Coal Pulverizer Prognostics Data Challenge in PHMAP 2017 and Suggestions for Future Studies

Hyunseok Oh[1], Yun Ho Seo[2], Min-Hwan Mo[3] and Joo-Hyung Kim[4]

[1]*School of Mechanical Engineering, Gwangju Institute of Science and Technology, Gwangju, 61005, Korea*
hsoh@gist.ac.kr

[2]*Department of System Dynamics, Korea Institute of Machinery and Materials, Daejeon, 34103, Korea*
yhseo@kimm.re.kr

[3]*Technology Research and Development Institute, KEPCO Plant Service and Engineering Co., LTD, Naju, 58217, Korea*
saintmo@kps.co.kr

[4] *Department of Mechanical Engineering, Inha University, Incheon 22212, Korea*
joohyung.kim@inha.ac.kr

ABSTRACT

Pulverizers in a power plant are used to grind coal into the form of a fine powder for combustion in a power plant. To secure reliable operation, redundant pulverizers should be installed in power plants and monitored. Pulverizers can be operated and maintained in a cost-effective manner by correctly estimating the current health condition and remaining useful life of the pulverizer's gearbox system. To this end, the Data Challenge Committee of the PHM Asian Pacific 2017 (PHMAP 2017) conference organized an open competition on the topic of coal pulverizer health estimation based on a real working power station. This paper presents the original problem and given facts, as well as the list of winners of the Data Challenge Competition. We anticipate that this paper can be used as a reference in the development of a prognostic method that can accurately predict the health conditions of coal pulverizers.

1. INTRODUCTION

Pulverizers in a power plant must perform as intended during their expected design life. Fault diagnosis of numerous systems in power plants has received significant attentions from both academia as well as industry, with the goal to realize condition-based maintenance strategies. However, pulverizers found in power plants have received relatively little attention. This paper focuses on the fault diagnosis and prognosis of gearboxes in pulverizers.

Already numerous fault diagnosis methods have been proposed to analyze vibration signals, including time-domain, frequency-domain, and time-frequency domain analysis (Tian, Morillo, Azarian and Pecht, 2015; Chow and Tan, 2000; Cabal-Yepez, Garcia-Ramirez, Romero-Troncoso, Garcia-Perez and Osornio-Rios, 2013). Also a review of prior work can be found in the publications (Lee, Wu, Zhao, Ghaffari, Liao and Siegel, 2014) with regard to the state-of-the-art of PHM for rotating machines. Existing methods can provide accurate results when relevant vibration signal data are used during the development stage. Prior methods typically incorporate data from test beds that emulate real-world power plant systems. Unfortunately, it is unclear whether diagnostic and prognostic methods developed with test bed data can perform as intended when data from a real power plant is incorporated. Testbed data do not incorporate uncertainties, such as variability and randomness, which are often observed in data from real power plants in the field.

To address this issue, a committee was organized to devise a problem for the PHMAP 2017 Data Challenge Competition. The PHMAP 2017 Data Challenge was a fully open competition in which collaboration was encouraged. Participating teams might be composed of any combination of students, researchers, and industry professionals. Solutions were proposed by teams, results were evaluated by the PHMAP 2017 Data Challenge Committee, and all teams were ranked. The top three scoring teams were invited to present at a special session of the conference and are recognized during the conference.

Corresponding author: J.-H. Kim (e-mail: joohyung.kim@inha.ac.kr).

This paper describes the details of the Data Challenge. The remaining sections are organized as follows. Section 2 presents the problem, data sets, and scoring metrics. Section 3 presents the results and discussion. Section 4 concludes the paper with suggestions for future studies.

2. PROBLEM, DATASETS, AND SCORING

The problem for the 2017 PHMAP Data Challenge Competition is described in Section 2.1. The details of the data sets are presented in Section 2.2, followed by the scoring function, which is proposed in Section 2.3.

2.1. Problem Definition

The goal of the problem is to develop a prognostic method that (1) estimates the date for a replacement of pulverizer's gearbox and (2) determines the most severely degraded component in the gearbox. Specifications regarding the gearboxes can be found in Table 1.

The committee provided data sets from six pulverizers from a real power plant in South Korea. The data sets are available at http://www.phm.or.kr/info/board.php. Data sets from three pulverizers did not include any faults or failures. The data sets were used for training of proposed prognostic methods. Data sets were also acquired from three other pulverizers that were repaired with gearbox replacement. However, specific sets of data collected before the date of gearbox replacement were intentionally eliminated by the organizers of the Data Challenge Competition since abnormal deviations in the data of the specific sets from nominal values were attributed to other systems than the gearbox. Each Data Challenge team was asked to predict the actual date for the gearbox replacement of the three pulverizers of interest. The participants were also expected to answer what specific component of the gearboxes was most vulnerable to failure. A list of the gearbox components, and the number of gear teeth, are shown in Figure 1 and Table 2, respectively.

2.2. Data Sets

Pulverizers grind up coal into the form of fine powder for combustion in a furnace of a power plant. The power plant of interest has Pulverizers A to F that have operated for more than 20 years. To run the pulverizer, an electric motor is connected to a large-size grinding system (*i.e.*, mill) through a gearbox (See Figure 2).

Time series data were collected from Pulverizers A to F. They include displacement, velocity, and acceleration signals measured at predetermined locations. The measuring points are indicated in Figure 2. In the data sets, file names were given in the format of:

[6PULV_x_y_z_YYYYMMDD_HHMMSS]

where x indicates the pulverizer whose data were collected including "C", "D", "F", "A", B", and "E"; "y" is the location where sensors were placed, including "Body Case-1", "Body Case-2", "MTR-I_B", and "MTR-O_B"; "z" is a dummy character used for data record; YYYYMMDD and HHMMSS represent the date and time when data were recorded, respectively.

For example, "6PULV_E_가속도_Body-CASE-1-H_20120105_142806" indicates that, for Pulverizer E, acceleration data were recorded from the "Body-CASE-1-H" location at 2:28 PM on January 5, 2012. Individual files contain 4,096 data points with a sampling frequency of either 2,048 or 7,680 Hz.

Table 3 shows the time period of data measurements that were revealed to the participants of the Data Challenge. For training, sensor data from Pulverizers C, D, and F were provided for a given period of time. The gearboxes in the pulverizers performed as intended without any failure during that period of time, although there may have been

Table 1. Specifications of the gearboxes

Specification	Quantity
Weight (ton)	120
Power (kW)	448
Mill speed (rpm)	35.1
Motor speed (rpm)	887
Motor bearing type	Sleeve

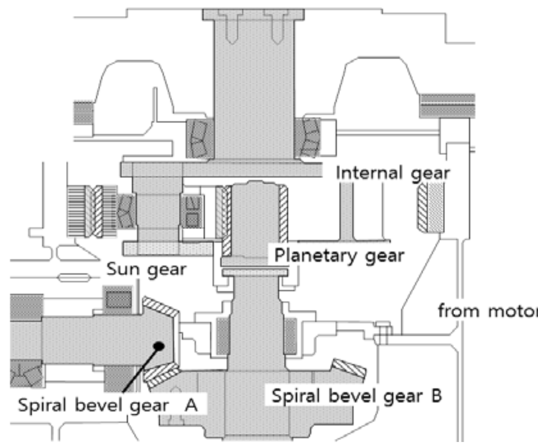

Figure 1. Detailed image of gearbox components.

Table 2. Detailed information: Number of gear teeth

Gear	Quantity
Spiral bevel gear A	15
Spiral bevel gear B	58
Sun gear	15
Planetary gear	33
Internal gear	84

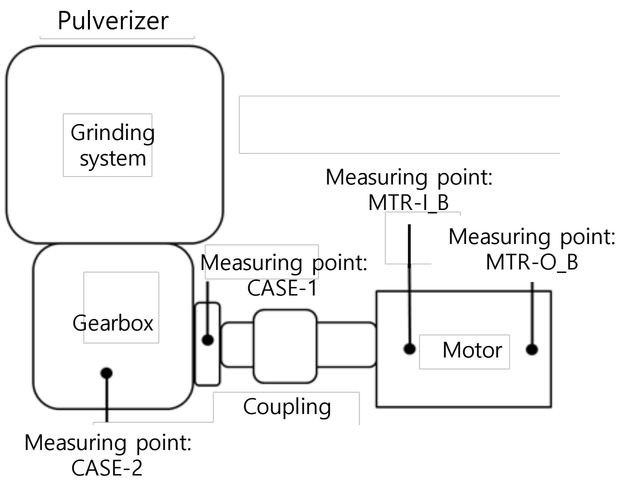

Figure 2. Power transmission of pulverizers: from an electric motor through a gearbox to a grinding system.

normal degradation. The data recorded on September 24 and 26, 2012 represent the idle and load conditions, respectively. For testing, sensor data from Pulverizers A, B, and E were provided. It should be noted that, for Pulverizer E, the data recorded on July 12, 2012 shows an increase of motor current due to a stuck bowl roll during operation. At that time, the roll assembly part was replaced with a new one. After replacement, the pulverizer operated normally.

For Pulverizers A, B, and E, other data sets were available to the organizers. However, only a limited amount of data in Table 3 associated with normal conditions of the gearboxes was revealed to the participants. This is because the problem of the Data Challenge Competition was defined to predict the failure times of Pulverizers A, B, and E.

The participants were also provided with additional data sets associated with faulty conditions of the pulverizers. One data set (*i.e.*, Failure data set 1) was recorded 81 days before the gearbox failure. Another data set (*i.e.*, Failure data set 2) was recorded one day before the gearbox failure. The failure data was from Pulverizer B.

Table 3. Data sets given to Data Challenge participants

Pulverizer	Data recorded		Remark
	From	To	
C	June 14, 2011	June 12, 2013	Training
D	June 14, 2011	June 12, 2013	
F	June 14, 2011	June 12, 2013	
A	June 14, 2011	July 31, 2012	Test
B	September 8, 2011	June 12, 2013	
E	June 14, 2011	July 31, 2012	

2.3. Scoring

The metric for calculating an absolute score was defined as:

$$\text{Absolute score} = \sqrt{w_A \times \left(t_{A,\text{actual}} - t_{A,\text{team}}\right)^a + w_B \times \left(t_{B,\text{actual}} - t_{B,\text{team}}\right)^b + w_E \times \left(t_{E,\text{actual}} - t_{E,\text{team}}\right)^c}$$

where, t_{actual} is the actual date for gearbox replacement of the pulverizer; t_{team} is the gearbox replacement date predicted by each team; and w is the weight factor. If the most severely impacted component (closest to failure) of each pulverizer was correctly answered, a weight factor value of 0.7 was multiplied to the squared term. Otherwise, another value of 1.0 was used for the calculation. a, b, and c are the constant of "2". Here, the a, b, c values were not suggested to the participants throughout the competition and the preliminary winners were determined based on the absolute score.

Another metric for computing team scores is:

$$\text{Team score} = \sqrt{w_A \times \left(t_{A,\text{actual}} - t_{A,\text{team}}\right)^a + w_B \times \left(t_{B,\text{actual}} - t_{B,\text{team}}\right)^b + w_E \times \left(t_{E,\text{actual}} - t_{E,\text{team}}\right)^c}$$

where a, b, and c are constants between 1.6 and 2.4 randomly assigned to each team.

The "team score" was calculated with a particular set of a, b, and c values assigned each team. The a, b, and c values assigned to each team were not changed throughout the competition. It may not be useful to compare the performance between teams. However, each team can track the change of its own scores between submissions and use the team score to improve their algorithm.

3. RESULTS AND DISCUSSION

Actual dates for gearbox replacement and the most severely-degraded components are listed in Table 4. The results of the top three teams are listed in Table 5. A total of sixteen teams from different countries, including South Korea, China, United States, India, Russia, and Singapore, were registered in the Data Challenge Competition as shown in Fig. 3.

Table 4. Answers for the problems in Data Challenge

Actual date for gearbox replacement	
Pulverizer A	Sep. 25, 2012
Pulverizer B	Jul.21, 2014
Pulverizer E	Dec. 9, 2013
Most severely degraded component	
Pulverizer A	Spiral bevel gear (SBG) A
Pulverizer B	SBG A
Pulverizer E	SBG B

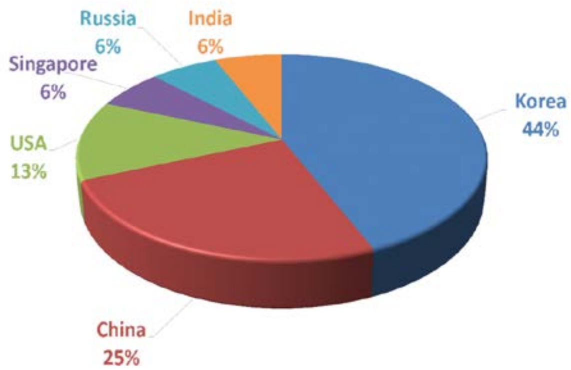

Figure 3. Countries of the participating teams in PHMAP 2017 Data Challenge.

Table 5. Answers of top-three teams

	SHRMer	Starling	PHM_YDDS
Pulverizer A (Error; days), Faulty component	Dec. 14, 2012 (+80) SBG A	Dec. 14, 2012 (+80) Planetary gear	Feb. 28, 2013 (+156) Planetary gear
Pulverizer B (Error; days) Faulty component	Dec. 14, 2013 (-219) SBG A	Jan. 9, 2014 (-193) SBG B	Feb. 2, 2014 (-169) Planetary gear
Pulverizer E (Error; days) Faulty component	Dec. 27, 2013 (+18) SBG A	Nov. 22, 2013 (-17) SBG A	Feb. 12, 2014 (+65) Planetary gear

The absolute scores after final submissions are listed in Table 6. The winners summarized the prognostic methods their team used to estimate the date of gearbox replacement. Park, *et al*. (2017) proposed a method based on a signal processing technique applicable to external noise and disturbance operating conditions. Two features, including an RMS value in the higher frequency zones (HRMS) and an amplitude of a period for high-speed shaft in the quefrency domain (QA_{HSS}), were associated with faulty conditions of gearbox components.

Table 6. Scoring after final submission

Team name	Absolute score	Team name	Abs. score
Josey Mathew	Not submitted	Smart-Inside	455
SHRMer	196	PHM_YDDS	239
PRANA	Not submitted	Prognostia	Not submitted
Starling	210	IAI-KOR-GUA	Not submitted
ISR_TIM	Not submitted	IAIK	Not submitted
PHM_MDSCS	Not submitted	THX	353
Super Chargers	406	PDX_DNS	395
PHM_XS	Not submitted	H.Y.I.E.	483

A regression analysis was used to extrapolate available data points. Details of this proposed method are well specified in the reference by Park, Jeon, Park, Cui, Kim, and Youn (2017). Lee, *et al*. (2017) proposed a prognostic method based on autocorrelation and wavelet packet decomposition techniques. Health indices, including variance, square root amplitude value, and absolute mean amplitude, were determined. A Bayesian regression analysis was conducted to predict dates for gearbox replacement. Song, Ding, and Lu (2017) attempted to solve the problem using a spectrogram and deep learning.

4. CONCLUSIONS AND SUGGESTIONS FOR FUTURE STUDIES

This paper presented the problem, data sets, scoring metrics, and results of the PHAMP 2017 Data Challenge Competition. The gearboxes of pulverizers in real power plants were the focus of the 2017 PHMAP Data Challenge. The goals of the competition were to predict dates for gearbox replacement and determine faulty components of the gearboxes. Three winners out of 16 registered teams were selected based on the calculated scores. The methods of the top-scoring teams are well described in the published references.

The key contributions of PHMAP 2017 Data Challenge and this paper are two-fold. First, state-of-the-art algorithms were introduced and compared by the participants. It was shown that the results proposed by the winning teams are promising. The errors between actual gearbox replacement dates and predicted dates were minimized to be as small as tens of days. Second, the data measured from a real power plant in South Korea was revealed to the public domain. Typically, organizations do not disclose power plant data for proprietary reasons. However, through close collaboration between academia and industry, this data was opened to the public so that the research community can benefit from it. We anticipate that the PHMAP 2017 Data Challenge Competition and this paper can be a reference for developing an improved method for PHM of power plants.

The PHMAP 2017 Data Challenge is not the only competition in the PHM research community. A couple of Data Challenges can be found in Annual Conferences of the PHM Society, in the IEEE International Conferences on PHM. Potential readers who are interested in will be able to find problems, data sets, and results in the past.

The scientific value of so-called "*big data*" has received a great deal of attention as PHM techniques get evolved. The big data may include any type of data used to control components/systems, to record their operational and environmental conditions, and to monitor health conditions such as vibration, acoustic emission, temperature, electrical current, and so on. In the past, there are some excellent data sources for PHM studies, including NASA Prognostics Data Repository, CWRU Bearing Data Center, PHM Society Data Challenge. Nevertheless, for the purpose of PHM research, it is extremely difficult to find open sources for the "big" and "useful" data. The "big" data are often unavailable, while the domain knowledge may improve the lack of the big data in the real problem.

The PHMAP data challenge in this paper is one of our efforts to provide valuable data for the PHM community. The efforts are strongly encouraged among the PHM research community. Ultimately, the efforts would make significant contributions towards the ultimate goal of PHM of mission-critical and safety-related systems, *i.e.*, reliable operation of the systems.

ACKNOWLEDGEMENT

The Data Challenge Committee of the PHMAP 2017 conference would like to express our sincere gratitude to Yong-Wan Kim for recording vibration data from a real power plant and senior engineers in the Technology Research and Development Institute, KEPCO Plant Service and Engineering Co., LTD for allowing us to access the data. This work was partially supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) Grants No. 20172010105590 funded by the Korean Ministry of Trade, Industry and Energy (MOTIE) and also supported under the framework of 2017 international cooperation program (GRDC) through National Research Foundation (NRF) by Ministry of Science, ICT and Future Planning of Korea (MSIP, Grant No: 2017K1A4A3013662).

REFERENCES

Cabal-Yepez, E., Garcia-Ramirez, A. G., Romero-Troncoso, R. J., Garcia-Perez, A., and Osornio-Rios, R. A. (2013). Reconfigurable monitoring system for time-frequency analysis on industrial equipment through STFT and DWT, *IEEE Transactions on Industrial Informatics*, vol. 9(2), pp. 760-771.

Chow, T. W. S. and Tan, H. Z. (2000). HOS-based nonparametric and parametric methodologies for machine fault detection, *IEEE Transactions on Industrial Electronics*, vol. 47(5), pp. 1051-1059.

CWRU Bearing Data Center, https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website, Accessed on Dec. 16, 2017.

Korean Society for Prognostics and Health Management (KSPHM) Data Repository for PHMAP 2017 Data Challenge, http://www.phm.or.kr/info/board.php, Accessed on Apr. 11, 2018.

Lee, G., Kim, S., Song, J., and Kim, T. (2017). Failure prognosis of coal pulverizer gearbox using autocorrelation wavelet packet decomposition method. *In Proceedings of the Asian Pacific Conference of the Prognostics and Health Management Society*, July 12-15, Jeju, Korea.

Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., and Siegel, D. (2014). Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications, *Mechanical Systems and Signal Processing*, vol. 42(1-2), pp. 314-334.

NASA Prognostics Data Repository, https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/, Accessed on December 16, 2017.

Park, J., Jeon, B., Park, J., Cui, J., Kim, M., and Youn, B.D. (2017) Failure prediction of a motor-driven gearbox in a pulverizer under external noise and disturbance. *Smart Structures and System, in press*.

PHM Society Data Challenge Competition, http://www.phmsociety.org/competition, Accessed on December 16, 2017.

Song, D., Ding, Y., and Lu, C. (2017). A prognostics approach for gearbox based on spectrogram and deep learning. *International Journal of Prognostics and Health Management, in press*.

Tian, J., Morillo, C., Azarian, M. H., and Pecht, M. (2016). Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with k-nearest neighbor distance analysis, *IEEE Transactions on Industrial Electronics*, vol. 63(3), pp. 1793-1803.

BIOGRAPHIES

Hyunseok Oh received the B.S. degree from Korea University, Seoul, Republic of Korea, in 2004, the M.S. degree from KAIST, Daejeon, Republic of Korea, in 2006, and the Ph.D. degree from the University of Maryland, College Park, MD, USA, in 2012. Dr. Oh is an Assistant Professor, Gwangju Institute of Science and Technology, Korea. His current research area includes prognostics and health management and model verification and validation. He has received several awards, including the A. James Clark Fellowship (2007), the IEEE PHM Data Challenge Competition Winner (2012), the PHM Society Data Challenge Competition Winners (2014, 2015), and the ACSMO Young Scientist Award (2016).

Yun Ho Seo received the B.S. degree from Yonsei University, Seoul, Korea, in 2003 and the Ph.D. Degree from KAIST, Daejeon, Korea, in 2009. Dr. Seo is a senior researcher in the Korea Institute of Machinery and Materials (KIMM). His current research topics are monitoring, diagnosis, and prognosis of machinery and dynamic analysis of structures.

Min-Hwan Mo received the B.S. and M.S. degrees from Chonnam National University, Gwangju, Korea, in 2000 and 2004, respectively. He is a Senior Research Engineer, Technology Research and Development Institute, KEPCO Plant Service and Engineering Co., LTD, Korea. His current research area includes Predictive Maintenance (PM) and

Prognostics and Health Management (PHM) for Power Plants.

Joo-Hyung Kim received the B.S. and M.E. degrees from the Department of Mechanical Engineering, Inha University, Incheon, Korea, in 1993 and 1995, respectively, and the Ph.D. degree from the Department of Microelectronics and Information Technology, KTH (Royal Institute of Technology), Stockholm, Sweden, in 2005. From 1995 to 2002, he was a senior research engineer with Daewoo and Samsung SDI central research centers, Korea. From 2006 to 2008, he was a senior scientist in the Fraunhofer Institute, Germany, for novel material research in microelectronics. Currently, he is an associate professor in Mechanical Engineering in Inha university and director of the INHA IST-NASA Joint Research Center, 3D Printing Center and GM-PACE Center. His research interests are renewable energy systems, intelligent sensor fabrication, advanced 3D printing, and advanced smart mechanical systems.