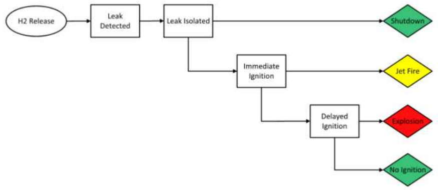
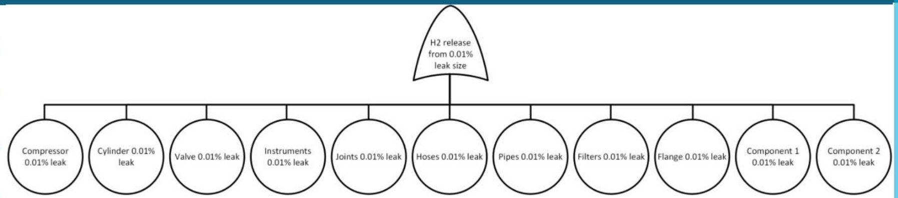
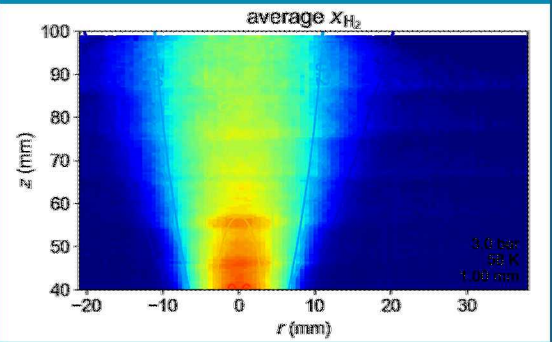
SAND2020-4083PE

Bayesian Leak Frequency Methodology







Presented by

Dusty Brooks, Brian Ehrhart

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.

Leak Frequency Estimation

Original work: SAND2009-0874 (LaChance et. al.)

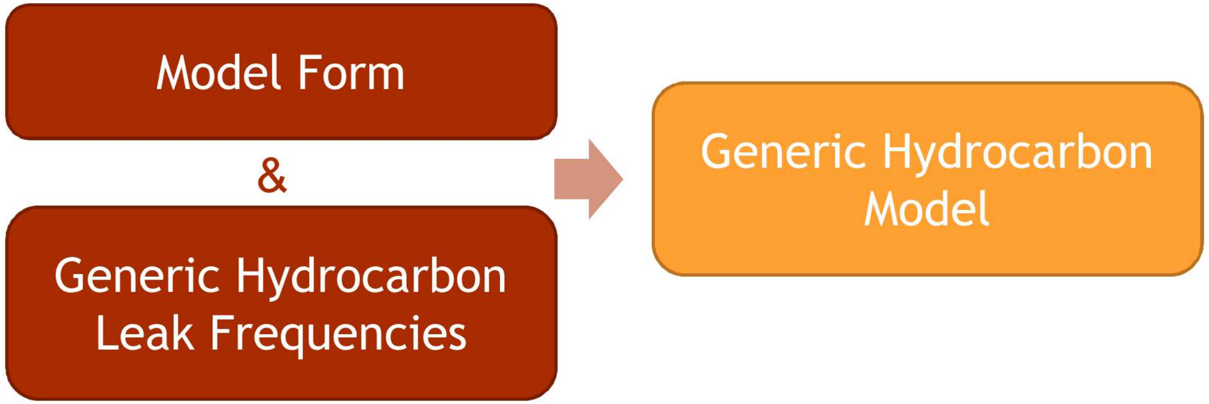
- Bayesian hierarchical model
- Used uninformed priors, updated twice: first update with generic hydrocarbon leak frequencies, second update with gaseous hydrogen leak event data (occurrences over time)
- Leak frequencies estimated based on size as a fraction of the component cross sectional area
  - 0.01%, 0.1%, 1%, 10,% or 100%

Our goals:

- Duplicate the 2009 work to verify that we implemented the methodology correctly ✓
- Use the methodology with different data sets to derive leak frequencies for CNG, LNG, and LH2 (in progress)

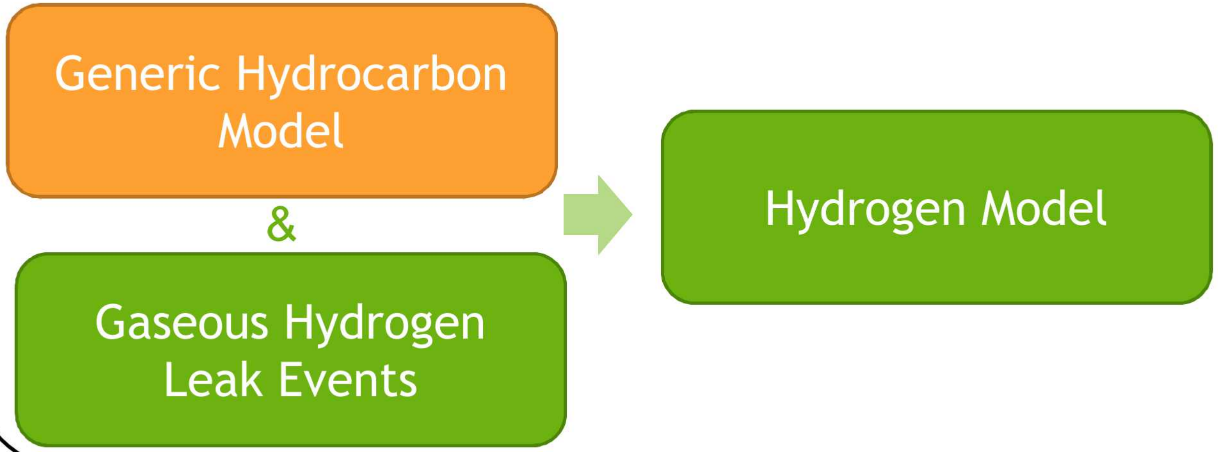
Leak Frequency Estimation Methodology

First Model Update – Frequency Data



$$LN(\mu_{LF,j}) = \alpha_2 LN(FLA_j) + \alpha_1$$
$$\alpha_1 \sim N(0,10^{-3})$$
$$\alpha_2 \sim N(0,10^{-3})$$
$$LN(LF_j) \sim N(\mu_{LF,j}, \tau_j)$$
$$\tau_j \sim Gamma(3,1)$$

Second Model Update – Events Over Time Data



$$LN(\mu_{LF,j}) = \alpha_2 LN(FLA_j) + \alpha_1$$
$$\alpha_1 \sim N(\mu_{\alpha_1}, \tau_{\alpha_1})$$
$$\alpha_2 \sim N(\mu_{\alpha_2}, \tau_{\alpha_2})$$
$$LN(LF_j) \sim N(\mu_{LF,j}, \tau_j)$$
$$\tau_j \sim Gamma(\mathrm{a}_j, b_j)$$
$$x_j \sim Poisson(\lambda_j)$$
$$\lambda_j = LF_j \times Time$$

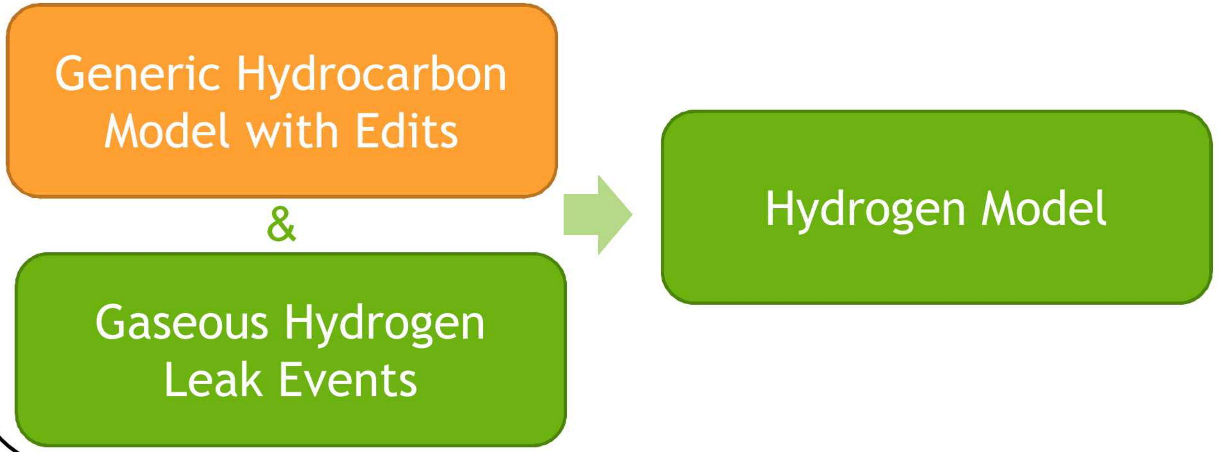
Leak Frequency Estimation Methodology

The report described the model as base 10 and $\tau_j \sim Gamma(1, 1)$ but this model did not converge
A base e model was clearly used, as base 10 logs are not implemented in WinBUGS.

Incrementing the first parameter of the gamma distribution enabled convergence of the model while still keeping the precision distribution "low"

$$
\begin{cases}
LN(\mu_{LF,j}) = \alpha_2 LN(FLA_j) + \alpha_1 \\
\alpha_1 \sim N(0, 10^{-3}) \\
\alpha_2 \sim N(0, 10^{-3}) \\
LN(LF_j) \sim N(\mu_{LF,j}, \tau_j) \\
\tau_j \sim Gamma(3,1)
\end{cases}
$$

Second Model Update – Events Over Time Data



$$
\begin{aligned}
& LN(\mu_{LF,j}) = \alpha_2 LN(FLA_j) + \alpha_1 \\
& \alpha_1 \sim N(\mu_{\alpha_1}, \tau_{\alpha_1}) \\
& \alpha_2 \sim N(\mu_{\alpha_2}, \tau_{\alpha_2}) \\
& LN(LF_j) \sim N(\mu_{LF,j}, \tau_j) \\
& \tau_j \sim Gamma(a_j, b_j) \\
& x_j \sim Poisson(\lambda_j) \\
& \lambda_j = LF_j \times Time
\end{aligned}
$$

Leak Frequency Estimation Methodology

First Model Update – Frequency Data

Model Form

&

Generic Hydrocarbon Leak Frequencies

→ Generic Hydrocarbon Model

$$LN(\mu_{LF,j}) = \alpha_2 LN(FLA_j) + \alpha_1$$
$$\alpha_1 \sim N(0, 10^{-3})$$
$$\alpha_2 \sim N(0, 10^{-3})$$
$$LN(LF_j) \sim N(\mu_{LF,j}, \tau_j)$$
$$\tau_j \sim Gamma(3,1)$$

The report did not describe the inclusion of the Poisson distribution, but:

- It is necessary to fully utilize the events/time data for hydrogen
- The final hydrogen model results in the report are not-linear – this cannot be accomplished with the model from the first update
- This Poisson implementation was found in model files from the authors of the report

$$LN(\mu_{LF,j}) = \alpha_2 LN(FLA_j) + \alpha_1$$
$$\alpha_1 \sim N(\mu_{\alpha_1}, \tau_{\alpha_1})$$
$$\alpha_2 \sim N(\mu_{\alpha_2}, \tau_{\alpha_2})$$
$$LN(LF_j) \sim N(\mu_{LF,j}, \tau_j)$$
$$\tau_j \sim Gamma(\mathrm{a}_j, b_j)$$
$$x_j \sim Poisson(\lambda_j)$$
$$\lambda_j = LF_j \times Time$$

Leak Frequency Estimation Methodology

First Model Update – Frequency Data

Model Form

&

Generic Hydrocarbon Leak Frequencies

→ Generic Hydrocarbon Model

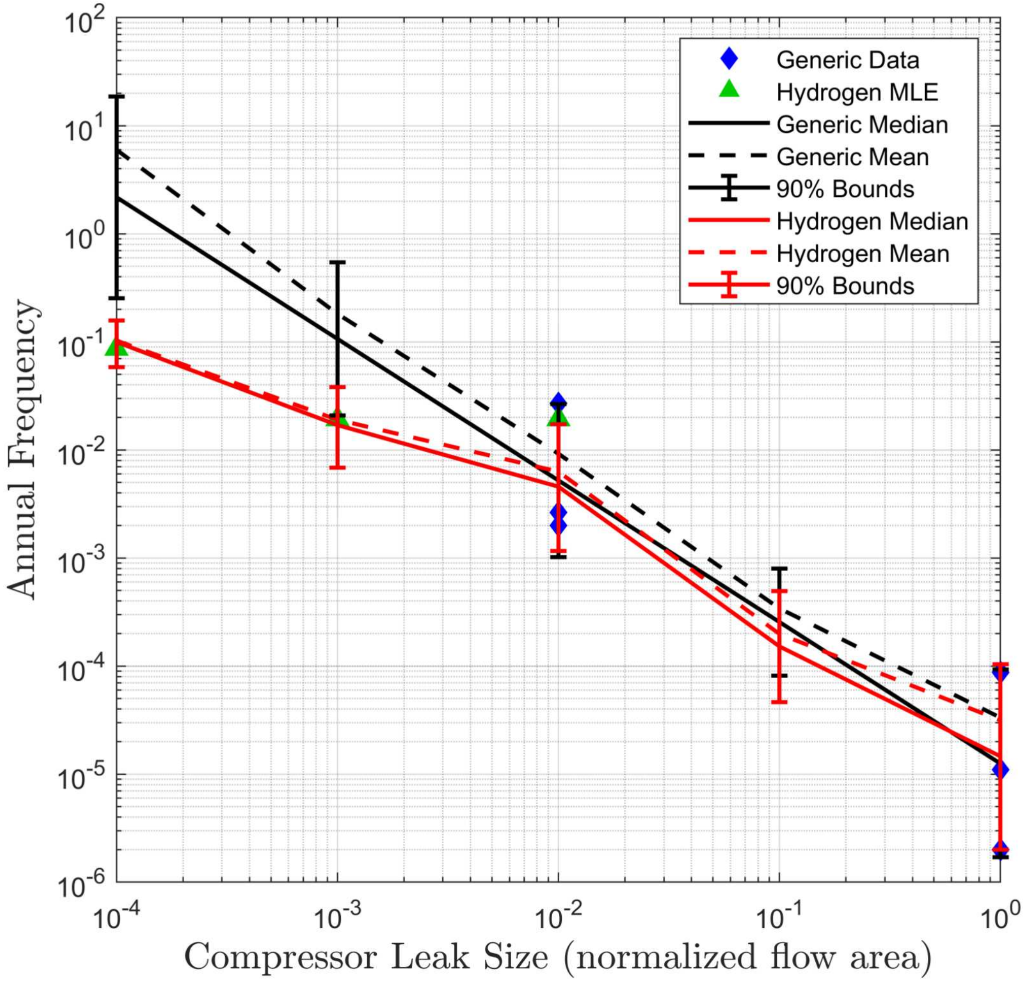
$$LN(\mu_{LF,j}) = \alpha_2 LN(FLA_j) + \alpha_1$$
$$\alpha_1 \sim N(0, 10^{-3})$$
$$\alpha_2 \sim N(0, 10^{-3})$$
$$LN(LF_j) \sim N(\mu_{LF,j}, \tau_j)$$
$$\tau_j \sim Gamma(3,1)$$

The report did not describe the inclusion of the Poisson distribution, but:

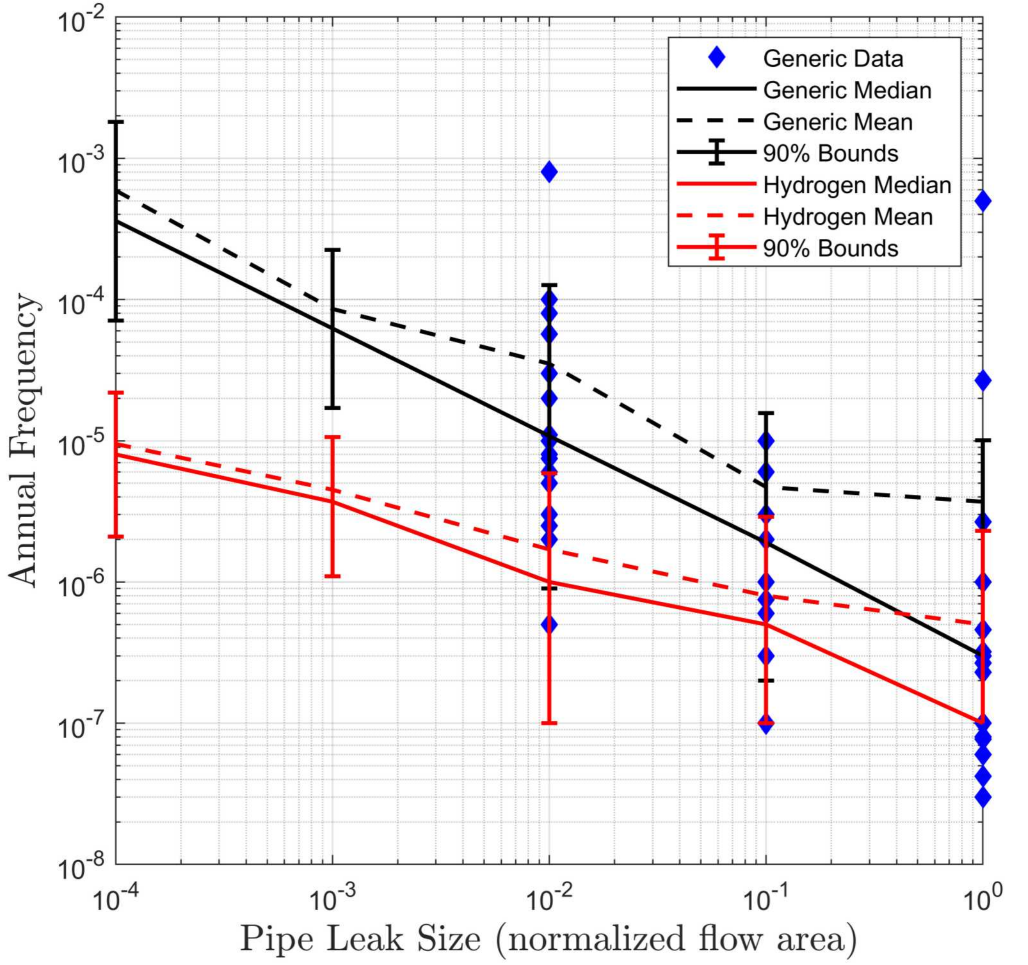
- It is necessary to fully utilize the events/time data for hydrogen
- The median hydrogen model results in the report are not-linear – this cannot be accomplished with the model from the first update
- This Poisson implementation was found in model files from the authors of the report

$$LN(\mu_{LF,j}) = \alpha_2 LN(FLA_j) + \alpha_1$$
$$\alpha_1 \sim N(\mu_{\alpha_1}, \tau_{\alpha_1})$$
$$\alpha_2 \sim N(\mu_{\alpha_2}, \tau_{\alpha_2})$$
$$LN(LF_j) \sim N(\mu_{LF,j}, \tau_j)$$
$$\tau_j \sim Gamma(\mathrm{a}_j, b_j)$$
$$x_j \sim Poisson(\lambda_j)$$
$$\lambda_j = LF_j \times Time$$

Results - Examples



* Hydrogen MLE was 0 for pipes



Discussion Questions

The original methodology used the data to generate informed priors. The same data was then used to update the model. Does this make sense, or does this inflate the influence of that data? Is it more appropriate to use uninformed priors but use the data to estimate initial values for sampling?

Because the model is not truncated, the mean can "explode". Does instability in the mean invalidate the model or is it still valid to use the median and stable percentiles?

Does it make sense to use generic hydrocarbon data to update specific models? How do we decide if there is enough hydrogen data to use it exclusively in the update?