IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

Feature Extraction o Condition Monitoring Data on Heavy Equipment's Component Using Principal Component Analysis (PCA)

To cite this article: M A Yudha *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **598** 012088

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

This content was downloaded from IP address 120.234.63.196 on 18/09/2019 at 07:31

IOP Conf. Series: Materials Science and Engineering **598** (2019) 012088 doi:10.1088/1757-899X/598/1/012088

Feature Extraction o Condition Monitoring Data on Heavy Equipment's Component Using Principal Component Analysis (PCA)

M A Yudha[1], I Surjandari[1], and Zulkarnain[1]

[1]Department of Industrial Engineering, Universitas Indonesia, Depok 16424, Indonesia

m.andi.yudha@gmail.com

Abstract. The maintenance strategy is significantly important to minimize risk and impact on equipment productivity from component failure. The mechanical transmission on heavy equipment has a function to change speed and torque from engine to final drive. Because of the function that carries high loads which leads to an increase in wear particles, a condition monitoring (CM) approaches is employed. CM data is consisting of 26 parameters and need to reduce the dimension for simplifying correlated variables into fewer independent principal components (PCs). Principal component analysis (PCA) method has been applied to this dataset and deciding 10 PCs with explaining 73.62% variability of the data.

1. Introduction

The main concern at heavy equipment on long-term utilization is maintenance and repair costs, especially in major component replacement. A maintenance strategy is needed to minimize risk and impact on equipment productivity from component failure. Wear debris analysis method has been successful to improve and examine equipment reliability[1]. The mechanical transmission on heavy equipment has a function to change speed and torque from the engine to the final drive. Because of the function that carries high loads which leads to an increase in wear particles, condition monitoring approaches are employed. Condition monitoring (CM) has the advantage to reduce cost when abnormal behavior of the component takes place with maintenance intervention [2].

In reality, a big volume of data has been collected from various sources in the CM process, i.e oil analysis (wear trend), visual inspection, and site condition. Oil analysis to predict transmission degradation have been carried in much numerous research [3] but considering a single concentration data in oil analysis [4]. Characterizing the condition of transmission with extracting a composite degradation index with multiple data has been developed and providing an accurate failure time prediction [5]. Principal component analysis (PCA) has an aim to explain the structure's correlation using a smaller set combination of variables [6]. The PCA method has been applied in the previous study to extract the features of vibration signals in gear malfunctions as predictors in fault detection and achieve a high accuracy [7] and a study for dimension reduction of 94 variables to extract the important factors with PCA in the pre-processing stage with 65.58% of the variability the data has been developed [8].These achievements of the PCA approach for dimension reduction could be leveraged in the mechanical transmission research to extract the important faeatures from multiple resource's data.

Content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
Published under licence by IOP Publishing Ltd

This research has an aim to extracting the features from CM data, i.e oil analysis (wear trend) and visual inspection, on heavy equipment's transmission using the PCA method. These data have more than 20 elements and need to reducing dimension before using as a predictor on modeling and prediction method. This condition would be easily interpreted and avoided overfitting in a model or prediction. The paper sections are organized as follow. Section 2 presents the importance of mechanical transmission in heavy equipment with condition monitoring dataset which has been collected from the health monitoring database. Section 3 describes how the PCA method was applied in the study with 10 principal components has been extracted. Finally, the paper is concluded in Section 4.

2. Materials and Methods

2.1. Power-shift transmission
Power train on heavy equipment has a function to transmit power from engine to final drive and this system consists of the torque converter, transmission, differential, and final drive. The transmission has an important role in heavy equipment when failure will significantly impact and losses [9]. The highest failure rate among other components in heavy equipment at coal mining is the transmission [10]. A Power-shift transmission is an object for this study which applied on the machine that often changes speed and varied workload. This type of transmission consists of planetary gear and clutches and can be illustrated in figure 1. The power-shift transmission has an advantage with on the go shifting which the gear shifting using a fluid clutch, so that the displacement is carried out directly without breaking the connection between the engine and transmission.

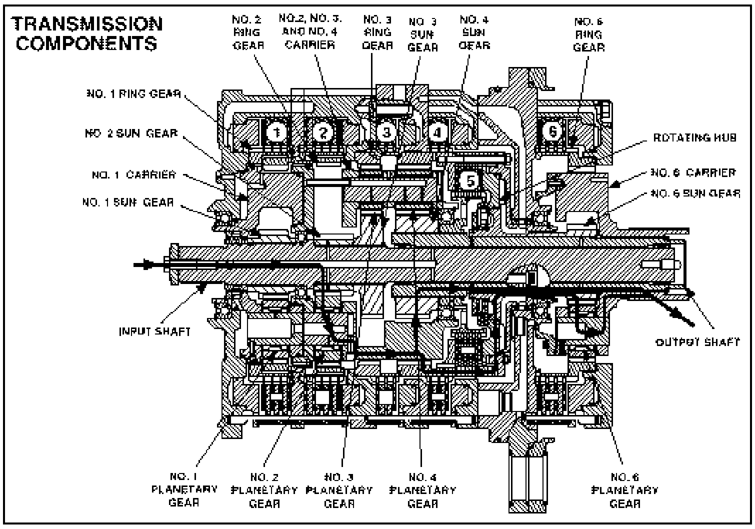

Figure 1. Power-shift Transmission

2.2. Condition monitoring data
CM strategy is a proactive process based on heavy equipment utilization's data and information and will generate a recommendation to repair or maintenance a component. Avoiding an unnecessary maintenance task and labor standby is using CM process, therefore the first challenging step is determining the set of parameters. Degradation of power-shift transmission has been considered in this paper which using wear trend analysis and visual inspection of magnetic plug and filter cut. This paper is using a data set that consists of 26 variables from 27 units and shown in table 1. The data set comes from a health monitoring database which is the equipment's data and visual inspection recorded. Replacement of lubrication oil and filter in the data set can be described as TRUE (replacing) and

FALSE (not replacing) and a value in each parameter described as a number of particle or compound has been detected in the transmission oil.

Table 1. (a) Oil analysis data and visual inspection on 2017-2018

#Sample	oilchanged	filterchanged	Si	Al	Cr	Fe	Pb	Cu	Sn	Ni	Na	K	Mo	Zn	Mg
01	FALSE	TRUE	3	3	0	6	0	2	0	0	2	0	91	965	28
02	TRUE	TRUE	4	2	0	6	0	2	0	0	4	1	81	985	42
03	FALSE	TRUE	2	1	0	4	1	2	0	0	1	0	77	925	72
04	FALSE	TRUE	6	3	0	5	0	3	0	0	0	0	70	852	16
05	FALSE	TRUE	4	2	0	5	1	7	0	1	0	0	52	958	17
06	TRUE	TRUE	4	2	0	6	0	16	0	1	0	0	48	964	14
…															
724	FALSE	TRUE	4	1	0	4	0	9	0	0	0	0	49	884	19

Table 1. (b) Oil analysis data and visual inspection on 2017-2018

#Sample	Ca	P	B	V100	PQ	PC06	PC14	ISO.6	ISO.14	Filter Cut	Magnetic
01	4406	938	190	10.85	1	5784	86	20	14	A	A
02	3732	1011	194	10.66	1	2603	106	19	14	A	A
03	3636	938	173	11.06	0	3492	281	19	15	A	A
04	3507	863	161	10.91	0	29048	4710	22	19	A	A
05	3617	987	118	10.59	1	8214	952	20	17	A	A
06	3396	894	100	10.98	10	13359	1452	21	18	A	A
…											
724	2924	889	104	10.8	2	11202	1314	21	18	A	A

Oil sample and visual inspection data are collected every 250 hours during operation of the machine. The Transmission sample was run to replace with two conditions, replace before failure due to a maintenance strategy and replace after failure occurred.

2.3. Principal component analysis (PCA)

The interpretation of analysis can be unnecessary complicate when using too many predictor variables and need to reduce the dimension. Dimension-reduction that popular and effective for modeling is the PCA method [11]. PCA technique has been developed in many studies, i.e feature extraction on gearbox condition [12],on steam turbines before processed using BPNN [13], construction degradation index (DI) on spectral oil data[5], designing a degradation index on roller bearing [14]. Extracting components can be following these criteria:

a. Only components with an eigenvalue greater than 1 (one) should be retained
b. How much proportion of the variance that components to explain?
c. Minimum communality from the components based on researcher's need
d. The maximum number of components should be retained based on a scree plot as shown in figure 2.

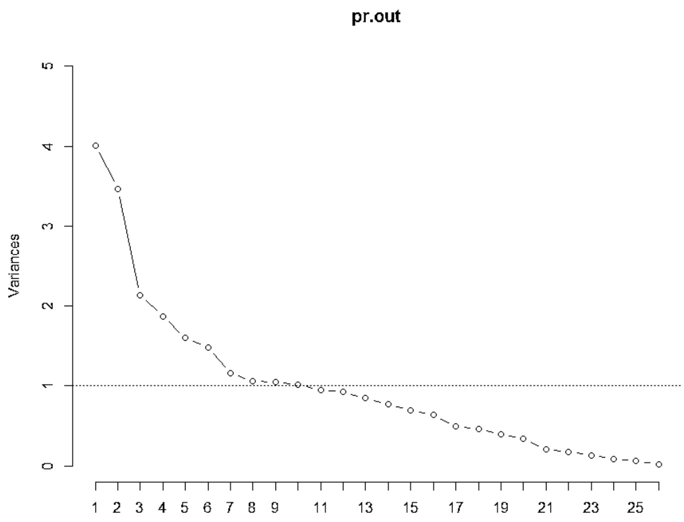

Figure 2. Example of a scree plot

Projecting the dataset onto the subspace of lower dimensionality is the basic concept of PCA [12] The procedure of PCA can be described as follows [14]:

1. Defining an array from original data, CM dataset, as multiple characteristics as

$$X = \begin{bmatrix} X_1(1) & X_1(2) & \dots & X_1(m) \\ X_2(1) & X_2(2) & & X_2(m) \\ \dots & & & \dots \\ X_k(1) & X_k(2) & \dots & X_k(m) \end{bmatrix}, i = 1,2,\dots,k; j = 1,2,\dots,m. \quad (1)$$

where k is the number of wear or rate analysis, m is the number of parameters, and $X_i(j)$ is the jth parameter of ith wear analysis, i.e Magnetic plug rating, filter cut rating, oil changed, filter changed, wear element analysis, oil viscosity, etc.

2. Calculating the corresponding correlation coefficient as

$$R_{jn} = \left(\frac{Cov(x_i(j), x_i(n))}{\sigma_{x_i}(j)\sigma_{x_i}(n)}\right), j = 1,2,\dots,l; n = 1,2,\dots,l. \quad (2)$$

$\mathrm{Cov}(\mathrm{x_i(j)}, \mathrm{x_i(n)})$ is the covariance of sequences $\mathrm{x_i(j)}$ and $\mathrm{x_i(n)}$, $\sigma_{\mathrm{x_i}}(\mathrm{j})$ and $\sigma_{\mathrm{x_i}}(\mathrm{n})$ are the standard deviation of sequence $\mathrm{x_i(j)}$ and $\mathrm{x_i(n)}$, respectively.

3. From the correlation coefficient array, eigenvalues and eigenvectors were determined as

$$(R - \lambda_m I_k)V_{im} = 0, \quad (3)$$

$\lambda_\mathrm{m}(m = 1,2,\dots,n), \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_\mathrm{n} \geq 0$ is the eigenvalue and $\mathrm{V_{im}} = [b_{m1}b_{m2} \dots b_{mn}]$ is the eigenvectors and correspond to the eigenvalue λ_m.

4. Determining the contribution rate of each component as follows

$$b_m = \frac{\lambda_\mathrm{m}}{\sum_{j=1}^{n}\lambda_\mathrm{j}} \; (m = 1,2,\dots,n), \quad (4)$$

$$\mathrm{BC} = \textstyle\sum_{\mathrm{m=1}}^{\mathrm{n}} \mathrm{b_m}, \quad (5)$$

where b_m is the contribution rate of each component and BC is cummulative contribution rate.

5. The last step is obtaining the principal component from

$$Z_{km} = \sum_{i=1}^{n} x_k(i) \, . \, V_{im} \tag{6}$$

where Z_{k1} is the first principal component, Z_{k2} is the second, etc. The most contribution rate in the data is the first principal component, therefore the principal components are aligned in descending order.

3. Result and Discussion

This study using oil analysis and visual inspection data on transmission at 27 unit of dump trucks 100-tons type from 2017-2018 and consists 26 variables.

3.1. Data Preprocessing

Data preprocessing in this study consists of data cleaning and scaling, i.e z-score normalization. The first step of this process is data cleaning where the missing data for processing is replacing with the field mean (numerical data) or the mode (categorical data). The data has 26 variables and length 406 observations is used in this study. Data scaling is the next step of this process with normalizing their numeric variables. The aims of this process are avoiding bigger scale domination over smaller scale data [11]. Scaled data that processed in this study was applied with z-score standardization which the formula as follows:

$$Z - score = \frac{X - mean(X)}{SD\ (X)} \tag{7}$$

In statistical analysis, Z-score standardization is very widely used in the world. This method is scaling the differences between value and mean value with the standard deviation (SD).

3.2. Feature Extraction

Feature extraction from the CM dataset was used to reduce the dimension of variables into a smaller set of linear combinations, called as components, while described the maximum variability of the data. PCA in this paper has an aim to simplifying correlated variables into fewer independent principal components (PCs). CM datasets are selected and shown in table 1, then a principal component analysis is applied to this dataset. The cumulative contribution and contribution rate of each component are obtained and displayed in table 2.

Table 2. Cumulative contribution and contribution of principal components

Principal Components	Eigenvalue	Contribution (%)	Cum. Contribution (%)
1	3.66	14.58	14.58
2	3.34	13.29	27.86
3	2.19	8.72	36.59
4	1.82	7.26	43.85
5	1.60	6.36	50.21
6	1.37	5.46	55.66
7	1.27	5.05	60.72
8	1.15	4.56	65.28
9	1.09	4.34	69.62
10	1.01	4.01	73.62
11	0.91	3.63	77.25

IOP Conf. Series: Materials Science and Engineering **598** (2019) 012088 doi:10.1088/1757-899X/598/1/012088

12	0.85	3.37	80.62
13	0.81	3.23	83.85
14	0.70	2.77	86.62
15	0.67	2.68	89.30
16	0.54	2.14	91.45
17	0.45	1.80	93.25
18	0.42	1.68	94.92
19	0.31	1.22	96.14
20	0.28	1.11	97.25
21	0.21	0.84	98.09
22	0.17	0.67	98.77
23	0.15	0.59	99.36
24	0.09	0.35	99.71
25	0.06	0.23	99.94
26	0.02	0.06	100.00

Deciding the appropriate number of retained principal components could be subjective. The eigenvalues of each component are greater than 1 add up to 10 PCs that shown in figure 3. Following the eigenvalue criterion about extracting the component which eigenvalues greater than 1 and consider the factor loading of each parameter, 10 (ten) principal components are selected. Based on table 2, the first PCs explaining around 14.58% variability of the data and drops significantly on the third PCs. Retaining the ten PCs, 73.62% explaining the variability of the data.

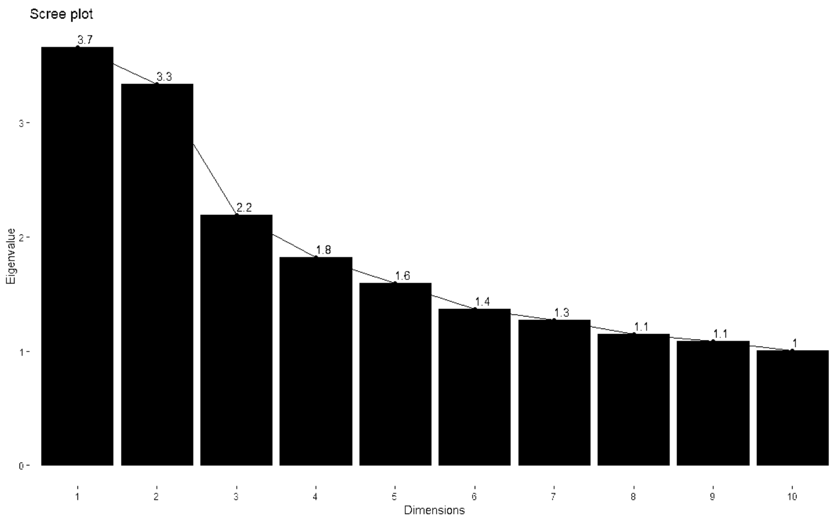

Figure 3. Scree plot for an eigenvalue of each component

Table 3. Component matrix for extracting 10 components

	Component Matrix									
	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**	**PC7**	**PC8**	**PC9**	**PC10**
oilchanged2		0.45								
filterchanged2		0.82								
Filter.Cut		0.99								
Magnetic		1.07								
Si.1	0.58									
Al.1								0.53		
Cr.1					-0.73					

Fe.1				0.68						
Pb.1				0.47						
Cu.1							-0.72			
Sn.1									0.62	
Ni.1					-0.50					
Na.1				0.55						
K.1							0.42			
Mo.1			0.61							
Zn.1	0.61									
Mg.1									-0.53	
Ca.1						-0.74				
P.1	0.61									
B.1	-0.61									
V100.1										-0.48
PQ.1	0.15									
PC06.1			0.66							
PC14.1	0.48									
ISO.6.1			0.63							
ISO.14.1			0.56							

Factor loading of each parameter at 10 PCs is shown in table 3 and could be examining the salient characteristic of each component. For ease of interpretation when giving a title of each component as follows:

1. PCs 1 - The Dirt Component: Strong tendency among parameters Si, Zn, P, B, PC14, and PQ are indicating that dirt from the environment has been entering in the transmission compartment.
2. PCs 2 – The Inspection Component: Parameter oil changed, filter changed, filter cut, and a magnetic plug are representing a visual inspection of field inspector.
3. PCs 3 – The Oil additive Component: Strong tendency among parameters Mo, PC06, ISO.6, and ISO.14 are indicating that lube additive has been mix with transmission oil.
4. PCs 4 – The Gears Wear Component: Fe (iron), Pb (lead), and Na (sodium) are base material for a planetary gear in the transmission.
5. PCs 5 – The Roller bearings Component: Strong tendency on parameters Cr (chrome) and Ni (nickel) indicating a problem in roller bearing which leads material could be plugged in filter system and/or adhere on magnetic screen.
6. PCs 6 – The Calcium Component: Ca (calcium) is indicating an airborne contaminant that could be shown at some site.
7. PCs 7 – The Copper Component: Cu (copper) is a parameter that indicating a problem in clutch plates in the transmission.
8. PCs 8 – The Aluminum Component: Al (aluminum) is a parameter that indicating a problem in pumps in the transmission.
9. PCs 9 – The Steel Alloy Component: Strong tendency on parameters Sn (tin) and Mg (magnesium) indicating an airborne contaminant and could be rubbing the bearing which based on Sn.
10. PCs 10 – The Viscosity Component: V100 (oil viscosity) is a parameter that indicating degradation on oil viscosity and recommended to drain and change the transmission oil.

4. Conclusions and future research

This study has been developed feature extraction CM data using PCA method and can be used for power-shift transmission type. PCA method is used to determine the principal combination of parameters which can be used for a variable in modeling and predictor. The parameter in CM data

IOP Conf. Series: Materials Science and Engineering **598** (2019) 012088 doi:10.1088/1757-899X/598/1/012088

consists of 26 parameters can be reduced into 8 principal components and explaining 65.28% variability of the data. The limitation of this study is the type of component on heavy equipment and sample range in a year's data. Therefore, for future research, we should be considering various type components, e.g engine, final drive, etc, and sample range greater or equal 5 year's data.

References

[1] Kumar A, Ghosh SK 2013 *Measurement: J. Int Meas Confed* **131** 412–8.
[2] Mosallam A, Medjaher K, Zerhouni N 2013 *Chem Eng Trans* **33** 139–44.
[3] Lee J, Wu F, Zhao W, Ghaffari M, Liao L and Siegel D 2014 *Mechanical Systems & Signal Processing* **42** 314-334
[4] Liu Y, Ma B, Zheng C S and Xie S Y 2015 *Spectroscopy and Spectral Analysis* **35** 2620-4
[5] Yan S, Ma B, and Zheng C 2018 *Industrial Lubrication and Tribology*
[6] Larose D T and Larose C D 2015 *Data Mining and Predictive Analytics* (New Jersey: John Wiley & Sons) 92.
[7] Yang CY, and Wu TY 2015 *Meas J Int Meas Confed* **61** 75–87.
[8] .Wang J and Wang X. 2013 *Wear* **304** 96–102.
[9] Li Z, Jiang Y, Hu C and Peng Z 2016 *Meas J Int Meas Confed* **90** 4–19.
[10] Qian P P 2015 *Fault Diagnosis and Reliability Analysis for Transmission System of Shearer Cutting Part* Ph. D Thesis (Xuzhou: China University of Mining and Technology).
[11] Dhini A, Kusumoputro B and Surjandari I 2017 *Int Conf Aware Sci Technol* 149–54.
[12] Li W, Shi T, Liao G and Yang S 2003 *J Qual Maint Eng* **9** 132–43.
[13] Gupta A Barbu A 2018 *Pattern Recognit* **78** 215–27.
[14] Wu J, Wu C, Lv Y, Deng C and Shao X 2017 *Ind Manag Data Syst* **117** 713-28
[15] Zekić-Sušac M, Šarlija N and Pfeifer S 2013 *Croat Oper Res Rev* **4** 306–17.