Cadent
Your Gas Network

13/01/2025

Webinar will begin shortly

DPLA
DIGITAL PLATFORM FOR LEAKAGE ANALYTICS

Lead Organisations

Cadent
Your Gas Network

Guidehouse
outwit complexity

SIF Project Partners

SGN
Your gas. Our network.

WALES&WEST UTILITIES

Northern Gas Networks

national gas transmission

Cadent
Your Gas Network

13/01/2025

DPLA

DIGITAL PLATFORM FOR LEAKAGE ANALYTICS

Machine learning modelling for gas leak detection

Lead Organisations

Cadent
Your Gas Network

Guidehouse
outwit complexity

SIF Project Partners

SGN
Your gas. Our network.

WALES&WEST UTILITIES

Northern Gas Networks

national gas transmission

This webinar will provide an overview of the machine learning modelling component of the DPLA including background, model development and results achieved so far

Agenda

- **Introduction**
  - Project goals
  - Modelling benefits
  - Model development timeline
  - Definitions of key terms
- **Model Overview**
  - Introduction to model architecture and methodology for DPLA
- **Results**
  - Results obtained throughout the beta phase to date on Cadent's high and intermediate pressure tiers
- **From results to business outputs**
- **Conclusion and next steps**

Cadent
Your Gas Network

1.
Introduction

Digital Platform for Leakage Analytics (DPLA) aims to significantly reduce gas network leaks and emissions in a cost-effective way

Aim: develop and demonstrate a functional Minimum Viable Product (MVP) for how **data, analytics and models can be used to identify and locate gas leaks in the gas distribution network**.

Core functionality: **data-driven leakage modelling**, unlocking proactive leak detection capabilities, combined with testing the application of novel gas sensor technologies.

Mission: reduce **carbon emissions**, realise **customer benefits** and **improve safety** in a **cost-effective way**

Funding: **DPLA is one of SIF (Strategic Innovation Fund) projects** for the Gas Transmission and Distribution sectors in the UK and it has been developed according to the following phases:

Discovery	Alpha	Beta
• March-April 2022 • Feasibility studies	• August 2022-February 2023 • Experimental development	• September 2023 to June 2025 • Build, operation and/or demonstration

Project partners:

Cadent – Your Gas Network: **Lead Network**

Guidehouse – Outwit Complexity: **Delivery Partner**

national gas transmission, Northern Gas Networks, Wales & West Utilities, SGN: **Network Partners**

This project presents substantial financial, environmental, safety, and consumer benefits

- **Financial** benefits due to lower gas leakage volumes, achieved by targeting larger leaks sooner, leading to lower volumes of gas lost per year and lower shrinkage costs
- **Environmental** benefits as in a 10-year period DPLA could facilitate up to a 58% reduction in methane emissions from pipes and Above Ground Installations (AGIs)
- **Customer** benefits linked to the monetary and social value of the volume of natural gas that would have leaked from the network

Additional benefits and use cases:

Provide Leak Alerts

Proactive Emergency Intervention

Reduce the risk to humans and properties

Improve Network Understanding

Condition Based Monitoring

Improve the understanding of the state of network assets by proactively detecting leaks rather than relying on models alone

Improved Asset Replacement and Maintenance

Leverage improved understanding of the network and leakage hotspots to tailor and better target maintenance cycles and AGI replacement

Accurate Modelling of the Network Leakage

Gas Leakage Regulatory Reporting

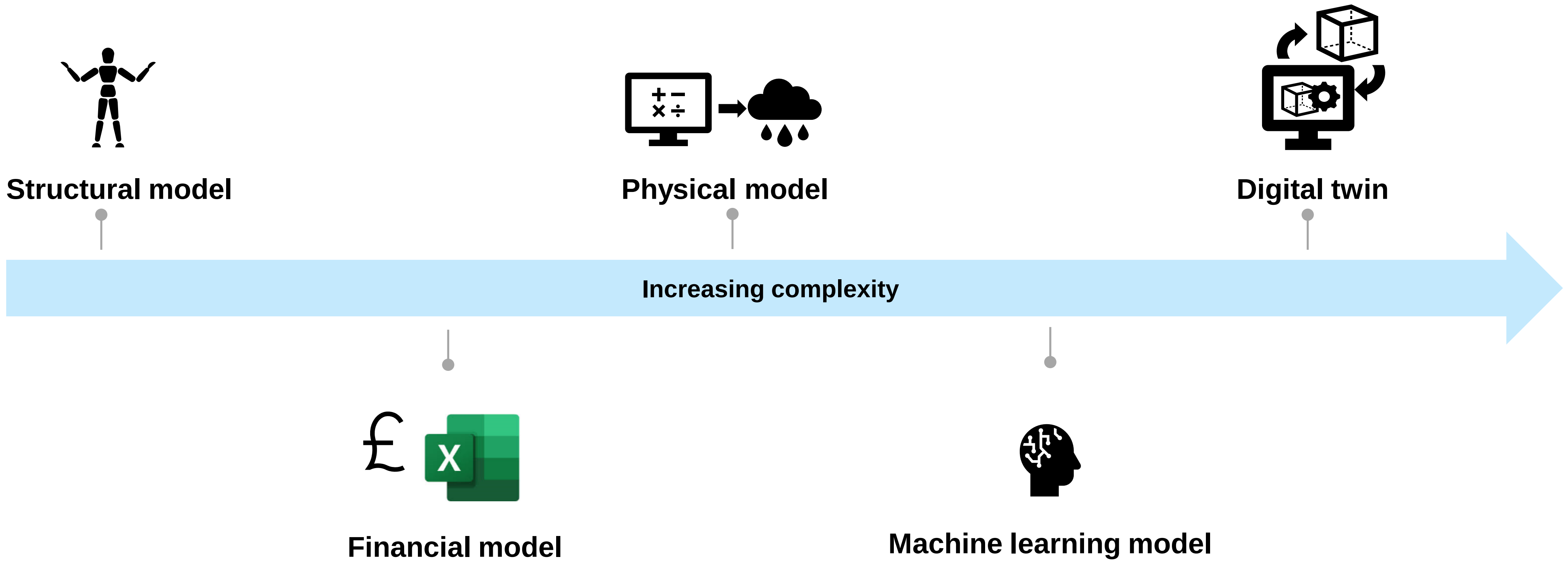
Provide more accurate annual reports of gas leakage compared to the current Shrinkage and Leakage Model (SLM)

Regulatory Performance and Revenue Generation

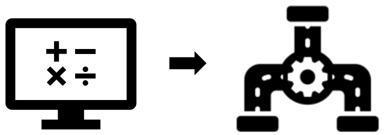
Measure accurately network performance in reducing shrinkage/leakage and develop a fair incentive mechanism

The definition of a model is broad as it encapsulates different approaches to representing any system, process, or object

Models can be any physical, mathematical, or otherwise logical representation of any system, process, or object. Therefore, there are thousands of types of model available which can vary from simple to highly complex dependent on the approach and aim of the model.

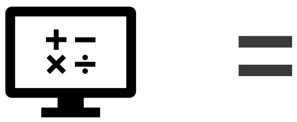


Physical and machine learning models have been combined to develop models for the DPLA



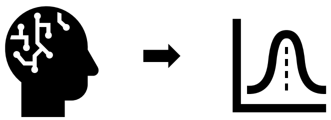
Hydraulic model

Hydraulic models are physics-based models which describe the dynamics of fluids. For gas distribution networks they describe the pressure and flow of gas through pipe networks.



Deterministic model

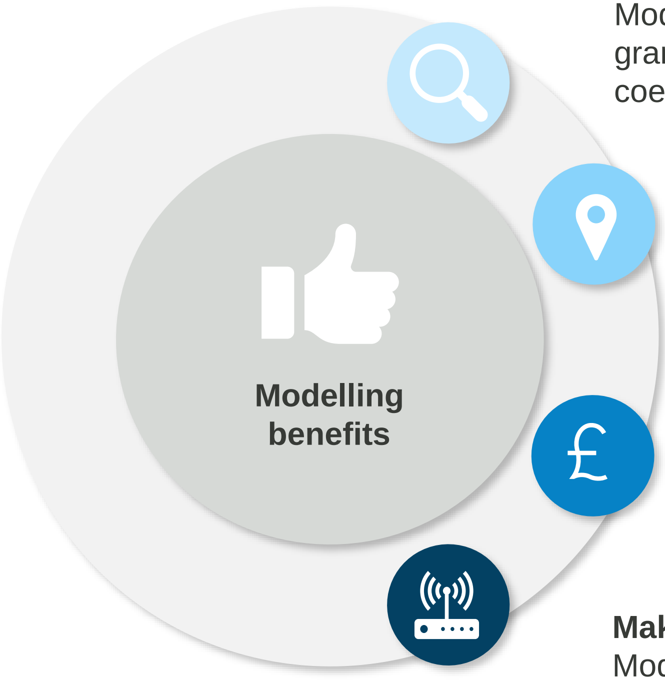
A deterministic model is one which always produces the same definite outcome from a set of inputs. For example, the equations which form hydraulic models can be solved for an exact solution given sufficient inputs. This solution (pressure and flow values) will be the same every time it is calculated from the same inputs.



Probabilistic model

Probabilistic models involve randomness and express outputs as probability distribution describing the most likely result. For example, the DPLA probabilistic model is a machine learning model which learns patterns in training data then applies the 'knowledge' to unseen test data to predict the most likely outcome.

Modelling gas distribution networks for leak detection has many benefits for proactive response and emissions management



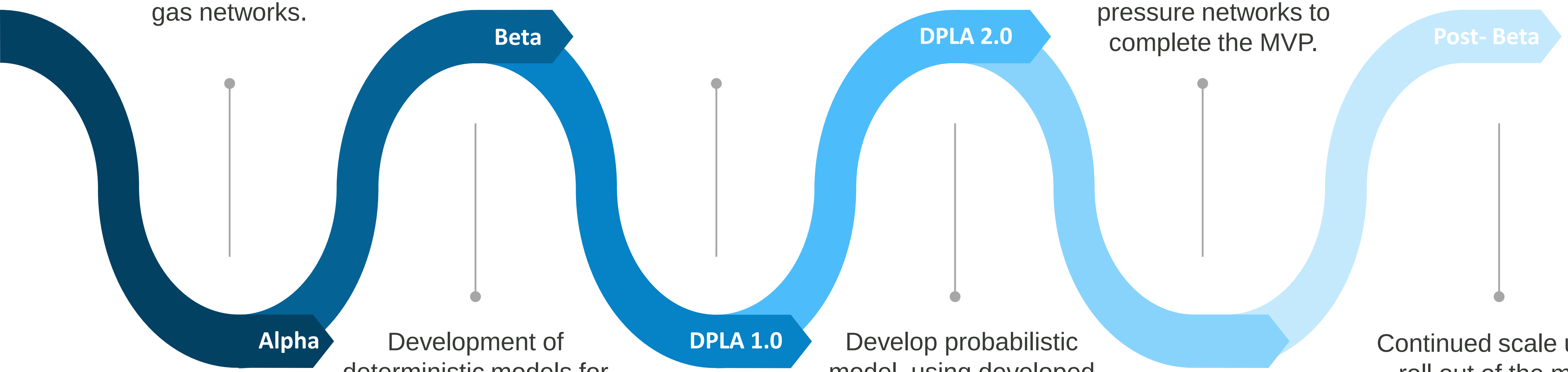
Improves understanding of network dynamics and leakage
Modelling the network for leakage detection improves understanding of leakage across the network and at a more granular, pipe level in comparison to the current Shrinkage Leakage Model (SLM) which extrapolates estimated asset coefficients across the network.

Accurately detects and quantifies leaks
Models can accurately detect and quantify leaks. Leaks can be detected as soon as they occur before they are measured or reported and effectively quantified.

Avoids sampling errors and ongoing costs of periodic measurements
Intermittently surveying gas networks with in-field methane detection technologies introduces error as measurements only provide snapshots of network leakage unlike models which can be run continuously. These surveys also have high ongoing costs whereas modelling involves lower initial investment.

Makes use of pressure sensors already installed
Models can be applied to detect leaks using only pressure data from across the network. Pressure sensors are already installed for operational monitoring across gas networks which provide the necessary data.

As an innovation project, the modelling methodology has developed and adapted to balance business needs and data limitations



Proof of concept for deterministic models for leak detection on gas networks.

Development of deterministic models for leak detection on gas networks.

Assessment of applicability to Cadent's network.

Develop probabilistic model, using developed deterministic model to produce training data, for example network segments.

Optimisation of model pipeline and scale up across Cadent's high and intermediate pressure networks to complete the MVP.

Continued scale up and roll out of the model across other networks.

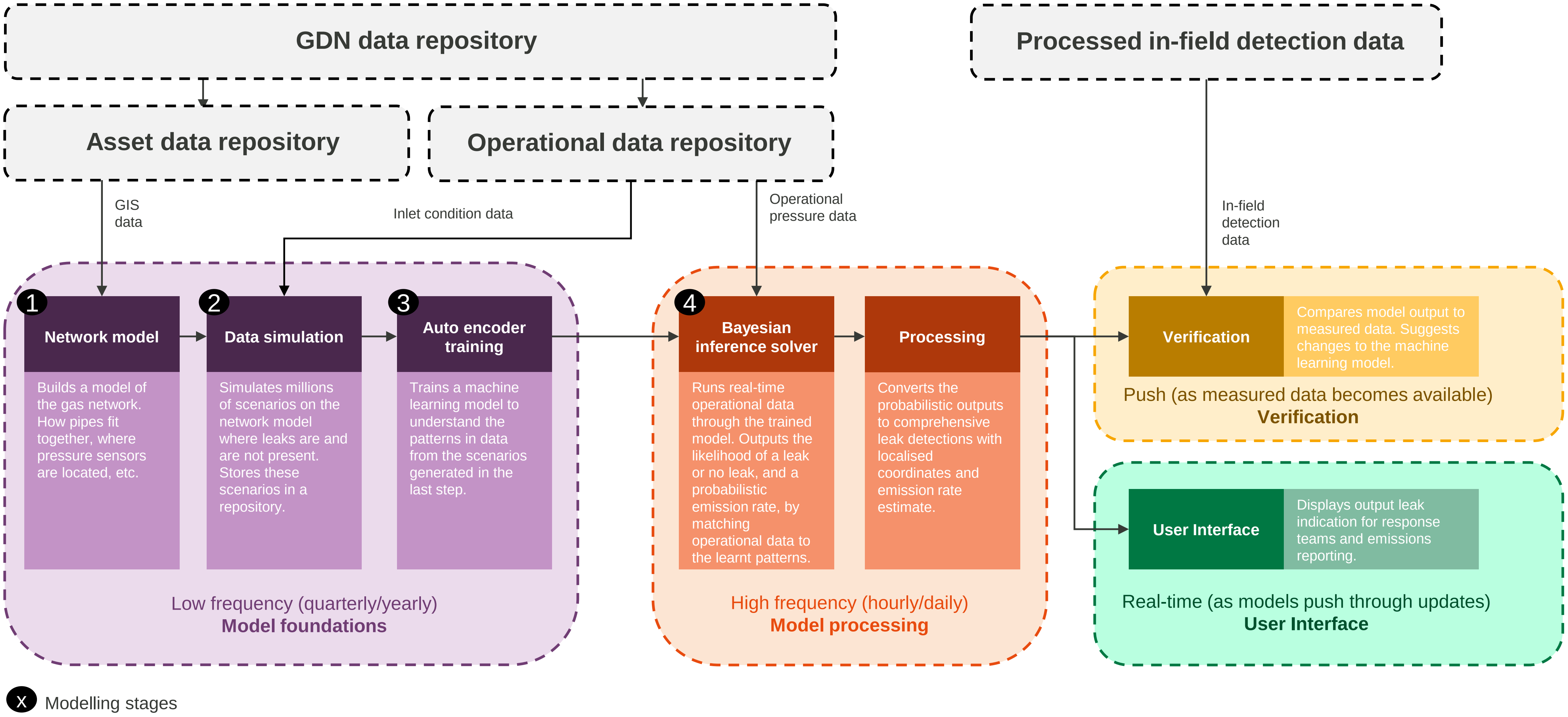
There is lots of terminology within the field of probability theory, so definitions of common terms are provided:

Probability/Certainty	Probability is the certainty of an event. In the case of leak detection, a probability of 0% means there is certainly no leak, 100% means there is a leak, and 50% is unsure either way.
Accuracy	Accuracy is a measure of how often the model makes a correct prediction as a fraction of the total number of predictions. To calculate the accuracy, a threshold must be defined for a correct prediction.
Prior	The prior is a probability informed by the previous information seen by the model. It can be thought of as an initial best guess.
Data Likelihood	The likelihood is a probability informed by the most recent information given to the model.
Posterior	The posterior is the resulting probability from updating the prior using the likelihood. The calculated posterior probability is the certainty of a leak occurring at a given location.
Bayesian Inference	Bayesian inference is the most advanced form of probabilistic analysis. It is used in our probabilistic model to generate predictions for leak locations.

Cadent
Your Gas Network

2.
Model overview

End-to-end modelling process overview

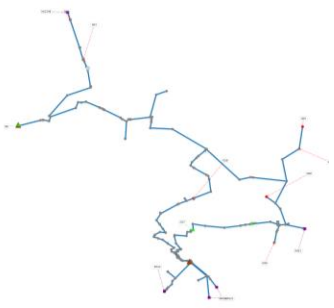


Stages 1 & 2 break down: configuring the network model and generating data to train the probabilistic model

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.

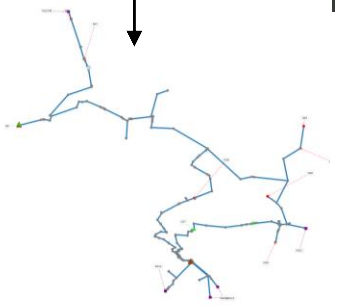


Network models are created for each pressure tier within each network. These are then broken down into 'segments' (smaller sections of the network which are contained by pressure sensors) to reduce the computational modelling load.

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

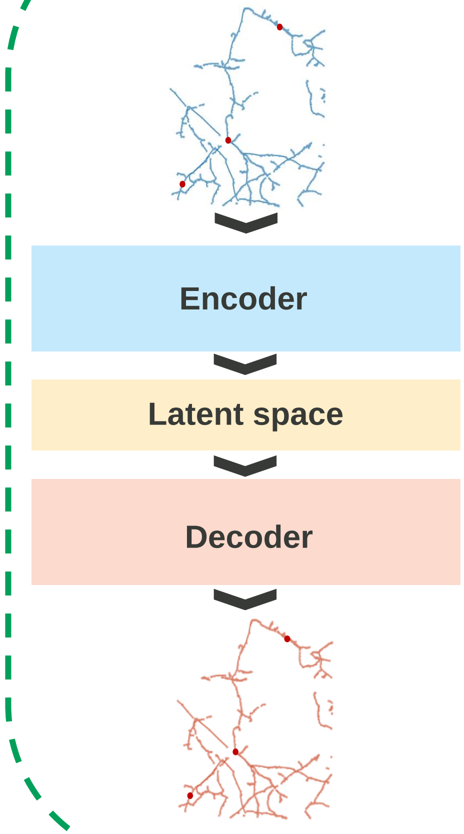
The possible inlets are input into hydraulic equations to simulate pressure are flow values across the segment.

Training scenarios

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

Stages 3 & 4 break down: training the auto-encoder and applying it to operational data to perform Bayesian inference

Stage 3: Training the auto-encoder

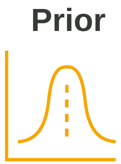


Millions of samples are generated for various leak scenarios.

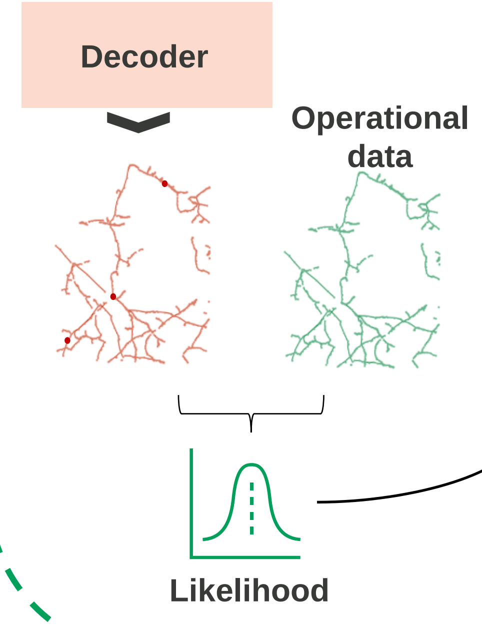
A machine learning algorithm is trained to learn patterns in the training data and compress the millions of scenarios to a smaller representation (latent space).

The algorithm also learns to take compressed representations and reconstruct the approximate pressure and flow values for a given leak location and time.

Stage 4: Probabilistic leak detection

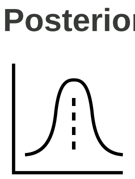


The prior distribution is generated and fed through the trained decoder to produce reconstructed data



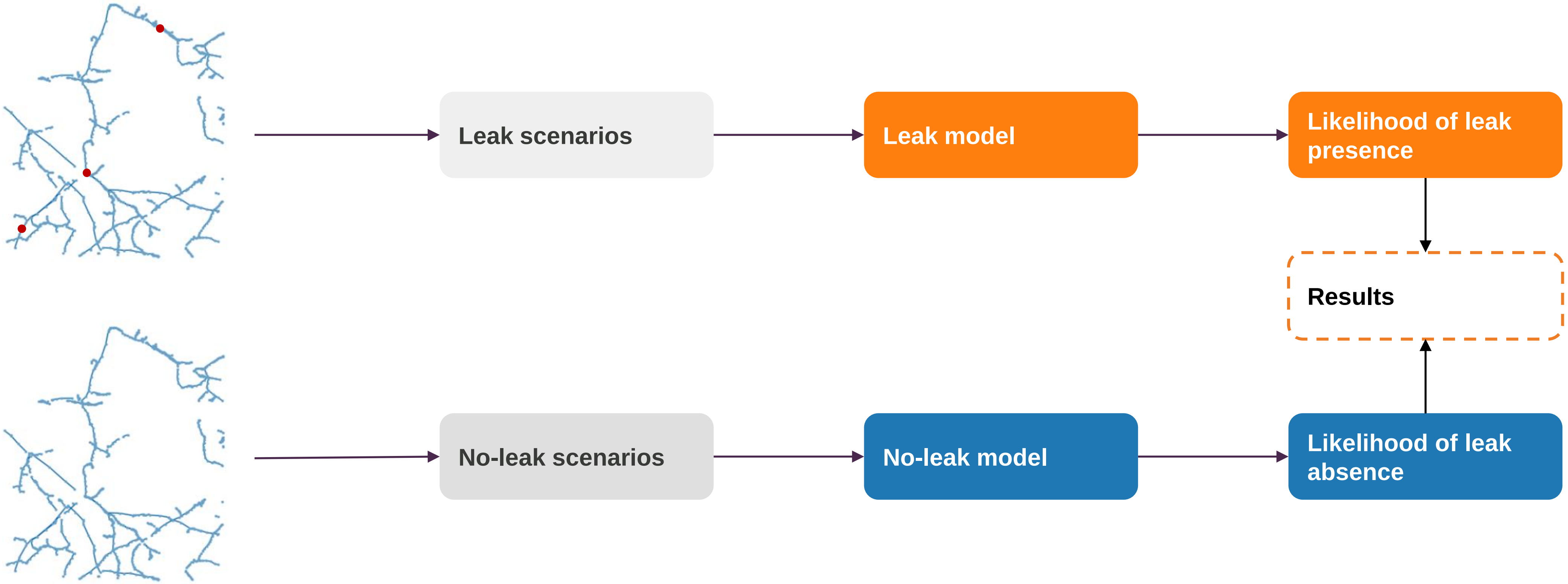
This data is combined with the real operational data to produce the likelihood

The likelihood is used to update the prior and produce the posterior



The posterior replaces the prior as the model moves on to iterate through the next time step

Two models are trained to produce separate probabilities for the presence and absence of a leak which are compared for leak detection

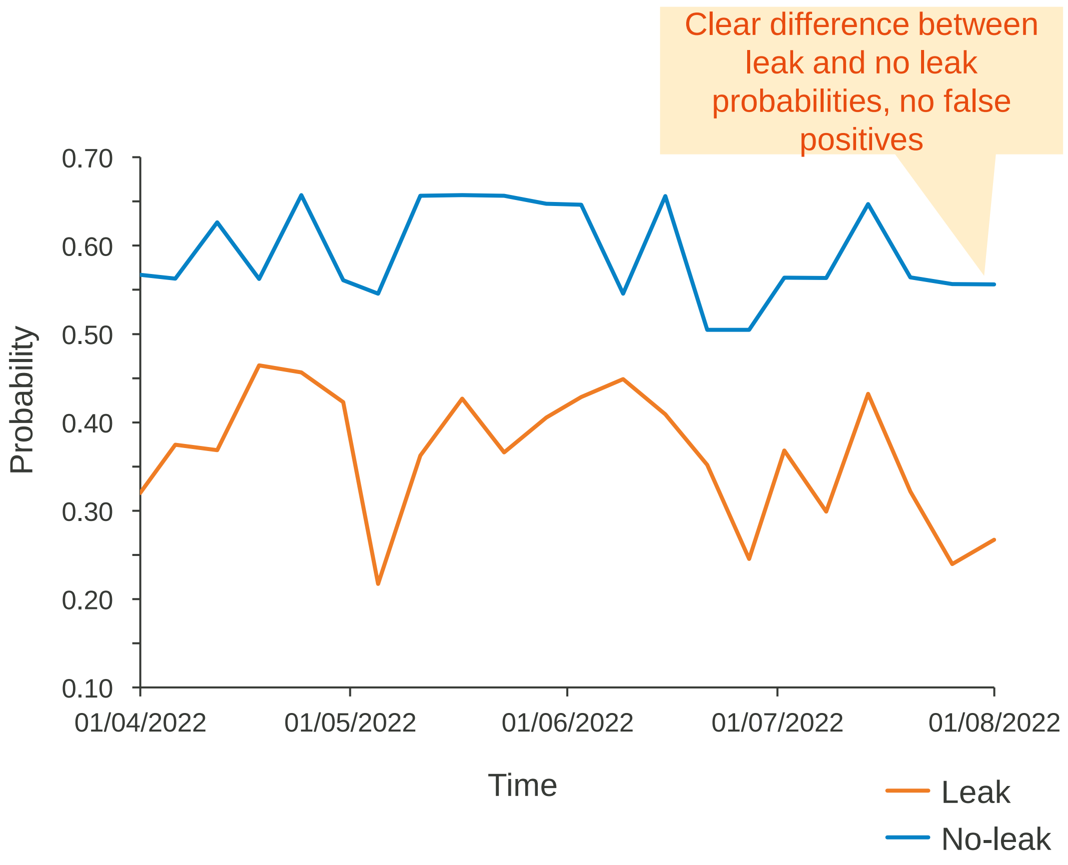


Cadent
Your Gas Network

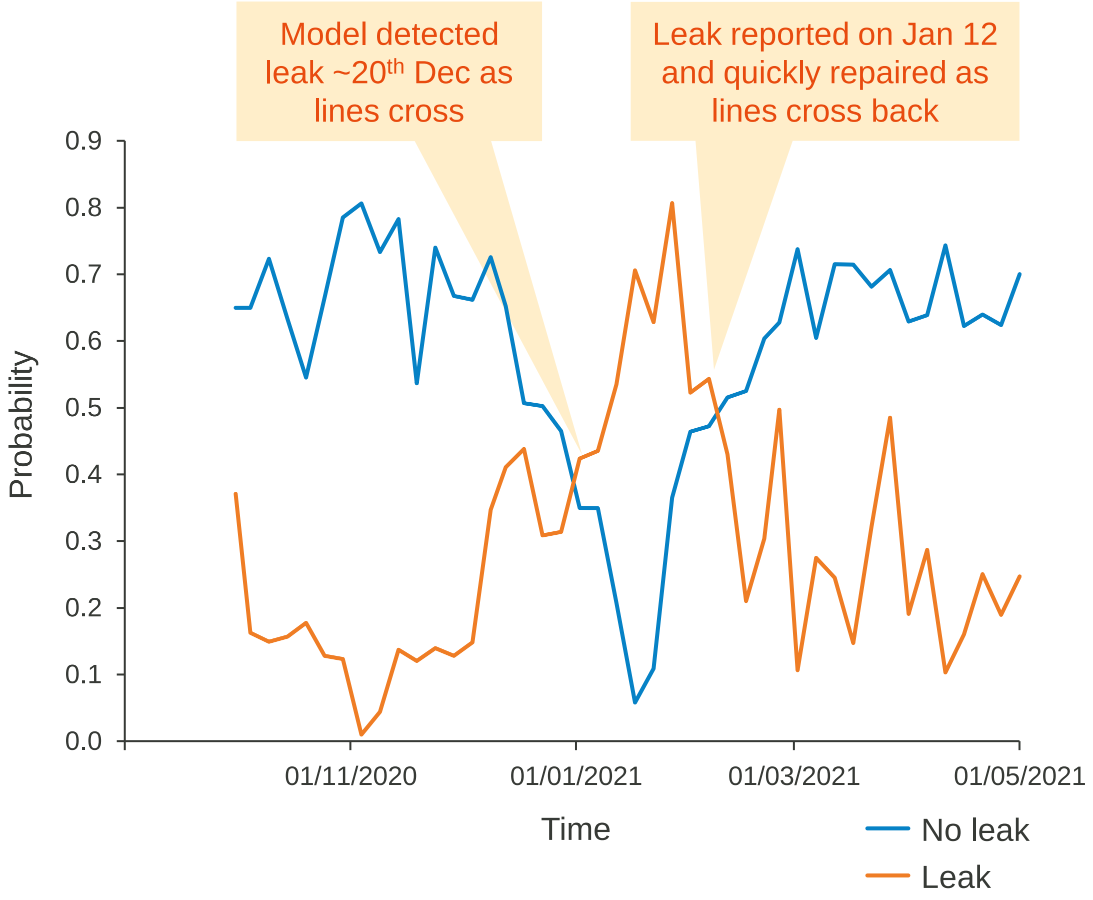
3.
Results

The model outputs two probabilities which give the likelihood of a leak in each pipeline and inform leak detection

Status quo

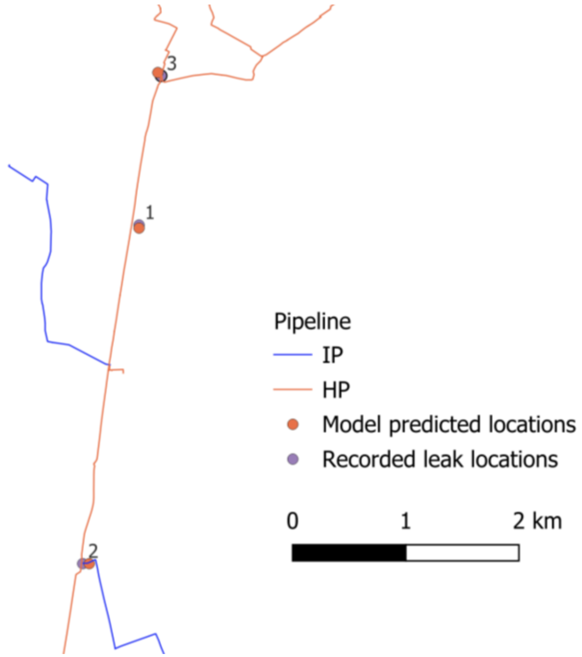


Leak detection



The model accurately detected four historically recorded leaks and localised them within 60m

Leak localisation – High pressure

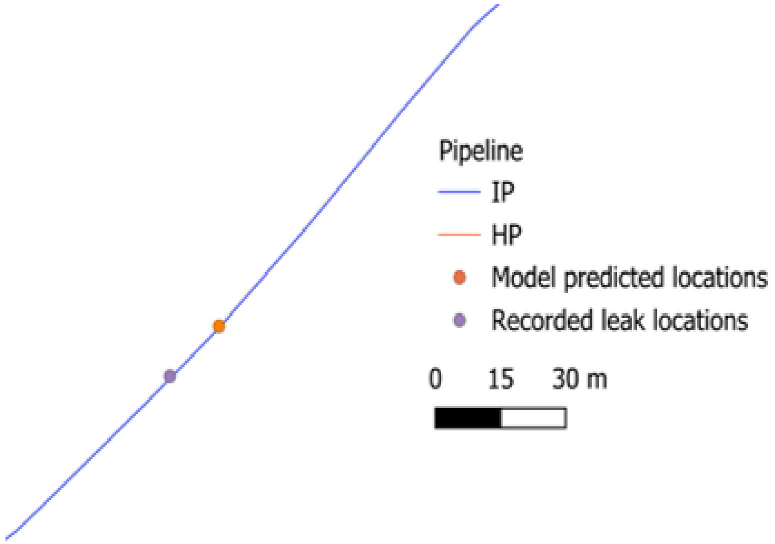


The coordinates of the predicted leaks can be plotted geospatially with the pipelines and recorded leak locations

Leak #	Distance between model and actual (m)
1	27
2	58
3	40

The distance measured between the model result and the actual leak location was <60m for all three historical leaks.

Leak localisation – Intermediate pressure

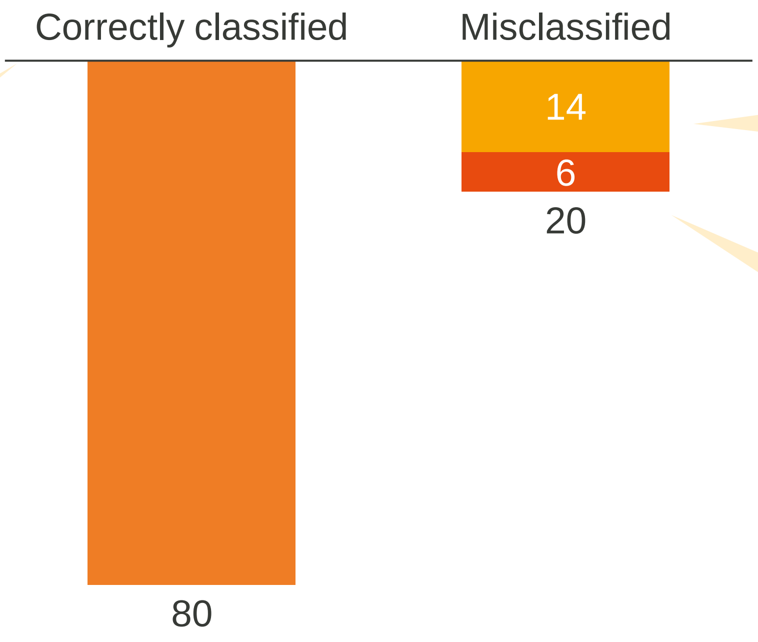


The coordinates of the predicted leaks can be plotted geospatially with the pipelines and recorded leak locations.

Historic emission rates are not recorded so the model quantification was tested on 100 synthetically generated leak scenarios of varying size

Leak quantification

The model quantifies leaks by selecting the emission rate range with the highest probability



14 leaks were estimated the adjacent range to their actual value

6 leaks were in the smallest size range making accurate quantification more challenging

Across the test set of 100 leaks 80 were correctly quantified and 20 misclassified.

Cadent
Your Gas Network

4.

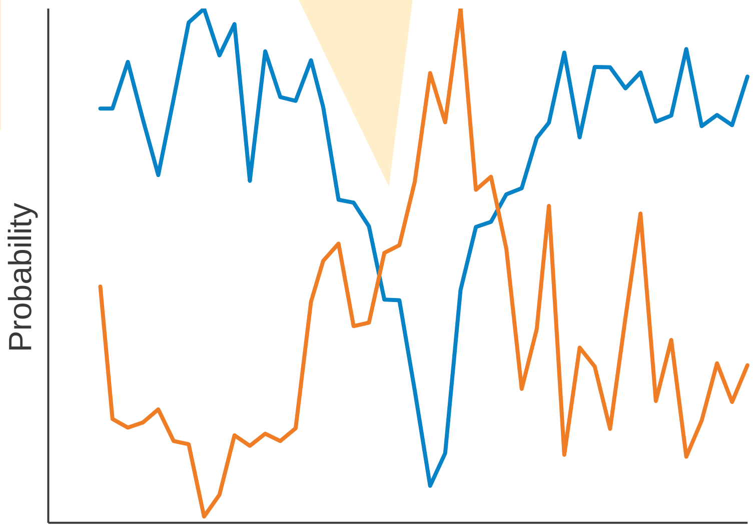
From model results to business outputs

The results from the probabilistic model will be processed to inform leak indications which are provided to users of the DPLA

Results processing

The model outputs are in the form of probability distributions for each pipeline which can be challenging to interpret and analyse for untrained users.

A leak indication will be triggered when a probability threshold is reached



Leak indications will be produced in a format that can be displayed in a user interface and interpreted for proactive intervention and emissions reporting.

Leak #
X coordinate
Y coordinate
Asset
Leak rate

Cadent
Your Gas Network

5.
Next steps

The focus for the remainder of the project is to scale-up the model and ensure the outputs meet the business needs

Results processing

Exploring, developing, and testing methodologies to convert probabilistic outputs to leak indications will be key to ensure the probabilistic model successfully meets the business needs of the DPLA solution.

Pipeline automation

Continuing to automate the model pipeline and ensure the process can be run from end-to-end will enable the successful scale up of the model to across networks.

Scale-up process

Having the computing infrastructure in place for developing the model at a larger scale will also be key for enabling the rollout across networks.

Q&A