Title: - Leak Frequency Analysis for Hydrogen-based Technology using Bayesian and Frequentist Methods

Authors:

Mahesh Kodoth[a], Shu Aoyama[a], Junji Sakamoto[b], Naoya Kasai[c,d], Yehia Khalil[f], Tadahiro Shibutani[c,e,*], Atsumi Miyake[c,d,e]

[a] *Graduate School of Environment and Information Sciences, Yokohama National University, 79-7 Tokiwadai, Hodogaya-ku, Yokohama, Kanagawa 240-8501, Japan*

[b] *Graduate School of Natural Science and Technology, Okayama University, 3 Chome-1-1 Tsushimanaka, Kita Ward, Okayama, 700-8530, Japan*

[c] *Institute of Advanced Sciences, Yokohama National University, 79-5 Tokiwadai, Hodogaya-ku, Yokohama, Kanagawa 240-8501, Japan*

[d] *Faculty of Environment and Information Sciences, Yokohama National University, 79-7 Tokiwadai, Hodogaya-ku, Yokohama, Kanagawa 240-8501, Japan*

[e] *Center for Creation of Symbiosis Society with Risk, Yokohama National University, 79-5 Tokiwadai, Hodogaya-ku, Yokohama, Kanagawa 240-8501, Japan*

[f] *United Technologies Research Center (UTRC), East Hartford, CT 06108 USA*

* Corresponding author: E-mail: shibutani-tadahiro-bj@ynu.ac.jp, Tel & Fax: +81-45-339-3597.

Abstract

Dealing with hazardous environments such as hydrogen poses considerable risks to property, people, and the environment. Leak frequency analysis is a method of understanding the characteristics of risks at hydrogen refueling stations (HRSs). This paper proposes leak rate estimation using time-based evaluation methods that utilize historical HRS accident information. In addition, leak frequency estimates from another two methods (non-parametric and leak-hole-size) were examined. In the non-parametric approach, the leak frequency is estimated based on a Bayesian update. The results from these three approaches are summarized to understand the trend of leak rate data. The leak rate data from the time-based method displays a similar trend to the leak size based method. However, the non-parametric method tends to be conservative due to high failure observations (new evidences) during the Bayesian update. Finally, the unrevealed leak time was calculated as a function of the leak frequency. The quantitative insights of this study can be used to set performance standards for the availability and reliability in the operation and maintenance of HRSs.

Keywords: Leak frequency, unrevealed leak time, hydrogen refueling station, time-based model, Bayesian update

1. Introduction

Hydrogen refueling stations (HRSs) are a key infrastructure in the fuel supply chain for fuel cell vehicles (FCVs). Several hundred stations have been planned worldwide for implementation before 2020 (IEA, 2015). Because pressurized hydrogen is used in FCVs that require sufficient cruising distance, a large amount of pressurized hydrogen is stored at HRSs. Moreover, there are risks associated with the pressurized hydrogen stored at HRSs. When an accident occurs with respect to the high-pressure gas, a notification report should be submitted to the prefectural governor or police official as per the High Pressure Gas Safety Act (KHK, 2015). Accident information such as hydrogen leakage at an HRS is available in the high-pressure gas incidents database of The High Pressure Gas Safety Institute of Japan (KHK, 2012). This database contains a compilation of high-pressure gas accidents, including accident information for HRSs. It also provides information on which facility tends to fail and on the accident count over the years. Considering the accident statistics of natural gas stations, there are concerns that HRS accidents may increase as more HRSs are implemented in the future.

It is well known that there is a possibility of abnormal events occurring due to increased activities and operations performed at the HRS. As HRSs store and dispense hydrogen at relatively high pressure, they are controlled by the aforementioned High Pressure Gas Safety Act. This Act defines "accident" as follows: (i) Explosion, (ii) Fire, (iii) Leak, (iv) Degradation, and (v) Others. In legal terms, explosions, fires, spouting or leaks, rupture or damage, and loss or burglary are defined as "accidents" (Yamada et al., 2015). For example, a minor leakage at an HRS is recognized as a reportable accident. In the case of hydrogen fuel, even a small leak in a confined space can potentially lead to a catastrophic event. Most accidents at HRSs are due to hydrogen leaks. In fact, almost all accidents at HRSs reported in the database are hydrogen leaks (KHK, 2012). In this case, the accident rate can be considered to be almost equivalent to the leak rate. In this study, the "leakage or leak rate" refers to an "accident", as defined in the High Pressure Gas Safety Act.

In the general quantitative risk assessment (QRA), risks are calculated from the impacts and frequencies of leak scenarios. The failure (leak) rate estimation is a key component of QRA. However, hydrogen failure data for QRA is extremely limited. One possible solution for this is to use generic failure data from public or commercial facilities, such as nuclear power plants, chemical plants, and offshore platforms (Redbook, 1997). Casamirra et al. (2009) used fault tree analysis (FTA) to determine the occurrence frequency of accidental scenarios based on generic failure data. Another solution is to employ a Bayesian statistical approach to estimate the failure rate from a limited data source.

Sakamoto et al. (2016) carried out a qualitative study on leakage-based analysis of accidents in Japan. In their study, leakage was classified based on the components and cause of accident. One of the characteristics of HRS accidents in Japan is that a high percentage of leak accidents occur at pipe joint sections. Because there are many joints and seals in a hydrogen refueling station, and the station's hydrogen compressor produces mechanical vibrations, small leaks from joints and seals are a major concern at HRSs. The results revealed that the main cause of leakage among flanges, valves, and seals is screw joint failure. Leakage associated with the filling hose and dispenser is mainly due to human error. Although their study makes an important contribution to leakage analysis of HRSs, it is limited to only a qualitative assessment of the leakage analysis.

The leak frequency of HRSs has been reported by several researchers. One of the ways to estimate leak frequency is based on the hole size. LaChance et al. (2009) developed a Bayesian model for leak frequency in various components used in HRSs. The leak frequency is assumed to be a function of the fractional flow area of the leak. The leak frequency was estimated as a function of leak size, which is the ratio of the leak area divided by the total cross-sectional flow area. For a leak area of 0.1% of the total flow area, the corresponding system leakage frequency would be 0.03 per year and 0.06 per year for the 20.7 MPa and 103.4 MPa systems, respectively. Since even a small leak from joints and seals are a major concern at HRSs, most leak failures can be classified as a very small leak with leak area of 0.01% of the total flow area. When the leak area is 0.01% of the total flow area, the system's leak frequency would be 0.2 per year.

A more traditional approach to leak frequency estimation is provided in the Dutch Redbook model (Redbook, 1997). The model employs a non-parametric approach using the Nelson–Aalen estimator. By using a non-parametric approach, the leak rate can be estimated as a function of the number of fillings in the HRS. The relationship between the cumulative hazard and the number of fillings should be understood and equated to calculate the leak rate of the system. The method developed is quite generic, and is often implemented in the oil and gas industry where more traditional approaches using constant failure rates are widely adopted in leak and failure analysis. However, the characteristics of HRSs are different from those of the oil and gas industry. For example, in the oil and gas industry, accidents can occur due to a wide range of causes resulting in leaks, toxic effect, fire, and/or explosions. However, in HRSs, fire and/or explosions are not observed on a large scale. As mentioned earlier, leakage is a major event in accidents reported in HRSs.

Some advanced probabilistic models have been employed by some researchers to quantify the risks caused by leaks. Khalil (2017) employed a state-of-the-art visual flowchart methodology to develop a probabilistic model that can quantify occupational risks of fire and explosion events initiated by leaks that ignite within enclosed spaces. The functionality of this proposed model was demonstrated by a HRS case study in which gaseous hydrogen leaked from the compressor system. The application of various existing techniques for leak modelling is restricted due to the scarcity of data. The issue of scarce data is modelled using a precursor data and hierarchical Bayesian methodology (Yang et al., 2013, Gheriani et al., 2017).

The Bayesian network has been widely used in important fields involving safety assessments. A related study focused on an estimation method for accident rates calculated for equipment failure of a nuclear power plant (in Japan's nuclear probabilistic risk assessment). Japan Nuclear Technology Institute introduced the Bayesian method to enable updates to the uncertainty width of failure rate with data storing, which had a fixed value until then (JNTI, 2009). This method is still used in the latest report (JNTI, 2016). This latest report considers the failure rate to be constant over time and the probabilistic variance is updated by new data. In addition, the observation probability $\{p\}$ is taken into account in the nuclear report for failure rate estimation.

Focusing on the HRS operation time from its start may reveal characteristics of leak occurrence. In other words, collecting data about leaks from the past will allow better understanding of the trend in the possibility of leaks by identifying its operation time. Under such circumstances, time-based evaluation is important. Studies have been performed on HRS accidents or leak frequency by organizations such as Sandia National Laboratories (LaChance et al., 2009). However, only a few studies have been conducted on the time series effect. This is one of the reasons why the operation time is given more importance and is discussed at length in this paper.

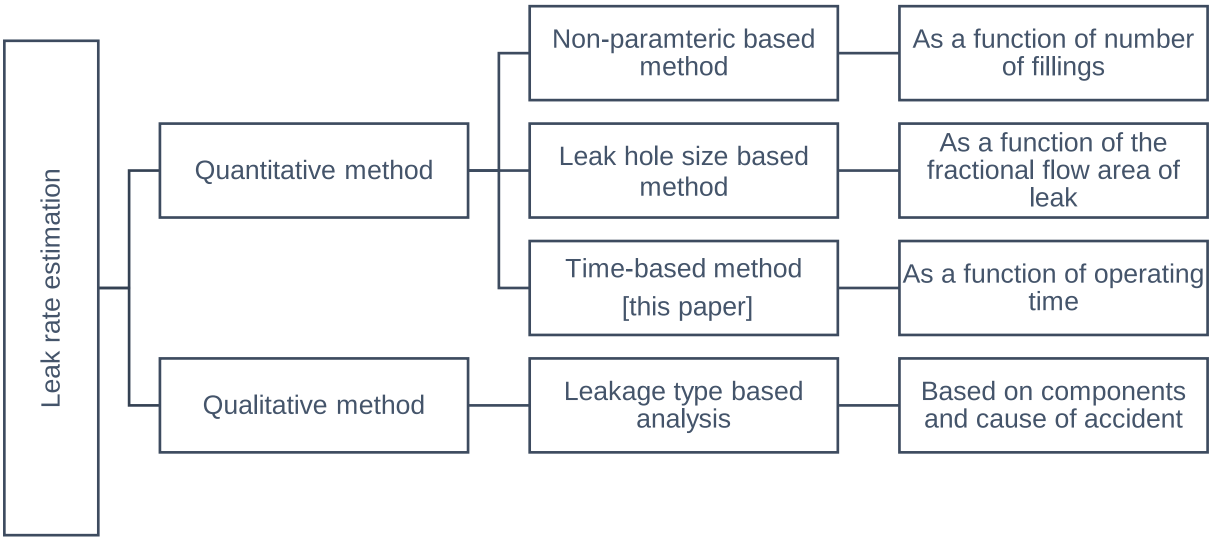

Fig. 1. Review of methods developed for estimating leak frequency

Fig. 1 summarizes various methods that are currently adopted in the leak frequency estimation of HRS. The leak-rate estimation methods can be classified as follows:

1. Qualitative method
    a. Leakage-type-based analysis (Sakamoto al., 2016)
2. Quantitative method
    a. Leak-hole-size method (LaChance et al., 2009)
    b. Non-parametric method (Redbook, 1997)
    c. Time-based method (this paper)

Note: All of the above quantitative methods employ Bayesian update for leak data evaluation

In this study, the models described above are analyzed to identify trends. Failure and operating data of HRSs are collected and analyzed in detail to estimate the leak rate using frequency (time series effect) and Bayesian based evaluation methods. The results will describe trends in the leak rate: whether they increase or decrease over operation time, or whether they are peaking and declining. Leak frequency estimations from various methods are examined with the present method to understand the different ways of modelling leak frequency. The algorithm based model implemented through statistical interpretation and WINBUGS tool provides a new way of dealing with accident data in safety and risk management. The study results will help asset managers make an engineering judgement on the appropriate leak rate data of systems.

In addition, unrevealed leak time is calculated as a function of leak rate and inspection interval. Unrevealed leak time is one area within safety and risk management of hydrogen stations that has not yet been addressed in any research paper. The authors believe that in addition to process safety time, unrevealed leak time is an equally critical parameter that needs to be considered in the engineering safety designs. It determines the time period when the leak exists at the installation due to an unrevealed leak failure. This is considered to be an important characteristic of HRSs. The quantitative insights of this study can be used to set performance standards for the availability and reliability of safety critical systems, such as leak detectors, during the operation and maintenance of the HRS.

2. Leak data evaluation at HRSs based on operation time

In this case, the number of accidents (leaks) at an HRS over time was determined from the start of its operation. The term "leak rate" is used in this paper in reference to the accident occurrence per unit time per HRS. "HRS operation start" denotes the start of HRS operation used in either test research or commercial operation. The operation start time of the HRS does not include the construction time of the HRS infrastructure. "Through operation time" indicates that the data is treated with the time elapsed from the start of operation. The accident count is based on the events listed in the high-pressure gas incidents database of The High-Pressure Gas Safety Institute of Japan (KHK, 2012). Operation time is the period between the operation start month and the accident occurrence month. In this paper, a "month" signifies a unit of time measurement.

Kodoth et al. (2018) has already collated data based on the events listed in the high-pressure gas incidents database (KHK, 2012) to determine the data uncertainty in accident rate estimation. The same data will be referred to in this paper, because the data source is common to both studies. In total, 26 accidents were reported for these HRSs. The data source is limited to 35 MPa and 70 MPa systems. The original data contains information on when the station operated and when an accident happened for each HRS. The length of time that elapsed from the operation start time to the accident occurrence is calculated using these data. Although the starting periods differ among the stations, they are assumed to be the same point for accident analysis. The unit of time is "month". The accident count for each month is estimated as "[Event count per station-month]".

However, the data collated from the database by Kodoth et al. (2018) contain many no-accident months. The statistical model that will be introduced in Sections 3.2.1 and 3.2.2, in which the function $f(x)$ describes each month's accident rate, is not suitable for the original data, which contains many no-accident months. Consequently, the input data were modified as follows: if the first accident occurred in the second operation month, it is distributed evenly over the first and second operation months, resulting in the input accident count for each month being 0.5 [event per (station-month)]. In brief, the number of accidents in an operation month is divided by the length of the non-accident period starting from the earlier accident and is estimated as the average accident input data over the period.

The converted input data are shown in Fig. 2. A relatively large value for the input data corresponds to accidents in rapid succession, whereas a relatively small value for the input data corresponds to accidents occurring over a long period. Although the data presented in Fig. 2 are similar to the original data collated by Kodoth et al. (2018), note that there is no zero-accident month in Fig. 2. This paper emphasis lack of data treatment and re-organizing them to make it suitable for the model.

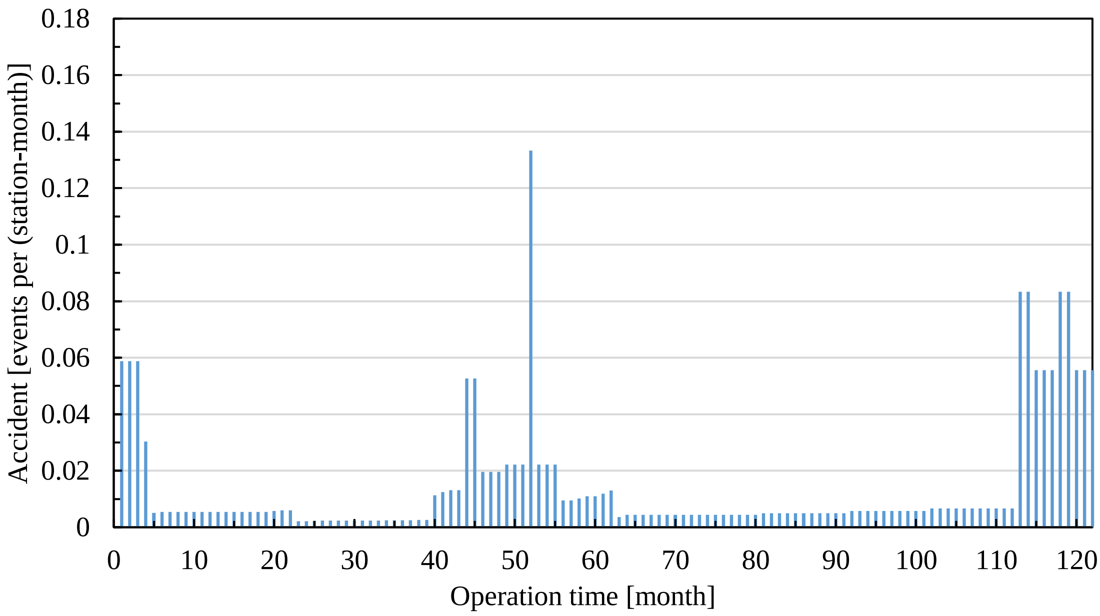

Fig. 2. Input data for estimation by function of time

3. Methods for Estimation and Interpretation of Leak rate

In this section, two statistical models (Lognormal and Weibull) are applied to the time series station events data in order to understand their characteristics. As each model has a different application, suitable care should be taken to apply the correct model to the data.

3.1 Flow of accident rate analysis as a function of time

The flow of accident rate analysis by function of time is shown in Fig. 3. The analysis flow is divided into the two parts. Part 1 is related to organizing data in the format suitable to the model. The input data (referring to accident data in Fig 2) is analyzed by operation time (mean). Input data is given as an input to Part II. Part II performs statistical analysis based on the model described in Section 3.2. The output from the model is the posterior data.

1. Part I: Input data preparation for statistical analysis software - Data processing is needed in accident

analysis using either the log-normal function or the Weibull function of time. The prior distribution for the dataset is represented in Fig.2. The prior (input) data reported is given as an input to the model.

2. Part II: Statistical analysis using WINBUGS software (Ntzoufras, 2009) - Using the prior (input) data, the accident rate for each month is estimated. The model in WINBUGS is written in a series of commands as shown in Appendix.

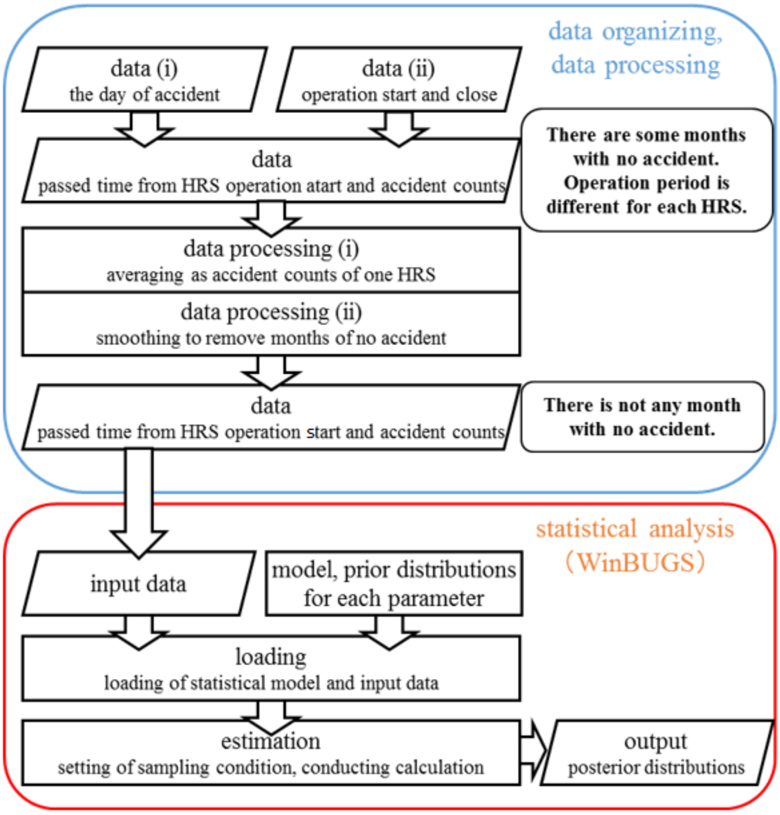

Fig. 3. Flow of analysis using the log-normal function or Weibull function of time.

3.2 Estimation of leak rate based on time function

In the early stages of a hydrogen station's operation, human factors can cause accidents because workers may not operate or maintain the system well, and this may cause frequent accidents. During the intermediate stages of the operation, equipment component failures may cause leaks. Thus, the leak rate can be considered to be variable. This is similar to the bathtub curve used in reliability engineering. It is intuitively supposed that when computing leak data, as shown in Fig. 2, the result of the early operation period is reliable, but the estimation of the late operation period is not very convincing. The conditionally autoregressive (CAR) model described in Kubo (2014), which is often used to describe spatial correlations, is suitable for application to these data (Barua, 2014).

It should be noted that the question is not whether the leak rate is constant. Perhaps, the method of estimation and modelling of the leak rate leads to the differences. The non-parametric method presented in Redbook treats "leak rate" as a constant value. However, the leak rate under certain conditions can change with time, which will not be taken into account in that method. If the constant value is sufficiently appreciated, time series analysis need not be conducted. To estimate the leak rate change over time, two time-based methods are adopted in this paper. Both methods use statistical models to describe the leak rate as a function of time (NIST, 2012).

3.2.1 Leak rate description by function of time: A log-normal function

First, a log-normal function is introduced that models the time-changing leak rate. The variable has a log-normal distribution if the logarithm of the variable follows the normal distribution.

The probability density function for log-normally distributed positive x is shown in Eq. (1):

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right), 0 < x < \infty \tag{1}$$

where,

μ - the log-normal distribution mean value parameter

σ - the log-normal distribution standard deviation parameter

x – a positive random variable

The accident rate $f(x)$ is described by multiplying the function in Eq. (1) by a coefficient a:

$$f(x) = a\frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2} \right), 0 < x < \infty \tag{2}$$

In Eq. (2), coefficient a is multiplied by the original distribution. The operation time and leak rate correspond to x and $f(x)$, respectively. A detailed explanation of the estimation is presented in Appendix A.

3.2.2 Leak rate description by function of time: Weibull function

Weibull distribution can take a more flexible shape of a graph than the log-normal distribution, even a nearly constant one. In this case, estimation is conducted with the Weibull function instead of the log-normal function. The probability density function for Weibull distributed positive x is shown in Eq. (3):

$$f(x) = \left(\frac{\alpha}{\beta}\right)\left(\frac{x}{\beta}\right)^{\alpha - 1} \exp\left(-\left(\frac{x}{\beta}\right)^{\alpha} \right) \tag{3}$$

where,

α - Weibull distribution shape parameter

β - Weibull distribution scale parameter

x – Positive random variable

The equation applied to estimate the leak rate using the Weibull function model is shown below:

$$f(x) = a\left(\frac{\alpha}{\beta}\right)\left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{x}{\beta}\right)^{\alpha}\right) \tag{4}$$

This is the function in Eq. (3) multiplied by coefficient a. As in Eq. (2), operation time and leak rate correspond to x and $f(x)$, respectively. A detailed explanation of the estimation is presented in Appendix B.

4. Results and Discussions

4.1 Leak rate estimation by time function: log-normal function

The result of leak rate estimation via the log-normal function is shown in Fig. 4. Although the estimated values are individual points for each month, they are represented using a smooth curve, as shown in the figure.

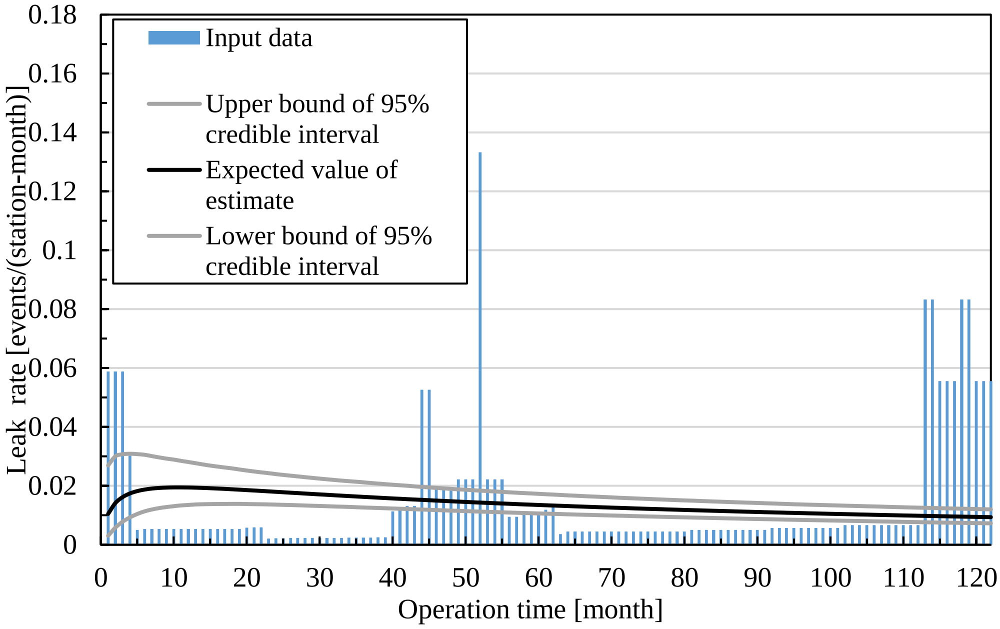

Fig. 4. Leak rate estimation via log-normal time functon

The horizontal axis in the graph of Fig. 4 is the operation time from the beginning of the HRS. The vertical axis is the leak rate, i.e., the average number of accidents per station per month. In the graph legend, the expected value of the estimate (shown in black) is the value that the leak rate is expected to follow, and the 95% interval is the range within which the leak rate is likely to lie with a 95% credibility. For example, from the graph, the estimated 10^{th} operation month's expected value of leak rate is 0.0194 [event per (station-month)]. In addition, because the lower bound of the 95% credible interval in the 10^{th} month is 0.01336 [event per (station-month)] and the upper bound is 0.0284 [event per (station-month)], there is a 95% probability of the leak rate having a value between these two bounds. As shown in the graph, the peak of the expected value of estimation falls on the 10^{th} and 11^{th} month. The leak estimate of 0.0132 [event per (station-month)] is estimated by the lognormal type function. This equates to 0.16 leaks per year.

4.2 Leak rate estimation by time function: Weibull function

The result of the Weibull type function estimation is shown in Fig. 5. The input data used are the same as that for log-normal function estimation. As with the log-normal curve, this estimation resulted in a smooth curve for the overall time.

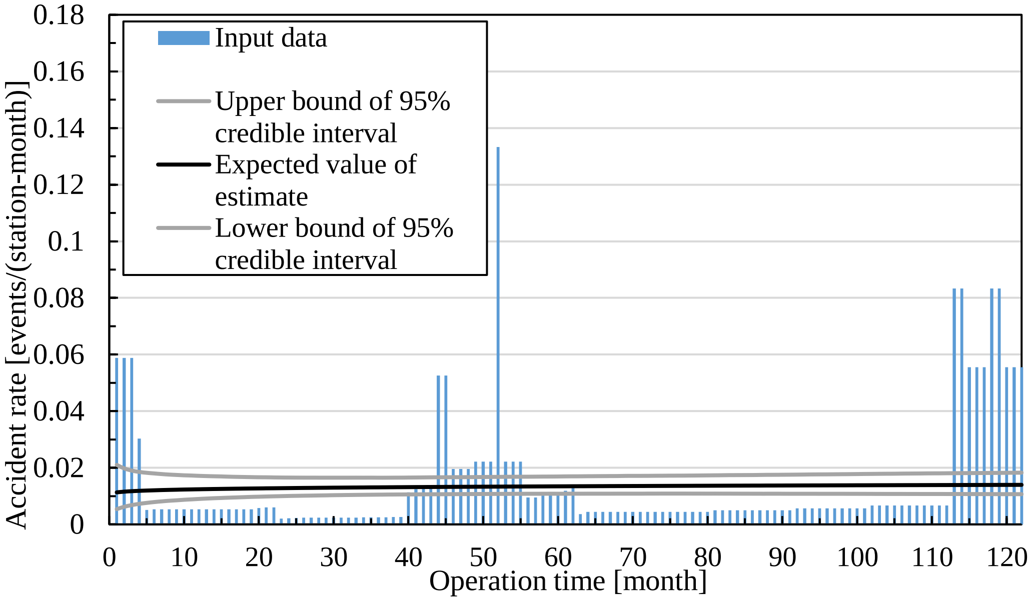

Fig. 5. Accident rate estimate using Weibull function of time

From Fig. 5, it can be observed that the expected value of estimation is virtually constant. Compared to the log-normal distribution, Weibull distribution's form is flexible in accordance with the parameter value. Thus, the steady result may suggest that the leak rate does not increase or decrease gradually but has a virtually constant value. It should be noted that the selected input data might have an impact on the expected value estimated as there are many periods with equal input data, for example, the data from the 5^{th} to the 19^{th} month. Processing of the input data may affect the result. A leak estimate of 0.0134 [event per (station-month)] is estimated by the Weibull type function.

4.3 Total leak rate estimation using a non-parametric approach with the Bayesian update

A non-parametric approach was employed by Kodoth et al. (2019) to estimate the failure data of the HRS. It adopted a non-parametric analysis to estimate the leak frequency as a function of the number of fillings using JHFC data for 35 MPa systems (JHFC, 2011). To be consistent with the time-based approach, the non-parametric failure analysis results for the 35 MPa systems will be used as the initial data and succeeded by the Bayesian update approach for the 70 MPa systems. This solution can help engineers to deal with data transformation from old system (e.g. 35MPa system) to new system (e.gg. 70 MPa system) when data is extremely limited. Bayesian update is used to update the initial failure information and provide updated failure data based on new observations and evidences. This study employs the Bayesian update in accordance with the Dutch model (Redbook, 1997). The probability density function of gamma distribution takes the form,

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(x)} x^{\alpha-1} e^{-\beta x} \quad \text{for } x > 0 \tag{5}$$

Here, α and β of the parameters can be determined by the mean and variance of the prior failure rate.

$$\alpha = \frac{E(\lambda)^2}{V(\lambda)} \quad (6)$$

$$\beta = \frac{E(\lambda)}{V(\lambda)} \quad (7)$$

The total failure rate $E(\lambda)$ of 6.7×10^{-4} per day and its variance $V(\lambda)$ of 6.0×10^{-8} per day estimated from Kodoth et al.'s (2019) findings will be used as prior knowledge in the Bayesian update. Substituting these figures in Eq. (6) and Eq. (7), the initial values of α and β are calculated to be 7.48 and 11166, respectively. Bayesian update is performed by calculating the parameters of posterior distribution, α' and β' based on new observations summarized in Table 1. The parameters can be updated as follows:

$$\alpha' = \alpha + n_f \quad (8)$$

$$\beta' = \beta + T_s \quad (9)$$

Here, n_f is the number of failures and T_s is the observed time. By using Eq. (6–9), the posterior distribution's mean and variance can be obtained from the updated parameters. Table 1 summarizes the number of accidents in the 70 MPa stations based on new observations. The observed time and number of leaks are collected from January 2011 to December 2015 extracted from the literature (KHK, 2015).

Table 1. Accidents in the 70 MPa hydrogen refueling stations in Japan from 2011 to 2015

ID	**Observed time [days]**	**Number of leaks**	**Leak rate [day^{-1}]**	
			E(λ)	V(λ)
1	30	2	8.4×10^{-4}	7.5×10^{-8}
2	50	1	7.5×10^{-4}	6.7×10^{-8}
3	11	2	8.4×10^{-4}	7.5×10^{-8}
4	409	1	7.3×10^{-4}	7.0×10^{-8}
5	126	1	7.5×10^{-4}	6.6×10^{-8}
6	2374	2	7.0×10^{-4}	1.0×10^{-8}
7	1678	1	6.6×10^{-4}	5.1×10^{-8}
8	1585	2	7.4×10^{-4}	5.8×10^{-8}
9	71	1	7.5×10^{-4}	6.7×10^{-8}
Sum.	6334	13	1.1×10^{-3}	6.6×10^{-8}

Table 1 shows the updated leak rate and its variance for each station. Since the number of leaks for each station is one or two, Bayesian update depends on the prior distribution. When the total number of leaks and observed time are used as specific data for the 70 MPa stations, the updated leak rate is 1.1×10^{-3} per day, which is about twice the leak rate estimated from other methods.

4.4 Summary of results from three methods

The results from time series method can be verified with those of the other two methods to make an engineering judgement on the leak rate estimation. The obtained results are summarized as follows:

1. Time-based method: The leak rate follows lognormal distribution with a mean value of 1.84×10^{-5} per hour. The estimated leak rate as a function of time is 0.16 per year.
2. Non-parametric (Bayesian) method: The leak frequency is estimated from failure data by means of

Bayesian update. The total leak rate using the non-parametric approach is estimated to be 1.1×10^{-3} per day, which is equivalent to 0.42 per year. This value is conservative and almost twice the leak rate compared to the results from other two methods. It should be noted that the initial data used in this approach and time based method is same however, the evidence (posterior data) of 70MPa for bayesian update is different from the time-based method. Based on the evidence, the result can vary by small to large margin.

3. Leak-hole-size method: The observed failure data (collated from the JHFC project report) is associated with leaks from threaded joints and seals. Under such conditions, it is assumed that the failure rate is equivalent to the leak frequency. Most of the failures are classified as “very small leak” of which the leak area is 0.01% of the total flow area, and the system frequency can be estimated to be 0.20 per year in line with the study by LaChance et al. (2009).

5. Unrevealed leak time forecast based on Leak Frequency Estimation

In general, an odorant is added to the gas (such as natural gas), to make it easy to detect leaks. However, pure hydrogen is used for FCV and it is not easy to detect a hydrogen leak. Then, the sensors used to detect hydrogen are used to detect leaks. Unrevealed leaks can occur at HRSs. There are two possibilities for hydrogen leak: either the leak will be detected by the hydrogen leak sensor within the inspection interval or the leak will not be revealed until the next scheduled inspection interval, as shown in Fig. 6. Based on these two possibilities, the parameter of interest from a safety point of view is unrevealed leak time.

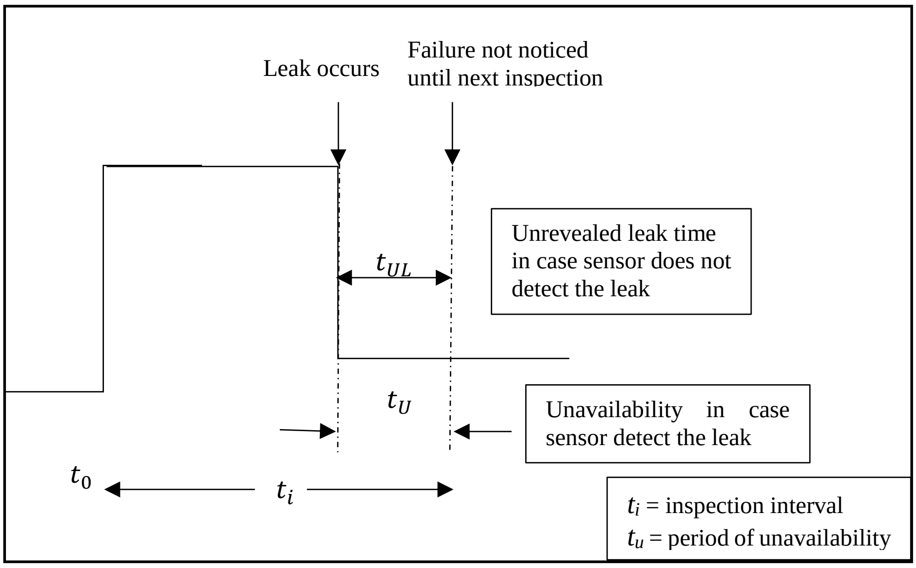

Fig. 6. Inspection cycle for revealed and unrevealed leaks

For unrevealed failures, the failures become obvious only after regular inspection. Unrevealed leak time is the difference from the point when the unrevealed leak occurs and the next inspection time. Failure probability is the measure of unreliability of the installation. The unavailability is the downtime of the process when the sensor detects the leak resulting in station shutdown.

Unrevealed leak time based on leak rate estimate

Using the leak estimate per station-month given by the lognormal type function from the previous section, the unrevealed leak time t_{UL} is obtained from the failure (leak) rate λ and inspection interval t_i.

$$t_{UL} = \int_0^{t_i} \lambda(t_i - t)dt = \frac{1}{2}\lambda t_i^2 \tag{10}$$

The leak rate estimated value from the previous section is used as a basis for λ. The unrevealed leak time forecast based on leak rate estimation and inspection interval is presented in Table 2.

Table 2. Unrevealed leak time forecasts based on leak rate and inspection interval

Methods	**Leak Rate (per year)**	**Inspection Interval**	**Unrevealed leak time**
Log-Normal (time-based)	0.16	Daily	19.08 s
Weibull (time-based)		Monthly	17043 s
Non-parametric Analysis	0.42	Daily	49.70 s
		Monthly	44738 s
Leak-Hole-Size Approach	0.20	Daily	23.67 s
		Monthly	21304 s

The unrevealed leak time is directly proportional to the leak rate and inspection interval. For example, in the case of the time series method, when the leak rate is 0.16 per year and the inspection interval is 24 h (daily inspection), the unrevealed leak time is 19.08 s. It means that hydrogen sensors are required to detect minor leaks at short intervals to reduce the unrevealed leak time. The unrevealed leak time using the non-parametric approach is estimated to be 49.70 s, which is conservative and almost twice the value compared to the results from other two methods. Perhaps this is due to the influence of evidence posterior data used in the non-parametric method for 70 MPa system. In leak hole size method, most of the failures are classified as "very small leak" of which the unrevealed leak time is estimated to be 23.67 s, which is significantly lower than the non-parametric method. In addition, for each method, the unrevealed leak time can increase drastically if the inspection interval is moved from daily to monthly routine. All these factors should be taken into account during the leak rate analysis and design of hydrogen sensors. The leak rate and unrevealed leak time data estimated with respect to inspection test in the paper can provide useful insights to engineers working in the reliability quantification of hydrogen energy system.

6. Conclusions

This paper examined the manner in which the leak rates are modelled using various methods. A time-based Bayesian estimate method was proposed, in which leak rates were modeled using operating time data on HRSs. One of the main results is that for the log-normal and Weibull models, the leak rate changes according to the time function. Parameters for the two statistical models were determined based on a Bayesian update. Even if accident events are rare, two statistical models can provide a range of leak rates as a function of time. The results from the time series method were then examined with other two methods to make an engineering judgement on the leak rate estimation.

To summarize, the leak rate is estimated to be 0.16 per year, 0.20 per year, and 0.42 per year based on the time-based, leak-hole-size, and non-parametric methods, respectively. It can be observed that even though the values do not exactly match, there is no large margin between the results obtained by the time-based and leak-hole-size methods. However, the leak rate obtained from the non-parametric method is the most conservative among the three. Perhaps, this is because of the more frequent failures observed in the new evidences for the 70 MPa system. The leak rate data from the time-series method shows a similar trend with the non-parametric and leak-hole-size method. The asset manager can select appropriate leak rate data based on the accident data and method availability. One of the possible solutions is to consider a conservative value for the design, in which case, the non-parametric model leak rate of 0.24 per year can be used. The base value selected can be used in design to set performance standards for the availability and reliability in the operation and maintenance of HRSs.

Unrevealed leak time was assessed from the estimated leak frequency. It can be concluded that if the leak rate is estimated to be high, the inspection interval should be more frequent to reduce the unrevealed leak time and increase the process safety. The unrevealed leak time can be used to the specification of hydrogen sensors to detect leaks of hydrogen. This will ensure the component and process both meet the requirements in the performance standard, leading to increased process safety in HRSs.

Acknowledgment

Funding: This research was supported by the Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), “Energy carrier” (Funding agency: Japan Science and Technology Agency (JST)). Authors would like to thank technical experts of IEA Hydrogen Safety Task 37 for fruitful discussion.

References

Barua, S., Basyouny, K., Islam, MT., 2014. A full Bayesian multivariate count data model of collision severity with spatial correlation. Anal Methods Accid Res, 2014; 3-4:28–43.

Bedrick, J., Christensen, R., Johnson, W., 2017. A new perspective on priors for generalized linear models. J. Am. Stat. Assoc. 1996; 91(436):1450–1460, http://www.jstor.org/stable/2291571/; [accessed September 2017].

Casamirra, M., Castiglia, F., Giardina, M., Lombarado, C., 2009. Safety studies of a hydrogen refueling station: Determination of the occurrence frequency of the accidental scenarios, International Journal of Hydrogen Energy, Vol. 34, pp. 5846-5854.

Redbook Dutch Model, 1997. Methods for determining and processing probabilities, Committee for the Prevention of Disasters (CPR 12E), The Hague, The Netherlands.

Gheriani, M., Khan, F., Chen, D., Abbassi, R., 2017. Major accident modelling using spare data, Process Safety and Environmental Protection, Volume 106, Pages 52-59, ISSN 0957-5820, https://doi.org/10.1016/j.psep.2016.12.004.

International Energy Agency (IEA), Technology Roadmap Hydrogen and Fuel Cells, 2015.

Japan Hydrogen & Fuel Cell Demonstration Project (JHFC), http://www.jari.or.jp/portals/0/jhfc/station/index.html; [accessed September 2017].

Japan Hydrogen and Fuel Cell Demonstration Project (JHFC Phase2), 2011. http://www.jari.or.jp/Portals/0/jhfc/data/report/pdf/tuuki_phase2_01.pdf

Japan Nuclear Technology Institute (JNTI). Estimation of domestic general equipment failure rate considering uncertainty of failure counts (1982~2002, for 21 years, 49 plants data), http://www.genanshin.jp/archive/failure_rate/; March 2009 [accessed September 2017].

Japan Nuclear Technology Institute (JNTI). Estimation of domestic general equipment failure rate considering uncertainty of failure counts (1982~2010, for 29 years, 56 plants data), http://www.genanshin.jp/archive/failure_rate/; June 2016 [accessed September 2017].

Jones, N., 1984. A Schematic Design for a HAZOP Study on a Liquid Hydrogen Filling Station. International Journal of Hydrogen Energy, Vol. 9, pp. 115-121.

Khalil, Y., 2017. A probabilistic visual-flowcharting-based model for consequence assessment of fire and explosion events involving leaks of flammable gases. Journal of Loss Prevention in the Process Industries 50 (2017) 190–204.

Kikukawa, S., Mistuhashi, H., Miyake, A., 2009. Risk assessment for liquid hydrogen fueling stations. International Journal of Hydrogen Energy, Vol. 34, pp. 1135-1141, 2009.

Kodoth, M., Aoyama S., Sakamoto, J., Kasai, N., Shibutani, T., Miyake, A., 2018. Evaluating uncertainty in accident rate estimation at hydrogen refueling station using time correlation model, International Journal of Hydrogen Energy, Volume 43, Issue 52, 2018, Pages 23409-23417, ISSN 0360-3199,https://doi.org/10.1016/j.ijhydene.2018.10.175.

Kodoth, M., Khalil, Y., Shibutani, T., Miyake, A., 2019. Verification of appropriate life parameters in risk and reliability quantifications of process hazards, Process Safety and Environmental Protection, 2019, ISSN 0957-5820, https://doi.org/10.1016/j.psep.2019.05.021.

Kubo, T., 2012. Introduction to statistical modeling for data analysis (generalized linear model, hierarchical Bayesian model, Markov chain Monte Carlo method). Tokyo: Iwanami; 2012.

LaChance, J., Houf, W., Middleton, B., Fluer, L., 2009. Analyses to support development of risk-informed separation distances for hydrogen codes and standards. SAND2009-0874 Sandia National Laboratories; 2009. http://prod.sandia.gov/techlib/access-control.cgi/2009/090874.pdf; [accessed September 2017].

Markowski, A., Siuta, D., 2017. Selection of representative accident scenarios for major industrial accidents, Process Safety and Environmental Protection, Volume 111, Pages 652-662, ISSN 0957-5820, https://doi.org/10.1016/j.psep.2017.08.026.

Ministry of Economy, Trade and Industry (METI). 2015. The condition of reviewing regulations corresponding to the demand of a new era (related to hydrogen and fuel cell vehicle), http://www.meti.go.jp/committee/sankoushin/hoan/koatsu_gas/pdf/007_05_01.pdf; March 2015 [accessed September 2017].

Ministry of Economy, Trade and Industry (METI). 2016. The project for investigating measures of promoting non-fossil energy introduction (formulating technological regulation of high-pressure gas for safe spread of new energy technologies), (3) Overseas survey, http://www.meti.go.jp/meti_lib/report/2016fy/000124.pdf; March 2016 [accessed September 2017].

Nakayama, J., Sakamoto, J., Kasai, N., Shibutani, T., Miyake, A., 2016. Preliminary hazard identification for qualitative risk assessment on a hybrid gasoline-hydrogen fueling station with an on-site hydrogen production system using organic chemical hydride. International Journal of Hydrogen Energy, Vol. 41, pp. 7518-7525.

Nelson, W., 2000. Theory and Applications of Hazard Plotting for Censored Failure Data, Technometrics, Vol. 42, No. 1, pp. 12–25.

NIST/SEMATECH, 2012. E-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/,April 2012.

Ntzoufras, I., 2009. Bayesian Modeling Using WinBUGS. John Wiley & Sons, Inc., Hoboken, NJ, USA, January 2009.

Pasman, H., Rogers, W., 2012. Risk assessment Risk assessment by means of Bayesian networks: A comparative study of compressed and liquefied H2 transportation and tank station risks. International Journal of Hydrogen Energy, Vol. 37, pp. 17415-17425, 2012.

Sakamoto, J., Sato, R., Nakayama, J., Kasai N., Tadahiro, S., Miyake, A., 2016. Leakage-type-based analysis of accidents involving hydrogen fueling stations in Japan and USA. Int. J. Hydrogen Energy, 2016; 41(46):21564–21570.

Sato, R., Nakayama, J., Sakamoto, J., Kasai, N., Shibutani, T., Miyake, A., 2016. Investigation of hydrogen leakage from joint part in hydrogen stand. The Collection of Presentation Papers at the Mechanical Engineering Congress 2016 Japan.

The California Energy Commission (CEC), 2004. Technical Consultant Report on Failure Modes and Effects Analysis for Hydrogen Fueling Options. CEC-600-2005-001, Nov 2004.

The High Pressure Gas Safety Institute of Japan (KHK), 2012. The high-pressure gas incidents database, https://www.khk.or.jp/english/accident_reports.html; [accessed Feb 2017].

The High Pressure Gas Safety Institute of Japan, 2015. Notes on Accidents related to Hydrogen Refueling Stations, 2015 (in Japanese).

Yamada, T., Kobayashi, H., Akatsuka, H., Hamada, K., 2015. Analysis of high-pressure gas incidents in hydrogen fueling stations. J High Press Gas Safety Institute Japan, 52 (10) (2015), pp. 23-29 [in Japanese].

Yang, M., Khan, F., Lye, L., 2013. Precursor-based hierarchical Bayesian approach for rare event frequency estimation: A case of oil spill accidents. Process Safety and Environmental Protection, Volume 91, Issue 5, Pages 333-342, ISSN 0957-5820, https://doi.org/10.1016/j.psep.2012.07.006.

Appendix

Appendix A. Detailed explanation of estimation using a log-normal time function

This appendix gives a detailed explanation of Section 3.2.1. In order to calculate accident rate over time, each parameter in Eq. (2) is estimated to describe the accident data. In Eq. (2), parameters a, σ, and μ are considered random variables. Bayesian statistics estimate a parameter's value by updating its distribution by some data; thus, each parameter is first given prior distribution. If prior information is available about these parameters, the prior distribution reflects this information and is usually called "informative prior" (e.g., (Bedrick, 1996)). In contrast, when there is no prior information, prior distribution with little information (e.g., normal distribution with mean zero and variance 10^4) is used as a "non-informative prior." In this case, no prior information is available, and thus, a non-informative prior is used.

The value for each parameter was estimated using the Bayesian statistics supporting software WinBUGS. To calculate the posterior distribution by updating the prior distribution with the accident data, WinBUGS uses Markov Chain Monte Carlo simulation, and it needs an initial value for each parameter. An appropriate initial value was chosen by judgement or automatically selected by software, and it was checked for calculation errors.

For this estimation, the following Bayesian model is introduced. Note that other methods such as least squares fitting can also suffice and so there is no special reason to use the Bayesian model. However, using the Bayesian model often enables complex modeling and utilization of other information in addition to the observed data. The time-series accident rate is described by following logical relationship:

$$\overline{\lambda}_j = a \frac{1}{\sqrt{2\pi}\sigma t_j} \exp\left(-\frac{(\ln t_j - \mu)^2}{2\sigma^2} \right) \tag{A1}$$

where,

$\overline{\lambda_j}$: expected value of accident rate for the j^{th} month

a: coefficient

σ, μ: parameters of the log-normal function

t_j: j^{th} operation time

j: index of the operation time

Each month's accident rate is considered as a random variable following the log-normal distribution below:

$$\lambda_j \sim LN(\mu_{2,j}, \tau) \tag{A2}$$

where,

λ_j: accident rate of the j^{th} operation time t_j

$LN(\mu_{2,j}, \tau)$: log-normal distribution with mean $\mu_{2,j}$ and inverse square of standard deviation τ

$\mu_{2,j}$: expected value of the log-normal distribution of the accident rate for the j^{th} operation month

τ: inverse square of the standard deviation of the log-normal distribution

To connect Eq. (A1) and (A2), the relation between parameter $\mu_{2,j}$ and the expected value of accident rate $\overline{\lambda_j}$ is utilized as follows:

$$\mu_{2,j} = \ln\left(\overline{\lambda}_j\right) - \frac{1}{2\tau} \tag{A3}$$

Prior distribution ("non-informative prior") of each parameter is set as follows:

$a \sim Gamma(1,1)$

$\tau \sim Gamma(1,1)$

$\sigma \sim Unif\,(0,10)$

$\mu \sim Unif\,(0,10)$

where,

$Gamma(a,b)$: gamma distribution with shape parameter a and rate parameter b

$Unif\,(a,b)$: uniform distribution with lower bound a and upper bound b

Using this statistical model and the accident data, the accident rate was estimated as shown in Fig. 4.

Appendix B. Detailed explanation of estimation using a Weibull time function

This appendix gives a detailed explanation of Section 3.2.2. It differs from Appendix A in its description of the time-series accident rate and each parameter's prior distribution and initial value, but the flow of modeling and estimation are virtually the same as in Appendix A. Firstly, the time-series accident rate is described by following logical relationship:

$$\overline{\lambda}_j = a\left(\frac{\alpha}{\beta}\right)\left(\frac{t_j}{\beta}\right)^{\alpha-1} \exp\left(-\left(\frac{t_j}{\beta}\right)^{\alpha}\right) \tag{B1}$$

where,

$\overline{\lambda_j}$: expected value of the accident rate for the j^{th} month

a: coefficient

α, β: parameters of the Weibull function

t_j: j^{th} operation time

j: index of the operation time

Each month's accident rate is considered as a random variable following the log-normal distribution below:

$$\lambda_j \sim LN(\mu_{2,j},\tau) \tag{B2}$$

where,

λ_j: accident rate of operation time t_j

$LN(\mu,\tau)$: log-normal distribution with mean μ and inverse square of standard deviation τ

$\mu_{2,j}$: expected value of the log-normal distribution of accident rate for j^{th} operation month

τ: inverse square of the standard deviation of the log-normal distribution

To connect Eq. (B1) and (B2), the relation between parameter $\mu_{2,j}$ and the expected value of accident rate $\overline{\lambda_j}$ is utilized as follows:

$$\mu_{2,j} = \ln\left(\overline{\lambda}_j\right) - \frac{1}{2\tau} \tag{B3}$$

Prior distribution ("non-informative prior") of each parameter is set as follows:

$a \sim Gamma(1,0.00001)$

$\alpha \sim Gamma(0.1,0.00001)$

$\beta \sim Gamma(0.1,0.00001)$

$\tau \sim Gamma(1,0.00001)$

where,

Gamma(*a*,*b*): gamma distribution with shape parameter *a* and rate parameter *b*

Using this statistical model and the accident data, the accident rate was estimated, as shown in Fig. 5.