root-cause variables are spotted if they have the largest contributions during the early stages of degradation. The identified faulty variables are used subsequently to build a prognostic feature using the aforementioned CVA method. With this process, a new prognostic feature based only on faulty variables is constructed and then fed into a MGFM-PF prognostic model.

C. Enhancement of MGFM using PF

The main idea of using PF for prognostics is to use a state transition model to predict the future state of a system:

$$x_k = g(x_{k-1}, x_{k-2}, x_{k-3}, x_{k-4}) + u_k \quad (15)$$

where g represents the state transition function, x_k denotes the system state at time k, and w_k is a noise term. u_k can be determined by the state transition model's modelling errors during the state tracking phase. In this work, we use the exponential regression model to describe the state evolution as formulated in (15).

Mathematically speaking, during the state tracking phase, PF uses a weighted set of random particles to approximate the posterior marginal density of the system state:

$$p(x_k|y_{1:k}) \approx \sum_{j=1}^{N} \pi_k^j \delta(x_k - x_k^j) \quad (16)$$

where N is the number of particles, and $\sum_{j=1}^{N} \pi_k^j = 1$. When new measurements become available, the weights π_k^j are updated in line with the principle of importance sampling:

$$\pi_k^j \propto \pi_{k-1}^j p(y_k|x_k^j) \quad (17)$$

According to [14], [15], a canonical representation of the measurement model can be used when the system state x_k is not directly measurable:

$$y_k = x_k + v_k \quad (18)$$

where both y_k and x_k represent the system state (in this case the HI obtained via CVA), and v_k is a noise term. After the state tracking phase, state prediction is implemented by projecting the current particle population based on the last estimated particles and the associated weights along all possible pathways.

In summary, Fig. 1 illustrates a flowchart of the proposed prognostic method. Once the prediction start time is determined by the CVA method, the fault identification and prognostic modules will be triggered. The EWMA-Pearson model identifies faulty variables after the prediction starts, which will help in the construction of the prognostic feature. The proposed MGFM-PF method is then utilized to predict at future times the value of the prognostic feature until it reaches the failure threshold. The RUL and its associated uncertainty are computed based on the propagated particle population.

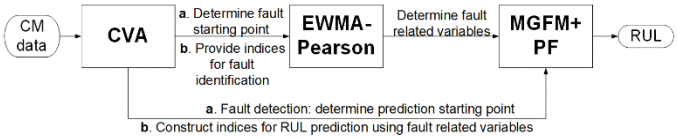


Figure 1. Flowchart of the proposed method

III. INDUSTRIAL APPLICATION

A. Data collection

In this subsection, condition monitoring (CM) data obtained from an industrial centrifugal pump were captured to test the performance of the proposed approach. The pump serves at a European refinery. The first data set consists of 197 samples and the second contains 380 samples. The names of different sensors are listed in Table I.

The data were recorded at a sampling rate of one sample per hour. Figure 2 (a) shows that for the first data set, the pump was at normal condition during the first 138 hours of operation, and then symptoms of fault degradation were observed. The fault developed in magnitude until the pump was forced to shut down. Similar behaviors can be observed in Figure 2 (b) which demonstrates the performance of the pump corresponding to data set 2.

TABLE I. MEASURED VARIABLES

Variable ID	Variable Name	Units
1	Speed	rpm
2	Suction pressure	bar
3	Discharge pressure	bar
4	Discharge temperature	degree C
5	Actual flow	kg/h
6	Radial vibration overall X 1	mm/s
7	Radial vibration overall Y 1	mm/s
8	Radial bearing temperature 1	degree C
9	Radial vibration overall X 2	mm/s
10	Radial vibration overall Y 2	mm/s
11	Radial bearing temperature 2	degree C
12	Active thrust bearing temperature	degree C
13	Inactive thrust bearing temperature	degree C

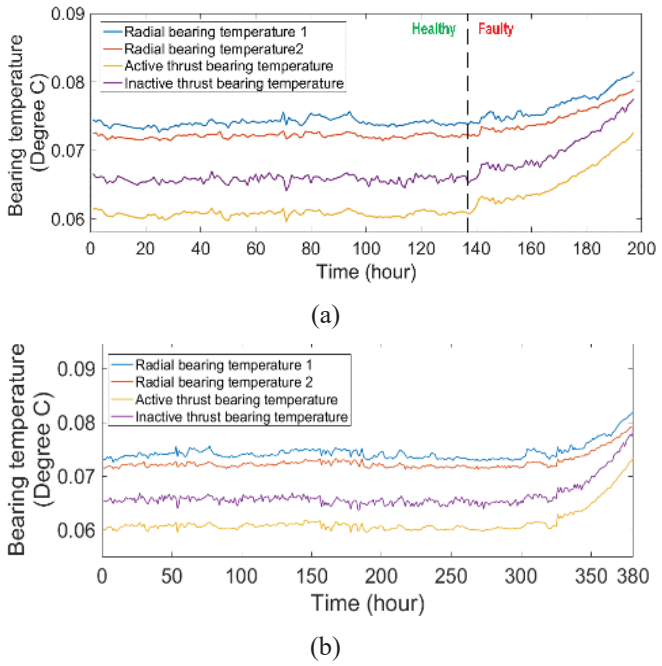


Figure 2. Trend of four different bearing temperature sensor measurements of pump cases 1 and 2 (signals are normalized). (a) fault case 1; (b) fault case 2.

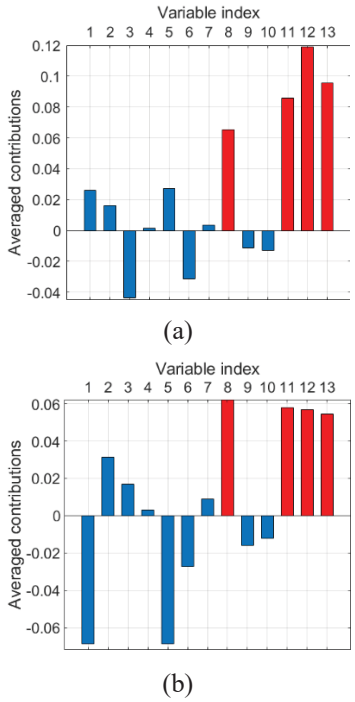


Figure 3. Identification of root-cause variables for pump A. (a) fault case 1; (b) fault case 2.

B. *Prediction start time and root-cause variables*

CVA is used to extract the main performance features that represent both health state and degradation trend of the engines. CVA can also reduce the data's dimensionality. A training dataset obtained from different machines at normal operation conditions are used to train the CVA model. By doing this, the various operating conditions of the fleet of are fully accommodated. The control limit for healthy operational conditions is determined using the Kernel density estimation method [26] with a 99% confidence level. The trained CVA model is then employed to calculate a HI for each of the testing engine. The prediction start times for data set 1 and 2 were t=138 and t=353, respectively.

The proposed EWMA-Pearson was employed to find faulty variables once the prediction start time is determined. To achieve this goal, PCCs were calculated for each process variable with regard to the fault indicator. Then the EWMA model was applied to enhance the contributions from faulty variables. The forgetting factor δ and the moving window width W were set to 0.9 and 10, respectively. We take the averaged contributions over a period of twenty-five hours (starting from the prediction start point) as the final contributions from each variable. The fault identification results are shown in figure 3. It can be observed that variables no. 8, 11, 12 and 13 are successfully identified as faulty variables in both fault cases, which is in line with the time series observations as demonstrated in figure 2.

C. *RUL prediction*

The RUL was predicted by propagating in the PF framework the trend of the prognostic feature along all possible pathways. The prognostic feature was obtained by fusing only fault root-cause variables through CVA method. Similar to [21], a locally weighted regression filter (LOESS) was employed to eliminate the noisy components. The predictive results are demonstrated in figure 4 for different prediction starting points. The areas between the dashed black curves denote the one standard deviation tolerance interval, which was calculated assuming the particles follow a normal distribution.

It is observable from the figure that the RUL tends to fluctuate around the true value at the beginning and gradually approaches the true value as the prediction starting point moves closer to the failure point. Meanwhile, the associated uncertainty boundaries tend to become narrower as the prediction starting point moves closer to the failure point as well.

In order to compare our method with existing prognostic techniques, 4 famous methods namely MGFM [22], LSTM [23], and AR model [24] were selected to perform the prognostic task using the same data and the comparison results are shown in Table II. Prognostic performance metrics are provided, and three evaluation criteria were employed in this work to quantitatively benchmark the predictive accuracy of the methods compared:

Root mean square deviation (RMSD):

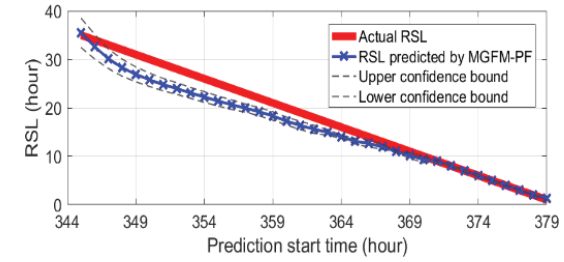
$$RMSD = \sqrt{\sum_{i=1}^{N}(RUL_{pre,i} - RUL_{true,i})^2/N} \qquad (20)$$

Mean absolute deviation (MAD):

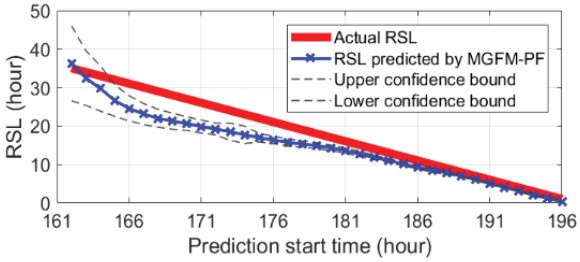
$$MAD = \sum_{i=1}^{N} |RUL_{pre,i} - RUL_{true,i}|/N \quad (21)$$

Cumulative relative accuracy (CRA):

$$CRA = \sum_{i=1}^{N} \left(\frac{|RUL_{pre,i} - RUL_{true,i}|}{RUL_{true,i}}\right)/N \quad (23)$$



(a)



(b)

Figure 4. Predictive results of the pump obtained using the MGFM-PF method. (a) fault case 1; (b) fault case 2.

TABLE II. PREDICTIVE PERFORMANCE

	Metrics	MGFM-PF	LSTM	AR	MGFM	MGFM-PF (all variables)
Fault 1	RMSD	2.44	20.82	4	3.69	3.81
	MAD	1.92	17.49	3.57	2.97	2.86
	CRA	0.0051	0.046	0.0094	0.0078	0.0075
Fault 2	RMSD	3.14	20.64	3.9	8.18	7.75
	MAD	2.35	17.87	3.2	4.2	5.81
	CRA	0.012	0.095	0.017	0.022	0.031

IV. CONCLUSION

In this paper, a fault detection and prediction method is presented under an integrated framework. Using two industrial pump case studies, the proposed approach was shown to accurately identify faulty variables and predict RUL with small errors. Our experimental results show RMSD values of 2.44 and 3.14 and MAD values of 1.92 and 2.35 between the actual RUL and that predicted using the MGFM-PF method. The proposed approach was shown to be superior to MGFM, LSTM and AR models. Through experiments, the necessity of using only faulty variables for prognostics was also demonstrated.

REFERENCES

[1] S. Wook, C. Lee, J. Lee, J. Hyun, and I. Lee, "Fault detection and identification of nonlinear processes based on kernel PCA," Chemom. Intell. Lab. Syst., vol. 75, pp. 55–67, 2005.

[2] J. Fan and Y. Wang, "Fault detection and diagnosis of non-linear non-Gaussian dynamic processes using kernel dynamic independent component analysis," Inf. Sci. (Ny)., vol. 259, pp. 369–379, 2014.

[3] W. Li and S. J. Qin, "Consistent dynamic PCA based on errors-in-variables subspace identification," J. Process Control, vol. 11, no. 6, pp. 661–678, 2001.

[4] S. Yin, X. Zhu, S. Member, and O. Kaynak, "Improved PLS Focused on Key-Performance- Indicator-Related Fault Diagnosis," IEEE Trans. Ind. Electron., vol. 62, no. 3, pp. 1651–1658, 2015.

[5] C. Ruiz-cárcel, Y. Cao, D. Mba, L. Lao, and R. T. Samuel, "Statistical process monitoring of a multiphase flow facility," Control Eng. Pract., vol. 42, pp. 74–88, 2015.

[6] G. Stefatos and A. Ben Hamza, "Dynamic independent component analysis approach for fault detection and diagnosis," Expert Syst. Appl., vol. 37, no. 12, pp. 8606–8617, 2010.

[7] X. Li, X. Yang, Y. Yang, I. Bennett, A. Collop, and D. Mba, "Canonical variate residuals-based contribution map for slowly evolving faults," J. Process Control, vol. 76, pp. 87–97, 2019.

[8] K. Liu, N. Z. Gebraeel, and J. Shi, "A Data-Level Fusion Model for Developing Composite Health Indices for Degradation Modeling and Prognostic Analysis," IEEE Trans. Autom. Sci. Eng., vol. 10, no. 3, pp. 652–664, 2013.

[9] J. Chen, C.-M. Liao, F. R. J. Lin, and M.-J. Lu, "Principle component analysis based control charts with memory effect for process monitoring," Ind. Eng. Chem. Res., vol. 40, no. 6, pp. 1516–1527, 2001.

[10] H. Ji, X. He, J. Shang, and D. Zhou, "Incipient fault detection with smoothing techniques in statistical process monitoring," Control Eng. Pract., vol. 62, pp. 11–21, 2017.

[11] H. Ji, X. He, J. Shang, and D. Zhou, "Incipient Sensor Fault Diagnosis Using Moving Window Reconstruction-Based Contribution," Ind. Eng. Chem. Res., vol. 55, no. 10, pp. 2746–2759, 2016.

[12] B. Wei, N. Xie, and A. Hu, "Optimal solution for novel grey polynomial prediction model," Appl. Math. Model., vol. 62, pp. 717–727, 2018.

[13] Z. Chen, "Bayesian filtering: From Kalman filters to particle filters, and beyond," Statistics (Ber)., vol. 182, no. 1, pp. 1–69, 2003.

[14] C. Chen, B. Zhang, G. Vachtsevanos, and M. Orchard, "Machine condition prediction based on adaptive neuro-fuzzy and high-order particle filtering," IEEE Transactions on Industrial Electronics, vol. 58, no. 9. pp. 4353–4364, 2011.

[15] H. Hotelling, "New Light on the Correlation Coefficient and its Transforms," J. R. Stat. Soc. Ser. B, vol. 15, no. 2, pp. 193–232, 1953.

[16] P. E. P. Odiowei and C. Yi, "Nonlinear Dynamic Process Monitoring Using Canonical Variate Analysis and Kernel Density Estimations," IEEE Trans. Ind. Informatics, vol. 6, no. 1, pp. 36–45, 2010.

[17] J. Benesty, S. Member, J. Chen, and Y. A. Huang, "On the Importance of the Pearson Correlation Coefficient in Noise Reduction," IEEE Trans. Audio. Speech. Lang. Processing, vol. 16, no. 2008, pp. 757–765, 2008.

[18] K. Pearson, "Mathematical contributions to the theory of evolution.—III. Regression, heredity and panmixia," Philos. Trans. R. Soc., vol. 187, pp. 253–318, 1895.

[19] K. Pearson, "LIII. On lines and planes of closest fit to systems of points in space," London, Edinburgh, Dublin Philos. Mag. J. Sci., vol. 2, no. 11, pp. 559–572, 1901.

[20] S. Liu, Y. Yang, and J. Forrest, Grey data analysis. Springer Singapore, 2017.

[21] K. Javed, R. Gouriveau, and N. Zerhouni, "Enabling Health Monitoring Approach Based on Vibration Data for Accurate Prognostics," IEEE Trans. Ind. Electron., vol. 62, no. 1, pp. 647–656, 2015.

[22] S. Liu, J. Forrest, and Y. Yang, "A brief introduction to grey systems theory," Grey Syst. Theory Appl., vol. 2, no. 2, pp. 89–104, 2012.

[23] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[24] Robert F. Engle, "Autoregressive Conditional Heteroscedacity with Estimates of variance of United Kingdom Inflation," Econometrica, vol. 50, no. 4, pp. 987–1008, 1982.