DEVELOPMENT OF LIQUID HYDROGEN LEAK FREQUENCIES USING A BAYESIAN UPDATE PROCESS

Brooks, D.M.[1], Ehrhart, B.D.[2], and LaFleur, C.[3]
[1] Structural & Thermal Analysis, Sandia National Laboratories, PO Box 5800, MS 0748, Albuquerque, NM, 87185-0748, USA, dbrooks@sandia.gov
[2] Fire, Risk & Transportation Systems, Sandia National Laboratories, PO Box 5800, MS 0748, Albuquerque, NM, 87185-0748, USA, bdehrha@sandia.gov
[3] Fire, Risk & Transportation Systems, Sandia National Laboratories, PO Box 5800, MS 0748, Albuquerque, NM, 87185-0748, USA, aclafle@sandia.gov

ABSTRACT

To quantify the risk of an accident in a liquid hydrogen system, it is necessary to determine how often a leak may occur. To do this, representative component leakage frequencies specific to liquid hydrogen can be determined as a function of the normalized leak size. Subsequently, the system characteristics (e.g., system pressure) can be used to calculate accident consequences. Operating data (such as leak frequencies) for liquid hydrogen systems are very limited; rather than selecting a single leak frequency value from a literature source, data from different sources can be combined using a Bayesian model. This approach provides leakage rates for different amounts of leakage, distributions for leakage rates to propagate through risk assessment models to establish risk result uncertainty, and a means for incorporating liquid hydrogen-specific leakage data with leakage frequencies from other fuels. Specifically, other cryogenic fluids like liquefied natural gas are used as a baseline for the Bayesian analysis. This Bayesian update process is used to develop leak frequency distributions for different system component types and leak sizes. These leak frequencies can be refined as liquid hydrogen data becomes available and may then inform safety code requirements based on the likelihood of liquid hydrogen release for different systems.

INTRODUCTION

Risk assessments are generally used to analyze a specific system by identifying potentially high-risk activities, scenarios, or components. A risk assessment can also be used to quantify the risk of an entire facility, which enables comparisons to alternative options. Alternatively, risk assessment can be performed with a representative facility in order to better inform regulations, codes, and standards requirements and attempt to improve safety for many facilities at once. For example, this was done in the U.S. National Fire Protection Association (NFPA) fire codes by LaChance et al. [1] for gaseous hydrogen bulk storage systems. In this case, a quantitative risk assessment was used to estimate the risk for a representative system, which was then used by the code committee (along with information about leak frequencies in particular) to select a design leak scenario by which certain code requirements would be determined. A similar analysis is needed for liquid hydrogen systems to better inform similar code requirements.

To quantify the risk of an accident in a liquid hydrogen system, it is necessary to determine how often a leak may occur. To do this, representative component leakage frequencies specific to liquid hydrogen can be determined as a function of the normalized leak size. Normalized leak sizes are used because component sizes within and between systems vary and there are insufficient data to derive leak frequencies for each size of each type of component. The intent of the normalized leak size is to enable leak frequency estimation at the component level. Once leak frequency estimates are established, system characteristics (e.g., system pressure) can be used to calculate accident consequences.

Operating data (such as leak frequencies) for liquid hydrogen systems are very limited, which presents a challenge for leak frequency estimation. In this analysis, we use leak frequency data from similar fuels to demonstrate the estimation methodology, as well as the need for liquid hydrogen system data to improve the estimates. Rather than selecting a single leak frequency value from a literature source, data from different sources were combined using a Bayesian model, as in previous analyses [1, 2, 3]. This approach provides leakage rates for different amounts of leakage (characterized by the fractional leak area), uncertainty distributions for leakage rates to propagate through risk assessment models, and a means for incorporating liquid hydrogen-specific leakage data with leakage frequencies from other fuels.

Gaseous hydrogen and other cryogenic flammable fluids like liquefied natural gas were used as a baseline for the Bayesian analysis. The assumption underlying use of this data is that these systems and the mechanisms causing leaks within them should be similar to those in liquid hydrogen systems. Hence, these analogous systems can be used to inform belief about how liquid hydrogen systems behave with respect to leaks, but the assumption remains untested until more liquid hydrogen data are obtained.

The Bayesian update process is used to develop leak frequency distributions for different system component types and leak sizes. Because data for gaseous hydrogen systems and liquefied natural gas systems were both used, only components common to both systems were analyzed. The resulting leak frequencies can be refined as liquid hydrogen data becomes available and may then inform safety code requirements based on the likelihood of liquid hydrogen release for different systems.

MATHEMATICAL CHARACTERIZATION

The goal of this analysis is to obtain estimates of liquid hydrogen leak frequencies using limited data. At its core, this is a problem of statistical inference. Operating experience from systems for different types of fuel provides data on leaks which is assumed to be representative of the future operation of similar systems. Though data are not available for all systems, this analysis treats the available data as a representative sample and uses it to infer conclusions about the expected annual leak frequency for a typical system. The representativeness of available data is a key assumption underlying this analysis which cannot yet be shown to be true but can be re-assessed and improved as future data becomes available.

Given the sparsity of data, the goal is not just to obtain single point leak frequency estimates. Such estimates, while necessary for some types of risk analysis, should be interpreted with uncertainty. While this analysis seeks to make use of limited data, the methods cannot alleviate the need for more high-quality data. Hence, this is not just a problem of inference, but also a problem of uncertainty quantification.

Uncertainty has classically been dichotomized into epistemic uncertainty and aleatory uncertainty. Epistemic uncertainty (usually considered hypothetically reducible) describes the state-of-knowledge about something that is assumed to have a fixed value, whereas aleatory uncertainty describes inherent variability that cannot be reduced by gaining knowledge. Uncertainty in leak frequencies for liquid hydrogen systems is largely philosophically epistemic; components in these systems will leak with some frequency but that frequency is not known exactly. However, because the analysis estimates industry-wide frequencies, aleatory uncertainty may also be present. There is inherent variability between different types of systems, sizes of components, and individual sites which likely lead to different actual leak frequencies. As long as the leak frequency estimate is meant to be applied industry-wide, it will contain these inherent differences [4].

For the purposes of this analysis, the epistemic and aleatory uncertainties are not analyzed separately; the uncertainty is all treated mathematically as if it is epistemic, though it is a combination of aleatory and epistemic uncertainties. Characterization of epistemic uncertainty can be accomplished in multiple ways, which can be categorized broadly by probabilistic methods or non-probabilistic methods. The non-probabilistic category includes methods such as interval analysis, possibility theory, and evidence theory and the probabilistic category can be divided into frequentist (i.e. classical statistics) or Bayesian methods [4]. Probabilistic methods are ideal for this analysis because the goal is to provide leak frequencies that can be used in the context of risk analysis, which is often probabilistic. However, the choice between frequentist and Bayesian methods in this case is less straightforward.

Bayesian Framework

Bayesian statistics are not altogether different from frequentist statistics; the theorem underlying the field is a basic result within frequentist statistics. Bayes' Theorem is a statement of conditional probabilities, which in its simplest form, states that:

$$p(\theta = \theta^* | x) \propto p(x | \theta = \theta^*) p(\theta = \theta^*) \tag{1}$$

where p denotes probability, θ is a parameter characterized as a random variable, θ^* is a specific value of the parameter θ, and x denotes observed data. In essence, the theorem states that we can understand the probability of a specific parameter value given data by decomposing the problem into 1) the probability of the data assuming that specific parameter value, and 2) the probability that the parameter takes on that specific value. In Eq. (1), $p(\theta = \theta^* | x)$ is called the posterior, $p(x | \theta = \theta^*)$ is the likelihood, and $p(\theta = \theta^*)$ is the prior. Typically, the posterior is what we want to understand, the likelihood is the information we can observe, and the prior is our initial state of belief about θ [5]. The prior allows objective information to be included when needed, as is the case when supplementing data with expert judgement.

With respect to this leak frequency analysis, however, we do not want to supplement objective leak frequency data from leak events with subjective judgements. Instead, we chose to use Bayesian methods because of how uncertainty is characterized. Note in Eq. (1) that the prior does not just enable subjective information to be included; the prior assumes that that parameter θ is itself a random variable. This assumption is the vehicle through which the subjective information is included.

The treatment of parameters as random variables is the key conceptual difference between frequentist and Bayesian methods that leads us to prefer Bayesian methods for leak frequency estimation with uncertainty. In a frequentist framework, the parameter θ above would be considered to have a fixed value that is unknown. By collecting data, we could hypothetically learn its true value. In the Bayesian framework, θ is treated as a random variable with a distribution rather than a fixed value. By collecting data, we could hypothetically learn its true distribution. This concept leads to hierarchical models in Bayesian statistics, which are useful tools for uncertainty propagation and can be calibrated using Bayes' theorem.

The specific model applied for leak frequency estimation is described in the next section, but the concept of the hierarchical model is as follows. Leak frequencies are assumed to arise from a specific model and that model defines the base of the hierarchy. We can collect data (leak frequencies from literature) at this level. The next level of the hierarchy consists of distributions on the parameters of the leak frequency model and these distributions have their own parameters, which can either have fixed values or correspond to another level of distributions. The parameters at the highest level of the hierarchy have fixed values, which are estimated using Bayes' theorem. Hence, even when the model is calibrated, it propagates uncertainty from the top of the hierarchy down through the base model, which in our case describes leak frequencies.

The reader is referred to [5, 6] for more detailed treatments of Bayesian statistics.

Mathematical Model

The mathematical model employed in this analysis assumes a normal distribution on the log-leak frequency for each fractional leak area. The means of these distributions for each fractional leak area are assumed to be related to each other linearly in log-space, as illustrated in Fig. 1. The normal distribution and linear relationship define the base of the hierarchical model which is assumed to govern the leak frequency data in the literature. Uncertainty in the model is incorporated via the assumed normal distribution; uncertainty is included in the mean through uncertainty distributions on the parameters of the line and through uncertainty distributions on the precision of the normal distribution at each fractional leak area.

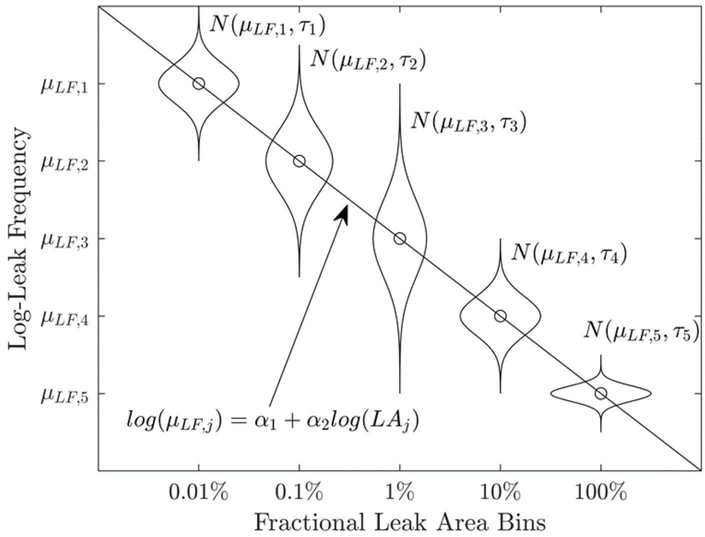

Figure 1. The mathematical model applied to estimate leak frequencies assumes 1) a linear relationship in log-space between the fractional leak area bins and 2) normal distributions on the log-leak frequency within each bin. Though the means of the distributions are related linearly, the precisions are not.

The distribution on the log leak frequency is defined as:

$$\log(LF_j) \sim normal(\mu_{LF,j}, \tau_j), \tag{2}$$

where LF_j is the leak frequency for the j^{th} leak frequency bin, $\mu_{LF,j}$ is the mean of the normal distribution on $\log(LF_j)$, and τ_j is the precision of that normal distribution. The $\log(\mu_{LF,j})$ is related to the log leak area ($\log(LA_j)$) linearly by:

$$\log(\mu_{LF,j}) = \alpha_1 + \alpha_2 \log(LA_j), \tag{3}$$

where α_1 is the intercept of the line and α_2 is the slope. Uncertainty is propagated through this linear relationship, and hence through the normal distribution, by assigning uncertainty distributions to these linear parameters as follows:

$$\alpha_1 \sim normal(\alpha_{11}, \alpha_{12}) \tag{4}$$

$$\alpha_2 \sim normal(\alpha_{21}, \alpha_{22}) \tag{5}$$

where α_{11} and α_{21} are the means of the respective normal distributions and α_{12} and α_{22} are the precisions.

Finally, uncertainty is also included in the precision of each normal distribution:

$$\tau_j \sim gamma(s_j, r_j), \tag{6}$$

where s_j and r_j are the shape and rate parameters of the gamma distribution on τ_j.

With this mathematical structure, the normal distributions fit for the log leak frequency at each leak size will be different, but they are not wholly independent due to the assumption of a linear relationship in log space between the means. The distributions on α_1, α_2, and τ_j must be specified as priors and leak frequency data will be used to estimate final posterior distributions for α_1, α_2, and τ_j, which will be propagated through the model to produce final distributions on leak frequency. The model is presented in Fig 2.

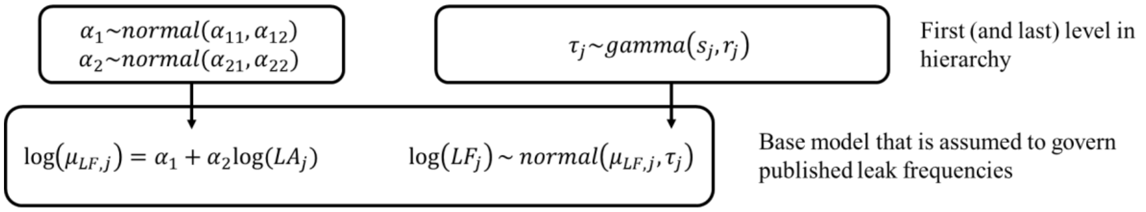

Figure 2. A simple hierarchical model applies a normal distribution to model the log leak frequency with a single level above that incorporates uncertainty.

The distributions selected for the priors were also chosen based on the previous work [2, 3]. They were defined as:

$$\alpha_1 \sim normal(0, 10^{-3}) \tag{7}$$

$$\alpha_2 \sim normal(0, 10^{-3}) \tag{8}$$

$$\tau_j \sim gamma(4, 1) \tag{9}$$

The first two priors can be described as uninformed priors; they are centered at zero with low precision. This means that the state-of-belief as input to the model allows α_1 and α_2 to be positive or negative and allows their values to fall within a wide range of magnitudes. The distribution on τ_j is also uninformed in that it is defined to represent very little knowledge. A $gamma(1,1)$ distribution would be a reasonable uninformed prior for this model as it would concentrate values of the precision closer to zero, but with a wide enough domain to allow for higher precision in the final model if suggested by the data. However, this distribution assumes too low a precision for the model to converge with limited data. Therefore, the first parameter was incremented, as has been done in previous leak frequency analyses [1, 2, 3], to shift the distribution towards higher precisions just until the model could converge.

Once the priors were specified, multiple data sets were used to calibrate the model in sequential updates. This is because there were multiple data sets to inform the model. This is a repetition of the same calibration process except that the uninformed priors were only used for the first update; subsequent updates used the posteriors from the previous update as informed priors.

The leak frequency model was implemented in the R programming language using the rjags package to call JAGS (Just Another Gibbs Sampler) [7, 8, 9]. The sampling was performed using 5×10^5 samples for initialization, 5 parallel chains, and 1×10^5 samples to update the model. These sample sizes were used for each sequential update. A final 1×10^5 samples were taken from the posterior distributions and of the predicted leak frequencies. These sample sizes were shown to be more than sufficient in a previous analysis that applied this model to estimate leak frequencies for liquid natural gas systems [3].

DATA

There is a need for leak frequency estimates that can be used in risk assessments for liquid hydrogen systems. However, data from liquid hydrogen systems are not currently available to inform these leak frequency estimates. Until such data can be obtained, the estimates rely on data from systems that were assumed to be informative and that have already been collected for previous analyses [1, 2].

The first data set contains two categories of data: 1) generic data from the chemical processing, compressed gas, nuclear power plant, and offshore petroleum industries, and 2) data from gaseous hydrogen systems. These data were originally compiled for the analysis in LaChance et al. [1] and were again used for Glover et al. [2]. They are applied within this analysis for the first update in the leak frequency model calibration. These leak frequency data for gaseous hydrogen are informative to the current analysis because they contain hydrogen-specific data. Thus, the leak-related behavior of gaseous hydrogen, containment material selection, and system design help inform the current analysis for liquid hydrogen systems.

The second data set used in this analysis was originally compiled for estimation of liquid natural gas system leak frequencies in Mulcahy et al. [3]. This data set also contains two categories of data: 1) data from non-liquid natural gas systems that are judged in the original source to be applicable to liquid natural gas systems, and 2) data that originates from liquid natural gas systems. These leak frequency data for liquefied natural gas are informative to the current analysis because they contain cryogenic-specific data. This means that cryogen-specific behavior, operating temperatures, and system design help inform the current analysis for liquid (cryogenic) hydrogen systems.

This analysis estimates leak frequencies only for system components that are represented in the data sets for both updates: flanges and gaskets, hoses, joints, pipes, valves, and vessels. Results for each of these components are presented in the following sections and Table 1 describes the number of data points in each data set for each component.

Table 1. Number of data points in each of the two data sets used to sequentially update the leak frequency model.

Component	**Points in 1st Data Set**	**Points in 2nd Data Set**
Flanges and Gaskets	17	32
Hoses	12	16
Joints	12	6
Pipes	56	86
Valves	35	61
Vessels	9	168

RESULTS

Leak frequency distribution estimates are plotted in Fig. 3 through Fig. 8. The distributions are represented using violin plots, which are rotated and mirrored estimates of the probability density functions estimated using 10^5 leak frequency samples at each fractional leak area. Note that the model is described for the log leak frequency but the results are plotted for the leak frequency. The log leak frequency samples are exponentiated to convert them into leak frequency samples, but the distributions of these exponentiated samples still appear normal in plots because they are plotted on a log scale.

The data points from each data set are shown as separate markers on each of the violin plots. The two data sets do not appear to trend in one direction in particular; as seen in Fig. 3, the GH2 (Data Set 1) and LNG (Data Set 2) leak frequencies seem to be interspersed. Importantly, neither data set contains frequencies for every leak size within every component. For example, the two smallest leak sizes (0.01% and 0.1%) for flanges (Fig. 3) appear to only have frequencies from LNG (Data Set 2). Thus, the resulting leak frequency distributions may be different for the combination of both data sets compared to distributions resulting from either data set individually.

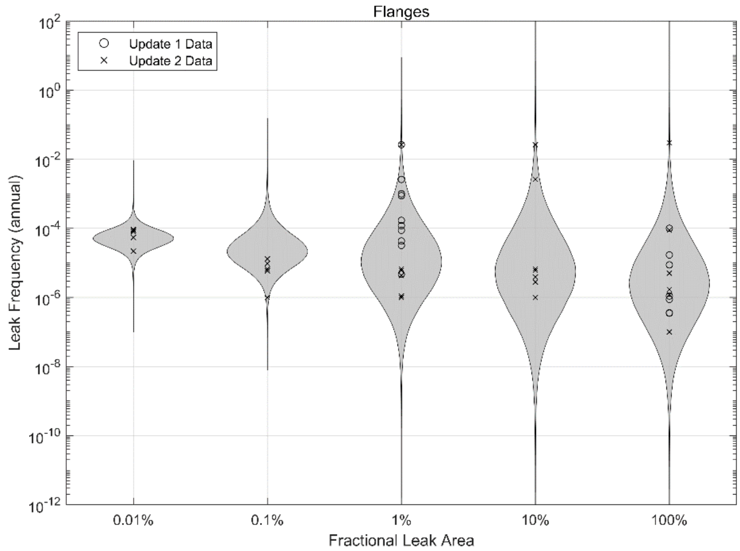

Figure 3. The final estimates for flanges in liquid hydrogen systems demonstrate decreasing leak frequency with respect to leak size, but more variation in estimates for larger leaks. The increased spread in distributions for large leaks compared to small leaks may be due to the limited quantity of small leak data.

The distribution estimates for flanges (Fig. 3), pipes (Fig. 6), valves (Fig. 7), and vessels (Fig. 8) show that large leaks occur less frequently than small leaks. This is consistent with the intuition that systems are designed specifically to prevent large leaks, so they may occur less by design. However, the opposite trend is seen for hoses (Fig. 4) and joints (Fig. 5). It is not clear whether the increasing frequency of leaks relative to leak size for these components is a realistic result reflecting physical differences between these components and the others, or whether this result arises due to insufficient data. Neither conclusion can be drawn without additional data, so leak frequencies for these components should be used cautiously.

In addition to the general linear trends in model results, some of the distributions have unusually long tails. While normal distributions are defined on an infinite domain, these distributions are estimated using samples and the tails of these distributions should not be sampled frequently with just 10^5 samples. This result arises due to the form of some of the data in the first data set. Most of the data for this model are leak frequency estimates. However, the gaseous hydrogen data set was originally in the form of operational time and number of leaks of each size. Such data were included in the model by converting into maximum likelihood estimates of leak frequency. This results in a data point at 0 annual leak frequency for some leak sizes, which drives the overall precision down so the spread in the final estimate is high. Though this is not ideal, as it likely overestimates uncertainty, hydrogen data are so uncommon in the open literature that it is reasonable to accept the cost of overestimating uncertainty for the benefit of more accurately estimating the centers of the distributions specifically for hydrogen systems.

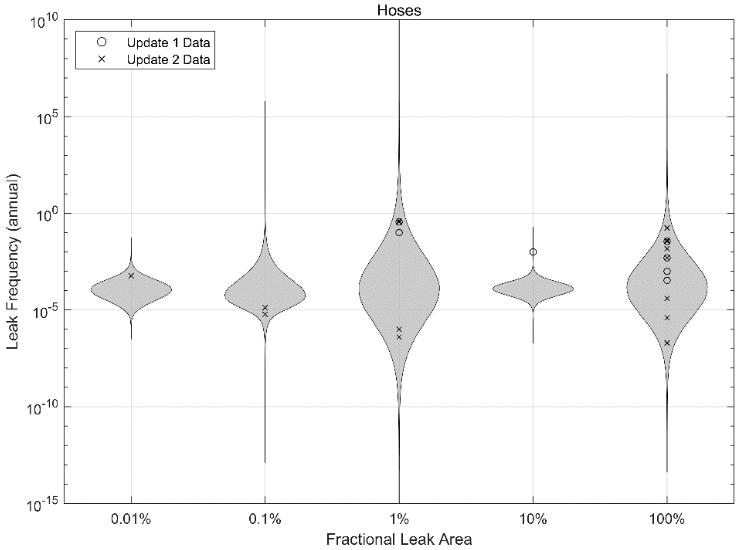

Figure 4. The final estimates for hoses in liquid hydrogen systems demonstrate a slightly increasing leak frequency with respect to leak size. This was also observed in results for liquid natural gas systems in Mulcahy et al. [3]. This trend may be explained by the sparsity of data or may reflect physical phenomenology that causes leaks in hoses to quickly expand compared to components. It is not clear which interpretation is appropriate; liquid hydrogen system data may shed light on this in the future.

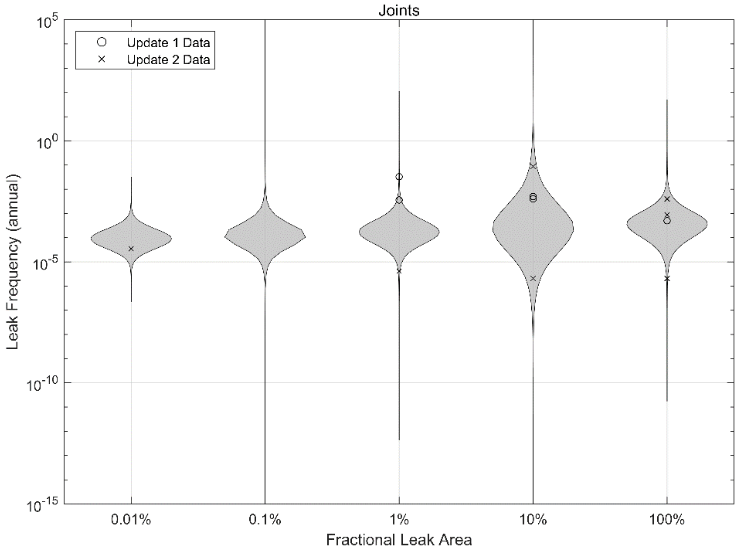

Figure 5. The final estimates for joints in liquid hydrogen systems demonstrate a similar trend of increasing leak frequency for large leaks as seen for hoses. As is the case for hoses, there is insufficient data and context to determine whether large leaks are truly more frequent than small leaks in joints, or whether this result is due to insufficient data.

The data points from Data Sets 1 and 2 appear to be fairly interspersed for most of the components and leak sizes; that is, there does not appear to be a clear trend in which frequencies from one data set are always higher or lower than the other. One possible exception is for Pipes; as shown in Fig. 6, the frequencies from Data Set 2 (LNG) appear to generally be lower than the frequencies from Data Set 1 (GH2). This type of trend can have two separate effects; first, the center (median) of the combined distribution is likely to be different than the distributions for each of the data sets individually. Second, the uncertainty spread of the distribution is likely to be wider than for each of the data sets individually. It is not clear exactly why the leak frequencies are different for pipes and not for other components. Potential contributing causes are hydrogen embrittlement causing more leaks for GH2 and vacuum-jacketed pipes leaking less often for cryogenic LNG. However, it is expected that other components would show similar trends if these are the only causes. This is a specific instance in which liquid hydrogen-specific data are likely to be especially informative, since the two proxy data sets appear to trend to different values.

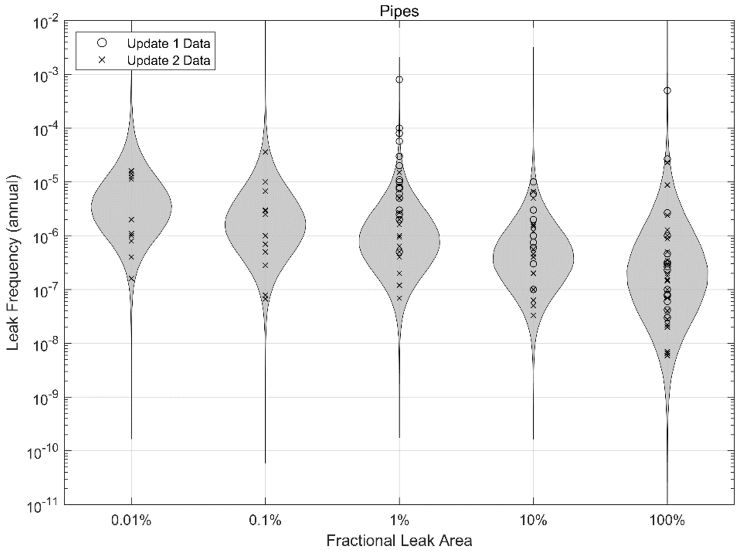

Figure 6. The final estimates for pipes demonstrate the expected trend with large leaks being less frequent than small leaks. These results appear consistent with the observed data and there is a sufficient quantity of data to conclude that the relationship is likely accurate.

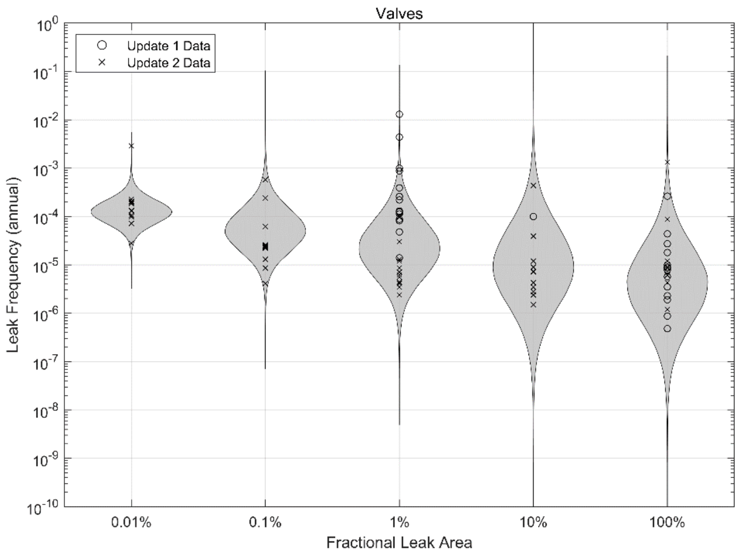

Figure 7. The final estimates for valves in liquid hydrogen systems reflect a similar trend as for pipes, though there is increasing variation in the leak frequency for larger leaks. This variation may be due to the gaseous hydrogen data, which included maximum likelihood estimates of 0 annual leaks for leaks with a fractional leak area of 1% or greater. Though these data are not plotted on the log scale, they are

included in the calibration and increase the spread in the final leak frequency distribution in both directions due to the normal distribution symmetry.

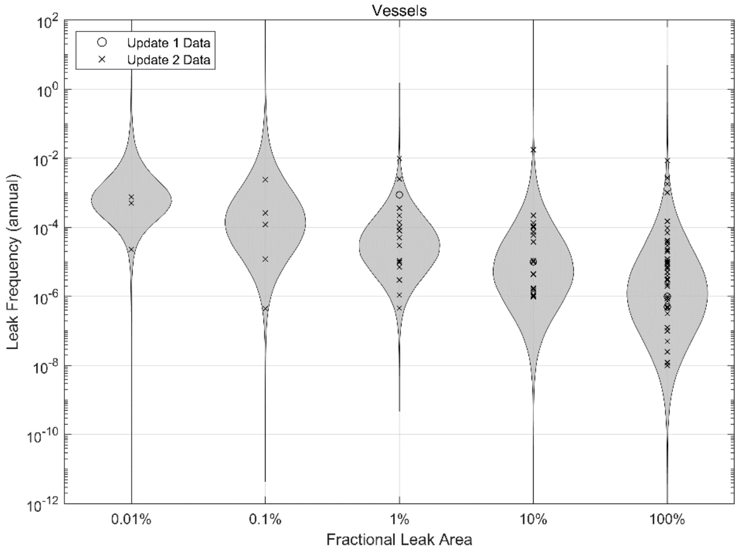

Figure 8. The final estimates for vessels in liquid hydrogen systems demonstrate a similar trend of decreasing frequency with respect to leak size as the results for pipes and valves. As with those components, the data appear consistent with this trend and numerous enough to judge it reasonable. The long tails on these distributions are also the result of maximum likelihood estimates of 0 from the gaseous hydrogen data in prior analyses [1, 2].

Statistics from the distributions plotted in Fig. 3 through Fig. 8 are included below in Table 2. Because the normal distribution is symmetrical about the log leak frequency, it is asymmetrical in linear space. This means that the arithmetic mean of samples from this distribution is likely biased towards the upper ends of the leak frequency distributions, whereas the median is a more appropriate representation of the center of each distribution. Hence, we provide the median rather than the mean.

Table 2. Statistics from the leak frequency distributions estimated for each component can be used in liquid hydrogen system risk assessment, though the accuracy of these estimates is difficult to judge without data from liquid hydrogen systems for validation (and improved calibration).

Component	Leak Size	5th	Median	95th	Component	Leak Size	5th	Median	95th
Flange/ Gasket	0.01%	1.6E-05	5.0E-05	1.4E-04	Pipe	0.01%	3.3E-07	3.4E-06	3.6E-05
	0.1%	3.1E-06	2.2E-05	1.8E-04		0.1%	1.4E-07	1.6E-06	2.0E-05
	1%	1.7E-07	1.1E-05	6.6E-04		1%	7.4E-08	8.0E-07	8.5E-06
	10%	3.3E-08	5.0E-06	6.9E-04		10%	3.5E-08	3.9E-07	4.2E-06
	100%	1.8E-08	2.4E-06	2.8E-04		100%	6.3E-09	1.9E-07	5.4E-06
Hose	0.01%	1.3E-05	1.1E-04	7.3E-04	Valve	0.01%	4.6E-05	1.2E-04	3.4E-04
	0.1%	5.7E-06	9.0E-05	4.3E-03		0.1%	8.8E-06	5.2E-05	3.2E-04
	1%	6.1E-08	1.2E-04	2.2E-01		1%	2.1E-06	2.2E-05	2.4E-04
	10%	3.1E-05	1.2E-04	4.9E-04		10%	3.3E-07	9.3E-06	2.5E-04
	100%	2.9E-07	1.3E-04	5.6E-02		100%	1.5E-07	3.9E-06	9.6E-05
Joint	0.01%	1.9E-05	9.0E-05	4.3E-04	Vessel	0.01%	4.3E-05	6.5E-04	1.3E-02
	0.1%	1.4E-05	1.2E-04	1.1E-03		0.1%	2.1E-06	1.4E-04	8.9E-03
	1%	2.2E-05	1.7E-04	1.3E-03		1%	9.6E-07	2.8E-05	8.1E-04
	10%	2.2E-06	2.4E-04	2.3E-02		10%	6.4E-08	5.7E-06	4.9E-04
	100%	2.6E-05	3.3E-04	3.3E-03		100%	6.8E-09	1.2E-06	2.0E-04

CONCLUSIONS AND FUTURE WORK

Liquid hydrogen leak frequencies have been estimated using a combination of proxy data for gaseous hydrogen and liquefied natural gas. This was done with a hierarchical model using a Bayesian update process with two data sets: one based on previous work for gaseous hydrogen and another based on previous work for liquefied natural gas.

The distribution estimates for flanges, pipes, valves, and vessels are consistent with the intuition that systems are designed against large leaks (large leaks being more risk-significant), and so large leaks occur less frequently in industry experience than small leaks. However, the opposite trend is seen for hoses and joints; larger leaks are slightly more common than smaller leaks. It is not clear whether the increasing frequency of leaks relative to leak size for these components is a realistic result reflecting physical differences between these components and the others, or whether this result arises due to insufficient data. Neither conclusion can be drawn without additional data, so leak frequencies for these components should be used cautiously.

The data points from both gaseous hydrogen and liquefied natural gas appear to be fairly interspersed for most of the components and leak sizes; that is, there does not appear to be a clear trend in which leak frequencies from one data set are always higher or lower than for the other. One possible exception is for pipes; the leak frequencies from the liquefied natural gas data set appear to generally be lower than the frequencies from the gaseous hydrogen data set. This affects both the center (median) of the estimated distribution, as well as increasing the uncertainty spread. This is a specific instance in which liquid hydrogen-specific data is likely to be especially informative, since the two proxy data sets appear to trend to different values.

The most important future work to follow on this analysis is to update these distributions with liquid hydrogen-specific data. The data used in this analysis are proxy data meant to estimate the effects of liquid hydrogen component leaks, but data from liquid hydrogen systems specifically would be much more accurate. The leak frequency distributions estimated in this analysis can be used as the priors and updated with the specific data in order to obtain the appropriate posterior distributions. However, not all components are included in this analysis; pumps, for example, were not included in the liquefied natural gas data set, and so were not included in this analysis. An alternative prior would need to be identified for these types of components.

Another future study of interest would be to explore the hoses and joints components more closely. These components may have higher leak frequencies for larger leaks than smaller leaks due to a physical explanation, such as flexible hose material being much more likely to fail catastrophically rather than develop small leaks. It may also be better informed with more specific data. However, these types of components are also somewhat different than some of the other components considered. While a vessel or pipe may be in contact with liquid hydrogen for the entire operating year, other components may only come into contact with the liquid hydrogen periodically (such as during a transfer). Therefore, it may be more appropriate and informative to obtain additional data about transfers for these components, and estimate leak frequency distributions per transfer rather than per year of operation.

ACKNOWLEDGMENTS

This work is financially supported by the U.S. Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy (EERE) Hydrogen and Fuel Cell Technologies Office (HFTO). The authors wish to thank Austin Glover and Bobby Middleton at Sandia National Laboratories for reviewing this work. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

1. LaChance, J., Houf, W., Middleton, B., and Fleur, L., Analyses to Support Development of Rik-Informed Separation Distances for Hydrogen Codes and Standards, Sandia National Laboratories Report No. SAND2009-0874, March 2009.
2. Glover, A.M., Baird, A.R., and Brooks, D.M., Final Report on Hydrogen Plant Hazards and Risk Analysis Supporting Hydrogen Plant Siting Near Nuclear Power Plants, Sandia National Laboratories Report No. SAND2020-10828, October 2020.
3. Mulcahy, G.W., Brooks, D.M., and Ehrhart, B.D., Using Bayesian Methodology to Estimate Liquefied Natural Gas Leak Frequencies, Sandia National Laboratories Report No. SAND2021-4905, April 2021.
4. Helton, J.C., and Johnson, J.D., Quantification of Margins and Uncertainties: Alternative Representations of Epistemic Uncertainty, *Reliability Engineering and System Safety*, **96**, No. 9, 2011, pp. 1034-1052
5. Kruschke, J.K., Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, 2015, Elsevier, London.
6. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari A., and Rubin, D.B., Bayesian Data Analysis, 2013, CRC Press, Boca Raton.
7. R Core Team, R: A Language and Environment for Statistical Computing, 2013, R Foundation for Statistical Computing, Vienna.
8. Plummer, M., Stukalov, A. and Denwood, M., rjags: Bayesian Graphical Models Using MCMC, 2019.
9. Plummer, M., JAGS: Just Another Gibbs Sampler, 2012.