# Bayesian-based Component Lifetime Prediction Model Using Workshop and Telematics Data

Seungyoung Park[1,2], Jihyeon Lee[1]

[1] *HL Mando Corp. WG Campus F3 Lab, Seongnam-si, Gyeonggi-do, 13486, Republic of Korea*
*seungyoung.park@hlcompany.com*
*jihyeon.lee@hlcompany.com*
[2] *Department of Mechanical Engineering, Seoul National University, Seoul, 08826, Republic of Korea*
*seungyoung,park@snu.ac.kr*

## ABSTRACT

This paper presents a Bayesian approach to predicting brake pad and battery life based on field service data from a fleet management system(FMS). The data includes vehicle driving data collected via telematics and maintenance record data managed by the workshop. The proposed approach consists of three modules: component health diagnosis, workshop data analysis and driving pattern analysis. Using KL divergence, the health diagnosis module detects changes in domain-based transformed features from driving data. The maintenance record data from the workshop analysis module estimates the prior probability of maintenance cycles. The censored nature of workshop data is validated by updating the posterior probability using driving patterns from driving data. The driving pattern analysis module classifies driving patterns for lifetime prediction. This study develops a predictive maintenance model for brake pads and batteries without additional sensors using the data required for fleet operation. The mileage-based maintenance approach commonly used for fleet management is improved by this model. Future FMS systems are expected to make extensive use of this concept.

## 1. INTRODUCTION

To reduce maintenance time and costs, Prognostics and Health Management (PHM) technology, which diagnoses conditions and predicts future failure time, is becoming increasingly important. There are a variety of methods, including data-driven and physics-based, for PHM diagnostics and prognostics. Most of these methods typically require real-time sensor data to be collected and stored or uploaded for analysis.



IoT technology is being actively applied in the mobility sector. In particular, for fleet companies providing services to various customers, IoT technology is essential for service operation and management(Arena, Collotta, Luca, Ruggieri, & Termine, 2021). For example, connectivity between in-vehicle IoT devices allows customers to control their vehicles with a smartphone application, and vehicle owners to manage their entire fleet online. Fleet owners not only want to know how well their fleets are operating, but they also want to monitor their health online. Online health monitoring can prevent losses due to unexpected breakdowns. This requires Mobility PHM, a technology that diagnoses and prognosis vehicle health.

However, there are challenges in applying PHM to mobility. Diagnostic capabilities in the mobility domain are limited due to signal noise, dependence on environmental and operating conditions, lack of fault data, and uncertainties in maintenance records(Vogl, Weiss, Helu, & moneerhelu, 2019). In mobility, faults are detected by maintenance records, but the use of maintenance records has the disadvantage that they are easily censored. There are parametric and non-parametric estimation methods for estimating the distribution of maintenance data to deal with censored characteristics, both methods are not simple and emphasize the importance of an accurate definition of the data (Yang, Kanniainen, Krogerus, & Emmert-Streib, 2022). Variability in operating and environmental conditions is particularly important in fleet management. This makes it difficult to accurately estimate the remaining useful life (RUL).

Previous research has proposed different approaches to address each of these challenges. Voronov, Frisk, and Krysander (2018) describes a novel methodology to correct the censored nature of maintenance data. A Bayesian approach is proposed to improve the accuracy of RUL prediction by updating the degradation curve in real time as operating conditions change (Gebraeel, Lawley, Li, & Ryan, 2005). Others present methods that use more than one type of data, and are proposed to predict the remaining useful life of an automotive component using multiple features, such as component wear, usage, and beam-searched from maintenance record data and logged vehicle data(Prytz, Nowaczyk, Rögnvaldsson, & Byttner, 2015). We have separately derived the following implications from these papers. 1) Vehicle lifetime prediction can be achieved from maintenance record data with censored attributes removed. 2) Usage patterns have an influence on component lifetime. 3) The accuracy of lifetime prediction can be improved by updating usage or operational data.

In this paper, we propose a novel methodology that simultaneously incorporates all the findings derived from previous research. The methodology is based on Bayesian approaches to compensate for the censored nature of the maintenance records that are used for the residual life distribution parameters, to improve prognosis accuracy using vehicle driving data as a likelihood to update the lifetime distribution considering the variational operating conditions. We also propose a simple diagnostic method using vehicle driving data to determine the uncertain initial state and monitor unexpected component failures. Our contribution is that we overcome the problem of limited fault data using maintenance records which are well defined and refined by our novel method, and all the data used are acquired in real-world fleet operating situations, making them suitable for future services. The results are compared to the heuristic planned maintenance method by skilled mechanics and validated with a physically measured residual amount of each component.

The structure of the paper is as follows: In the ”System and Data Definition” section, we explain details of the overall system, the target components, and the characteristics and limitations of the dataset we used. The ”Bayesian Modeling” section then describes the overall structure and details of the model to overcome the aforementioned limitations. In the ”Results” section, we discuss the experimental setup and its results, as well as the validation process and limitations of the model proposed in this paper. Finally, the ”Conclusion” section summarises the methodology proposed throughout this paper, the main findings and contributions.

## 2. System and Data Definition

This paper is based on vehicle and maintenance records from 30 operational taxis. Each vehicle is equipped with a data acquisition system called ’SmartLink’. This system is used for fleet management vehicles by SK Rent-a-Car, a car rental service company in South Korea. This system collects vehicle data online in real-time. All vehicles from which we collected data are maintained by qualified mechanics who are hired by the taxi company and the maintenance history is recorded to ensure serviceability. In this study, we used these maintenance records together with the vehicle driving data. For verification, we collected additional data that measured the actual residuals of the target components.

### 2.1. Target Components: Brake Pad and Battery

In this study, we selected brake pads and 12V batteries as target components based on their safety, operational importance, and maintenance frequency.

The brake system is responsible for stopping the vehicle and is usually considered the most important safety component. Brake system performance problems cause significant traffic accidents, so it is necessary to diagnose abnormal conditions and take proactive management. The brake works by pressing down on the spinning disc with a pair of pads to stop it. This creates friction between the disc and a pair of pads, resulting in heat loss and pad wear. Taxis have a high daily mileage and require frequent brake pad changes.

The 12V batteries have relatively few safety issues compared to the brakes. However, undetected battery faults can cause problems in vehicle operation, such as failure to start or engine shutdown while driving. To prevent this, vehicles are equipped with battery sensors that provide information on the state and health of the battery. However, the vehicles used in this study do not use the OEM-recommended batteries for financial and durability reasons, which prevents or limits the sensors from collecting battery status and health signals, so an alternative monitoring method is required.

### 2.2. Maintenance Records from Workshop

All vehicles are maintained manually in the fleet owner’s own workshop. When maintenance is carried out, information such as the date of component replacement and mileage is recorded. We refer to these as ’maintenance records’ in this study.

The maintenance records can be classified as time-to-event data. An event is a maintenance action, and in this study, we used records of when brake pads and batteries were replaced. Assuming that all maintenance actions occur at the end of the component’s life, we can obtain the lifetime distribution of the components directly from these records. In the real world, however, it is difficult to know the actual remaining life from maintenance records, which introduces a data censoring problem.

Censoring is generally divided into two types: left-censored and right-censored. The left-censored means that we cannot determine if it was $t_s$ at the nth observed time $m_n$. Another type is right-censored, which means that $t_e$ cannot be determined at the nth observed time $m_n$ (Yang et al., 2022).

$t_s$ : start time of the component life

$t_e$ : end time of the component life

$m_i$ : the ith maintenance time $(i = 0, 1, \ldots, n, )$

In the maintenance records, the component was installed at $m_n$, and replaced at $m_{n+1}$. If the mechanic always replaces the component with a new one at $m_n$, it can be assumed that $m_n$ is the start of the component's life, $t_s$. But the component was not replaced with a new one, but with a used one whose remaining life is unknown, it is left-censored. At the replaced time, $m_{n+1}$, if the mechanic does not know or measure the remaining life, the replaced component can be one of three states: it still has remaining life, it has already exceeded its end of life, or it is at the end of life; $t_1 < t_e$, $t_1 > t_e$, $t_1 = t_e$. Both left-censored and right-censored cases are common in the real-world workshop.

To determine the lifetime of components, we need the start time and end time, but if the data is censored, we don't know the exact time. One of the basic approaches to analyzing censored data is a likelihood-based approach (Leung, Elashoff, & Afifi, 1997). In this study, we use this likelihood, which is calculated from the vehicle driving data described below, to deal with censored maintenance records.

### 2.3. Vehicle Driving Data from Telematics System

In general, three types of data are commonly employed in a fleet management system: 1) vehicle driving data to monitor vehicle usage costs; 2) GPS (Global Positioning System) sensor data to prevent or detect theft; and 3) IMU (Inertial Measurement Unit) sensor data to verify the occurrence of accidents. In this paper, we use vehicle driving data collected by the aforementioned telematics system.

There are several challenges to using this data. First, this data is not collected directly from the target component: brake pads do not have sensors to collect signal data, and the battery receives data from a battery sensor, but it is connected to other internal vehicle systems, so the sensor signal includes environmental factors and noise. Secondly, since the ultimate purpose of collecting this data through a telematics system is a fleet management service 'business', the cost of data collection should be minimized by reducing the column or frequency. Thirdly, there is the uncertainty of environmental conditions. As vehicles move, the external environment is constantly changing, as are the drivers, so driving patterns are different from moment to moment.

In this study, to address these challenges, we generate transformed physics-based features using indirect data affecting the wear of the target components and distribution parameters differentiated by the drivers to calculate the likelihood in real-time. For this purpose, vehicle driving data is divided into two types: driver input data and vehicle output data. Driver input data depends on the driver's intention, such as steering angle, throttle position, and brake pedal position. Vehicle output data is the vehicle's response to the driver's input data, such as vehicle speed, acceleration, and brake pressure. The key is to exploit the complex correlations between these two types of data to improve the accuracy of the model.

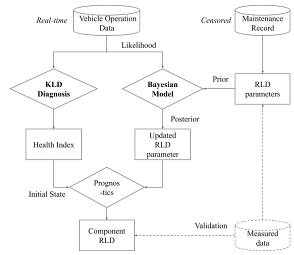


Figure 1. Model Architecture (inspired by (Youn & Wang, 2008))

## 3. Bayesian-based Component Lifetime Prediction Model

This section looks at degradation characteristics, Bayesian updates, and health diagnosis and explores the influence of maintenance records from the workshop and vehicle driving data from the telematics system on component lifetime. We examine the harshness of driving and its impact on the component's lifetime, discuss Bayesian update strategies using driver input data as likelihood and maintenance records as prior, and suggest a simplified method for current health diagnosis. The end goal of the model is to predict the optimal time for vehicle component replacement.

### 3.1. Degradation Characteristics

Brake pad degradation is caused by brake pad wear, which occurs during braking when the brake pads and discs dissipate braking energy through friction. The amount of wear has the characteristic of a non-linear monotonic increase depending on the temperature. The relationship between temperature and amount of wear generally uses experimental results obtained from brake dynamometer tests. To predict the amount of brake pad wear, a physical model is used to calculate the braking energy applied by the driver and the temperature generated in the brake pad. Estimating temperature increases is a relatively straightforward process. However, estimating temperature decreases can prove more complex due to the influence of various factors. These factors, such as vehicle speed, outside temperature, wheel shape, and the specific characteristics of the brake caliper, impact the cooling coefficients, making accurate predictions a challenging task.

In the case of batteries, degradation is caused by the decreasing capacity of the battery. One of the main causes of capacity reduction is the oxidation of the anode, which is due to physicochemical changes caused by the number of charge and discharge cycles and temperature. Since battery capacity cannot be measured by sensors, internal resistance can be used as an alternative. Internal resistance tends to increase as the battery degrades.

### 3.2. Bayesian Updating for Residual Life Distribution

We performed Bayesian updates by fusing two important sources of data to improve the accuracy of component status and prediction. Assuming that the degradation curve of the component lifetime curve is linear,

$$Y = \beta t + e,\ e \sim \mathcal{N}(0, \sigma^2) \tag{1}$$

$$\begin{cases} \beta_{brake} \simeq f(T(D)) \\ \beta_{battery} \simeq f(R(D)) \end{cases}$$

The coefficient $\beta$ of the lifetime curve is different for different applications. In the case of the brake system, the temperature can be estimated using the braking energy and cooling coefficients. This estimation can be simplified as a function of the driving data. For batteries, the internal resistance strongly depends on usage patterns such as the frequency of charge and discharge and environmental factors, which can be expressed as a function of the driving data. The stochastic process, Y, is the observed or to-be-realized value by the maintenance personnel, denoted as $y_t$. Considering censoring, the maintenance personnel do not necessarily replace the component at the degradation deadline. That is, it is replaced with the probability in equation (2).

$$P(y_t) \sim P(\mathrm{h - c < Y < h + c|M}) \tag{2}$$

$$Y \sim \mathcal{N}(\mu, \sigma^2) \tag{3}$$

$$\begin{cases} \mu + C < M & \kappa = 0 : \text{eco} \\ \mu - C < M < \mu + C & \kappa = 1 : \text{normal} \\ M < \mu - C & \kappa = 2 : \text{hard} \end{cases}$$

Based on the maintenance records, the probability distribution of lifetime is derived for each component. We divide the lifetime into specific groups, taking into account factors related to the component's lifetime. The eco group uses components longer than others, and the hard group has the shortest component life compared to other groups. C should be decided by the operational strategy of fleet management.

The prior in equation (5) would be determined by the metadata of the driver and the environmental status, such as age, driving skills, driver characteristics responding to emergency status, traffic, and if the type of data is available. In this study, the mechanics are fixed for each vehicle and individual driver information is not available, so the D was used. We updated the posterior from the likelihoods of the transformed data, which are features for each component that affect lifetime, and the prior from maintenance records and metadata. By repeating the above and continuously updating the input from the driver, we aimed to predict the replacement time of the component more accurately.

$$\tilde{D} = g(D) \tag{4}$$

$$P(\kappa_1|\tilde{D}_0) \propto P(\tilde{D}_0|\kappa_0) \times P(\kappa_0) \tag{5}$$

$$P(\kappa_{i+1}) = P(\kappa_i|\tilde{D}_i), \tag{6}$$

$$\delta y \sim \int_{i,\kappa} P(\kappa_{i+1})\mathbf{E}(\tilde{D}_i) \tag{7}$$

### 3.3. Health Diagnostics

Components that are directly related to safety, such as brakes, can gradually deteriorate or enter a state that is difficult to predict with a very low probability. This can happen if the pads wear unevenly or if there is a problem with the material properties of the brake oil. In order to anticipate problems, it is important to accurately assess the current condition of the component. As previously defined, we have used a method of monitoring the output status of the vehicle in response to driver input.

However, due to the large variability in the environmental conditions of the vehicles, this method can have significant variations, so it was assumed that it would be more advantageous to consider the long term rather than the short term without all relevant information being secured.

We used KLD by comparing the probability distribution of performance over a given period. If there is a significant difference from the reference probability, we make a comprehensive assessment of whether this is a problem compared to the previously calculated lifetime and use it as a way of identifying potential problems in advance. As KLD is somewhat sensitive to outliers based on the reference probability, we considered the conditional probability by introducing environmental factors that may affect brake performance.

## 4. Results

To evaluate the proposed method in this paper, it would be most appropriate to compare it with other examples where PHM techniques are applied, but it was difficult to find applicable cases. Therefore, we observed the advantages of the proposed method compared to the existing maintenance method. We monitored the data of several vehicles for a period of time while observing the forms managed by the existing maintenance personnel for several months. During the observation period, we used separate equipment to periodically measure the exact current state (true value) of each component to determine the difference (c) between the time checked by the maintenance personnel and the actual state. We carried out a comparative evaluation between the existing method managed by mechanics (scheduled maintenance) and the method proposed in this paper, using the measured true value as the basis for evaluation.

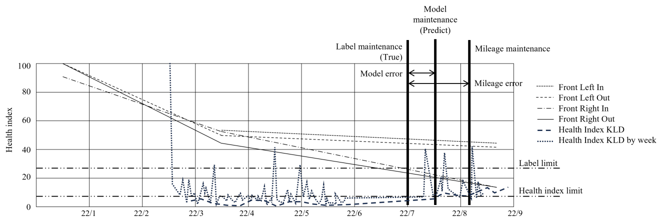


Figure 2. Compare model prediction vs mileage maintenance vs label data (true data) 1: early broken

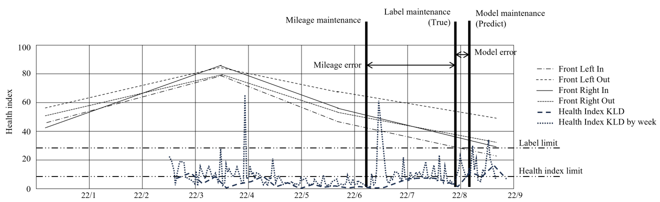


Figure 3. Compare model prediction vs mileage maintenance vs label data (true data) 2: early maintenance, Maintenance likely occurred prior to March 2022

### 4.1. Experimental setup and results

To verify this, we measured the true values of the brake pads and the battery. For the brake pads, we measured the height of the friction surface at four points and calculated the remaining thickness by averaging the values. We have developed equipment to measure pad thickness directly, and this equipment can measure the four-point heights of two types of pads simultaneously. The measurements can also measure uneven wear, but no such labels were found in this experiment. For the battery, there are many ways to accurately measure its health, but conventional methods are dependent on temperature and regular discharge cycles and are difficult to perform in fleet operating conditions. Therefore, we used precise internal resistance measuring equipment, which is also one of the factors in determining battery health and can be used to distinguish excessive abnormal states. A few excessively high internal resistance labels were found as the batteries were constantly managed in fleet operating environments.

The figures above compare the brake pad measurements and replacement dates based on the mileage of two vehicles. The straight line at the top of the figure represents the trend of the measurement changes of four types of front pads (front left in, front left out, front right in, front right out). Due to the limited availability of daily measurements, each data point has been approximated by linear interpolation to create a continuous trend. The dotted lines at the bottom of the graph represent the daily and weekly average health index.

The validation results using the true labels showed an accuracy of over 0.9 and confirmed the superiority over the existing mileage-based maintenance. The first graph relates to cases where the actual pads wear faster than the expected replacement point based on mileage, requiring early replacement. Compared to the mileage model, the predictive model was able to determine relatively quickly when to actually replace the parts. The second graph relates to cases where actual failures occur later than the expected replacement point based on mileage. If the actual measurement values do not decrease steadily, it is estimated that replacement will occur before March 2022. In this case, replacement based on mileage occurs approximately one month earlier than the actual wear replacement point. In contrast, the model estimation results indicate replacement at a similar time to the actual measured values.

### 4.2. Discussions

The methodology proposed in this paper is expected to be more accurate than the existing method, which was determined by the average driving of the general population, as it reflects the uncertainties caused by driver operation. However, there are several factors that need to be considered: The workshop data used in the study partly reflected the mechanic's intention. It may be difficult to use the same method if the mechanic changes or the operating environment changes. The fleet studied in this paper had experienced drivers who did not change the driver of each vehicle and drove in similar environments, so there is a limitation that the operating characteristics were not varied. Given these limitations, future research should address conditional probabilities encompassing diverse uncertainties, as they could play a key role in extending the applicability of our methodology to drivers with different driving tendencies.

We modeled the component degradation characteristics as a linear function and combined them as a Bayesian update to estimate the used life for specific time periods. In the background of the linear function, high-order degradation factors, or experimental coefficient results, coefficient $\beta$ needs to be implemented. If we test components independently, such as a rig test, then acquire the distribution of the degradation factors, it can be used for different components. By leveraging this simplification strategy, we aim to efficiently handle the complexity associated with different components within the mobility system.

## 5. CONCLUSION

In this paper, we propose a Bayesian-based prediction technique that applies maintenance data to actual driving data in order to apply PHM techniques to real-world data. Accordingly, it is expected that an infrastructure will be established that can combine data, including various environmental factors that may occur in operational situations. In such a scenario, a clear understanding and approach to the background and procedures for collecting maintenance data is essential.

As mobility is made up of many different components, only a few critical parts cannot be representative of the safety of the whole system. Therefore, it is crucial to use a methodology that can integrate models for each component into a coherent whole, allowing for a systemic evaluation. In our study, we have adopted a system that uses multiple linear models for each component, which is designed to facilitate the application of coefficients $\beta$ derived from different methodologies and domain knowledge bases. In future work, we aim to integrate models developed by specialized component companies using the approach proposed in this study, thereby enabling a comprehensive system-level evaluation.

To validate the methodology in this study, we compared it to the mileage-based management method used by skilled mechanics by measuring uncensored true values over several months. This confirmed that performance can be improved by combining data used in fleet operations without the need for expensive, high-precision sensors commonly used in PHM. These findings open up new possibilities for the future use of PHM techniques in mobility applications.

## ACKNOWLEDGMENT

This work was supported by the Industrial Strategic Technology Development Program (No. 20017301) funded By the Ministry of Trade, Industry & Energy(MOTIE, Republic of Korea). We gratefully acknowledge the support of Minwoo Park in HL Mando Corp. and Minho Yoon in SK Rent-a-Car.

## NOMENCLATURE

| | |
|---|---|
| $t_s$ | Component life start time |
| $t_e$ | Component life end time |
| $m_n$ | nth observation by inspectors |
| $Y$ | residual life of a component |
| $\beta$ | Linear coefficient of lifetime |
| $t$ | time or load factors to lifetime |
| $e$ | error in linear model |
| $f$ | probability density functions |
| $T$ | temperature |
| $R$ | internal resistance |
| D | Driver input |
| $y_t$ | realized maintenance time |
| $h$ | degradation limit or deadline |
| c | censoring error |
| M | maintenance person, mechanic |
| $C$ | user-defined range |
| $g : D \to \tilde{D}$ | transformation function |
| $\kappa$ | specific groups considering factors, t |

## REFERENCES

Arena, F., Collotta, M., Luca, L., Ruggieri, M., & Termine, F. G. (2021, 12). Predictive maintenance in the automotive sector: A literature review. *Mathematical and Computational Applications 2022, Vol. 27, Page 2*, *27*, 2. doi: 10.3390/MCA27010002

Gebraeel, N. Z., Lawley, M. A., Li, R., & Ryan, J. K. (2005, 6). Residual-life distributions from component degradation signals: A bayesian approach. *IIE Transactions (Institute of Industrial Engineers)*, *37*, 543-557. doi: 10.1080/07408170590929018

Leung, K.-M., Elashoff, R. M., & Afifi, A. A. (1997). Censoring issues in survival analysis. *Public Health*, *18*, 83-104.

Prytz, R., Nowaczyk, S., Rögnvaldsson, T., & Byttner, S. (2015, 5). Predicting the need for vehicle compressor repairs using maintenance records and logged vehicle data. *Engineering Applications of Artificial Intelligence*, *41*, 139-150. doi: 10.1016/J.ENGAPPAI.2015.02.009

Vogl, G. W., Weiss, B. A., Helu, M., & moneerhelu, M. H. (2019). A review of diagnostic and prognostic capabilities and best practices for manufacturing. *Journal of Intelligent Manufacturing*, *30*, 79-95. doi: 10.1007/s10845-016-1228-8

Voronov, S., Frisk, E., & Krysander, M. (2018, 6). Data-driven battery lifetime prediction and confidence estimation for heavy-duty trucks. *IEEE Transactions on Reliability*, *67*, 623-629. doi: 10.1109/TR.2018.2803798

Yang, Z., Kanniainen, J., Krogerus, T., & Emmert-Streib, F. (2022, 5). Prognostic modeling of predictive maintenance with survival analysis for mobile work equipment. *Scientific Reports 2022 12:1*, *12*, 1-20. doi: 10.1038/s41598-022-12572-z

Youn, B. D., & Wang, P. (2008). A generic bayesian approach to real-time structural health prognostics. *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, MAO*. doi: 10.2514/6.2008-6053

## BIOGRAPHIES

**Seungyoung Park** is a Senior Research Engineer at HL Mando Corp. since 2008. Currently focusing on a new business for mobility, he previously worked on brake and e-traction systems. He is pursuing a Ph.D. in mechanical engineering at Seoul National University and holds B.S. and M.S. degrees in mechanical engineering from Pusan National University (2006 and 2008 respectively). His research interests include prognostics and health management in mobility.

**Jihyun Lee** is the Research Engineer at HL Mando Corp. since 2017. Currently focusing on new business on new business for mobility, she previously worked on e-traction systems. She received a B.S. degree in mechanical engineering from Ulsan National Institute of Science and Technology, South Korea, in 2018. Her research interests include prognostics and health management in mobility.