



CFMS Services Ltd.

0498 Predictive Tool for Unaccounted for Gas Identification Technical Report

D3.1 Final Technical Report

© 2026 CFMS Services Ltd.

The Copyright in this work is vested in CFMS Services Limited. The document is issued in confidence solely for the purpose for which it is supplied. Reproduction in whole or in part or use for tendering or manufacturing purposes is prohibited except under an agreement with or with the prior written consent of CFMS Services Limited and then only on the condition that this notice is included in any such reproduction.

Commercial In Confidence

Document Management

Reference	D3.1 Final Technical Report
Issue No.	1.0
Issue Date	27/3/2026
Submitted to	Doug Millington-Smith, National Gas Transmission
Classification	Commercial In Confidence

Authorisation

	Name	Title/Role	Signature
Prepared	Joanna Tumelty	Senior Research Engineer	
Approved	Jory Griffin	Senior Research Engineer	
Authorised	Alexander Ballisat	Head of Engineering	

Amendment History

Issue	Section	Description of Change	Date	Initials
0.1	All	First Draft	13/03/2026	
1.0	All	Release Version	27/03/2026	

Distribution

Name	Organisation
Doug Millington-Smith	National Gas Transmission

Executive Summary

Unaccounted for Gas (UAG) is gas that is lost, or otherwise not accounted for, on the National Transmission System (NTS). UAG is a significant problem for the UK's gas network and forms a dominant component in NTS shrinkage. There are significant financial, reputational and operational impacts associated with UAG. Methods to better understand and identify key sources of UAG have the potential to reduce this impact.

Unaccounted for Gas predominantly arises from metering errors in meters across the network. Compounded uncertainty due to these errors from the large numbers of meters across the network means that using typical data-driven approaches to identify metering errors is difficult. Physics-based simulations, which can predict the expected pressure and flow dynamics of the gas from the underlying governing physical equations, however, could be used to identify potential metering errors by identifying meters with discrepancies between the recorded telemetry data and the predicted flow dynamics.

The main aims of this project were to prototype a forward predictive physics-based model that simulates the gas flow within sections of the NTS given Supervisory Control and Data Acquisition (SCADA) inputs, perform statistical analysis on the results to identify discrepancies that may indicate metering errors and sources of UAG. This report details the technical work undertaken during the project and outcomes from the four phases of the project—Phase 1: Data Discovery and Formatting; Phase 2: Model Development; Phase 3: Model Testing and Validation; and Phase 4: Standards and Reporting.

National Gas Transmission provided CFMS with a number of datasets defining the sites and pipelines within a section of the network; historic SCADA data for sites in this section of the network; and energy usage data across the NTS. Exploratory analysis was performed on these datasets to identify missing information and identify appropriate modelling assumptions to account for the absence of these data. The SCADA data required preprocessing to improve the quality of the data to enable it to be incorporated within the model. This involved handling a range of features in the data, including missing values, irregular recorded measurements and pressure transmitter errors.

A forward predictive physics-based model simulating the gas flow within sections of the NTS using SCADA inputs was developed. The simulation is implemented using the Portable Extensible Toolkit for Scientific Computing (PETSc), which enables its extension to the full NTS in future through the use of more compute resources. The tools required to take SCADA and other network definition information were also developed to enable the automated instantiation, execution and analysis of simulations. The simulated results were compared to physical measurements in the SCADA data to identify discrepancies that could correspond to metering errors and hence impact UAG.

The performance of the model was assessed on a series of test cases. This included simple test cases to validate the physics of the model, small subsections of the NTS to validate the model pipeline, two test cases of historic metering errors, and finally it was extended to a significant portion of the North East section of the NTS. The results demonstrated a good qualitative agreement between simulated and SCADA data, demonstrating the feasibility of this approach. Furthermore, the effect of metering errors was demonstrated through the application of synthetic perturbations to the SCADA data, showing that it should be possible to identify these.

Overall, the project successfully demonstrated the viability of this approach. However, it has also highlighted a number of avenues of future work required before this approach can be extended the full NTS. The primary challenge in this work has been the parsing and processing of SCADA and network topology data, which has required significant manual intervention to undertake, and affected the quantitative accuracy of the model's output. Developing automated pipelines to do process SCADA data robustly in the future, and handling more edge cases, will be critical to improve confidence in the metering error identification and for extension to the full NTS. Other areas of future work include further development of the physics model and validation of the assumptions made, improving the computational efficiency of the model and executing a wider range of test cases.

Contents

1	Introduction	6
1.1	Purpose and Scope	6
1.2	Document Structure	6
1.3	Background	6
1.4	Project Objectives	7
2	Phase 1: Data Discovery and Formatting	8
2.1	Network Definition	8
2.2	Network Modelling Assumptions	8
2.2.1	Short Pipelines	8
2.2.2	██████████	8
2.3	SCADA Data	8
2.4	NTS Energy Usage Data	10
2.5	Key Deliverables	10
3	Phase 2: Model Development	12
3.1	Model Overview	12
3.2	Gas Flow Simulation	12
3.2.1	Network Simulation Integration	14
3.2.2	Extension to Gas Networks	16
3.2.3	Statistical Analysis of SCADA and Simulated Data for Metering Error Identification	16
3.3	Synthetic Metering Errors	17
3.4	UAG Calculation	18
3.4.1	Introduction	18
3.4.2	Energy Difference And Throughput	19
3.4.3	Linepack And Delta Linepack	19
3.4.4	Compressor Fuel Usage	20
3.4.5	Calorific Value Shrinkage	21
3.5	Key Deliverables	21
4	Phase 3: Model Testing and Validation	22
4.1	Verification Test Cases	22
4.2	██	25
4.2.1	Test Case Definition	25
4.2.2	Baseline Simulation Results	26
4.2.3	Baseline Energy Balance Results	26
4.2.4	Application of Synthetic Metering and Transmitter Errors	28
4.2.5	Metering Errors on Flow Meters	33
4.3	██████████ Test Case	37
4.3.1	Historic Metering Errors at ████████ Power Station	37
4.3.2	Simulation Results	38
4.3.3	Metering Error Identification	38
4.3.4	Unaccounted for Gas	40
4.4	Application to the North-East of England Section of the Network	42
4.4.1	Problem Overview	42
4.4.2	Test Case Definition	42
4.4.3	Simulation Results	42
4.4.4	Energy Balance Calculation	42
4.5	Discussion and Interpretation of Results	46
4.6	Key Deliverables	47

5	Next Steps	49
5.1	Improving Model Accuracy	49
5.2	SCADA Data Processing	49
5.3	Application to Additional Test Cases	50
5.4	Extension to National Transmission System	50
6	Cost Benefit Analysis	51
6.1	Costs Associated with Unaccounted for Gas	51
6.2	Benefits of the Solution	51
6.3	Replication Potential in a Wider Roll-Out	51
7	Conclusions	53
A	Convergence Study	55

1 Introduction

1.1 Purpose and Scope

This report serves as the final proof-of-concept report and governance pack for the project: Predictive Tool for Unaccounted for Gas (UAG) Identification, which CFMS undertook in collaboration with National Gas Transmission (NGT) between September 2025 and March 2026. The purpose of this document is to demonstrate how the key deliverables of the project, outlined in the Statement of Work (SoW) [1], have been achieved, or, the modifications to the deliverables that were required due to the reduced scope of the project. The report focuses on the final model developed, and is not an exhaustive account of the investigative and exploratory work undertaken during the project.

1.2 Document Structure

This document is structured as follows. Sections 1.3– and 1.4 summarise the background and the objectives that this proof-of-concept project aims to address. Section 2 details outcomes from Phase 1 of the project, which includes outlining datasets provided by NGT, and how these were processed during model development. Section 3 provides an overview of the developed model prototype and technical details about the modelling approach and assumptions made. Section 4 summarises the outcomes and performance of the model when applied to a set of test cases of varying complexity. Each of Sections 2–4 conclude with a summary of the contributions to the key deliverables that were documented in that section. These summaries includes any limitations of the current model that meant that these deliverables could not be fully addressed. Section 5 outlines recommendations for next steps. Section 6 provides the Cost Benefit Analysis (CBA) for the solution developed in this project. Section 7 concludes the report. An appendix is included to document additional results derived in this project.

1.3 Background

Unaccounted for Gas (UAG) is gas that is lost, or otherwise not accounted for, on the National Transmission System (NTS). Unaccounted for Gas is a significant problem for the UK's gas network and forms a dominant component in NTS shrinkage—estimated at 46% in the 2025/2026 formula year [2]. There is a significant financial impact associated with UAG, which is typically passed on to end-customers, as well as reputational and operational impacts as NGT are required to investigate and reconcile sources of UAG for regulatory compliance. Reducing this impact motivates the development of methods to improve understanding about when and how UAG arises.

It has been recognised that UAG predominantly arises from metering errors in measured data from meters distributed across the network, rather than physical gas loss. These metering errors can arise from a variety of sources, including:

- Data and telemetry inconsistencies,
- Transmitter drift or failure,
- Physical meter errors,
- Human error when returning meters back to service following an inspection,
- Incorrect flow computer calculations.

These metering errors are the cause of UAG and NTS shrinkage. The compounding effects of these meter errors across the large numbers of meters in the network mean that it is challenging to identify isolated metering errors and localise these sources of UAG.

Previous approaches to identifying sources of UAG have typically been data-driven, for example by identifying statistical variations in operation from the baseline UAG behaviour and analysing trends. However, given competing effects from the large numbers of meters across the network and variability in metering errors, reliably isolating individual metering errors and sources of UAG using these methods has been

found to be difficult. Physics-based simulations, however, which can predict the expected pressure and flow dynamics of the gas from the underlying governing physical equations, could be used to identify potential metering errors by identifying meters with discrepancies between the recorded telemetry data and the predicted flow dynamics.

The main aims of this project were to prototype a forward predictive physics-based model that simulates the gas flow within sections of the NTS given Supervisory Control and Data Acquisition (SCADA) inputs and perform statistical analysis on results to identify discrepancies that may indicate metering errors and sources of UAG. This report details the technical work undertaken during the project and outcomes from the four phases of the project:

- Phase 1: Data Discovery and Formatting.
- Phase 2: Model Development.
- Phase 3: Model Testing and Validation.
- Phase 4: Standards and Reporting.

This includes technical documentation of the modelling approach, modelling assumptions and limitations, the performance of the model when applied to validation and test cases, and recommendations for next steps required to scale the model and approach to the entire NTS.

1.4 Project Objectives

The primary objective for this project was to demonstrate the viability of using physics-based simulations of the gas flow dynamics to identify high-risk meters and potential sources of UAG. The key steps to achieve this are as follows:

- Develop a forward predictive model that takes in historic telemetry data and simulates the expected network dynamics.
- Quantify the difference between simulated data and telemetry data to identify potential metering errors.
- Demonstrate the model's ability to isolate and attribute UAG.
- Validate the model's accuracy and functionality through the application of the model on a selection of test cases.
- Identify next steps to develop the model to scale the model across the wider network in a pilot deployment phase.

2 Phase 1: Data Discovery and Formatting

National Gas Transmission provided CFMS with a number of datasets defining the sites and pipelines within a section of the network; historic SCADA data for sites in this section of the network; and energy usage data across the NTS. This section summarises the information contained within the provided datasets and the work undertaken during Phase 1 of the project to preprocess and standardise these datasets so that they could be incorporated into the model. Over the course of the project it was recognised that some of these datasets were missing information that was required to develop the model. This section also outlines these gaps and the modelling assumptions that were made to account for the unavailable data.

2.1 Network Definition

A section of the NTS in the North-East of England, shown in Figure 1, was considered in this project. National Gas Transmission provided CFMS with 18 months of historic SCADA data between 01/04/2023 and 01/10/2024 for sites with Above Ground Installations (AGIs) within this section of the NTS. The pipeline databook [3] was used to define the lengths and diameters of pipelines that connect these sites.

2.2 Network Modelling Assumptions

It was not possible to define the lengths and diameters for all pipelines in this section of the NTS directly from the pipeline databook, and consequently the following modelling assumptions were made.

2.2.1 Short Pipelines

Detailed site schematics and the precise locations of AGIs within each site and the lengths and diameters of the pipelines connected these AGIs are not included in the pipeline databook. To account for this, sites that did not have a pipeline in the pipeline databook directly associated to it were linked to other sites in the network according to the site name (for example [REDACTED]). The pipeline associated with this site name (e.g. [REDACTED]) in the pipeline databook was connected to the junction site within this group of sites and short pipelines with default length 0.1km and diameter 900mm were introduced to connect sites of other types (e.g., offtake, power station, compressor) to this junction. Details about the pipelines that connect offtakes, and other exit sites, to the main feeders in the NTS were also not specified in the pipeline databook. Consequently, short pipelines were also added to the network in these cases, with Section 3.2 providing further details about this.

2.2.2 [REDACTED]

Modifications to the network were also made around Easington due to missing SCADA data for [REDACTED]. An additional site at [REDACTED] was introduced to connect the subterminals [REDACTED] to the NTS. This site was connected to [REDACTED] through the single pipeline specified in the pipeline databook. Through discussions with NGT, however, it was identified that in fact three pipelines connect [REDACTED] for redundancy purposes to handle the high flow volumes in this region of the network. These additional pipelines were not included in this version of model but could be in future work.

2.3 SCADA Data

The SCADA data was provided in the form of two Excel spreadsheets, and contained historic telemetry data for measurements of the following quantities:

- Pressure (P).
- Volumetric flow rate (F).

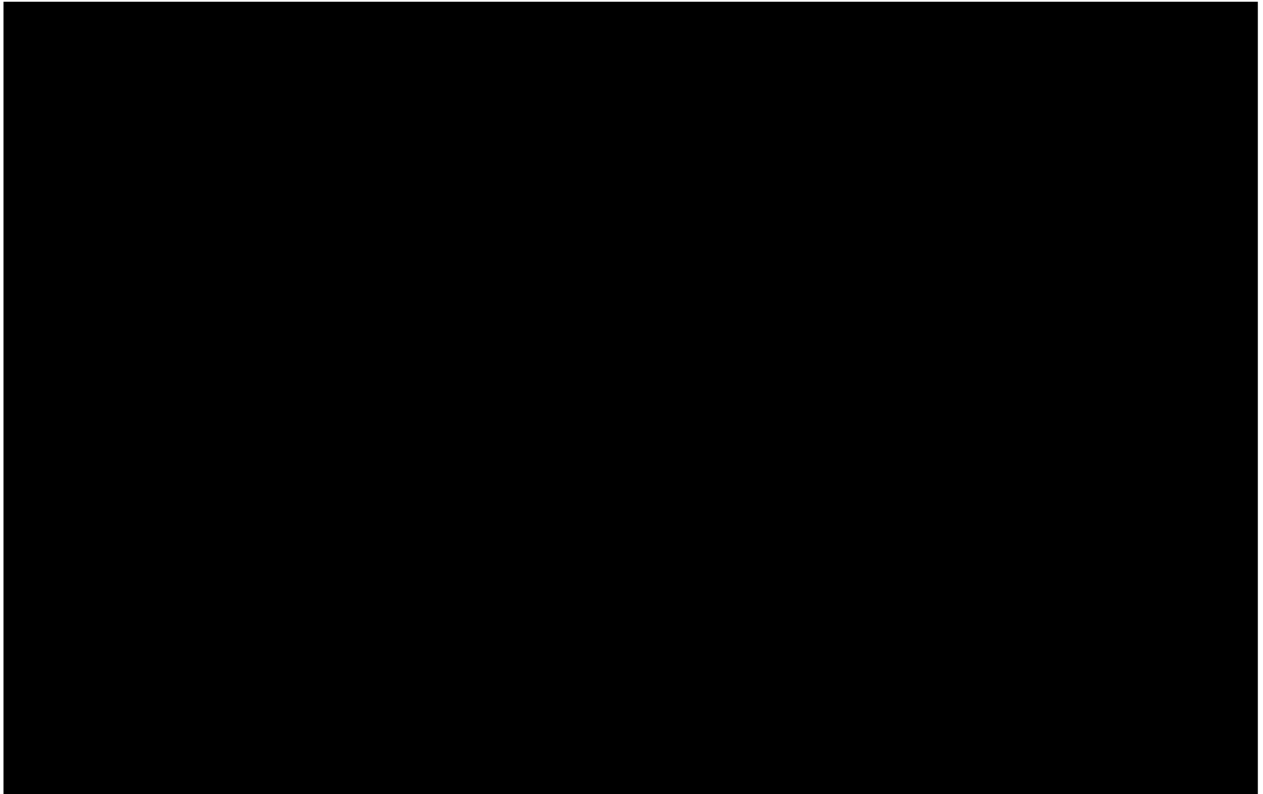


Figure 1: Map showing sites and pipelines of the section of the NTS in the North-East of England.

- Energy flow rate (EF).
- Integrated volumetric flow (IntegF).
- Integrated energy flow (IntegEF).
- Daily integrated volumetric flow (DVOL).
- Daily integrated energy flow (DVOLEF).
- Calorific value (CV).
- Temperature (T).

A subset of these fields, as outlined in Table 1, were used in the model. The SCADA data for these fields required pre-processing in order to identify and exclude instances of low quality data. This included identifying and removing data points with the following features:

1. Pressure transmitter errors, when $P = -9999$.
2. Pressure values outside of the range $[0, 100]$ bar.
3. Flow values below -0.1 mcm/d.
4. Negative energy flow values.
5. Values with reliability indicated as 'Unreliable'.

A summary of the number of the data points removed and the reason for their removal can be outputted in a data cleaning summary, stored in a Comma Separated Variable (CSV) file, both for the full data set, and for a subset of sites over a reduced time period. Following the removal of the erroneous data points, the fields were interpolated using a Piecewise Cubic Hermite Interpolating Polynomial to handle the irregular

Abbreviation	Data Field	SCADA Data Units
F	Volumetric Flow Rate	mcm/d
EF	Energy Flow Rate	MMJ/d
P	Pressure	barg

Table 1: Summary of SCADA data fields and units for variables that were used in the model. It should be noted that the units for this telemetry data are in UK standard units, which are related to the standard conditions at 1.01325 bar and 15°C.

recorded measurement frequency in the SCADA data. This ensured that each field had values on a regular time base with frequency of a minute.

Additional data cleaning steps could be undertaken to further improve the quality of the SCADA data input to the model. This includes the removal of static data, integer data and large spikes or discontinuities in the data fields. However, given the variability in the SCADA data, performing these additional steps was determined to be out of scope in this project, due to the additional domain knowledge required to reliably remove these features. Instead, for the purposes of this project, the input SCADA data was manually checked for quality and the dates for each of the test cases was chosen for the gas flow simulation to run without instability.

2.4 NTS Energy Usage Data

National Gas Transmission also provided a set of commercial data associated with the daily energy usage across the NTS, which is used to calculate UAG. These datasets consisted of:

- Supply and Demand Energy data [4], detailing the site-level daily volume and energy associated with the supply of gas from entry sites and the demand of gas at exit sites.
- Energy Report Template [5], detailing the total daily measured and billed energy, and Calorific Value Shrinkage (CVS) per Local Distribution Zone (LDZ).
- Compressor Usage [6], detailing the total daily volume, energy and calorific value for compressor usage in the network.
- Offtakes By LDZ [7], listing the offtakes associated with each LDZ.

2.5 Key Deliverables

The Statement of Work [1] identified the following three key deliverables for Phase 1: Data Discovery and Formatting. This section summarises the extent to which these deliverables were achieved in this project.

- **1.1: Cleaned and Formatted Dataset:** This deliverable was fully met and evidenced in Section 2.

Section 2.3 described the dataset consisting of SCADA pressure, volumetric flow and energy flow data and the data cleaning and formatting steps undertaken to validate and standardise SCADA data input in the model. The processing steps undertaken, specifically the removal of pressure transmitter errors and interpolation to fill in missing data, allowed the simulation to run without instability for each of the test cases in the model and enabled a working prototype of the model to be developed. This deliverable was therefore achieved for the purposes of this proof-of-concept project. However, the significant variability in the SCADA data across different sites in the network, and particularly the inconsistent presence of large spikes and discontinuities in the data, proved to be a challenge within this project due to limited time and domain-specific knowledge required to process the data. Over the course of the project, it was found that anomalous features in the input SCADA data significantly affected the simulation results, and presented similarly to metering errors. Additional pre-processing of the SCADA data will be essential in any future work to remove these anomalous features, improve

confidence in the validity of input data and the quantitative conclusions that can be drawn from the model.

- **1.2: Feature Set Specification:** This deliverable was fully met and evidenced in Sections 2 and 3.

The key variables that were used in the predictive model were the network sites, pipelines and pressure and volumetric flow SCADA data, as described in Sections 2.1 and 2.3. The metrics that calculated within the model to identify UAG and metering errors will be summarised in the system architecture diagram, shown in Figure 2 in Section 3.

- **1.3: Data Quality and Coverage Report:** This deliverable was fully met and evidenced in Section 2.

Section 2 provided a summary of data completeness, gaps, and anomalies. A significant data gap is that not all pipelines have lengths and diameters recorded in the pipeline databook which necessitated the modelling assumptions discussed above.

3 Phase 2: Model Development

3.1 Model Overview

Phase 2 of the project covered the development of a forward predictive model to simulate the dynamics of the gas flow within the network and quantitatively compare the simulation outputs with the input SCADA data to identify potential metering errors. This model was developed with the knowledge that it should scale to the full national network in the future, a consideration that had a significant impact on the simulation architecture. The model also includes capability to calculate terms in the network energy balance to relate the effect of metering errors to sources of UAG. This section provides a technical overview of the developed model, modelling assumptions made, and the current limitations of the approach.

The simulation architecture for the developed model is illustrated in Figure 2. This architecture was informed by the following key requirements for this prototype model:

- Take historic SCADA data as an input to set the initial state and boundary conditions of the model.
- Simulate expected network behaviour to predict the expected measurements at meters.
- Allow comparison of simulated and SCADA data to identify, characterise and assess discrepancies.
- Introduce controlled degradations to simulate metering errors.
- Calculate terms in the network energy balance to quantify the relationship between metering errors and UAG.

The general flow of the model is as follows. Firstly, the network is defined from a set of pipelines, sites and gas properties. This information is used in conjunction with the pipeline data book and cleaned SCADA data (see Section 2 for more details) to instantiate a network model. From this, a simulation of the network is created and executed, providing baseline data where the presence of metering errors is initially unknown. The simulated data is compared to the SCADA data to identify potential sources of metering error. This is achievable as the physics-based simulation assumes that the SCADA data corresponds to the actual gas flow in the network. Thus, in a fully calibrated model, any discrepancy between the baseline and simulation data indicates a difference in the gas flow, and hence the presence of a metering error.

The architecture was chosen to enable the two primary use cases: modelling baseline network performance and modelling network performance with metering error, to be easily executed with minimal duplication of functionality. The architecture also enables a direct comparison between the baseline network model and the network model with synthetic metering errors to provide a forward prediction of the effect that metering errors have on the error. This direct correspondence allows the identification of appropriate metrics that could be used to identify candidate metering errors and quantifying the amount of UAG associated with a given metering error by calculating the difference in network energy balance terms calculated from the original SCADA data in the baseline model, with those from the modified SCADA data with the metering error applied.

3.2 Gas Flow Simulation

A critical part of this model is the simulation of the physics of gas flow through the network and implementing a model that is sufficiently fast to meet the needs of the project. The dynamics of gas transport within pipelines is fundamentally governed by the laws of conservation of three key properties, mass, momentum and energy. As the pressure and flow dynamics along the length of the pipeline are the quantities of interest and pipelines within the network are much longer than their diameters, the equations are considered in one dimension only. The network is also considered to be isothermal, an assumption that is commonly found within literature [8, 9] and one that NGT also use in UAG calculations. The governing equations for this reduce to the one dimensional isothermal Euler equations [9]:

$$\frac{\partial p}{\partial t} = -\frac{1}{a} \frac{\partial q}{\partial x}, \quad (1a)$$

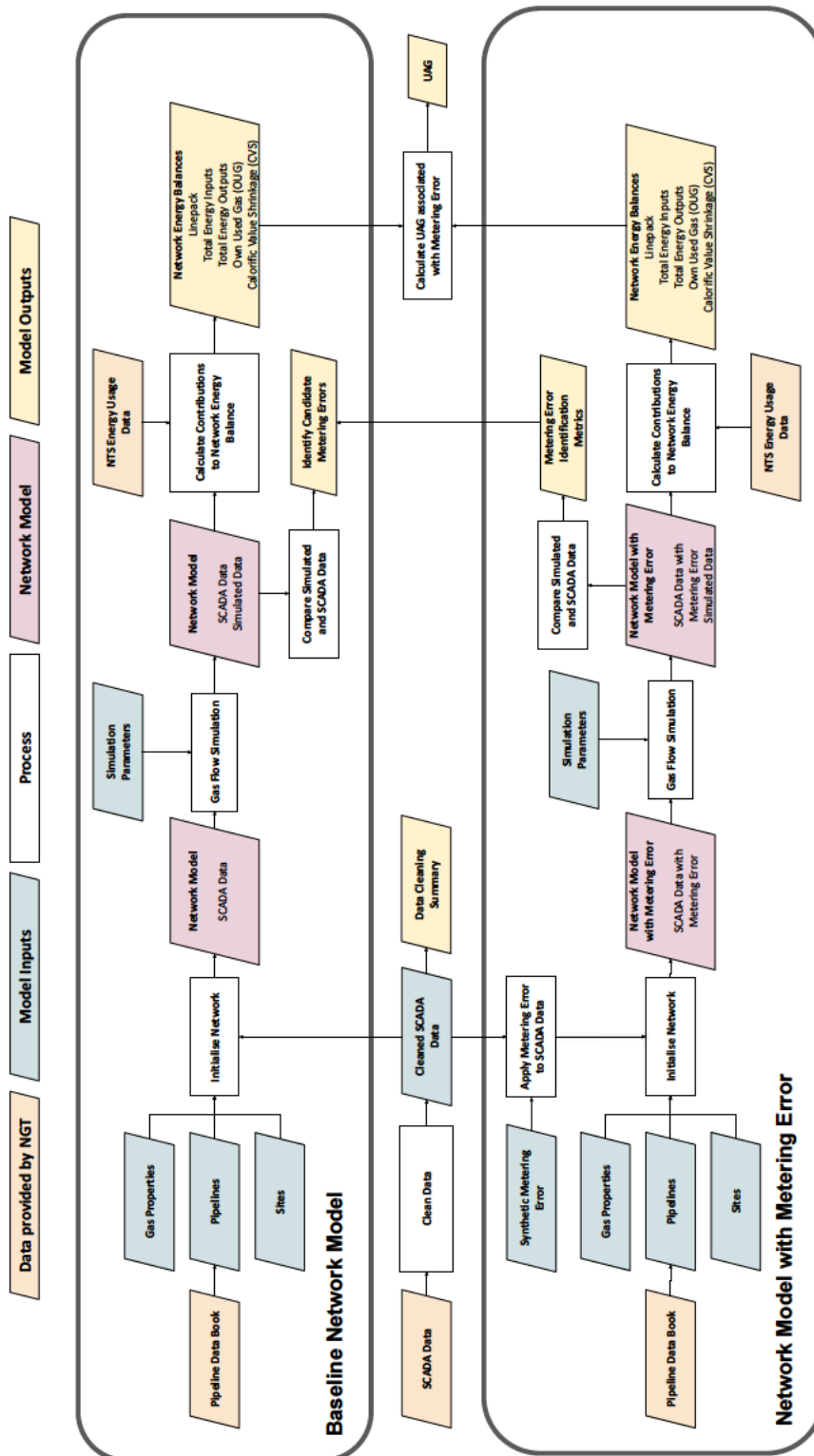


Figure 2: Simulation architecture diagram of the model to simulate the gas flow dynamics within a network and identify candidate metering errors.

$$\frac{1}{a} \frac{\partial q}{\partial t} = -\frac{\partial p}{\partial x} - \frac{\lambda(q)}{2da^2} \frac{q|q|}{\rho} - \frac{1}{a^2} \frac{\partial}{\partial x} \frac{q^2}{\rho} - g\rho \frac{\partial h}{\partial x}, \quad (1b)$$

$$p = \lambda_0 z_0(p) \rho, \quad (1c)$$

where ρ is the density of the gas, q is the mass flow, d and a are the diameter and cross-sectional area of the pipeline respectively, λ is the friction factor of the gas, g is the gravitational constant and p is the pressure of the gas. In this initial implementation, the compressibility factor z is set to the constant value of 1, as NGT noted a typical value of 0.9978 citeDataQuality. The gas constant is calculated using:

$$P = R_s T \rho, \quad (2a)$$

$$\lambda = R_s T, \quad (2b)$$

where R_s is the specific gas constant of methane (J/kgK) and T is the absolute temperature of the gas (K). The absolute temperature of the gas is taken to be $15^\circ C$ ($288K$), as outlined in NGT documentation [10] and the specific gas constant is taken to be $518.3J/kgK$. From Equation (2b), this results in a gas constant of 1.49×10^5 (J/kg). The friction factor relates to the pressure drop across a pipeline due to surface roughness of the pipeline and the Reynold's number of the flow. The friction factor is approximated using the Haaland equation [11] and set to the constant value of 0.008.

The last two right-hand side terms in Equation (1b) represent the impact of inertia and gravity within the pipeline respectively. The handling of each of these contributions is outlined below. The inertia term within the 1D Euler equations can be shown to be [8, 9]

$$\frac{1}{a^2} \frac{\partial}{\partial x} \frac{q^2}{\rho} \approx 10^{-3} \frac{\partial p}{\partial x}. \quad (3)$$

As the contribution of the inertia term is three orders of magnitude smaller than spatial variations in mass flow and pressure, the impact on the overarching flow and pressure behaviour in the network is assumed to be negligible, and thus the term is removed. This simplification is commonly performed within literature [9, 12]. The gravity term within the 1D Euler equation is also removed with the assumption of homogenous pipeline elevation. In reality, there is likely some variation in the pipeline elevation; however, this data is not readily available.

To solve the above equations along the length of the pipeline, the pipeline must be temporally and spatially discretised. The Finite Volume Method (FVM) is used to spatially discretise the domain into small cells called control volumes. The conservation equations, Equation (1a) and Equation (1b), are integrated over each cell in the domain resulting in a linear system of equations that are solved to calculate the value of pressure and flow in each control volume. The boundary conditions at each end of a pipeline are defined by a supply pressure into a pipeline and a demand flow exiting the pipeline. FVM is a commonly used approach in literature and is a discretisation that suits the nature of the problem well. In addition to this, FVM is generally less computationally expensive than other applicable methods [13] and due to the aim of scaling the model to large gas networks, computational expense is a key priority for the selected method.

3.2.1 Network Simulation Integration

The NTS is modelled as a directed graph where the sites are represented by nodes and the pipelines between them as edges. Nodes in the graph consist of three types, entry, exit and interior. Terminals and subterminals are entry sites where gas is added to the system through application of pressure boundary conditions on the corresponding nodes. NTS assets such as offtakes, powerstations and industry sites are exit sites where gas is allowed to escape through application of flow boundary conditions on the corresponding nodes. Junctions are interior sites where gas passes through the site and boundary conditions are not applied to the corresponding nodes. It should be noted that storage sites are treated as exit sites by calculating the net flow contribution from both injection and withdrawal of gas from the network.

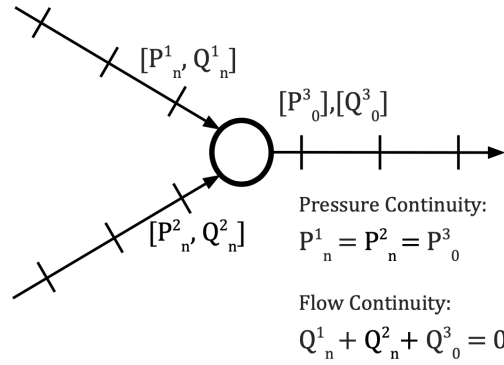


Figure 3: Schematic of the applied nodal constraints, pertaining to a node with two incoming pipelines and one outgoing pipeline, where P_j^i represents the pressure value in the j th control volume for pipeline i . Similarly, Q_j^i represents the flow value in the j th control volume for pipeline i .

The directionality of the pipelines in the model dictate the gas flow direction at the initialisation of the model. The directionality of the network is determined through a breadth-first search method to ensure that all non-entry sites possess a pipeline that feeds into the site and all non-exit sites possess a pipeline that carries flow away from the site. In reality, the flow direction within the NTS is not fixed and is often bidirectional along sections of the network. To handle this, the flow within the simulation is treated as a vectorised quantity, where the sign of the flow is the direction of the flow relative to the pipeline direction: positive flow indicates flow travelling in the same direction as the pipeline and negative indicates the opposite direction. The ability to model bidirectional flow means that storage sites may inject gas into the system as well as withdraw it.

To preserve continuity of pressure and flow throughout the network, additional algebraic constraints must be added to sites that possess connections to multiple pipelines, see Figure 3. To ensure continuity of pressure, when multiple pipelines meet at a given site the pressure values at the corresponding control volumes at the end of each pipeline are assumed to be equal to that site's pressure value. Figure 3, shows these constraints around a node in the graph where pipelines 1 and 2 enter the node and pipeline 3 exits. Therefore, the pressure values in the last control volumes of pipeline 1 and 2, P_n^1 and P_n^2 , must be equal to each other and the first control volume of pipeline 3, P_0^3 . To preserve the continuity of flow, the sum of all flows entering and exiting a site must be equal to zero. Again, for pipelines with a direction moving flow away from a site, this imposes flow constraints on the initial control volume and for pipelines that inject flow to a site, this imposes flow constraints within the final control volume. This is shown within Figure 3, where the flow values in the last control volumes of pipeline 1 and 2, Q_n^1 and Q_n^2 , along with the flow in the first control volume of pipeline 3, Q_0^3 , must sum to zero.

To assemble the network, the lengths and diameters of the pipelines are fed in from the pipeline data-book along with site connectivity data to assemble a representative graph of the subsection of the network. Each edge is then discretised according to a user defined spatial resolution. The discretised edges are then assembled in to a nonlinear system of equations. The Portable Extensible Toolkit for Scientific Computation (PETSc) is used to solve this system of equations [14]. PETSc was chosen for speed and scalability given its C++ implementation and parallelisable nature in the aim of ensuring a simulation tool that could be implemented on a nationwide scale. The configuration of the matrices to be solved is dependent on the timestepping scheme implemented. To reduce implementation complexity, an explicit timestepping scheme was chosen. The stability of simulation is dependent on the relationships between the timestep, spatial discretisation and velocity of the gas in the model. The choice of timestep and spatial step used in the simulation is determined through the results of a convergence study found in Appendix A, and is primarily driven by the stability constraints enforced by the explicit scheme.

3.2.2 Extension to Gas Networks

Several edge cases must be considered to enable stable and representative simulation results. Within the network, several sites do not possess data for their connections in the SCADA pipeline databook. This is particularly prevalent in sections of the network with several sites closely clustered together, e.g. around the terminal at [REDACTED]. Modelling assumptions are made for the cases where pipeline data is not readily available, as previously discussed in Section 2.

To aid in the scalability of the simulation, where possible, a selection of non-SCADA sites are removed from the network to reduce the total number of nodes within the network graph and the number of nodal constraints enforced in the simulation. As telemetry data is not readily available for these sites, directly modelling them provides little benefit. When a non-SCADA site is removed and the pipelines flowing into and out of the site are combined into an aggregated pipeline that spans between SCADA sites. In order for a site to be eligible for removal, the site must not split flow within the network and any connected pipelines must have the same diameter so that the pipelines can be aggregated into a simplified combined pipeline.

In addition to the above changes in the network topology, when extending the simulation to incomplete networks the key assumption of conserved mass within the network is violated. To combat this, a synthetic site is added to the network to account for the net flow difference in the subnetwork. This synthetic site represents the flow dynamics external to the subnetwork. The synthetic site is applied as an exit site attached to the network via a short pipeline and possesses the time series data corresponding to the net flow difference within the subnetwork. This approach ensures the conservation of mass within the network and rectifies discrepancies in the flow data. The location of the additional flow site for a particular test case is chosen through NGT domain knowledge of the larger network surrounding the subsection. Future work could span to applying multiple additional sites with a proportion of the net flow difference to more accurately capture the additional flow out of the subsection of the network,

Once a representative network is assembled and entry and exit sites for the subsection of the network are identified, the boundary condition data must be applied. Meter data provided within SCADA are used as boundary conditions. For exit sites, the exit flow boundary condition is defined as the flow value at an instantaneous volumetric flow meter. In the case of the SCADA data possessing multiple instantaneous flow meters, the first is chosen by convention. For entry sites, a pressure meter from the sites SCADA data is assigned as boundary condition data. At all remaining sites, initial pressure and flow values are also taken from the SCADA meter data at the defined simulation start time. The control volumes within the pipelines also require initialisation. To do so, linear interpolation along the length of the pipeline using the pressure and flow data at the nodes connected to the pipeline is performed. In the case where sites do not possess flow or pressure data for initial conditions, a distance based weighted average of neighbouring sites is applied to estimate an initial condition. Where a site and all of its neighbours do not possess initial condition data, a network-wide average is performed to provide a sensible starting value. As a result of this approximation, the simulation requires an initial warm-up time to allow for the pressure and flow conditions within the simulated network to become representative of those in the network.

3.2.3 Statistical Analysis of SCADA and Simulated Data for Metering Error Identification

The primary outputs of the gas flow simulation are simulated pressure and flow data at each of the sites in the section of the network being considered. Since the simulated data represents the expected physical behaviour of the gas flow dynamics, discrepancies between these results and the corresponding SCADA data can be used to identify potential metering errors. Differences between the SCADA and simulated data at all sites in the network are quantified by calculating the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) between the two time series. Sites with larger values for these metrics indicate the presence of potential metering errors. Other comparison metrics were considered within this project; however, following application on the test cases, they provided minimal additional insight.

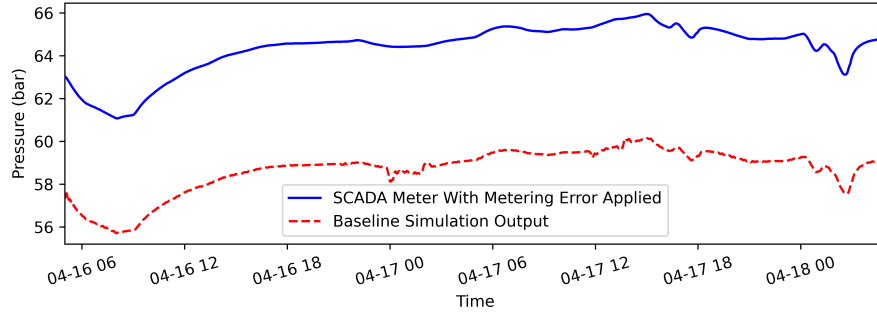


Figure 4: Example of synthetic multiplier error applied to SCADA data (blue).

3.3 Synthetic Metering Errors

Metering errors lead to differences between true and measured values and are the cause of much of the uncertainty present in the SCADA data. Synthetic metering errors are used to investigate the effect of controlled and predictable metering errors on the network's behaviour. Specifically, the aim is to examine the impact of metering errors on comparison metrics between SCADA and simulated data, and hence to identify metrics that are likely to indicate the presence of a metering error. Additionally, synthetic metering errors are used to simulate perturbed flow at a specific site by considering the erroneous data to be representative of the genuine flow and then observing its propagation through the network.

To ensure physicality, the types of metering errors defined are drawn from physical metering errors observed and documented in large metering error report documents provided by NGT [15, 16, 17, 18, 19]. In this project, the following three types of synthetic metering error are considered to demonstrate the functionality. More realistic types of metering errors, including out of operating range errors, could be similarly integrated into the model in future work.

Multiplier Error: Multiplier error is the amplification of a measurement by a constant factor over time. Synthetic multiplier errors use parameterised multiplier values to model the effect of constant amplification on measured values. The base equation for the application of synthetic multiplier error is:

$$x(t) = a \cdot x_0(t),$$

where $x(t)$ is the erroneous reading at time t , a is some constant multiplier and $x_0(t)$ is the original data point. An example of synthetic multiplier error applied to SCADA data is shown on Figure 4.

Calibration Error: Calibration error is an offset between genuine and measured values, linearly increasing overtime to represent metering drift. Synthetic calibration errors are defined by an initial offset value and a linear gradient describing the increase of the offset. The base equation for the application of synthetic calibration error is:

$$x(t) = x_0(t) + a + b(t - t_0),$$

where $x(t)$ is the erroneous reading at time t , $x_0(t)$ is the original data point, a is the initial offset, b is the gradient, and t_0 is the start time of the metering error. An example of synthetic calibration error applied to SCADA data is shown on Figure 5.

Noise: Noise is unwanted, unpredictable disturbance that can obscure and distort measurements. The effect of noise on a meter is modelled by superimposing Gaussian random fluctuations onto the original data set. Synthetic noise is defined by specifying the mean and standard deviation of the underlying Gaussian probability distribution, which is then used to weight the randomised selection of the amplitude of noise at each point. The base equation for the application of noise is:

$$x(t) = x_0(t) + \varepsilon_t,$$

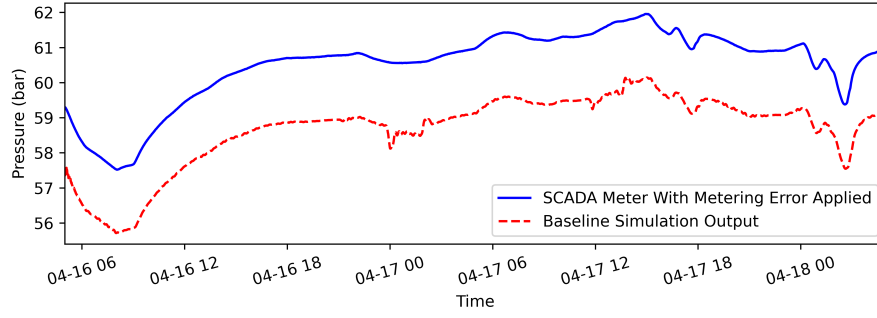


Figure 5: Example of synthetic calibration error applied to SCADA data (blue).

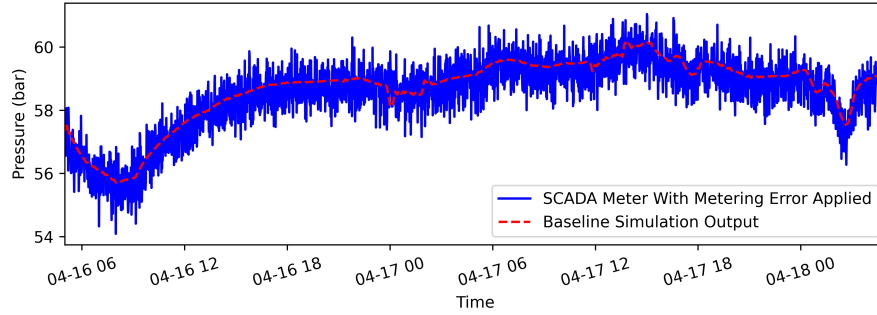


Figure 6: Example of synthetic noise applied to SCADA data (blue).

where $x(t)$ is the erroneous reading at time t , $x_0(t)$ is the original data point, and the ε_t are independent Gaussian random variables. An example of synthetic noise applied to SCADA data is shown on Figure 6.

3.4 UAG Calculation

3.4.1 Introduction

Unaccounted For Gas is the quantity of gas that remains following the balancing of all inputs and outputs to the NTS. For a closed network on gas day t , this can be expressed as:

$$E_{bal,t} = (E_{in,t} - E_{out,t}) - \Delta Lp_t - (OUG_t + CVS_t), \quad (4)$$

where $E_{bal,t}$ is the energy balance of the network, $E_{in,t}$ is the total energy of gas entering the network, $E_{out,t}$ is the total energy of gas exiting the network, ΔLp_t is the delta linepack (difference between the opening and closing linepack energy), OUG_t is the Own Use Gas (OUG) from compressor fuel usage and CVS_t is the Calorific Value Shrinkage (CVS). To calculate each of the components in Equation (4), the pressure and flow data from SCADA and simulation outputs were used in combination with the NTS energy data usage [4, 6, 5, 7].

UAG is typically calculated across the full NTS, which can be considered a closed system. For an ideal closed system, the energy entering the NTS is equal to the energy exiting the NTS and no metering errors are present. This leads to an energy balance with $E_{bal,t} = 0$, and no UAG. However, in the presence of metering errors across the network, the calculated energy balance is non-zero, and this difference corresponds to the amount of UAG.

In this project, the network energy balance $E_{bal,t}$ can be calculated using the actual SCADA data, SCADA data with synthetic metering errors applied or simulated data. A difference in $E_{bal,t}$ calculated using the actual SCADA data and calculated with the associated simulation data indicates the presence of UAG. A difference in $E_{bal,t}$ calculated using SCADA data with and without synthetic metering errors enables the amount of UAG associated with the applied metering error to be quantified.

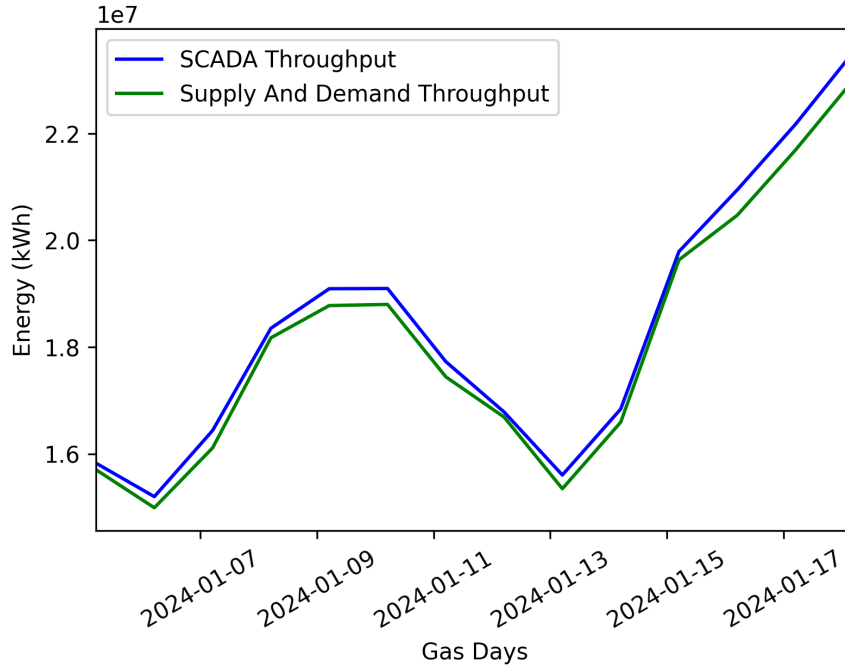


Figure 7: Total Energy Throughput calculated from SCADA telemetry data and Supply And Demand data between 05/01/2024 and 18/01/2024 for a network section consisting of the sites: [REDACTED]

3.4.2 Energy Difference And Throughput

The energy difference ΔE is defined as the difference between the total energy entering the network and the total energy exiting the network:

$$\Delta E_t = \sum_{\text{entry sites}} E_{in,t} - \sum_{\text{exit sites}} E_{out,t}. \quad (5)$$

The energy throughput $E_{throughput}$ is defined as the total energy flow rate through the network:

$$E_{throughput,t} = \frac{1}{2} \left(\sum_{\text{entry sites}} E_{in,t} + \sum_{\text{exit sites}} E_{out,t} \right). \quad (6)$$

To calculate the total energy input, $E_{in,t}$, and total energy output, $E_{out,t}$, for an arbitrary network on gas day t , from either volumetric flow or energy flow, the following steps are undertaken. For each entry and exit site in the network, the mean volumetric or energy flow is calculated to get the daily volumetric or energy flow rate. Next, the volumetric flow or energy flow rates are summated for entry and exit sites to get the total entry / exit volumetric / energy flow for gas day t . To convert the total entry and exit flows to kWh, the conversions detailed at [20] are followed.

The quantities E_{in} and E_{out} can also be calculated directly from supply and demand data [4]. Comparing the throughput calculated using the supply and demand data with that from the telemetry data for a subsection of the NTS across two weeks, shows a good agreement which, demonstrates the validity of the calculation of E_{in} and E_{out} from telemetry data (Figure 7).

3.4.3 Linepack And Delta Linepack

Linepack is defined as the amount of gas within the NTS at any one time. Linepack is typically calculated as a volume m^3 before it is converted to linepack energy to be used in the UAG calculation. Linepack

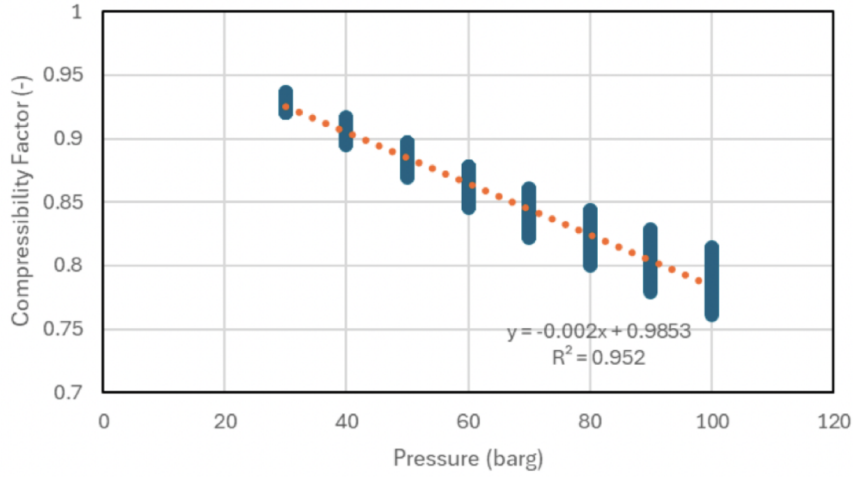


Figure 8: Provided relationship between pressure (barg) and compressibility factor [10].

is calculated for each pipeline in the NTS, which is then summated to get total linepack. Linepack is calculated using the following methodology [21]. On gas day t , the linepack volume (m^3) of the pipeline can be expressed as:

$$Lp_t = \frac{P_{avg,t}}{P_{std}} \cdot \frac{Z_{std}}{Z_{m,t}} \cdot \frac{T_{std}}{T_{m,t}} \cdot V_{pipe}, \quad (7)$$

where $P_{avg,t}$ is the average pressure of the pipeline, P_{std} is the standard NTS pressure, Z_{std} is the standard compressibility factor of the NTS, $Z_{m,t}$ is the measured compressibility factor of the pipeline, T_{std} is the standard NTS temperature and $T_{m,t}$ is the measured temperature of the pipeline and V_{pipe} is the volume of the pipeline. The standard values for Z_{std} and temperature T_{std} were provided by NGT [21] and are 0.9978 and 288K (15 degrees C) respectively. The standard pressure, P_{std} , is assumed to be 50barg [10]. In this implementation, T_m is assumed to be constant across the network, and is set to be $T_{std} = 288K$.

The average pressure in the pipeline P_{avg} is calculated according to [21]:

$$P_{avg} = \frac{2}{3} \cdot \frac{P_{in}^2 + P_{in}P_{out} + P_{out}^2}{P_{in} + P_{out}}, \quad (8)$$

where P_{in} and P_{out} are obtained from the pressure telemetry data at the start of gas t at both ends of the pipeline. If pressure telemetry data is missing, it is set by default to be P_{std} . Having calculated P_{avg} , the measured compressibility factor, Z_m , can then be computed using the pressure telemetry data from the linear relationship:

$$Z_m = -0.002P_{avg} + 0.9853, \quad (9)$$

shown in Figure 8 [10].

Delta linepack is calculated from the difference between the opening and closing linepack energy values for each gas day t . The opening linepack is calculated using the telemetry data at the time closest to the start of the gas day t . Closing linepack is defined as the linepack calculated from telemetry data at the time closest to the start of gas day $t + 1$.

3.4.4 Compressor Fuel Usage

Compressor Fuel Usage, or OUG, is the energy required to run the compressors used to push gas at high pressure around the NTS. The Compressor Fuel Usage (kWh) per gas day for each compressor in the network is extracted directly from [6] and summed to get the total Compressor Fuel Usage per gas day for the network.

3.4.5 Calorific Value Shrinkage

The Calorific Value Shrinkage (CVS) is defined as the gas that cannot be billed due to the application of Gas Regulations [22]. The CVS for a network is calculated on each gas day by taking the daily CVS per LDZ from [5]. As this value has contributions from all offtakes in each LDZ, to calculate the CVS for a network subsection, a proportion of this value is taken for a given network, making the assumption that each offtake contributes equally to the total daily CVS.

3.5 Key Deliverables

The Statement of Work [1] identified the following deliverables for Phase 2: Model Development. This section summarises the extent to which these deliverables were achieved in this project.

- **2.1: Working Prototype of the Model:** This deliverable was fully met and evidenced in Section 4. Section 4.6 will provide further information after presenting the results from executing the model on test cases.
- **2.2: Technical Documentation:** This deliverable was fully met and evidenced in Sections 3 and 4. This section provided a detailed technical explanation of the model architecture and assumptions. Section 4.6 will detail the limitations of the model in the context of the obtained result.
- **2.3: Initial Performance Benchmarks:** This deliverable was partially met and will be evidenced in Section 4. Section 4.6 will provide further information after presenting the results from executing the model on test cases.

Parameter	Value	Units
dt	0.2	s
dx	0.4	km
Individual Pipeline Length	11.5	km
Pipeline Diameter	600	mm
Friction Factor	0.008	(-)

Table 2: Simulation parameters used for the suite of synthetic test cases.

4 Phase 3: Model Testing and Validation

The model was applied to several test cases to demonstrate the functionality of the model and assess its performance. In Section 4.1, a set of verification test cases were performed to verify the modelling approach. In Section 4.2.5, the results from applying the model to a small section of the network between [REDACTED] are presented. The purpose of this test case was to demonstrate that the model can simulate the expected network behaviour under ideal conditions given input SCADA data. Synthetic metering and transmitter errors were applied to this baseline model to simulate the effect of different types of errors, including metering drift and noise, and demonstrate the capability of this model to identify these metering errors and calculate the UAG associated with these errors. In Section 4.3.4, the results from applying the model to two historical metering errors at [REDACTED] are presented to assess the performance of the model with realistic metering errors. In Section 4.4.4, the model was applied to the North-East Section of the NTS to demonstrate scalability to larger networks. Section 4.5 concludes this section by providing a discussion of the key results and conclusions that can be drawn from the results of these test cases and the limitations of the current prototype of model.

4.1 Verification Test Cases

Before integrating the model with SCADA data and network topologies, a set of small test cases were performed to validate the flow and pressure behaviour in networks with representative flows and network features. For the following suite of test cases, the simulation parameters are outlined in Table 2.

The first test case undertaken was the simplest network that could be simulated: a straight pipeline consisting of three sites connected by two pipelines. The configuration for the first pipeline is shown in Figure 9a. The empirical steady state pressure drop for this network is described by the Darcy–Weisbach equation [11]:

$$\Delta P = f_d \cdot \frac{\rho v^2}{2D_h} L, \quad (10)$$

where ΔP represents the pressure drop along the length of the pipeline, f_d represents the Darcy friction factor (an empirical factor input to the simulation and the equation), D_h is the hydraulic diameter of the pipeline, v is the velocity of the flow and L is the length of the pipeline. For the configuration shown in Figure 9a, the Darcy–Weisbach outputs an expected pressure drop of 1.206 bar along the total length of the pipeline. Figure 9b shows the pressure at site C through time. After 200 seconds, the system reaches steady state and the pressure at site C is found to be within 1% of the pressure predicted by the Darcy–Weisbach equation.

The next test case considered was performed to assess flow and pressure behaviour in a network that possessed branching junctions and internal pipelines forming cycles within the network. The configuration of the second test case is shown in Figure 10. This feature is common within the NTS and can be seen for example between Easington and Paull Multi-Junction. Figure 11 shows the resultant flows at site A, site B and site C. As can be seen in Figure 11a, the resultant flow at site A reaches the prescribed flow rate at site E of 2,000,000mcm/d. Figures 11b and 11c indicate that the flow is split between sites B and C evenly which is expected for a symmetric network. This result shows that the simulation can effectively handle branching in the network.

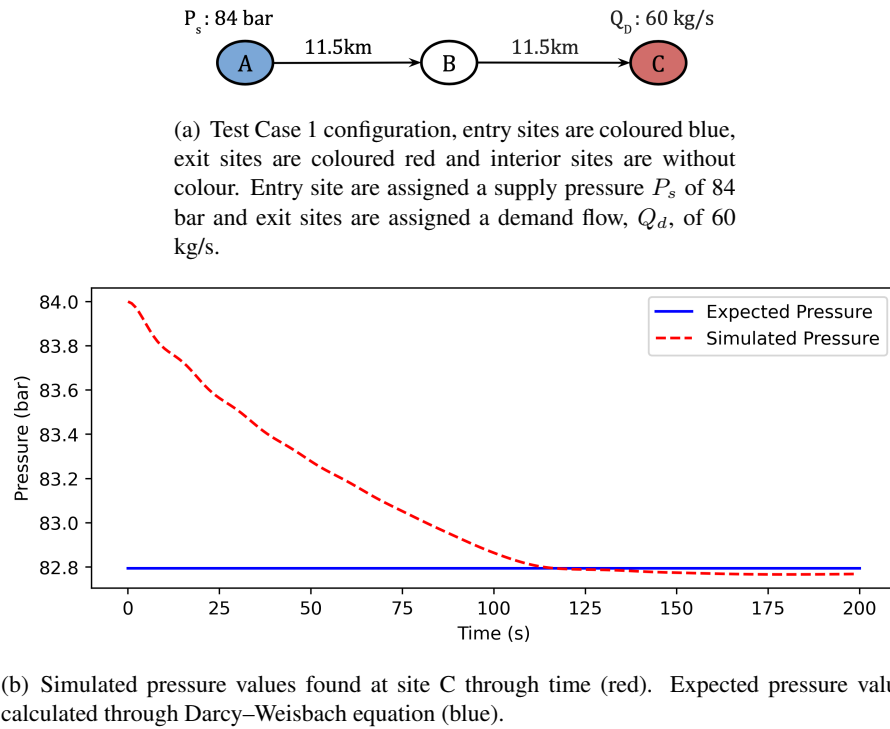


Figure 9: Test case 1 network configuration (a) and simulated pressure results (b).

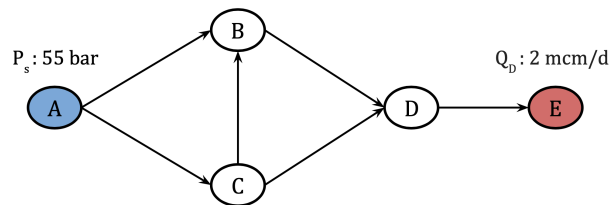
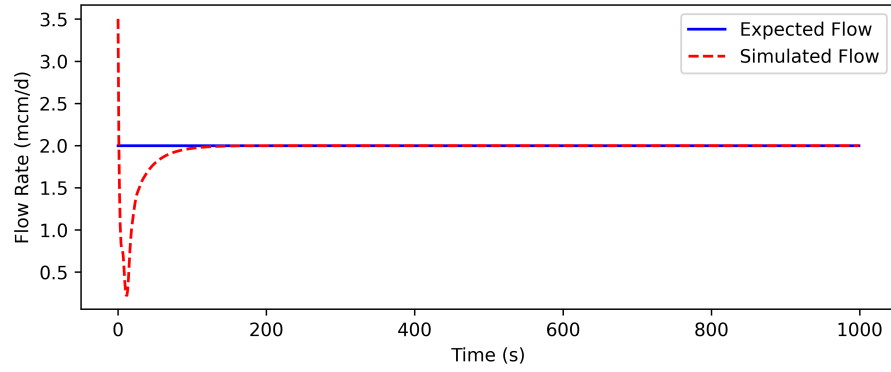
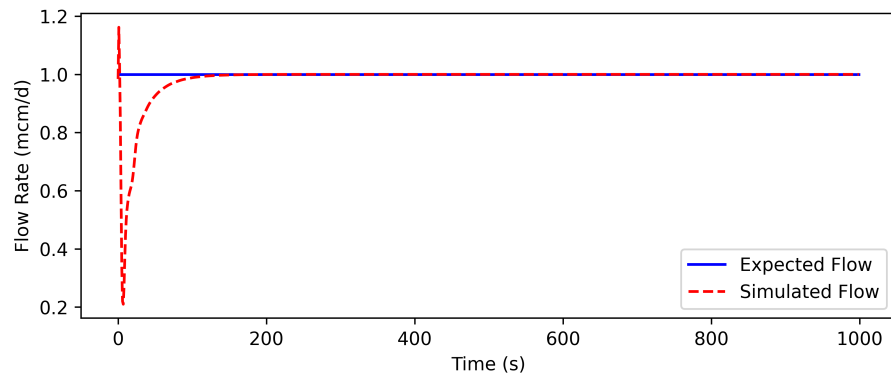


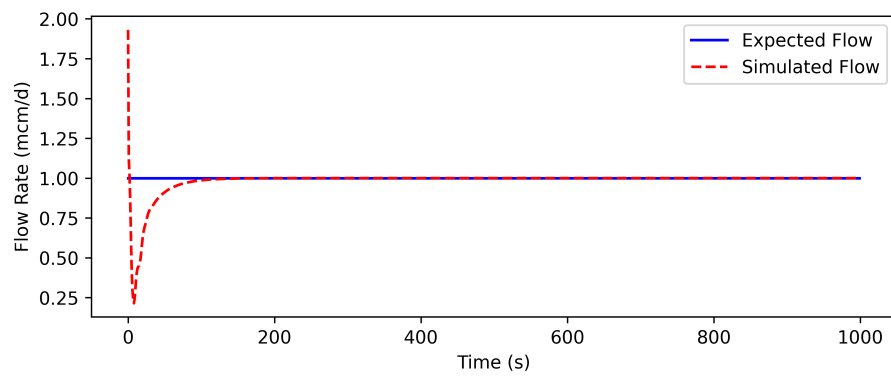
Figure 10: Test case 2 configuration, entry sites are coloured blue, exit sites are coloured red and interior sites are without colour. Pipeline BC allows for gas transfer between each side of the network in either direction.



(a) Simulated pressure meter at site A.



(b) Simulated pressure meter at site B.



(c) Simulated pressure meter at site C.

Figure 11: Test Case 2 results. The resultant flows at site B and site C sum to the entry flow found at site A, indicating the conservation of momentum in this test case.



Figure 12: Section of network used in the [REDACTED] Test Case.

4.2 [REDACTED]

The [REDACTED] Test Case serves to demonstrate the core functionality of the model on a small section of the NTS. In particular, this test case demonstrates the capability of the model to take in historical SCADA data as an input and simulate the expected gas flow on a subsection of the network under ideal conditions (Sections 4.2.1 and 4.2.2). Synthetic metering errors were applied to the input SCADA data in this test case to simulate the effect that the presence or absence of these errors have on the expected gas flow and demonstrate how the model allows identification of metering errors by comparing SCADA data with the simulation results (Section 4.2.4).

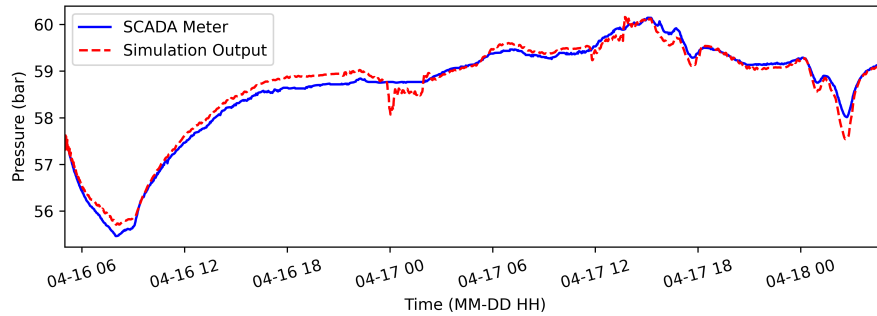
4.2.1 Test Case Definition

This test case considers a small section of the NTS comprising [REDACTED], as shown in Figure 12. The gas flow simulation was run on this section for two gas days between the 16th and 18th of March 2024 as SCADA data was available for every site in this network section during this time period.

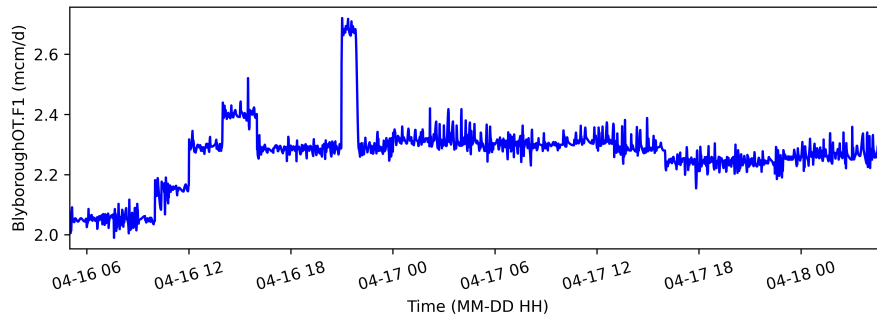
The following modifications were required to initialise the network for this test case. As the model assumes that gas flow into the system is equal to gas flow out, the flow difference in the system was calculated from SCADA data and this extra flow was injected into the network via an external flow site, as described in Section 3.2. This external flow site was connected to [REDACTED] to represent the typical direction of gas flow in this region of the network, flowing southwards from [REDACTED]. Within the model of the network used in this test case, [REDACTED] were assumed to be connected by a short pipeline as information about this pipeline was not contained within the data pipeline book (cf. Section refsec:Phase1). From an initial analysis of the SCADA data, it was recognised that typical pressures at [REDACTED] were approximately 1.5 bar higher than the pressures at [REDACTED] and other sites in this test case. To account for this pressure difference with the assumed pipeline properties, the prescribed pressure at [REDACTED] was calibrated by applying a corrective metering error of -1.5 bar.

Parameter	Value	Units
dt	0.2	s
dx	1	km
Default Pipeline Length	0.1	km
Default Pipeline Diameter	600	mm
Friction Factor	0.008	(-)
Calibration at	-1.5	bar

Table 3: Simulation parameters used for the [REDACTED] Test Case.



(a) SCADA vs Simulated Pressure.



(b) Prescribed SCADA Flow.

Figure 13: Simulation outputs at [REDACTED]

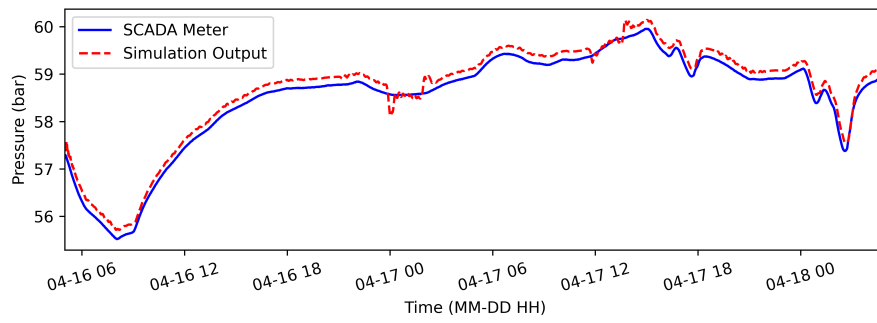
4.2.2 Baseline Simulation Results

The gas flow simulation was applied to the test case using the parameters in Table 3, with dt , dx and friction factor decided based on a convergence study described in Appendix A. Simulating the two days of the gas flow dynamics took approximately 1.9 minutes running on 16 cores.

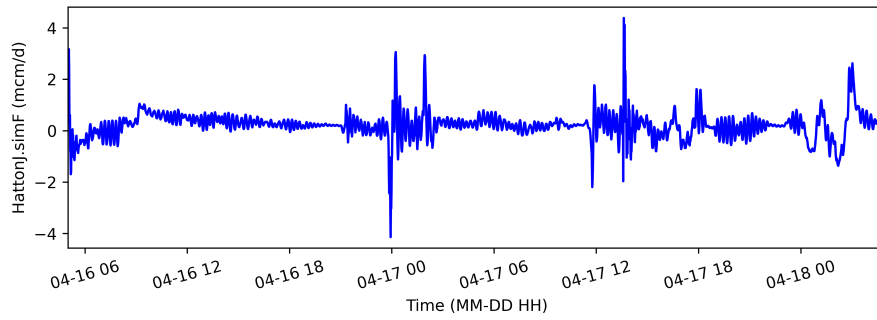
Figures 13, 14 and 15 show the simulated pressure and flow data at [REDACTED] respectively. The simulated pressure data shows good qualitative agreement with the SCADA data as can be seen by the simulated pressure being able to capture the peaks and troughs in the SCADA data. Simulated flow at [REDACTED] oscillates around the value in the SCADA data (Figure 15). The fluctuations seen are most likely due to the modelling assumption surrounding the addition of additional flow into the network.

4.2.3 Baseline Energy Balance Results

Constituents in the network energy balance were calculated for the first gas day of this test case (16/04/2024), excluding any contributions from the additional flow site, with the results tabulated in Table 4. From these results, the following conclusions can be drawn: across the board, it can be seen that there is a good match

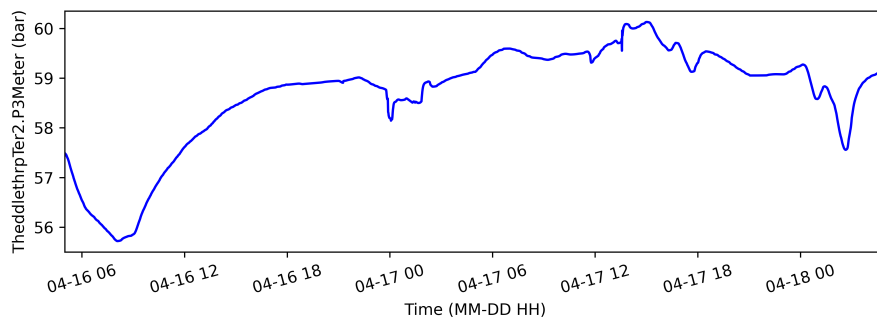


(a) SCADA vs Simulated Pressure.

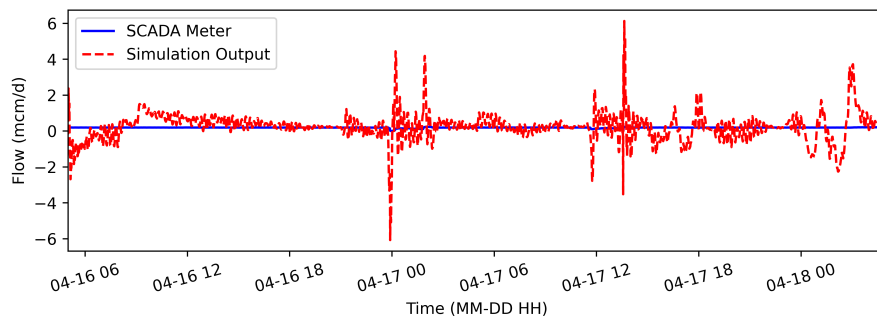


(b) Simulated Flow.

Figure 14: Simulation Outputs at [REDACTED].



(a) Prescribed SCADA Pressure.



(b) SCADA vs Simulated Flow.

Figure 15: Simulation outputs at [REDACTED].

	E_{bal}	Energy Inputs - Outputs (ΔE)	Delta Linepack (ΔLp)	CVS	OUG
SCADA	-24,495,367	-23,159,907	1,339,311	-3,851	0.0
Simulated	-23,611,612	-22,255,671	1,359,792	-3,851	0.0
Absolute Difference	883,775	904,236	20,481	0.0	0.0

Table 4: Energy Balance constituents, measured in kWh, for 16/04/2024 calculated from SCADA and simulated data. The absolute difference is the difference between constituents calculated using SCADA and simulated data, and indicates the presence of UAG.

between E_{bal} and the subsequent constituents for simulated and SCADA data, with the network energy balance, E_{bal} , the energy difference, ΔE , and delta linepack, ΔLp , possessing a small relative percentage difference of 3.67%, 3.95% and 1.51% respectively. The energy difference, ΔE , is negative and of the order of 2×10^7 kWh, indicating that the amount of energy exiting the system is significantly larger than the amount of energy entering the system. This results in the energy difference ΔE being the dominant constituent contributing to E_{bal} . The Calorific Value Shrinkage, CVS, is the same for both simulated and SCADA results due to the fact that it is calculated independently of telemetry data, using data from [5]. Compressor Fuel Usage, OUG, is zero as this network does not contain any compressor sites. The difference between the value of the network energy balance E_{bal} from simulated and SCADA data is 883,775 kWh. In a fully calibrated model, the non-zero value for this difference would indicate the presence of UAG in this test case.

4.2.4 Application of Synthetic Metering and Transmitter Errors

Synthetic metering and transmitter errors were applied to the baseline simulation of gas flow across this test case, and the resulting simulated network data compared to the baseline case. The application of synthetic metering and transmitters errors is separated into two distinct parts—the application of the three defined synthetic transmitter error types (see Section 3.3) at [REDACTED] pressure transmitter, and a synthetic noise metering error at [REDACTED] flow meter. In the first case, the aim is to examine whether there exists a measurable impact of the transmitter error when comparing SCADA and simulated pressure data. In the second case, the aim is to examine how perturbations to the gas flow propagate through the network. In both cases, identifying discrepancies between the simulated and SCADA data, both at the site at which the metering or transmitter error is applied, and at other sites within the network, can be used to demonstrate the ability of the model to identify potential metering and transmitter errors. The synthetic metering and transmitter errors applied have higher magnitudes than typical physical errors encountered in meter error reports [15, 16, 17, 18, 19] in order to facilitate the identification of key trends in the gas flow model behaviour.

Pressure Transmitter Errors

Synthetic transmitter errors are applied at [REDACTED] pressure transmitter over simulation period of two gas days between 16/04/2024 and 18/04/2024. The transmitter errors are defined as follows:

1. Error 1: Calibration error with initial offset of 2 bar and a linear gradient of 0.000017 bar/year (to reflect a 3% increase over the test case period).
2. Error 2: Multiplier error with a multiplication factor of 1.1.
3. Error 3: Noise with a mean of 0 bar and a standard deviation of 0.5 bar.

The transmitter errors are applied by modifying the SCADA pressure data at [REDACTED]. The resulting pressure measurements for each error are respectively shown in blue in Figures 16, 17 and 18.

The effect of a transmitter error on the agreement between SCADA and simulated data can be seen in the difference between the two datasets over time. Figure 19 shows the difference between the SCADA pressure

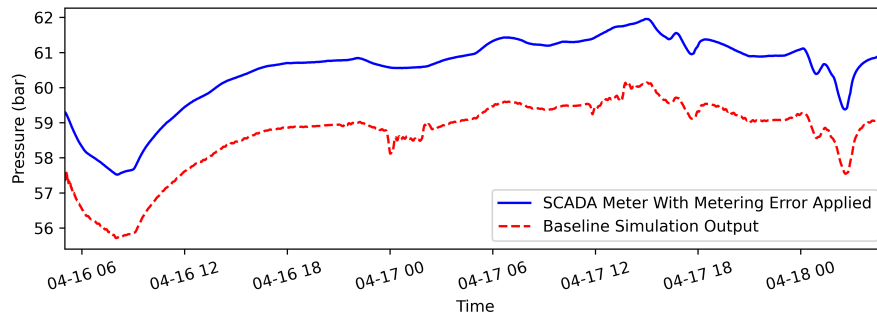


Figure 16: Synthetic calibration error with initial offset of 2 bar and gradient of 0.000017 bar/year, applied at [REDACTED] pressure transmitter (blue) compared to baseline simulation output (red).

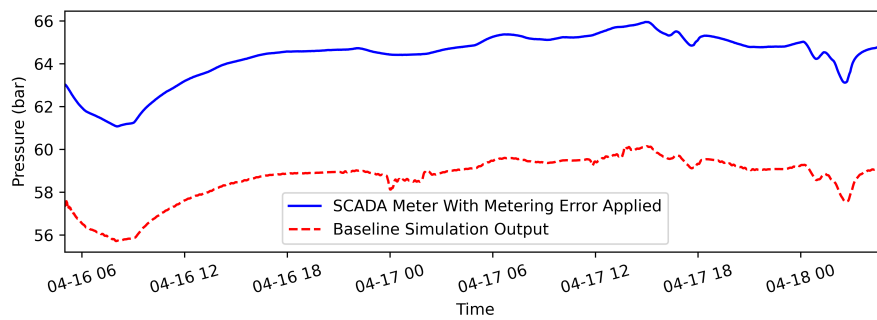


Figure 17: Synthetic multiplier error, with multiplication factor of 1.1, applied at [REDACTED] pressure transmitter (blue) compared to baseline simulation output (red).

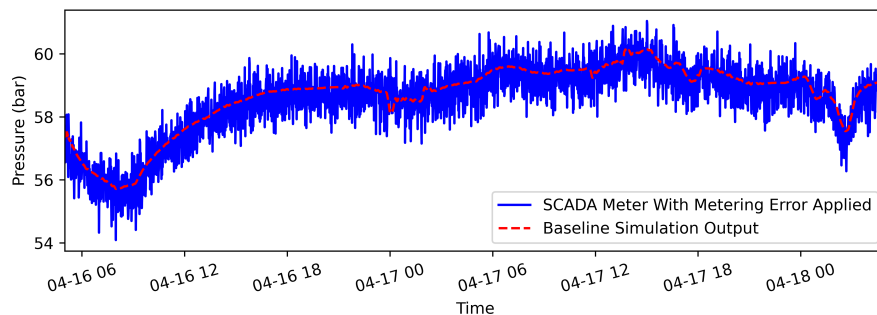
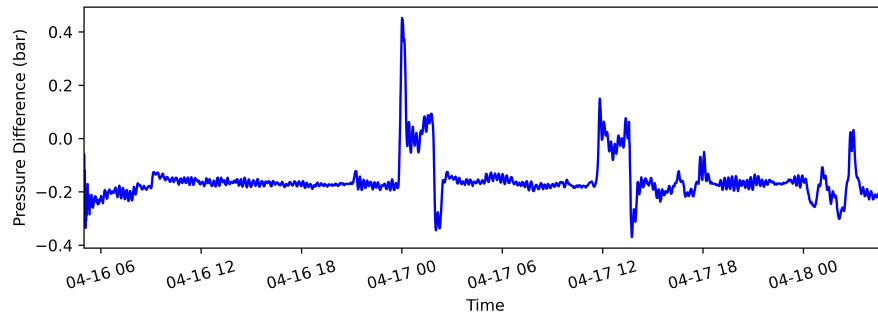


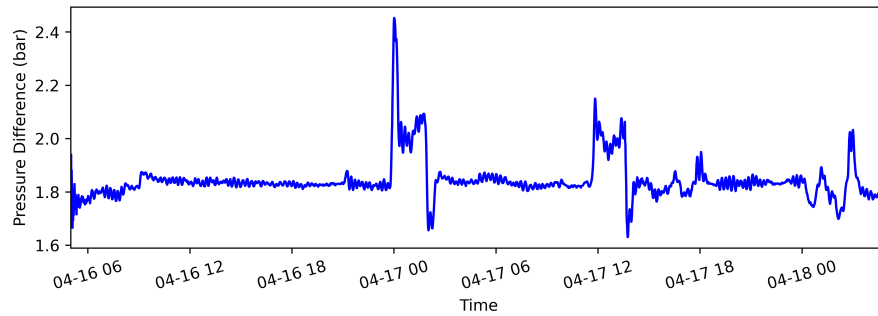
Figure 18: Synthetic noise with mean of 0 bar and standard deviation of 0.5 bar, applied at [REDACTED] pressure transmitter (blue) compared to baseline simulation output (red).

data affected by the transmitter error compared to the unaffected simulated pressure data (Figure 19a). Comparing Figures 19(a) and 19(c), the multiplier error preserves the overall shape of the signal but significantly amplifies the magnitude of the discrepancy by a factor of around fifteen. A similar effect is observed for the calibration error (Figure 19(b)), although the increase in magnitude is smaller and grows gradually over time. Finally, noise introduces random variance compared to baseline difference case, resulting in the original trend becoming unrecoverable. These results are consistent with the type of transmitter error applied and demonstrate that the effect of transmitter errors can be qualitatively seen in comparing the simulation results to the SCADA data with and without the transmitter error.

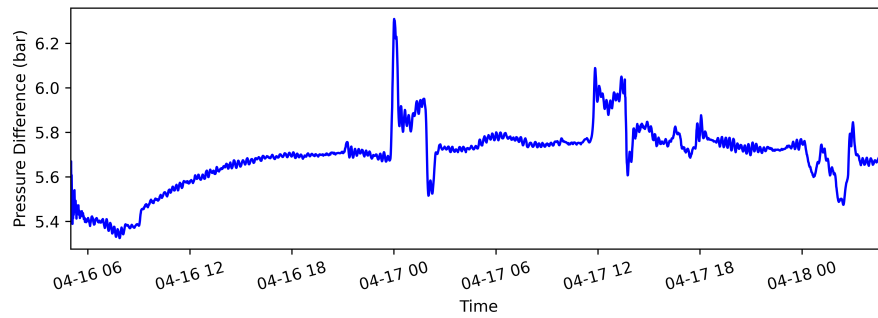
In order to quantify the impact of the applied transmitter error, the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) between SCADA and simulated data are calculated. Figure 20 shows the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) between these two datasets for the baseline case, described in Section 4.2.2, and for the erroneous flow simulation with transmitter error at [REDACTED], at each site in the test case. As seen in the figure, both RMSE and MAPE at [REDACTED] are significantly increased compared to the baseline case for each synthetic transmitter error test case. This indicates a discrepancy between simulated and SCADA values and hence a potential transmitter error. As the value of the comparison metrics at other sites remain similar in magnitude with and without the transmitter error applied, the increased value at [REDACTED] demonstrates the capability of the model to detect and localise the applied transmitter error.



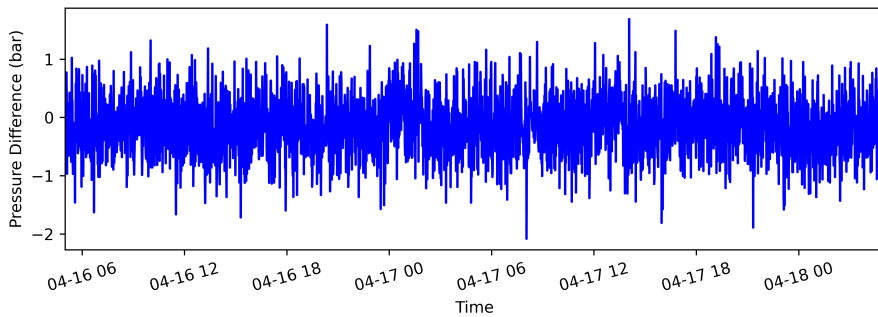
(a) Baseline Case (No error)



(b) Calibration Error

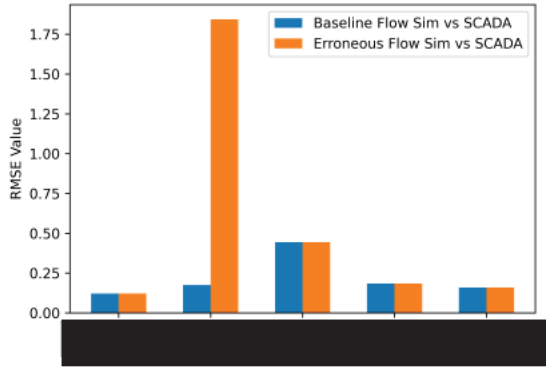


(c) Multiplier Error

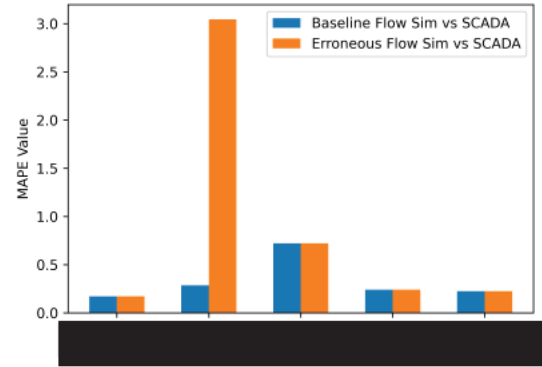


(d) Noise

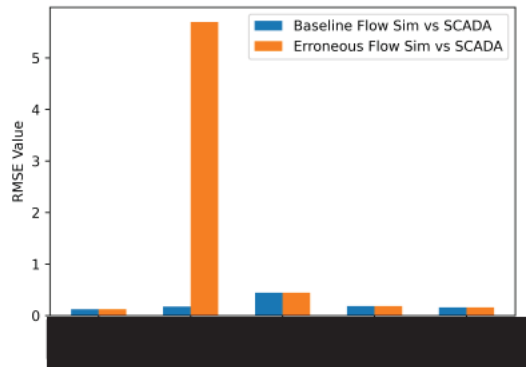
Figure 19: Difference between and SCADA and erroneous simulated pressure meter measurements at [REDACTED] over time for all synthetic transmitter errors.



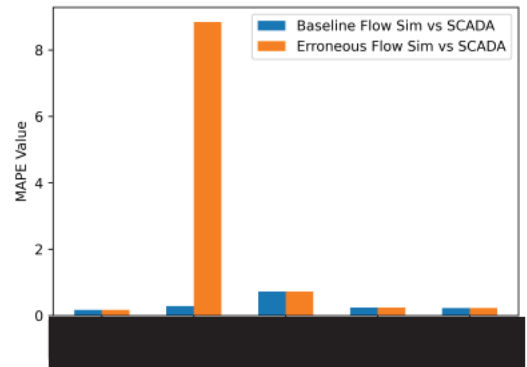
(a) RMSE with and without calibration error



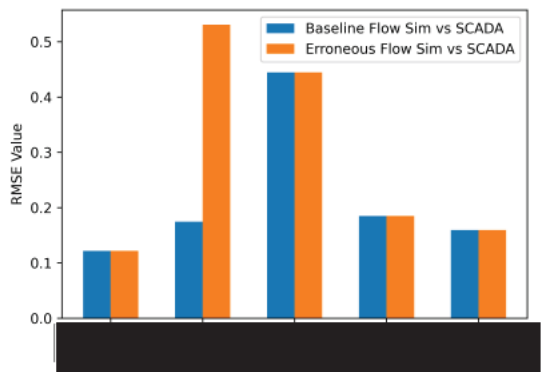
(b) MAPE with and without calibration error



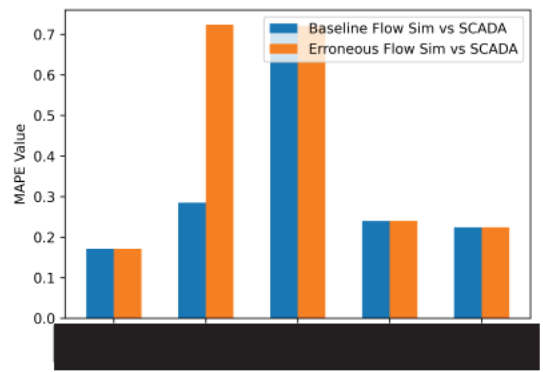
(c) RMSE with and without multiplier error



(d) MAPE with and without multiplier error



(e) RMSE with and without noise



(f) MAPE with and without noise

Figure 20: Baseline against SCADA pressure (blue) and erroneous against SCADA (orange) pressure root mean squared error (RMSE) (*left*) and mean absolute percentage error (MAPE) (*right*) for all synthetic metering errors.

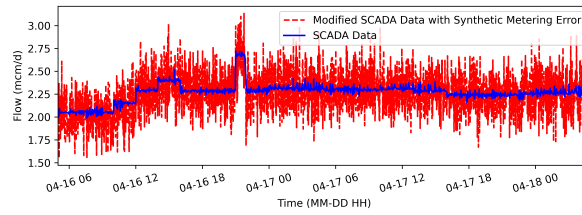


Figure 21: Original SCADA data (*blue*) and SCADA data at [REDACTED] with a synthetic noise metering error (*red*) applied.

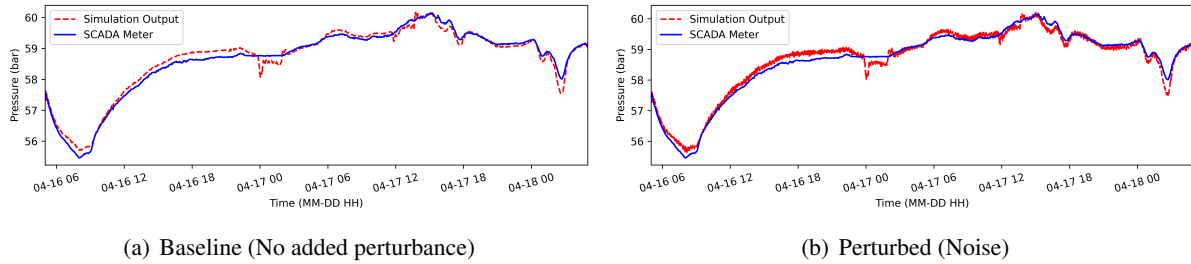


Figure 22: Comparison between SCADA data (*blue*) and simulated (*red*) pressure at [REDACTED] over time, for baseline and perturbed simulations.

4.2.5 Metering Errors on Flow Meters

The following metering error is applied at the [REDACTED] flow meter for a simulation period of two gas days between 16/04/2024 and 18/04/2024:

1. Noise with a mean of 0mcm/d and a standard deviation of 0.2mcm/d.

The application of the noise synthetic metering error perturbs the flow prescribed at [REDACTED] as shown in Figure 21. By comparing simulated data under this perturbation with the SCADA data generated by the unperturbed flow, it can be shown that the model is able to distinguish the two flow regimes, and hence identify when such a metering error may be present within the system. For illustrative purposes, the following discussion focuses on a noise metering error; however, a similar approach can be applied for the other metering errors, including calibration and multiplier errors, and similar conclusions can be drawn.

The gas flow perturbed by the effect of this noise at [REDACTED], shown in Figure 21c, is simulated for the test case. The effect of the flow perturbation is seen in the simulated pressure results at [REDACTED] through the small-amplitude fluctuations (Figure 22b) compared to both the baseline case (Figure 22a) and the SCADA data. These small-amplitude fluctuations are also seen in both the simulated flow and pressure at other sites in the network, as can be seen in Figures 23 and 24 respectively. As the distance from [REDACTED] increases, the magnitude of these oscillations is diminished, as would be expected—in particular, at [REDACTED] the baseline and noise-perturbed cases are almost indistinguishable.

The propagation of the flow perturbation to other sites in the network can be quantified through the comparison metrics between baseline simulated and perturbed simulated data. Figure 25 shows the value of RMSE between baseline and perturbed simulated pressure meters for the synthetic noise test case. The decrease of RMSE coincides with increasing distance from [REDACTED], which aligns with the conclusion drawn from Figures 23 and 24. The non-zero values in RMSE across multiple sites indicate that the gas flow simulated is different between the test cases, and hence that a metering error on a flow metrics likely to be present. Further, the relative magnitudes of RMSE over different sites can be used to localise the metering error to [REDACTED], where the value is largest.

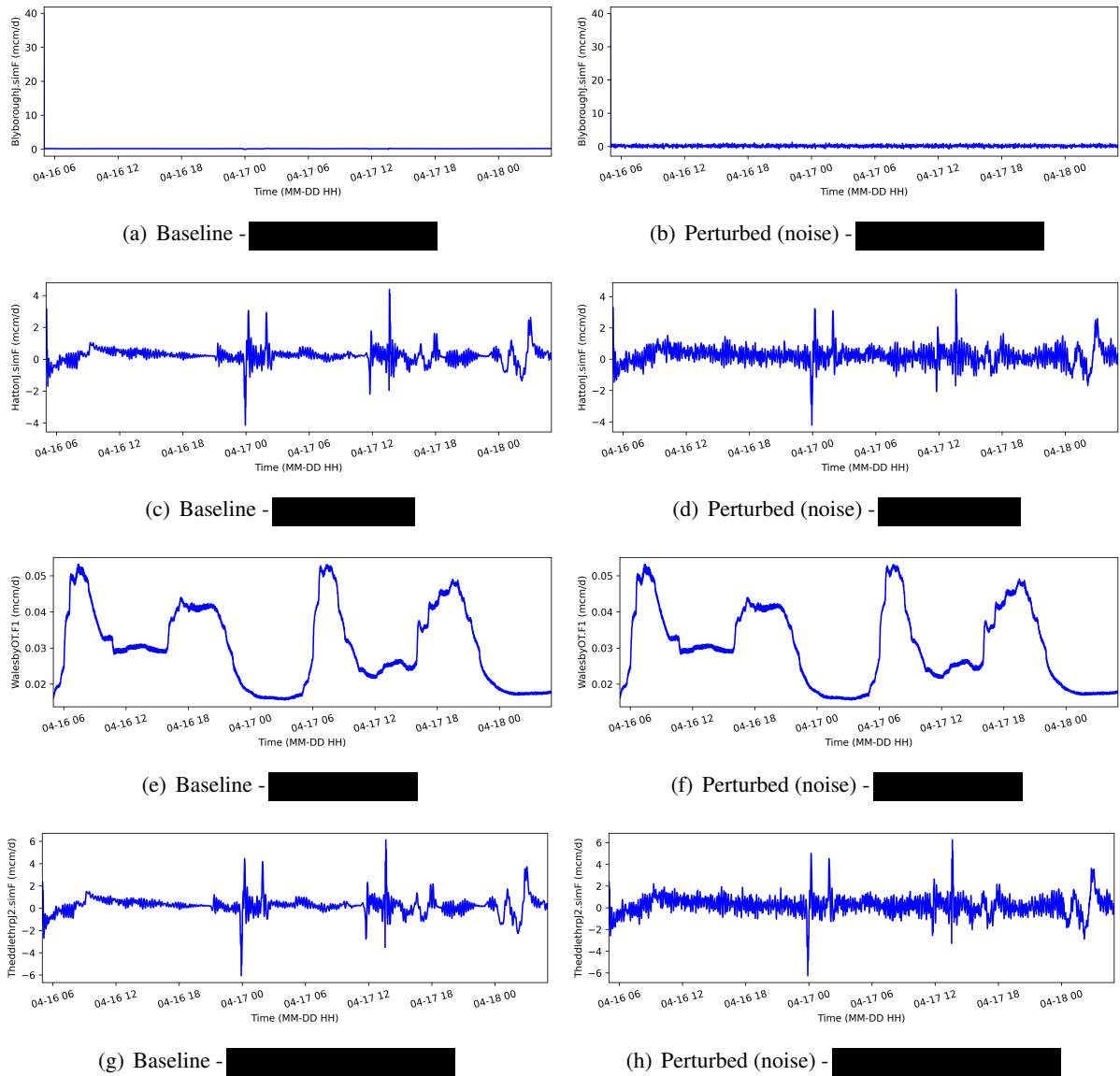


Figure 23: Baseline (*left*) and perturbed (noise) (*right*) simulated flow over time at all sites.

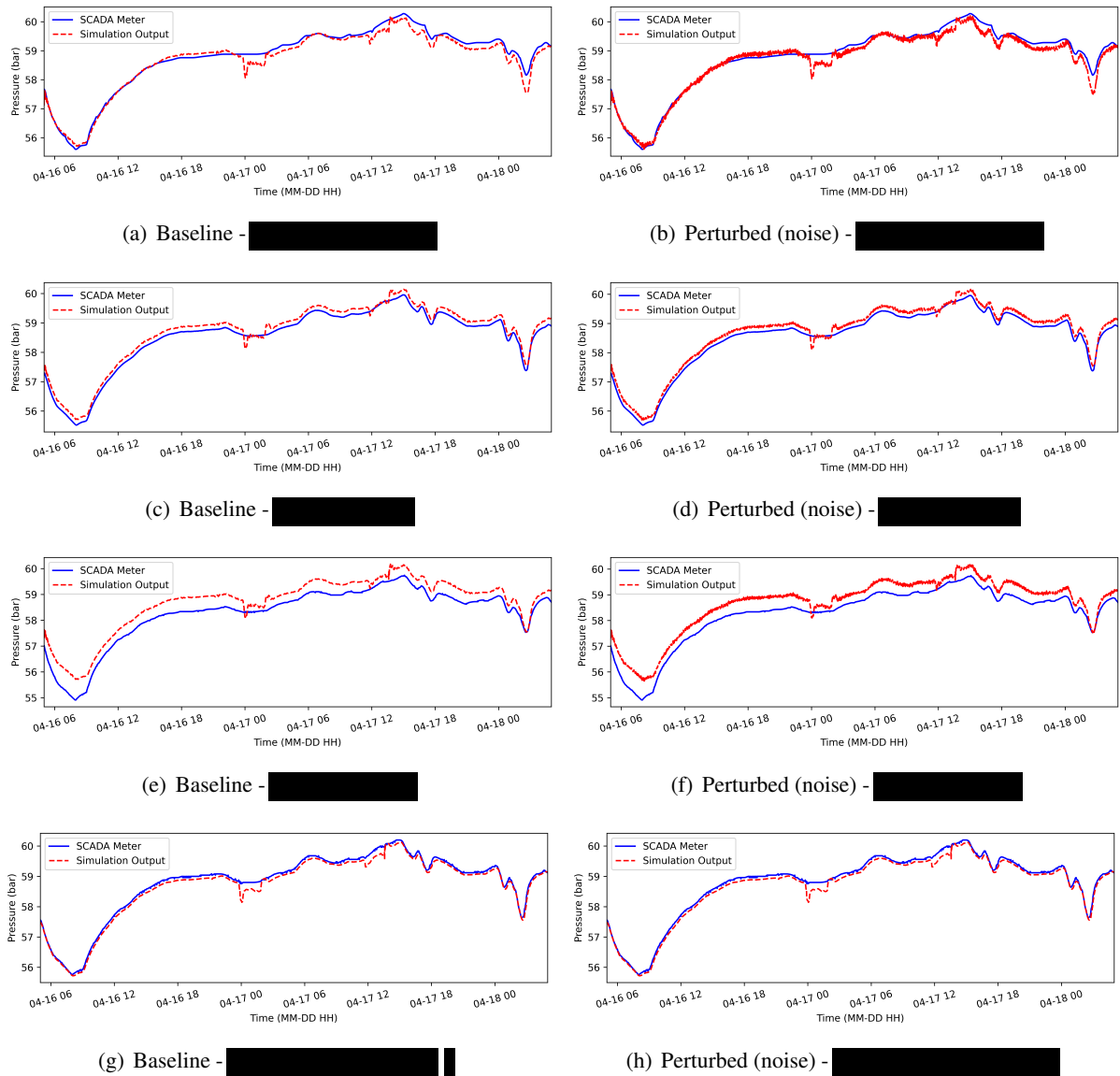


Figure 24: Baseline (*left*) and perturbed (noise) (*right*) compared to SCADA pressure over time at all sites.

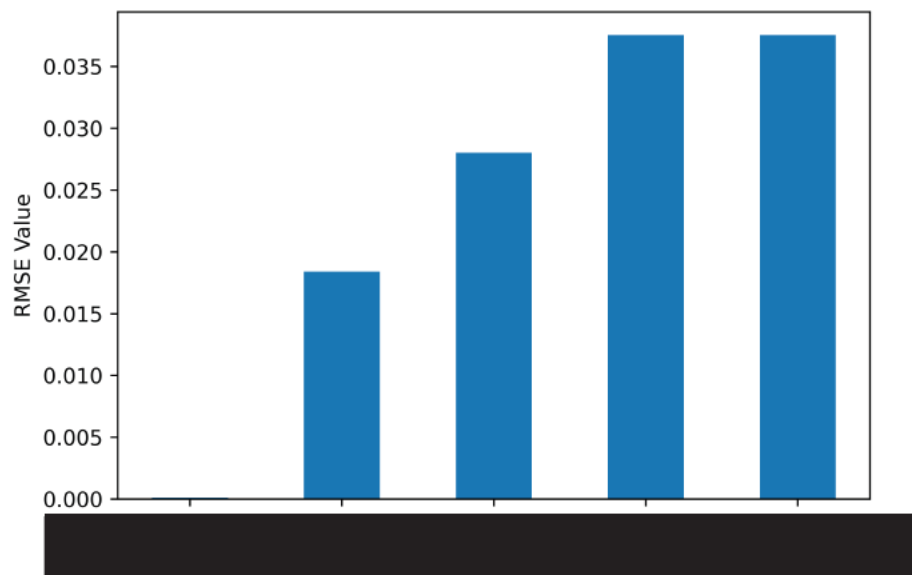


Figure 25: Root mean squared error (RMSE) between baseline and perturbed (noise) simulated pressure across all sites.



Figure 26: Section of network used in the [REDACTED] test case.

4.3 [REDACTED] Test Case

The [REDACTED] Test Case extends the [REDACTED] Test Case by adding [REDACTED] as shown in Figure 26. This allowed the model's behaviour to be validated on power stations and the model's behaviour to be tested using real metering error data recorded at [REDACTED].

The same simulation parameters, pressure calibration at [REDACTED] and external flow site setup were used as in the previous section, with modifications to the external flow to account for the extra contribution of [REDACTED]. A larger two-week time period between 5th January and 19th January 2024 was used in order to assess the change in simulation metrics over time when a metering error occurs for only part of the simulated time period. This time period was chosen as it contains SCADA data for all the sites within the test case.

A baseline simulation was run using the original SCADA data with no known metering errors present. This was followed by two cases where the SCADA flow data at [REDACTED] was modified to include real metering error data recorded by National Gas at [REDACTED]: [REDACTED] Error One and [REDACTED] Error Two. This allows for comparisons between the simulation results if the flow at [REDACTED] is as expected (Baseline), and the results if the meter data for [REDACTED] was genuine and corresponded to the flow perturbed by a contribution equivalent to historic metering error. These comparisons were used to demonstrate that the model could detect the presence of metering errors and quantify the associated amount of UAG.

4.3.1 Historic Metering Errors at [REDACTED]

[REDACTED] has two flow streams operated in a duty-standby arrangement. The metering supervisory system summates the flow rates from the two streams for an overall flow rate to be sent to SCADA. Two historic metering errors have been identified at this site: [REDACTED] Error One, which occurred between 26/09/24 and 09/10/24, and [REDACTED] Error Two, which occurred between 01/09/25 and 09/10/25. These were both caused due to noise from the standby meter being picked up in total station flow. These errors were seen in the SCADA data as spurious flow.

As these metering errors occurred outside of the time period for which SCADA data was provided, the flow data at [REDACTED] for the test case needed pre-processing. Twelve days of spurious flow data

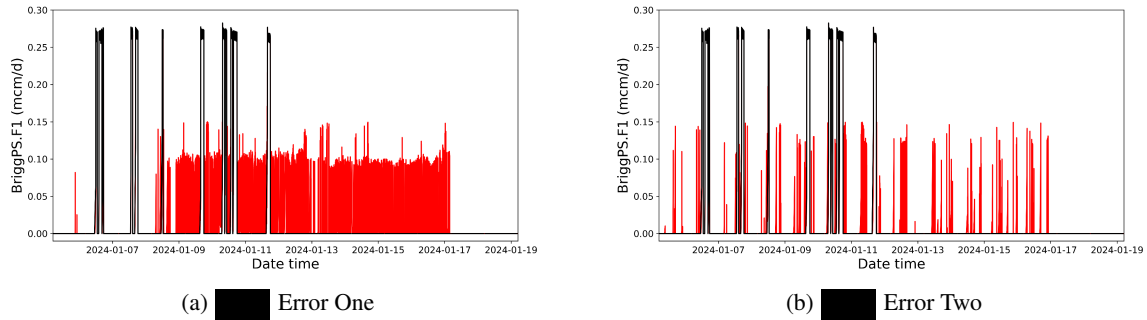


Figure 27: SCADA data for [REDACTED] with historic meter errors added. The baseline flow is shown in black and the spurious flow, corresponding to the metering error is shown in red.

were extracted from the SCADA flow data associated with each of the metering errors based on a threshold of 0.15 mcm/d. The data was added to the actual flow data within the two week time period to generate modified flow data for [REDACTED] for the two test cases, leaving the final two days in the processed test case data with no metering error in both cases. Figure 27 shows the resulting flow profiles at [REDACTED] on for both [REDACTED] Error One and [REDACTED] Error Two.

In these test cases, the aim is to calculate the root mean squared error (RMSE) between the baseline simulation and the simulations of [REDACTED] Error One and Two to determine whether the errors can be detected from differences between results from the two simulations and quantify the amount of UAG.

4.3.2 Simulation Results

All simulations on the [REDACTED] Test Case ran on 16 cores in approximately 16 minutes. The baseline simulation shows qualitative agreement in pressure and flow, which was comparable to the results in the [REDACTED] to [REDACTED] Test Case. Notably, the simulated pressure at [REDACTED] shows qualitative agreement with the SCADA data (Figure 28).

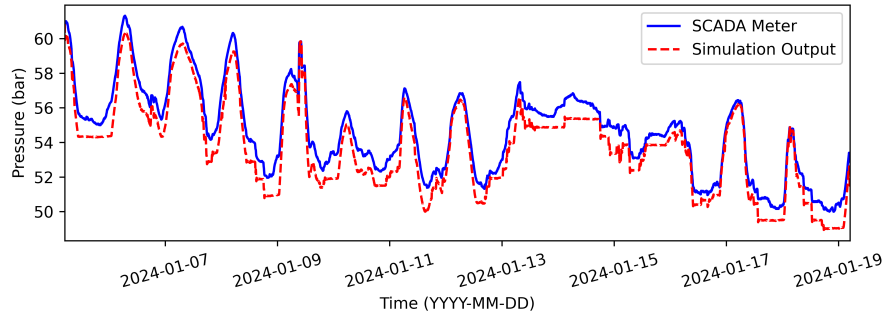
4.3.3 Metering Error Identification

As in the synthetic metering error cases, calibration errors due to anomalous features in the input data in the current version of the model increased the typical discrepancy between simulation results and SCADA data. These discrepancies affect the identification of metering errors, so the simulated baseline results were compared with the simulated results of the perturbed flow for each of [REDACTED] Error One and [REDACTED] Error Two. This is equivalent to assuming that the model is fully calibrated with the baseline simulation results replacing the input SCADA data.

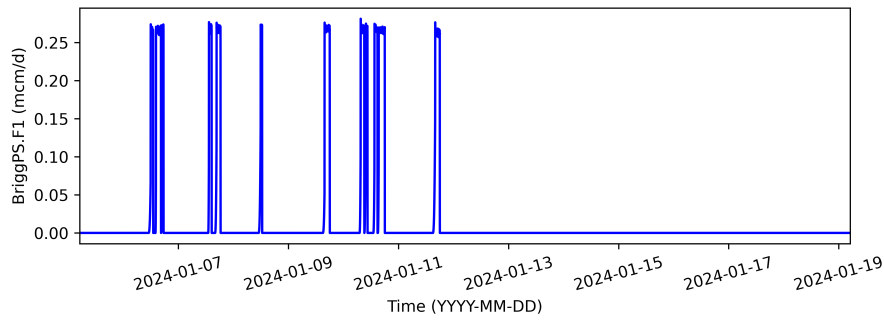
For both [REDACTED] Error One and Two, the extra noise in the flow meter at [REDACTED] is picked up in the simulation results through increased noise in the simulated pressure at [REDACTED] which is not present in the SCADA data. This can be seen through the increased fluctuations seen in the time traces of the difference in pressure between the baseline simulated pressure and the simulated pressure from each of the [REDACTED] Error test cases (Figures 29 and 31).

These figures show a difference of zero over the final two days, where no spurious flow was present, in both cases. The pressure differences also show characteristics of the associated metering errors, with the difference for [REDACTED] Error One showing more continuous fluctuations, and the difference for [REDACTED] Error Two showing more intermittent fluctuations.

The noise is picked up in the daily average RMSE metric when comparing the baseline simulation results with the simulation results for [REDACTED] Error One (Figure 30) and [REDACTED] Error Two (Figure 32). In both cases, the RMSE magnitude increases over the periods in which the historic metering error was applied to the existing SCADA data and decreases to zero at the end of the time period, when the metering error is not present, demonstrating the model's metering error detection capabilities.



(a) SCADA vs simulated pressure



(b) SCADA flow

Figure 28: Pressure and flow at [REDACTED].

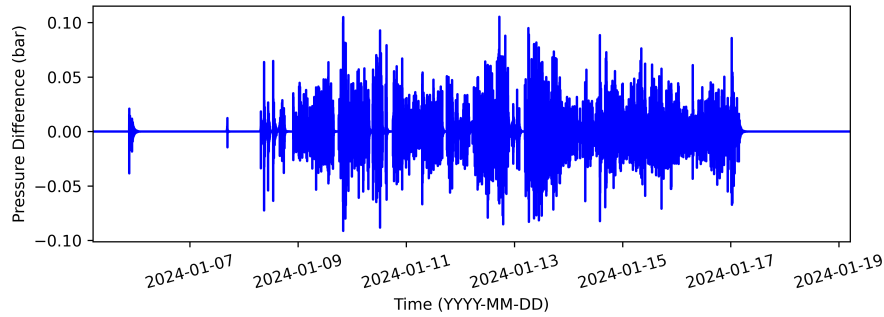


Figure 29: Baseline vs [REDACTED] Error One simulated pressure difference.

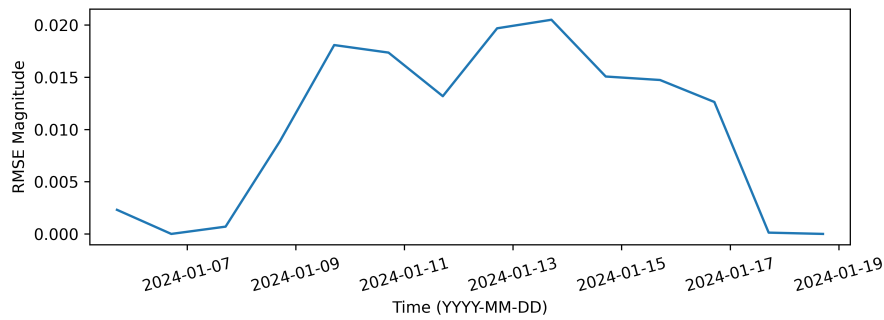


Figure 30: Baseline vs [REDACTED] Error One RMSE over time.

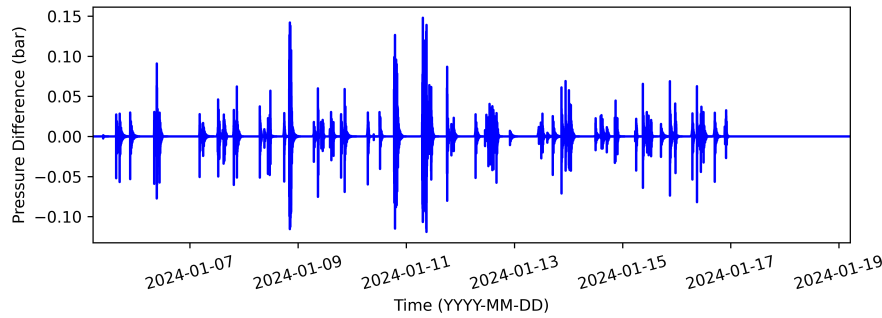


Figure 31: Baseline vs [REDACTED] Error Two simulated pressure difference

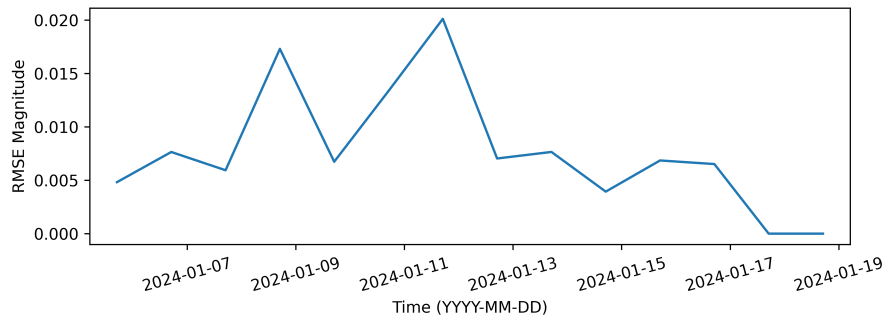


Figure 32: Baseline vs [REDACTED] Error Two RMSE over time

Characteristics of the metering errors can also be interpreted from the daily-averaged RMSE time traces. [REDACTED] Error One has a more consistently high RMSE value for the time period over which the error was applied, whereas [REDACTED] Error Two shows more variability in RMSE, with the decreased daily magnitude being consistent with the more intermittent nature of the error.

4.3.4 Unaccounted for Gas

Having identified the presence of these historic metering errors from the simulation results, the network energy balance can be calculated to demonstrate that these errors contribute to UAG. Figure 33 presents the total network energy balance, E_{bal} , calculated using results from simulating the baseline flow (blue) and flow perturbed by [REDACTED] Error One (red). A difference between the two values can be qualitatively seen in this figure, which Figure 34 quantifies. This difference arises because the two different flow regimes—baseline and perturbed—are being simulated. This difference in flow leads to a difference in the total energy contained within the network, which indicates the presence of UAG within the network. The good agreement between the time period where this difference is larger with the presence of [REDACTED] Error One supports attributing this energy difference to this metering error. To quantify the associated amount of UAG following NGT's methodology, however, the difference should also be calculated using SCADA data with and without this metering error. Similar results were obtained when applied to [REDACTED] Error Two, although the magnitude of the difference was smaller owing to the increased intermittency in the spurious flow.

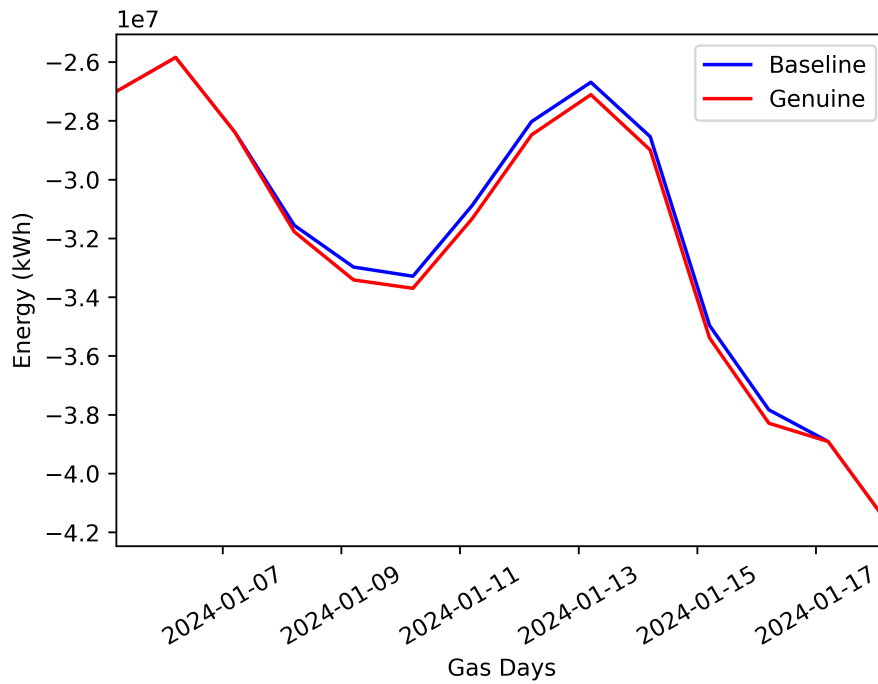


Figure 33: Total daily network energy balance calculated using simulation results from the unperturbed baseline flow, (blue) and flow perturbed by the effect of Error One (red).

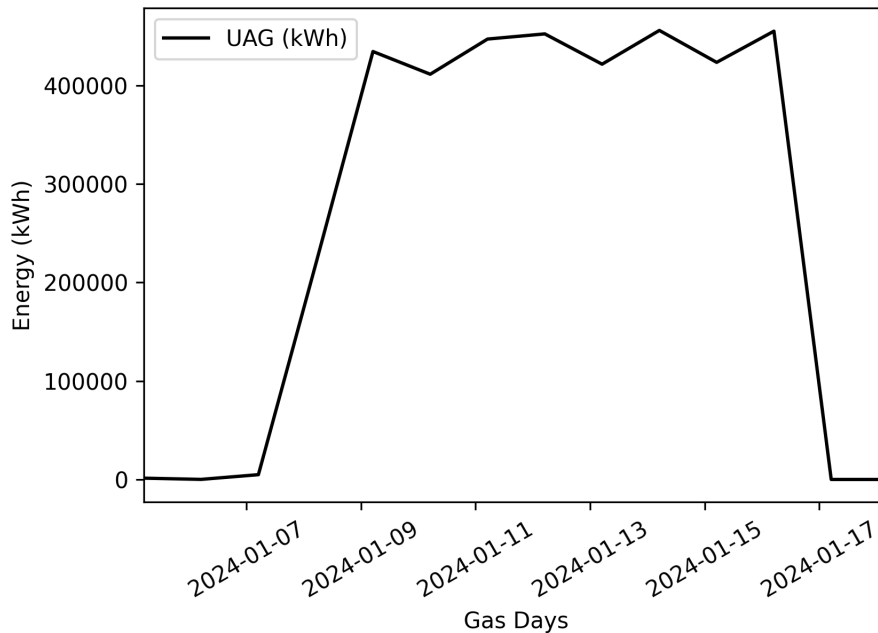


Figure 34: Difference between total daily network energy balance calculated using simulation results from the unperturbed baseline flow, and flow perturbed by the effect of Error One.

4.4 Application to the North-East of England Section of the Network

4.4.1 Problem Overview

National Gas calculate UAG on the entire NTS, rather than subsections of the network as considered in this project. Therefore, to demonstrate feasibility of this approach for future integration with typical operations, the model has been applied on a larger section of the network to demonstrate that the approach is scalable. In particular, the model should be capable of simulating the gas flow across a network with a larger number of assets, in practicable time frames, and obtain qualitatively correct behaviour. The model should also be able to calculate contributions to the network energy balance to show that the methodology can extend to calculating UAG on the entire NTS.

4.4.2 Test Case Definition

The section of the NTS in the North-East of England spanning from [REDACTED], shown in Figure 1, has been used to demonstrate this feasibility. This region consists of 800km of pipeline and approximately 50 assets and includes additional complexity compared to the previous test cases. These features include the addition of storage sites allowing bidirectional flow, closed loops of pipelines, and a wider range of flow regimes, including larger volumes of flow with increased gas flow rates, which can reach as high as 75mcm/d. The gas flow simulation was run on the North-East section for two gas days between the 5th January 2024 and the 7th January 2024, and executed in 73 minutes on 8 cores.

4.4.3 Simulation Results

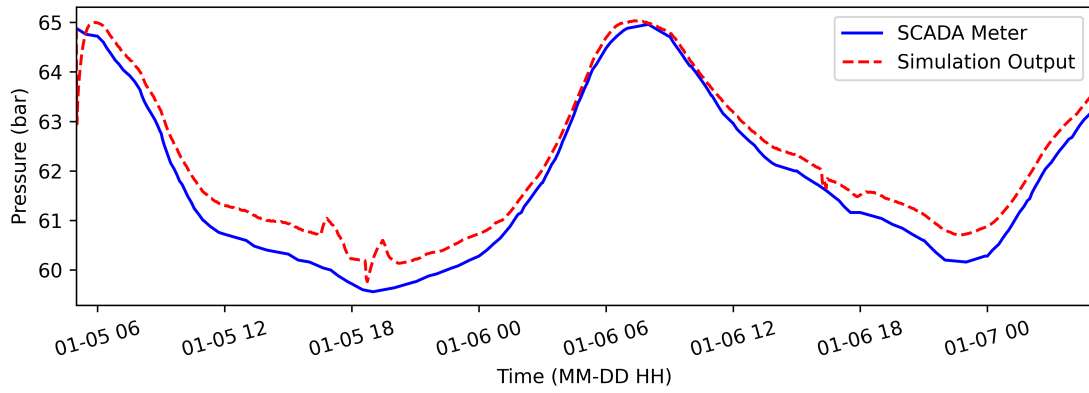
The simulation results on the North-East Section are promising and showed a good qualitative agreement with the general trends in the SCADA pressure data throughout the region. There was particularly good quantitative agreement at sites near to the entry sites in this region, as can be seen in Figures 35(a) and (b) for [REDACTED], respectively. The main discrepancy in pressure at [REDACTED] at 6pm on the 5th January can be attributed to pressure spikes at [REDACTED], which highlights the need for a more comprehensive data cleaning process to improve the quality of the input data used in the simulation. The quantitative agreement between simulated and SCADA pressure data was weaker further into the region, for example at [REDACTED] (Figure 35(c)). These results suggest that developing an improved understanding of the additional flow out of the region at [REDACTED] and specifics of the network topology, including pipeline lengths and diameters, are required to improve the accuracy of the model.

4.4.4 Energy Balance Calculation

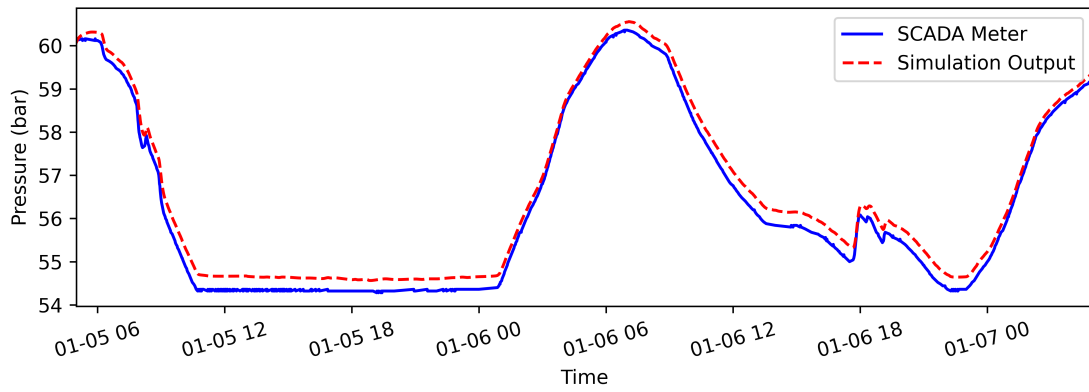
The energy balance was calculated across the full North-East section using three months worth of SCADA data between the 01/02/2024 and 30/04/2024. Figure 36 shows the energy balance and its subsequent constituents calculated. This also includes a contribution from the compressor fuel usage at [REDACTED].

Figure 37 shows the total energy entering and exiting the system. Similarly to the energy balance calculated for the [REDACTED] Test Case in Section 4.3.4, the energy balance is dominated by the energy difference. In Figure 37, the magnitude of the mean total energy Input is approximately 1.0×10^9 kWh and the magnitude of the mean total energy output is 2.5×10^8 kWh. This difference can be explained by the significant amount of gas entering the network through [REDACTED], and missing information about flow leaving this section of the NTS, for example near [REDACTED].

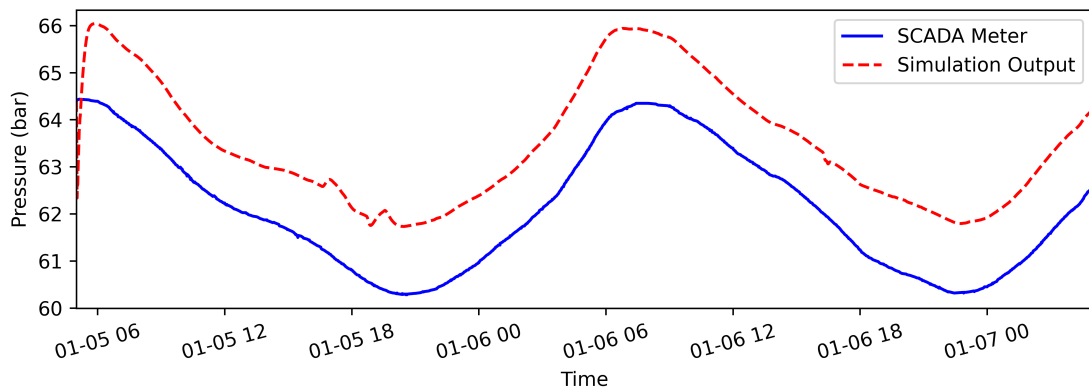
Figures 36 and 37 demonstrate that the energy balance can be calculated across the full North-East region, and that scaling up the energy balance calculation to calculate UAG over the full NTS is feasible. It should be noted that calculating UAG over the full NTS will in fact be easier than on subsections of the network considered in this project for the following reasons:



(a) [REDACTED]



(b) [REDACTED]



(c) [REDACTED]

Figure 35: Simulated pressure (red dashed) compared against SCADA pressure data (blue) at three sites in the North-East section.

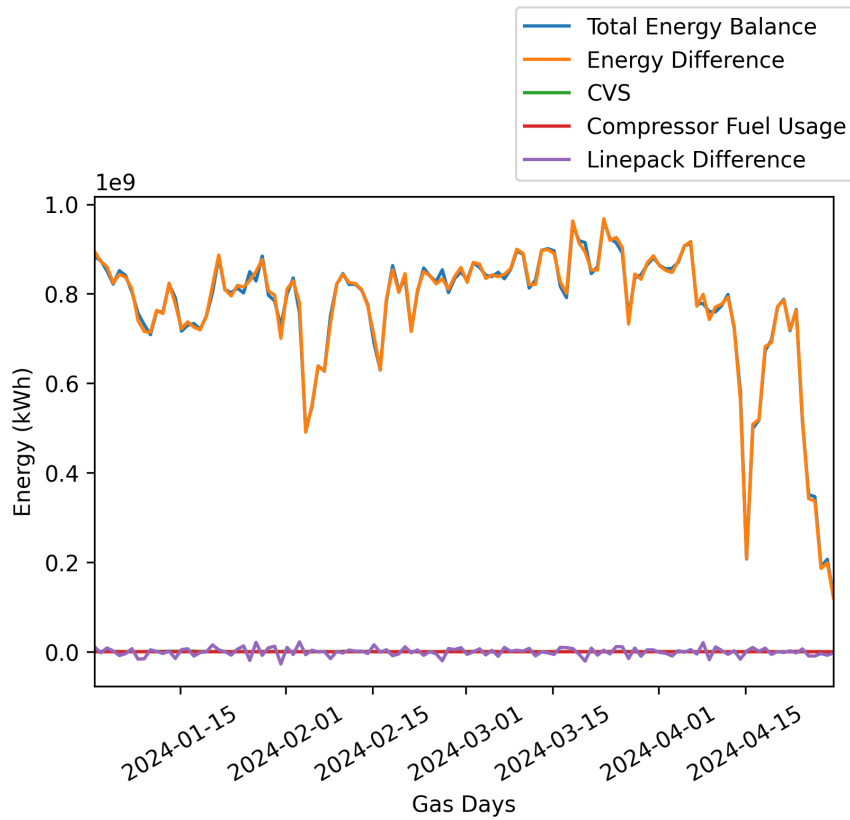


Figure 36: Energy Balance and the constituents of the Energy Balance calculated across the full North-East network section between 01/02/2024 and 30/04/2024.

- The total CVS can be calculated directly from the Energy Report Template [5], removing the need to assume that each offtake in an LDZ contributes equally.
- The calculated network energy balance will directly correspond to UAG: as the NTS is a closed system, the difference between inputs and outputs is expected to balance the change in energy contained within the NTS.

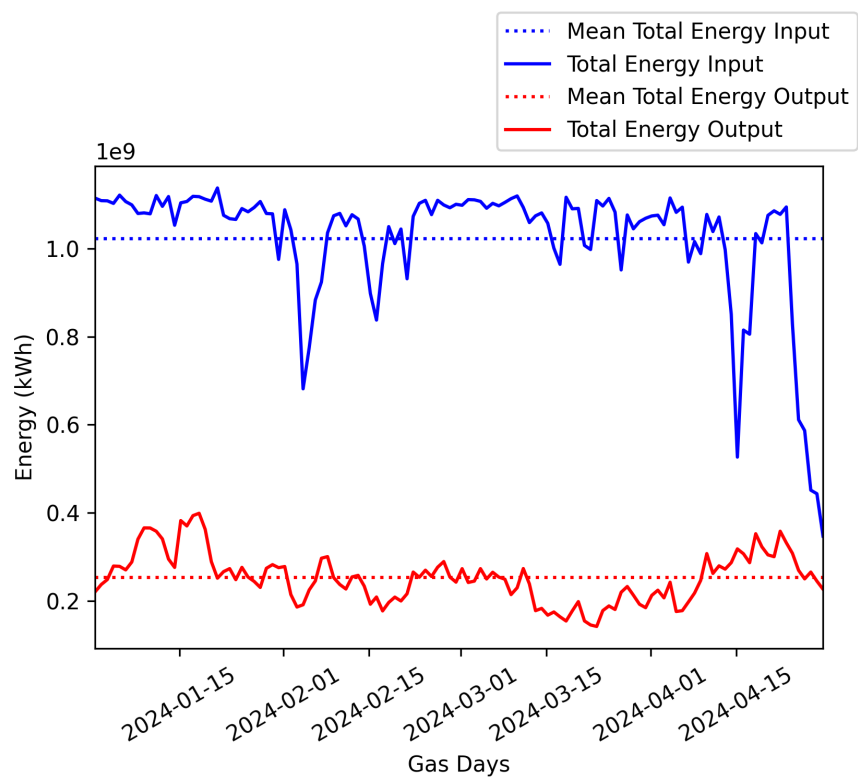


Figure 37: Total Energy Input and Output calculated across the full North-East network section between 01/02/2024 and 30/04/2024.

4.5 Discussion and Interpretation of Results

The [REDACTED] Test Case (Section 4.2.5) demonstrated the core functionality of the model, namely simulating the gas flow within sections of the network given input historic SCADA data, calculating terms in the network energy balance to identify the presence of UAG, and applying synthetic metering and transmitter errors to demonstrate how comparing simulated and SCADA data can be used to identify the presence of metering and transmitter errors. Results from this simulation showed good qualitative agreement with SCADA data across sites in the section of the network and captured key trends and features in the gas flow dynamics.

Synthetic metering and transmitter errors were applied at two different sites in this test case: firstly, at the pressure transmitter at [REDACTED]; and secondly at the flow meter at [REDACTED]. The root mean squared error (RMSE) between the simulated and SCADA data with a transmitter error applied at [REDACTED], showed a significant increase in magnitude in this value at [REDACTED] only for each of the transmitter errors applied. This showed that the model could detect and localise these transmitter errors. The perturbed flow due to the additional noise at [REDACTED] was simulated and showed that this perturbation propagates to other sites in the network. This increased the RMSE between simulated and SCADA pressure data at multiple sites in the network, which enabled the metering error to be detected. Further, initial results indicated that the model results could be used to localise this flow metering error to [REDACTED] from identifying the sites with highest RMSE.

The [REDACTED] Test Case (Section 4.3.4) demonstrated the performance of the model on two historical metering errors. The simulation results showed the propagation of the flow perturbation representing the applied metering errors to the pressure meter at [REDACTED]. This increased RMSE between simulated and SCADA data at this meter over the time period when each metering error was applied demonstrated that the model could detect these errors. Calculating the difference in the total network energy balance using the baseline flow, and using the flow perturbed by the effect of the historical metering error, demonstrated that the metering errors were a source of UAG. These results demonstrated that the model was able to detect realistic metering errors and identify them as sources of UAG.

The North-East of England Test Case (Section 4.4.4) demonstrated the capability of the model to scale to larger networks. Results from this test case showed that the model was capable of simulating the gas flow across a network with a larger number of assets, in practicable time frames, and obtained qualitatively correct behaviour across various sites. The results also showed that contributions to the network energy balance could be calculated, and the methodology naturally extends to calculating UAG on the entire NTS. This provides strong evidence that this approach could be extended to the national network and provide useful results in a practicable time.

Overall, results from applying these test cases have demonstrated the feasibility of using a forward predictive model to simulate the expected network dynamics from input telemetry data, and use the results to identify potential metering errors as sources of UAG. However, the significant complexity of the problem meant that various challenges were faced over the course of the project. The primary challenges included the significant variability in the quality of the input SCADA data and missing a complete definition of the network topology. This resulted in manually validating the SCADA data used in the test cases to ensure model stability, and making various modelling assumptions related to lengths and diameters of missing pipelines. Over the course of this project, it was recognised that these assumptions should be validated. For example, in the model, [REDACTED] and [REDACTED] are assumed to be connected by such a short pipeline. During the project it was identified through discussions with NGT that the pipeline between the two sites is far longer, at a length of 10km, and possesses a much smaller diameter than was modelled. This difference could partially explain the larger than expected pressure drop of 1.5 bar between the two sites. Small sections of the network were used in this project to support initial model development. However, information about additional flow entering or exiting these sections of the network through other pipelines, was not readily available. To address this missing information, an external site was added to the simulation provide a net flow balance. While this resulted in representative typical flow values, the simulated flow was also significantly noisier than might be expected. This motivates further validation of this modelling assumption and for future test cases to focus on closed systems within the network, such as

the full NTS. These challenges affected the ability to fully calibrate the model against the input SCADA data. Future work to undertake additional pre-processing of the input SCADA data and validation of these modelling assumptions could improve the accuracy of the model and also the confidence in the predictions of metering errors and sources of UAG identified by the model.

A number of additional areas for future development to improve model robustness and to scale the model to the entire NTS, which will be discussed in Section 5.4.

4.6 Key Deliverables

The Statement of Work [1] identified the following deliverables for Phase 2: Model Development and Phase 3: Concept Validation. This section summarises the extent to which these deliverables were achieved in this project.

- **2.1: Working Prototype of the Model:** This deliverable was fully met and evidenced in Section 4.

Results from the application of the model on a collection of test cases were presented in this section and it was demonstrated that the developed model is functional and able to simulate expected network behaviour, identify metering errors and the presence of UAG.

- **2.2: Technical Documentation:** This deliverable was fully met and evidenced in Sections 3 and 4.

Section 3 provided a detailed technical explanation of the model architecture and assumptions that were made. Section 4.5 detailed the limitations of the model in the context of the results from executing the test cases.

- **2.3: Initial Performance Benchmarks:** This deliverable was fully met and evidenced in Section 4.

By applying the model to a range of test cases of varying scale and complexity, it was demonstrated that the model was able to qualitatively capture key features and trends in the gas flow dynamics. The model was also able to identify the metering errors applied to these test cases by quantifying the discrepancy between simulated and SCADA data. By calculating terms in the network energy balance, the presence of UAG in the [REDACTED] test case could be identified. These results demonstrate that the model is largely qualitatively accurate and reliable.

However, it was not possible to provide a complete quantitative assessment of the model accuracy and reliability in the limited project time. This is primarily due to the challenges related to the quality and completeness of the input data to the model, as described in Section 2.

- **3.1: Validated Model:** This deliverable was fully met and evidenced in Section 4.

Towards the start of the project, it was identified that the 18-months of provided SCADA data did not contain any known metering errors. Consequently, NGT identified and provided data for two historical metering errors at [REDACTED] outside this time period and to test the model's performance on historical anomalies.

Section 4.3.4 presented the results from applying the model to two historical metering errors at [REDACTED]. The results from these test cases demonstrated that the metering errors could be identified qualitatively, from the differences between simulated and input pressure, and quantitatively, from the positive values of RMSE over the time period over which each error was applied. The test cases also demonstrated that there was a change in the network energy balance, which indicated that these metering errors correspond to UAG. Given the limitations on input data, the model was therefore accurate when applied on these test cases; however, a more thorough analysis is needed in future using a wider range of test cases to quantify the accuracy of the model.

- **3.2: Model Performance Report:** This deliverable was fully met and evidenced in Section 4.

Section 4.5 provided a summary of the model's performance, accuracy and limitations on a set of test cases.

- **3.3: Recommendations for Enhancements:** This deliverable was fully met and will be evidenced in Section 5.

Section 5 will document a series of recommendations for improving model robustness and operational readiness.

5 Next Steps

This work was a Proof-of-Concept project which aimed to demonstrate the feasibility of the key methods required to take a simulation based approach to understanding metering error and sources of UAG and shrinkage in the NTS. The results documented in this report have demonstrated the feasibility of this approach and, unsurprisingly, have identified a number of avenues of future work to improve the model and value of the insights it provides. This section summarises these areas of future work which should be incorporated into any follow-on project to improve the robustness of the model for operational readiness. This section serves as evidence for the key deliverable: Phase 3: Recommendations for Enhancements.

5.1 Improving Model Accuracy

As discussed throughout this report, the initial simulation of the gas network implemented in this project makes various assumptions. The initial results clearly show generally good qualitative agreement between the simulation and SCADA data; however, there are some discrepancies, which affects the confidence in the model predictions of sources of UAG. These discrepancies are primarily due to the network-wide assumptions made in the current version of the simulation and the quality of the input data to the model. A first step will be to validate how well these assumptions, such as constant gas properties, hold across the network and where these assumptions need to be relaxed and spatially and/or temporally varying values across the network used.

In addition to the modelling assumptions made on the gas properties, more detailed network topology data is required to accurately predict gas dynamics at all sites across the network. In particular, this includes assessing the validity of the modelling assumption to connect sites by short pipelines with default diameter and length when pipeline data is missing. For example, in the model, [REDACTED] are assumed to be connected by such a short pipeline. However, during the project it was identified through discussions with NGT that the pipeline between the two sites is far longer, at a length of 10km, and possesses a much smaller diameter than was modelled. This difference could partially explain the larger than expected pressure drop of 1.5 bar between the two sites. The use of a default pipeline diameter should also be addressed when connecting sites with large flow rates, for example near [REDACTED]. The pipeline diameter has a significant impact on the velocity of the gas within and by extension the largest stable time step and hence the execution time of the simulation.

The approach to modelling the gas flow in sections of the network, where the volume of flow entering or exiting through connected pipelines out of these sections, should also be validated. Currently, this was achieved through adding an artificial external flow site to ensure net flow balance. The general direction of flow and location of net flow leaving and entering the sections of the network used in these test cases were known by subject-matter experts at NGT and used to determine appropriate locations. However, when extending to larger more complex networks, this domain knowledge may be insufficient in capturing the gas behaviour extending beyond the network modelled. Alternative approaches to account for the missing flow data out of sections of the network should be considered to understand the impact that this assumption has on results.

All of these are possible to achieve within the current model architecture and will just require time to implement, validate and demonstrate improvements in the model accuracy. This should form a self-contained work package in any future project.

5.2 SCADA Data Processing

Section 2 discussed the challenges associated with preprocessing and cleaning the SCADA data. This is a critical step in the use of the model as the SCADA data is used for a range of tasks including network definition, model calibration and providing measured data against which to compare simulated data. Over the course of this project, this step required significant effort and manual intervention to convert the raw SCADA data into a consistent, usable format. Going forward, with the knowledge gained from this project, work should be undertaken to automate these methods and account for the additional edge cases that will

undoubtedly arise. This will be essential to improve the accuracy of the model and enable the model to be expanded to cover more test cases, larger networks and longer time periods.

5.3 Application to Additional Test Cases

This project has focused on a small number of test cases to demonstrate the viability of the model. Whilst these have shown promising results, the model should be validated against a wider set of test cases covering different features of interest across the NTS. Specific test cases have not yet been defined; however, the following features have been identified as being of interest:

- Multiple metering errors to determine whether more than one metering error can be identified and understand the impact of cancellation effects from different metering errors on UAG.
- A wider range of metering errors to include different types of error and magnitude.
- Different flow regimes across the NTS and at different times of the year to assess seasonal variability.
- Longer time periods to detect potential metering drift and deterioration.
- Larger sections of the NTS to determine whether certain assets and metering systems are more susceptible to metering error.
- Uncertainty in model inputs to better understand the key drivers of variability in network behaviour and assess confidence in the predictions.

Applying the model of these additional test cases will enable assessing the extent to which the current comparison metrics can detect different types of metering errors, at different sites and time of year. Results from these test cases could indicate modifications to these metrics to improve the reliability of metering error identification and localisation. This will be essential for building confidence in the model's predictions of sources of UAG, which will be required if applying the model for operations.

All of these test cases are feasible, although it should be noted that there will again be challenges associated with preprocessing the required input data. Future projects should define these cases and their value as early as possible to mitigate these risks.

5.4 Extension to National Transmission System

The choice of simulation architecture, especially the use of PETSc, naturally enables the simulation to be extended to the entire NTS at the expense of increased computational cost. However, there is still room for improvement to reduce the simulation run time and computational burden which will enable faster uncertainty quantification through executing more scenarios in a shorter time frame. There are a number of ways this could be achieved, including looking at improving the preconditioners used, investigating the use of implicit solving schemes and general optimisations to the simulation code.

This project has identified some additional complexities when scaling up to more complex networks through building a model of the north-east subsection of the NTS. Currently, only passive network components have been considered. When extending to a gas network on a national scale, additional complexity may be introduced through the modelling of active network components such as compressors and control valves. The inclusion of these assets may improve model accuracy through the correct modelling of their impact on pressure and flow behaviours within the network.

The other significant challenge to scaling to the national network is the need for large scale cleaning of the input SCADA data. This challenge has already been extensively discussed in Section 2, and it is worth emphasising that a significant body of work will have to go into extending this to the national network.

Overall, extending to the national network presents some challenges but should be feasible from a technical point of view.

6 Cost Benefit Analysis

This section constitutes the Cost Benefit Analysis (CBA) for the UAG Predictive Tool developed in this project. In particular, it demonstrates the potential to deliver a net benefit to consumers.

6.1 Costs Associated with Unaccounted for Gas

Unaccounted for Gas is a significant problem for the UK's gas network and forms a dominant component of the total NTS shrinkage, which comes with associated financial implications. There is no direct cost to NGT from the amount UAG as these costs are passed on to gas shippers and consumers. However, UAG does in fact have indirect financial implications for NGT. When errors in metering or data are discovered, a reconciliation process is initiated, which can result in significant billing adjustments. These costs may fall on either NGT, asset owners, or gas shippers, depending on the error. Additionally, NGT bears operational costs related to investigating, validating, and processing these errors. This includes the expense of engaging third-party independent meter consultants for significant discrepancies. Beyond direct financial costs, there are also soft costs such as reputational risk, regulatory compliance obligations, and the administrative burden of managing disputes and customer communications. As financial costs were not available to provide quantitative evidence to support this CBA, the following focuses on providing qualitative evidence for the benefits that this solution provides.

6.2 Benefits of the Solution

The solution developed in this proof-of-concept project was a predictive tool to identify causes of UAG. This solution provides a number of benefits to NGT and other recipients, and could be realised on a national scale with a wider roll-out of the solution. These benefits include:

- **Improved understanding of typical network behaviour:** Simulating the gas flow dynamics within the network can provide NGT and asset owners with a better understanding of the expected behaviour in a range of scenarios, e.g. variable flow regimes, different times of year, different locations within the network. This could help asset owners to identify deteriorating assets before they cause significant discrepancies and billing errors.
- **Easier identification and localisation of metering errors and sources of UAG:** Quantifying the discrepancy between telemetry data and simulated data enables potential metering errors and sources of UAG to be identified and localised. This can improve operational efficiency for NGT through targeted investigations into the most likely sources of error, proactive maintenance and targeted meter calibration, reducing downtime and improving network reliability. This could lead to earlier identification of metering errors, which could reduce billing errors for customers.
- **Improved reporting of network performance:** Understanding and quantifying how the telemetry data differs from the expected behaviour across the network can provide a better understanding into the full network performance. This can support NGT activities into regulatory compliance by improving measurement accuracy and reporting, and support any long-term strategic planning activities.

These benefits provided by this predictive tool could allow NGT to identify sources of UAG more efficiently, ultimately reducing the amount of UAG and the associated financial, operational, environmental and regulatory implications.

6.3 Replication Potential in a Wider Roll-Out

The solution was developed to naturally enable future extension to the full NTS using the same methodology. Section 5.4 outlined the following main developments required to achieve this wider roll-out:

- Improving model accuracy to improve performance of the model.

- SCADA data processing to improve model robustness.
- Application to additional test cases to improve confidence in model predictions.
- Extension to full NTS to incorporate all assets into the model and identify metering errors across the full network.

While this extension would be a significant undertaking, this approach would provide a much deeper understanding of the typical behaviour of gas flow across the NTS and more efficient identification of metering errors and sources of UAG, and realise the benefits outlined above on a national scale.

7 Conclusions

This proof-of-concept project aimed to develop a methodology to use a combination of historic SCADA data and a forward simulation of the transient behaviour of the gas flow in the NTS to identify sources of UAG. This should enable meter errors to be better identified and quantified, yielding a better understanding of the behaviour of the NTS. This is a novel approach which requires the development and validation of various model components.

Following a requirements capture phase, an appropriate model architecture was defined. The model workflow consists of preprocessing and cleaning input SCADA data to define a network and its inputs, instantiating a simulation of the transient dynamics of the network, executing the simulation to predict expected measures at meters, and finally comparing the simulated and measured meter data to identify discrepancies. Most of this tooling is written in Python; however, the transient dynamics simulation is built upon the PETSc framework to allow arbitrary scaling to larger amounts of compute required for larger networks. This model was also developed with the future in mind, and is readily extensible to networks and analyses of a larger scope.

The model was verified against a set of test cases identified in published literature. These test cases showed good agreement to expected results and provide confidence in the simulation. The complete model pipeline was then applied to various test cases of differing scale and complexity. This included a test case to demonstrate the core functionality of the model, which demonstrated that the simulation can qualitatively capture a range of key features and behaviour of the gas flow dynamics across small sections of the network. They also demonstrated that the model can detect the presence of metering errors in the network and additionally indicate the most likely locations for these errors. The application of the model on two historic metering errors that NGT identified at [REDACTED] demonstrated that the approach can detect realistic metering errors and identify these errors as source of UAG. A model of the entire North-East of England section of the NTS was also implemented to test the viability of modelling larger sections of the network. This demonstrated the viability of this approach to scale to larger sections of the network and eventually to the full NTS.

A key challenge that was identified over the course of this project was the extent to which the SCADA data and other network definition sources needed to be preprocessed and cleaned. Significant manual work was undertaken in this project to prepare the data for use to enable the test cases to be executed, which impacted the project outcomes. Most notably, the uncertainty in the quality of the input SCADA data and complete network definition meant that the simulation could not be fully calibrated against SCADA data, which reduced the accuracy of the gas flow simulation and led to larger discrepancies between the simulated and SCADA data. This meant that metering errors in test cases in this project were identified manually from model outputs. Overcoming this challenge will be essential to any further work to identify a wider range of metering errors and improve confidence in the results obtained from the model.

Overall, this work has demonstrated the feasibility of this approach and identified a number of areas for future development to improve model robustness and to scale the model to the entire NTS. While this extension would be a significant undertaking, this approach would provide a much deeper understanding of the typical behaviour of gas flow across the NTS and the effect that different metering errors have on the network. This method has significant future potential to provide more efficient identification and localisation of metering and transmitter errors and sources of UAG. Using this approach could therefore improve operational efficiency through targeted investigations into the most likely sources and reduce the financial costs passed on to gas shippers and consumers through faster identification of errors.

References

- [1] D. Millington-Smith, “Statement of Work: Predictive Tool for Unaccounted-For Gas (UAG) Identification,” 2025.
- [2] N. G. Transmission, “UAGCVS Report,” November 2025.
- [3] N. G. Transmission, “Pipeline Databook.”
- [4] N. G. Transmission, “Supply and Demand Data - Site level.xlsx.”
- [5] N. G. Transmission, “Energy Report Template.xlsx.”
- [6] N. G. Transmission, “COAS Compressor V5_1.xlsm.”
- [7] N. G. Transmission, “Offtakes By LDZ.xlsx.”
- [8] A. Herran-Gonzalez, J. De La Cruz, B. De Andres-Toro, and J. Risco-Martin, “Modeling and simulation of a gas distribution pipeline network,” *Applied Mathematical Modelling*, vol. 33, no. 3, pp. 1584–1600, 2009.
- [9] Y. Qiu, S. Grundel, M. Stoll, and P. Benner, “Efficient numerical methods for gas network modeling and simulation,” *arXiv preprint arXiv:1807.07142*, 2018.
- [10] N. G. Transmission, “Gas Quality Trends Sheet.”
- [11] J. Kiijarvi, “Darcy friction factor formulae in turbulent pipe flow,” *Lunowa Fluid Mechanics Paper*, vol. 110727, no. 2011, pp. 1–11, 2011.
- [12] A. Zlotnik, M. Chertkov, and S. Backhaus, “Optimal control of transient flow in natural gas networks,” in *2015 54th IEEE Conference on Decision and Control (CDC)*, pp. 4563–4570, 2015.
- [13] S. M. Ahmadi, M. Behbahani-Nejad, and Y. Shekari, “Transient modeling of natural gas in pipeline networks by two non-iterative explicit and implicit finite volume methods,” *Applied Mathematics and Computation*, vol. 478, p. 128793, 2024.
- [14] S. Balay, W. D. Gropp, L. C. McInnes, and B. F. Smith, “Efficient management of parallelism in object oriented numerical software libraries,” in *Modern Software Tools in Scientific Computing* (E. Arge, A. M. Bruaset, and H. P. Langtangen, eds.), pp. 163–202, Birkhauser Press, 1997.
- [15] B. Kirkman, “Measurement Error Report EA018 (MER/CAD/225/22) Royston,” 2022.
- [16] P. D. Ben Kirkman, “SMER EM011 Thornton Curtis,” 2023.
- [17] B. Kirkman, “Horndon Significant Meter Error Report,” 2011.
- [18] K. Vugler, “Summary Report for Aberdeen SMER SC006,” 2013.
- [19] K. V. Ben Kirkman, “INDEPENDENT EXPERT SIGNIFICANT METER ERROR (SMER) - FINAL REPORT,” 2010.
- [20] N. G. Transmission, “Conversion Factors Technical Guidance.”
- [21] N. G. Transmission, “LinePack (Fixed temp & Z).xlsx.”
- [22] N. G. Transmission, “NTS Shrinkage.”

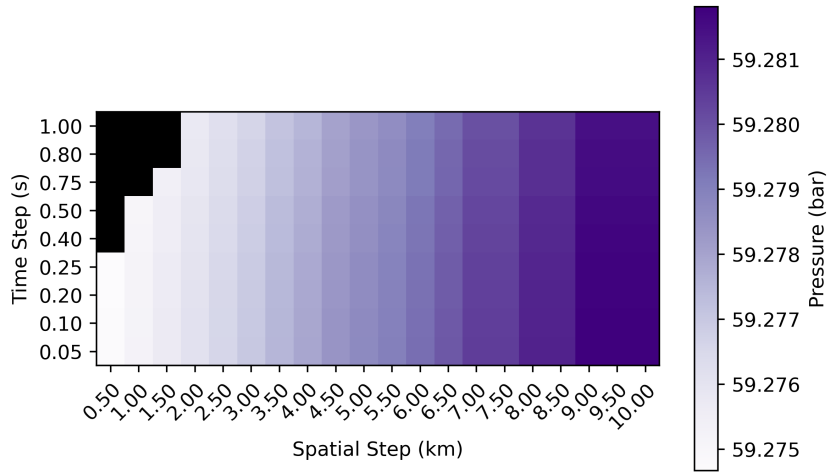
A Convergence Study

In order to determine an appropriate temporal and spatial resolution for the simulation a convergence study must be performed. The aim of this study is to remove control volume size and timestep size dependence from the simulation solution by choice of an appropriate timestep and spatial step. Due to the relationship between flow conditions, timestep size and control volume size a test case representative of the conditions within the SCADA test case must be used. For this reason the [REDACTED] Test Case (Section 4.2.5) was executed with SCADA data spanning the period between the start of the gas day on the 5th of January 2024 and the start of the gas day on the 6th of January 2024. The simulation is run for a total duration of 24 hours.

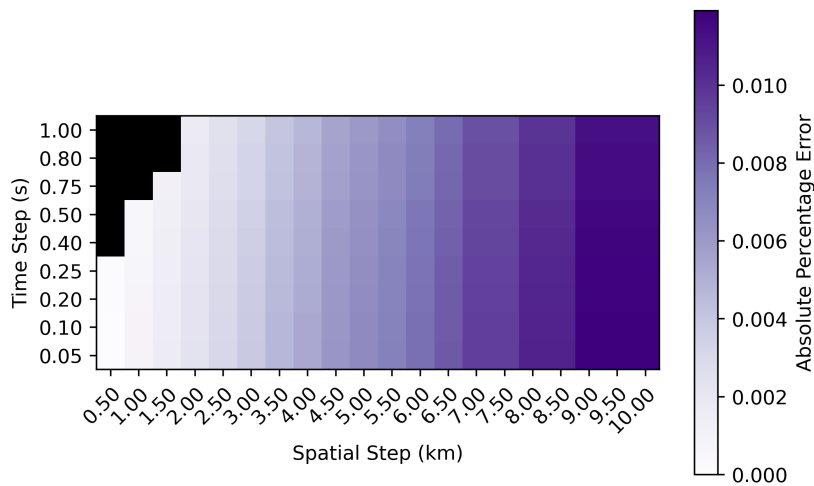
To execute this study, the timestep size used in the simulation, dt , was varied between 0.05 and 1 seconds alongside the variation of the control volume size, dx , between 0.5 and 10 kilometres in increments of 0.5 kilometres. Due to the stability criteria imposed by an explicit timestepping scheme, a number of samples become unstable as some combinations of time and spatial steps do not meet the stability criterion.

In order to compare the outcomes of the simulation and thus determine convergence, a metric of the simulation at a fixed time was chosen and compared to the value obtained when using the smallest timestep and smallest control volume size. Specifically, the pressure and flow at the end of the simulation for a selection of sites were used as the metric to determine convergence. As this study's focus is on model convergence, rather than model's adequacy to capture the underlying physics, a comparison to the SCADA data was not required.

Figures 38–41 show the variation in the pressure and flow results at [REDACTED] for each sample. It is clear from these plots that the variation in simulation results is minimal across all the temporal and spatial steps considered. The maximum percentage variation in flow at [REDACTED] of 2% and 0.05% respectively. Similarly, pressure variation across the samples is found to be $1 \times 10^{-6}\%$ at [REDACTED] and 0.01% at [REDACTED]. Therefore, when determining an appropriate spatial and temporal resolution, simulation execution time becomes the most important factor. However, due to the variation of pipeline length within the network, and each pipeline requiring at least three control volumes, the largest stable dt and dx for each test case is chosen. Specifically, a time step of 0.2 s and a spatial resolution of 1 km were chosen. This is to ensure the fastest possible and stable simulation execution.

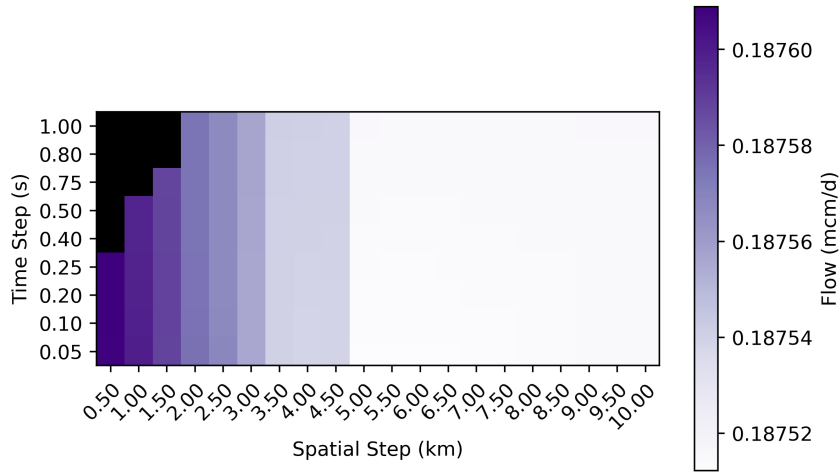


(a) Pressure value at simulation end for each sample.

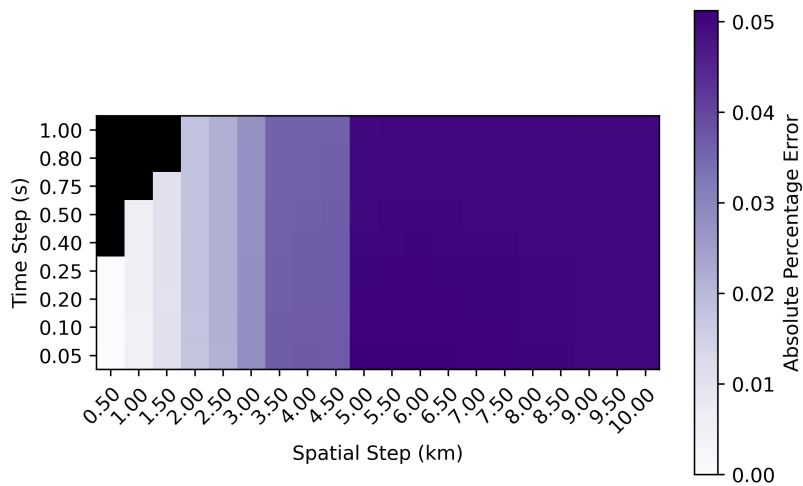


(b) Pressure absolute percentage error at simulation end for each sample.

Figure 38: Pressure results at the end of the simulation at [REDACTED] for each sample within the convergence study. Percentage error is calculated against the results of the simulation with the smallest Δt and Δx combination. Black squares denote an unstable set of simulation parameters.

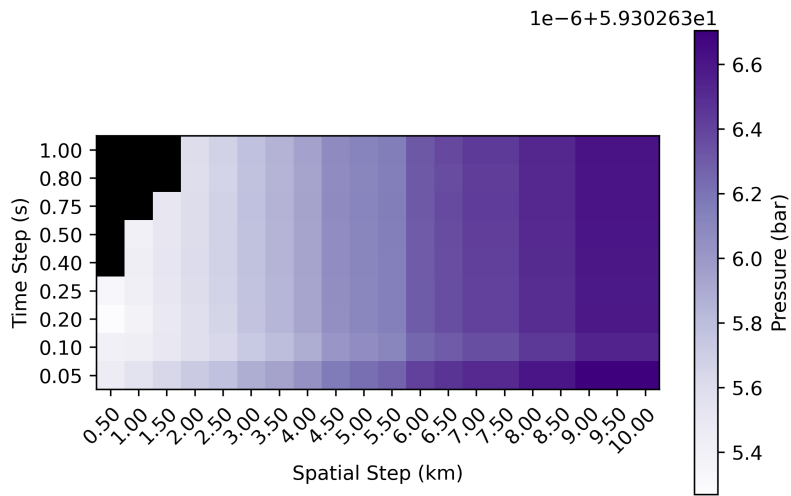


(a) Flow value at simulation end for each sample.

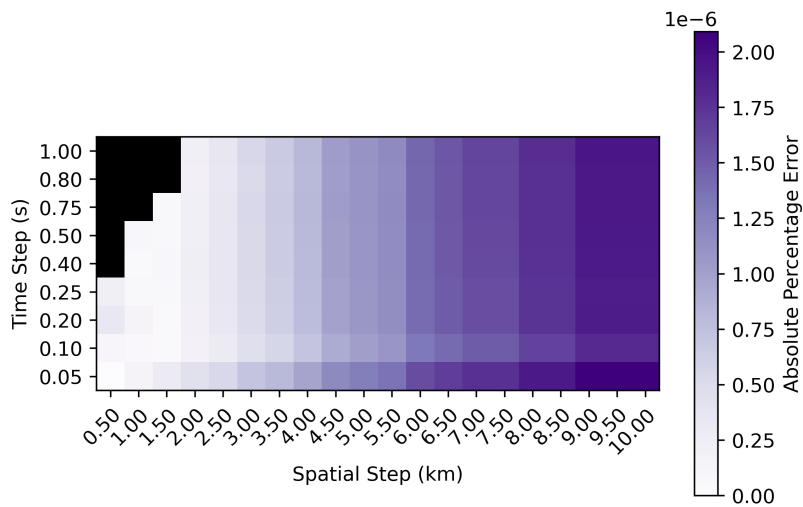


(b) Flow absolute percentage error at simulation end for each sample.

Figure 39: Flow results at the end of the simulation at [REDACTED] for each sample within the convergence study. Percentage error is calculated against the results of the simulation with the smallest dt and dx combination. Black squares denote an unstable set of simulation parameters.

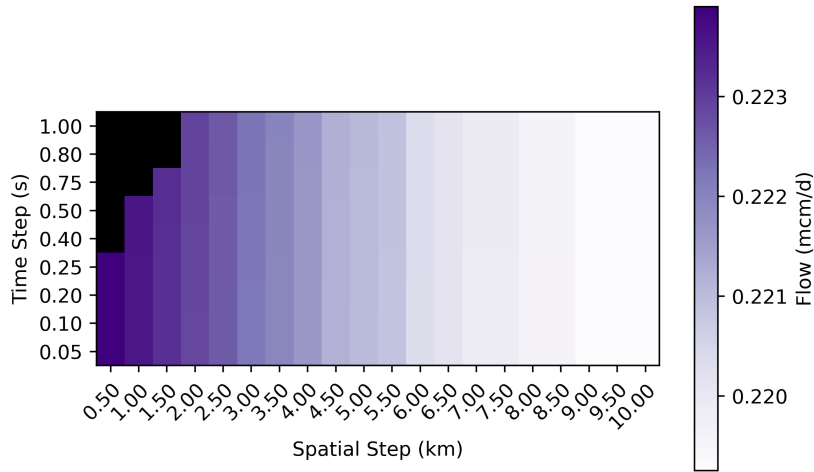


(a) Pressure value at simulation end for each sample.

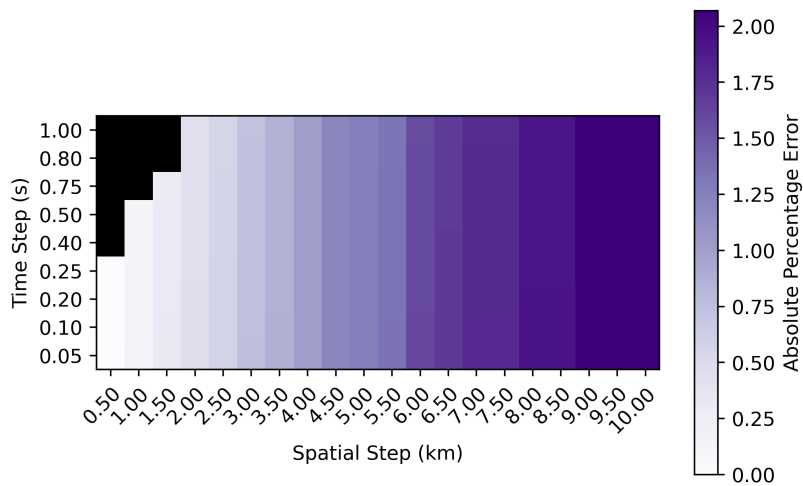


(b) Pressure absolute percentage error at simulation end for each sample.

Figure 40: Pressure results at the end of the simulation at [REDACTED] for each sample within the convergence study. Percentage error is calculated against the results of the simulation with the smallest dt and dx combination. Black squares denote an unstable set of simulation parameters.



(a) Flow value at simulation end for each sample.



(b) Flow absolute percentage error at simulation end for each sample.

Figure 41: Flow results at the end of the simulation at [REDACTED] for each sample within the convergence study. Percentage error is calculated against the results of the simulation with the smallest Δt and Δx combination. Black squares denote an unstable set of simulation parameters.