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.

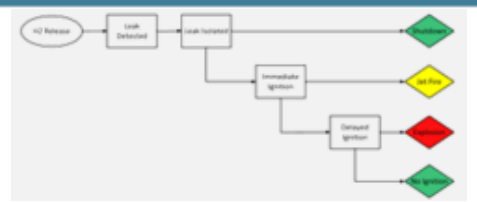
SAND2021-11595C

Sandia National Laboratories

Development of Liquid Hydrogen Leak Frequencies Using a Bayesian Update Process









Dusty M. Brooks, Brian D. Ehrhart, Chris LaFleur

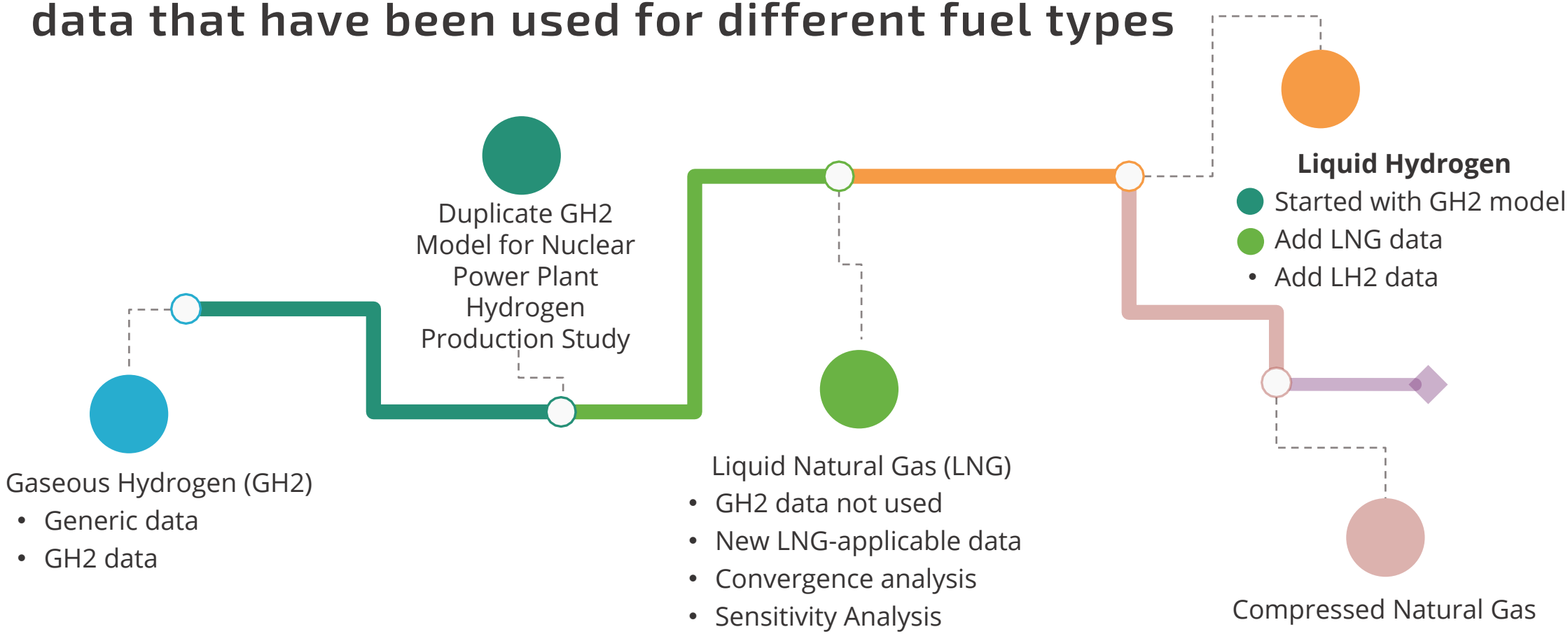
International Conference on Hydrogen Safety 2021

September 21, 2021

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.

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.

The liquid hydrogen prediction model is based on a model and data that have been used for different fuel types



Our liquid hydrogen leak frequency prediction model is built upon methodology and data that have been used for multiple energy systems and development is ongoing.

Leak frequencies are uncertain but necessary for risk analysis

- Per-component annual leak frequencies can be applied broadly to facilities of different sizes
- We can propagate leak frequencies through full system models to predict the frequency of risk-significant events
- Many sources of uncertainty affect our ability to estimate leak frequencies

Aleatory uncertainty: inherent variation between the designs, materials, maintenance, operating conditions, ages, etc. of different facilities

Epistemic uncertainty: lack of data for new systems, lack of reporting or inconsistent reporting for existing systems, measurement errors

Bias: from detection (larger leaks are easier to detect), reporting requirements

The prediction model should include state-of-knowledge uncertainties that may be reduced over time with more data, but it must also include the within-population variation.

The modeling strategy aims to maximize the amount of data that can be used but this requires simplifications

Problem	Solution	Consequence
Leak sizes are not always specified quantitatively or by the same metrics	Bin the data by order of magnitude and assign qualitative measures to each bin	Increased variation in predictions, potential miscategorization favoring higher frequencies of larger leaks
The dimensions of components vary but there are not enough data to model each size separately	Convert leak sizes to fractional leak areas relative to component cross- sectional area	Geometries ignored; cross-sectional area may not represent common leak mechanisms (e.g., circ. weld vs. valve component)
Data are not available for all fuel sources (like LH2)	Use data from similar fuel/system types as proxy	Estimates may be systematically biased conservatively or non-conservatively, and this bias cannot be understood without specific data

Data from GH2 and LNG were used to estimate LH2 leak frequencies by proxy

GH2 Data Set

- Mostly consists of generic leak frequencies (non-GH2 sources)
- Includes a small amount of GH2 data
- Maximum likelihood estimates used for GH2 data include 0.0 values which increase uncertainty in estimates

Component	Data Points
Flanges and Gaskets	17
Hoses	12
Joints	12
Pipe	56
Valves	35
Vessels	9

GH2 Data From: SAND2020-7946

LNG Data Set

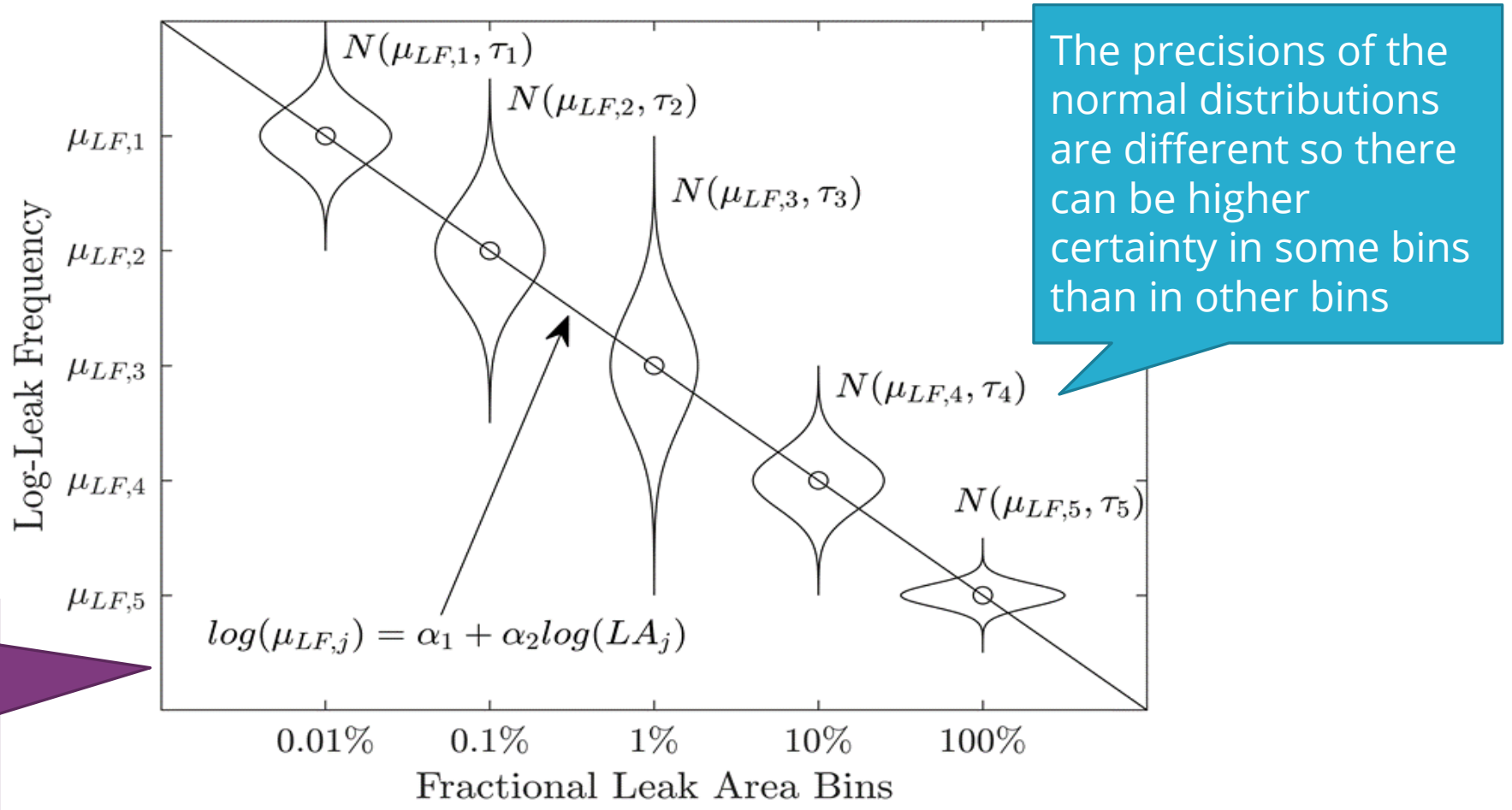
- Includes data from an LNG analysis that was deemed LNG-applicable (from LNG systems or judged to be applicable to LNG systems in the original reference

Component	Data Points
Flanges and Gaskets	32
Hoses	16
Joints	6
Pipe	86
Valves	61
Vessels	168

LNG Data From: SAND2021-4905

The model estimates leak frequencies for each fractional leak area using distributions that are related by their means

The distributions are fit in log space, so $\mu_{LF,1}$ is the mean of the normal distribution on the log-leak frequency for the smallest leak size bin



The precisions of the normal distributions are different so there can be higher certainty in some bins than in other bins

The linear relationship between means allows data from one bin to influence other bins that may have no data

The model is implemented as a trivial Bayesian hierarchical model that is updated in stages

Initial values are provided for $\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, \tau_j$
These values are calibrated using data and the definitions at the bottom of the hierarchy

$$\alpha_1 \sim normal(\alpha_{11}, \alpha_{12})$$
$$\alpha_2 \sim normal(\alpha_{21}, \alpha_{22})$$

$$\tau_j \sim gamma(s_j, r_j)$$

First (and last) level in hierarchy

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

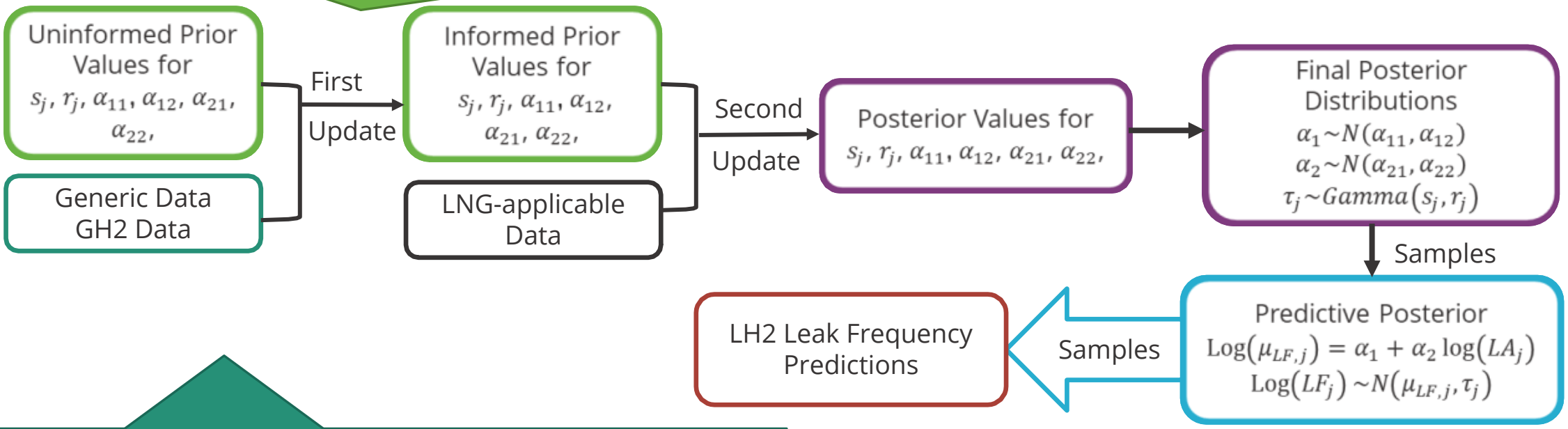
$$\log(LF_j) \sim normal(\mu_{LF,j}, \tau_j)$$

Base model that is assumed to govern published leak frequencies

Once the values for $\alpha_{11}, \alpha_{12}, \alpha_{21}, \alpha_{22}, \tau_j$ are calibrated, the model is implemented by sampling from the top of the hierarchy, propagating to the bottom, and exponentiating the final estimate

The full analysis flow

The first update is technically two stages because the GH2 data were provided in a form incompatible with the update structure; this means the linear component of the model was not strictly enforced in this stage.

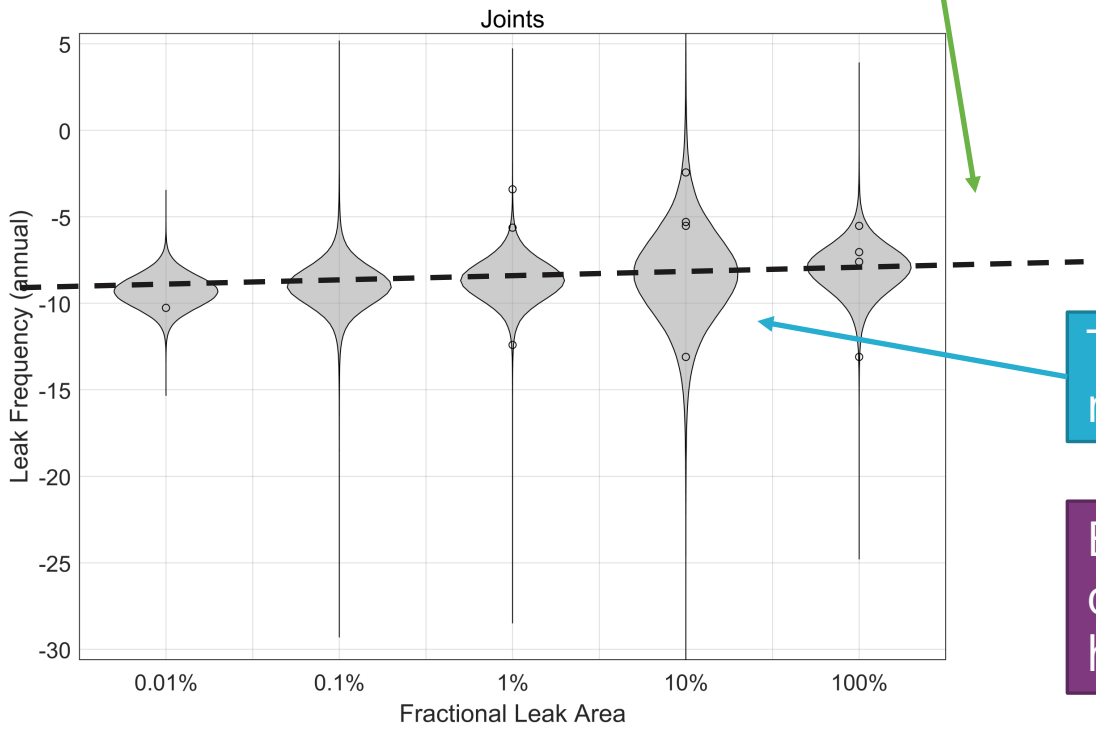


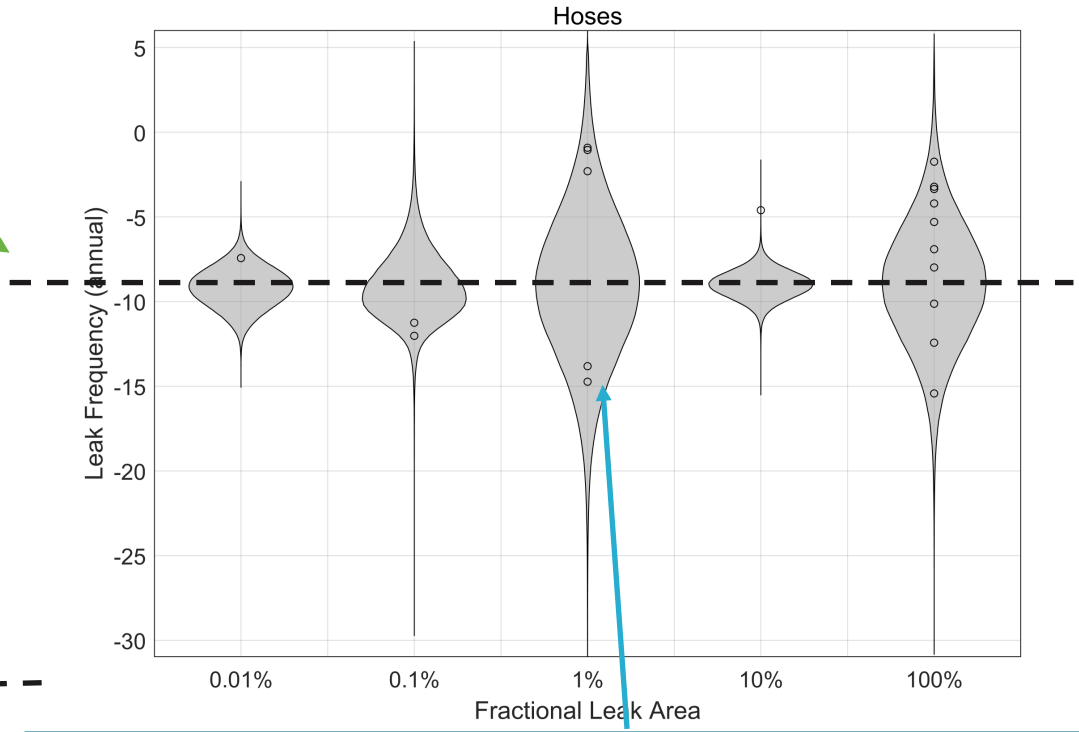
LH2-specific data can be implemented as another update stage, combined with the current data in the second update, or can replace less-applicable data

Uncertainty in the parameters leads to uncertainty in the normal distributions; two layers of uncertainty are incorporated into the leak frequency estimates

There are insufficient data to draw reliable conclusions about leak frequencies for hoses and joints

Linear relationships indicate similar or higher frequencies for large leaks compared to small leaks; it is unclear if this is due to lack of data or physics related to these components

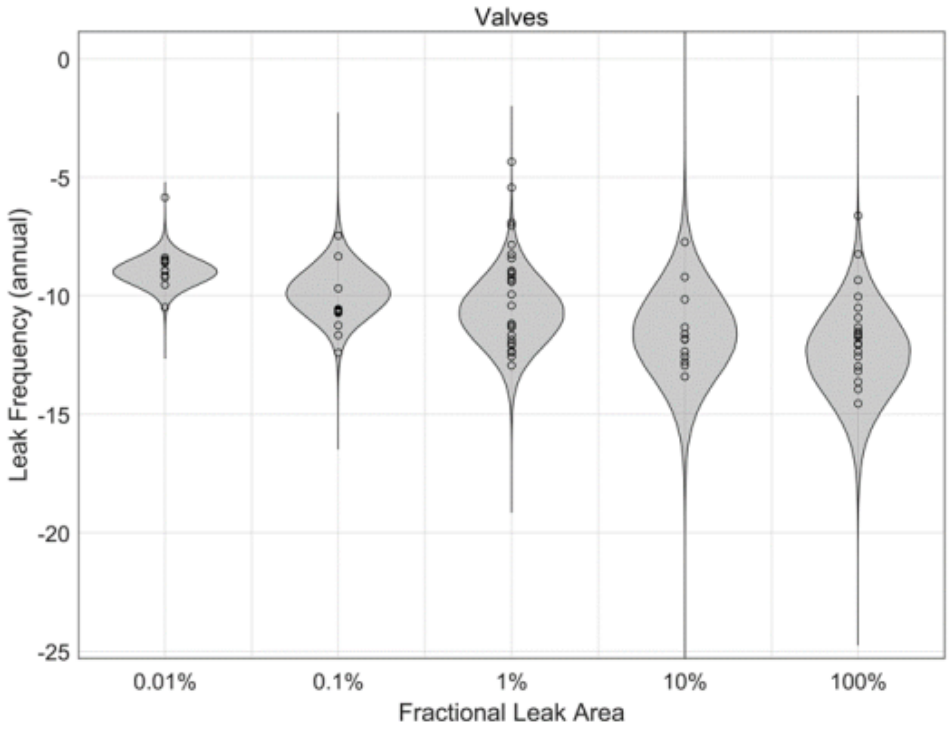


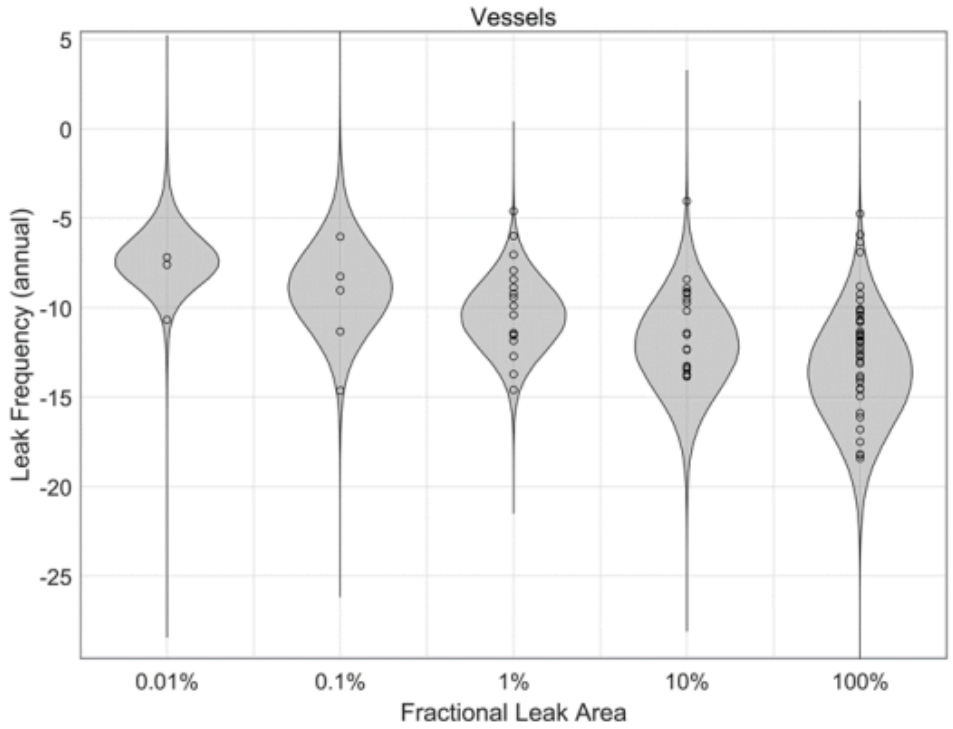


The data we have may not be consistent with a normal distribution

Estimates for both components span multiple orders of magnitude; hose estimates can be very high for some leak sizes

The models for valves and vessels demonstrate increased uncertainty in large leak frequencies which may reflect bias

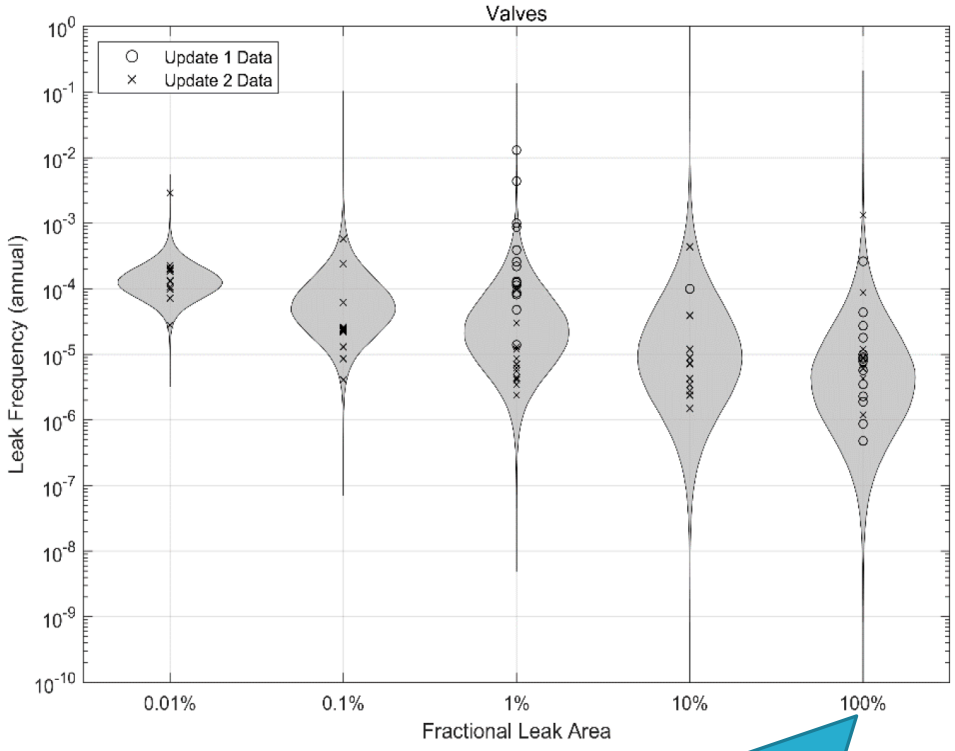


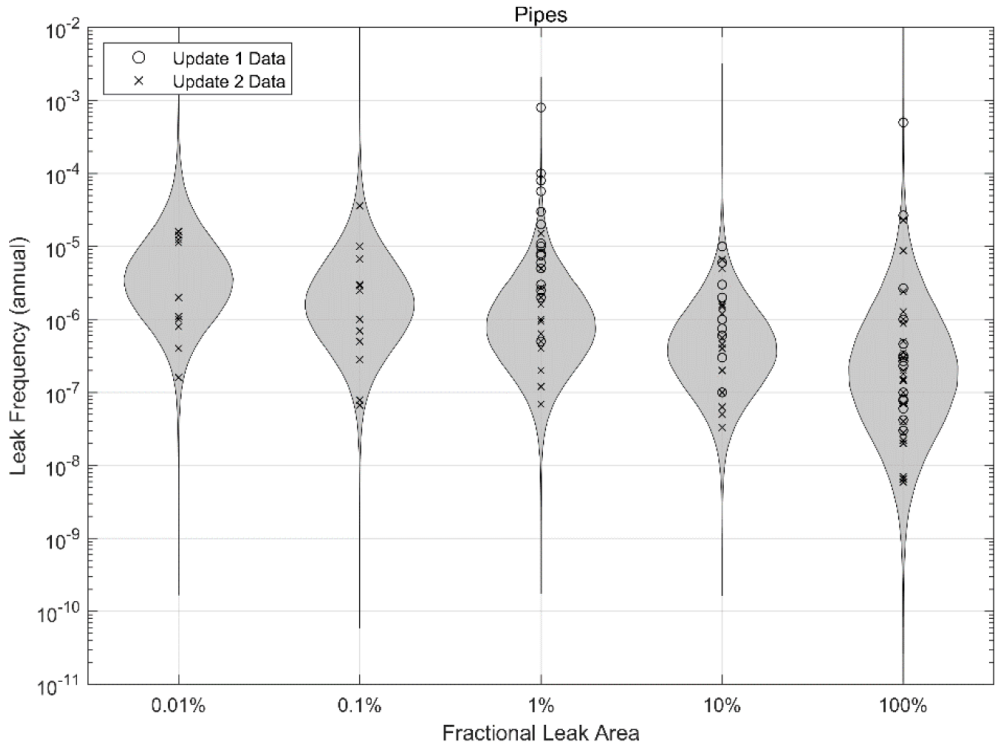


There is lower uncertainty at the 0.01% leak area but this may be because there is insufficient data for this leak size to capture the within-population variation

The large standard deviations for some of the normal distributions may be dominated by outliers, in particular the GH2 maximum likelihood estimates that were equal to 0.0

GH2 and LNG data points appear to overlap for most components, except possibly pipes





Leak frequency data points from Data Set 1 (GH2) and Data Set 2 (LNG) appear to generally be overlapping for most components and leak sizes

Frequencies from Data Set 2 (LNG) appear to generally be lower than the frequencies from Data Set 1 (GH2) for Pipes; LH2-specific data can help clarify this

The highest estimated median leak frequencies are on the order of 0.001 per year

Component	Leak Size	5th	Median	95th
Flange	0.01%	1.6E-05	5.0E-05	1.4E-04
	0.1%	3.1E-06	2.2E-05	1.8E-04
	1%	1.7E-07	1.1E-05	6.6E-04
	10%	3.3E-08	5.0E-06	6.9E-04
	100%	1.8E-08	2.4E-06	2.8E-04
Hose	0.01%	1.3E-05	1.1E-04	7.3E-04
	0.1%	5.7E-06	9.0E-05	4.3E-03
	1%	6.1E-08	1.2E-04	2.2E-01
	10%	3.1E-05	1.2E-04	4.9E-04
	100%	2.9E-07	1.3E-04	5.6E-02
Joint	0.01%	1.9E-05	9.0E-05	4.3E-04
	0.1%	1.4E-05	1.2E-04	1.1E-03
	1%	2.2E-05	1.7E-04	1.3E-03
	10%	2.2E-06	2.4E-04	2.3E-02
	100%	2.6E-05	3.3E-04	3.3E-03

Component	Leak Size	5th	Median	95th
Pipe	0.01%	3.3E-07	3.4E-06	3.6E-05
	0.1%	1.4E-07	1.6E-06	2.0E-05
	1%	7.4E-08	8.0E-07	8.5E-06
	10%	3.5E-08	3.9E-07	4.2E-06
	100%	6.3E-09	1.9E-07	5.4E-06
Valve	0.01%	4.6E-05	1.2E-04	3.4E-04
	0.1%	8.8E-06	5.2E-05	3.2E-04
	1%	2.1E-06	2.2E-05	2.4E-04
	10%	3.3E-07	9.3E-06	2.5E-04
	100%	1.5E-07	3.9E-06	9.6E-05
Vessel	0.01%	4.3E-05	6.5E-04	1.3E-02
	0.1%	2.1E-06	1.4E-04	8.9E-03
	1%	9.6E-07	2.8E-05	8.1E-04
	10%	6.4E-08	5.7E-06	4.9E-04
	100%	6.8E-09	1.2E-06	2.0E-04

The medians characterize the centers of the distributions but due to the within-population variation, higher or lower percentiles may be more appropriate for specific sites.

The distributions may shift/stretch/shrink as more data become available, but they will never reduce to a point even with perfect knowledge due to this variation.

Key questions

- Are all of the data (general fuel systems, GH2 systems, LNG systems) applicable to LH2 systems?
- Do these leak frequencies seem reasonable in magnitude?
- Does it make sense that large leaks are more likely than small for hoses and joints?
- Is the level of uncertainty acceptable? Will systems meet requirements at the higher percentiles?
- Will leak frequency data for LH2 systems significantly modify these leak frequency distributions?



THANK YOU

Questions?

Brian Ehrhart
bdehrha@sandia.gov

This work funded by the U.S. DOE Hydrogen and Fuel Cell Technologies Office (HFTO) Safety, Codes, and Standards (SCS) sub-program