SAND2021-13725C

ANS PSA 2021
2021 International Topical Meeting on Probabilistic Safety Assessment and Analysis

Development of Leak Frequencies Using a Bayesian Update Process

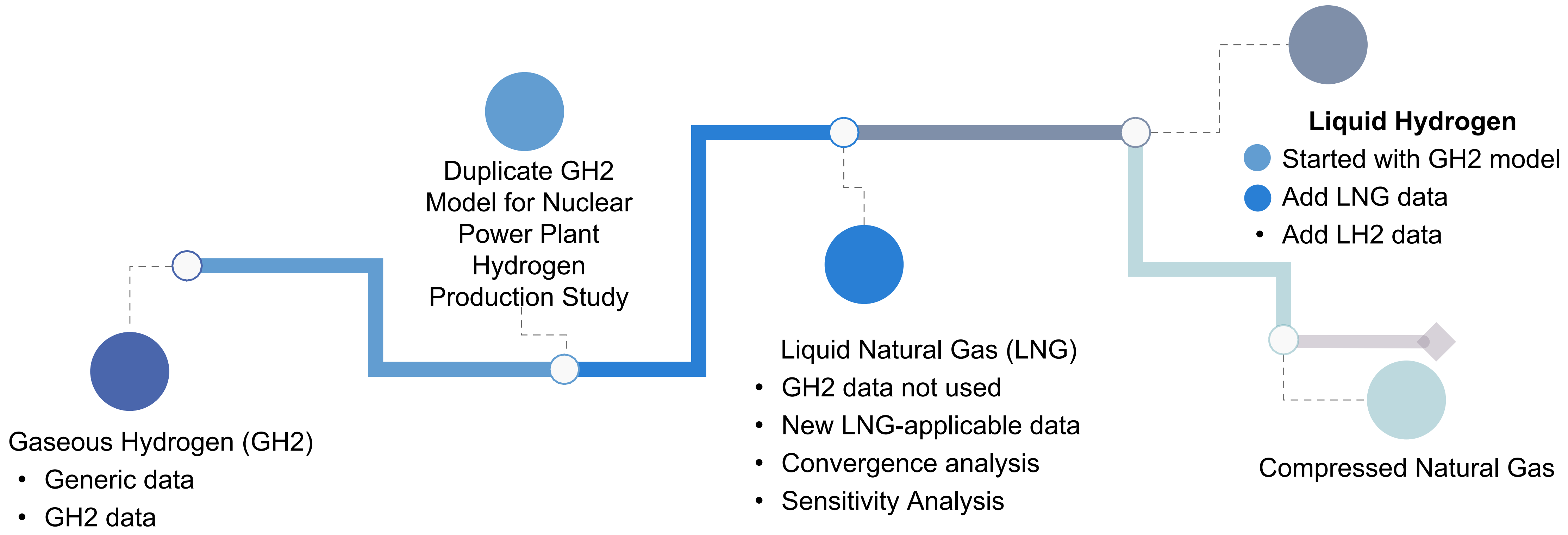
Dusty M. Brooks, Brian D. Ehrhart, Garrett W. Mulcahy, Chris LaFleur

Sandia National Laboratories

Sandia National Laboratories

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 leak frequency prediction model is based on a model and data that have been used for different fuel types



The model is built upon methodology and data that have been used for multiple energy systems; it also contributes to ongoing development for additional fuel types.

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.

ANS

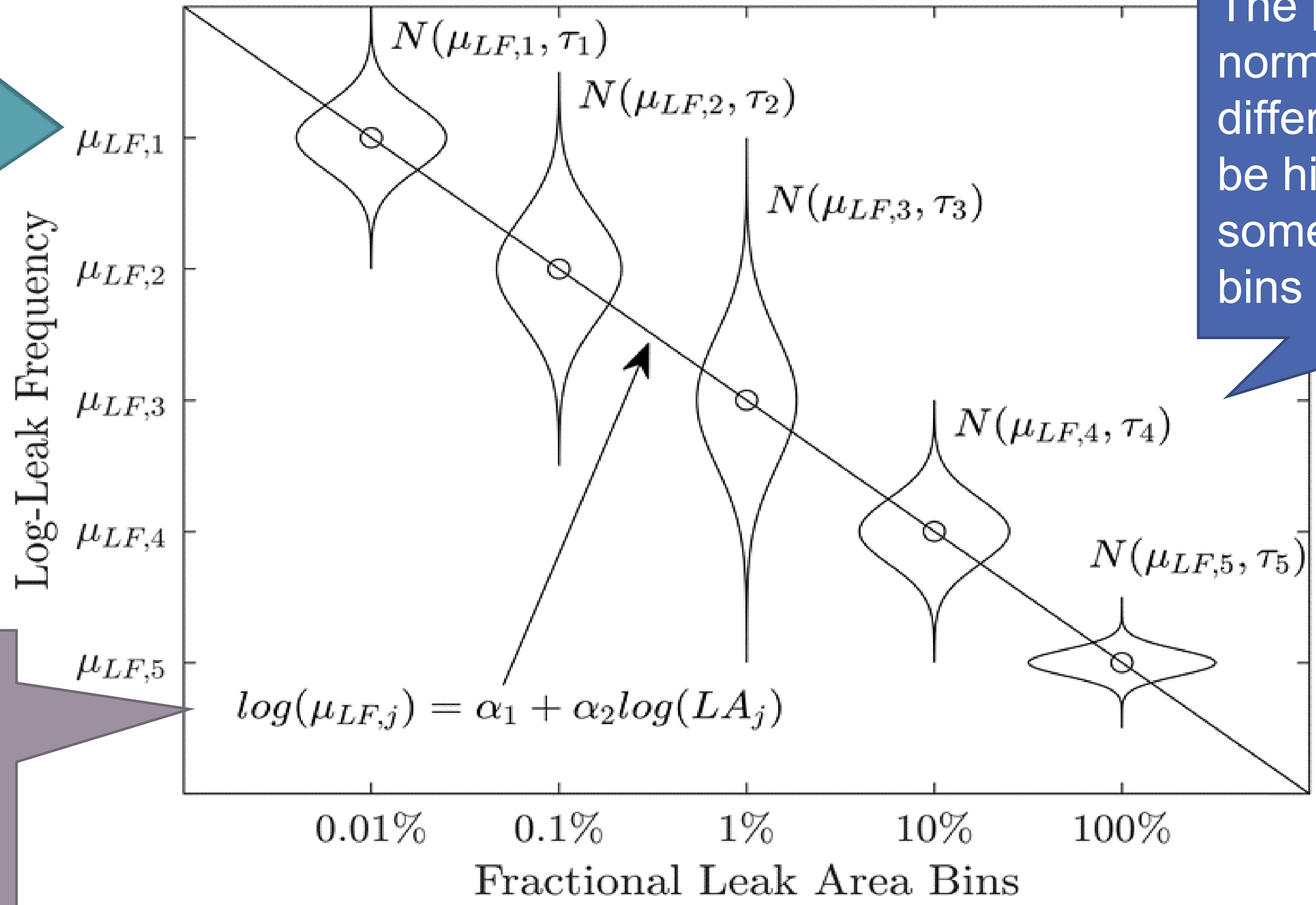
The modeling strategy aims to maximize the amount of data that can be used but this requires simplifications

Problem	Solution	Impact
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 mis-categorization 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 or sparse for some fuel sources	Use data from similar fuel/system types to supplement fuel-specific data	Estimates may be systematically biased conservatively or non-conservatively

ANS

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 α_{11}, α_{12}, α_{21}, α_{22}, τ_j
These values are calibrated using data and the definitions at the bottom of the hierarchy

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

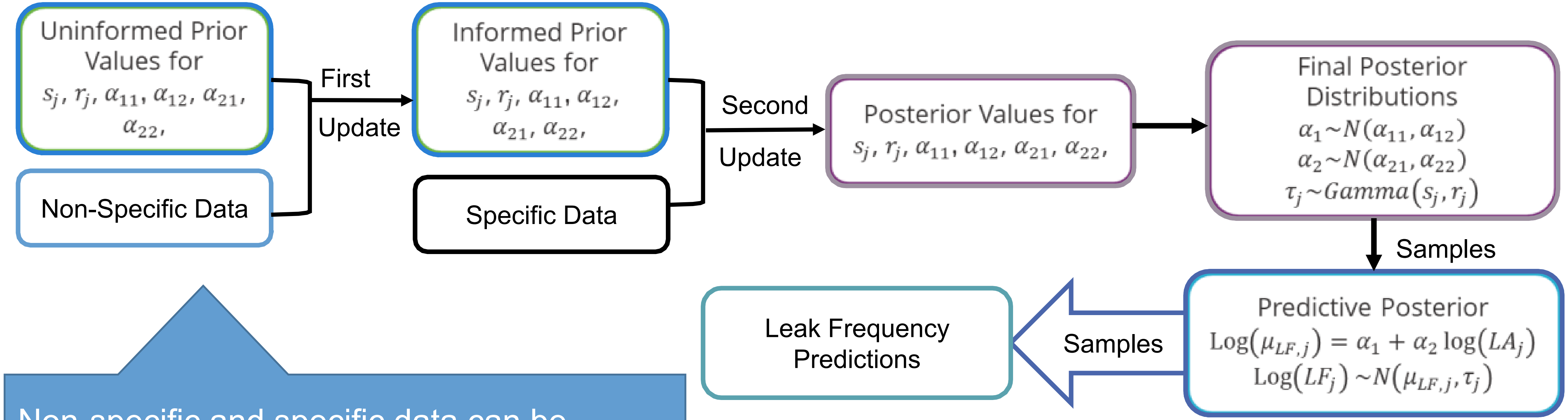
$$\tau_j \sim gamma(s_j, r_j)$$

$$\log(\mu_{LF,j}) = \alpha_1 + \alpha_2 \log(LA_j) \qquad \log(LF_j) \sim normal(\mu_{LF,j}, \tau_j)$$

Base model that is assumed to govern published leak frequencies

Once the values for α_{11}, α_{12}, α_{21}, α_{22}, τ_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 update process can be repeated many times to include different data sets.

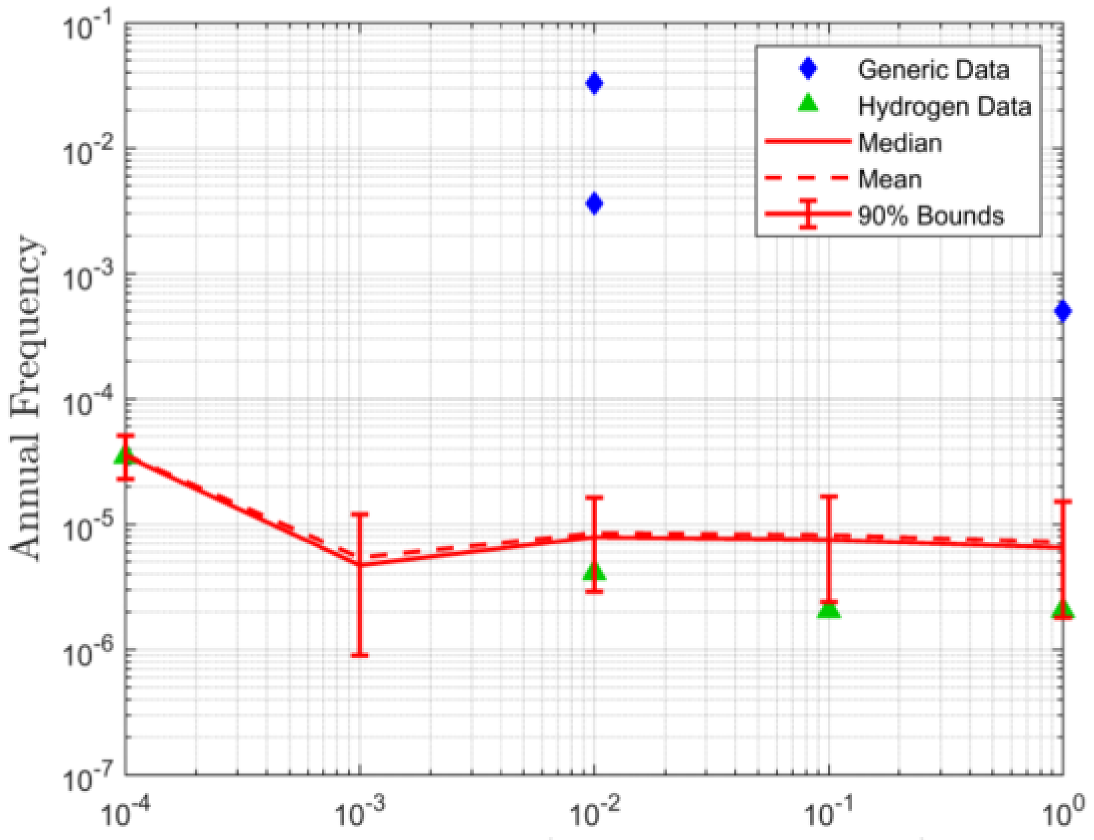
The full analysis flow



Non-specific and specific data can be implemented as single update stage if the data are in the same format, but separating the updates shows the individual impact of each data set on the final model.

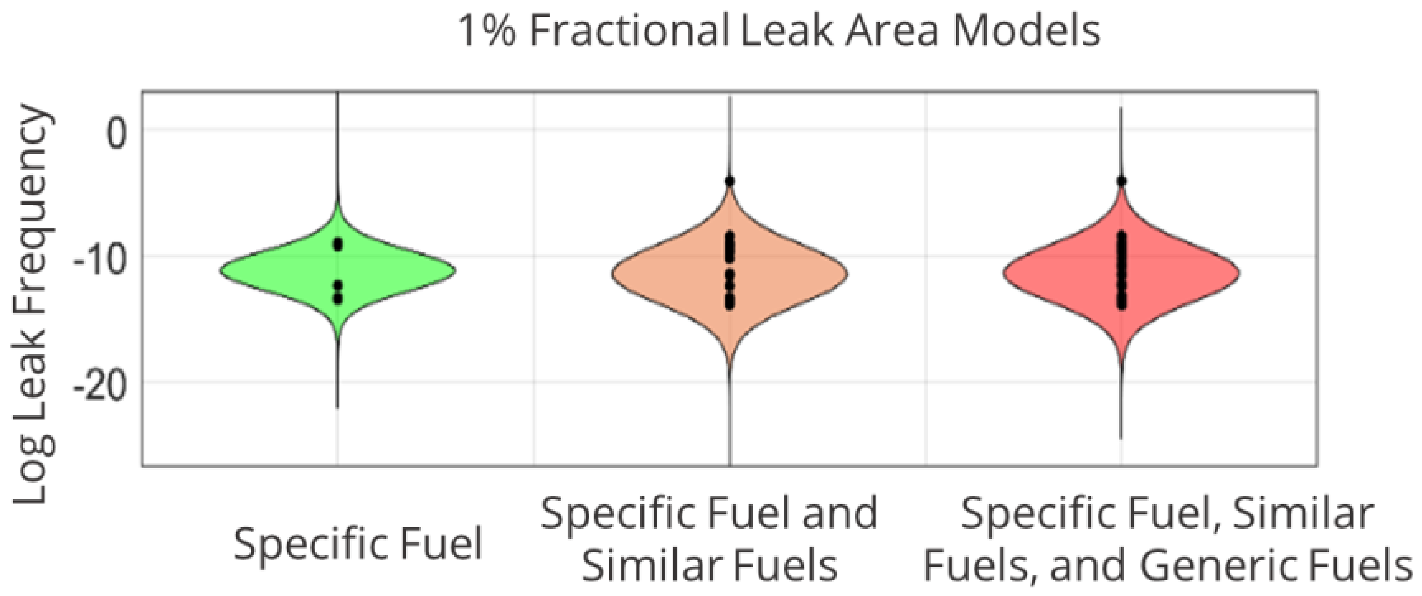
Uncertainty in the parameters leads to uncertainty in the normal distributions; two layers of uncertainty are incorporated into the leak frequency estimates.

Case Study 1: Gaseous Hydrogen (GH2)



- GH2 case was unique because we had two data sets in different formats:
  - Generic data in the form of frequencies
  - Specific data as number of incidents over time (some zero for long periods of time)
- The structure of the model had to be modified in the second update to accommodate the GH2 specific data, but this had benefits:
  - The linear assumption was no longer imposed
  - The fuel-specific data dominated the model so it was not overly conservative due to higher generic leak frequencies
- Another option would be to convert the GH2 data into frequency estimates, though the zeroes would need to be excluded or estimated with nonzero values

Case Study 2: LNG

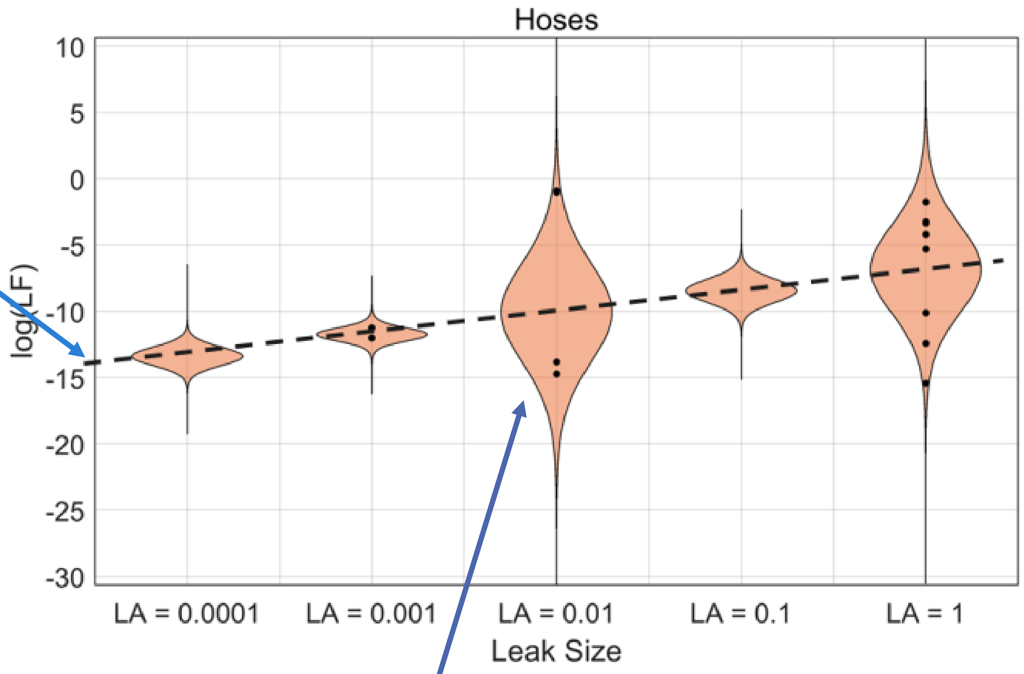


- Data from LNG systems were sparse but there were data from similar fuels
- All data were in the form of frequency estimates, so the linear assumption was maintained
- Our analysis included sensitivity studies to evaluate the effect of each data set
- Generic data were not included in the final model, but may still be needed in analyses of fuel types with less data from specific and similar sources

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

Linear relationship indicates 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 hoses

- Violin plots show the rotated and reflect distribution of results. This:
- gives more detail than a mean or median with uncertainty bounds,
- emphasizes the normal distribution assumption, and
- discourages interpolation between the discrete leak size bins.
- Estimates span multiple orders of magnitude; hose estimates can be very high for some leak sizes



The data we have may not be consistent with the assumed normal distribution, but we need more data to know

The highest estimated median leak frequencies were uniquely high for joints in the LNG analysis

Component	Leak Size	5th	Median	95th
Flange	0.01%	1.13E-05	4.18E-05	1.14E-04
	0.1%	3.52E-06	2.26E-05	1.82E-04
	1%	2.84E-07	1.40E-05	7.14E-04
	10%	8.81E-08	8.68E-06	7.30E-04
	100%	4.03E-08	5.24E-06	5.40E-04
Heat Exchanger	0.01%	5.41E-04	2.34E-03	1.22E-02
	0.1%	1.03E-04	8.93E-04	7.27E-03
	1%	3.11E-05	3.24E-04	3.22E-03
	10%	2.69E-06	1.17E-04	5.14E-03
	100%	3.13E-06	4.18E-05	6.27E-04
Hose	0.01%	4.49E-07	1.52E-06	5.13E-06
	0.1%	3.16E-06	7.89E-06	1.99E-05
	1%	4.38E-08	4.13E-05	3.82E-02
	10%	4.65E-05	2.14E-04	1.00E-03
	100%	2.96E-06	1.10E-03	4.34E-01
Joint	0.01%	9.89E+02	3.51E+04	1.25E+06
	0.1%	3.20E+01	4.77E+02	7.09E+03
	1%	9.98E-01	6.46E+00	4.18E+01
	10%	2.78E-02	8.76E-02	2.76E-01
	100%	4.32E-04	1.19E-03	3.26E-03

Component	Leak Size	5th	Median	95th
Pipe	0.01%	3.08E-07	2.67E-06	2.30E-05
	0.1%	1.39E-07	1.44E-06	1.52E-05
	1%	1.17E-07	7.86E-07	5.22E-06
	10%	4.59E-08	4.25E-07	3.91E-06
	100%	1.15E-08	2.30E-07	4.63E-06
Valve	0.01%	2.40E-05	8.43E-05	2.48E-04
	0.1%	8.76E-06	4.20E-05	2.18E-04
	1%	3.54E-06	2.16E-05	1.53E-04
	10%	4.72E-07	1.18E-05	2.69E-04
	100%	2.34E-07	6.42E-06	1.29E-04
Vaporizer	0.01%	1.27E-04	8.19E-03	5.24E-01
	0.1%	1.24E-03	2.63E-02	5.57E-01
	1%	1.15E-02	8.46E-02	6.23E-01
	10%	8.65E-02	2.72E-01	8.57E-01
	100%	2.79E-01	8.75E-01	2.75E+00
Vessel	0.01%	8.18E-05	4.77E-04	3.41E-03
	0.1%	3.69E-06	1.39E-04	5.25E-03
	1%	1.65E-06	3.90E-05	9.14E-04
	10%	2.03E-07	1.10E-05	5.80E-04
	100%	1.67E-08	3.05E-06	5.77E-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.

ANS

Conclusions

- Case studies demonstrate that the method is effective in characterizing uncertainty using limited data sets
- The staged approach to updated the model emphasizes the effects that fuel-specific data have over non-specific
- The model may underestimate uncertainty in some cases due to lack-of-data, and overestimate in other cases due to generic data incorporating non-representative variability from other fuel types
- The quality and comprehensiveness of data, or lack thereof, should always be considered critically before model results are applied in risk analysis.

Key Questions

- Are all of the data (e.g. from similar fuels) applicable to the systems being analyzed?
- Do these leak frequencies seem reasonable in magnitude?
- Does it make sense that large leaks are more likely than small for hoses?
- Is the level of uncertainty acceptable? Will systems meet requirements at the higher percentiles?
- Will additional fuel-specific leak frequency data significantly modify these leak frequency distributions?

ANS