Risk Prediction of Engineering Assets: An Ensemble of Part Lifespan Calculation and Usage Classification Methods

Hyunjae Kim[1], Taewan Hwang[2], Jungho Park[3], Hyunseok Oh[4], and Byeng D. Youn[5]

[1,2,3,4,5]*Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul, 151-742, Republic of Korea*
secutus07@snu.ac.kr
michel4546@gmail.com
hihijung@snu.ac.kr
hyunseok52@gmail.com
bdyoun@snu.ac.kr

ABSTRACT

For the 2014 Prognostics and Health Management (PHM) Data Challenge Competition, the PHM Society proposed a problem surrounding risk prediction of engineering assets. We worked to address this problem by statistically analyzing the maintenance records, extracting key data features, and proposing an ensemble method for accurate prediction of imminent failure of assets. The data analysis of maintenance records provided two key pieces of information: 1) parts and part replacement reasons were able to be classified into corrective and scheduled maintenance actions, and 2) a linear relation was found between failure frequency and usage time. Based on this information, we proposed two risk-prediction methods, namely, a method based on part lifespan calculation and a method based on usage classification. Further work showed that the ensemble approach, which combined these two methods with a risk assignment formulation, provided more accurate risk prediction. The score predicted by the ensemble approach ranked in the second place in the 2014 PHM Data Challenge Competition.

1. INTRODUCTION

As sensing and data acquisition techniques have advanced, the amount of sensory data available from engineered assets has increased dramatically. The data are collected for different purposes, such as to aid in optimizing operation, condition monitoring, and maintenance. Among the data sources, maintenance data are of great importance for reliability analysis of engineering assets. In principle, analysis of the data should help optimize decision making in terms of maintenance scheduling and cost. However, in reality, it is extremely challenging to extract useful information from the maintenance data (Heng, Zhang, Tan, & Mathew, 2009). In practice, this process often suffers from a lack of relevant techniques, limited budgets, and time constraints.

For the 2014 Prognostics and Health Management (PHM) Data Challenge Competition, the PHM Society proposed a problem of risk assessment for engineering assets that makes use of a maintenance record database. The primary task of the problem was to predict the risk of the selected assets with the limited information available in the maintenance data. The limited data were to be interpreted as maintenance data without any description about the type of asset, usage loading, failure modes and mechanisms. This extra challenge made it even more difficult to predict imminent failure of assets than in prior studies. In previous work, and in practice, some of this additional information is generally available.

Health management of engineering assets has received significant attention due to its potential value. About two decades ago, Crow (1990) developed a Weibull-Poisson process model that represents a repairable system's reliability. Lawless, Hu, and Cao (1995) proposed relevant distribution models by fitting reliability event data. Even more recently, a systematic approach on how to select a relevant model was investigated by Louit, Pascual, and Jardine (2009). These studies, however, only focused on statistical modeling of maintenance data and did not address how to predict the risk.

Recently, prediction of the future condition of assets started to be actively addressed. Zhou (2011) adopted the auto-regression function to model the failure number of a bus fleet. The fitted model can be used to predict the failure number in the future. Taghipour, Banjevic, and Jardine (2011) attempted to analyze maintenance records of a medical device. The authors presented a method that classifies the different types of failure including soft and hard failure. Hjartarson and Otal (2006) developed techniques that quantify the impact of preventative maintenance on health indices and assess future

asset conditions. Braglia, Carmignani, Frosolini, and Zammori (2012) presented a multivariate statistical approach that calculates the mean time between failures. Jahromi, Piercy, Cress, Service, and Fan (2009) presented a condition-based asset management tool that quantifies power transformer degradation. Most of the approaches in the literature relied on the scheme of statistical modeling and machine learning. In actual field applications, the accuracy achieved on paper is not guaranteed due to invalidity of the assumptions and a lack of sufficient prior knowledge.

In this competition, we attempted to predict failure with extremely limited information; risk prediction in our situation involved a large number of unknown parts in the given assets, combined with a short span of training datasets. These additional challenges were not addressed in any previous study. Therefore, the goal of this study was to extract useful information utilizing the limited maintenance data offered by the PHM Data Challenge Committee and to accurately assess the risk of the engineering assets for the near future. To effectively present the procedures of achieving this goal, this paper is organized as follows. Section 2 provides an overview of the original problem described by the PHM Data Challenge Committee and the raw datasets (i.e., maintenance records). Sections 3 presents the analysis of the datasets to find clues related to the health condition of the assets. Section 4 explains the methods proposed to assess the risk of the assets using the features extracted from the analysis in Section 3. Section 5 offers conclusions and suggests future work.

2. PROBLEM AND DATASETS

This section summarizes the problem given by the PHM Data Challenge Competition Committee. First, the basic information (the engineering assets, reasons for part consumption, and objective) of the problem is given to provide context for the risk prediction. Second, the training and test datasets are described in terms of data types and sizes. Finally, the scoring metric is defined to evaluate the risk prediction performance of the test instances.

First, there are approximately 1800 engineering assets in the data set, but essential information (i.e., name, type, function of the asset) is not provided. Also, each "asset" is composed of multiple "parts." As in the case of the assets, no part descriptions are provided. The maintenance records of the assets include failure time, part replacement time, identification number of the replaced parts, quantity of the replaced parts, reason for the part replacement, and asset usage.

Second, the records are composed of two distinct datasets: training and test. The training dataset encompasses all maintenance records listed above, whereas the test dataset excludes the knowledge of failure time from the maintenance records. Thus, with the test dataset, the participants need to answer if the risk of an asset is high or low at a certain

	Number of records	Number of assets	Number of part consumption reasons	Number of part types
Training data	270830	1913	14	3192
Test data	176509	2076	13	2863
Sum	-	2077	14	3868
Intersection	-	1912	13	2187

Table 1. Summary of the training and test data

instance defined by the PHM Data Challenge Competition Committee. The training and test datasets are briefly summarized in Table 1. A test instance for a particular asset is given at a specific time. The number of test instances is approximately 10,000. The term "high risk" is defined to be when a failure occurs within three time units; in other words, when a failure is imminent. Otherwise, an asset is considered to have "low risk."

Third, the performance of a risk prediction method for the test cases must be carefully evaluated. The following scoring metric is given by the PHM Data Challenge Competition Committee and is used for scoring:

$$Score = L / N + H / N \tag{1}$$

where N is the number of test instances that are drawn from the categories of "high risk" and "low risk;" the total number of sampled test instances is $2N$; L is the number of test instances that the participant answered correctly in the category of low risk; and H is the number of test instances that the participant answered correctly in the category of high risk. With regard to N, it is worth noting that the chance to have high or low risk is similarly maintained in the problem. However, in real applications, the frequency of the low-risk time instances is typically much higher than that of high-risk time instances. Using this scoring system, when the risk is arbitrarily assigned, the score tends to converge to "one" out of two. When the risk is perfectly predicted, the score should be "two." If the risk prediction is completely wrong, the score should be "zero." Using specific knowledge about usage and part consumptions after the test time instance should be avoided, because this "after the fact" data is not available in real field applications.

3. DATA ANALYSIS

Based on the fact that none of the asset failure modes, causes, and mechanisms were revealed, a data-driven approach was determined to be the most suitable for the risk prediction of the assets. This section analyzes the given training datasets to extract key features in conjunction with the health conditions of the assets.

3.1. Failure time

Figure 1 is a histogram that presents the failure time of all the assets given in the training dataset. As an initial trial, assuming that the amount of failure increases with the accumulation of usage, we conducted a linear regression between the failure count and time. The number of failures in the ranges of one to 365 and 366 to 730 increased to 4329 and 4870, respectively. This result showed a 10% increase between the first and second halves. Curve fitting using high-order polynomial functions is also possible to minimize the residual. However, it is difficult to interpret the physical meaning of those functions.

We calculated the ratio of high to low risk using information about failure time of all the assets in the training datasets. The calculated relative frequency of high risk is 0.027, which indicates that a randomly selected asset is in high risk with the probability of 2.7 %.

3.2. Part consumption

Fourteen reasons for part consumption exist in the training dataset. These reasons can be divided into two groups—corrective maintenance (CM) and scheduled maintenance (SM). CM is the maintenance carried out after detecting failure. It aims at restoring an asset so that the asset continues to perform its intended function. In contrast, SM is the maintenance that is performed to maintain an asset in satisfactory operating conditions. SM is conducted on a predefined schedule regardless of the health condition of the asset.

As a representative example of the CM group, let us consider a part consumption time labeled with the reason of R193. The main characteristic of CM is that CM events typically follow part failure. Figure 2 is a scatter plot of failure time and R193 recording time from fifty randomly chosen assets. It can be easily observed that most corrective maintenance times labeled with R193 are marked right after failure. To quantify the degree of coincidence, a time gap between the failure and a R193 recording near the failure was calculated. The R193 recording that followed failure within five time units was found to be 4,016 out of the total occurrences of 9,199. The rest of the CM group includes R364 and R446.

As a representative example of the SM group, let us consider a part consumption time labeled with the reason of R707. Figure 3 shows the time difference for part consumption labeled with R707 from fifty randomly chosen assets. It is apparent that part consumptions labeled with R707 occur periodically. This implies that R707 is related to a scheduled maintenance activity. Other reasons (including R119, R417, R565, and R783) showed similar behavior, and were thus classified into the SM group.

Table 2 shows the rank of replaced parts for six predominant reasons. For example, for R417, the parts 566684, 663583, and 102391 ranked first to third, respectively. The part
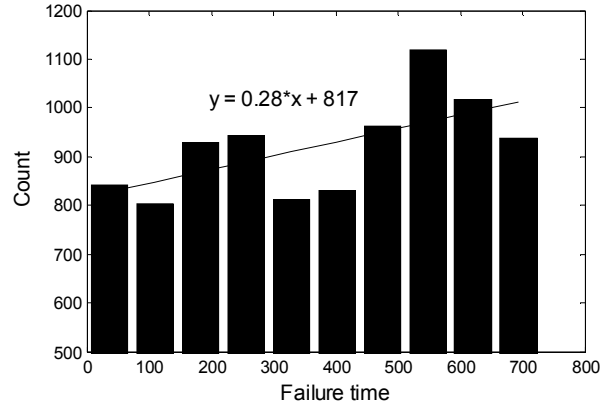

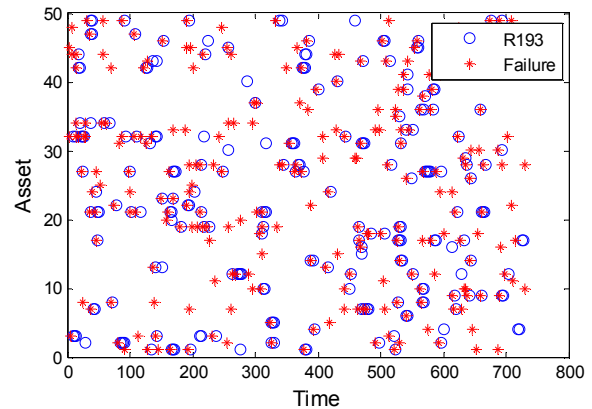Figure 1. Count of failure time

Figure 2. Scatter plot of failure time and R193 recording time from fifty randomly selected assets
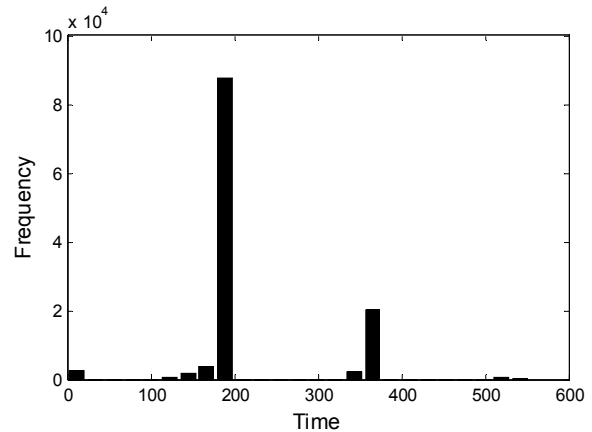

Figure 3. Scheduled maintenance: frequency of R707 time difference

566684 also ranked first for R565 and R707. The number of times that the part 566684 was replaced was 3192, which is about 8% of all replaced parts. For R193 and R446, the part 953340 ranked first. We were able to draw a conclusion from

the observations in the table: replaced parts are distinct according to the reason of part consumption. For example, the part 566684 can be assumed to be replaced based on scheduled maintenance. We assume that the replacements suggested by the manufacturer (scheduled maintenance) are related to wearout failure. In contrast, the part 953340, which was replaced based on corrective maintenance, is an example of a part replaced for reasons related to unexpected failure.

Rank	1	2	3
R193	953340	97048	991287
R364	738107	357852	953340
R417	566684	663583	102391
R446	953340	97048	169319
R565	566684	102391	738107
R707	566684	663583	102391

Table 2. Rank of replaced parts for six major reasons
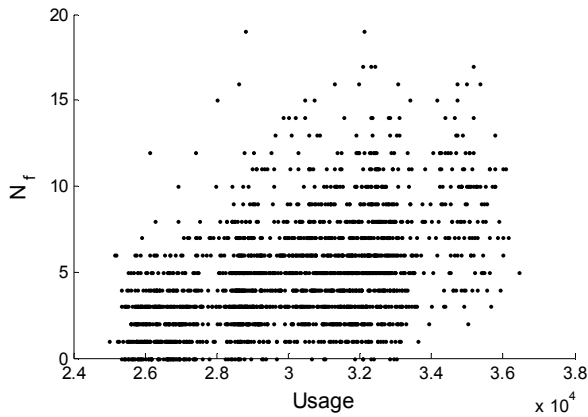
a)
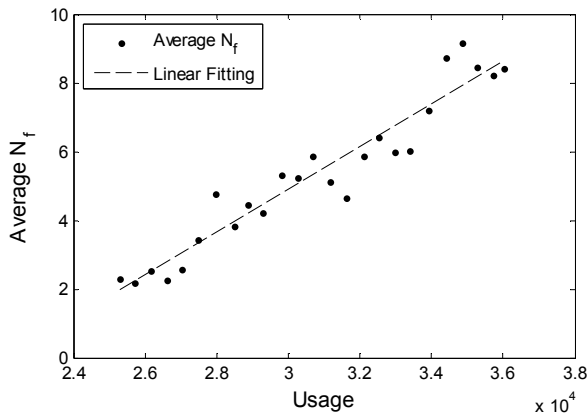
b)

Figure 4. Relation between failure frequency and usage a) scatter plot and b) linear fit of averaged failure frequency and its usage

3.3. Relation between failure frequency and usage

It is assumed that failure occurs more frequently as usage accumulates. To verify this assumption, we plotted the relationship between the failure frequency of assets (N_f) and their usage in Figure 4 a). It was observed that the failure frequencies roughly increase as the usage increases. However, the large deviation in N_f makes it difficult to extract a trend. To overcome this difficulty, the usage range from 25,000 to 36,000 was equally divided into 25 subsections. In each individual subsection, the mean value of N_f was calculated. The mean value of N_f and its usage are shown in Figure 4 b), which clearly shows a linear increase. To quantify the degree of linearity, Pearson's correlation coefficient of the mean and the usage was calculated to be 0.9589. This indicates that there is a strong linear relationship between failure frequency and usage.

3.4. Summary of data analysis

The features extracted through data analysis of the maintenance records can be summarized as follows. First, reasons related to part consumption can be divided into two categories: CM and SM. The parts related to SM are of interest for risk prediction of test assets. Parts related to CM have little value for risk prediction. Second, failure occurs more frequently with the accumulation of usage. It was found that the usage is proportional to the frequency of asset failure.

4. RISK PREDICTION USING AN ENSEMBLE METHOD

In the previous section, it was identified that the replacement of particular parts and the accumulation of usage are closely related to failure occurrence. In this section, two methods are proposed to predict future risk based on the results of the analysis above. An ensemble approach is also proposed to utilize the two different methods more effectively. We compare the performance of the three methods in terms of prediction accuracy.

4.1. Method 1: Part lifespan calculation

The main assumption of the first method is that the event of part replacement is followed by an occurrence of failure. Time gaps between part replacement events and failure occurrences were investigated statistically. First, an individual time gap for the part, p, was calculated between the replacement instance of the part and the time instance of failure occurrence. If a time gap is within three time units prior to failure, that time gap is indicated with T_H. If not, it is indicated with T_L. Second, time gaps for T_H and T_L are plotted in individual histograms. Then, the frequency of T_H is compared to that of T_L at each time section. As presented in Figure 5, the frequency of T_H is higher than that of T_L at the 3rd and 6th time sections indicated by the shade. In other words, those sections have a higher probability of failure than other sections.

A systematic approach is required to select time sections that have a higher probability of failure. The relative frequency of high risk at a time section, k, followed by the replacement of the part, p, is defined as:

$$R^T_{p,k} = N^T_{H,p,k} / (N^T_{H,p,k} + N^T_{L,p,k}) \qquad (2)$$

where $N^T_{H,p,k}$ and $N^T_{L,p,k}$ are the frequencies of T_H and T_L in the range of the kth section, respectively. As discussed in Section 3.1, the probability of high risk at a randomly-chosen time instance is 0.027. The time section which has a higher relative frequency than the threshold (i.e., 0.027) is assigned to be high risk. Otherwise, low risk is assigned.

As an example, the time gaps between part replacement events and high-risk (or low-risk) occurrence were calculated using the part 76398. The relative frequency of high risk after the replacement of the part 76398 is illustrated in Figure 6. The time section that has larger relative risk than this threshold is shown shaded in black.

We had to choose the most dominant parts related to failure for risk assessment. The selection rule was determined as follows: (1) if the probability mass function (pmf) of $R^T_{p,k}$ *larger than the threshold* is significantly different from the pmf of $R^T_{p,k}$ *smaller than the threshold*, it was determined
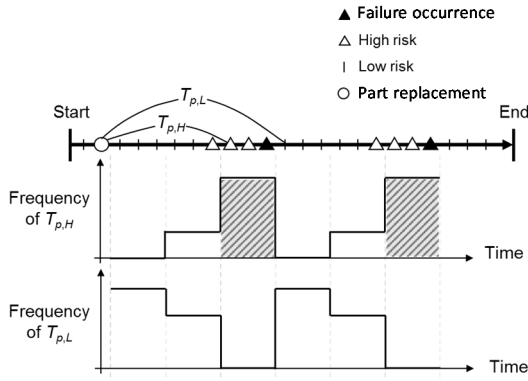

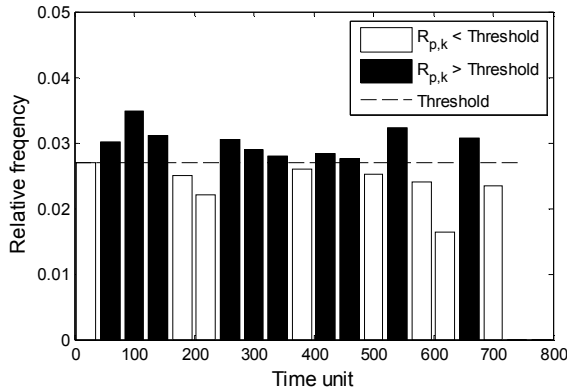Figure 5. Illustration of the concept of the method based on part lifespan calculation

Figure 6. Relative frequency of high risk after the replacement of part 76398; R76398

that the part has a clear trend for replacement over time. (2) If Rule 1 is not true, the part fails randomly regardless of the time instance.

We expected that the prediction score would increase as more parts were employed in this method. We chose the top 10 out of a few thousand parts that had a large gap between the pmfs. The top 10 parts are shown in Table 3. Employing more parts was possible. However, it was found that considering more than 10 parts did not help increase the accuracy of risk prediction.

Part number	224501, 318604, 374736, 501319, 632374, 669023, 692017, 711803, 804669, 844732

Table 3. Top 10 parts selected in Method 1

4.2. Method 2: Usage classification

Here, we propose a method based on usage classification since the finding in Section 3.3 showed a strong linear relationship between the number of failures and part usage. A strategy for risk prediction of the assets is herein devised: the relative frequency of high risk in a particular usage range is calculated; if the usage range has a higher relative frequency than the threshold of 0.027, the instance from this range is assigned to high risk. To be more specific, this usage range was decided using the following steps.

Step 1: The usage values of all time units in the training data (one to 730) from all the assets were calculated. It was observed that the usage information was not always available since they were recorded intermittently. This difficulty was overcome by extrapolating (or interpolating) the existing usage data.

Step 2: The usage values obtained from Step 1 were equally divided into n sections. In the example, this usage data ranged from 22,257 to 37,272 units was divided into 20 sections. Each section had a range of 750 units. The relative frequencies of high risk of R^U_k were calculated for the individual sections:

$$R^U_k = N^U_{H,k} / (N^U_{H,k} + N^U_{L,k}) \qquad (3)$$

where $N^U_{H,k}$ and $N^U_{L,k}$ are the frequencies of high and low risk that belong to the usage section k, respectively.

Step 3: Find the section k where R^U_k is greater than the threshold.

The relative frequencies of the sections are shown in Figure 7. The threshold for assessing risk of the test assets was determined to be the relative frequency of the section with the mean usage of 29,800 units or more. The ten sections that correspond to the usage of 29,800 to 38,000 units have higher relative frequency values than the threshold. Therefore, high
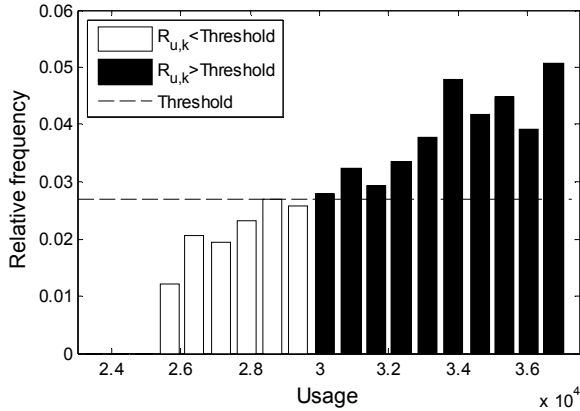
5

Figure 7. Risk assessment based on usage classification

Table 4. Ensemble method using ratio of high-risk frequency to low-risk frequency

Part lifespan classification	Usage classification	Relative frequency of high risk	Final submission
High risk	High risk	0.0364	High risk
High risk	Low risk	0.0218	High risk
Low risk	High risk	0.0315	High risk
Low risk	Low risk	0.0153	Low risk

risk was assigned to those sections. As the relative frequencies of the sections from zero to 29,800 units were observed to be lower than the threshold, the risk of those sections was assigned to be low. It is worth noting that the risk at the instances of the last usage record of the assets was assigned to be low. This assignment was based on analysis of the training data that showed the absence of usage records was correlated with infrequent failure of assets.

4.3. Ensemble approach

An ensemble approach, which combines the part lifespan calculation and the usage classification methods, is presented in this section. The ensemble approach was developed in the machine learning community to improve the accuracy and robustness of individual methods. The ensemble of different methods is accomplished based on the consensus or learning strategy (Hu, Youn, & Wang, 2012). More details are available in Gao, Fan, and Han (2010). In this paper, we adopted the consensus strategy since it is more appropriate for use when combining the outcomes from a limited number of available methods.

Table 4 displays four combinations using the outcomes from the two methods. Ten parts were selected in the part lifespan calculation, while the usage classification method adopted a classification boundary of 29,800 units. In each case, the relative frequency of high risk was calculated. It is reasonable

to assign high (or low) risk if both of the outcomes are larger (or smaller) than a predefined threshold. Therefore, the combination of "high risk" and "high risk" (or "low risk" and "low risk") is assigned to have high (or low) risk. For the two cases where the outputs from the two methods were in conflict, we assigned high risk if either of the two methods determined high risk since the performance evaluation gave the best score from that determination.

5. CONCLUSION

Two features were identified from statistical analysis of the maintenance record database. Those features led us to propose two distinct methods for predicting imminent failure of engineering assets: a method based on part lifespan calculation and a method based on usage classification. Finally, an ensemble of the two methods was developed based on the scheme of "consensus." The ensemble method achieved higher accuracy for risk prediction compared to either single method.

The ensemble method scored the second highest rank in the 2014 PHM Data Challenge Competition. Yet, there is still room for improvement. The score of 1.1413 achieved (out of the perfect score, 2.0000) may not be sufficient for real industrial applications in the field. We expect that additional methods can be developed if specific knowledge about the assets (e.g., asset structure, loading conditions, and physics-of-failure) is utilized. The ensemble method has flexibility that can incorporate additional methods for risk prediction. A "learning" strategy for the ensemble can be also considered when additional methods become available. Considering the performance, it is expected that the ensemble method can help reduce maintenance costs and increase availability by offering decision makers accurate risk prediction for engineering assets.

REFERENCES

Bazovsky, I. (2004). *Reliability Theory and Practice*. NY: Courier Dover Publications.

Braglia, M., Carmignani, G., Frosolini, M., & Zammori, F. (2012). Data classification and MTBF prediction with a multivariate analysis approach. *Reliability Engineering*

& System Safety, 97(1), 27-35. doi:10.1016/j.ress.2011.09.010

Crow, L. H. (1990). Evaluating the reliability of repairable systems. *IEEE Annual Reliability and Maintainability Symposium* (275–279), January 23-25, Los Angeles, CA. doi:10.1109/ARMS.1990.67969

Gao, J., Fan, W., & Han, J. (2010). On the power of ensemble: Supervised and unsupervised methods reconciled, *Tutorial on SIAMD at a Mining Conference*, Columbus, OH, 2010

Heng, A., Zhang, S., Tan, A. C., & Mathew, J. (2009). Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*, 23(3), 724–739. doi:10.1016/j.ymssp.2008.06.009

Hjartarson, T., & Otal, S. (2006). Predicting future asset condition based on current health index and maintenance level. *IEEE 11th Conference on Transmission & Distribution Construction, Operation and Live-Line Maintenance*, October 15-19, Albuquerque, NM. doi:10.1109/TDCLL M.2006.340747

Hu, C., Youn, B. D., & Wang, P. F. (2012). Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life, *Reliability Engineering & System Safety*, 103, 120–135. doi:10.1016/j.ress.2012.03.008

Jahromi, A., Piercy, R., Cress, S., Service, J., & Fan, W. (2009). An approach to power transformer asset management using health index. *IEEE Electrical Insulation Magazine*, 25(2), 20–34. doi:10.1109/MEI.2009.4802595

Lawless, J., Hu, J., & Cao, J. (1995). Methods for the estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Analysis*, 1(3), 227-240. doi:10.1007/BF00985758

Louit, D. M., Pascual, R., & Jardine, A. K. S. (2009). A practical procedure for the selection of time-to-failure models based on the assessment of trends in maintenance data. *Reliability Engineering & System Safety*, 94(10), 1618-1628. doi:10.1016/j.ress.2009.04.001

Taghipour, S., Banjevic, D., & Jardine, A. K. S. (2011). Reliability analysis of maintenance data for complex medical devices. *Quality and Reliability Engineering International*, 27(1), 71-84. doi:10.1002/qre.1084

Zhou, Y. (2011). The auto regression model of bus fleet failure number. *International Journal of Reliability and Applications*, 12(2), 95-102

BIOGRAPHIES

Hyunjae Kim received his B.S. degree in Biosystem Engineering from Seoul National University (SNU), Seoul, Korea, in 2012. He is a Ph.D. student at the Laboratory for System Health & Risk Management in SNU. His research interests involve battery thermal management and battery power management.

Taewan Hwang received the B.S degree from Seoul National University (SNU), Seoul, Korea, in 2013. He is a M.S. student at the Department of Mechanical and Aerospace Engineering in SNU. His current research area is prognostics and health management (PHM).

Jungho Park received his B.S. degree from Seoul National University (SNU), Seoul, Republic of Korea, in 2012. He is a Ph.D. student at the Department of Mechanical and Aerospace Engineering in SNU. His research area includes fault diagnostics of planetary gears and model verification and validation (V&V).

Hyunseok Oh is 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. He is a Research Professor in the Laboratory for System Health & Risk Management, Seoul National University. His current research area includes prognostics and health management and model verification and validation. He was with the Hyundai MOBIS Technical Research Institute as a Research Engineer from 2006 to 2007. He was a Postdoctoral Research Associate at the Center for Advanced Life Cycle Engineering (CALCE), University of Maryland from 2012 to 2014. Dr. Oh received the A. James Clark Fellowship (2007). He was recognized as winner of the IEEE PHM Data Challenge Competition (2012).

Byeng D. Youn received the B.S. degree from Inha University, Incheon, South Korea, in 1996, the M.S. degree from KAIST, Daejeon, Republic of Korea, in 1998, and the Ph.D. degree from the University of Iowa, Iowa City, IA, USA, in 2001. He is an Associate Professor of mechanical and aerospace engineering at Seoul National University (SNU), Seoul, Republic of Korea. Before joining SNU, he was an Assistant Professor in the Department of Mechanical Engineering, University of Maryland, College Park. His research goal is to develop rational reliability and design methods based on mathematics, physics, and statistics for use in complex engineered systems, mainly focused on energy systems. His current research includes reliability-based design, prognostics and health management (PHM), energy harvester design, and virtual product testing. Dr. Youn's dedication and efforts in research have garnered substantive peer recognition resulting in four notable awards including the ASME IDETC Best Paper Awards (2001 and 2008), the ISSMO/Springer Prize for a Young Scientist (2005), the IEEE PHM Competition Winner (2014), etc.