Cadent
Your Gas Network

Guidehouse

NEURAL INFORMATION PROCESSING SYSTEMS

Probabilistic modelling for methane leak detection in gas distribution networks

Rubab Atwal and Katie Green

Introduction

Methane emissions from gas distribution networks account for an estimated 1% [1] of total UK greenhouse gas emissions. Timely leak detection and proactive response are key to reducing emissions from the sector.

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

Developing and deploying these methodologies in the real world is challenging due to:

- Sparse data availability across pipe networks
- Sensor readings being inherently noisy, variable and affected by environmental conditions
- Gas Distribution Networks having specific business needs

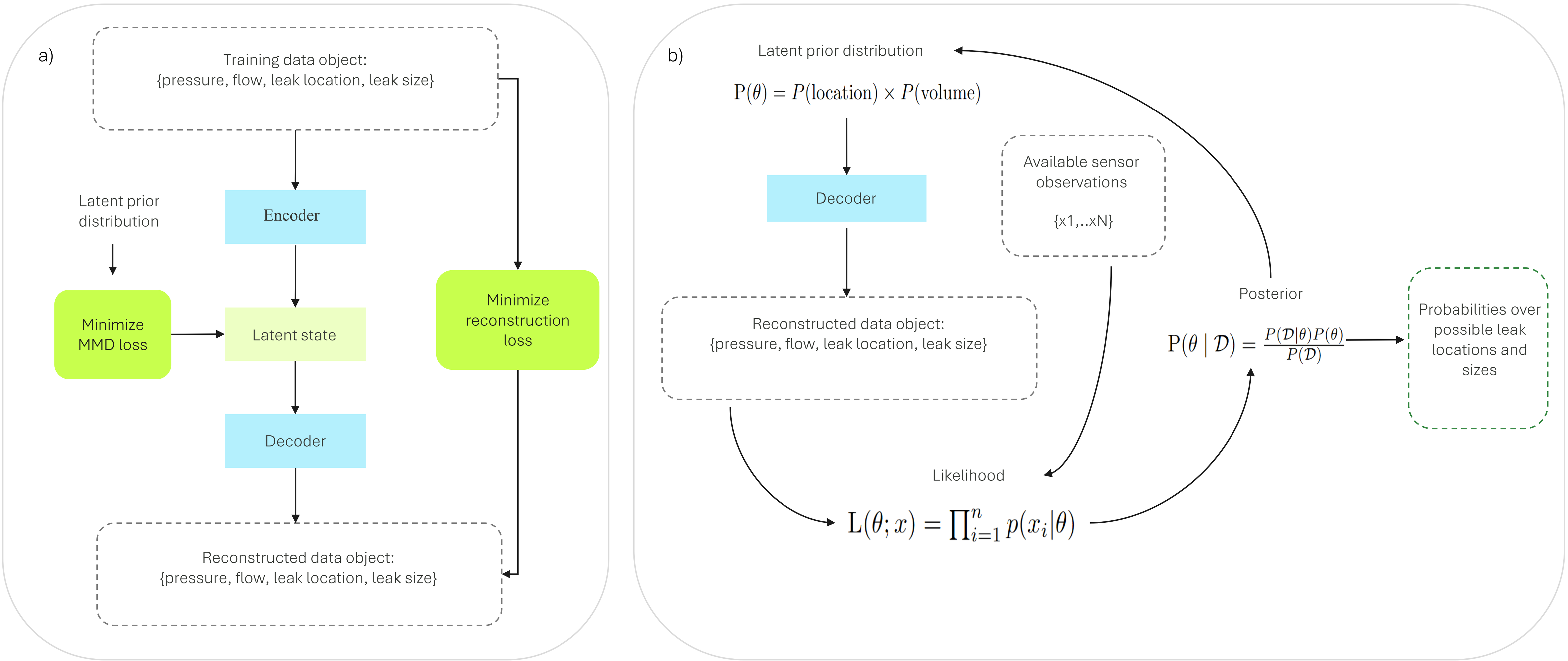
Methodology

We developed a probabilistic model building on the work of Mücke et al. (2023) [2]. Our method involves constructing a graph-based representation of a section of pipe network and simulating training scenarios of operational pressure and flow values under standard operating and leaking scenarios. These scenarios are used to train a Wasserstein autoencoder to act as a surrogate model and approximate the likelihood function required for Bayesian inference.

At run time, operational pressure data is input to the trained decoder and Bayesian inference solver to produce the probabilities across possible leak locations and sizes. After generating probabilities for each pipeline during inference, a statistical t-test is applied to detect potential leaks. When a positive leak signal is identified, post-processing estimates both the leak's location and size. The estimated location is then converted into geospatial coordinates, enabling easy mapping and visualization for end users

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

Model Architecture



(a) Architecture diagram for the process of training the WAE to encode the network dynamics in the latent space and reconstruct the data object.
(b) Schematic for the inference process for probabilistic leak detection through iterative calculation of the posterior from available sensor data

Results

1. The first test case involved running our model on a set of 100 simulated leaks. These simulations assume that data is available from all 12 sensors installed across the network.

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

2. We then tested our model on operational pressure data collected from three historical time periods when leaks were recorded on the network. Across these periods the data available varied from 5-7/12 sensors.

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

3. Due to the importance of sensor data availability demonstrated in experiment 2, we conducted an ablation study to determine the minimum number of sensor readings required for each gas network segment while maintaining the model's expected accuracy. Although 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.

Conclusion

We have developed a probabilistic machine learning framework to detect, localize, and quantify methane leaks in gas distribution systems. Testing on a section of the UK gas distribution network demonstrates that 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.

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, and reduced emissions. Furthermore, the model's ability to accurately quantify detected leaks will help distribution networks enhance emissions estimates, track changes over time, and improve regulatory reporting.

Acknowledgements
The authors acknowledge and thank the Guidehouse and Cadent Gas 'Digital Platform for Leakage Analytics' (DPLA) project teams for their support and contributions to the work. This work has been developed under the Strategic Innovation Fund (SIF) Beta framework funded by UKRI and Innovate UK

References

[1] National Physical Laboratory. (2017). *Understanding the UK greenhouse gas inventory*. Technical report for the Committee on Climate Change. https://www.theccc.org.uk/wp-content/uploads/2017/06/Understanding-the-UK-Greenhouse-Gas-Inventory.pdf
[2] Mücke, N. T., Pandey, P., Jain, S., Bohté, S. M., & Oosterlee, C. W. (2023). A probabilistic digital twin for leak localization in water distribution networks using generative deep learning. *Sensors, 23*(13). https://doi.org/10.3390/s23136179

Contact us

ratwal@guidehouse.com

kmgreen@guidehouse.com