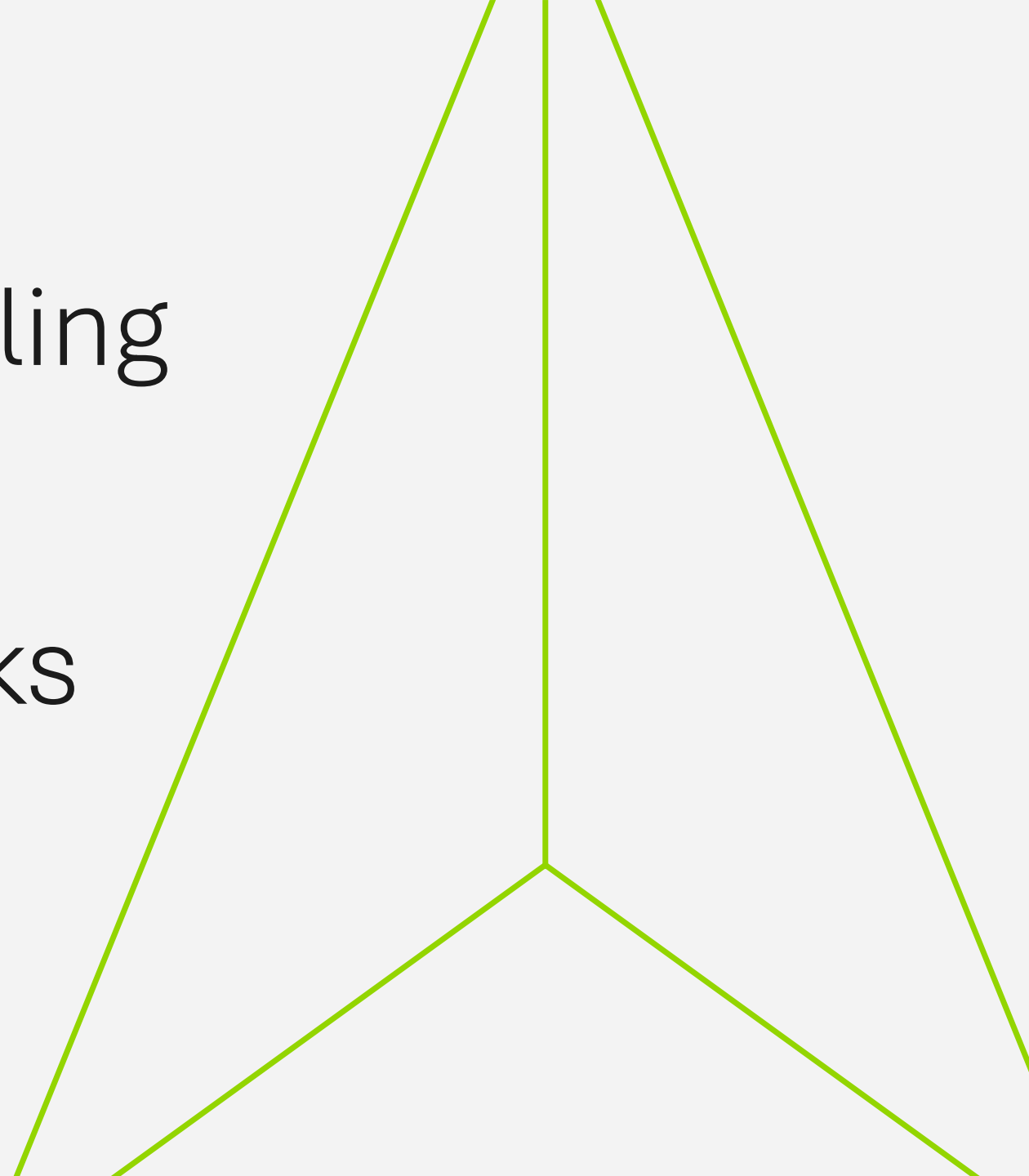


Probabilistic modelling for methane leak detection in gas distribution networks

Rubab Atwal, Katie Green

outwit complexity™



Machine learning can be used for leak detection in gas distribution networks to reduce methane emissions



Environmental problem

Methane emissions from gas distribution networks account for an estimated 1% [1] of total UK greenhouse gas emissions.

Machine learning methodologies offer the potential to continuously monitor pipelines, allowing for:

- Real time detection, localization, and quantification of leaks
- Swifter response and therefore reduced emissions and increased safety



Technical challenges

Developing and deploying machine learning solutions for leak detection in gas distribution networks is challenging due to:

- Sparse data availability across pipe networks
- Sensor readings being inherently noisy and affected by environmental conditions
- Gas distribution networks have specific business needs for their emissions management



Proposed solution

We developed a probabilistic model, building on the work of Mücke et al. (2023) [2], which utilizes a Wasserstein autoencoder and Bayesian inference to detect, localize, and quantify leaks in gas distribution networks from real-time pressure readings.

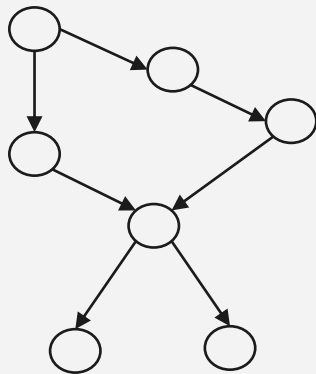
We developed and tested our methodology on a 42km gas distribution section in the UK.

Representative training data is simulated for each network segment for both standard operational and leaking scenarios

Stage 1: Network model

Asset data repository

Asset data is drawn from various sources across the GDN data repository to create a detailed, graph-based representation of the network.

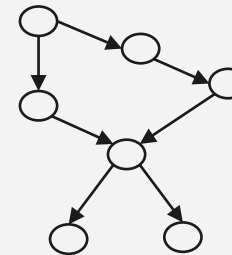


Network models are created for each pressure tier within each network. These are then broken down into 'segments' (smaller sections of the network which are contained by pressure sensors) to reduce the computational modelling load.

Stage 2: Data simulation

Inlet conditions

The inlet conditions for each segment are used to generate a distribution of possible inlet pressure and flow values



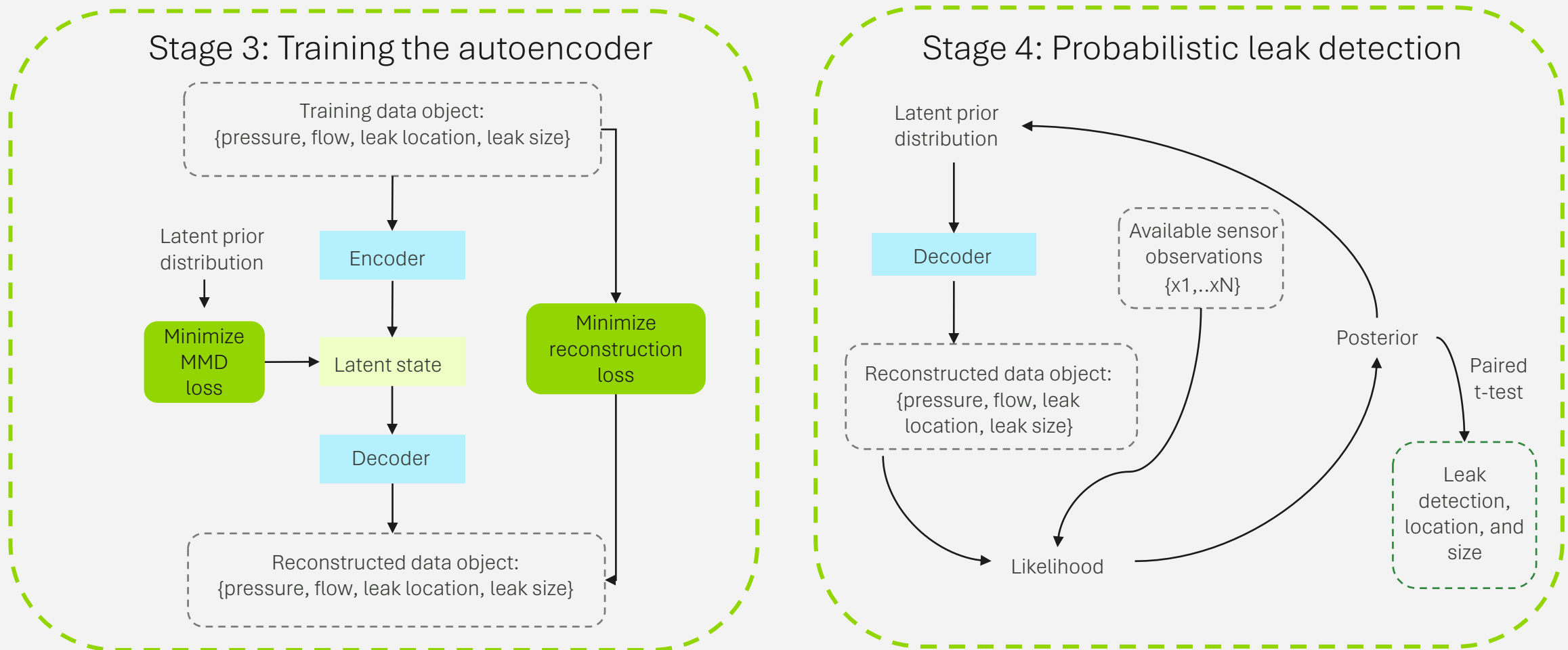
Hydraulic equations

The possible inlets are input into hydraulic equations to simulate pressure and flow values across the segment.

Training scenarios

Leaks are artificially introduced across the segment to simulate operational conditions for millions of different scenarios representing potential real network states.

The simulated data is used to train an autoencoder to perform probabilistic leak detection



The model (covering 42km of network) can successfully detect, localize, and quantify leaks but is heavily dependent on sensor coverage

Experiment 1: 100 simulated leaks assuming pressure data is available from all 12 installed sensors across 42km of network.

Leak detection accuracy	Localisation accuracy	Leak size accuracy	Working sensors
82%	80%	76%	100%

Experiment 2: 3 historically recorded leaks with variably available pressure sensor data from the 12 installed sensors.

Leak	Pipeline detected	Distance along pipeline (m)	Volume (kscm/hr)	Working sensors
1	Correct	400	0.021	58.3%
2	Correct	250	0.286	58.3%
3	Not detected	-	-	38.4%

Experiment 3: Ablation study to determine the minimum sensor count for acceptable model accuracy

Model performance declined after removing 30% of sensors, we observed that pipes located near active sensors still detected synthetic leaks accurately, even when overall network coverage was sparse. This indicates that spatial coverage, in addition to sensor count, plays a critical role in scaling probabilistic modeling effectively.

Through further deployment, validation, and optimization the benefits of probabilistic modelling for leak detection can be realised



What we have achieved

We successfully developed a probabilistic machine learning framework for methane leak detection, localization, and quantification and tested it on a section of UK gas distribution network.

The model can identify leaks in near real time, even with limited data. However, the results also highlight the minimum data requirements for reliable performance.



Benefits

Deploying these models in distribution networks with sufficient sensor coverage will enable continuous pipeline monitoring, providing an early warning system for leak detection.

This allows for proactive interventions, improved safety, reduced emissions, and more accurate reporting.



Next steps

- Conduct further testing and validation of the model across additional network sections.
- Deploy the framework more broadly and monitor real-time performance.
- Optimize model accuracy by recommending sensor placements and configurations.
- Investigate the use of weighted priors and emulators to enhance data generation and model robustness