

# An Effective Predictive Maintenance Approach based on Historical Maintenance Data using a Probabilistic Risk Assessment: PHM14 Data Challenge

Seyed Mohammad Rezvanizani, Jacob Dempsey, Jay Lee

*NSF I/UCRC for Intelligent Maintenance Systems, University of Cincinnati, Cincinnati, OH 45221, USA*

*rezvansd@uc.mail.edu*

*dempsejb@uc.mail.edu*

*jay.lee@uc.edu*

## ABSTRACT

This paper presents an effective health assessment and predictive maintenance technique for industrial assets. The technique and algorithms applied to data sets provided by the Prognostics and Health Management Society 2014 Data Challenge. The data contains usage and part consumption for three years. In short, the usage data contains a parameter that roughly measures asset usage, and the part consumption data includes information regarding part replacement and maintenance actions. The first two years of data are considered as "Training" data and the third year is considered as "Testing" data. The proposed method built on the probability of the failure risk during training dataset. The main objective is to develop a model based on first two years data set (training) and determine the high risk and low risk times of failure for each individual asset for the third year.

Training data shows many maintenance activities with 14 different codes. The principle difficulty is to detect the Preventive Maintenance (PM) in the training data. The paper presents the method in three main steps: the first step is to recognize the PM pattern based on time and type of maintenance activity via the training data. The second step is to determine the high-risk time intervals based on PM times by checking the frequency of the failures at specific times between each PM. The third step is to predict the high risk time intervals in the testing data using the information acquired from the training data. The score predicted by this probabilistic risk assessment method won the first place in the PHM Data Challenge Competition.

---

Rezvanizani et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## 1. INTRODUCTION

Failures occur as materials and machines are continuously used. When a small failure occurs due to a subsystem, an entire system may fail and require maintenance. This may cause an extended downtime to occur when an unpredicted failure happens; however, this can be avoided by performing maintenance tasks and part replacement before failure occurs (Endrenyi et al, 2001). Failure data plays an important role in optimization of the maintenance schedules for machinery. Moreover, each maintenance activity contains abundant raw data that shows some aspects of the asset health. Appropriate analysis on assets failure and maintenance data can help to improve the reliability of the engineering assets.

Classical maintenance practices can be divided into two main categories a) Preventive Maintenance and b) Corrective Maintenance (Arunraj & Maiti, 2007). Preventive Maintenance (PM) is a periodical planned downtime, in which a clear set of action items such as part replacement, lubrication and inspection is performed (Ebeling, 1997). Corrective Maintenance (CM) is performed after failure occurs to restore the equipment or machine to the operational condition in which it can perform its planned function (Ebeling, 1997). In recent years, risk evaluation of system or machine failure, referred to as risk-based maintenance, presents a more practical modeling of equipment failure and estimation of expected consequences (Khan & Haddara, 2003). In a work proposed by Lin, Ming, and Richard, (2001) the analysis and optimization of PMs is studied for repairable machines with two failure modes classes, namely the maintainable and non-maintainable failure modes. Only the system failure rate matching to the maintainable failure mode is changed whenever PM actions are executed. A similar work to that presented by El-Ferik and Ben-Daya (2006) to build up degradation based fusion

model for PM for repairable machines. In (El-Ferik & Ben-Daya, 2006), an analysis is performed to establish the existence and distinctiveness of optimal PM strategy. Statistical modeling is the main approach that has been applied in the literature.

The NSF I/UCRC on Intelligent Maintenance Systems (IMS) mission is to develop intelligent techniques to improve reliability, productivity, and asset utilization. In support of this vision, the IMS Center at the Univ. of Cincinnati participated in the Data Challenge competition that is part of the annual conference of the Prognostics and Health Management Society held in 2014. This year's data challenge focused on asset health calculation, an industry problem found in remote monitoring and diagnostics. The paper is prearranged in the following sections: after the introduction, Section 2 discusses the problem statement for the 2014 Prognostics and Health Management Society Data Challenge. Section 3 describes the general methodology of risk assessment based on maintenance activities. This is followed in section 4, where the model development as well as feature extraction method is explained. Section 5 describes the prediction algorithm to forecast the high risk and low risk time intervals for each asset during the testing period. Section 6 presents an evaluation metric for this data challenge and measures the performance of predicted results. Lastly, conclusions and future works are discussed in Section 7 and Section 8 respectively.

## 2. PROBLEM STATEMENT AND CHALLENGES

The objective of the data challenge is to design an algorithm that would evaluate whether or not a system would fail in the next three days after a given sample time. The dataset consisted of three training datasets and two test datasets. The training data sets included part consumption, usage and failures. The test data sets included part consumption and usage. The part consumption data contained a record of what parts were replaced on an asset, the time that the part were replaced, the reason the asset was being worked on, and the quantity of parts that were replaced. For example, the structure of a small portion of the part consumption data for asset ID 18997 is shown in Table 1.

Asset ID	Time	Reason code	Part Number	Number of Parts
18997	727	707	133495	1
18997	727	707	133495	1
18997	727	707	133495	1
18997	727	707	396485	11
18997	727	707	396485	1
18997	728	707	2198	1
18997	728	707	366285	3
18997	729	707	201954	1

Table 1. Sample of part consumption data structure

The usage file contained a record of a parameter that roughly measures asset usage and the time that measurement was taken. The main objective was to determine if the equipment was going to fail within 3 time units using certain times within the testing data (Garvey, 2014).

### 2.1. Data description

The range of time in the entire training data set (part consumption, usage, and failure) is between 0 and 730 and for testing, is between 731 and 1095. Based on this information we assumed there were 3 years of data, the first two years were training and the third year for testing. To better understand the training data set Figure 1 shows a sample of maintenance tasks for asset ID 6787. The 'x' axis shows the time in days for three years. The blue dashed line separates the testing part from the training part at the end of the second year.

The 'y' axis shows the repeated number of part numbers replaced (RNPNR) in the maintenance time period with green color indicating replacement during training time and yellow color indicating replacement during testing time. The RNPNR means how many times part numbers have been repeated at each maintenance time in the part consumption regardless of the unique number of part numbers. For example, in Table 1 for the asset at time 727 the RNPNR would be 5 and at time 728 it would be 2.

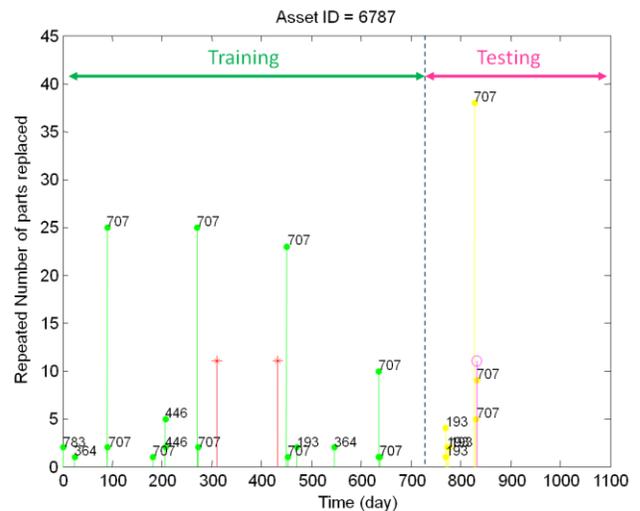


Figure 1. Example of maintenance tasks

The digit at each maintenance action shows the reason that asset was worked on (column 3 of Table 1). An asset would be worked on for 14 different reason categories. These categories have been coded with two or three digit numbers. Table 2 shows the list of all reasons. In Figure 1, the failure time is depicted with a red asterisk during the training. 'Test time', showing with pink circle, is the time during the testing period that has been asked to identify if there is any failure

in the next 3 days (time units). Even if there are some maintenance tasks after the test time, the method needs to define the risk level of the test time just by looking at data before 'Test time'.

No	Code	No	Code
1	44	8	446
2	64	9	565
3	119	10	575
4	193	11	606
5	364	12	707
6	396	13	782
7	417	14	783

Table 2. 14 reasons that maintenance performed on assets

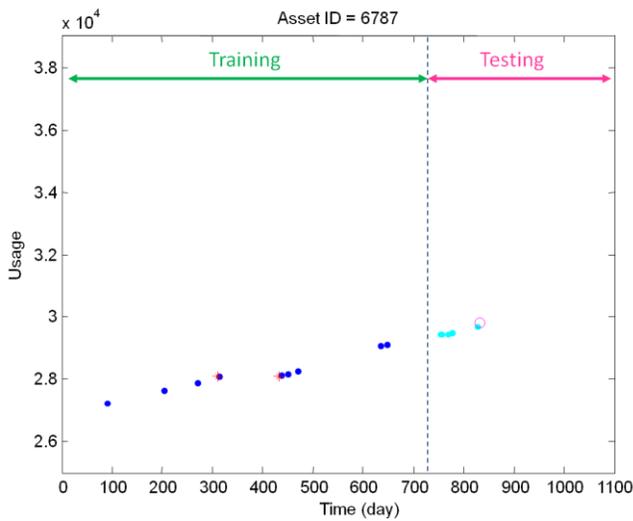


Figure 2. Example of usage data

Figure 2 illustrates the usage in training (dark blue) and testing part (light blue). Red asterisks have indicated the time when the two failures occurred and the test time is marked by a pink circle. The unit of usage is unknown; however, the PHM 2014 mentioned it is similar to an odometer. The usage measurement is irregular and not same as the time the maintenance action takes place. Moreover, the amount the usage is increasing at each time intervals is not linear. This can be interpreted that the asset might be in rest, downtime, or standby for several days or months. The range of the usage for all assets is between 25000 and 39000.

## 2.2. Challenges and Pre-processing

The subsequent processing steps in the risk assessment algorithms for this data set depend on quality data inputs. For all inputs, data filtering is performed to remove erroneous data samples and provide a more suitable data set for further processing.

The first issue is that the unique number of asset IDs in "Test Instances" are 1866 that is needed to identify at least 1 test time; however, just 1738 of these assets has "Train Part Consumption" information. Therefore there are 128 unique assets IDs with no train part information. Moreover, the training data can be varied from complete two years up to just several days at the end of the second year. Other assets, which are not being considered in the testing data, are not included in this analysis.

The second issue is the number of parts replaced for each part number. Logically the number of parts should be an integer positive value. Consequently, after sorting the Train number of part replacements in ascending order, those with values lower than one has been removed from the analysis.

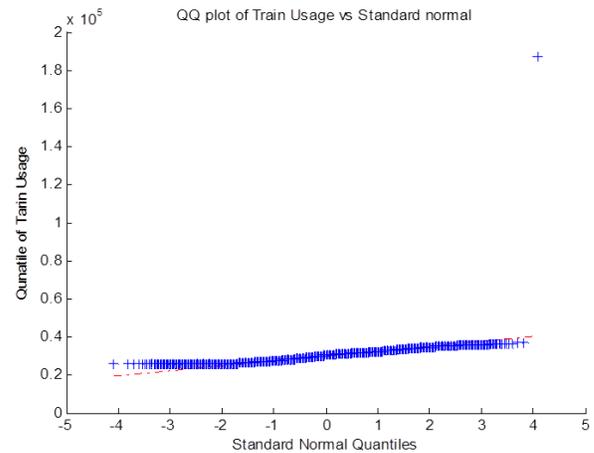


Figure 3. Q-Q plot on usage to remove outlier and define usage distribution

The third challenge is outlier detection and removal. In this paper, the Grubb's test is used to detect outliers before any further analysis occurs. Figure 3 illustrates an example of outlier detected by Grubb's test and verified by the Q-Q plot. The Q-Q plot is a graphical method for comparing usage distributions and normal distribution by plotting their quintiles against each other. The graph clearly detects an outlier in the usage and removing the outlier declares the distribution of usage is close to normal.

The performance of a risk prediction method for the test cases evaluates based on the Eq. (1). The PHM Data Challenge Competition gives the following scoring metric:

$$score = \left(\frac{L}{N}\right) + \left(\frac{H}{N}\right) \quad (1)$$

where N is the number of samples for each type "low risk" and "high risk"; so the number of samples from "low risk" and "high risk" are equal. L is the number of correct low risk samples and H is the number of correct high risk samples.

### 3. RISK ASSESSMENT METHODOLOGY

The overall approach for assessing the low risk or high risk time intervals for this specific case study consists of a series of algorithmic processing steps. A flow chart of the health assessment algorithm for this study is provided in Figure 4. These steps include data pre-processing, feature extraction, risk assessment in training data and risk prediction in testing data. The risk assessment algorithms developed for this data set follow that step by step processing methodology.

Section 3.1 presents an overview of the classical maintenance risk assessment based on Mean Time between Repair (MTBR) and bathtub curve. Before our method can be explained, it is important to have a solid foundation of these concepts. Section 3.2 explains how this classical method can be used in this specific case study to estimate the high risk time intervals for each asset. More specific details of each processing module along with intermediate results from each step are shown in the subsequent sections to further illustrate this risk assessment method.

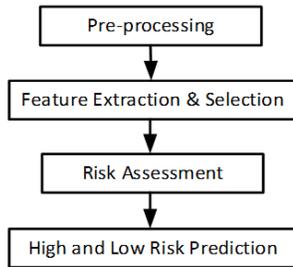


Figure 4. Risk assessment algorithm flowchart

#### 3.1. Preventive maintenance background

Usually PMs are designed based on time or usage to reduce the risk of failure in the time ahead. In other words, maintenance tasks are based on elapsed time or days of operation that are founded on statistical or historical data for specific types of machine or system. Figure 5 a) illustrates an example of the statistical life of a system or equipment. The Mean Time Between Failure (MTBF) or bathtub curve indicates that the probability of failure during the first few time units of operation is high for new equipment, typically caused by manufacturing or installation problems. The first period, called infant mortality, is characterized by a decreasing failure rate (Wang, Hsua & Liub, 2002). Following this first period, the chance of failure is low for an extended period. Following this regular system life period, the probability of failure increases promptly with elapsed time of operation. In preventive maintenance management, machine inspections, replacement, repairs or rebuilds are scheduled based on the MTBF statistic (Klutke, Kiessler & Wortman, 2003).

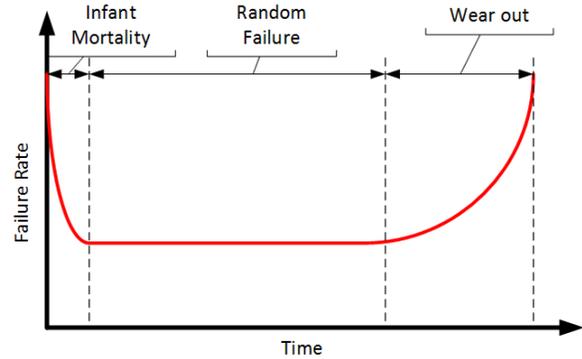


Figure 5 a). General Bathtub curve of industrial machines

#### 3.2. Proposed methodology

Dealing with complex reparable systems, PMs usually are executed on the wear out section to reduce hazard rate. Such PM operations when performed bring the system to operate close to new state (Rezvanianiani, Barabady, Valibeigloo, Asghari & Kumar, 2009). The next PM will be performed when the hazard rate reaches the maximum acceptable level. Figure 5 b) illustrates the location of PMs and their affect on decreasing hazard rate. The time between two PMs referred is time between repairs (TBR). In the proposed method it is important to identify, the PM exact time during maintenance activates to define the value of TBR. In this case, the probability of failure before TBR is higher because of a larger hazard rate.

On the other hand, after each maintenance activity there is a short period of high hazard rate same as the infant mortality in the Figure 5 b). These maintenance actions can be preventive maintenance with large number of parts replaced or even corrective maintenance (CM) with a few parts replacement. The main difference is that for PMs the higher failure rate is before the PM occurs, however, for CMs it is after that. The proposed methodology uses both techniques and builds a model that infers high risk before the PMs and after CMs. Section 4 elaborates these techniques in details.

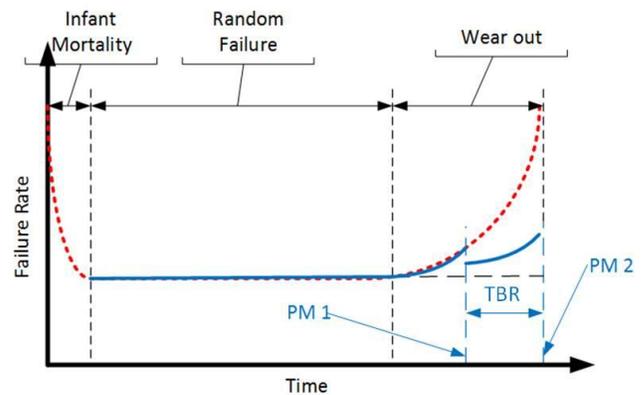


Figure 5 b). PM affect on decreasing hazard rate

4. MODEL DEVELOPMENT

4.1. Feature extraction & selection

Referring to the bathtub curve in Figures 5 a) and 5 b), the higher probability of failure occurs during the wear out section where the PMs should be performed. Therefore, the time PM are scheduled can play an important role in defining the high risk failure for each asset. To deal with this problem the three features of part consumption including time of part replacement, reason the asset was worked on and number of parts replaced are used.

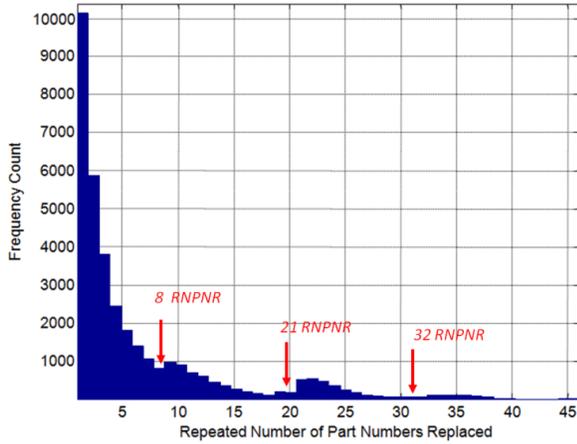


Figure 6. Histogram of repeated part number replaced

Figure 6 shows the histogram of repeated number of part numbers replaced (RPNR) at each maintenance time. The frequency decreases when the quantity of repeated part numbers increases. A jump of frequency occurs in three locations. The first jump happens at 8 RPNR and the second and third happens at 21 and 31 respectively. These higher quantity part usages can determine specific maintenance tasks. To identify these tasks, the maintenance tasks with lower than 8 RPNR have been removed. For each asset, the time between each repair was calculated. Figure 7 a) illustrates the measurement of time between Repair for one sample. Figure 7 b) demonstrates the histogram distribution of time between repair for all assets with higher than 8 repeated part number replacements.

The figure 7 b) shows very important information about maintenance activities. There is a peak at TBR near to 183. It means that the maintenance activities, which take place after 183 days of the previous maintenance, has higher occurrence. The distribution of the peak seems symmetric usually the planned maintenance activities need replace lots of parts and they might take several days so that can be a reason of symmetric shape of the peak. The peaks has been shown in the green rounded rectangular and it will be zoomed later in Figure 9.

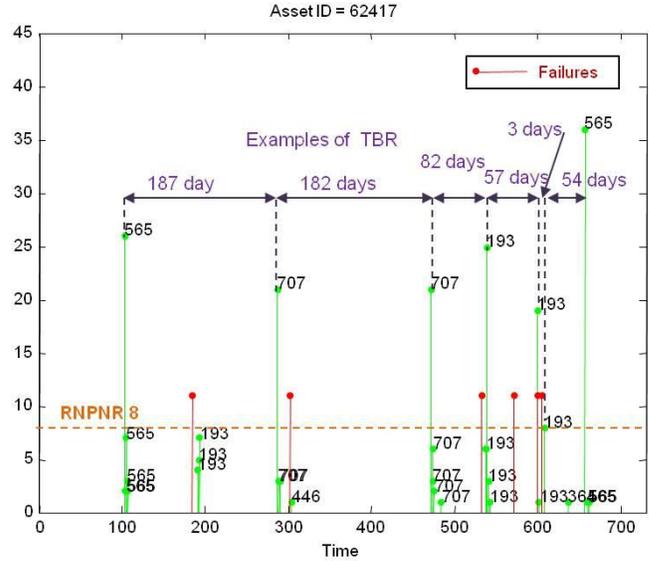


Figure 7 a) Measurement of Time between Repair for one asset ID with RPNR greater than 8

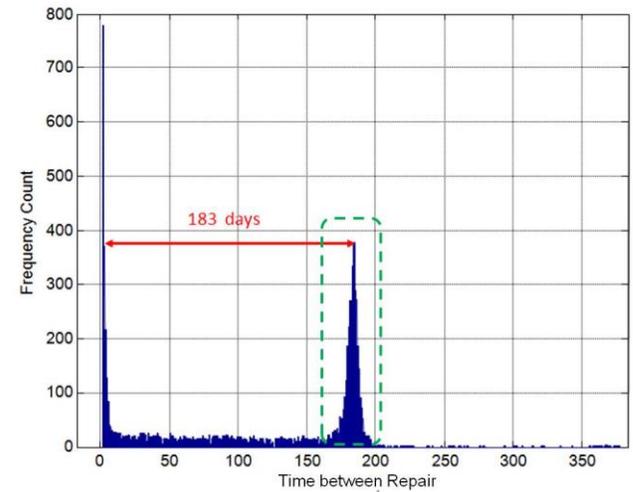


Figure 7 b) Time between Repair histogram for more than 8 RPNR

By increasing the RPNR threshold from 8 to 21 some TBRs are merged together and create larger TBR. This fact has shown in Figure 8 a) match up to Figure 7 a). The histogram graphs can help us to find where the maximum frequency of maintenance tasks takes place and what the maximum TBR is. The same analysis can be performed for repeated parts more than 31 as shown on Figure 8 b). Although the numbers of maintenance actions with more than 21 repeated part numbers are very low, the mode of distribution happens at approximately every 183 days.

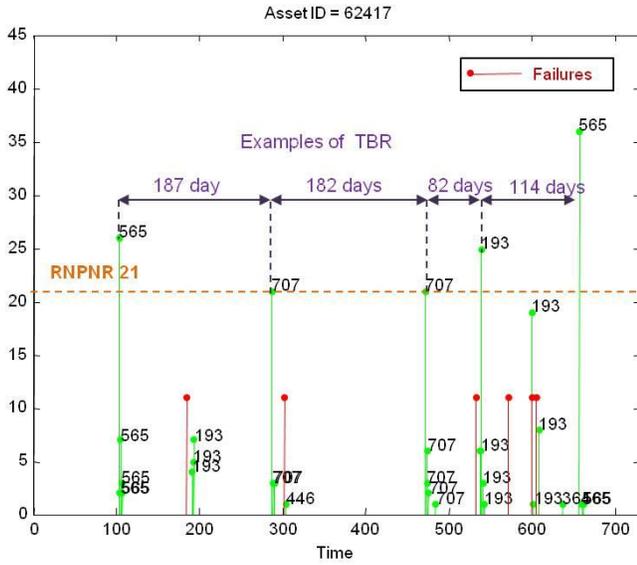


Figure 8 a) Measurement of Time between Repair for one asset ID with RNPNR greater than 21

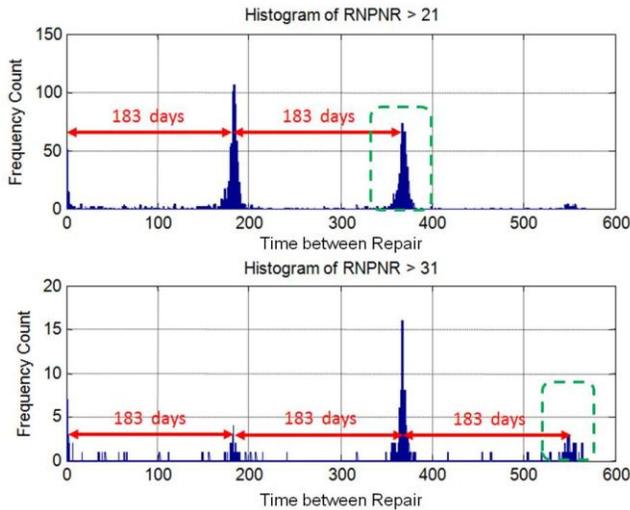


Figure 8 b) Time between Repair histogram for RNPNR higher than 21 and higher than 31

Filtering the RNPNR shows that with higher than 31 replacements it is possible to have 550 days in between two repair times; however, the statistical mode of these time between repairs just happens at 183 days. In Figures 7 b) and 8 b), the mode of Time between Repair distribution, shows approximate estimation of PM location. Even the graphs can give approximate understanding of TBR. However, to identify the PM locations, deep understanding of PM distributions is necessary. To better understand the distribution of time between repair in figures 7 b) and 8 b), the last mode (peak) of the graph, which has been highlighted with the green rounded rectangular, has been zoomed separately in Fig. 9. Figure 9 shows the distribution of time between repair for the part replacement greater than

8, 21 and 31 respectively. The distribution is considered as normal and the mean and standard deviation is shown in Table 3.

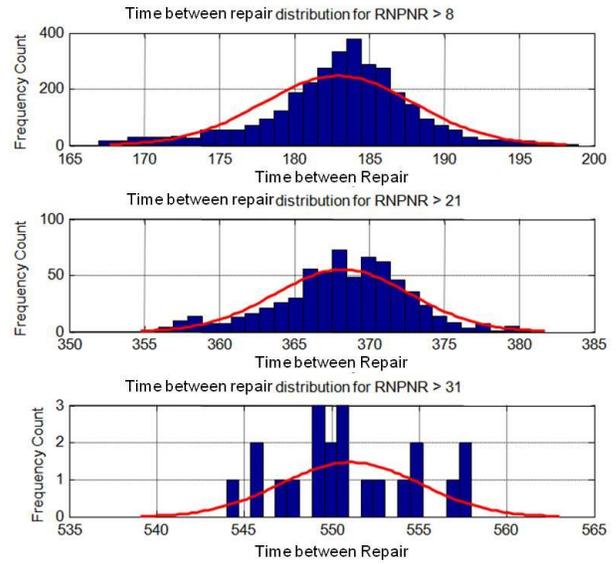


Figure 9. Distribution of time between repair for RNPNR greater than 8, 21 and 31

RNPNR	mean	std
>8	182.94	5.07
>21	368.26	4.50
>31	551.09	3.99

Table 3. Distribution characteristics of three distributions

Equation (2) shows the tolerance interval for the normal distribution:

$$F(\mu + n\sigma) - F(\mu - n\sigma) = \operatorname{erf}\left(\frac{n}{\sqrt{2}}\right) \quad (2)$$

where  $\mu$  is the mean and  $\sigma$  is the standard deviation and  $\operatorname{erf}$  is the error function, which is twice the integral of the Gaussian distribution with 0 mean and variance of 0.5. Equation (3) shows the error function:

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt \quad (3)$$

Based on these equations the probability of PM located at the range of  $\mu \pm \sigma$  is 68.2%.

#### 4.2. Data-driven approach for PM detection

The crucial step in this analysis is to define the location of PMs based on the extracted features. According to the above analysis, the PM features can be summarized as:

1. RNPNR greater than 8
2. The MTBR is 183 days +/- 5 days

Based on the above conditions, the flowchart in Figure 10 has been designed to detect the PMs during two years training data set.

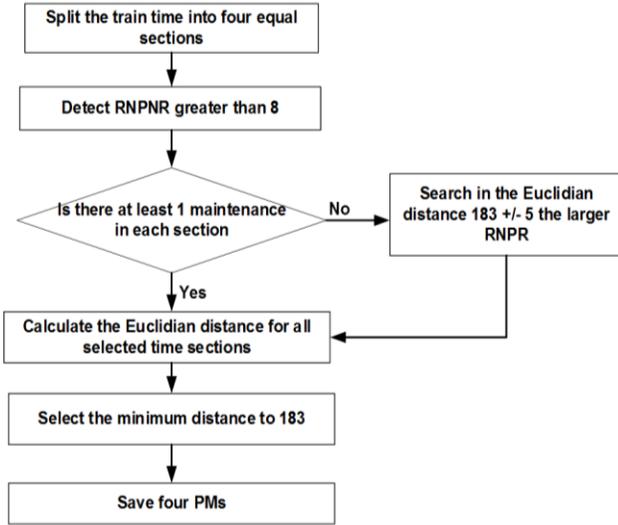


Figure 10. Flow chart of "PM detection algorithm"

The algorithm at first divide the training part time to four equal sections and defines all maintenance activities with RNPNR greater than 8. Then, the Euclidian distance matrix for all detected points calculated based on Eq. (4):

$$d_{ij} = \|x_i - x_j\|_2^2 \quad (4)$$

where  $x$  is the maintenance time and  $i$  or  $j = 1, \dots, N$ , individually addresses all the maintenance detected. To find the closest maintenance time to PM all  $d_{ij}$  subtracted from Mean Time between Repair (MTBR):

$$D_{ij} = |d_{ij} - MTBR| \quad (5)$$

The minimum  $D_{ij}$  value within each section is considered as the PM location. If there is not at least one maintenance action with greater than 8 RNPNR in a section, then the algorithm detects PMs in other sections to estimate the time of missing PM. This is done by finding the range of  $MTBR \pm \sigma$  in the section without a detected PM. Within the newly found range, the maintenance with largest RNPNR is considered as a PM. If a PM is still not found, the range is slowly increased to  $2\sigma$  until maintenance falls within this range. Otherwise, we estimate where a PM would fall and create a simulated PM at that location. For those assets, which their training part is less than two years, the PMs are less than four. Table 4 shows the summary of PM detection for all 14 reasons. Totally, 5995 PMs were detected in which 4502 of them are with the reason 707. More than 97% of the maintenances classified as PMs have a reason code of 417, 565 or 707; therefore, we classified those reasons as PMs and considered the other reasons for maintenances as CMs.

Reasons	Number of PMs	Maintenance type
44	0	CM
64	1	CM
119	25	CM
193	73	CM
364	21	CM
396	1	CM
417	338	PM
446	38	CM
565	985	PM
575	3	CM
606	5	CM
707	4502	PM
782	0	CM
783	3	CM
<b>Total</b>	<b>5995</b>	

Table 4. Results of PM detection

The set of Figures numbered 11 illustrate the four examples of PM detection using this technique. Each figure selected to discuss some aspects of PM detection. The blue triangles show the time of PM, which labeled in order from  $PM_1$  to  $PM_4$ . The orange dashed line is the threshold for 8 RNPNR. Figure 11 a) shows the correct performance of the algorithm since all PMs have a distance close to MTBR from each other. Close to the time of  $PM_1$  there are several other maintenance activities with same reason which represent the PM tasks might take several days.

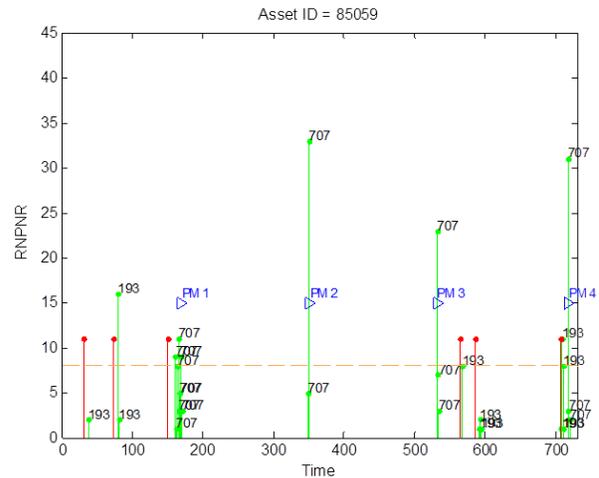


Figure 11 a) Example of PM detection

Figure 11 b) is an example that "PM detection algorithm" is robust enough to distinguish the PM action event he RNPNR is lower than 8. Therefore, the algorithm at first selects  $PM_2$  to  $PM_4$  and then estimates the location of  $PM_1$  based on  $MTBR \pm \sigma$  with RNPNR lower than 8 and searches for a maintenance task with higher RNPNR. The selected  $PM_1$  has a code of 707, which verifies based on the results of Table 4. Figure 11 c) shows one of the samples which does not have a complete two years of data. In this

case, the "PM detection algorithm" assign zero value to PM<sub>1</sub> and PM<sub>2</sub> and the other PMs are detected properly.

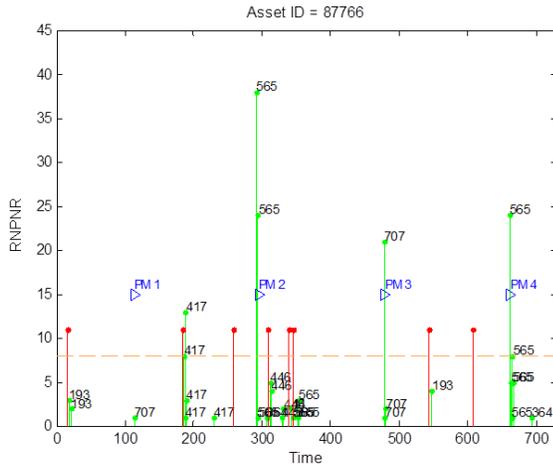


Figure 11 b) Example of PM detection

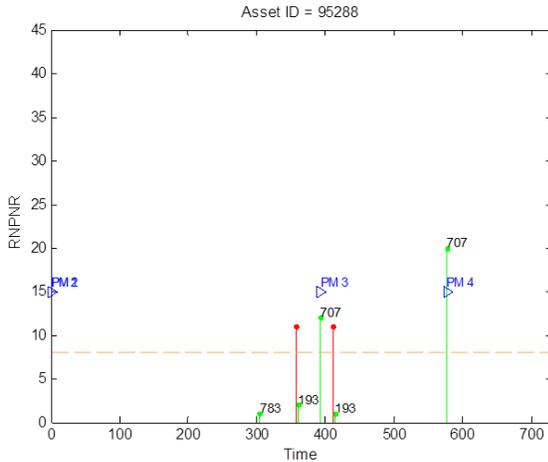


Figure 11 c) Example of PM detection

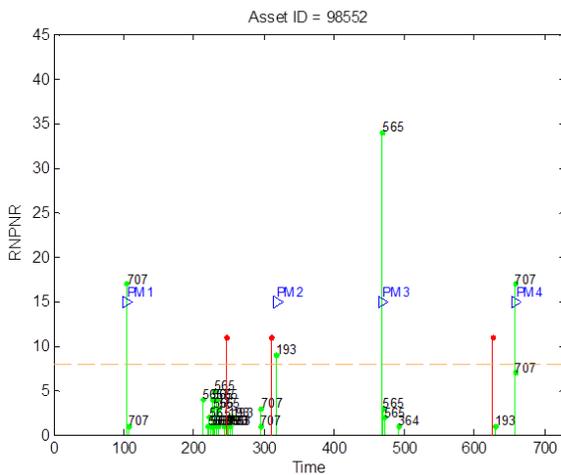


Figure 11 d) Example of PM detection

The most significant issue is observed in Figure 11 d) at PM<sub>2</sub> where the maintenance with reason 707 has RPNR lower than 8, but there is a very close CM with reason 193 that the algorithm detects as the PM.

### 4.3. Probabilistic risk assessment

There are two approaches to define the high risk time intervals in this algorithm. The first approach uses PMs as the main feature and detects the probability of failure between two PMs. The second approach uses CMs to evaluate the correlation between CM reasons and the probability of failure in the first few days after the CM is performed. In other words, these two approaches can be summarized as follows:

- 1- Building a model that infers risk based on Preventive Maintenance
- 2- Building a model that infers risk by looking for high risk Corrective Maintenance

The first approach is to define the failure rate during each MTBR. For this approach, the amount of failures in the four MTBR sections for all assets has been calculated and added together. The histogram of failure rate is shown on Figure 12 where each bar represents the accumulation of the number of failures in a three day time interval.

In addition to having a failure rate, it is also necessary to have a corresponding threshold for alerting high risk time intervals (Siegel & Lee, 2011). A more appropriate way of handling this aspect for threshold setting is to use Chebyshev's inequality since that relationship is appropriate for any distribution. For an unknown distribution, Chebyshev's inequality is given by the following expression provided in Eq. (6), in which  $\mu$  is the mean of the failures at each bar time,  $\sigma$  is the standard deviation, and  $k$  is a parameter that is directly related to the false alarm rate (Siegel, 2013). Effectively, the false alarm rate is given by  $\frac{1}{k^2}$ , in which a large value of  $k$  would mean a lower false alarm rate and a threshold that is more standard deviations away from the mean.

$$\Pr (|X - \mu| \geq k\sigma) \leq \frac{1}{k^2} \quad (6)$$

In this study, a threshold that is approximately 1.5 standard deviations greater than the mean of failure rate was selected. Therefore, the threshold set to 160 and indicates three high risk time intervals that are highlighted on Figure 12.

The graph shows that the probability of failure is lowest immediately after the PM and the probability of failure is the highest 8 days before the PM. Based on the classical bathtub curve, the highest failure rate is expected close to the next PM. However, the graph also shows some periodic high and low failure rates which might associate to a time of a year, environmental condition or other physical aspects of

the system. Besides the 8 days before the PM, the failures higher than the threshold take place approximately at 12 and 42 days after PM.

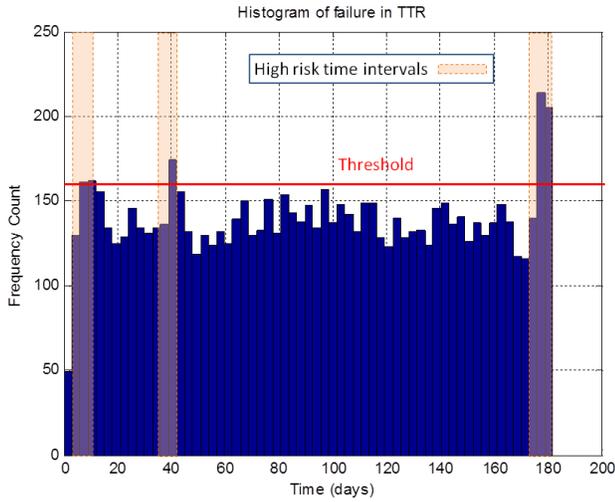


Figure 12. Histogram of failure rate in union of four TBR

The second approach is to evaluate the probability of failure after CMs. Table 5 shows the summary of all maintenance actions classified based on their reason. The second column shows total number of maintenance actions in the entire train data set for each reason. The third column represents the number of failures that occur after the maintenance reason before any other maintenance activities take place. Therefore, the rate of number of failures over total maintenance actions can represent the probability of failure for each reason.

Reasons	Total maintenance actions	Number of failures after maintenance	Pr (failure)	first quantile $\frac{\ln(\frac{4}{3})}{\lambda}$
44	172	86	50.00%	10.28
64	5	0	0.00%	NaN
119	567	129	22.75%	11.22
193	9578	2324	24.26%	9.58
364	2447	789	32.24%	11.29
396	2	1	50.00%	31.65
417	1370	271	19.78%	12.07
446	2927	646	22.07%	10.40
565	7263	864	11.90%	13.45
575	99	33	33.33%	10.47
606	69	10	14.49%	0.29
707	11399	2347	20.59%	16.75
782	0	0	0.00%	NaN
783	668	259	38.77%	10.64
<b>Total</b>	<b>36566</b>	<b>7759</b>		

Table 5. Summary of CMs and thresholds

Equation (7) represents probability of failure:

$$\Pr(\text{failure}) = \frac{\text{Number of failures}}{\text{Number of maintenance actions}} \quad (7)$$

Except for reason 396 that has a very low number of maintenance actions, there is a high probability of failure after a few reasons. However, this high probability needs to be checked based on the number of days after the maintenance action takes place and a failure occurs. Figure 13 shows the histogram of the failure after four high ranked failure CMs.

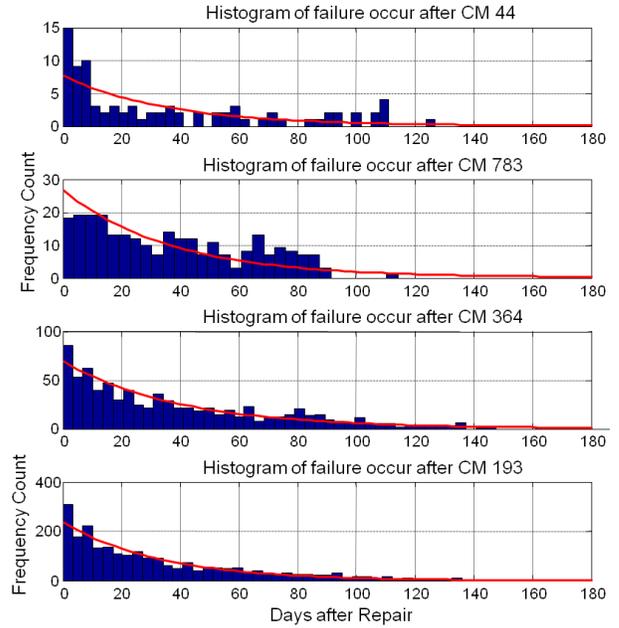


Figure 13. Histogram of failure after CMs

To define the best fitted distribution, the distributions are ordered by the Kolmogorov-Smirnov test statistic, and the best fitting distribution selected as exponential. Equation (8) shows probability distribution function for exponential distribution:

$$f(x) = N\omega\lambda e^{-\lambda x} \quad x \geq 0 \quad (8)$$

where  $N$  is the number of time to failure observations,  $\omega$  is the class interval width,  $\lambda$  is a hazard rate and  $\lambda^{-1}$  is the mean of distribution. Figure 14 illustrates the cumulative distribution function (CDF) of failures after CM 193. Using Tukey's criterion (Kotz & Seier, 2009) the CDF can be split into four parts where the quantiles can be reached based on Eqs (9) to (11). The first quantile represents 25% of failures and the second quantile (median) represent 50% of failures.

$$\text{first quantile} = \frac{\ln(\frac{4}{3})}{\lambda} \quad (9)$$

$$\text{median} = \frac{\ln(2)}{\lambda} \quad (10)$$

$$\text{third quantile} = \frac{\ln(3)}{\lambda} \quad (11)$$

The first quantile and the median calculated for all 14 reasons and represented in the Table 5. The first quantile is considered as the threshold for the high risk time interval after each CM reason.

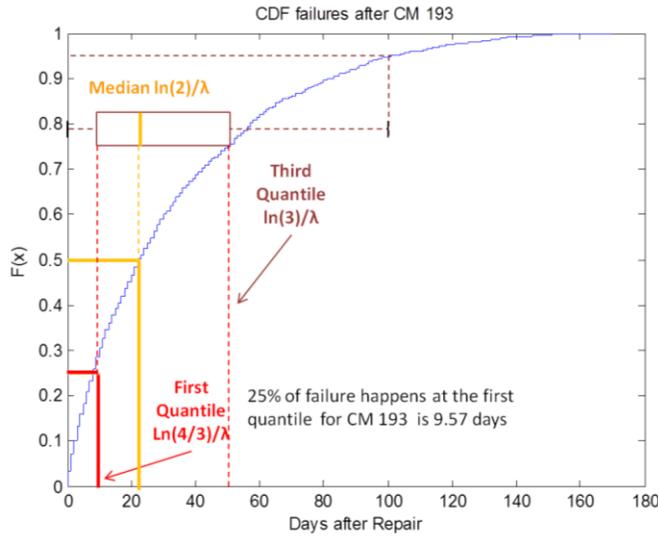


Figure 14. CDF for CM 193 and high risk indication based on Tukey criterion

## 5. PREDICTION OF HIGH FAILURE RATE

Similar to the risk assessment part, there are two approaches for prediction as well. The first approach is to define the high risk time intervals after the last PM. In this approach according to Figure 12 there are several high risk areas between two PMs. High risk test times should be detected 3 time units (days) before failure happens, so the thresholds is shifted 3 days before the high risk time intervals initiate. The three high risk time intervals highlighted by light orange color on Figure 12 and the start and end points summarized on the Table 6.

	$T_1$	$T_2$	$T_3$
$T_h$	[3 12]	[36 42]	[173 183]
$T_h \pm \sigma$	–	–	[168 188]

Table 6. High risk time intervals after last PM

According to the normal distribution of PM times, the mean value +/- the standard deviation is used to predict the next PM. Therefore, the  $T_3$  time interval has been set to  $T_h \pm \sigma$ . The process of prediction is shown by Eq. (12):

$$\text{High risk time intervals} = [PM_{last} + T_1] \cup [PM_{last} + T_2] \cup [PM_{last} + T_3 \pm \sigma] \quad (12)$$

where  $PM_{last}$  is the last PM during training or the first PM before the test time,  $T_1$ ,  $T_2$  and  $T_3$  are the high risk intervals from Table 6 and  $\sigma$  is the standard deviation ( $\sigma = \pm 5$  form Table 3). The second approach is to define the high risk days after each CM based on the reason of the CM. According to Table 5, the range of first quantile is considered as the high risk time interval after each reason. Figure 15 shows the prediction algorithm in the flow chart steps:

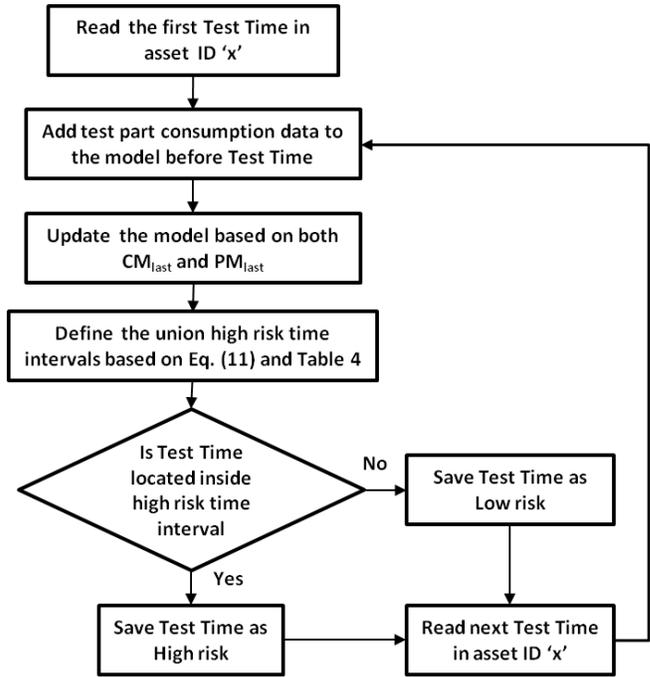


Figure 15. Flowchart of high risk prediction

The algorithm starts by reading the first Test Time for each asset and insert all part consumption data and features from Test Part Consumption before the Test Time. Then according to the PMs features the last PM is selected that can be  $PM_4$  in train section or even  $PM_5$  or  $PM_6$  in test part depends on the value of the Test Time. Using Eq. (12) to define the high risk time intervals based on PM and Table 5 to define the high risk time interval after each CM. The Test Time is considered high risk if it is located in the union of the high risk time interval for both PM and CM. Then continue the process until the Test Times of each of the assets are determined to be high or low risk.

An example of prediction process is shown in Figure 16. The upper plot shows the Test Time (1) at time 944 in pink color and the maintenance action with reason 782 detected as  $PM_5$  as a last PM. The high risk time intervals are shown by vertical orange dashed lines. The Test Time is located in the range of the high risk time interval where the estimation of  $PM_6$  is predicted. Consequently, the Test Time (1) is considered as high risk. The lower plot shows the Test Time (2) for the same asset at time 993. By moving the test time, ahead more data before the new test time pull out from Test

Part Consumption, which reveals new maintenance action with reason 417. For this reason, the last PM location changed from PM<sub>5</sub> to PM<sub>6</sub> with reason 417 and the high risk time intervals updated based on PM<sub>6</sub>. Since Test Time (2) is not located inside, any high risk time intervals it is consider as low risk.

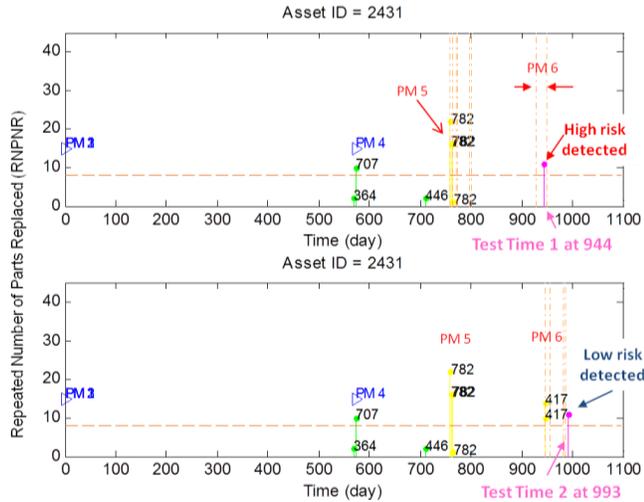


Figure 16. Example of risk assessment and prediction for two test times

## 6. EVALUATION AND RESULTS

Evaluation is the most important step to obtain feedback about goodness of the method. Usually in the competitions, this feedback can be acquired by submitting the results; however, the number of submissions is restricted. Therefore, it is better to have an evaluation method to determine if the method works well.

To evaluate the proposed method, we divided the training data into two sections; section one is the first year and section two is the second year. We trained the model based on section one and tested by section two. To assess how the results generalize to independent failures and avoid over-fitting the cross validation technique used. There are totally 4870 failures in the second year of the training data set. These failures randomly classified into five groups. Each group contains 974 failure targets which distributed uniformly over second year. A uniform random numbers generated in a way that 50% of the random numbers represent the failures in each group. Because according to Eq. (1) the number of "low risk" and "high risk" is equal. Thus, in each group there are 1948 targets that 50% of them are real failures.

The number of failures that has been predicted correctly in each section is considered as the correct high risk (cH). Based on this method, we measure the effectiveness of the prediction by the ratio of cH over the number of failures in each section ( $H_n$ ). The closer the value of cH to the number

of failures in each section ( $H_n$ ), the better the result is. The same process has been applied for Low risk (or non-failure) sample times as well. Because for each section the algorithm selects less than 974 sample times as high risk other sample times consider as low risk, therefore the probability to have better accuracy on low risk is higher. In other words, low risk is assigned to the majority of given times. To converge the results this process applied several times. Table 7 summarizes the results from one of the iterations.

Sect.	Number of failures in 2 <sup>nd</sup> year	Failures in each section ( $H_n$ )	Correct predicted failures (cH)	Eff. cH (%)	Correct predicted Non failures (cL)	Eff. cL (%)
1	4870	974	222	0.23	836	0.86
2	4870	974	213	0.22	782	0.80
3	4870	974	207	0.21	893	0.92
4	4870	974	234	0.24	891	0.91
5	4870	974	201	0.21	863	0.89
<b>Mean</b>				<b>0.22</b>		<b>0.88</b>

Table 7. Prediction accuracy of evaluation technique

The results show the accuracy for High risk is about 22% of failure prediction and for Low risk is about 88%. According to Eq. (1) the evaluation criteria can be obtained by sum of cH and cL. Although this method gives some hints about the accuracy of the prediction, still has an uncertainty. For some assets, there is not enough data for the first year, which causes building poor model.

## 7. CONCLUSION

The health assessment of the data from the PHM 2014 data challenge was completed through the analysis of the Preventive and Corrective Maintenances. The preliminary processes of the data analysis included the removal of outliers, feature extraction selection methods. After those methods were completed, a health assessment rating was given to the asset ID's at specified times to determine if that specific asset was at high risk.

Pre-Processing functions, included data observation, sorting and irregularities in the usage data, were used to eliminate data that does not have the quality required to accurately determine features. The concept of a bathtub curve was used to extract and select features on how soon after a maintenance a failure would occur during the training data, which was then applied to the testing data to provide an accurate health assessment rating and therefore determine if the asset was at high risk. A crucial part of using the bathtub curve is determining where the preventative maintenances fall in the data we were given. Through observation and

analysis of the repeated number of part numbers replaced, and MTBR, function were used to determine where the preventative maintenances fell, which provided most of the information needed to determine the health risk of assets in the given data set.

The other approach is to use the corrective maintenances to evaluate the correlation between corrective measures and the probability of failure afterward. Using the number of failures immediately after maintenance actions in the testing data, the probability of failure after maintenance actions was calculated.

## 8. SUGGESTIONS FOR FUTURE WORK

The proposed risk assessment algorithm provided encouraging results but there are several refinements to be considered for future work. The first idea is to implement usage data in the analysis. The usage measurements can be considered for estimating more precise period for the PM location; however, usage measurements are not exactly at the maintenance actions times. In addition, it needed to be synchronized based on usage per day or the slope of usage for each asset.

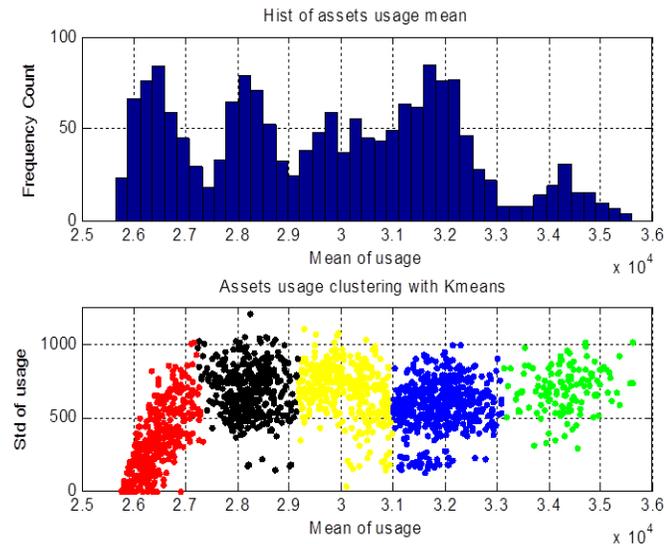


Figure 17. Asset clusters based on usage features (mean and standard deviation)

The second thought is to classify the assets into different clusters based on the usage features. We did some works in this area and extracted two main statistical features from train usage data including mean and standard deviation. Figure 17 upper plot shows the histogram of average usage of each asset. The high and low frequency of average value inspire some cluster existing based on usage. Figure 17 lower plot demonstrates standard deviation of usage versus mean of usage for each asset using K-means technique to

classify them and using mean of Silhouette value as a criterion of the optimal number of clustering. The work was not completed because of the lack of time for this competition. However, the probability of failure in each cluster would provide more extension to the proposed framework and would provide a way of detecting high risk assets based on usage.

The third suggestion is deep analysis of the unique part numbers and the number of parts replaced (the fifth column of part consumption). The RPNR in this analysis just measures the repeated number of parts replaced and for example does not make any differentiation between part number 5566684 and part number 953340. The correlation between the number and time of failures occurred and the exact part numbers can depict more information on the higher risk of some parts because of their material quality, different work condition, installation affect or higher human error on their maintenance.

## ACKNOWLEDGEMENT

The work is supported by NSF Industry/University Cooperative Research Center on Intelligent Maintenance systems (IMS) at the University of Cincinnati.

## REFERENCES

- Arunraj, N., Maiti, J. (2007). Risk-based maintenance—Techniques and applications. *Journal of Hazardous Materials*, 653-661.
- Ebeling, C. (1997). *An Introduction to Reliability and Maintainability Engineering*. New York: McGraw-Hill.
- El-Ferik, S., Ben-Daya, M. (2006). Age-based Hybrid Model for Imperfect Preventive Maintenance. *IIE Transactions*, 365-375.
- Endrenyi, J. e. (2001). The Present Status of Maintenance Strategies and the Impact of Maintenance on Reliability. *IEEE TRANSACTIONS ON POWER SYSTEMS*, 638-646.
- Garvey, D. (2014 August). From phm data challenge: <http://www.phmsociety.org/events/conference/phm/14/data-challenge>
- Khan, F., Haddara, M. (2003). Risk-based maintenance (RBM): a quantitative approach for maintenance/inspection scheduling and planning. *Journal of Loss Prevention in the Process Industries*, 561-573.
- Klutke, G.-A., Kiessler, P., Wortman, M. (2003). A Critical Look at the Bathtub Curve. *IEEE TRANSACTIONS ON RELIABILITY*, VOL. 52, NO. 1, 125-129.
- Kotz, S., Seier, E. (2009). An analysis of quantile measures of kurtosis: center and tails. *Statistical Papers*, 553-568.
- Lin, D. Ming, Y., Richard, M. (2001). Sequential Imperfect Preventive Maintenance Models with Two

- Categories of Failure Modes. *Naval Research Logistics*, 172-183.
- Rezvanizani, S., Barabady, J., Valibeigloo, M., Asghari, M., Kumar, U. (2009). Reliability Analysis of the Rolling Stock Industry: A Case Study. *International Journal of Performability Engineering*, Vol. 5, No. 2, pp. 167-175.
- Siegel, D. (2013). Prognostics and Health Assessment of a Multi-Regime System using a Residual Clustering Health Monitoring Approach. *Doctoral Dissertaion. University of Cincinnati*.
- Siegel, D., Lee, J. (2011). An Auto-Associative Residual Processing and K-means Clustering Approach for Anemometer Health Assessment. *International Journal Of Prognostics And Health Management*, ISSN 2153-2648,.
- Wang, K., Hsua, S., Liub, P. (2002). Modeling the bathtub shape hazard rate function in terms of reliability. *Reliability Engineering & System Safety*, 397-406.