ALLOCATION OF UNIDENTIFIED GAS

Proposed Allocation of Unidentified Gas Statement for 2020/21

Report No.: 1, Rev. 1.0
Date: 24 December 2019

Project name:	Allocation of Unidentified Gas	DNV GL - Software
Report title:	Proposed Allocation of Unidentified Gas Statement for 2020/21	Software Consulting
Customer:		Holywell Park
Customer contact:	Fiona Cottam	Ashby Road
Date of issue:	24 December 2019	Loughborough
Project No.:	10161618	LE11 3GR
Organisation unit:	Software	Email:AUGE.software@dnvgl.com
Report No.:	1, Rev. 1.0	

Objective: This document is the proposed (first draft) AUG Statement for 2020/21 and contains details of the methods developed by the AUG Expert for allocating daily UIG between Product Class and End User category (EUC) including details of the data requested to support this analysis, data received following such requests and any assumptions made.

This document also includes an estimate of the total quantity of permanent Unidentified Gas and identifies each Unidentified Gas source with its proportionate contribution to the total.

This document contains a first estimate of the AUG weighting factors based on the proposed methodology. The factors are based on the latest available data and will change when updated data is used.

Prepared by:	Verified by:	Approved by:
Tony Perchard Principal Consultant	Clive Whitehand Head of Section	Anthony Gilbert Global Delivery Manager
Andy Gordon Senior Consultant		
Mark Lingham Group Leader		

DNV GL Distribution:

☒ Unrestricted distribution (internal and external)
☐ Unrestricted distribution within DNV GL Group
☐ Unrestricted distribution within DNV GL contracting party
☐ No distribution (confidential)

Keywords:
Unidentified Gas, AUGE, AUGS

Rev. No.	Date	Reason for Issue	Prepared by	Verified by	Approved by
0.1	2019-12-13	Draft for CDSP Review	Mark Lingham, Andy Gordon & Tony Perchard	Clive Whitehand	
0.2	2019-12-19	Updated draft	Mark Lingham, Andy Gordon & Tony Perchard	Clive Whitehand	
0.3	2019-12-20	Draft for final review	Mark Lingham, Andy Gordon & Tony Perchard	Clive Whitehand	
1.0	2019-12-24	Proposed AUGS for consultation	Mark Lingham, Andy Gordon & Tony Perchard	Clive Whitehand	

EXECUTIVE SUMMARY

Project Nexus was implemented in June 2017, replacing key IT systems for gas settlement and supply point administration in the gas industry ('UKLink') and changed the way that the gas settlement is handled. Project Nexus introduced individual meter point reconciliation for all meter points including those on CSEPs and a rolling monthly AQ process. It also introduces 4 new meter point 'classes'.

As a result of these changes, there is a requirement to fairly apportion the daily total UIG estimate (daily imbalance calculated from the settlement process) between Product Classes and EUC. Mod 0473 was raised to allow the appointment of an independent expert (AUG Expert) to develop a methodology to do this and provide a table of weighting factors that will target the correct amount of UIG to different classes of meter points, based on an assessment of their relative contribution. The table of weighting factors will be used in the daily gas nomination, allocation and reconciliation processes.

This document is the proposed AUG Statement for 2020/21 and contains details of the methods developed by the AUG Expert for allocating daily UIG between Product Class/EUC including details of the data requested to support this analysis, data received following such requests and any assumptions made. This document also describes how the AUG Expert has followed the published guidelines and provides the following additional supporting information

a) identifies each Unidentified Gas Source

b) estimates the total quantity of permanent Unidentified Gas

c) specifies the proportionate contribution of each Unidentified Gas Source to the total quantity of Unidentified Gas

d) provides a first estimate of the AUG weighting factors (based on the latest available data at the time)

Following publication of this proposed AUG Statement there will be a 21-day consultation period allowing the industry to provide feedback to the AUG Expert and raise any questions/issues. Where appropriate, the methodology will be updated based on this feedback.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 Background

Most of the gas consumed in Great Britain is metered and registered. However, some gas is lost from the system, or not registered, due to theft, leakage from gas pipes, consumption by unregistered supply points and other reasons. Some elements of the gas that is not directly consumed/measured are currently modelled, and hence the gas consumed by these can be estimated. The gas that is lost and not recorded or modelled is referred to as Unidentified Gas.

The Great Britain gas industry is segmented into two market sectors; Larger Supply Points (LSP) and Smaller Supply Points (SSP). Prior to April 2012 there was no methodology in place to determine the allocation of Unidentified Gas between the LSP and SSP market sectors: Unidentified Gas was ultimately borne by the SSP market sector following reconciliation (an interim amount was allocated for 2011/12). Through the approval of UNC Modification 0229 and the appointment of DNV GL as the Allocation of Unidentified Gas Expert (AUGE) in 2011, a methodology was developed to calculate and apportion Unidentified Gas equitably to the relevant gas market sectors. This approach involved an annual estimate of Unidentified Gas and a monthly transfer of costs between market sectors to address the misallocation of Unidentified Gas that occurred under that regime. DNV GL carried out this annual process until 2014 when it became clear that the requirements for the AUGE would change given the upcoming implementation of project 'Nexus' (Mod 0432).

Project Nexus involved the replacement of key IT systems ('UKLink') for gas settlement and supply point administration in the gas industry. It introduced individual meter point reconciliation for all meter points (previously SSP meters were subject to reconciliation by difference) and a rolling monthly AQ process. It also introduced four new meter point 'classes'. Following Project Nexus implementation, an amended NDM Algorithm (with scaling factor removed) uses actual weather data to derive a bottom-up estimate of NDM Demand. This allows the calculation of a daily Balancing Factor, referred to as UIG. This UIG is shared out to all live sites, on the basis of the AUG weighting factors and their recorded/estimated throughput for the day.

Note that this quantity is not the same as the Unidentified Gas figure calculated previously by the AUGE. UIG is calculated 'live' on a daily basis using estimated NDM demand based on prevailing AQs and the NDM algorithm; the AUG Expert estimate of Unidentified Gas was calculated retrospectively (and then projected forward) on an annual basis using estimated NDM demand based on meter reads/volumes for approximately 86% of sites with the remaining 14% based on average EUC demands or interpolation.

As a result of these changes, the industry noted a requirement to be able to fairly apportion this total UIG between Product Classes and End User Category (EUC). Mod 0473 was raised to allow the appointment of an independent expert (the AUG Expert) to develop a methodology to do this and provide a table of weighting factors that will assign the correct amount of UIG to different classes of meter points, based on an assessment of their relative contribution to Unidentified Gas. This table of weighting factors is used in daily gas allocation processes. Daily measured or estimated gas throughput in each sector are weighted using the AUG table factors to assign daily UIG to Shippers based on their throughput by meter point class and EUC.

DNV GL were appointed to the new role of AUG Expert in July 2016, with Project Nexus go-live occurring on 1st June 2017.

1.2 High Level Objectives

The AUG Expert's high-level objectives are:

- To assess the sources of Unidentified Gas and the data available/required from industry bodies to evaluate Unidentified Gas
- To gather data as required from the CDSP, from Gas Shippers, or drawn from other sources, as deemed appropriate by the AUG Expert
- To develop the methodology to assess the relative contribution to Unidentified Gas of different Classes and sizes of sites
- To publish the methodology in the AUG Statement (this document) and present findings to the industry
- To consult with the industry bodies and respond to questions/issues raised, assess the impact of questions on the methodology and update as appropriate
- Produce the table detailing the weighting factors for each Product Class and EUC

1.3 Scope

This document is the Proposed AUG Statement for 2020/21 and contains the following:

- A detailed description of the proposed methodology
- A summary of data requested, received and used, and associated assumptions
- The table of weighting factors for apportioning UIG between Product Classes and EUCs
- Details of the database used to hold information associated with Unidentified Gas and used to develop the methodology

This document will be published to the industry for review and comment during the consultation period.

The following are out of scope but may be referenced where they may impact on the methodology to calculate the AUG factors.

- The AUG Expert is not concerned with issues regarding the deeming algorithm or the former RbD mechanism.
- The AUG Expert is not concerned with resolution of fundamental gas industry business process issues.
- The AUG Expert process is not an opportunity to provide permanent solutions to issues within the gas industry that should be addressed by other workgroups (e.g. Shrinkage Forum). This includes Shrinkage Error (i.e. bias in the Shrinkage and Leakage Model), although it is acknowledged that any such errors feed through into the daily UIG calculation.
- The AUG Expert is not concerned with transportation charges.

2 COMPLIANCE TO GENERIC TERMS OF REFERENCE

This section describes how DNV GL has adhered to the Generic Terms of Reference described in Section 5 of the AUGE Framework[43].

The AUG Expert will create the AUG Statement and AUG Table by developing appropriate, detailed methodologies and collecting necessary data.

The AUG Expert has developed a detailed methodology for estimating factors to apportion UIG between EUC and Product Classes. To calculate the factors, total Unidentified Gas is also estimated using meter read and consumption data for all meters, which has been obtained from the CDSP. Further detailed datasets are used to directly estimate some components of the total Unidentified Gas where this is possible e.g. Shipperless sites. The AUG Expert has also developed a methodology to account for elements of Unidentified Gas which are Temporary in nature.

Additional data regarding theft of gas was sought and a subset of the requested TRAS data was provided. Two industry data requests were also issued, one relating to meter maintenance and volume converter capabilities, the other relating to internal building temperatures.

The decision as to the most appropriate methodologies and data will rest solely with the AUG Expert taking account of any issues raised during the development and compilation of the AUG Statement and AUG Table.

For the avoidance of doubt although UIG includes any LDZ Shrinkage Error, the AUG Expert acknowledges that the process for determining LDZ Shrinkage is laid out in the relevant DNO licences. To avoid dual governance of any LDZ Shrinkage Error, the AUG Expert's role in respect of any LDZ Shrinkage Error is therefore limited to confirming that there are controls in place to ensure that DNOs discharge their licence obligation (that is that there is a methodology and that it is periodically reviewed for confirmation that the methodology remains relevant). The AUGE will present any comments or observations on the LDZ Shrinkage model through the annual consultation carried out by the DNOs.

The proposed methodology and assessment of what constitutes Unidentified Gas has been decided solely by the AUG Expert based on available information. Comments raised by shippers will be considered and additional analysis undertaken where necessary. Having considered all views, the final decisions are the AUG Expert's own.

LDZ Shrinkage Error is not considered by the AUG Expert. However, the AUG Expert has assessed IGT Shrinkage as this is not included within the current LDZ Shrinkage models.

The AUG Expert will determine what data is required from Code Parties (and other parties as appropriate) in order to ensure appropriate data supports the evaluation of Unidentified Gas.

The AUG Expert has assessed what data is required to support the proposed methodology and has requested information from relevant parties. Updated data sets have been requested from the CDSP together with any relevant new datasets available following Nexus implementation.

The AUG Expert will determine what data is available from parties in order to ensure it has appropriate data to support the evaluation of Unidentified Gas.

The AUG Expert has determined what data is available following discussions with the CDSP. A request was also made to SPAA for further information regarding theft. Information has also been provided by ICoSS on AMR uptake not included within the BEIS dataset used by the AUG Expert.

Two industry data requests were also issued, one relating to meter maintenance and volume converter capabilities, the other relating to internal building temperatures.

The AUG Expert will determine what relevant questions should be submitted to Code Parties in order to ensure appropriate methodologies and data are used in the evaluation of unidentified error.

The AUG Expert issued two industry data requests and raised a number of questions at the AUG Sub-committee Early Engagement Meeting. Further communication will take place as and when necessary.

The AUG Expert will use the latest data available where appropriate.

Most of the data requested by the AUG Expert has been provided, and the latest data has been used where possible. Further data updates will be used in future re-calculations of the table of factors.

Where multiple data sources exist the AUG Expert will evaluate the data to obtain the most statistically sound solution, will document the alternative options and provide an explanation for its decision.

For the Consumption Method of estimating total Unidentified Gas, both meter reads and metered volumes are provided. Over time LSP metered volumes may be corrected, but the meter reads are not. The AUG Expert's analysis has shown that metered volumes can be erroneous, particularly for non-corrected SSP data. The decision was therefore taken to use meter reads for SSP and metered volumes for LSP. This approach will be reviewed in future, as post-Nexus it may be more appropriate to treat LSP and SSP meters the same.

Where data is open to interpretation the AUG Expert will evaluate the most appropriate methodology and provide an explanation for the use of this methodology.

Throughout the statement the AUG Expert has described how data will be used and why e.g. the explanation of the availability of gas temperature data and the decision of which to use.

Where the AUG Expert considers using data collected or derived through the use of sampling techniques, then the AUG Expert will consider the most appropriate sampling technique and/or the viability of the sampling technique used.

The Consumption Method for estimating the Unidentified Gas total is one part of the analysis where a sample rather than the full dataset is used. This calculation will be at its most accurate when the largest possible representative subset of the meter point population is used. To achieve this, a validation process was developed that was designed to maximise the sample size whilst removing any meter points with invalid data. Appropriate methods are then applied to scale up for any meter points which have been excluded.

The gas temperature data used is also based upon a sample. Unfortunately, the AUGE does not have access to the full dataset which limits the statistical analysis which can be performed. However, the AUGE has taken into account sample sizes and confidence intervals when deciding the most appropriate data to use.

The AUG Expert will present the AUG Statement, including the AUG Table, in draft form (the "proposed AUG Statement"), to Code Parties seeking views and will review all the issues identified submitted in response.

The AUG Expert will present the AUG Statement at the AUG Sub-committee meeting to be held on 10th Jan 2020. Code parties can submit any feedback and issues during the 21-day consultation period.

The AUG Expert will then document and review all feedback as in previous years. Section 9 of this document refers to the publication of feedback from previous consultations with details of the issues raised, with the full text of the comments from the Code Parties and the AUG Expert responses contained in separate documents published on the Joint Office of Gas Transporters website.

The AUG Expert will provide the proposed and any modified AUG Statement, including the AUG Table, to the Gas Transporters for publication who will then provide the AUG Statement and Table to the CDSP.

This Proposed AUG Statement for 2020/21 is provided to the Gas Transporters in advance, for publication on 1st January 2020.

The AUG Expert will ensure that all data that is provided to it by parties will not be passed on to any other organisation, or used for any purpose other than the creation of the methodology and the AUG Statement and Table.

On receipt of data, the AUG Expert stores the data in a secure project storage area with limited access by the consultants working on the project. The AUG Expert can confirm data used in the analysis has not and will not be passed on to any other organisation. The data used will be made available to all bona fide industry participants in order to review the methodology, and in this dataset all MPR information has been replaced by 'dummy' MPR references by the CDSP so that the anonymity of the consumer is protected.

The AUG Expert shall ensure that all data provided by Code Parties will be held confidentially, and where any data, as provided or derived from that provided, is published then it shall be in a form where the source of the information cannot be reasonably ascertained.

Data is stored in a secure project storage area with access limited to those working on the project. Any data that contains market share or code party specific information has been and will be made anonymous to ensure the source of the information cannot be ascertained.

The AUG Expert will act with all due skill, care and diligence when performing of its duties as the AUG Expert and shall be impartial when undertaking the function of the AUG Expert, ensuring that any values derived will be equitable in their treatment of Code Parties.

When developing the methodology, the AUGE does not consider the implications of the weighting factors generated from a Code Party impact perspective i.e. it is an independent, evidence based approach. This is noted by OFGEM when describing the role of the AUGE [52].

The AUG Expert will compile the methodology and AUG Statement and AUG Table in accordance with this Framework.

The AUGE has, following the revision of the AUGE Framework in 2018[43], modified its approach to the development of the methodology. This has included early engagement meetings with Code Parties and more regular progress updates to ensure the industry are fully up to date with progress. All activities will be based on the latest AUGE Framework.

3 TERMINOLOGY

This section aims to provide clarity surrounding the use of terminology relating to Unidentified Gas.

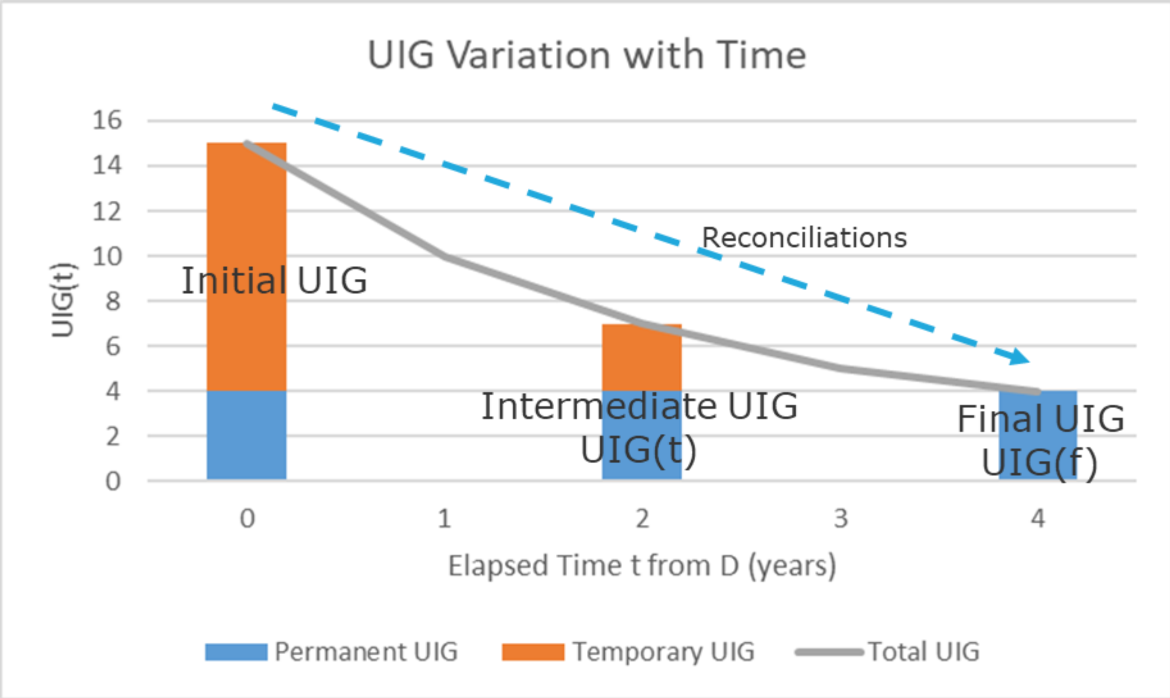


Figure 1: Unidentified Gas Definitions

Unidentified Gas refers to gas which is consumed in an unrecorded manner at or beyond the meter. Estimates of Unidentified Gas for a given gas day (D) are initially made on the day as part of the settlement process. This "initial" value of Unidentified Gas is defined in UNC Section H 2.6.1 as UIG. It is calculated by taking the total LDZ input and subtracting all the LDZ outputs (either known values or estimates).

At the time of calculating initial UIG, most meter reads are not available. This initial value therefore includes an element relating to the uncertainty in the estimate of demand (model error). Ideally this error would be random and have a zero bias, but this is dependent on the estimation algorithm and its inputs being unbiased. Over time, as meter reads are received, this part of UIG will reduce towards zero as shown in Figure 1. This diagram is indicative only to explain how UIG varies over time. The levels of UIG shown do not represent true values of UIG. Although the diagram shows UIG reducing over time, UIG can also increase.

The model error described above is considered as temporary Unidentified Gas as it is corrected for over time through reconciliation. There are further sources of Unidentified Gas which are also temporary in nature e.g. LDZ metering errors which are identified and corrected for. UIG can therefore be recalculated at any point in time after the initial gas day taking into account this updated information. A Code Cut Off Date or 'line in the sand' is defined in UNC, beyond which no further reconciliations will be applied. Once a gas day falls beyond this date the value of UIG at that point is referred to as Final UIG or UIG(f) and consists of only permanent Unidentified Gas.

Intermediate UIG or UIG(t) refers to the value of UIG calculated at some point in time (t) between D and the line in the sand. This value will contain a mix of permanent and temporary Unidentified Gas.

Throughout this document, the term Unidentified Gas will refer to the general concept, i.e. any gas burnt in an unrecorded manner. Where the reference needs to be more specific, the following terms will be used:

- UIG – Initial UIG as defined in UNC Section H 2.6.1
- Intermediate UIG or UIG(t) – UIG recalculated at some point in time, t between D (initial allocation) and the line in the sand
- Final UIG or UIG(f) – UIG as at line in the sand

The methodology described in this statement aims to provide an estimate of the Final UIG and provide factors to apportion this between EUC and product classes.

4 HIGH LEVEL OVERVIEW OF METHODOLOGY

This section provides a high-level overview of the methodology. For each of the areas of Unidentified Gas presented here a more detailed discussion is given in Section 7.

Under the pre-Nexus settlement regime, an independent forecast of permanent Unidentified Gas split by market sector was made to allow correct allocation of Unidentified Gas costs. Following implementation of Project Nexus, UIG is calculated daily as part of the settlement process. This UIG amount is then apportioned based on the weighting factors calculated by the AUG Expert using the methodology described in this document.

As part of the AUG methodology, the AUG Expert makes an independent estimate of Final UIG. The Initial UIG value calculated by the CDSP includes elements which are temporary in nature and are removed over time through the reconciliation process. The Final UIG calculated by the AUG Expert to derive the AUG weighting factors is an estimate of the value of UIG following all reconciliations.

4.1 LDZ Load Components

The Unidentified Gas calculations for the 2017/18 gas year were complicated by the fact that the UIG Factors, which are the ultimate output of the work, are required to be split by post-Nexus market sector definitions, whilst the available data was from the pre-Nexus regime with its associated market sectors. The load components are different in these two scenarios, and this created a requirement to map from one to the other in as accurate a manner as possible as part of the calculation process.

This is no longer the case for the 2020/21 analysis, for which actual Product Class data is available. For completeness, and to provide a link with the historic Unidentified Gas analysis, both pre-Nexus and post-Nexus regimes are described below.

4.1.1 Pre-Nexus Regime

Daily load (as measured or calculated at the Supply Meter Point) fell into three relevant categories as far as the reconciliation process was concerned. These are as defined in Section A of the Uniform Network Code (UNC) [2]:

1. **Smaller Supply Point Component Load**

   Load from Supply Point Components (SPCs) which are part of a Smaller Supply Point (SSP). This is defined as a supply point where the AQ is not greater than 73,200 kWh.

2. **Larger Non-Daily Metered Supply Point Component Load**

   Load from Non-Daily Metered (NDM) SPCs which were part of a Larger Supply Point (LSP). This is defined as a supply point where the AQ is greater than 73,200 kWh but less than the mandatory daily metering threshold of 58,600,000 kWh. Note that historically (prior to the implementation of Mod 0428), Larger NDM SPCs may have contained individual meters that fell below the SSP AQ threshold.

3. **Daily Metered Supply Point Component (DM SPC) Load**

   Load from Daily Metered (DM) SPCs. This includes Daily Metered Mandatory (DMM) sites, which are those above the 58,600,000 kWh threshold, Daily Metered Voluntary (DMV) and Daily Metered Elective (DME) sites.

4.1.2 Post-Nexus Regime

Following Project Nexus implementation, the population of supply points is instead split into four different Product Classes, each of which have different meter read frequency requirements and reconciliation rules. A list of products and associated details (including approximate equivalence to historic services) is shown in Table 1 below. Information in this table is taken from UNC Modification 0432 [3].

Process Description	**Basis of Energy Allocation**	**Basis of Energy Balancing**	**Shipper Read Submission**	**Market Sector**	**Historic Service Equivalent**
Product Class 1: Daily Metered Time Critical Readings	Daily Read	Daily Read	Daily by 11 am on GFD+1	DM	DM Mandatory
Product Class 2: Daily Metered not Time Critical Readings	Daily Read	Daily Read	Daily by end of GFD+1	DM	DM Voluntary / DM Elective
Product Class 3: Batched Daily Readings	Allocation Profiles	Allocation Profiles	Periodically in batches of daily readings	NDM	Non-Daily Metered
Product Class 4: Periodic Readings	Allocation Profiles	Allocation Profiles	Periodically	NDM	Non-Daily Metered

Table 1: Product Classes

Each site is classified as subscribing to one of these product classes, and meter read submissions, settlement and reconciliation is then carried out for each site in the manner suitable for its class.

In addition to splitting the UIG figure between the four products, Mod 0473 also includes a requirement to include a split between EUCs [4]. The final output of the AUG process will therefore be a table of UIG allocation adjustment factors with the following structure:

Supply Meter Point Classification	**Unidentified Gas Weighting Factor**			
	Class 1	**Class 2**	**Class 3**	**Class 4**
EUC Band 1				
EUC Band 2				
EUC Band 3				
EUC Band 4				
EUC Band 5				
EUC Band 6				
EUC Band 7				
EUC Band 8				
EUC Band 9				

Table 2: Example AUG Table

4.1.3 Unidentified Gas

Daily Metered load (i.e. Product Classes 1 and 2) is, by definition, metered and known on an ongoing daily basis. Like all metered load it can be subject to metering error, and data for known errors is used to correct it. NDM load (Product Classes 3 and 4) for a given day is estimated using the NDM deeming algorithm, to which reconciliation is applied when the meter readings become available. The estimation process is described in Mod 0432 for the post-Nexus regime [3].

Even after reconciliation and meter corrections are applied, the sum of these load components does not equal the gas intake into the LDZ due to the presence of two further factors:

1. Shrinkage

LDZ shrinkage occurs between the LDZ offtake and the end consumer (but not at the Supply Meter Point - the LDZ shrinkage zone stops immediately before this point). It covers:

- Leakage (from pipelines, services, AGIs and interference damage)
- Own Use Gas
- Transporter-responsible theft

The majority of shrinkage is due to leakage, and the overall LDZ shrinkage quantity is calculated using the standard method defined in the UNC [2].

2. Unidentified Gas

Unidentified Gas occurs downstream of shrinkage, i.e. at the Supply Meter Point. It potentially covers:

- Unregistered and Shipperless sites
- Independent Gas Transporter Connected System Exit Point (IGT CSEP) setup and registration delays
- Errors in the shrinkage estimate
- Shipper-responsible theft
- Meter errors – this includes both LDZ offtakes and consumer meters
- Volume conversion errors

Unidentified Gas is currently unknown and hence must be estimated.

The relationship between these components of daily load can therefore be expressed as follows:

$$\textit{Total Unidentified Gas} = \textit{Aggregate LDZ Load} - \textit{Product 1 to 4 Load} - \textit{Shrinkage} \qquad (4.1)$$

4.2 Permanent and Temporary Unidentified Gas

Unidentified gas can be divided into two categories:

Permanent Unidentified Gas is consumed in an unrecorded fashion and costs are never recovered e.g. undetected theft.

Temporary Unidentified Gas is initially consumed in an unrecorded fashion, but volumes are later calculated directly or estimated, and the cost is recovered via back billing or reconciliation. This includes:

- Errors in DM meter reads which are later corrected
- Errors in estimated NDM consumption which are later corrected when new meter reads are available

4.3 Unidentified Gas Methodology

4.3.1 Estimation of Total Unidentified Gas for Historic Years

The overall concept of calculating total Unidentified Gas using metered consumption data is simple and is centred on the basic principle of the allocation process. Note that the historic consumption data used by the AUG Expert covers both the pre- and post-Nexus periods, and hence the calculation is designed to be applicable to both regimes. The method outputs an aggregate national figure for Unidentified Gas, and this output can be extrapolated to the 2020/21 gas year for which the factors are being produced. Pre-Nexus, the NDM Allocation can be written as follows:

$$\textit{NDM Allocation} = \textit{Aggregate LDZ Load} - \textit{DM Load} - \textit{Shrinkage}$$

Post-Nexus, the Scaling Factor (SF) used during allocation has been removed such that

$$\textit{NDM Allocation} = \textit{Aggregate LDZ Load} - \textit{PC1 Load} - \textit{Shrinkage} - \textit{UIG}$$

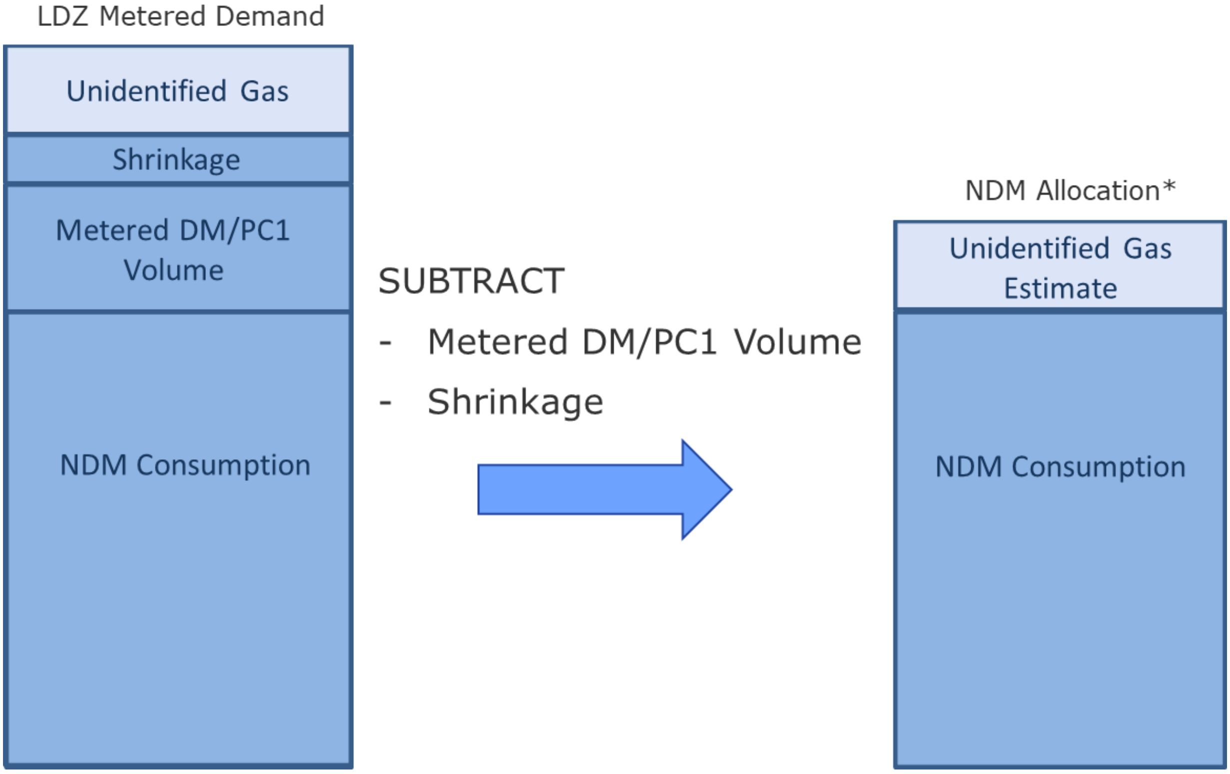
This is shown graphically in Figure 2.

As the NDM load is the sum of all metered NDM consumptions, Total Unidentified Gas can be calculated as

$$\textit{Total Unidentified Gas} = \textit{NDM Allocation}^{*} - \textit{Metered NDM Consumption} \quad (4.2)$$

** denotes the pre-Nexus equivalent NDM Allocation i.e. post-Nexus allocations have the daily UIG value added back in*

This formulation of Unidentified Gas means that the total will include any Permanent Unidentified Gas arising from unresolved errors in aggregate LDZ load, DM/PC1 load and shrinkage.



** denotes the pre-Nexus equivalent NDM Allocation i.e. post-Nexus allocations have the daily UIG value added back in*

Figure 2: NDM Allocation and Unidentified Gas

The metered NDM consumption is calculated for each meter point and gas year using meter reads or metered volumes, and meter asset information. There are several complexities in this calculation that must be accounted for in the methodology and a fall-back approach must be developed for those meter points without sufficient data of suitable quality. This is summarised in Figure 3. The full details of the Consumption Methodology can be found in Section 7.2.

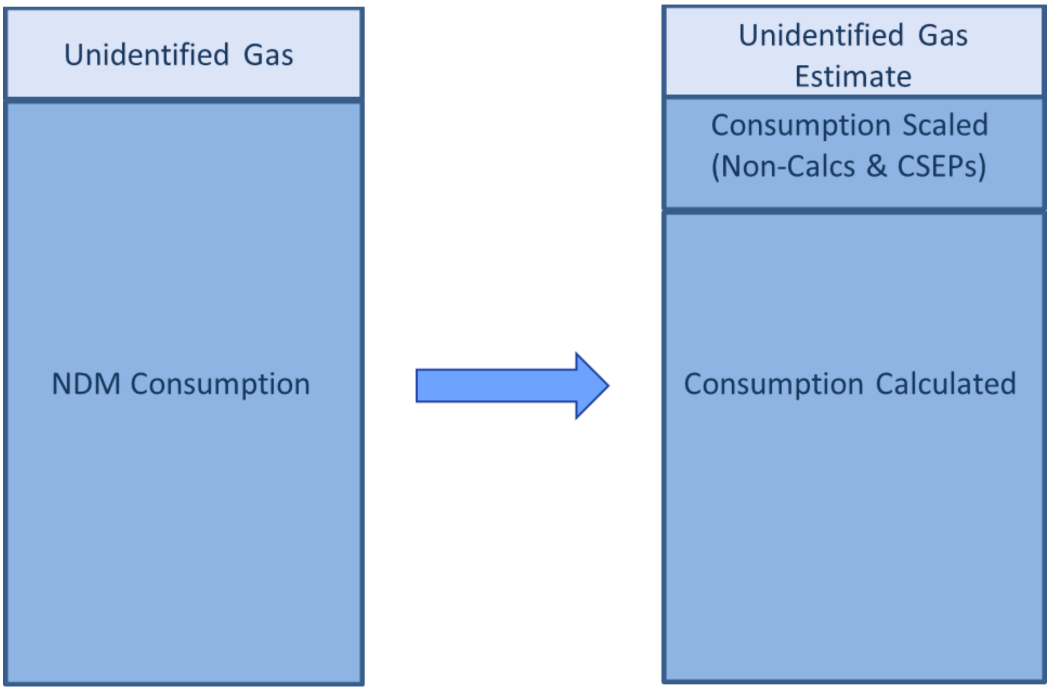


Figure 3: Calculation of Unidentified Gas from Consumptions and Allocations

This method is used to estimate total Unidentified Gas on a gas year basis. This is initially done on an LDZ by LDZ basis due to the very high volume of data required (i.e. all meter reads for all sites). The AUG Expert then determines which years are most suitable for use in the process to estimate the total Unidentified Gas. Recent data is excluded as the number of meters where consumption can be successfully calculated is much lower due to fewer meter reads being available. The use of this data is therefore subject to a large degree of uncertainty. Older data may also be excluded as this is less relevant and may have data quality issues.

Note that at this stage the total Unidentified Gas figure contains both Permanent and Temporary Unidentified Gas.

4.3.2 Calculating Components of Total Unidentified Gas

Having obtained the total Unidentified Gas figure using the Consumption Methodology described above, the value of individual components that make up the Unidentified Gas total are calculated where this is possible. This also includes the calculation of the amount of this Unidentified Gas which is temporary for each component and how the Unidentified Gas is split between market sectors. "Market sector" is now defined in the post-Nexus terms of Product Class and EUC, even though the training period still lies mostly in the pre-Nexus period. This approach therefore requires a method of applying post-Nexus Product Class information to the pre-Nexus period, and this is defined in Section 4.3.3 below. Splitting each directly calculated source of Unidentified Gas by EUC/Product in this way allows calculations for both the historic period (now a mix of both pre-Nexus and post-Nexus data) and forecast period (post-Nexus) to be carried out as follows:

1. Pre-Nexus data is required as a total for each Unidentified Gas component (but split into permanent and temporary elements). As described below, however, the Product/EUC split for this time period is still required for training and extrapolating values for the forecast year, and hence the calculations are still carried out on this basis even for pre-Nexus years. The output that feeds directly into the calculations for the training period is the sum across the categories and so the final figures for the

training period are not dependent on the categorisation process used. It remains the most efficient approach to work with the Product/EUC split, however, because it is required for extrapolation to the forecast period anyway.

2. The post-Nexus Unidentified Gas forecast must be carried out separately not only for each Unidentified Gas component but every individual EUC/Product combination within each component. This is because each EUC/Product combination can follow its own trend over time (depending on market conditions, Mods that have been created to address individual issues, and so on). Therefore, each must be calculated across the whole historic period so that the trend can be identified and extrapolated to the forecast period. A full split into EUC/Product categories is required for this work.

It is known that data for each of the six potential components of Unidentified Gas (Unregistered and Shipperless sites, IGT errors, CSEP shrinkage, Shipper-responsible theft, metering errors and volume conversion errors) is available. The availability and quality of this data varies from component to component, and the AUG Expert has therefore identified the best method of calculating each Unidentified Gas component based on the quality of information available for that component.

Brief descriptions of each Unidentified Gas element are given below.

1. Unregistered and Shipperless Sites

The data available for this element consists of the details for every site that is either Shipperless or Unregistered at a given point in time. This point in time is the snapshot date, and snapshots are provided on a monthly basis, allowing the trends in each such Unidentified Gas category to be monitored over time. The details for each site include AQ, which allows each site to be assigned to the correct EUC and allows its gas usage whilst Unregistered/Shipperless to be estimated. Unregistered and Shipperless sites that contribute to Unidentified Gas are split into the following sub-categories:

- Shipper Activity
- Orphaned Sites
- Unregistered <12 Months
- Shipperless Passed to Shipper (PTS)
- Shipperless Shipper Specific Report (SSrP)
- Sites Awaiting GSR Visit

2. IGT CSEP Setup and Registration Delays

Gas consumed in an unrecorded manner due to IGT CSEP setup and registration delays is also included in the Unidentified Gas calculation. Unidentified Gas from this source is due to gas networks owned by IGTs but not present in the CDSP’s records and comes from Unregistered sites on known CSEPs. The data available for this analysis consists of the number and composition of these unknown projects (number of sites and AQ split by EUC), and the number and AQ of each Unregistered site associated with a known project. Unknown Projects data is again provided in monthly snapshots, allowing trends over time to be established.

3. Shrinkage Error

Shrinkage errors affect the total Unidentified Gas calculation in that estimated Shrinkage is deducted from the LDZ input total (along with DM load) to give the total NDM allocation from which metered load is then removed to calculate total Unidentified Gas. The Shrinkage estimate comes from the Shrinkage Model, and if this is biased it will affect the Unidentified Gas estimate.

Following a recent change to the AUGE Framework[43] the estimation of Shrinkage Error by the AUGE has been ruled out of scope. Therefore, the Shrinkage estimate provided by the GTs will be used without further adjustment.

CSEP Shrinkage is not included in the Shrinkage model. This is relatively small but non-zero and is not taken account of anywhere in the settlement process, and therefore filters through into the daily UIG figure. The value of CSEP Shrinkage must therefore be estimated and split by EUC and Product Class as appropriate.

4. Shipper-Responsible Theft

Whilst detailed information for investigations and confirmed theft is available, theft by its nature is often undetected. Undetected theft levels are very difficult to quantify accurately, and estimates from different sources vary widely, from 0.006% of throughput (based on detected theft only) to around 10%. As it is difficult to accurately estimate theft levels directly, undetected theft is calculated by subtraction once known levels of detected theft have been accounted for. Undetected theft is part of the Balancing Factor (see 7 below), and considered over time, it forms the vast majority of that figure.

5. Meter Errors

Meter errors affect Unidentified Gas in different ways depending on their source. Errors in LDZ offtake metering and DM supply metering affect the estimate of total NDM demand including Unidentified Gas for the training period, whilst NDM metering errors contribute to Unidentified Gas by affecting the metered total. Corrections are applied to LDZ offtakes, DM and unique site meters using detected error data supplied by the CDSP. In addition, the effects of consumer meters (all EUC/Product combinations) under- and over-reading due to operating at the extremes of their range are modelled and included in the calculations.

6. Volume Conversion Errors

The volume conversion errors referred to here are the result of using assumed static values for the atmospheric pressure (1013.25mbars) and temperature (12.2°C) at the meter. In reality, both quantities will vary daily. If the assumed values do not represent the true average (consumption weighted) annual values, permanent Unidentified Gas will result. Meters with volume conversion equipment will not be subject to this error. An estimate of the permanent Unidentified Gas can be made based upon

- Details of meters with volume conversion equipment
- Historical atmospheric pressure data
- Historical meter temperature data

7. **Balancing Factor**

The Balancing Factor is calculated by taking the difference between the calculated total Unidentified Gas and the sum of the directly estimated components. The Balancing Factor is comprised of Unidentified Gas elements that cannot be calculated directly because data is either unavailable or unreliable and is believed to be mostly undetected theft.

The Permanent component of total Unidentified Gas is then given by the sum of the Balancing Factor and the Permanent components of the directly calculated components (see Figure 4).

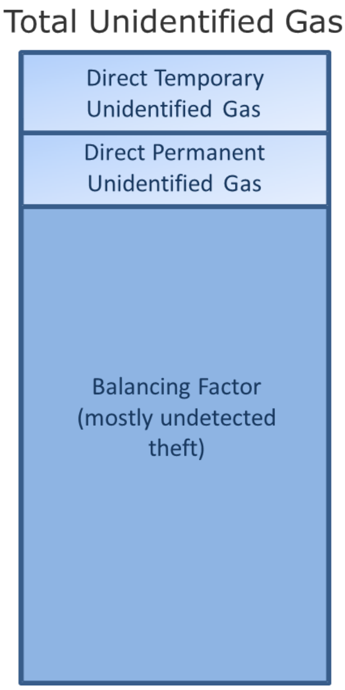


Figure 4: Unidentified Gas Components

4.3.3 Mapping to Post-Nexus Product Classes

The classification of Unidentified Gas into EUC/Product classes is referenced in several places in the descriptions above. The reason for having to create this split is described in Section 4.3.2. The need to do this creates a challenge, however, in that most of the training data comes from historic (i.e. pre-Nexus) years, during which time the Product Classes did not exist.

Therefore, a method is required for mapping sites as they were pre-Nexus to post-Nexus Product Classes. For the first AUG Statement of the new AUG period (i.e. the statement for 2017/18), the analysis took place before the implementation of Nexus and so all assignments of sites to Product Classes were handled using a rule set developed for this purpose. For subsequent years actual Product Class information is available for each site. This information can therefore be used to map sites to Product Classes individually, which provides far more accurate output. In some cases, this approach is not possible (e.g. Unregistered Sites, which by definition do not have a Product Class assigned): for these, the actual Product Class data can be used to create an approximate mapping that supersedes the rule sets developed for 2017/18 and provides a much higher level of accuracy.

The rules used to assign Product Classes to each site for each Unidentified Gas category are therefore as follows:

Detected Theft, Consumer Meter Errors, Volume Conversion Errors:
Map each site to its actual Product Class using the most recent data available.

Shipperless and Unregistered Sites, IGT CSEPs:
These sites do not have actual Product Classes recorded. Therefore, for each EUC, calculate the split by Product Class for all sites in the total population (for which data has been provided). This is done by number of sites to prevent unusual AQs from skewing the figures. For each EUC this creates a percentage split by Product Class. The next step is to use as many instances of this population dataset as are available to create a trend over time. Data currently exists for five points in time from Nexus go-live to the present. The calculated trend is then used to extrapolate the population split to the forecast year. Finally, this split is applied to the Unidentified Gas calculated as arising from each EUC to give the estimated Product Class split.

Balancing Factor:
The Balancing Factor is assumed to be largely composed of undetected theft, for which a direct Product Class mapping is not available. The calculation process for the UIG factors for the 2020/21 gas year contains the AUGE's new method of estimating the split of undetected theft by EUC and Product Class. This is based on historic data for theft investigations and confirmed thefts and includes a technique for removing the effect of Supplier theft detection activity from these figures. The output resulting from this removal of bias from the confirmed theft figures is a set of theft levels by EUC/Product Class based on population and propensity to steal only – i.e. one that is representative of the split of undetected theft.

Population is a key element of this calculation, and the data used is necessarily from the past. Theft levels will change over time in proportion with any change in population of relevant categories of meter point, and hence in order to provide an accurate split by EUC/Product Class for the forecast year, the outputs from the analysis of historic theft data must then be extrapolated to this year.

Theft levels from Smart Meters/AMR and traditional meters are different and so each of these populations must be extrapolated individually to the forecast year. Smart Meter installation rates and population totals are taken from the latest BEIS Quarterly Smart Meter Report [18], and these are used to calculate a best estimate of the Smart Meter populations for the midpoint of the forecast year based on the trend in installation rates and the current populations. These figures are used to estimate the Smart Meter/traditional meter mix in Product Class 4 for EUCs 01B-03B. The Standard Conditions of Gas Supply Licence [5] state that all sites in EUCs 04B and above must have an advanced meter. Therefore, for these EUCs, the entire population is assumed to be AMR. In addition, all sites in Product Classes 1-3 must also have a Smart Meter or AMR. AMR installation figures for Product Class 4 EUCs 02B and 03B are supplied by ICoSS. All this information is used together to create a map of the population by meter type and Product/EUC for the forecast year, which is then used to extrapolate the calculated unbiased theft split to this point in time. This process is described in detail in Section 7.9.

4.3.4 Projection of Permanent Unidentified Gas to Forecast Period

Having calculated the best estimates of Permanent and Temporary Unidentified Gas for each historic year for which reliable data is available (the training period), it is then necessary to calculate the projected values of Permanent Unidentified Gas for the forecast year (see Figure 5). Note that the estimated values for the forecast year are calculated based on seasonal normal weather. The projection is carried out individually for each Unidentified Gas component category and EUC/Product class, in each case using the most suitable data and extrapolation technique. Extrapolation to the forecast period is carried out for each of:

- Shipperless and Unregistered
- IGT CSEPs
- Consumer Metering Errors
- CSEP Shrinkage
- Volume Conversion Errors
- Balancing Factor

The methods used differ based on the observed behaviour of each category of Unidentified Gas and are in many cases affected by several UNC modifications introduced in order to address various Unidentified Gas issues. The Balancing Factor is calculated for each of the six most recent historic years with reliable meter read data (2011/12 to 2016/17) and projected forward based on the pattern observed in this time period. Input data for the directly estimated components of Unidentified Gas is reliable throughout and so all available data is used. Properties of the Balancing Factor and full details of the extrapolation techniques used in all cases are described further in Section 7.

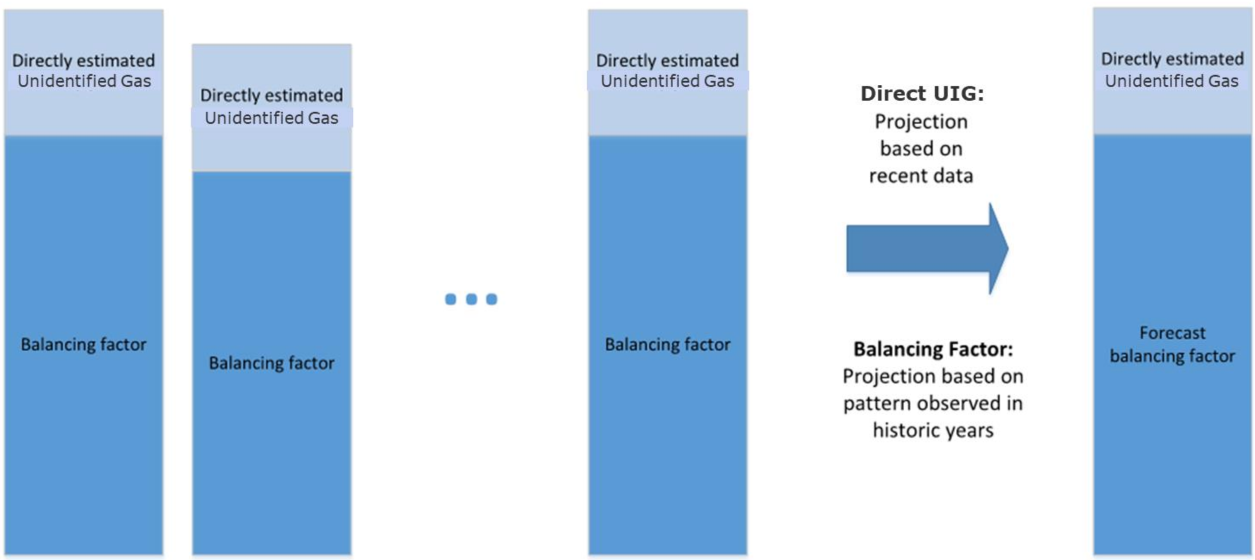


Figure 5: Projecting Unidentified Gas

As part of the estimation of the directly calculated Unidentified Gas components for the training years, an estimate of the amount of Temporary Unidentified Gas for each component is made as described above. The values projected forward to the forecast year are the permanent part of the Unidentified Gas only, however, for each EUC/Product combination. Note that detected theft (up to the 3-4 year cut-off date) is treated as a directly measured component of Unidentified Gas (100% Temporary and hence not taken forward to the forecast year).

4.3.5 Unidentified Gas Factors

The final output of the Unidentified Gas analysis is a set of UIG factors rather than direct estimates of the magnitude of Unidentified Gas itself. These factors can be applied to the population (defined in terms of the aggregate AQ for each EUC/Product) to give the relative magnitude of Unidentified Gas from each: these relative figures can then be applied to the independent daily UIG estimate to give the final breakdown (in energy terms) by EUC and Product Class.

The advantage of this approach is that this allows the effect of changing population to be taken account of in the UIG split without the need for the factors to change: when the number (and hence AQ) of sites for a particular EUC/Product category goes up or down, the fact that this AQ is then multiplied by the relevant factor ensures that the value of UIG from this source also goes up or down accordingly. This means that fixed factors can be generated that last a full year, until the results of the subsequent AUG Expert analysis become available, with the effects of changing population during that year still taken into account.

The factors themselves are a fundamental link between population and the Unidentified Gas from it, however, and so they must be calculated using the detailed estimates of the value of Unidentified Gas (for the year in which the factors will be in force) described above. Once the Unidentified Gas for each EUC/Product combination for the forecast year has been estimated, this is converted into a factor by dividing by the relevant aggregate AQ (i.e. the best estimate of the AQ for that EUC/Product combination for the forecast year):

$$\textit{UIG Factor}_{PRODUCT,EUC} = \textit{Unidentified Gas (GWh)}_{PRODUCT,EUC} \,/\, \textit{Aggregate AQ (TWh)}_{PRODUCT,EUC} \times 50 \quad (4.3)$$

Note that in this equation the Unidentified Gas and Aggregate AQ have different units, and the raw output is also scaled up by a factor of 50. These steps ensure that the resulting factors give sufficient precision when expressed to two decimal places as required by CDSP systems.

The final step in the UIG factor calculation is to apply a smoothing process. This is done on a Product by Product basis to ensure a smooth transition from each EUC to the next and avoid any step changes or up-and-down effects. The Unidentified Gas calculations involve a large amount of statistical modelling, and like any such models these will always contain a certain amount of common cause variation (i.e. natural variability that cannot be removed). This variability can, in some cases, cause the factors within a single Product Class to exhibit minor changes from one EUC to the next that fall outside the underlying pattern. A cubic fit (one per Product Class) is used to remove any such random effects and ensure a smooth pattern for the factors.

It should be noted that UIG factors are calculated by Product Class and EUC. As such they represent the average contribution to Unidentified Gas from any site within that Product Class and EUC. It is recognised that when a site changes Product Class and/or EUC it will be deemed to contribute a different quantity of Unidentified Gas as a result, even though there may be no physical change at the site. This is a consequence of the use of Product Class and EUC groupings, which creates an assumption that each site within any given group contributes Unidentified Gas in the typical manner for that group.

5 SUMMARY OF ANALYSIS

This section contains details of analyses carried out and updates made to the Unidentified Gas calculation methodology since the previous AUGS for 2019/20[58] was published. It also summarises new UNCC modifications and industry initiatives that could potentially impact on the Unidentified Gas calculation methodology in the future.

5.1 Consumption Calculation

The proposed methodology involves the calculation of Seasonal Normal (SN) consumption for several historic years (the training period). Although the approach to calculating these consumptions remains largely unchanged this year, it is worth noting a few differences from the previous AUGS for 2019/20 as follows.

- During analysis for the 2019/20 AUGS, a significant number of gaps were identified in the meter read data (missing meter reads). This was investigated by the CDSP and the root cause was found to be reads held within their systems which did not have the relevant LDZ field populated. As reads were extracted on an LDZ by LDZ basis, any meter reads without this field populated were not extracted and were therefore missing from the AUGE's dataset. This additional data was provided and included an additional 115,149 LSP reads and 12,889,436 SSP reads. Meter read data provided for the current AUG year included a further 151,745 LSP & 114,481 SSP meter reads without the LDZ field populated.
- There have been industry initiatives to improve the quality of the meter asset data held on the CDSP's systems. This updated data is used in the latest consumption calculations by the AUGE.
- All metered volumes are now provided to the AUGE in metric units. The AUGE therefore no longer needs to convert volumes based on the metric/imperial flag in the meter asset data.
- An additional year of consumption is now calculated (gas year 2016/17). This year spans the Nexus implementation date of 1 June 2017. The calculation of individual meter point consumptions is not affected by this, but in the estimation of UIG, the calculation of allocations needed to change (see 7.1 for details).
- The approach to any identified gaps and overlaps in the meter reads/volumes has changed due to the provision of the additional meter read data. Previously, the AUGE adjusted the metered volumes to allow for any gaps/overlaps. Given the number of gaps/overlaps has now reduced, the methodology now fails to calculate a consumption where gaps/overlaps exist (see section 7.2.1 for details) and a replacement value is used. Note that this change only affects LSP consumption.

The improvements to the data used by the consumption methodology have resulted in the changes to the AUGE estimates of total Unidentified Gas over the training period shown in Figure 6. The two most recent gas years show the highest levels of Unidentified Gas, but these are still subject to further reconciliation. The line in the sand is 1 April 2016 i.e. part way through gas year 2015/16. The 2016/17 gas year has the highest Total Unidentified Gas. This gas year had an anomalously high average pressure (see section 5.4). The estimated contribution to Unidentified Gas from atmospheric pressure for this year is 1,898GWh.

There are also several corrections which need to be applied to account for LDZ measurement errors and DM/PC1 errors. Pre-Nexus this was provided based on the period for which the correction applies. The

CDSP has confirmed that this data is only available by billing month post-Nexus. The AUGE will therefore need to decide how to apportion these corrections. The effect should be relatively small as this only applies to the period 1 June 2017 to 30 September 2017. These corrections will be included in future versions of the AUGS.

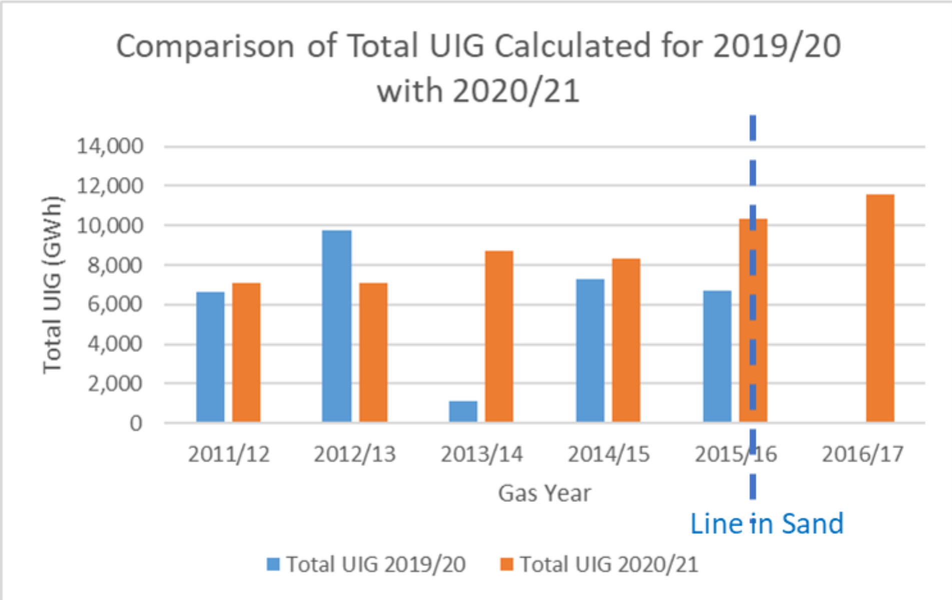


Figure 6: Change in Total UIG from 2019/20 AUGS

5.2 Fiscal Theft

Fiscal theft is an important issue due to fact that it is the only type of theft that does not contribute to UIG. Fiscal theft involves tampering with the pre-payment module of an Electronic Token Meter (ETM) so that it flows gas without payment being made. Therefore, the meter reading remains correct (and hence no Unidentified Gas is generated), but the gas has been stolen because it has not been paid for. Every other form of theft involves the falsification of the meter index in some way and hence the consumption of UIG.

Analysis carried out by the AUGE as part of the Joint Theft Reporting Review Group has shown that 78% of confirmed thefts in TRAS come from pre-payment meters, and that up to 74% of these thefts may be fiscal in nature. Fiscal theft therefore has the potential to be the single largest category of theft and given that it does not contribute to UIG it is important that it is identified as accurately as possible.

In addition to its impact on Unidentified Gas, fiscal theft is also important to wider industry processes, particularly settlement. This is because despite the meter readings of such thefts remaining correct, assessed losses are still calculated and entered into CMS. This means that they feed through into settlement and as such the consumption from these meters is double-counted for the theft period.

The impact of fiscal theft on both UIG and wider industry processes has been recognised, and as such the creation of a new Fiscal Theft tamper code in TRAS has been approved. This is currently being implemented by the industry but is not yet available. Hence a different method of identifying fiscal thefts is required until the point where the new code has been implemented and sufficient data incorporating it is available to allow reliable estimates using it to be made.

Whilst the removal of fiscal theft from detected theft calculations is an important issue, the most important application of this is in the split of undetected theft. A key assumption of the Unidentified Gas analysis is that the vast majority of the Balancing Factor, when considered over time, is composed of undetected theft. Therefore, TRAS theft data is analysed in order to create the best available estimates of the split of undetected theft between EUCs and Product Classes, and this split is applied to the Balancing Factor. The method for doing this is described in full in Section 7.9 below.

It is vital when creating the undetected theft split that only thefts that contribute to Unidentified Gas are included in the analysis, i.e. fiscal thefts must be identified and removed. This must be done despite the fiscal theft tamper code not yet being in operation, and so a different method of identifying these thefts must be used.

The work of the Joint Theft Reporting Review Group revealed that in the absence of a dedicated fiscal theft tamper code, these thefts are currently recorded as valve tampers. Whilst other types of valve tampers are possible on ETMs, this type of theft is relatively rare on traditional meters, accounting for less than 9% of thefts on traditional credit meters. Until the dedicated fiscal theft tamper code is available, it is therefore assumed that all valve tampers on ETMs relate to fiscal thefts, and these are now removed from the dataset as the first step of the overall theft method. This process is described fully in Section 7.9.5.

5.3 Metered Gas Temperature

UIG is calculated as an energy value. However, the gas consumed at supply points is measured as a volume and then converted to energy. For some larger sites, volume conversion devices are fitted to the meter which record the actual conditions (pressure, temperature and sometimes composition) of the gas. This allows an accurate conversion of the measured volume to energy under standard conditions. Overall, based on data provided by the CDSP, approximately 19% of total throughput energy is recorded by meters with a volume conversion device. For the remaining meters, several assumptions are used in order to perform the volume to energy conversion. These assumptions lead to UIG.

The AUGE has previously assessed this issue and findings were reported in the 'Final Allocation of Unidentified Gas Statement for 2019/20[58]. The previous analysis showed that the effect of temperature is significant and will vary based on geographical location and by time of year. At that time, the AUGE's opinion was there was too much uncertainty in the available gas temperature data to justify a change from the current assumption of an average gas temperature of 12.2°C and a laboratory study was proposed.

Based on additional data which is now available, the AUGE has reviewed the impact of temperature on UIG. The new information allows the estimation of the UIG resulting from using a fixed temperature of 12.2 °C and a corresponding update to the methodology for calculating UIG factors.

The following sections provide a summary of the data sources used, the results of the analysis and an overview of the approach. Full details of the approach can be found in the methodology description (Section 7.8.2).

5.3.1 Data Sources

The following sections provide an overview of the data on which the AUGE based its analysis. There are two main sources of gas temperature information, the Domestic Meter Temperature Study (DMTS) and the Industrial & Commercial Temperature Study (ICTS) which are described in detail sections 5.3.2.1 & 5.3.2.2. The remaining sections highlight additional sources of supporting data and information.

5.3.1.1 Domestic Meter Temperature Study (DMTS)

The DMTS was carried out by BG Technology (formerly the research division of BG plc, rebranded to Advantica, subsequently acquired by National Grid and then sold to Germanischer Lloyd and merged with DNV to become DNV GL) and was designed to estimate the temperature of the gas in a domestic meter during typical operation. The study was to support Transco in calculating LDZ level shrinkage factors. The DMTS was a multi-year project with an initial report in June 1999[59] and then a final report in August 2000[60]. Unfortunately, the AUGE is not able to locate the raw data collected. We are therefore reliant upon the summary level information contained within the two reports.

In total, 5,900 electronic gas meters with internal temperature sensors were fitted in homes. These were distributed as randomly as possible between LDZs and meter locations. The meters provided corrected and uncorrected reads which were used to calculate the Flow Weighted Average Gas Temperature (FWAGT) over the interval of the reads. For each installation a pro forma survey form was completed giving information on the location of the meter and other details such as the type/size of the property and the gas appliances installed.

Meters were installed across the country from late 1997 to mid-1998. The meters were then read in April 1999. On average, these meter reads covered a period of 15 months. The intention was to read the meters again in April 2000, but delays resulted in most meters being read in May 2000 such that the period covered was ~13 months. There are therefore three sets of temperature data which could be used

- The 1999 Data – valid reads taken in 1999 providing a FWAGT over the first year only (15 months on average). There was a total of 3,998 valid reads across all meter locations (1,249 internal meters, 2,420 external meters and 329 meters in unheated locations)
- The 2000 Data – valid reads taken in 2000 providing a FWAGT over the whole two year period (28 months on average). This data is referred to in the DMTS as the "2000 Data after Survey Form Analysis". There was a total of 3,900 valid reads (1,425 for internal meters and 2,475 for external meters)
- The 1999-2000 Common Set – valid reads taken in 1999 and 2000 providing a FWAGT over the second year only (13 months on average). There was a total of 3,083 valid reads (1,022 for internal meters and 2,061 for external meters)

During the first year of the study, analysis was broken down by three meter locations, namely internal, external and internal unheated (meters in unheated spaces such as a garage). The analysis in the final report[60] was based only on an internal/external meter location split (the previous unheated category being included in internal meters). The report does not explicitly state why this change was made but does refer to the uncertainty in the meter location between internal and internal unheated categories. This may be one reason for removing the internal unheated category. The internal unheated category is also more sparsely populated, having only 329 meters with valid reads (8% of the 3,998 total).

The DMTS included an analysis of meter temperature sensor accuracy during the first year. As a result, temperature corrections were applied differently during the final analysis. The AUGE therefore believes that although the results reported in the initial report[59] are useful, the results from the final analysis[60] should be used in preference. One use for the initial analysis results is to understand the temperatures in meters located in unheated spaces.

Several consistency checks carried out during the DMTS final analysis highlighted some discrepancies for meters which had been read in both 1999 and 2000 leading to additional meters being removed as invalid. The AUGE therefore believes that the most reliable dataset is the "1999-2000 Common Set". Unfortunately, this dataset has the smallest sample sizes. The other advantage of using this dataset is that the meter reads were taken approximately 13 months apart and with the additional month being April which is not a peak heating month. This means that the FWAGT will be closer to a true annual average (See analysis in Section 5.3.2.4).

Initially, the DMTS aimed to estimate gas temperature for each meter location (internal or external) by LDZ to an accuracy of +/-0.5°C with 95% confidence. This was achieved in the case of external meters but not for internal meters in all LDZs (only achieved in two LDZs). This is in part because the internal temperatures are more variable, but also because only a small sample of internal meters was available in some LDZs. The DMTS notes the greater difficulty of access for internal meters.

Appendix C provides summary level FWAGT data from the DMTS including sample size, standard deviation and 95% confidence intervals (as reported in the DMTS, not calculated by the AUGE).

5.3.1.2 I&C Temperature Study (ICTS)

In order to estimate the LDZ shrinkage resulting from the assumption of a fixed temperature value for volume conversion, Transco procured a study from XKO and Advantica Ltd (BG Technology, the research division of BG plc was rebranded to Advantica, subsequently acquired by National Grid and then sold to Germanischer Lloyd and merged with DNV to become DNV GL). This work involved taking data from approximately 7,500 daily read metering sites that use volume conversion devices over a four year period (gas years 1996 to 1999). Using the converted and unconverted meter reads, it was possible to calculate the FWAGT. The AUGE is not able to locate the raw data from the original study, but summary results are provided in "The derivation of LDZ gas temperatures for the period 1996-2000"[36].

The ICTS included a comparison of gas temperatures with the DMTS to assess consistency between the two studies which used two different approaches. The FWAGT calculated from all meters in the ICTS was "un-flow weighted" using the appropriate consumption data. This Unweighted Average Gas Temperature (UAGT) was then "flow-weighted" using the appropriate consumption data for domestic properties (presumably using 01B consumption) to obtain an estimate of domestic temperatures which was compared with the earlier DMTS work. The ICTS domestic gas temperature estimates were in good agreement with the gas temperatures measured for external meters from the DMTS and it was noted that "the conditions under which external domestic meters operate, will be closer to I & C meters and this is reflected in the much closer correlation between external, and I & C derived, domestic temperatures".

The ICTS provides FWAGT data based on the following groups

- DM
- Large I&C
- Small I&C
- Domestic (derived from I&C data)

At the time of this study, the CDSP believes there were ~2,000 DM sites. The data for the I&C groups (small + large) is therefore based on a sample size of ~5,500 meters but no information is available on the split of the sample between small and large. The CDSP believes that these will be NDM sample sites. Large I&C sites refers to those consuming >732MWh/annum i.e. EUC04 and above. The AUGE is currently liaising with the CDSP to identify how many NDM sites still have volume converters and are collecting daily reads (corrected and uncorrected).

Unfortunately, the ICTS provides no information on meter location. The reported temperatures therefore represent the sample split of meter locations and may not be representative of the overall population split of meter locations. It is likely that the split of meter locations in the sample is biased towards external meters as these provide easier access.

5.3.1.3 Other Data Sources

The following sections give a brief overview of other sources of supporting data considered by the AUGE and used within the analysis where appropriate.

Air Temperature

The CDSP have provided the AUGE with historic actual hourly air temperature by weather station from 1996 onwards to provide an overlap with the DMTS and ICTS work. This is based on the weather stations in use at that time. The AUGE has used this data to

- assess the relationship between gas temperature in the meter and LDZ air temperature
- assess the level of error in the DMTS gas temperatures due to collecting data over a period not equal to a single year (FWAGT will be biased)
- assess whether a correction is required to adjust the reported gas temperatures in the meter for a specific historic year to Seasonal Normal (SN) conditions

Historic Allocations

The available information on gas temperature in the meter spans the period 1996-2000 (gas years). This data is flow weighted based on the consumption at the time. Historic allocation data covering this period has been provided by the CDSP to allow the AUGE to convert between Flow Weighted and Un-flow Weighted temperatures as required.

In the case of the DMTS, temperatures were measured over a period of more than a year. The FWAGT will not therefore represent a true annual FWAGT. Using historic consumption together with historic temperatures an assessment was made regarding the need for a correction.

Ground Temperature

Ground temperatures are highly correlated with air temperatures and relationships exist which can calculate one from the other given depth and assumptions about local soil composition. Ground temperatures are generally lagged by a few days[36] compared to air temperatures and the values are damped giving a lower variability.

Soil temperature data are collected and archived by the UK Met Office at several hundred weather stations and are made available for academic purposes via the British Atmospheric Data Centre (http://badc.nerc.ac.uk/home). The data are recorded at 09:00 each day at depths of 5, 10, 20, 30, 50 and 100 cm. The AUGE would need a commercial agreement to use this data, but there have been studies published based on this data[57].

The British Geological Society (BGS) have analysed soil temperatures from 106 Met Office weather stations, located across the UK[57]. Mean annual temperatures at 1m depth, reduced to sea level, were found to range from 12.7°C in southern England to 8.8°C in northern Scotland.

BGS also reported an average Urban Heat Island (UHI) effect at 1 m depth of 0.55°C at localities adjacent to urban green spaces. They go on to suggest that the UHI effect will be greater in densely developed city and town centres.

A key conclusion of the BGS analysis is that on average (across 12 UK locations), mean annual soil temperatures at 1m depth are 0.9°C higher than mean annual air temperatures.

The CDSP have confirmed that they do not currently have ground temperature data. The AUGE would therefore need to purchase ground temperatures if required. At this stage, the AUGE sees no benefit in obtaining ground temperatures given measured gas temperatures at the meter are available. Depending on the outcome of the metered gas temperature lab study currently being procured by the CDSP, ground temperature data may subsequently be required.

Internal Temperatures

It has been reported that the gas temperature within internally located meters is significantly affected by the temperature in the immediate vicinity of the meter. In fact, a study by KIWA noted that "It follows that the temperature of the gas in the gas meter is almost exclusively determined by the temperature in the room where the meter is installed"[42]. It is therefore important to gain an understanding of the likely temperatures to be experienced by internally located gas meters.

The AUGE has identified two reports relating to internal domestic temperatures[61][62]. To date, no equivalent information for I&C sites has been identified.

The AUGE issued an industry data request regarding internal temperature data in the hope that data collected from building energy management systems may be available. This request can be found on the JoT website (http://www.gasgovernance.co.uk/augenex/2021).

From the reports on internal domestic temperatures, the key findings are

- Variability between homes was large (see **Figure 7** below) from Energy & Buildings report[61] which shows temperatures from a sample of 248 homes.
- There has been little or no increase in living room temperatures during winter and spring over the last 40 years[62].

- There has been an increase in average bedroom temperatures during winter and spring over the last 40 years, probably due to growth in central heating ownership, together with reductions in whole-house heat loss[62].

As internal meters are likely to be located in rooms other than the main living room, they may be subject to the increase in temperature as a result of the growth in central heating and reduction in overall heat loss described above.

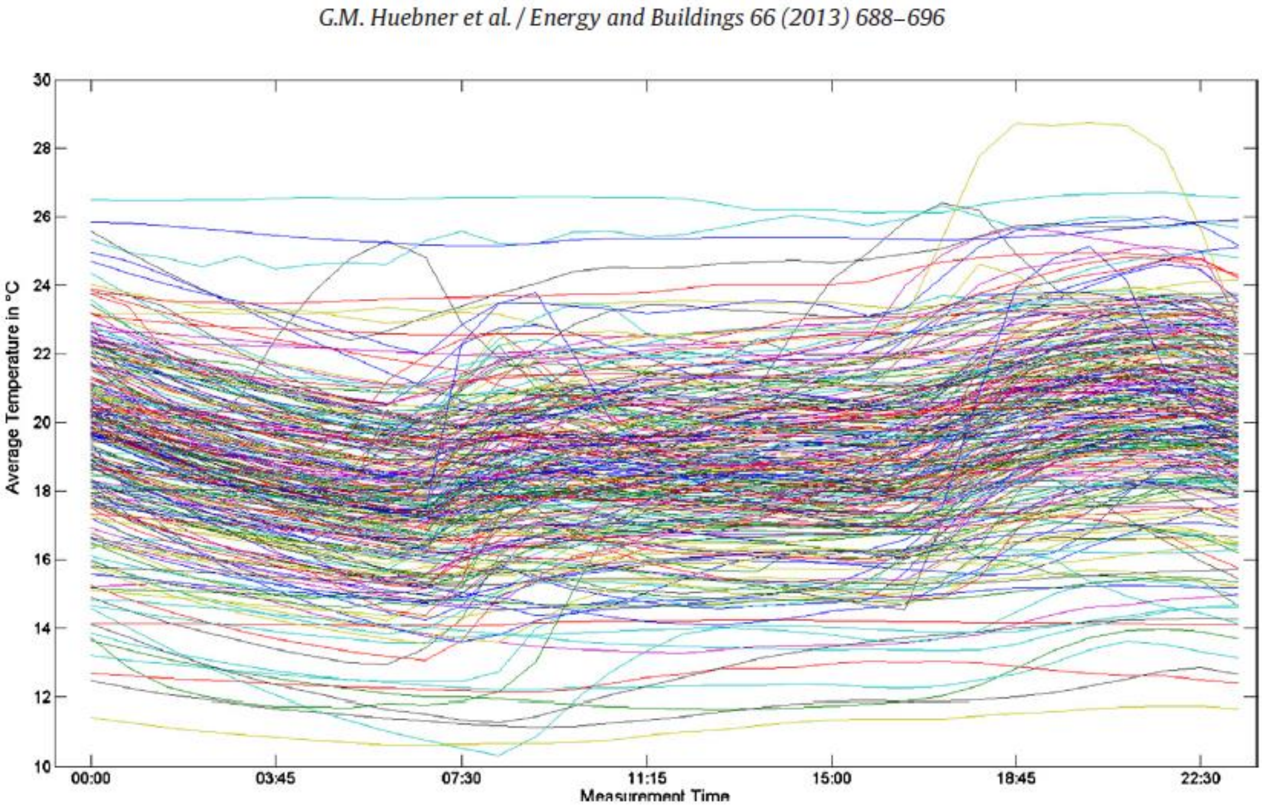


Figure 7: Average Weekday Winter Internal Temperature for 248 Homes[61]

Meter Location

Meter location information has been provided by the CDSP. Location data is available for ~77% of all meters. An analysis of the distribution of meter locations is given in Section 5.9.

Given that ~23% of meters are in an unknown location, the AUGE is assuming the same split for the unknown meters as for the known population as there is no evidence suggesting the unknown meters are split differently for some reason.

There is also the possibility that meter locations vary by property type e.g. more internal meters in flats. This would mean that the internal/external split from a consumption viewpoint would be different than the split based on meter numbers. No information is currently available on this.

The AUGE would like to see a push from the industry to improve the recording of meter location information.

5.3.2 Gas Temperature Data Analysis

Analysis was carried out around the two main meter temperature datasets (See Appendix C for actual FWAGT data)

- The DMTS 1999-2000 Common data
- The ICTS

Hourly LDZ air temperatures were sourced via the CDSP for the period 1996-2001 to coincide with the two meter temperature studies. This data was based on the weather stations used at the time and the relevant mapping to LDZs. From this data, average daily temperatures for each LDZ were calculated (simple averages, not flow weighted averages). These air temperatures were then compared to the meter temperatures in the DMTS and ICTS. These comparisons can be found in sections 5.3.2.1 and 5.3.2.2.

The remaining sections then cover the following analysis

- Assessment of gas temperatures in meters located in unheated locations based on the 1998-1999 data from the DTMS (Section 5.3.2.3)
- Assessment of the impact of calculating FWAT over an average 13 month period compared to one year (Section 5.3.2.4)
- Assessment of impact of using temperatures from 1999/2000 compared to current Seasonal Normal conditions (Section 5.3.2.5)

Consideration was also given to the fact that the measured FWAGTs are based on consumption profiles from the time of the DMTS and ICTS. If the consumption patterns have changed significantly from those of the sample sites used in the studies, then the FWAGTs will not be representative. However, without access to the profiles of the sites in the studies it is not possible to assess this effect and it must be assumed that the consumption profiles of the sample of the sites are representative of current consumption profiles.

5.3.2.1 DMTS

Actual LDZ air temperatures were compared to FWAGTs from the DMTS for internal and external meters over the same time period (1 April 1999 – 30 April 2000). The results are shown in **Figure 8**, with each data point representing an LDZ. The data shown in blue is for external meters, the data in orange for internal meters. A linear best-fit has been included for each dataset.

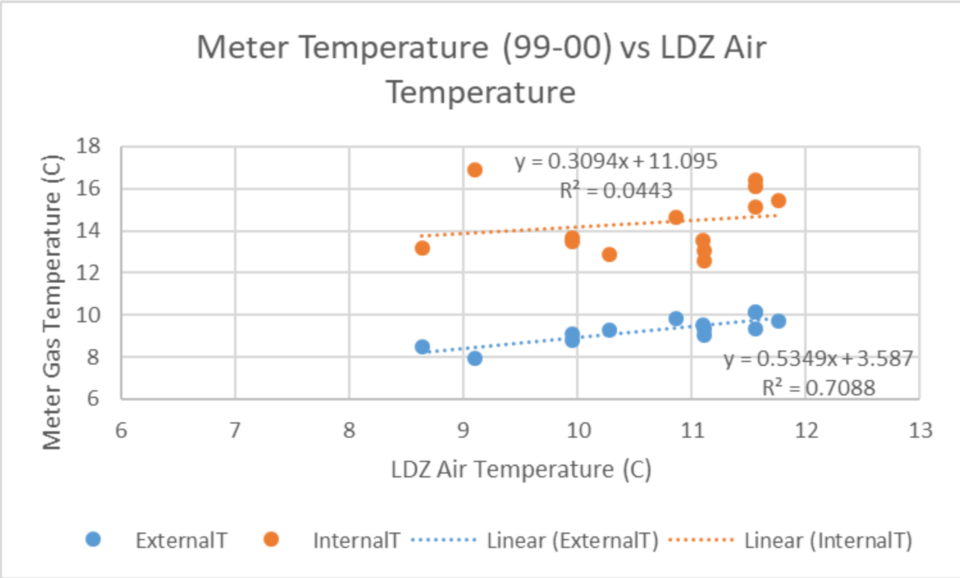


Figure 8: Relationship between FWAGT and LDZ Air Temperature for Internal & External Meters

Key observations from this data are:

- The average gas temperature in externally located meters is correlated to average LDZ air temperature (r^2=0.7088)
- A linear fit to the data suggests that for each degree C change in air temperature (simple average over 13 months), the gas temperature (flow weighted average over 13 months) changes by 0.53°C
- The gas temperature for internally located meters shows a very poor correlation with LDZ air temperature (r^2=0.0443). This is consistent with the conclusions of other studies[42] which found that the gas temperature of internally located domestic meters is primarily driven by the temperature in the vicinity of the meter
- The internal meter temperature for one LDZ looks anomalously high (16.92°C for an air temperature of 9.1°C in SC LDZ). There is no reason to suspect that this is a data error as the internal FWAGT is not significantly different to other LDZs, but it is based on a relatively small sample size of 58. Excluding this data point improves the correlation for internal temperatures to r^2= 0.357 which is still significantly lower than the correlation for external meters.

The high correlation between air temperature and external meter temperature is expected, but the gradient of the fit (suggesting only a 0.53°C change in meter temperature per degree change in air temperature) suggests that the relationship between air/ground temperature and metered gas temperature may be more complex. Possible reasons for the gradient not being 1 are

- Simple average air temperatures were used. Using FWA air temp then gradient increases to 0.62
- The fit is based on only seven points so is subject to a degree of variability
- Ground Temperature may give a different relationship to air temperature
- The gas may be subject to additional heating e.g. direct heating from solar radiation

- There may be changes in pressure (e.g. when gas passes through regulator) which affect the temperature

The lab test study proposed by the AUGE should provide additional information about this relationship.

5.3.2.2 ICTS

Actual LDZ air temperatures were compared to FWAGTs from the ICTS for the different load types (DM, Large I&C, Small I&C and Domestic) over the same time period (gas years 1996-1999). The results are shown in Figure 9, with each data point representing an LDZ and load type.

In all cases, there is a linear relationship between the FWAGT and the LDZ air temperature. The domestic, small I&C and large I&C categories are very similar, but the DM FWAGTs are higher.

Note that the ICTS domestic temperature is a derived temperature based on the I&C temperatures in order to compare and sense check the results against the DMTS. The ICTS report[36] noted that the derived domestic temperatures are generally very close to the external meter temperatures from the DMTS. This is expected as a high proportion of the meters in the ICTS work are external (higher proportion of external meters in higher EUCs), especially when expressed in terms of consumption split. A comparison of the ICTS derived domestic temperatures and the DMTS external meter temperatures is shown in Figure 10.

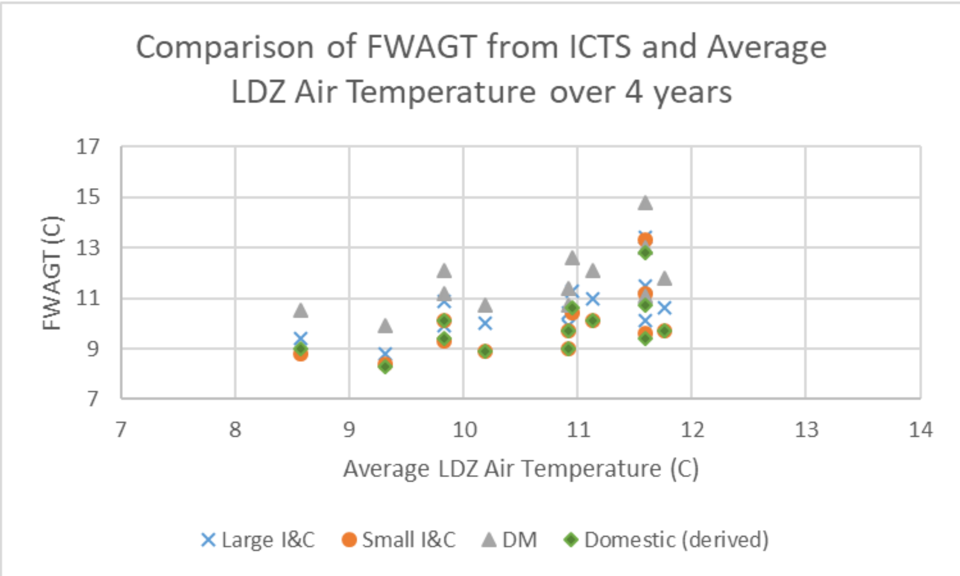


Figure 9: Relationship between FWAGT and LDZ Air Temperature for Different Load Types

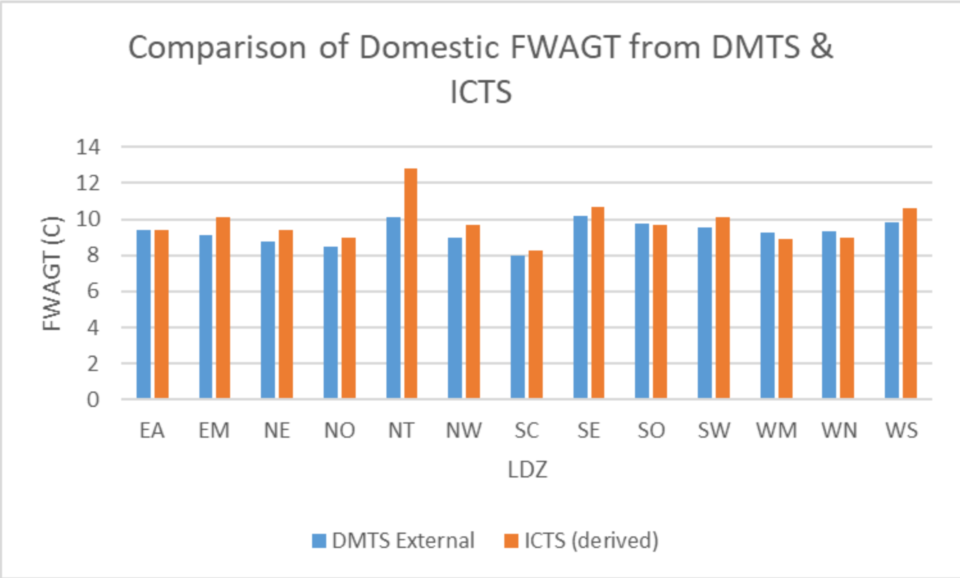


Figure 10: Comparison of Domestic Meter Gas Temperatures from DMTS and ICTS

Comparing the domestic FWAGTs between the two studies shows the average difference is -0.54°C (DMTS temperatures are on average colder). This may be due to the following

- Different temperatures between the time periods used in the two studies
- The ICTS contains some internally located meters
- Random variations due to sampling

NT is a significant anomaly and is probably the result of a significant number of internally located I&C meters in this LDZ used in the I&C temperature study work. This theory is supported by looking at the locations of I&C Meters with volume converters currently. **Figure 11** shows the percentage of external I&C meters (>01B) with volume converters. NT clearly has a much lower proportion of external meters with volume converters than other LDZs. Excluding NT, the average temperature difference reduces to -0.36C.

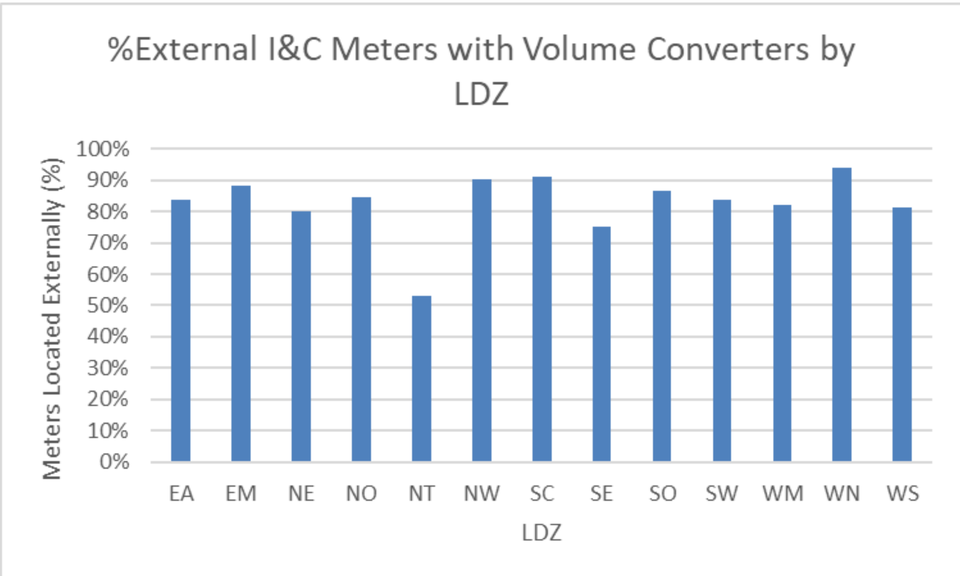


Figure 11: Percentage of Meters with Volume Converters Located Externally

This level of consistency between the two studies suggests that gas temperatures in external meters are relatively insensitive to the size of the meter (flow rate) and that it may be possible to use a single relationship between air/ground temperature and gas temperature for all EUCs. This will be confirmed or otherwise by the proposed lab tests.

5.3.2.3 Gas Temperatures in Meters Located in Unheated Locations

Based on data from the initial DMTS report[59], gas temperature in meters located in unheated spaces was on average 2.6°C warmer than gas temperature in external meters. However, this is based on a sample of just 329 meters in unheated spaces.

The final DMTS report[60] was based only on an internal/external meter location split (the previous unheated category being included in internal meters). The report does not explicitly state why this change was made but does refer to the uncertainty in the meter location between internal and internal unheated categories. This may be one reason for removing the internal unheated category. The internal unheated category is also more sparsely populated, having only 329 meters with valid reads (8% of the 3,998 total). This will result in a large degree of uncertainty in the estimated gas temperature. The AUGE has therefore made the decision to treat internal and internal meters in unheated locations as a single group.

5.3.2.4 The Impact of using 13 month FWATs

The temperatures measured as part of the DMTS were FWAGTs over an interval between two meter reads. The time period between meter reads will be different for each meter depending on when they were read, and this information is not available. The DMTS report does however state that on average, the meter reads cover a 13 month period from April 1999 to May 2000. The question is, "how different are the FWAGTs calculated over this 13 month period compared to annual FWAGTs?".

To estimate the size of this difference, LDZ air temperatures were used. A 13 month Flow Weighted Average Air Temperature (FWAAT) was calculated (April 1999-May2000) and compared to a 12 month FWAAT (April 1999-April 2000). The 12 month FWAAT was found to be 0.07C higher than the FWAAT over 13 months. This difference is negligible. The AUGE will therefore use the 13 month FWAGTs from the DMTS as annual FWAGTs without further adjustment.

5.3.2.5 Seasonal Normal Adjustments

The FWAGTs recorded as part of the DMTS and ICTS were under the specific prevailing weather conditions at the time. It is therefore important to understand if those conditions are representative or if the studies occurred during unseasonable weather conditions.

To assess this, the AUGE has obtained actual air temperature data by LDZ for the period 1 October 1996 onwards. This allows the calculation of a 5 year, 10 year and 20 year average temperature (based on the most recent gas years up to and including gas year 2017). These can then be compared to the actual periods over which the DMTS and ICTS data was collected. The results are summarised in **Table 3**. These results are averages across all LDZs.

	Avg Temperature (deg C)	Difference from 20yr Average
5yr Average	10.76	0.18
10yr Average	10.43	-0.15
20yr Average	10.58	0.00
DMTS year 1*	10.46	-0.11
DMTS year 2*	10.81	0.23
DMTS overall*	10.65	0.08
ICTS (4 years)	10.63	0.05

*Temperatures measured over a period <> 1year have been adjusted to a 12 month equivalent

Table 3: Comparison of Measured FWAGTs to Long Term Average LDZ Air Temperatures

In summary

- Temperatures during the first year of the DMTS were colder (-0.11C) than the 20 year average temperature whilst the second year of the DMTS was warmer (+0.23C)
- Temperatures over the period of the ICTS are very close (within 0.1C) of the 20 year average temperature
- The recent 5 year average temperature is slightly warmer (0.18C) than the 20 year average
- Temperatures during the second year of the DMTS were very close (0.05°C warmer) than the recent 5 year average

It is concluded that the second year of the DMTS and the complete period of the ICTS are representative of SN conditions and the meter temperatures can be used without further adjustment.

5.3.3 Summary of Gas Temperature Contribution to UIG

The Total UIG in GWh, split by EUC and PC due to temperature variations is shown in **Table 4**.

UIG_T (GWh)	PC1	PC2	PC3	PC4
01B	0.00	0.00	38.42	394.59
02B	0.00	0.00	8.97	28.74
03B	0.00	0.01	8.38	18.68
04B	0.00	0.04	6.53	21.83
05B	0.00	0.05	2.98	11.58
06B	0.00	0.15	1.12	6.97
07B	0.00	0.14	0.69	2.83
08B	0.09	0.34	0.06	1.51
09B	0.32	0.00	0.00	0.16

Table 4: Total UIG from Temperature Variations for 2020/21

5.4 Atmospheric Pressure

The 2019/20 AUG methodology introduced a correction due to variations in atmospheric pressure. This correction was at National level based on the average pressure for eight weather stations. Further analysis has been carried out to look in more detail at the effects of pressure, and in particular, the geographical variation.

Actual pressure data was provided by the CDSP. It consisted of hourly atmospheric pressure data for eight weather stations (two of the current weather stations do not have historic pressure data) covering a period of six gas years (2012 to 2017). The average pressure for each gas year and weather station is shown in **Figure 12**. This clearly shows the relationship between atmospheric pressure and latitude i.e. average pressure decreases from South to North. This relationship means that the choice of weather station for pressure data should be based on latitude rather than just distance. Details of which weather stations have been used for each LDZ are provided in Table 9.

Given that gas temperatures are now being treated on an LDZ by LDZ basis, it makes sense to do the same for atmospheric pressure. Note that the decrease in average pressure and the decrease in average temperature further North will have opposite effects on the volume conversion error (and therefore UIG) i.e. colder gas will be denser, but the lower pressure will make it less dense.

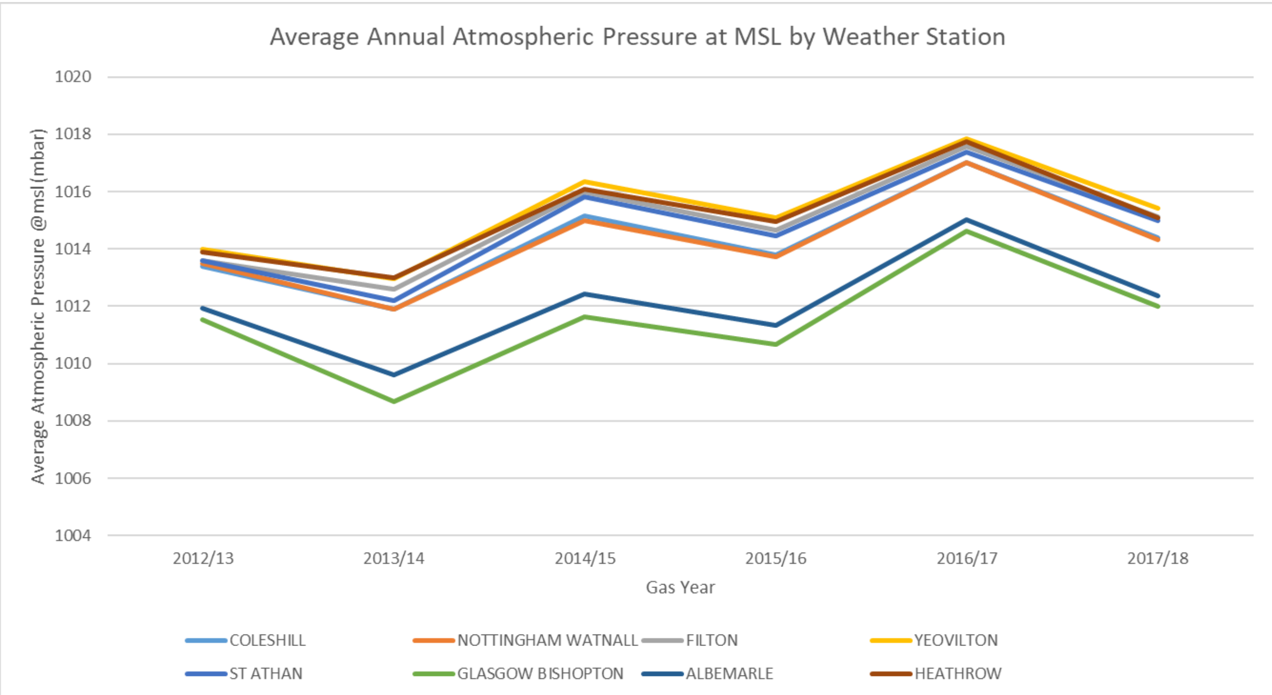


Figure 12: Average Annual Pressure Variation by Weather Station

Another change introduced for the 2020/21 gas year is the use of Flow Weighted Average Pressure (FWAP). This will account for any seasonal variations in average pressure. Total allocations at LDZ level were used to derive the FWAP.

Full details of the new methodology are given in Section 7.8.1.

The Total UIG in GWh, split by EUC and PC due to atmospheric pressure variations is shown in **Table 5**.

UIGp (GWh)	PC1	PC2	PC3	PC4
01B	0.00	0.00	3.84	39.47
02B	0.00	0.00	0.90	2.87
03B	0.00	0.00	0.84	1.87
04B	0.00	0.00	0.65	2.18
05B	0.00	0.01	0.30	1.16
06B	0.00	0.02	0.11	0.70
07B	0.00	0.01	0.07	0.28
08B	0.01	0.03	0.01	0.15
09B	0.03	0.00	0.00	0.02

Table 5: Total UIG from Atmospheric Pressure Variations for 2020/21

5.5 Permanent UIG and Reconciliation

Although the AUG Expert is not required to calculate total UIG, a value for UIG(f) (i.e. the figure at line in the sand) is required to calculate the UIG factors using the current methodology, and this value is included in the AUG Statement each year. The AUGE monitors the level of UIG(f) from year to year, particularly regarding whether it is consistent with the post-reconciliation UIG values calculated by the daily process.

The initial daily UIG values calculated under Nexus are estimates, based on the best available data on the day. Over time, as more information becomes available (e.g. consumer meter reads, LDZ offtake meter error estimates, updates to meter asset data etc.) corrections are made through reconciliation. The Total UIG calculated by the AUG Expert is an estimate of the permanent UIG remaining at line in the sand i.e. after all reconciliations. The line in the sand is reset each year to 1st April 3 years previously. There is therefore a period of 3-4 years over which reconciliations can occur (and Mod0429 allows certain adjustments after line in the sand).

In theory, over the long term the daily UIG and subsequent reconciliations should give an indication of the level of permanent Unidentified Gas. This process is still subject to a considerable degree of error, however, even though the reconciliations provide additional information and accuracy over and above the initial UIG estimate. A full discussion of this issue is contained in Section 5.6 of the 2018 AUGS for 2019/20[58], but the key differences between the daily UIG calculation process and the AUGE's calculation of UIG(f) are as follows:

- Unidentified Gas is a long-term concept, and due to its nature (being a small difference between two very large numbers), estimates of it for individual years are very volatile. It is only possible to estimate the prevailing level of Unidentified Gas accurately using analysis based on data over several years. The estimate of UIG calculated from the daily process is based on only two years data, which also contains step changes due to measures introduced to control the level and variation of daily UIG. This is not sufficient to produce a reliable estimate.

- The daily UIG analysis covers a time period that is different from the one the AUG Expert uses. Most of the time period covered by the daily UIG values has been removed from the training period for the AUGE's calculations because the reconciliation level (proportion of meter reads received) is too low to return results of the required accuracy.
- The meter reads used for reconciliation have been subjected to a lower level of validation than used by the AUG Expert.
- The effects of any Offline Adjustments that will be made up to (and potentially beyond) line in the sand are not included.

As such, even when line-in-the-sand has been reached, there may not be a 1:1 relationship between UIG(f) and the AUGE's estimate of Unidentified Gas. Such an analysis can still provide useful information, however, and conclusions from the comparison for this year are shown below.

Figure 13 below shows the CDSP figures for UIG as a percentage of total throughput. This shows the following features:

- For the first year, initial UIG is approximately 5.3% of throughput on average. This drops to approximately 4.4% of throughput after (partial) reconciliation. Line in the sand has not yet been reached for this time period and so further reconciliation is likely to occur.
- After this point, the level of initial UIG dropped due to a combination of data improvements, potential physical phenomena (such as pressure and temperature effects), and the introduction of ALP/DAF uplift factors in October 2018.
- Following the step change, initial UIG has been running at approximately 0% of throughput. This increases to approximately 1.8% of throughput after (partial) reconciliation.
- Reconciliation is not yet complete for any of the post-Nexus period, and it is clear that the step change in initial UIG has created an accompanying step change in the current best estimate of post-reconciliation UIG(f). This shows the influence of unreconciled sites on the UIG(f) estimate even when most of the reconciliation has been carried out, as it has been for most of the period being analysed.

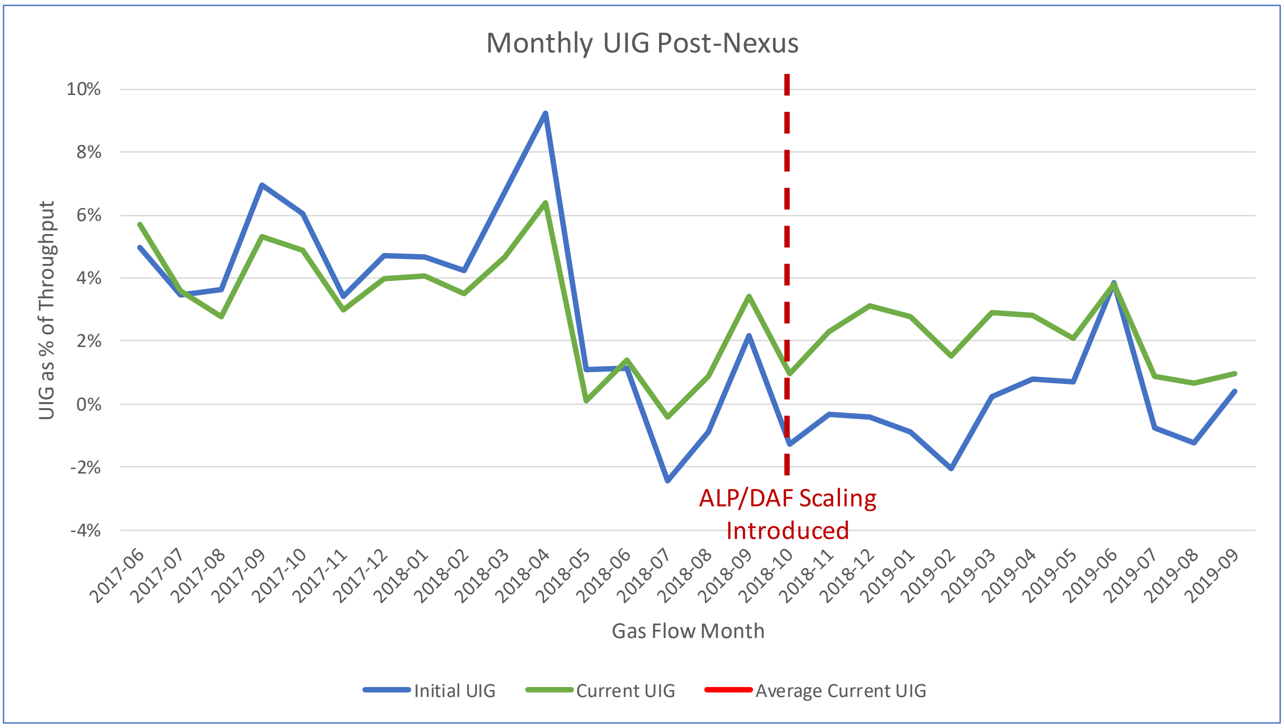


Figure 13: UIG as a % of Throughput (CDSP Analysis)

The AUGE's estimate of UIG(f) following the latest analysis is approximately 1.4% of throughput, which is somewhat lower than the estimate from the daily UIG process. The reasons for the difference between the two figures is the result of the greater degree of accuracy in the AUGE's calculations, due to:

- A longer time-frame being used to assess the genuine prevailing level of UIG.
- The restriction of training data to that which has either very full or extremely high levels of reconciliation.
- The more stringent validation procedures used by the AUGE.

5.6 Unidentified Gas Resulting from Meter Exchange

As a result of the SMART meter rollout programme, a significant proportion of the gas meter population is being replaced each year. Note that some meter exchanges are classed as cosmetic. In 2018, a sample of data associated with meter replacements was examined by the AUGE to see if there were any potential sources of Unidentified Gas. Potential issues identified included the following scenarios:

- Closing read but no opening read – when the new meter read is lower than the previous meter's closing read, a negative volume can be calculated
- An opening read but no closing read – the period between the last read of the old meter and the opening read of the new meter does not have a volume calculated
- Opening and closing reads with gap between – no volume calculated between reads. The example identified had ~3 months between the closing and opening reads

- Opening reads incorrectly flagged as final reads
- Opening and final reads out of sequence

At that time, the CDSP suggested that some of these issues were the result of missing meter reads from the AUGE's dataset. The CDSP has now provided additional data.

During this year's analysis the AUGE has examined more recent larger samples of meter asset data and corresponding meter read data where meter exchanges have occurred. The aim has been to find common issues that could give rise to unidentified gas in order to assess how material the impact of these would be.

With the data sets currently provided, the AUGE has observed situations where a meter exchange has occurred but there is no opening read or no closing read. There also appears to be an inconsistency with how round the clock indicators are used or detected. For example, a round the clock indicator may or may not be present for a meter exchange. There are examples where subsequent volume calculations result in large positive or negative consumption values.

The AUGE has sent some specific examples to the CDSP to verify whether the data is correct. For example, the opening/closing meter reads may be available, but we have not received them in our data sets. This is to ensure that we do not estimate levels of Unidentified Gas which turn out not to exist.

In addition, we have sought clarifications on some of the characteristics of meter exchange data to determine whether the patterns we are finding do indeed result in unidentified gas or whether they are resolved before line in sand (and would therefore form part of temporary Unidentified Gas).

At the time of writing the AUGS these queries are outstanding with the CDSP and therefore we are unable to draw any conclusions to the potential materiality of unidentified gas associated with meter exchanges.

Once these issues have been clarified / resolved we will revisit this issue and update the AUGS accordingly.

5.7 Accuracy of NDM Algorithm

The accuracy of the NDM algorithm is being assessed by the UIG taskforce. Inaccuracies will result in allocation errors and therefore contribute to UIG. This UIG should be temporary as all meters will eventually be read and the energy reconciled. The accuracy of the NDM algorithm therefore has little impact on the AUG methodology or factors.

However, the algorithm is used to convert measured consumptions to Seasonal Normal (SN) values and to apportion consumption between gas years. The AUG Expert therefore needs to be aware of any changes which will affect the NDM algorithm.

The AUG Expert is aware that the CWV definition is changing. The AUG Expert recommends that once all relevant data has been recalculated using the latest definitions, the calculation of UIG factors should be based on the new data. Until that time, the current definitions should continue to be used. The important thing is that any data based on CWV is calculated in a consistent manner i.e. it should all use the current or the new definition but not a mix. As Weather Adjusted Annual Load Profiles (WAALPs) will not be available until late 2020, the AUGE will continue to use the current definition for the calculation of the 2020/21 UIG factors.

5.8 New End User Categories

The AUG Expert is aware of the introduction of new EUC bands and Mod0711 has been raised to allow the UIG factors to be calculated and applied based on the new EUCs. In readiness, the AUGE requested the new EUC mapping data which would allow calculation based on the new EUCs. This data was received on 11 Dec 2019 which is too late for any analysis to feed into this version of the AUGS.

At the time of writing the AUGE has provided feedback with regards the timelines required for implementation of the new EUC bands. The decision on whether to implement Mod0711 will not be reached until February 2020 which does not allow sufficient time to modify/approve the methodology without some temporary changes to the AUGE Framework timelines.

5.9 Meter Locations

Meter location information is required to calculate the UIG related to metered gas temperature (See 5.3). Different temperatures are assumed for each LDZ and EUC. Meter location information is therefore required at this level. Meter locations by EUC are shown in **Figure 14**. There is also significant variation between LDZs.

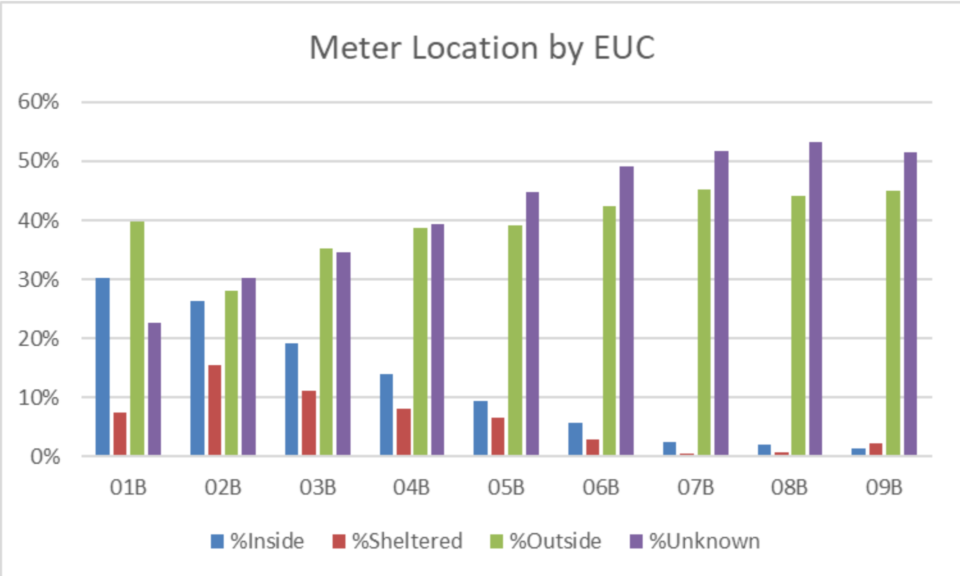


Figure 14: Meter Location by EUC

There are approximately 23% of meters without any location information. For these meters, the AUGE will assume that the location split is the same as for the known population as no other information is available. Note that the AUGE methodology only relies on the meter location for EUC groups 01-03 where the proportion of meters in unknown locations is lowest.

As described in Section 5.3.1.1, the inside and sheltered categories have been combined as the DMTS results are at this level.

5.10 CSEP Consumption Methodology Review

The AUGS for 2019/20 highlighted the estimation of CSEP consumption as an area of uncertainty in the calculation of Total Unidentified Gas. It was proposed that when enough post-Nexus data became available, an assessment of the uncertainty should be made. Prior to Nexus, the only information available for CSEPs was an aggregate AQ and supply point count by EUC from the annual CSEP invoice. Post-Nexus, CSEP data is available at individual meter point level.

Nexus go-live occurred on 1 June 2017, part way through the 2016/17 gas year. As the AUGE's consumption calculation method requires two meter reads spanning at least some of the gas year, it is not possible to calculate CSEP consumptions for 2016/17. The consumptions for all CSEPs were therefore calculated for gas year 2017/18. As this data is recent, it is expected that some meters will not yet have enough reads to allow consumption calculation. However, 76% of CSEP meters successfully had a consumption calculated.

For all CSEP meters where the AUGE was able to calculate a consumption for 2017/18, the AUGE compared the SN Consumption (Seasonal Normally adjusted Consumption) to the prevailing AQ (as at 1 Oct 2017). The difference was found to be 333GWh (1.59%), with calculated consumption being higher than AQ. This estimate is to provide information about the uncertainty in the CSEP Consumption estimates and is not intended as an estimate of additional Unidentified Gas. It represents the difference between two different estimation techniques, neither of which will be 100% accurate. It is also based on a sample of 76% of meters.

5.11 Corrected vs Uncorrected Meter Reads

Volume conversion is the process of converting the volume of gas measured at the actual prevailing metering conditions (Pressure P, Temperature T and gas compressibility Z) to the volume of gas which would have been recorded at standard conditions. This process can be performed by a volume conversion device fitted to the meter. This takes the pulse output from the meter and measures temperature and pressure at the meter to calculate reads and volumes which have been converted to standard conditions. Confusingly, the reads and volumes are usually referred to as uncorrected and corrected rather than unconverted and converted.

Approximately 19% of total throughput energy is recorded by meters with a volume conversion device. For these meters, it is the corrected reads that are used in settlement processes. Historically, both corrected and uncorrected data was received (mandatory fields), with some validation rules between the two datasets. During Nexus implementation, the validation between the meter index and the unconverted converter index was suspended[33]. The provision of uncorrected reads is also no longer mandatory where a volume converter is fitted, and no validation is carried out. This is unfortunate as the combination of uncorrected and corrected reads can provide some insight into the local conditions experienced by the meter.

The aim of this analysis was to take a sample of data for meters with volume converters and assess whether it is possible to extract useful information regarding gas temperature from the corrected and uncorrected meter reads and volumes.

Sections 5.11.1 and 5.11.2 provide background information about uncorrected and corrected data respectively. Section 5.11.3 provides an overview of the analysis and results.

5.11.1 Uncorrected Meter Reads and Volumes

Uncorrected reads are the raw output of the meter and are based on the gas flow at the prevailing conditions. The difference between two consecutive reads provides the volume of gas consumed over the period (after accounting for meter units) at the measured conditions. Uncorrected volumes are calculated from the uncorrected reads by using a fixed Correction Factor (CF_f). The uncorrected volumes are therefore an estimate of volume at standard conditions using a fixed CF_f provided by the Meter Asset Manager (MAM). The MAM calculates CF_f based on meter altitude and the meter gauge pressure assuming a temperature of 12.2°C. Unfortunately, the CDSP only holds the CF_f. A more comprehensive analysis of volume conversion could be performed if gauge pressure and (to a lesser extent) altitude information was available.

The uncorrected volume (V_m) is calculated as

$$V_m = (MR_2 - MR_1) \times units \times CF_f \qquad (5.1)$$

Where:

MR_1 and MR_2 are a pair of uncorrected meter readings

units is a correction for the units of the meter

CF_f is the fixed Correction Factor calculated by the MAM

The CF_f is in turn made up of three components representing Pressure (P), Temperature (T) and compressibility (Z) corrections as given by equation 5.2.

$$CF_f = \frac{P_m}{P_b} \times \frac{T_b}{T_m} \times \frac{Z_b}{Z_m} \qquad (5.2)$$

Where P is Pressure, T is temperature and Z is compressibility. The subscript m refers to the value at the meter and represents the conditions under which the volume measurement was made. The subscript b refers to base (standard) conditions.

The pressure and temperature corrections can be calculated as per equations 5.3 and 5.4. The compressibility factor is more complicated as it depends on pressure, temperature and gas composition. For meter pressures below 2 bar, the compressibility factor is close to 1.

$$CF_{temp} = \frac{T_b}{T_m} = \frac{(273.15+15)}{(273.15+T_{meter})} \qquad (5.3)$$

Where

CF_{temp} = Temperature Correction Factor

T_b = Temperature at standard conditions in K

T_m = Temperature of metered gas volume in K

T_{meter} = T_m converted to degrees Celsius

$$CF_{press} = \frac{P_m}{P_b} = \frac{(P_{msl}+P_{meter}+(A_{meter}\times ACF))}{P_b} \quad (5.4)$$

Where:

CF_{press} is the Volume Correction Factor relating to gas pressure effects

P_m is the pressure of the gas at the meter when the volume was measured

P_b is the pressure of gas under base or standard conditions. This is 1013.25 mbar according to GTER[26]

P_{msl} is the atmospheric pressure at mean sea level (msl)

P_{meter} is the regulator set pressure i.e. the additional pressure of the gas over atmospheric pressure at the meter (after the regulator)

A_{meter} is the altitude of the meter installation above msl in m

ACF is the Altitude Correction Factor. This is specified in the GTER[26] as -0.120208 mbar/m i.e. for every increase in altitude of 1m, the pressure will reduce by 0.120208 mbar.

5.11.2 Corrected Meter Reads and Volumes

Corrected reads and volumes are generated automatically by the volume converter based on raw meter reads and measurements of gas pressure and temperature. The use of corrected reads means no explicit volume conversion is required i.e.

$$V_c = (MR_{c2} - MR_{c1}) \quad (5.5)$$

Where:

V_c is the Corrected Volume

MR_{c1} and MR_{c2} are a pair of corrected meter readings

5.11.3 Analysis of Corrected & Uncorrected Reads/Volumes

Taken together, the corrected and uncorrected volumes provide information on the error resulting from using a fixed correction factor. The ICTS described in Section 5.3.1.2 used corrector data to estimate temperature for approximately 7,500 daily read meters over a four year period. Due to the complexity of the compressibility term (Z) in volume conversion, the ICTS excluded all sites operating at pressures of 2bar or above. This only excluded ~2.5% of the reads but allowed for much simpler calculations as it could be assumed that $Z_b/Z_m \sim 1$. The ICTS used the following equation to calculate the gas temperature.

$$T_g = \left\{\frac{V_m}{V_c} \times CF_f \times 285.35\right\} + \left\{\frac{V_m}{V_c} \times 0.284382 \times (P_m - 1013.25)\right\} - 273.15 \quad (5.6)$$

Where:

Tg is the derived gas temperature at the meter in degrees C

Vm is the uncorrected (measured) volume

Vc is the corrected volume

CFf is the fixed site specific Correction Factor calculated by the MAM

P_m is the pressure of the gas at the meter when the volume was measured

The ICTS was a comprehensive study and it is beyond the scope of the AUGE project to repeat this analysis. However, there may be some benefits in carrying out a similar analysis as follows

- Up to date estimates of FWAGT can be obtained based on recent weather conditions and demand patterns
- Meter location can be accounted for within the study (the ICTS made no distinction between internal and external meters)

The AUGE is currently liaising with the CDSP to identify what daily read information may be available for NDM sites to allow further analysis. We understand that some of the NDM sample meters have volume converters and that the data is collected by a third party.

Aside from daily read meters, there are also 6,446 NDM meters with volume converters for which the CDSP holds the data. These have varying meter read frequencies from weekly to annual.

The AUGE requested all meter read data for sites with volume converters in EA, SC and SW LDZs in order to assess whether these could provide useful information regarding gas temperature.

Using equation 5.6, the estimated temperature was calculated. An average pressure of 1013.25 was assumed for simplicity. Note that this approach assumes that the fixed CF is accurate and that the meter pressure is below 2bar. The temperature estimates were found to be reliable in only a small number of cases. **Figure 15** is an example of a meter with monthly reads with reliable data.

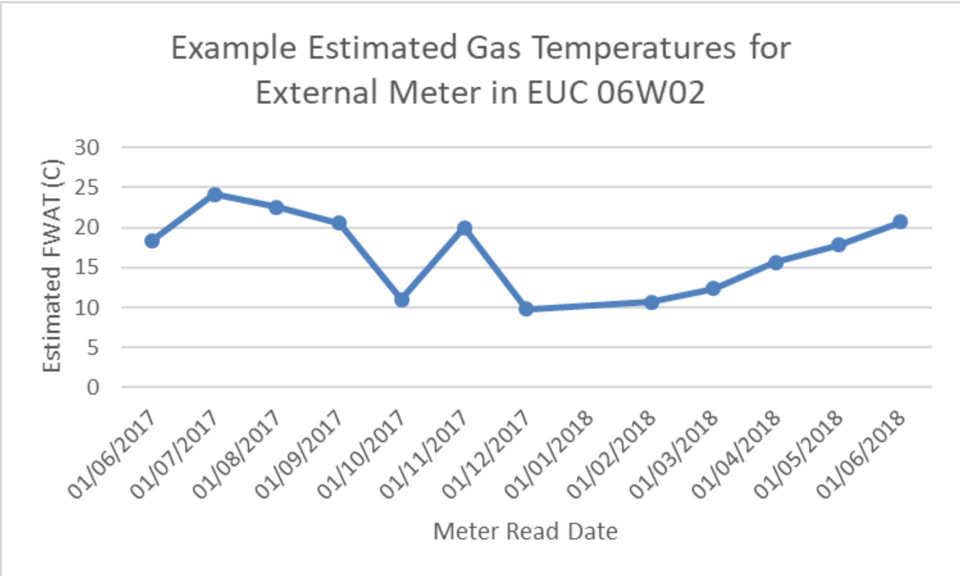


Figure 15: Example of Derived Gas Temperature for a Monthly Read Meter

Numerous examples of spurious temperatures were found using the above approach, quite often a mix of anomalous and sensible temperatures for the same meter as shown in **Figure 16**. This behaviour was also noted in the ICTS[36] but could not be explained. A significant amount of work was involved in the ICTS in developing filters to remove anomalous data.

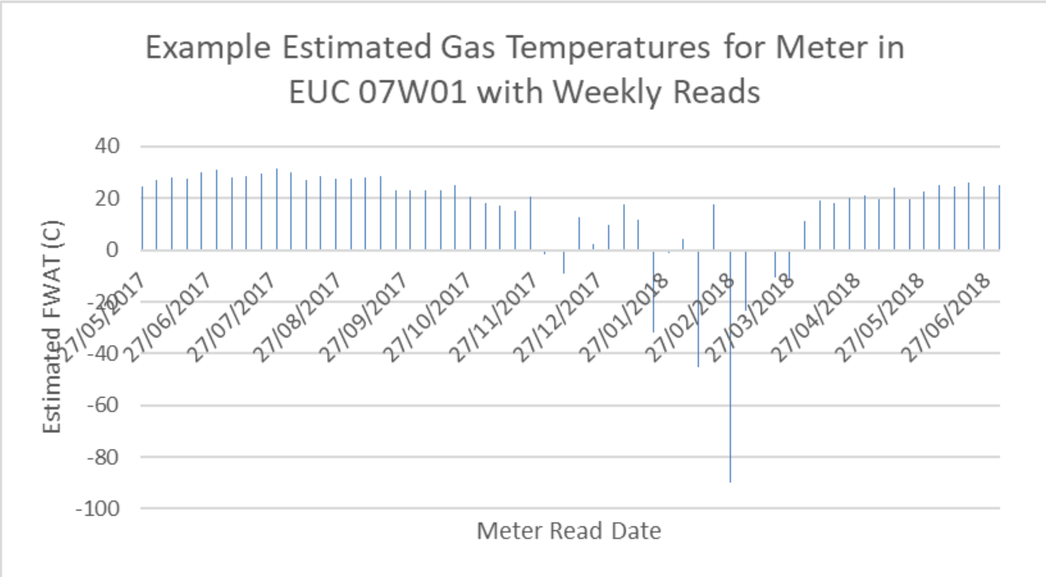


Figure 16: Example of Inconsistent Temperature Estimates for Meter with Weekly Reads

In summary, there is little useful information available from volume converter data for NDM meters without daily reads.

5.12 Changes to Product Class Populations

In order to calculate the UIG factors accurately for the forecast year, it is necessary to forecast both Unidentified Gas and the Product Class populations (number of sites and AQ) to this point in time. The methods used to calculate and forecast Unidentified Gas are documented in detail throughout Section 7 of this document. For the populations, the asset data provided by the CDSP is used.

This asset data has been provided for several points in time, from Nexus go-live to the most recent data set available, which is for November 2019. Each data set allows the population in terms of both the number of sites and the AQ to be calculated and split across EUCs and Product Classes for that point in time and is based on the EUC and PC definition for each individual site. Data is also available for CSEPs (this was previously on an aggregate basis but is now on a site-by-site basis), and this allows the CSEP population to also be included in these calculations.

Instances of this asset data exist for several points in time, which allows a trend to be established and used as a basis for the forecast. The trends are calculated at the Product Class level, and the forecast Product Class totals split by EUC using the split from the most recent instance of the asset data. The following rules are applied during this process:

- Values are projected to the midpoint of the forecast year, in this case April 2021.
- Trends are not necessarily calculated across the whole of the time period for which instances of the asset data are available. The rates of change of Product Class populations have shown significant variation, due to:
    - Product Class take-up rates changing as the new regime becomes more established, and as the result of compulsory movement of some sites between Product Classes.
    - The increased take-up of Product Class 3 during 2019.

In addition, it is expected that Mod 665 may result in step changes to the populations of Product Class 1 and Product Class 2, and the flexible trend methodology will also allow the effects of these to be extrapolated to the forecast period once asset data including these changes becomes available.

The process for calculating both the number of sites and aggregate AQ for each EUC/Product Class combination is as follows:

- Individual regression equations are trained for each EUC in Product Classes 1, 2 and 3.
- As many snapshots as possible are used for any categories for which the trend has been stable over time.
- Where the trend has not been stable, the appropriate number of historical snapshots is used to establish the trend. In the case of Product Class 3, only the two most recent snapshots are used, in order to ensure that the recent accelerated growth rate is picked up. In addition, we have capped the population of this Product Class at a maximum of approximately four million MPRNs. This cap reflects the fact that the current accelerated rate of change will not continue indefinitely, and its value is based on the number of notifications of shipper intent sent to the CDSP.
- In a similar manner to the above, trends are established for the total number and AQ of sites in each EUC and this is extrapolated to the forecast date.
- Values for each EUC in Product Class 4 are calculated by subtraction, i.e. PC4 = Total - ∑(PC1-PC3).

The population and aggregate AQ estimates that result from this process can be found in Section 7.10.

5.13 UNC/IGT Modifications

This section highlights UNC/IGT Modifications which do or could affect the level and/or split of Unidentified Gas. The AUG Expert monitors these modifications to ensure that the methodology takes these into account.

5.13.1 Mod0664 – Transfer of sites with low read submission performance from class 2 and 3 into class 4

If implemented, this modification could result in changes to the product class populations which cannot be predicted in advance by the AUG Expert. The UIG factors depend on the distribution of meters between product classes and the AUG methodology includes an approach to estimate the product class populations for the coming year (See Section 5.12 and Section 7.10). The AUG Expert will therefore monitor the progress of this modification and review the impact should it be implemented. If it is implemented, the AUGE may need to estimate the number of meters likely to move back to PC4 based on meter read performance statistics. This modification could reverse some of the current movements from PC4 to PC3.

5.13.2 Mod0665 - Changes to Ratchet Regime

This modification was implemented with effect from 1 July 2019. The modification was introduced because high ratchet charges were hindering uptake of Product Class 2. The modification introduces lower ratchet charges for PC2 but the transporters now have the ability to specify that certain supply points must be PC1 even if they are below the mandatory threshold. The AUGE envisages two effects

- A number of supply points will move into PC1. This may change the amount of UIG which needs to be apportioned to PC1. For example, the 2019/20 factors were based on no volume conversion error in PC1 as all meters were fitted with volume converters. This is no longer the case
- Given the lower PC2 ratchet charges going forward, there may be a migration of supply points to PC2. At the moment, there is no way to know how many supply points will move and the effect this will have on UIG apportionment. We will monitor the movement between Product Classes and will incorporate an appropriate trend to project forward as and when the data is available to support it. As for PC1, it may be that the proportion of meters in PC2 without volume converters will increase.

In terms of total UIG, the movement of sites into PC1 and PC2 should be positive as it will reduce the amount of allocation using the algorithm and reconciliation will occur sooner.

5.13.3 Mod0672 – Target, Measure and Reporting Product Class 4 Read Performance

If implemented, this modification should result in an improvement in the accuracy of allocation primarily due to the use of more up to date AQs. This should help reduce levels of UIG at the time of allocation and could also result in more timely reconciliation. However, this modification will have no impact on permanent Unidentified Gas (as it is assumed that all meters are currently reconciled before line in the sand) and will therefore not impact on the current AUG methodology.

5.13.4 Mod0677R - Shipper and Supplier Theft of Gas Reporting Arrangements

The Joint Theft Reporting Review Group met between March and September 2019 and resulted in several recommendations regarding theft reporting being made. These recommendations are in the following general areas:

- Mandating of theft reporting to ensure all Shippers and Suppliers populate TRAS and/or CMS (as appropriate) with information for all leads, investigations and confirmed thefts.
- Consistency of data between CMS and TRAS, potentially as a result of information sharing between the two systems.
- Introduction of a dedicated tamper code for fiscal theft, allowing this unique type of theft to be identified accurately and processed in the correct manner.

All these areas have the potential to improve the data available to the AUGE for the analysis of theft. As and when enough data is available, they will allow the full method for removing bias from TRAS outcome file data to be applied. This will improve the accuracy of the Balancing Factor split.

The introduction of a fiscal theft tamper code will allow this type of theft to be identified accurately, replacing the current assumption that all valve tampers on pre-payment meters are fiscal thefts. This will require a minor update to the methodology and will again result in greater accuracy in the Balancing Factor split.

5.13.5 Mod0681S/IGT127/XRN4932 - Improvements to the quality of the Conversion Factor values held on the Supply Point Register

The necessary CDSP system changes to support this modification are scheduled for implementation in the June 2020 UKLink release.

This modification will mean that the CDSP can make changes to the conversion factor where a meter crosses the 732,000kWh/annum consumption threshold. Above this threshold, a meter should have a specific CF based on the altitude and set pressure of the meter (but using the standard temperature conversion based on 12.2°C). Below the threshold, the standard CF should be used. The modification allows the CDSP to switch between using the site specific CF and the standard CF as appropriate when a meter's consumption crosses the threshold in either direction (assuming a site specific CF is available).

This approach is consistent with the GTER[26]. However, in terms of volume conversion error, switching from a site specific CF to the standard CF will make volume conversion less accurate and result in UIG. The difference between the site specific and standard CF is the use of the site specific altitude and set pressure information. Neither of these would change as a result of a decrease in consumption taking the site below the consumption threshold.

5.13.6 Mod0690S - Reduce qualifying period for Class 1

This Modification proposes that "the qualifying period for the requirement for a meter point to become Class 1 is reduced, to limit the time period when very large sites are subject to NDM Demand Estimation, as opposed to being Daily Metered".

This modification should have a positive impact, reducing Initial UIG and therefore subsequent reconciliation. In terms of the AUG methodology, no update is required. However, the modification may result in more supply points moving to PC1. Depending on the supply points in question, this may result in a change to the UIG allocated to PC1. For example, the proportion of meters in PC1 with volume converters may reduce, thus increasing the amount of UIG assigned to PC1.

5.13.7 Mod0691S - CDSP to convert Class 3 or 4 meter points to Class 1 when G1.6.15 criteria are met

This modification would allow the CDSP to take steps to convert a meter point to PC1 automatically when certain criteria have been met.

The impact of this modification is the same as for Mod0690S above.

5.13.8 Mod0692S/XRN4941 - Automatic updates to Meter Read Frequency

This modification places an obligation on the CDSP to automatically update the meter read frequency of a PC3 or PC4 meter to monthly if either the AQ increases to >=293,000kWh or the supply point register is updated to show that either Smart or Advanced metering equipment is in place. This will have no impact on the AUG methodology but may result in more frequent meter reads making Initial UIG more accurate and reducing reconciliation.

5.13.9 Mod0693R - Treatment of kWh error arising from statutory volume-energy conversion

This workgroup is assessing the impact of volume conversion errors as a result of using the statutory conversion factors. Currently, there is not enough information for the AUGE to assess the impact of any changes. The AUGE will continue to monitor the progress of this workgroup and will provide input if requested.

5.13.10 Mod0699 - Incentivise Read Submission Performance using additional Charges

This modification proposes a Shipper incentive mechanism to encourage the submission of valid meter readings into settlement. This modification should provide a benefit in terms of UIG volatility due to more timely reconciliation and improved AQs but will not impact the current AUG methodology.

5.13.11 Mod0700 - Enabling large scale utilisation of Product Class 3

This modification supports the movement of significant numbers of meters to PC3 and was implemented with effect from 28 September 2019.

To generate the AUG factors, the methodology needs to assume a Product Class split. This is based on latest statistics and a projection of further changes going forward to gas year 2020/21. This modification will allow a large number of meters to move to PC3. The AUGE therefore needs to take care when estimating the PC populations. For this Proposed AUGS, PC data from November 2019 has been used in addition to information regarding Suppliers intent to move meter to PC3. In order to make the final factors as accurate as possible, the latest data available will be used in the calculation.

5.13.12 Mod0711/XRN4665 Update of AUG Table to reflect new EUC bands

This modification would update the AUG table to include the new EUCs introduced by DSC change proposal XRN4665. The Workgroup is due to report to the Modification Panel on 20 February 2020.

Until this Modification is approved, the AUGE needs to continue to generate the table of AUG factors as defined in UNC. If this modification were approved in February, the AUGE would be unable to update the AUG table for the 2020/21 gas year under current framework arrangements. However, with industry agreement, the AUGE could work towards including this update for 2020/21. At the time of writing this draft the outcomes of the most recent meeting to discuss Mod0711 is unknown.

5.13.13 IGT126F - Alignment between the IGT UNC and UNC for Unregistered New Supply Meter Points

This modification ensures that the IGT UNC and the UNC contain equivalent provisions for handling Unregistered supply points. This may impact on the number of Unregistered sites in known CSEPs but does not affect any of the Unidentified Gas calculation processes.

5.13.14 XRN4645 - The rejection of incrementing reads submitted for an Isolated Supply Meter Point

This issue will not have any impact on UIG at allocation but may result in the quicker re-establishment of Isolated meters consuming gas if Shippers action the meter read rejections. This has the potential to improve the accuracy of the UIG calculation but does not impact on the AUG methodology.

5.13.15 XRN4690/4713 - Actual read following estimated transfer read calculating AQ of 1

This will remove one of the causes of AQs of 1, which has the potential to make allocation more accurate. It may also help with read submission levels due to fewer rejections for AQ tolerance failures. It may therefore improve the accuracy of initial UIG over time but does not impact on the AUG methodology.

5.13.16 XRN4740 - AQ Calculation for Class 4 sites with AMR Fitted

As above, this issue may improve the accuracy of initial UIG over time but does not impact on the AUG methodology.

5.13.17 XRN4742 - Incorrect AQ calculations due to migrated consumption

As above, this issue may improve the accuracy of initial UIG over time but does not impact on the AUG methodology.

5.13.18 XRN4772 - Composite Weather Variable (CWV) Improvements

This change is to create the mechanism to allow solar radiation and precipitation values to be loaded into UK link to support a revised CWV definition. This should make CWV more accurate which in turn will reduce allocation error and therefore UIG.

This change has no impact on the methodology, only the data used in the AUG factor estimation process. It is important that all data used by the AUGE is consistent i.e. CWV, SNCWV, WCF and WAALP must all be calculated on the same basis.

5.13.19 XRN4933 - Datafix CWV and recalculate WCF and WAALPs for SW LDZ

This may impact the AUGEs calculations of consumption (and total UIG) for SW LDZ. The AUGE has sought clarification from the CDSP about whether the data currently held by the AUGE is accurate or needs updating. When updated data is provided by the CDSP, the AUGE will recalculate consumptions for SW LDZ and update the UIG factors.

5.13.20 XRN4946 - Reporting on Installed Meters with Conversion Capability

This modification will have no impact on the AUG methodology but may provide an additional source of information for use in the factor calculation.

5.14 Innovation Areas

This section covers topics which the AUGE would like to raise with the industry regarding how UIG is calculated and apportioned. They are deemed as innovation as they either require changes to UNC or the AUGE framework or they may require a significant amount of analysis effort beyond the scope of the AUGE analysis.

5.14.1 Allocation and Reconciliation of UIG based on different factors

The AUGE is responsible for calculating a table of factors to split UIG between EUC and Product Class. The use of a single table of factors creates an issue as UIG varies over time. This is shown in **Figure 17**. At the point of calculation of UIG, a significant proportion of the UIG is actually temporary in nature. Over time, up to line in the sand, reconciliations are calculated which remove this temporary component of UIG to leave only the permanent part. If the split of temporary and permanent UIG are different, a single set of factors cannot correctly apportion UIG on the day whilst also ensuring the final position at line in the sand is correct. The current AUGE approach is to calculate factors to correctly apportion the final value of UIG (permanent only) at line in the sand, but it is recognised that this will not correctly apportion Initial UIG.

One approach to overcome this would be to use two sets of factors. The first set, referred to as F_{mixed} in **Figure 17**, would be calculated based on an estimated split of permanent and temporary UIG to correctly apportion initial UIG on the day. The second set of factors would represent the split of temporary UIG only (F_{temp}). This set of factors would be used for the process of reconciliation. Reconciliation can be thought of as the process to remove temporary UIG, so it is appropriate to use factors representing the split of temporary UIG only for this. If the ratio of permanent and temporary UIG assumed in the calculation of F_{mixed} is correct, then by line in the sand, the remaining permanent UIG should be split correctly based on the permanent only factors which would have been calculated under the current methodology (F_{perm}).

This approach is a significant change from the current approach and would require changes to UNC, CDSP systems and significant analysis work by the AUGE. Currently, UIG which is temporary is ignored by the AUGE. The AUGE would therefore need to develop factors specific to temporary UIG.

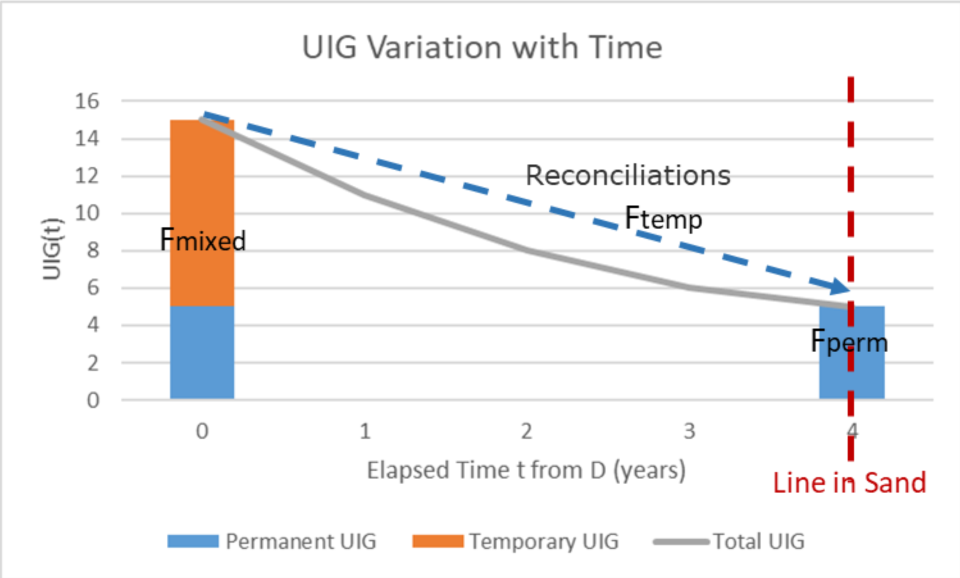


Figure 17: Different UIG Factors to account for UIG variation over time

5.14.2 Volume Converters

Volume conversion error is a significant contributor to UIG. Although the AUGE factors currently account for volume conversion errors, the approach is restricted by the current Product Class definitions. If a volume converter is fitted to a meter, this will currently have a negligible benefit in terms of UIG as the volume conversion error is calculated at the Product Class level. If the fitting of a volume converter resulted in that meter no longer being subject to any of the volume conversion portion of UIG, this could offer an incentive to fit volume converters. If the number of volume converters increased, this could reduce both the level and volatility of UIG.

There are a number of ways in which this could work. At this stage the AUGE would like feedback from the industry as to whether this is an issue to pursue further before investigating the best way forward.

5.14.3 Gas Temperature Derivation from Daily Read Meters

The ICTS used data from daily read DM and NDM meters to calculate metered gas temperatures. A similar study could be carried out to improve gas temperature estimates. The ICTS was a significant piece of work and would fall beyond the current scope of the AUGE.

Advantages of carrying out a similar study would be:

- Temperatures would be more representative of current conditions, both in terms of temperature (climate change) and consumption profiles (measured temperatures are actually flow weighted averages)
- Depending on the sample of daily read sites it may be possible to derive an internal/external meter location split
- Depending on the sample of daily read sites it may be possible to gain a greater understanding of the small I&C market sector

6 DATA USED

This section describes the data requested, received and used to derive the methodology to calculate the UIG factors. The AUG Expert has taken care to ensure that all datasets include all components of NDM consumption, i.e. CSEPs and Scottish Independent Networks are included throughout.

Section 6.1 below gives a summary of the data items requested and their current status. The subsequent sections give more detail about the data items for each individual element of the analysis.

As part of the AUG Expert's quality control process, several standard data checks have been defined which are run prior to performing any consumption calculations. Any anomalous data is reported to the CDSP for further investigation. At the current time, not all issues identified have yet been resolved, so by necessity the values contained in this document are based on the best data available. There are also several checks during the calculation process to ensure that where data is unreliable it will not be used in the estimation of the UIG factors.

6.1 Summary

Analysis Area	**Required Dataset**	**Status**
Total Unidentified Gas Calculation (Consumption Method)	NDM allocations/throughput	Received
	NDM meter reads and volumes	Received
	LDZ, DM and Unique Sites Metering Errors	Received
	Meter Asset Information	Received
	Algorithm data (ALPs, DAFs, EWCFs) including CV	Received
	Mod0429 Claims Data	Received
	CSEP AQ data	Received
	Non-CSEP AQ data	Received
	MMSP details	Received
	Prime and Sub-Prime meter details	Received
	New and Lost Sites	Received
Unregistered and Shipperless Sites	Connection details for orphaned sites	Received
	Gas Safety Regulations visit data	Received
	Further investigation results for large/suspicious sites	Supplied on request
	Mod 0410A supporting data	Supplied on request
	Shipperless sites supporting data	Supplied on request
	Snapshot files (including MPR details)	Ongoing

Analysis Area	Required Dataset	Status
IGT CSEPs	Known CSEP data	Partially received. Data query outstanding.
	Unknown Projects Snapshot files	Not received
Meter Error	Meter capacity report	Received
Theft	Detected and alleged theft updated to end March 2019 - CDSP	Received
	TRAS outcome file data	Received
Volume Conversion Error	Information on presence of volume converters and meter locations, historic pressure & temperature data	Received
Product Classes	Meter point Product Class	Received

Table 6: Data Status Summary

6.2 Total Unidentified Gas Calculation (Consumption Method)

Data has been requested from the CDSP in the following formats. In all cases, data has been provided for up to part way through gas years 2018/19.

- Allocation data on a day-by-day basis, split by End User Category (EUC). This data includes CSEP allocations.
- Meter read data on an MPRN-by-MPRN basis, with one record for each meter read. Therefore, the volume of data supplied for each MPRN is dependent on the meter read frequency for that meter.
- Aggregate meter error adjustments for LDZs, DMs and Unique Sites.
- Meter asset information on a MPRN-by-MPRN basis. This includes meter installation dates, metric/imperial flag, numbers of meter dials, meter index units and T&P correction factors. This information is used in several different parts of the consumption algorithm.
- NDM Deeming Algorithm factors and CVs for the analysis period.
- Aggregate MPRN count and AQ data by EUC for CSEPs. Pre-Nexus meter read data is not available for these sites, but knowledge of the number and AQ of MPRNs allows them to be included in the total Unidentified Gas calculations when the sample consumption is scaled up to cover the full population.
- A history of AQ and EUC data for each MPRN so that calculated consumptions can be validated against AQs and failed meter points can be replaced with an appropriate EUC average.
- Details of all meter points which are or have been part of a Prime and Sub configuration during the analysis period. This includes re-confirmation data to track the potential disaggregation of prime and sub configurations.
- Lists of all new sites and lost sites during the analysis period, including start/end dates. These are used to accurately track the population over time and to ensure that each new or lost site is only included in calculations for the period for which it was active.

The provision of this data allows the consumption for each individual meter point, for each gas year of interest, to be calculated using the method described in Section 7.2. The exact format of the data provided is described in Appendix A.

6.3 Unregistered/Shipperless Sites

The following information is supplied by the CDSP for all Unregistered and Shipperless sites (data supplied on a site by site basis). The CDSP have created a regular report to ensure that new data is collated and sent to the AUG Expert every month. This report covers the following categories of Unregistered and Shipperless sites:

- **Shipper Activity**

  These are new sites created more than 12 months previously, that a Shipper has declared an interest in (such as by creating the MPRN) but are nevertheless not registered to any Shipper. This data is split into sites believed to have a meter and those believed to have no meter.

- **Orphaned**

  These are new sites created more than 12 months previously, that no Shipper is currently declaring an interest in. This data is split into sites believed to have a meter and those believed to have no meter.

- **Shipperless Sites PTS (Passed to Shipper)**

  These are sites where a meter is listed as having been removed and 12 months later the gas transporter visits the site to remove or make the service secure (the GSR visit), but finds a meter connected to the service and capable of flowing gas. If it is the same meter as supposedly removed 12 months previously it is passed to the Shipper concerned to resolve.

- **Shipperless Sites SSrP (Shipper Specific Report)**

  Similar to Shipperless (Passed to Shipper) sites, these are sites where the GSR visit finds a new meter fitted and capable of flowing gas, in which case it is reported to all Shippers.

- **No Activity**

  These are sites currently being processed. They will end up in one of the other categories.

- **Legitimately Unregistered**

  These are sites believed to have no meter and hence are not capable of flowing gas.

- **Unregistered <12 Months**

  These are new sites that have been in existence less than 12 months and are not registered with a Shipper. Action is not taken on such sites until they have been in existence for 12 months. At this point they will move to either the Shipper Activity or the Orphaned category.

For all of these Unregistered/Shipperless Unidentified Gas categories, the following information is supplied for each site:

- Dummy MPRN
- LDZ
- AQ
- Meter Point Status

In addition, the following data is supplied for individual Unidentified Gas categories:

- Meter Attached Y/N
    - Shipper Activity, Orphaned, No Activity, Legitimate
- Meter Point Effective Date
    - Shipper Activity, Orphaned, Unregistered <12 Months, No Activity, Legitimate
- Shipperless Date
    - Shipperless PTS, Shipperless SSrP
- Isolation Date
    - Shipperless PTS, Shipperless SSrP

In addition, the following information is supplied on an annual basis:

- A summary of the remaining Shipperless sites, i.e. those that have been recorded as Isolated for less than 12 months and are awaiting their GSR visit. These sites do not yet appear in the Shipperless PTS or Shipperless SSP lists because sites only qualify for these after the GSR visit has found a meter at the site. This data comes from GSR visit records.
- Connection details for Orphaned sites, including asset and Shipper meter reads and information on whether the confirming Shipper is the same as the Shipper whose Supplier requested asset installation. This data is used to determine the proportion of sites that have been flowing gas prior to becoming registered and the proportion of these that can be back-billed.
- Shipperless sites supporting data. This is used to ascertain the final outcome for each Shipperless site that has appeared in any snapshot but has subsequently been either disconnected or (re)confirmed. This is used to determine whether the Unidentified Gas arising from them is temporary or permanent under the terms of Mods 0424 [7] and 0425 [8].

6.4 IGT CSEP Setup and Registration Delays

Data for IGT CSEP setup and registration delays consists of two elements, as follows:

- **Unknown projects summary, including**
    - the number of unknown projects by LDZ
    - a count of supply points and aggregate AQ of unknown projects by LDZ

  This data was supplied by the CDSP in monthly snapshot files on an ongoing basis. The last snapshot received was June 2017. The CDSP has confirmed that this information is not available post-Nexus and have suggested the IGTs as an alternative data source. A data request has been issued to the IGTs regarding this.

- **Known CSEP Data**

  This file contains data for both registered and unregistered sites on known CSEPs, with each listed separately. Data is supplied on an annual basis and contains the following data fields:

  - LDZ
  - EUC
  - Number of supply points
  - Aggregate AQ

6.5 Meter Errors

Data for meter error calculations consists of meter capacity, AQ and LDZ for all commercial sites. This report is supplied on an annual basis, with the latest one having been received by the AUG Expert in October 2019. This data is used to identify sites that due to the combination of AQ and meter capacity are likely to be operating at either a high or low extreme of their range, where bias in the readings starts to occur.

6.6 Theft

Data supplied by the CDSP consists of all recorded detected and alleged thefts from 2008 to March 2019. For each theft, the following key data items are supplied:

- Dummy MPRN
- Theft start and end dates
- LDZ
- Meter AQ
- Estimate of energy value of theft (kWh)

In addition, SPAA have supplied the SPAA Schedule 33 data for 2015 to 2017 (summary level) plus line-by-line data from the TRAS Outcome files covering the following:

  - Dummy MPRN (defined by the CDSP)
  - Meter Serial Number (supplied to CDSP only, not provided to the AUGE)
  - EUC (derived by matching MPRN with asset data)
  - Product Class (derived by matching MPRN with asset data)
  - Meter Type (traditional/Smart/AMR – derived by CDSP)
  - Meter installation date (derived by matching MPRN with asset data)
  - Source of lead (MAM, MRA, GT, TRAS, own analysis, tip-off)
  - Lead investigated? (Yes/No)
  - Theft detected? (Yes/No)
  - Assessed Losses
  - Tamper Code

The aspiration is for this dataset to include records for all leads (including ETTOS and TRAS qualified outliers) regardless of whether they were investigated. At this point in time, however, concerns with data quality and completeness have resulted in the data for the 2020/21 analysis being restricted to records involving actual investigations only. Whilst all requested fields are present for the theft records supplied, the following are missing:

- ETTOS leads that were not investigated
- TRAS Qualified Outliers that were not investigated

The loss of this data does not prevent the new theft analysis from being carried out, but it does limit its ability to remove some of the effects of Supplier theft detection targeting from the data.

6.7 Shrinkage

The data required to quantify CSEP Shrinkage is described below:

- GBNA network models (supplied by Cadent).
- Figures for the number of consumers in CSEPs (supplied by the CDSP as part of the IGT CSEPs data)
- National Leakage Test leakage rates (public domain)

6.8 Volume Conversion Errors

The data required to calculate volume conversion error due to atmospheric pressure variations is:

- AQ of all meters with Volume Conversion equipment fitted
- Historic actual atmospheric pressure data at MSL for a sample of locations across GB

The data required to calculate volume conversion error due to temperature variations is:

- Meter location information for all meters
- Gas temperature data, obtained from BG Technology Report GRTC 3895: Summary Report of the Domestic Gas Temperature Survey (2000) [60]
- Gas temperature data, obtained from ICTS report [36]
- Current and historic seasonal normal air temperature by LDZ

The data required to calculate the volume conversion error due to the use of standard CF rather than site-specific CF is

- AQ of all meters in EUC 04 and above without volume conversion equipment fitted
- CF data for all meters in EUC 04 and above without volume conversion equipment fitted
- Rate of update of CFs from standard to site-specific

6.9 Product Classes

This data set is supplied in conjunction with the asset data used in the Consumption Method and consists of an additional field for each site specifying the Product Class. This data is therefore available for each instance of the asset data that has been supplied.

- An initial list of the Product Class of each meter point at Nexus go-live (1st June 2017)
- Periodic updates listing all product class changes including the date of the change. The latest dataset available is from November 2019.

7 METHODOLOGY

This section describes in detail the methodology for estimating each element of Unidentified Gas.

The first stage in the calculation process is to use the Consumption Method to estimate the total Unidentified Gas for each year in the training period (using six gas years from 2011/12 to 2016/17). This process is very similar to that used by the AUGE previously [58] but has been updated to account for changes in data format and availability Post-Nexus.

All directly estimated Unidentified Gas categories are then calculated for the same period: this allows the amount of Temporary Unidentified Gas within the Consumption Method total for each year to be ascertained and allows the Balancing Factor (mostly undetected theft) to be calculated. All Unidentified Gas in the Balancing Factor is Permanent.

The data patterns observed in the training period for each Unidentified Gas component, including the Balancing Factor, are used to extrapolate to the forecast year (currently 2020/21) and provide the best estimate of each Permanent element of Unidentified Gas for this year. This is carried out individually for all 36 EUC/Product combinations for every Unidentified Gas category. Finally, these Unidentified Gas estimates are converted into factors by dividing the GWh Unidentified Gas estimates for the forecast year by the aggregate AQ for each EUC/Product combination, as per equation 3.3 in Section 4.3.5.

As given in equation 3.2 (Section 4.3.1), the Consumption Method can be stated in its simplest form as:

$$\textit{Total Unidentified Gas} = \textit{NDM Allocation} - \textit{Metered NDM Consumption}$$

This calculation involves correcting the allocations to take account of meter errors (LDZ offtake and DM) and calculating the metered consumption using meter reads, metered volumes or an average consumption value for sites where no reliable metered data is available.

The Total Unidentified Gas calculated as above includes both Permanent and Temporary Unidentified Gas. Therefore, Temporary Unidentified Gas (calculated from the individual component parts of Unidentified Gas) must be subtracted from the initial Unidentified Gas total, and it is this amended figure that then goes forward into the remainder of the calculations.

7.1 Correcting the NDM Allocation

The NDM allocation is calculated differently pre-Nexus and post-Nexus. Pre-Nexus, the NDM allocation is

$$Alloc_{NDM} = \textit{Aggregate LDZ Load} - \textit{DM Load} - \textit{Shrinkage}$$

Post-Nexus, the NDM allocation calculation no longer includes a scaling factor to ensure energy balance within the LDZ. The energy balance post-Nexus is the UIG, hence

$$Alloc_{NDM} = \textit{Aggregate LDZ Load} - \textit{PC1 Load} - \textit{Shrinkage} - \textit{UIG}$$

As the training data period now covers a mix of pre and post-Nexus data, the post-Nexus allocations have UIG added back in to provide consistency with pre-Nexus allocations.

Any subsequently detected significant errors in Aggregate LDZ Load, DM/PC1 Load or shrinkage will constitute Temporary Unidentified Gas which has since been reconciled. Therefore, the allocations are corrected to remove this element.

Meter error adjustment data is received on an LDZ by LDZ basis split by billing month. The total value of the error is given, and this is split so that the correct proportion of each meter error can be assigned to each gas year in which the error occurred.

These errors affect the Aggregate LDZ Load and the DM/PC1 Load, and have opposite effects on the allocation total, which is calculated at the gas year level of granularity. The result of applying corrections for the meter errors is as follows:

- LDZ meter under-reads *increase* the total NDM allocation
- LDZ meter over-reads *decrease* the total NDM allocation
- DM/Unique site meter under-reads *decrease* the total NDM allocation
- DM/Unique site meter over-reads *increase* the total NDM allocation

7.2 NDM Consumption Calculation

The consumption algorithm relies on a large quantity of data, summarised in Section 6.2. A full description of the raw data used to calculate consumption figures for each individual meter point is described in Appendix A. This raw data is then pre-processed to validate it and to derive additional information to help speed up the consumption calculation process, (details can be found in Appendix B). After the pre-processing the main algorithm is run to calculate consumption on a meter by meter basis. This calculation will not be successful in all cases, so a final step is required to scale up the consumption estimate to account for these 'failed' sites.

7.2.1 Algorithm

Figure 18 shows a flow chart of the process involved to calculate the consumption for a single meter and gas year with references to numbered steps, which are described in detail below.

Note, most gas years calculated are from the pre-Nexus period. Therefore, the methodology uses SSP/LSP market sector to determine the best method to use. This will be revisited as more post-Nexus data is used in future years to determine what updates are needed.

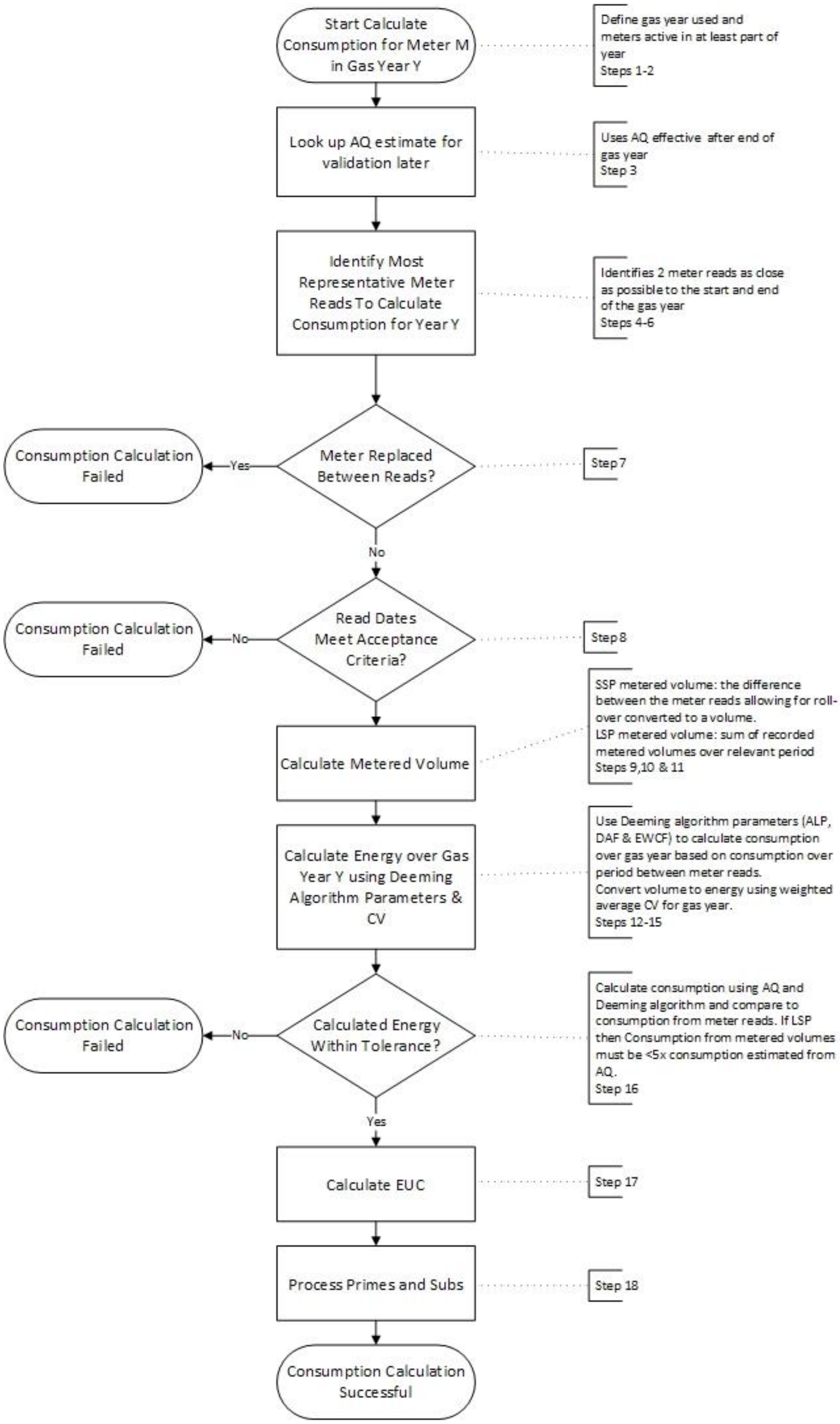


Figure 18: Consumption Algorithm Flow Chart

1. Given a gas year Y, define the start and end dates as 01 Oct Y and 30 Sep Y+1
2. Find all meter points that are not on a CSEP (as reads aren't available during the Pre-Nexus training period) and were active and NDM in a least part of year Y.
3. Look up the first AQ estimate effective after the end of the gas year. If none exists after the end of the gas year use the latest value. From this record store
   i. The AQ value
   ii. The EUC provided by the CDSP
   iii. The pre-calculated consumption band derived by the AUG Expert from the AQ value.
   iv. Market sector (SSP/LSP) based on the EUC from the CDSP
4. For each meter point find the meter reading date and value for:
   - LB1 (Lower Bound 1) – the latest meter reading prior to the start of the gas year
   - LB2 (Lower Bound 2) – the earliest meter reading within the gas year
   - UB1 (Upper Bound 1) – the latest meter reading within the gas year
   - UB2 (Upper Bound 2) – the earliest meter reading after the end of the gas year

   For SSPs those readings which have been flagged as bad by the pre-processing are excluded.

   Where a meter point has changed between NDM and DM or vice versa try to select meter reads from the period when it was NDM.

   Note that for any given meter point, only a subset of this full set of reads may be available. At least one lower bound and one different upper bound meter read are needed. Possible scenarios are shown in Figure 19 below:

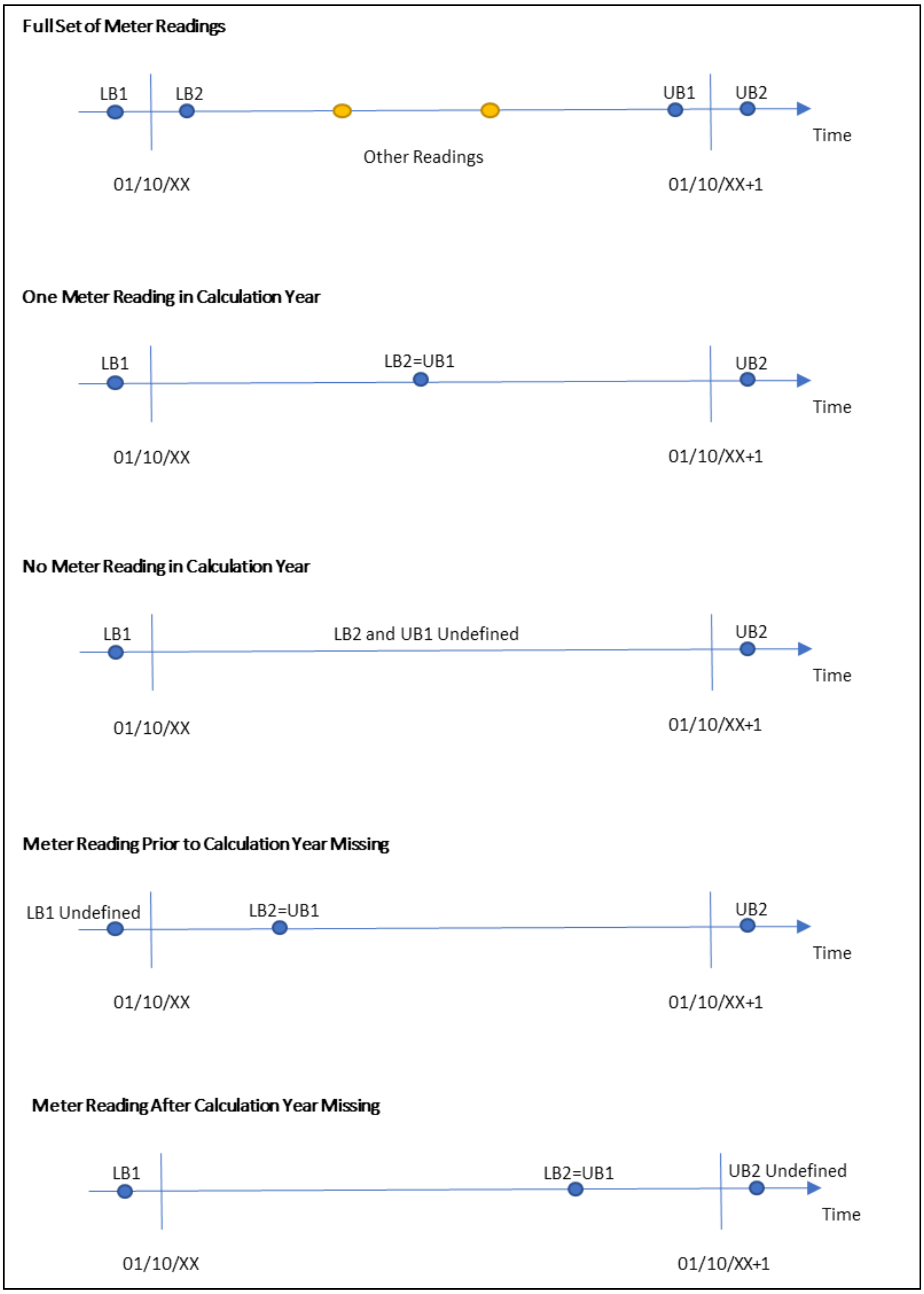


Figure 19: Meter Read Availability Scenarios

5. Set the start meter read date to LB1 unless

    A. the date of LB1 is more than 365 days from the start of the gas year, or

    B. the meter was replaced on or after LB1 and before LB2

    In which case set it equal to LB2.

6. Set the end meter read date to UB2 unless

A. the date of UB2 is more than 365 days from the end of the gas year, or

B. the meter was replaced after UB1 and on or before UB2

In which case set it equal to UB1.

7. If the meter is SSP and was replaced between LB2 and UB1 inclusive, then reject the meter point.

8. Check that:

A. The distance between the two chosen meter readings is at least 170 days

B. The overlap between the metering period and the gas year is at least 90 days

If this is true then proceed to calculating the metered volume, otherwise reject the meter point.

9. Apply either Rule A or Rule B depending on the market sector of the meter point:

A. If the site is SSP then calculate the volume consumed between the two chosen meter readings (mr_1, mr_2). If this gives a negative volume, then check if the meter index has rolled over (see subsection below).

B. Otherwise sum the metered volumes (mv_i) and volume corrections between the two chosen meter readings. If there are any negative volumes in the range, set the sum to -1. If any gaps or overlaps are present between the chosen reads the volume is also set to -1.

If this step produces a positive volume then proceed to the next step, otherwise reject the meter point.

10. Calculate the fraction of the year that the meter point was active and NDM weighted by the WAALPs. Note, currently the pre-Nexus definition of WCF is used to calculate the WAALPs.

11. Calculate the volume taken over the gas year (or fraction of year calculated in the previous step) by multiplying the volume from Step 9 by

$$\frac{\sum_{\substack{\text{Formula Year}\\ \text{or Part Thereof}}} WAALP^v}{\sum_{\substack{\text{Metered}\\ \text{Period}}} WAALP^v}$$

where $WAALP^v$ is the WAALP divided by the relevant CV value (i.e. a 'volume' WAALP rather than the usual energy WAALP).

12. Look up, in the meter asset information, whether the meter is/was metric or imperial and then apply either Rule A or Rule B to match the rule chosen in step 9.

A. If the meter point is SSP look up the read units (*U*).

- First choice is the units inferred from the meter read records.
- If this could not be calculated, then use the units provided by the CDSP.
- In the case where the read units from the CDSP are obviously wrong (i.e. are 0 or not a power of 10) use 1 for metric and 100 for imperial meters.

Combine this value with the default correction factor (*CF*) 1.022640 and relevant metric/imperial conversion factor to get a combined conversion factor.

B. Otherwise, if LSP no further information is required.

If no meter asset information can be found, reject the meter point.

13. Calculate the weighted average CV for the gas year, calculated as

$$\frac{\sum_{\substack{\text{Formula Year} \\ \text{or Part Thereof}}} WAALP}{\sum_{\substack{\text{Formula Year} \\ \text{or Part Thereof}}} WAALP^{v}}$$

14. Convert the gas year volume to energy in kWh by multiplying the output of Steps 11, 12 and 13 together. In summary, depending on the market sector of the meter point, this will be

$$Con = (mr_2 - mr_1) \times U \times CF \times \frac{CV}{3.6} \{\times 0.0283168466 \; if \; imperial\} \quad \text{for SSP}$$

$$Con = \sum mv_i * CV/3.6 \quad \text{for LSP}$$

15. Calculate an AQ from this consumption using the appropriate Cumulative Weather Adjusted Annual Load Profile (CWAALP)

$$AQ = Con * 365 / CWAALP$$

16. If a new AQ value has been calculated from the meter readings which is more than four times larger than the old AQ and the new AQ puts the site in the LSP market, then reject the meter point. If the old AQ is 1 then use four times the largest recorded AQ as the check instead.

17. If the consumption calculation was successful, calculate an EUC band based on the new AQ.

18. Carry out post-processing to avoid double counting of subs and deduct consumption. See subsection 7.2.1.2 below for details.

7.2.1.1 Meter Index Rollover Check

Given two reads mr_1 and mr_2 where (mr_2 - mr_1) < 0 the following process is used:

1. Estimate the number of dials from mr_1

$$num\_dials = max(ceil(log_{10}(mr_1)), 4)$$

2. Determine the maximum possible meter read

$$max\_read = 10^{num\_dials}$$

3. Calculate the period between the two meter reads in years

$$num\_years = \sum_{mr_1(date)+1}^{mr_2(date)} ALP / 365$$

4. Assume meter index roll-over and re-calculate the volume

$$tmp_1 = max\_read - mr_1 + mr_2$$

5. Calculate the new volume as a fraction of the max read per year

$$tmp_2 = (tmp_1 / max\_read) / num\_years$$

6. If tmp_2 < 0.25 then assume the meter index has rolled over and use tmp_1. Otherwise leave the calculated volume as negative and reject the meter point.

7.2.1.2 Prime and Sub Meter Post Processing

As the prime meter consumption is the difference between the total consumption (based on the prime meter reads) minus the sum of the sub-meter consumptions, issues can arise in cases where a full valid set of consumptions for all meters within a sub-prime configuration are unavailable. Note that the Consumption Methodology will not calculate consumption for a DM meter. There are four cases to consider:

1. If the prime meter is DM, no action is necessary as the methodology won't have calculated consumption for the prime meter (consumption not required for DM meters). Sub-meters will be calculated correctly based on available data.
2. If the prime meter is NDM and contains one or more DM sub-meters, then the prime meter consumption calculation is flagged as having failed so that an EUC average consumption is used (see 7.2.2).
3. If the consumption calculation fails for any of the sub-meters, then the prime meter calculation is flagged as having failed. An EUC average consumption is therefore used for the prime meter.
4. If the consumption calculation succeeds for the prime meter and all its sub-meters, then the prime meter consumption is calculated by subtracting the sub-meter consumptions from the total prime meter consumption.

Prime and sub meter arrangements may be disaggregated so data was requested from the CDSP to track the reconfirmation of these meters. Using this information, the necessary correction is made only for the relevant period.

7.2.1.3 Interpolation of Failed Consumptions

In the case where the consumption calculation fails for a meter point in a specific year but is successful for the two years either side, the average of the two successfully calculated consumption values is used.

Two validation checks are carried out on this process:

- In line with the standard consumption calculation the interpolated value must be less than four times the prevailing AQ value. This is unlikely to be an issue given the consumptions either side must have passed this check.
- The consumption values either side must differ by less than 40%. Otherwise the meter point's behaviour is changing significantly, and a simplistic linear model is unlikely to be appropriate.

7.2.2 Aggregation and Scaling-Up

When applied to each meter point in any given LDZ, the algorithm outputs a set of consumptions that can be aggregated to EUC level. The aggregated data for each EUC is also naturally split into the following categories by the algorithm:

- Meters for which a consumption could be calculated
- Meters for which the algorithm failed (failed to calculate consumption or calculated consumption failed validation)
- Meters in CSEPs (for which Pre Nexus meter reads are not available)

The sum of these three categories across all EUCs gives the total NDM population of the LDZ.

Where a consumption value was successfully calculated the EUC is based on this consumption, otherwise it is calculated by the AUG Expert based on the AQ.

Therefore, for each EUC band the following can be calculated:

1. The number of meter points with a successfully calculated consumption.
2. The number of meter points without a calculated consumption (i.e. calculation failed and interpolation not possible).
3. The average consumption for those meter points with a calculated consumption greater than zero.

The values for 3) are then used to estimate the consumption for meter points in 2). This involves several subtleties:

- In 3) attention is restricted to consuming meters only, in order to account for potential differences in the proportion of non-consuming meters within and outside the sample.
- Meter points where the consumption calculation fails are classified as consuming/non-consuming based on AQ, as this is the only reliable data available for such meters. It is recognised that due to changing circumstances for each meter, those with an AQ of 1 for Year X are not necessarily non-consuming during Year X. Likewise, those with an AQ greater than 1 for Year X are not necessarily consuming in Year X. Therefore, two figures have been calculated using available information (i.e. meters within the sample):
    - the proportion of meters with AQ = 1 for Year X that are consuming in Year X = A
    - the proportion of meters with AQ > 1 for Year X that are consuming in Year X = B
- The consumption for the non-calculated meter points is then calculated as

  Consumption = A x (meters with AQ = 1) x "AQ=1" average consumption

  + B x (meters with AQ > 1) x EUC average consumption

  Where:

  - *"AQ=1" average consumption* is the average consumption of meter points where the AQ=1, but our consumption estimate is greater than zero. This can arise when an AQ review produces AQ=1 yet the period of consumption being validated is actually non-zero.
  - *EUC average consumption* is the average consumption for successfully calculated meters in the corresponding EUC Band. The 01B EUC average excludes meters where AQ=1.

- Pre-Nexus CSEPs are treated differently to failed meters. This is because meter points are assigned to EUC band based on their maximum potential AQ which may not be the same as their current AQ. It is not appropriate to estimate their consumption using the number of meter points in each EUC band multiplied by the EUC band average consumption. The approach used is described in detail in Section 7.2.2.1.
- Where the sample size for a particular EUC for a given LDZ and gas year is less than 30 the national average is used in place of the LDZ average.
- Failed meter points which were only active for part of the year are assigned an average demand scaled based on the sum of WAALPs for that part of the year.

Figure 20 below summarises the process for obtaining a consumption value for each type of meter point.

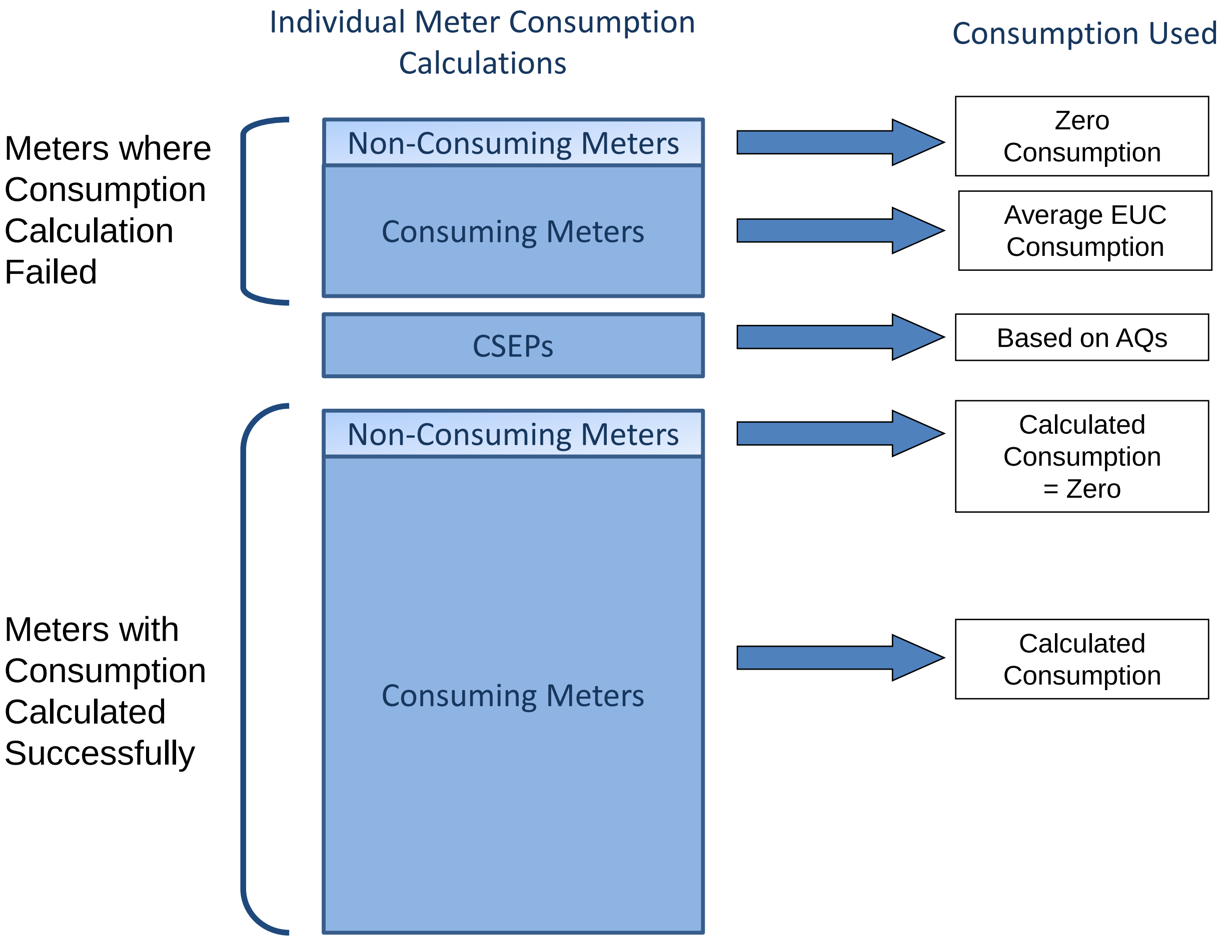


Figure 20: Consumption Method used for each type of Meter Point

Unidentified Gas for the LDZ for the gas year in question is then calculated by summing the metered NDM consumptions across all EUCs and subtracting these from the total combined allocations for the same period.

It is important to note that at this stage these figures still include Temporary Unidentified Gas. Therefore, whilst giving an indication of the order of magnitude of the Permanent Unidentified Gas total for that historic year, this is simply a step in the calculation process and not an estimate of the final value. The method for calculating the remaining Temporary elements is defined in detail in the relevant subsections below.

7.2.2.1 CSEP Consumption Calculation

The latest year used in the consumption calculation (2016/17) contains some post-Nexus data. However, there is insufficient Post-Nexus data at this stage to calculate CSEP consumption at a meter point level. The AUGE expects to be able to work at individual meter point level from gas year 2018/19 onwards. The 2016/17 gas year is therefore treated in the same way as earlier years in terms of the CSEP consumption calculation.

The following steps are used in the calculation of CSEP consumption for gas year Y. The steps are carried out separately for each LDZ and EUC for each gas year. All references to AQ mean the CSEP AQ for the corresponding LDZ and EUC unless stated otherwise.

CSEP AQ and supply point count are provided by the CDSP from invoicing files. The data therefore aligns to the start of each gas year (1 Oct). However, the AUG Expert has identified significant anomalies in the data for gas years 2014 and 2015. To overcome this issue, the AUG Expert has used the CSEP data from 2013 for both the 2013 and 2014 gas years. The 2016 CSEP data has been used for both the 2015 and 2016 gas years.

1. If the number of meter points in the CSEP is less than 100 in year Y then the CSEP consumption estimate for year Y is the aggregate AQ for year Y

   $$C_Y = AQ_Y$$

   Otherwise,

2. Calculate the average AQ for new meters

   For EUC01B before gas year 2012,

   $$NMAQ = AQ_{2007} / N_{2007}$$

   For EUC01B for gas years 2012 and 2013,

   $$NMAQ = AQ_{2012} / N_{2012}$$

   For all other EUCs and EUC01B from gas year 2014 onwards,

   $$NMAQ = AQ_{Y+1} / N_{Y+1}$$

   Prior to 2012, the CSEP NExA table had not been updated since 2007. A new table was calculated for 2012 and this is now updated annually following IGT 053 which was implemented from 1 October 2014 [13]. This removes the additional bias which previously existed for 01B meters as a result of using default AQ values.

3. Calculate the average AQ for lost meters

   $$LMAQ = AQ_Y / N_Y$$

4. Estimate the aggregate AQ for year Y by adjusting Y+1 for meter changes

   $$AQ'_Y = AQ_{Y+1} - Max(0, (N_{Y+1} - N_Y) * NMAQ) - Min(0, (N_{Y+1} - N_Y) * LMAQ)$$

5. Adjust the aggregate AQ to allow for the recalculation success rate

$$AQ''_Y = AQ_Y + (AQ'_Y - AQ_Y) / RR$$

where

AQ_Y is aggregate AQ in year Y. Note that for 2009, AQ_Y is adjusted to the new SNCWV definition.

AQ'_Y is the aggregate AQ in year Y+1 adjusted to allow for the different number of meters to year Y (as in step 4 above)

RR is the recalculation success rate expressed as a fraction

6. Estimate consumption for year Y by adding effect of new/lost meters

$$C_Y = AQ''_Y + Max(0, (N_{Y+1} - N_Y) * MAQ_{Y+1}) * YFrac + Min(0, (N_{Y+1} - N_Y) * MAQ_Y) * (1-YFrac)$$

where

C_Y is the final estimate of aggregate CSEP consumption for year Y

AQ''_Y is the estimate of aggregate AQ for year Y calculated from year Y+1 and adjusted for the recalculation success rate

N_Y is the number of meters in year Y

MAQ_Y is the average AQ per meter in year Y and is calculated as AQ_Y / N_Y

YFrac is an estimate of the proportion of a year's consumption which new meters contribute. It is assumed that new and lost meters will be consuming on average for half of the year so a default factor of 0.5 is used.

If the consumption estimate is negative, then use the AQ as the best estimate of consumption i.e. $C_Y = AQ_Y$.

7.3 Unregistered and Shipperless Sites

The magnitude of every Unregistered and Shipperless category of Unidentified Gas is affected by several Mods introduced between 2013 and 2014 to address and reduce these specific areas of Unidentified Gas. The Mods in question are as follows:

- Mod 0410A [9] applies to Shipper Activity, Orphaned and Unregistered <12 Months, and any site created on or after 01/09/2013 is subject to the terms of this Mod.
- Mod 0424 [7] applies to Shipperless PTS, and any site with an isolation date on or after 01/04/2013 is subject to the terms of this Mod. This results in all Unidentified Gas from post-Mod sites being Temporary rather than Permanent.
- Mod 0425 [8] applies to Shipperless PTS, and any site with an isolation date on or after 01/04/2014 is subject to the terms of this Mod. In theory, this should have resulted in all Unidentified Gas from post-Mod sites being temporary rather than Permanent, but the process required to implement this was not available in the pre-Nexus system. Therefore, all Unidentified Gas from post-Mod sites is only Temporary after the Nexus go-live date.

Note that Mods 0424 and 0425 apply to the PTS and SSrP elements of the "Awaiting GSR Visit" category in addition to their main categories.

A full description of the calculation method for all Unregistered and Shipperless Unidentified Gas categories is given below. Raw data for all of these except "Awaiting GSR Visit" is contained in snapshot files supplied by the CDSP every month. These are described in Section 6.3 above.

The following files also contain data that is used in the calculation process and are supplied on an annual basis.

- **Connection Details for Orphaned Sites**
  This dataset includes asset and Shipper meter reads and information on whether the confirming Shipper is the same as the asset Shipper. This data is used to determine the proportion of sites that have been flowing gas prior to becoming registered and the proportion of these that can be back-billed. Backbilling can only occur if the confirming Shipper is the same as the Shipper that carried out site works. The Connection Details dataset is split into two categories (pre-Mod 0410A sites and post-Mod 0410A sites) and different flow proportions and backbilling proportions are calculated for each. This is necessary because the terms of Mod 0410A affect how Unregistered sites are processed, and this leads to different conditions for pre-Mod and post-Mod sites.
- **Gas Safety Regulations Visit Details**
  The gas safety visit data is used to estimate the number and AQ of sites that have been recorded as isolated for less than 12 months and hence have not yet had their GSR visit and do not yet appear in the snapshots as Shipperless PTS or Shipperless SSrP, but are nevertheless still consuming Shipperless gas.
- **Shipperless Sites Supporting Data**
  This dataset contains the confirmation date of each Shipperless site that has appeared in any snapshot but has subsequently been (re)confirmed. It is used to ascertain the final outcome for each of the sites, i.e. whether it was (re)confirmed or whether it was disconnected. This is used to determine the proportion of Shipperless sites that have a meter and are *capable* of flowing gas that actually *are* flowing gas.

Further details of these data files are also given in Section 6.3 above.

The step by step calculation process for Shipperless and Unregistered Unidentified Gas is as follows:

1. Each MPRN in the snapshot files is assessed and flagged for further investigation by the CDSP if any of the conditions specified below are satisfied.
   - If a graph of AQs sorted by descending magnitude contains a "shoulder" point (i.e. a distinct change in gradient), any points to the left of the shoulder are flagged.
   - Any site with an AQ more than 100 times the average AQ for EUCs 02B-09B is flagged.
   - Any site with an AQ greater than 58.6 GWh is flagged.

   The resultant list of flagged sites is sent to the CDSP.
2. The CDSP respond with details where any of the flagged sites have been confirmed on their system, and the confirmed AQ of each such site is provided. Any differences between the queried AQs and the confirmed AQs are applied to the snapshot files. Sites where the CDSP have no further information are left as is.
3. All sites with a listed AQ above the VLDMC threshold (1.465 TWh) have their AQs replaced with the average EUC 02B-09B AQ. VLDMCs cannot be Unregistered or Shipperless due to the greater scrutiny the network code requires on such sites, and hence any AQ above this threshold in the

Unregistered or Shipperless lists must be erroneous (e.g. MPRN or phone number accidentally entered in AQ field).

4. Before the analysis is run, the following coefficients are also calculated using the latest available data:
    - Fraction of opening meter reads with gas flow for Unregistered sites (for the Permanent/Temporary split for Unregistered Unidentified Gas categories, with different fractions for pre-01/09/2013 and post-01/09/2013 sites).
    - Fraction of Unregistered Unidentified Gas not backbilled (for the Permanent/Temporary split for Unregistered Unidentified Gas categories, with different fractions for pre-01/09/2013 and post-01/09/2013 sites).
    - Proportion of Shipperless sites being disconnected rather than re-registered.
5. "Fraction of opening meter reads with gas flow" is calculated using the "Connection Details for Orphaned Sites" spreadsheet. This file contains a list of Orphaned meters and includes both their asset meter reading and their opening Shipper meter reading. The number of meters with gas flow (i.e. those where the reading has changed) is expressed as a proportion of the total number of meters in the sample. The dataset is split into two sections - pre-Mod 0410A sites and post-Mod 0410A sites – and separate factors are calculated for each in order to account for changes introduced in the Mod. The calculated proportions are applied to the AQs of each pre-Mod 0410A and post-Mod 0410A site with a meter in the snapshots, to give an estimate of the consumption from sites that are actually flowing gas in the Unregistered Unidentified Gas calculations.
6. "Fraction of Unregistered Unidentified Gas not backbilled" is also calculated using the "Connection Details for Orphaned Sites" spreadsheet. In addition to the meter readings, this file contains a flag that indicates whether the asset Shipper is the same as the confirming Shipper. This flag is used to calculate the proportion of sites with gas flow (as calculated in Step 5 above) that also have a different Shipper. This is the proportion of Unregistered sites that cannot be backbilled and hence contribute permanent Unidentified Gas. As for "Fraction of opening meter reads with gas flow", separate factors are calculated for pre-Mod 0410A and post-Mod 0410A sites.
7. The proportion of Shipperless sites that are disconnected rather than reconfirmed is calculated using information from the "Shipperless Sites Supporting Data" spreadsheet. Any site that disappears from the Shipperless lists without appearing in the "Confirmed" list in the supporting data has been disconnected and this is used to calculate the proportion that are disconnected rather than being reconfirmed. This figure is used as the best estimate of the proportion of sites capable of flowing gas that actually *are* flowing gas (i.e. it is assumed that if a site is flowing gas it is reconfirmed, and if it is not it is disconnected).
8. The raw Shipperless/Unregistered Unidentified Gas calculations are now carried out. This is carried out in a series of spreadsheets, with a set of spreadsheets for each individual Unidentified Gas category. Each of these sheets contains a full history of all available snapshot data for the relevant Unidentified Gas category, which at present runs from September 2011 to September 2019. The availability of snapshot data over such a period of time means that trends can be identified within each Unidentified Gas category and extrapolated to cover training years and the forecast year as necessary.
9. Every Shipperless and Unregistered category is affected by at least one Mod, and the implementation of each of these Mods therefore affects the magnitude and trend of the Unidentified Gas category it refers to. In addition, once the Mod has been implemented this creates a division of each category

into pre-Mod and post-Mod sites: only post-Mod sites are affected by the Mod, and hence pre- and post- sites will now behave in different ways and exhibit different trends over time. Hence, they must be analysed separately. In addition, the rules for Unidentified Gas being temporary or permanent can be different before and after the Mod and this must also be taken account of. This creates a set of four sub-categories of Unidentified Gas for each main Unidentified Gas category as follows:

- Pre-Mod permanent
- Pre-Mod temporary
- Post-Mod permanent
- Post-Mod temporary

The trends for each of these must be assessed individually. For each one, the standard 36-way split also exists (9 EUCs and 4 Product Classes), and it can also be seen from preliminary analysis of the data that there is an LDZ-by-LDZ effect where different LDZs can show different patterns in Unidentified Gas consumption. This creates a total of 1872 individual sub-categories for every main Unidentified Gas category (pre- or post-Mod/permanent or temporary/LDZ/EUC/Product).

The trend for each of these must be individually assessed in order to allow accurate calculation of Unidentified Gas for any time period covered by the snapshots, and accurate extrapolation to any required years that fall outside this range (e.g. the forecast year).

Finally, the introduction of each Mod and the reaction of the industry to it necessarily affects these trends, which will therefore not remain constant over the entire time period covered by the snapshots. Therefore, for each main Unidentified Gas category, the snapshot period is split into a small number of separate sub-periods where the trend is consistent: these differ for each Unidentified Gas category due to the timing and nature of each Mod. Individual trends are therefore calculated for each sub-period and apply to specific time periods only.

10. The Unidentified Gas estimate for each year of the training period and for the forecast period is calculated as follows:

- Where the training year lies within the snapshot period, the fitted trend for that year (calculated using the appropriate sub-period) is used.
- For the forecast year and where the training year lies outside the snapshot period, the appropriate fitted trend is extrapolated to cover the year in question. These extrapolated values are used for the calculation.
- Each point in the trend is the expected AQ of Unidentified Gas from that source at that particular point in time. Therefore, the best estimate of the Unidentified Gas consumed in Year Y is the average of the monthly AQ figures across the whole of the year.

Figure 21 below shows an example of the piecewise fit approach for one Unidentified Gas sub-category (permanent Pre-Mod Orphaned sites, Product 4, EUC 01B, EA LDZ).

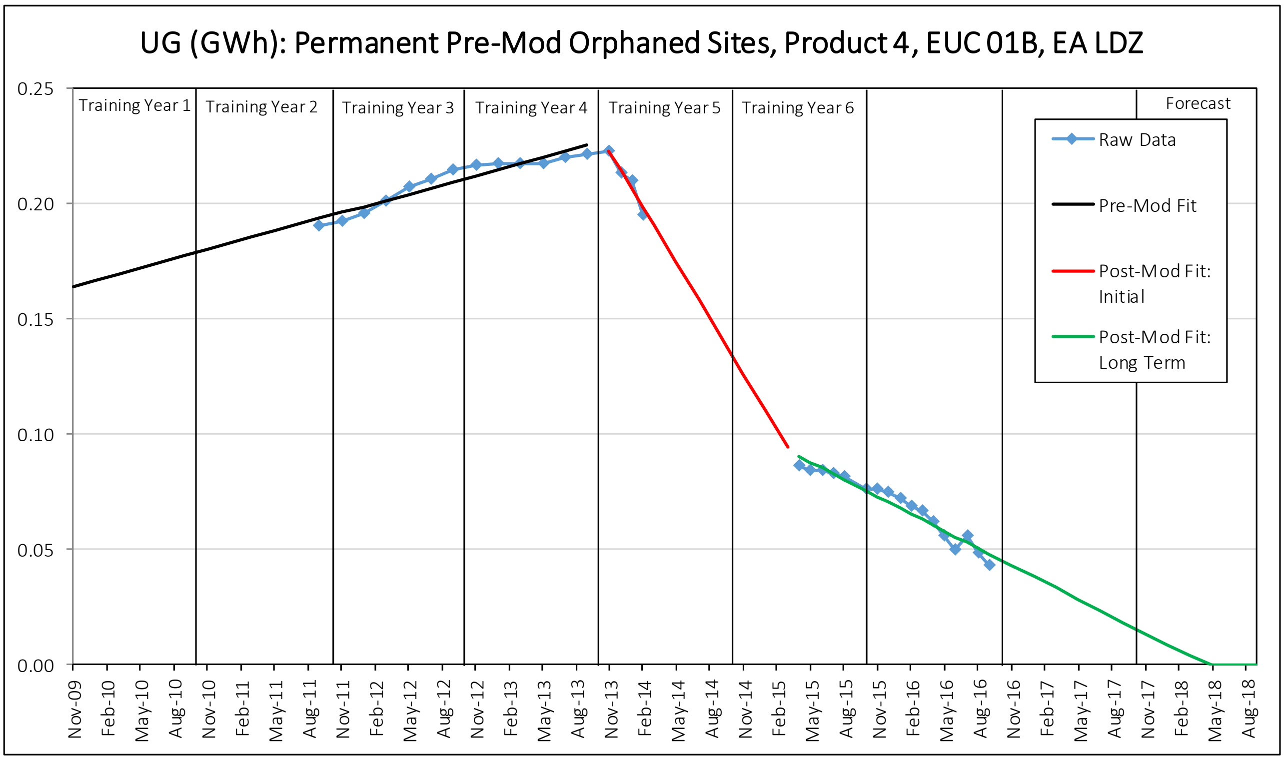


Figure 21: Piecewise Trend for Unregistered Unidentified Gas, Training and Forecast

11. Unregistered sites may or may not have a meter fitted. Where no meter is present, it is assumed that consumption will be zero. For meters in the Shipper Activity and Orphaned categories, the snapshot files contain data split into meter points with and without a meter present. Consumption for these categories is therefore calculated as described above only for meter points where a meter is actually known to be present. For the Unregistered <12 Months category, it is not recorded whether a meter is present or not. For these sites, it is therefore assumed that the fraction of meter points where a meter is present is the same as that found across the other two Unregistered categories.

12. The Unidentified Gas estimate for each type of Unregistered site is adjusted to account for the proportion of such sites with meters that actually flow gas whilst Unregistered, as described in Step 5 above. This adjustment is applied to the annual Unidentified Gas values calculated from the piecewise trends described in Step 10.

13. For Shipperless sites, the snapshots contain all sites found to be *capable* of flowing gas at the GSR visit. From these, the number *actually* flowing gas must be estimated. This is carried out using data in the "Shipperless Sites Supporting Data" file supplied by the CDSP. This contains the confirmation details of all sites that have appeared in the Shipperless report but have subsequently been confirmed and is used to determine the proportion of sites that were reconfirmed rather than being disconnected. The following assumptions are then made:

    - If the site was disconnected it was not flowing gas
    - If the site was reconfirmed it was flowing gas

    The reconfirmation percentage is therefore applied to the Shipperless Unidentified Gas totals (calculated as described in Step 10) to give the best estimate of the amount of Shipperless Unidentified Gas actually consumed.

14. For each Unregistered category, factors are used to convert from Requested AQ to Confirmed AQ and then from Confirmed AQ to AQ Following Review, as follows:

Requested AQ → Factor 1 → Confirmed AQ → Factor 2 → AQ Following Review

The "AQ Following Review" figure is regarded as a reliable indicator of the annual consumption at the site.

15. The Unidentified Gas estimates produced for each Shipperless/Unregistered category are multiplied by the appropriate combination of these factors. This is carried out as follows:
    - Shipperless sites (PTS, SSrP): no adjustment required for AQ bias
    - Unregistered (Orphaned, Shipper Activity and Unregistered <12 Months): adjust using composite *Factor1(n) x Factor2*, where *n* represents the Unidentified Gas category in question.
16. The resultant values are now split by Product Class using the principles defined in Section 4.3.3 above. For all Shipperless and Unregistered Unidentified Gas categories the *forecast year* population split by Product Class is also applied to the training period to remove the nuisance factor of changes in the underlying population and allow consistent trends to be created for extrapolation to the forecast year.
17. The above process allows Unidentified Gas estimates for all Unregistered and Shipperless Unidentified Gas categories apart from "Awaiting GSR Visit" to be calculated. At the time of calculation these are split by:
    - Unidentified Gas Category
    - Permanent and temporary
    - Pre-Mod and Post-Mod
    - Year
    - EUC
    - Product Class

    For use in further Unidentified Gas calculations, the estimates must be split by permanent / temporary, Year, EUC and Product Class and so figures are aggregated across categories and pre- or post-Mod status. Tables of figures split this way feed into the final UIG factor calculations.

18. "Awaiting GSR Visit" Unidentified Gas is calculated using Gas Safety Regulations visit data supplied by the CDSP. This file contains the details of each Shipperless site that has crossed the 12-month threshold during a period of a year and has subsequently been visited and found to be capable of flowing gas. The actual sites listed in this file by definition appear in the summarised data in the snapshot files because they have been Shipperless for more than 12 months. If it is assumed that sites become Shipperless at a steady rate, however, it can be assumed that the number and AQ of sites crossing the 12-month threshold in Year Y is a good approximation of the number and AQ that will cross in Year Y+1. At the end of Year Y these sites will have been (recorded as) Isolated for less than 12 months and hence make up the "Awaiting GSR Visit" Unidentified Gas category for this year. Therefore, in order to estimate the Unidentified Gas from this category, the AQs for each site from the GSR visit data are analysed and split by:
    - Permanent and temporary
    - Year
    - EUC
    - Product Class

The transient nature of this Unidentified Gas category means that there is no requirement to define sites as pre-Mod or post-Mod in this case. Whilst all sites in this category will go on to become either Shipperless PTS or Shipperless SSrP sites and hence are affected by Mods 0424 and 0425 (subject to the delay described above), all sites move on to their destination categories in a year and hence no legacy pre-Mod sites remain.

The permanent/temporary split for each year is defined by the implementation date of the relevant Mod, as follows:

- For the PTS element, all Unidentified Gas from this source is permanent up to the 2011/12 gas year, and all temporary from 2014/15 onwards. If a steady influx of sites is assumed for the remaining years, this results in 75% of Unidentified Gas being permanent for 2012/13 and 25% permanent for 2013/14.
- For the SSrP element, all Unidentified Gas from the pre-Nexus period is permanent due to the delay in implementation of back-billing. Following Project Nexus go-live, any site without a record of the installation date of the new meter still contributes permanent Unidentified Gas until its GSR visit regardless of the Isolation date. It is assumed that new meters are installed at a steady rate for these sites and as such the average period of the year for which a meter was present and capable of flowing gas is 6 months (i.e. 50% of the year). The following calculation steps are applied:
    - Multiply the Shipperless SSrP Unidentified Gas element by the "% of Year with Meter" factor. This gives the total (permanent plus Temporary) Unidentified Gas for the Shipperless SSrP element of "Awaiting GSR Visit".
    - Multiply this total by the "% of Meters without Install Date" factor. This gives the Permanent element of the Unidentified Gas total from this source. The remainder of the Unidentified Gas for the Shipperless SSrP element of "Awaiting GSR Visit" is Temporary.

Finally, the reconfirmation percentage (as described in Step 13 above) is applied to convert from the AQ of sites *capable* of flowing gas to the AQ of sites *actually* flowing gas and the data is split by Product Class (as per Step 16 above). This gives the final total Unidentified Gas estimate for this category, split by permanent/temporary, Year, EUC and Product Class.

The Unidentified Gas from this category is now taken through into the final UIG factor calculations.

7.4 IGT CSEPs

Connected System Exit Points (CSEPs) are typically small networks owned by Independent Gas Transporters (IGTs) that connect to the GTs' systems. They are often new housing estates, where the gas network for the estate has been built and is owned by an IGT. CSEPs can potentially contribute to Unidentified Gas where either sites within them or entire IGT networks are not recognised by the CDSP's system and are thus consuming gas in an unrecorded manner.

7.4.1 Overview

Unidentified Gas from CSEPs arises from two sources: Unknown Projects and Unregistered sites on known CSEPs.

Unknown Projects are CSEPs that are known to exist but for various reasons are not on the CDSP's systems. Regular meetings are held between the IGTs and the CDSP in order to resolve these issues and reduce the number of Unknown Projects.

Unregistered sites on known CSEPs lie in CSEPs that are on the CDSP systems, and the CDSP are notified of such Unregistered sites on them.

For both these sources of Unidentified Gas from CSEPs, in the pre-Nexus market the Unidentified Gas from LSP sites was backbilled and therefore temporary, whilst Unidentified Gas from SSP sites on CSEPs was not backbilled and therefore permanent. Under Project Nexus, all sites within CSEPs are individually recorded on the CDSP's systems and can hence be backbilled. Therefore, all CSEP Unidentified Gas from this source is temporary under Nexus.

It is necessary to calculate both the permanent and temporary elements of IGT CSEPs Unidentified Gas for the training period, whilst only permanent is required for the forecast year (although as described above, this is now zero). This allows temporary Unidentified Gas to be removed from the raw Unidentified Gas total (permanent plus temporary) calculated using the Consumption Method for the training period, leaving the final total as permanent Unidentified Gas only. In previous analysis years, the permanent element of Unidentified Gas was then extrapolated to the forecast year to facilitate the calculation of the UIG factors, but this is no longer required.

7.4.2 Data

Before Nexus go-live, *Unknown Projects* data was supplied by the CDSP in monthly snapshot files. These contained data for all Unknown Projects, split by LDZ and by the year in which the CSEP first came to the attention of the CDSP. For each LDZ, the total number of projects, the total number of supply points within them, and the sum of their AQ was given. As noted elsewhere in this document, since Nexus go-live the CDSP no longer has access to this information, and a data request has been sent to the IGTs as an alternative data source.

Unregistered sites on known CSEPs data is supplied in a file provided by the CDSP on request (typically 2-3 instances per year). This file contains data for all known CSEPs, summarised to the LDZ and EUC level. For each EUC, the count of supply points within CSEPs and their aggregate AQ is given. This is provided in separate tables for registered sites and Unregistered sites.

Data for registered sites is used to calculate the average CSEP throughput percentages by EUC for each LDZ, and this is used to split the Unknown Projects data by market sector.

Data for Unregistered sites is used to directly calculate the Unidentified Gas from this source for each LDZ.

It should be noted that the supply point count data for Unregistered sites is actually the number of times the CDSP have been notified that the supply point is Unregistered rather than the number of sites that are actually Unregistered. The CDSP are often notified about the same site on multiple occasions, and this artificially inflates the supply point count figures in this dataset. Therefore, further analysis is carried out on this data in order to estimate the actual number of Unregistered supply points. This is described in more detail below.

7.4.3 Process

Processing is carried out in spreadsheets, which are available to the industry on request to allow auditing of the AUG Expert's calculations to take place. The following process steps are performed:

1. Data for *registered sites on known CSEPs* and *Unregistered sites on known CSEPs* is imported into the calculation spreadsheets. The average AQ per site for each LDZ/EUC combination for registered sites is calculated. As noted above, for the Unregistered data the number of notifications is recorded rather than the number of sites, and hence the actual number of sites must be estimated.
2. The average AQ per site derived from the registered sites is used to estimate the number of Unregistered sites in each EUC using the aggregate AQ for each EUC in the Unregistered dataset. This gives an estimated number of Unregistered sites in each EUC under the assumption that each site has the average AQ for that EUC.
    - If this calculated figure is lower than the number of notifications, it is used as the best estimate of the number of Unregistered sites in that EUC.
    - If the number of notifications is lower, this is used as the best estimate of the number of Unregistered sites in that EUC.
3. The total site count and aggregate AQ by EUC and by LDZ is calculated for registered sites. These figures are then used to calculate the percentage split of CSEP site count by EUC and the percentage split of CSEP AQ by EUC. This split is used in the calculations for Unknown Projects, described below.
4. The total site count and AQ by EUC and by LDZ is calculated for Unregistered sites. The AQ figures produced are split by Product Class using the principles defined in Section 4.3.3 and are used directly in the Unidentified Gas figures: these represent the estimated annual contribution to Unidentified Gas from Unregistered sites in known CSEPs based on current conditions.
5. When each new monthly snapshot file becomes available, data for Unknown Projects is updated. An example of the snapshot table format is shown in Table 7 below. In these tables, the "Year" field refers to the year in which the CSEP came to the attention of the CDSP. For each LDZ the total number of Unknown Projects, their aggregate AQ and the total number of supply points within them is given. Each snapshot represents the situation at the point in time when it was produced.

YEAR	LDZ	Count of Unknown Projects	Sum Of AQ	Count of Supply Points
2016	EA	39	3,483,428	242
2016	EM	34	3,310,074	283
2016	LC	2	656,051	71
2016	NE	9	713,548	38
2016	NO	7	449,652	27
2016	NT	28	2,808,802	136
2016	NW	23	2,579,977	110
2016	SC	71	82,397,218	1,450
2016	SE	40	19,677,626	316
2016	SO	47	11,738,973	651
2016	SW	33	2,548,838	213
2016	WM	23	2,948,638	201
2016	WN	4	292,555	11
2016	WS	12	810,594	94
		372	134,415,974	3843

Table 7: Unknown Projects Snapshot

6. The total number and composition of Unknown Projects by LDZ is calculated by summing across all years. The total Unknown Projects supply point count and AQ for each LDZ is split by EUC using the percentages calculated from known CSEPs, described in Step 3 above.
7. In some cases, there may be additional Unknown Projects where the LDZ is unknown. These are assumed to have average composition by EUC (in terms of supply point count and AQ), with this composition again calculated from registered sites on known CSEPs.
8. The total AQ by EUC across all LDZs plus Unknown LDZ is calculated. These figures are split by Product Class using the standard rules and represent the best estimate of annual consumption in Unknown Projects at the time the snapshot was produced. As for the Shipperless and Unregistered categories, the *forecast year* population split by Product Class is also applied across the training period to allow consistent trends to be created for extrapolation to the forecast year.
9. The total IGT CSEPs Unidentified Gas is calculated for each LDZ as the sum of Unknown Projects Unidentified Gas for the LDZ (from Step 8 above) and the Unregistered Sites on Known CSEPs Unidentified Gas for the LDZ (from Step 4 above). Any Unknown Projects Unidentified Gas from Unknown LDZ is smeared across all LDZs.
10. The above process gives, for each point in time, an estimate of what annual Unidentified Gas from IGT CSEPs would be if conditions for the full year remained as they were in the snapshot. The estimates from successive snapshots show any trend that exists, which then requires extrapolation to the year for which Unidentified Gas is being forecast. An example of this trend across several snapshots is shown in Figure 22 below.

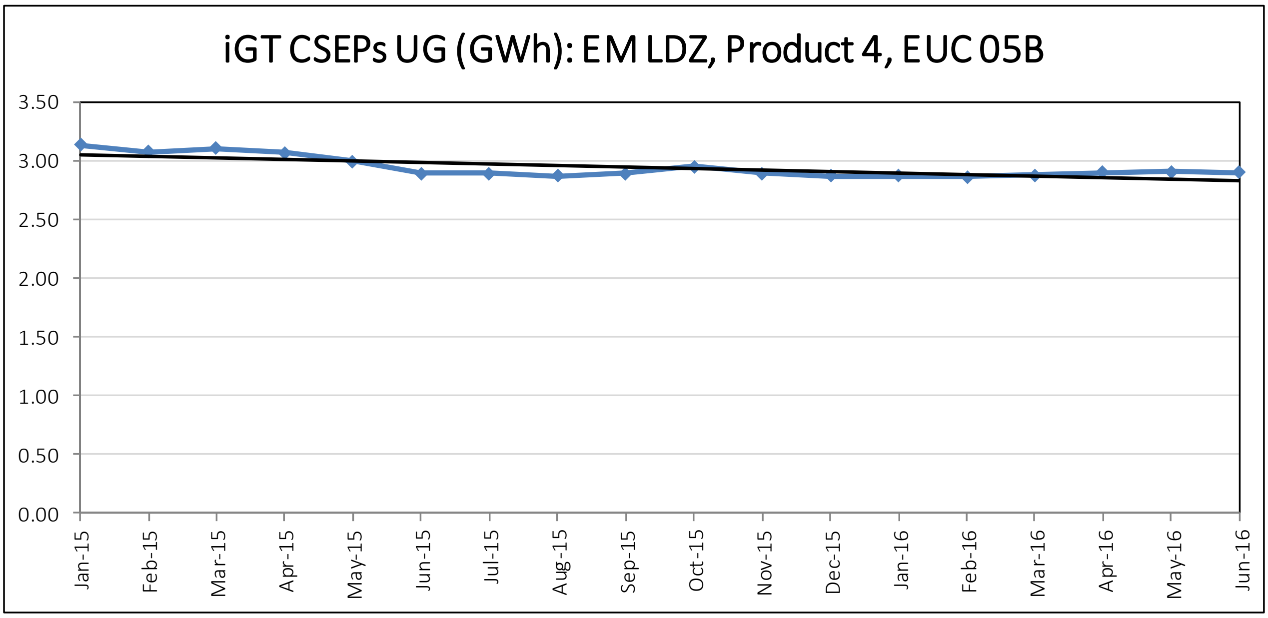


Figure 22: IGT CSEPs Unidentified Gas by Snapshot

11. The above process leads to the creation of a total of 468 individual trend lines for Unidentified Gas from IGT CSEPs. The identified snapshot-to-snapshot trend for each LDZ/Product Class/EUC combination is used to extrapolate either forwards or backwards to each time period of interest. The Unidentified Gas for each year used in the analysis is calculated using the fitted values for each snapshot date that falls within it. The time periods in question cover the Unidentified Gas forecast year and the historic Unidentified Gas training years, with values for each calculated using the fitted trend lines for each LDZ. The forecast year table is used directly in the final UIG factor calculations, whilst those for historic years are used in the calculation of total Unidentified Gas and the Balancing Factor, which are based on data from the training years (currently 2011/12 to 2016/17).

12. For the pre-Nexus period (i.e. the training period), the Unidentified Gas calculated in this way from LSP sites is temporary. The Unidentified Gas calculated in this way from SSP sites is permanent. Under Nexus, the CDSP have confirmed that all Unidentified Gas from this source will be reconcilable once the issue for the CSEP in question has been resolved. Therefore, for the forecast period all IGT CSEPs Unidentified Gas is temporary and does not form part of the permanent Unidentified Gas calculations.

7.5 Consumer Metering Errors

The effects of LDZ metering errors and known DM/Unique Site supply point errors are discussed in Section 7.1 above. In addition, undetected errors in all consumer supply point meters can cause gas to be burnt in an unrecorded or inaccurately recorded manner and hence have the potential to contribute to Unidentified Gas. An assessment of this area of metering error has therefore been carried out by the DNV GL Metering Team, and the conclusions drawn are presented here:

- Very little work has been done in the field of accurately assessing meter drift over time. Information is available about calibration curves taken at a particular point in time for certain meters, but there has never been any dedicated work looking at how these change over time. Therefore, conclusions drawn in this area are largely based on anecdotal evidence and/or extrapolation.

- Smaller sites (e.g. EUCs 01B and 02B) typically have diaphragm meters. The rubber diaphragm is known to warp over time, which causes drift in meter readings. Available evidence suggests that drift is equally likely to be up or down, which would result in a net bias of zero across each population. In the absence of any evidence to the contrary, this is therefore the assumption made throughout the Unidentified Gas calculations.
- For a more detailed analysis of such meter drift to be carried out, a large amount of data would have to be collected via a national meter survey. To carry out such a survey would be a significant undertaking as it would require a random sample of a sufficient size to cover many classes of meter (e.g. age of meter, type, model, level of consumption, capacity etc.), as well as co-operation of the customers and the physical testing of the meter itself with properly calibrated equipment. If such a survey was commissioned and carried out, the results could be used in future analyses of meter error. In the meantime, however, the evidence available leads to the assumption of a net zero drift over the population being used.
- Larger sites and offtakes generally have rotary/turbine meters that are constructed of metal and are unlikely to warp over time. These drift less than diaphragm meters, and again are equally likely to drift up or down, resulting in a net bias of zero across the population.
- Where large errors requiring an ad-hoc adjustment are found, these affect the Unidentified Gas calculations directly as described in Section 7.1. Data regarding such adjustments is supplied to the AUG Expert by the CDSP on a regular basis and is used to adjust the initial Unidentified Gas estimates.
- Calibration curves for both diaphragm and rotary/turbine meters follow a similar pattern. Such a curve for a NDM LSP Rotary Positive Displacement (RPD) meter is shown in Figure 23 below.

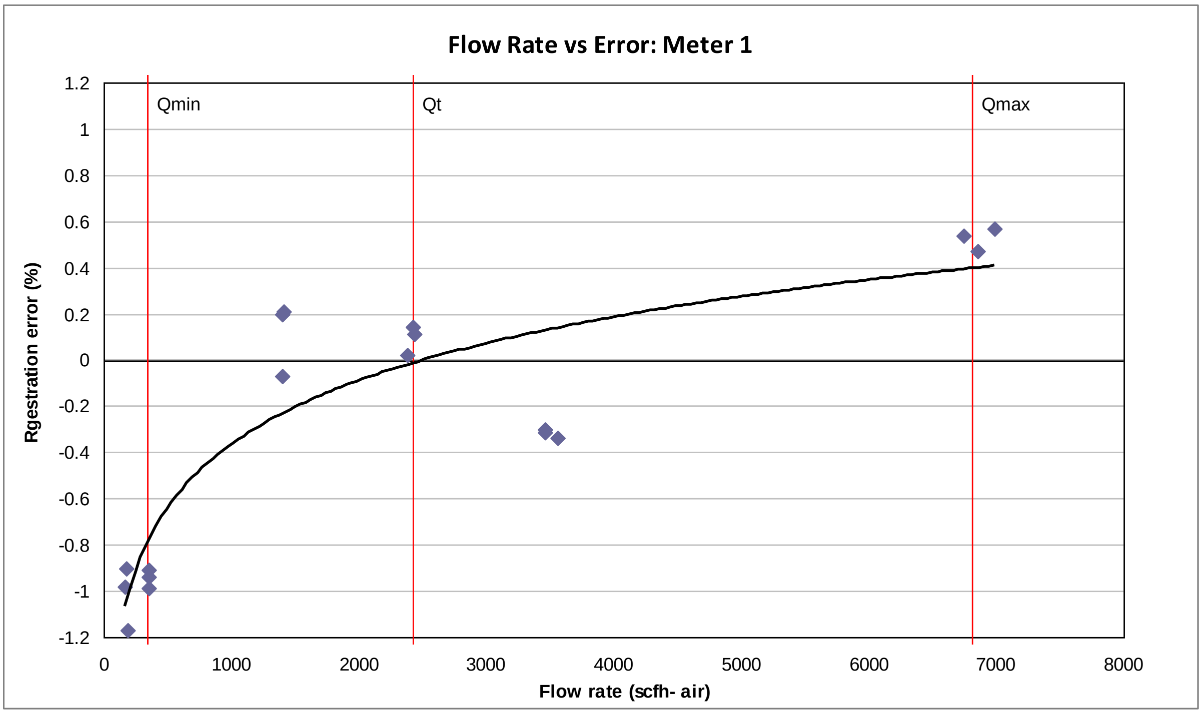


Figure 23: Typical Calibration Curve for an RPD Meter

Data for this graph was provided by the DNV GL metering team and comes from laboratory testing of a typical RPD meter. All identifying information has been removed for confidentiality purposes.

- The prominent features of this calibration curve are a consistent under-read of 1%-1.5% when operating at or below Q_{min}, unbiased readings around Q_t, and a consistent over-read at or close to Q_{max}.
- Meters are designed to operate at or around Q_t, ensuring that unbiased readings are obtained. This is not always the case, however, and circumstances may arise that cause some meters to operate close to Q_{min} or Q_{max}:
    - Loads at a particular site can drop over time, either due to changes in gas usage or because of economic conditions. This can lead meters to operate consistently close to Q_{min}.
    - Where businesses expand their operations without informing their gas supplier, the meter may no longer be appropriate for the load, causing it to run at or above Q_{max}.

Based on the above conclusions, an assessment of likely meter operating zones was carried out. Available data is limited to the meter capacity and AQ of each LSP site (EUCs 02B and above), and this requires the AQ to be used to estimate average hourly load, which can then be compared to meter capacity. This translation from annual load to hourly load necessarily introduces uncertainty into the analysis, but the comparison of average hourly load and meter capacity allows those meters that are likely to be operating at their extremes to be identified.

As stated in Section 4.3.3 above, for this Unidentified Gas category the latest available Product Class data can be used to directly look up the Product for each such site. The EUC of each site can be derived from its AQ and so the full 36-way split by EUC/Product Class can therefore be achieved.

- Sites with an average hourly flow of less than 1% of meter capacity are considered to be likely to be operating at or around Q_{min} when gas is flowing. These are assumed to be operating with an average under-read of 1.5%.
- Sites with an average hourly flow of more than 95% of meter capacity are considered to be likely to be operating at or around Q_{max} when gas is flowing. These are assumed to be operating with an average over-read of 0.5%.

The effects of under-reads and over-reads work in different directions, and the difference between them represents the net over- or under-read in the population.

- A net under-read results in Permanent Unidentified Gas equal to the value of the under-read.
- A net over-read results in the raw estimate of Unidentified Gas being over-stated, and it is therefore adjusted down by the value of the over-read.

The Meter Capacity data file is supplied annually. At present, four instances of this data are available, covering a period of four years, and this allows the trend across time for each Product Class/EUC category to be established. This set of trends is used to extrapolate the Unidentified Gas from this source for each individual category to the forecast year, and these are the figures used in the calculation of the factors.

7.6 Detected Theft

Detected theft that occurs within the back-billing window (3-4 years) is a temporary source of Unidentified Gas and it is therefore quantified and subtracted from the total Unidentified Gas for each part of the training period based on year of occurrence. For AUG years prior to 2017/18, the theft calculations were based entirely on CMS data provided by the CDSP. For more recent years, TRAS outcome file data going back to March 2016 is available, and so calculations are based on the combined set of theft records from both of these sources.

Whilst in theory the TRAS outcome files and CMS are meant to contain the same data, inconsistencies in how theft is recorded by Suppliers (TRAS) and Shippers (CMS) means that whilst there is overlap between the datasets, the two are different. Therefore, a superset is created from both sources, so that the full combined population of theft records is used. Theft values from these records are used directly in the Unidentified Gas calculation for the training period and are split into the relevant year of occurrence to ensure each theft applies to the correct gas year(s) – i.e. the year(s) in which it was active.

The detected theft analysis is carried out in Excel, where for each theft the proportion of it that falls into each training year it is estimated to have been active is first calculated. These figures are used to calculate total detected theft by gas year for the training period as follows:

1. For each theft record and for each training year:
    a. Calculate the number of days in each year that the theft was active. There are four scenarios:
        i. The theft starts and ends within the gas year: the full duration of the theft occurs in a single year and values for all other years are zero for this theft.
        ii. The period of theft spans the gas year in question: duration of the theft in this year is all days in the year.
        iii. The theft is estimated to start in the gas year in question and end in a subsequent gas year: in this case the difference between the end of the gas year and theft start date is used.
        iv. The theft is estimated to end in the gas year in question and had started in a previous gas year: in this case the difference between the start of the gas year and theft end date is used.

    For all years for which the theft is not active, the value is zero.

    b. Calculate the amount of theft estimated to have occurred in the year in question by splitting the total estimated energy value of the theft according to the number of days it was active in each year.
2. Aggregate all thefts by gas year for the training period. Dummy MPRNs that are consistent with all of the other datasets are supplied as part of both the TRAS and CMS data and so the Product Class for the site associated with each theft can be directly queried. The ultimate use of this data is as an aggregate adjustment to the output of the Consumption Method, however, and so neither a split by EUC nor Product Class is used in subsequent calculations.
3. The annual totals calculated in this way are then amended to account for the number of thefts for each training year that have not yet been detected but will be detected in the future up to the point where they can no longer be back-billed (i.e. those that are detected within the settlement period and are therefore temporary). This is up to year 5 prior to 2014 and year 4 afterwards following implementation of Mod 0398 [14] in April 2014. Thefts can often run for years before they are

detected, and so in many cases it can be several years before a theft active in Year Y is detected. Analysis of thefts and their time to detection has been carried out using the data supplied by the CDSP. Figure 24 below shows the percentage of thefts detected (i.e. as a percentage of the total thefts that will eventually be detected up to year 8) by year based on all years in the CDSP data.

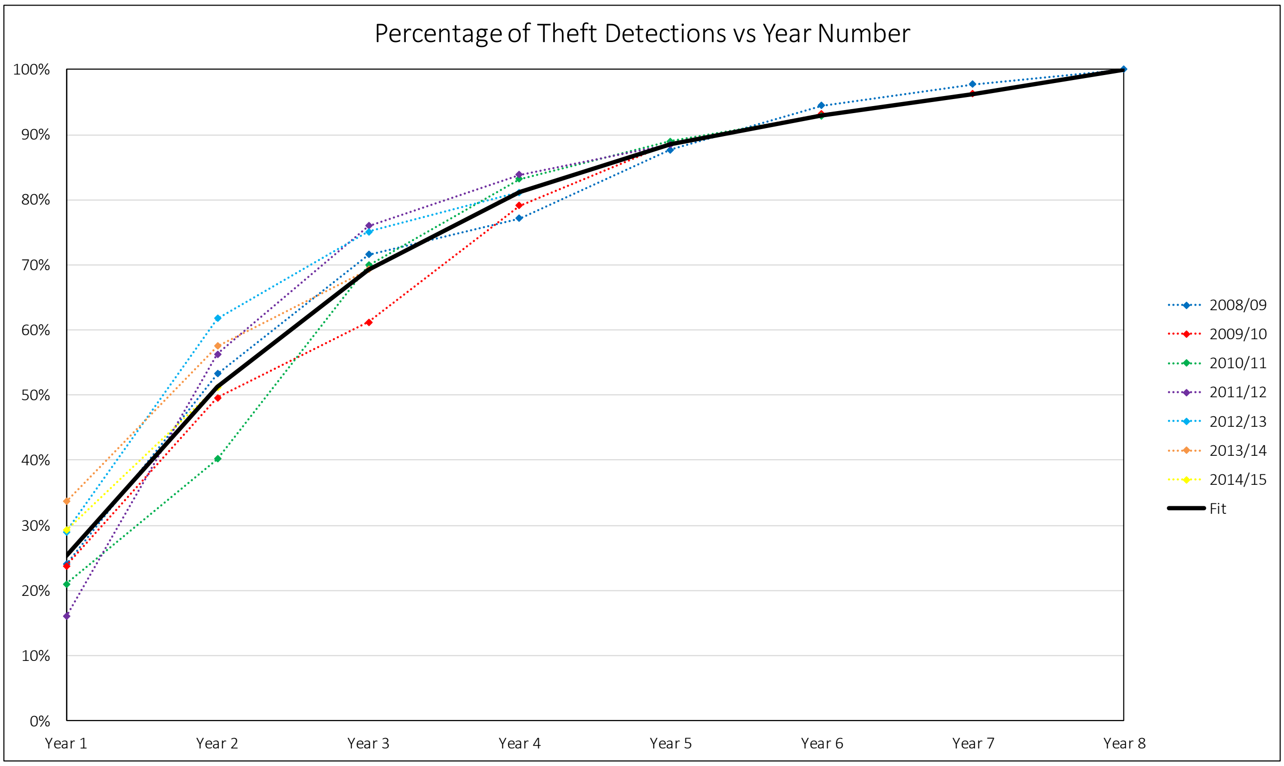


Figure 24: Percentage of Theft Detections by Year Number

The dotted lines show figures for thefts active in individual years, whilst the black line shows the overall fit. This shows that only around 25% of thefts that are going to be detected are detected in the year they become active, and that it takes six years before a detection rate of 90%+ (of the thefts that will eventually be detected) is reached. At Year 8 it can be regarded that the number of ongoing thefts that will be detected but have not been yet is non-zero but negligible.

Given that this full 8 year cycle has not completed for the majority of training years, the detected theft figures (which can now be regarded as representing thefts that have been detected so far) must be amended to incorporate an additional estimate of further thefts that will be detected. This is done using the values from the above graph, where detected thefts for active years less than 8 years ago are scaled up based on the estimated percentage to the level of theft which would be expected in total. Thefts detected up to and including year 4 contribute temporary Unidentified Gas, whilst the remainder are permanent.

4. The estimated annual temporary theft totals for the training period are then applied as corrections to the raw Unidentified Gas estimates from the Consumption Method. The permanent element becomes part of the Balancing Factor.

7.7 CSEP Shrinkage

The AUGE Framework has recently been updated to reflect the fact that Shrinkage Error (i.e. any bias in the Shrinkage and Leakage Model) is considered to be out of scope for the Unidentified Gas analysis. The facility to include an estimate of Shrinkage Error has therefore been removed from the calculation process.

CSEP Shrinkage must still be included in the Unidentified Gas calculations, however. It is acknowledged by the industry that CSEP Shrinkage is not accounted for anywhere in the settlement process and hence ends up in UIG. It is therefore taken account of in the Unidentified Gas calculations and hence the factors that are applied to the daily UIG figure. A method has been developed for estimating CSEP Shrinkage as accurately as possible, and this is documented below. If, in the future, CSEP shrinkage is estimated by the GTs (but still not accounted for elsewhere in settlement) then their figure will be used for Unidentified Gas and will replace the AUGE's own CSEP Shrinkage estimate.

The CSEP Shrinkage estimation method is based on the following principles:

- CSEP Shrinkage arises entirely from leakage.
- Leakage rates from the National Leakage Tests (as used in the Shrinkage and Leakage Model) are available and in the public domain.
- It can be assumed that all mains within CSEPs are PE.
- The calculation of leakage for CSEPs using these leakage rates requires data on the aggregate length of mains in CSEPs. Diameter is not important due to the assumption of all PE, because the leakage rates for all diameters of PE are the same. This information is not available, but it can be estimated as follows:
    - The network models used for validation and planning using the Graphical Based Network Analysis (GBNA) software contain all the information required to calculate both the number of demands and the total mains length in any given network or area of a network as defined by the user. Cadent have given the AUGE permission to use data from these models as part of the Unidentified Gas analysis.
    - GBNA allows the selection of given areas of any network by polygon, with results returned for that polygon only. Therefore, a sample of network areas with a composition similar to that of a typical CSEP can be identified.
    - For each such polygon, the total mains length and the number of customers can be calculated from the network model. This allows an average "mains length per customer" figure to be calculated across the whole sample.
    - The total number of sites in CSEPs is known from data provided to the AUGE by the CDSP. This can therefore be used to calculate an estimate of total CSEP mains length by applying the number of sites to the factor described above.
    - It is assumed that this mains length consists entirely of PE and hence the PE leakage rate from the National Leakage Tests is applied to the estimated CSEP mains length. This is the best estimate of CSEP Shrinkage.
    - The energy value of CSEP Shrinkage is converted to a percentage by dividing it by the total LDZ Shrinkage estimate from the Gas Transporters [19] for the equivalent period. The percentage of the Shrinkage total that arises from CSEPs is assumed to remain constant across the training and forecast period.

This approach involves the application of an LP leakage rate to all CSEP mains. Whilst it is known that not all CSEPs are low pressure, LP leakage rates are the only ones that currently exist, and these are used for both the LP and MP/IP pressure tiers in the Shrinkage and Leakage Model. The approach to CSEP Shrinkage is therefore consistent with the accepted approach to the calculation of overall LDZ Shrinkage.

7.8 Volume Conversion Error

For each year of the training period, the energy affected by the volume conversion issues (EAV) depends on the prevalence of volume conversion devices and is calculated as

$$EAV = Cons_{SN} * (1 - x)$$

Where:

EAV = Energy Affected by Volume conversion issues

$Cons_{SN}$ = Seasonal Normal Consumption calculated by the AUG Expert

x = Proportion of NDM AQ from meters with volume conversion devices

The total Unidentified Gas from assuming that average atmospheric pressure = 1013.25 mbar (UIG_p) is then calculated as described in Section 7.8.1 below and the total Unidentified Gas from assuming that average temperature = 12.2°C (UIG_t) is calculated as described in Section 7.8.2 below.

The total Unidentified Gas from volume conversion error due to pressure (UIG_p+ UIG_t) is then split by EUC and Product Class based on the total prevailing AQ excluding those meters with volume conversion devices. This gives the split in Table 8.

	PC1	PC2	PC3	PC4
01B	0.00%	0.00%	6.92%	71.07%
02B	0.00%	0.00%	1.62%	5.18%
03B	0.00%	0.00%	1.51%	3.36%
04B	0.00%	0.01%	1.18%	3.93%
05B	0.00%	0.01%	0.54%	2.09%
06B	0.00%	0.03%	0.20%	1.26%
07B	0.00%	0.02%	0.13%	0.51%
08B	0.02%	0.06%	0.01%	0.27%
09B	0.06%	0.00%	0.00%	0.03%

Table 8: Volume Conversion (Pressure & Temperature) Error Split

7.8.1 Average Atmospheric Pressure Assumption

The error in the standard correction factor due to the average pressure assumption is estimated by LDZ as

$$CF_{err}^{std} = \left[\frac{(P_{av} - 1013.25)}{1013.25}\right] * 1.0098/1.02264$$

Where:

CF^{std}_{err} = Error in the standard Correction Factor expressed as a ratio i.e. $(CF-CF_{std})/CF_{std}$

P_{av} = Average Flow Weighted actual pressure at MSL for the relevant gas year and LDZ. As pressure data is only available from 2012 onwards, the average value across all gas years (2012-2017) is used for gas years before 2012. Flow weighting uses total allocations i.e. NDM

The Unidentified Gas as a result of this pressure correction is then given by

$$UIG_{std}^{P} = EAV * CF_{err}^{std}$$

Where:

UIG^{p}_{std} = the pressure related Unidentified Gas assuming all meters use the standard Correction Factor

EAV = Energy Affected by Volume conversion issues

CF^{std}_{err} = Error in the standard Correction Factor expressed as a ratio

The estimate of UIG^{p}_{std} is a first order approximation assuming all meters without volume conversion operate at or near 21 mbar pressure. In reality, there are meters which operate at higher pressures. These meters are less sensitive to the changes in atmospheric pressure so UIG^{p}_{std} will be an over-estimate.

A further adjustment is therefore calculated for all meters with CF>1.03193 (roughly equates to meter pressures above 30 mbar). This affects only 1,470 meters. For each of these meters individually, the error in the CF is calculated as follows

$$CF_{err}^{act} = \left[\frac{(P_{av} - 1013.25)}{1013.25}\right] * 1.0098/CF$$

Where:

CF^{act}_{err} = Error in the actual Correction Factor expressed as a ratio

P_{av} = Average actual pressure at MSL over all LDZs for the relevant gas year. As pressure data is only available from 2012 onwards, the average value across all gas years (2012-2017) is used for gas years before 2012

CF is the actual CF for that meter.

The correction for each meter is then calculated as

$$UIG_{hp} = AQ * \left[1 - \frac{(1 + CF_{err}^{act})}{(1 + CF_{err}^{std})}\right]$$

Where:

UIG_{hp} = a correction to allow for meters operating at higher set pressures

AQ = Prevailing AQ of the meter

CF^{std}_{err} = Error in the standard Correction Factor expressed as a ratio

CF^{act}_{err} = Error in the actual Correction Factor expressed as a ratio

The total Unidentified Gas as a result of this pressure correction is then given by

$$UIG_p = UIG_{std}^{P} - UIG_{hp}$$

Where:

UIG_p = the total Unidentified Gas as a result of assuming average pressure is 1013.25 mbar

UIG^p_{std} = the pressure related Unidentified Gas as a result of assuming average pressure is 1013.25 mbar calculated based on the error in standard CF

UIG_{hp} = a correction to allow for meters operating at higher set pressures. This value reduces

the Unidentified Gas total as meters operating at higher set pressures are less sensitive to changes in atmospheric pressure.

Previously, the AUGE has estimated this over-estimate of pressure related UIG to be negligible (~0.2 GWh on average between 2011 and 2015)[58]. The latest estimate based on the 2011 to 2016 gas year average is ~0.033GWh.

The actual pressure data used in the estimation of UIG^p_{std} was provided by the CDSP. It consisted of hourly pressure data for eight weather stations covering a period of six gas years (2012 to 2017). The methodology has been updated this year to calculate UIG^p_{std} on an LDZ by LDZ basis using the actual local atmospheric pressure data. The LDZ specific UIG^p_{std} values were then summed to obtain a national figure (previously a simple average of the eight weather stations was used as an estimate of GB pressure and calculations carried out at national level). However, as data was only available for eight weather stations it was necessary to use a mapping to LDZ. The mapping used is shown in Table 9. The choice of weather stations was based roughly on the latitude of the LDZ given the atmospheric pressure variations shown in Section 5.4. The correction for meters operating at higher pressures (UIG_{hp}) was calculated at National level only and applied as an adjustment to the summed LDZ level UIG.

It should be noted that the average atmospheric pressure for gas year 2016/17 is higher than previous years and the 1013.25mbar value assumed by the standard CF. Including this additional year has therefore increased the UIG due to atmospheric pressure.

LDZ	Weather Station
EA	Heathrow
EM	Nottingham
NE	Average of Nottingham & Albemarle
NO	Albemarle
NT	Heathrow
NW	Nottingham
SC	Glasgow
SE	Heathrow
SO	Filton
SW	Yeovilton
WM	Coleshill
WN	Nottingham
WS	St Athan

Table 9: LDZ-Weather Station Mappings for Atmospheric Pressure Related UIG

7.8.2 Average Temperature Assumption

Four different sets of temperatures are used based on EUC category (groups of EUC bands) – 01B, Small I&C (EUC bands 02 and 03), Large I&C (EUC bands 04 to 08) and DM (EUC 09B). For each category, an average gas temperature is calculated as follows

- 01B – a weighted average of the internal and external temperatures from the DMTS based on the meter location split of 01B. The meter location split is based on the split of known meter locations only and sheltered meters are included as internal meters
- Small I&C – a weighted average of the internal and external temperatures from the DMTS based on the meter location split of all meters in EUC bands 02 and 03. The meter location split is based on the split of known meter locations only and sheltered meters are included as internal meters
- Large I&C – the Large I&C temperature from the ICTS is used
- DM - the DM temperature from the ICTS is used

This results in a set of assumed gas temperatures by LDZ and EUC category as shown in Table 10. These temperatures are taken as good estimates under SN conditions. Temperatures below the assumed 12.2°C are shown in blue, those above in orange.

LDZ	01B	02-03	04-08	09B
EA	11.70	12.29	10.10	11.10
EM	11.30	11.64	10.90	12.10
NE	11.03	11.71	9.90	11.20
NO	11.05	11.26	9.40	10.50
NT	13.96	14.28	13.40	14.80
NW	11.54	11.43	10.40	11.40
SC	10.79	11.11	8.80	9.90
SE	13.52	13.16	11.50	13.00
SO	11.97	12.18	10.60	11.80
SW	10.94	11.43	11.00	12.10
WM	11.29	11.23	10.00	10.70
WN	10.49	10.58	9.90	10.70
WS	12.41	12.45	11.30	12.60
Average	11.69	11.90	10.55	11.68

Table 10: Assumed FWAGTs to Calculate Temperature Related UIG

The error in the standard correction factor due to the average temperature assumption is estimated for each LDZ and EUC category as

$$CF_{err}^{std} = 1 - \left[\frac{288.15}{(273.15 + T_{av}) * 1.0098}\right]$$

Where:

CF^{std}_{err} = Error in the standard Correction Factor expressed as a ratio i.e. $(CF\text{-}CF_{std})/CF_{std}$

T_{av} = Average gas temperature in C for the relevant EUC category and LDZ

The Unidentified Gas for each LDZ and EUC category as a result of this temperature correction is then given by

$$UIG_{std}^{T} = EAV * CF_{err}^{std}$$

Where:

UIG^{T}_{std} = the temperature related Unidentified Gas assuming all meters use the standard Correction Factor

EAV = Energy Affected by Volume conversion issues for the relevant EUC category and LDZ

CF^{std}_{err} = Error in the standard Correction Factor expressed as a ratio for the relevant EUC Category and LDZ

The total UIG due to temperature conversion error is then calculated by summing across all LDZs and EUC categories.

7.8.3 Use of Standard Correction Factor

The GTER[26] mandates that all meters reasonably expected to consume in excess of 732,000 kWh (EUC 04 or higher) have a site-specific CF. Analysis carried out estimates a potential impact of using standard CFs rather than site specific CFs of 31.72GWh. This is based on the difference between the standard correction factor and the average site specific correction factor for meters in EUC 04 and above without volume converters.

The AUG Expert has assumed that this figure is unlikely to change significantly as most updates of correction factors from the standard value to a site specific value will already have been made as part of recent data cleaning initiatives.

The 31.72GWh of permanent UIG needs to be split by EUC and product class. This split is calculated based on relative throughput (AQ) for affected meters. Affected meters are those in EUC 04 and above with standard CF and without volume converters fitted. The resulting split of UIG by EUC and Product Class is shown in **Table 11**.

UIG Std CF (GWh)	PC1	PC2	PC3	PC4
01B	0.00	0.00	0.00	0.00
02B	0.00	0.00	0.00	0.00
03B	0.00	0.00	0.00	0.00
04B	0.00	0.00	1.80	16.80
05B	0.00	0.00	0.91	5.44
06B	0.00	0.00	0.40	3.45
07B	0.00	0.00	0.51	1.64
08B	0.00	0.00	0.00	0.77
09B	0.00	0.00	0.00	0.00

Table 11: UIG from Standard CF by EUC and Product Class

7.9 Balancing Factor

All of the analysis described above allows the total permanent Unidentified Gas to be calculated for each year of the training period. The permanent elements of the directly calculated Unidentified Gas categories (IGT CSEPs, Shipperless/Unregistered, consumer meter errors, CSEP Shrinkage, detected theft beyond line in the sand, and volume conversion error) can be subtracted from this to give the Balancing Factor for each year.

At present there are six years in the training period, which allows the pattern in the Balancing Factor to be assessed and the appropriate method selected to extrapolate it to the forecast year. This is discussed further in Section 7.11 below. In addition to extrapolating the Balancing Factor as accurately as possible, it must also be split by EUC and Product Class like all categories of Unidentified Gas. In order to do this, rules must be applied.

It is assumed that the Balancing Factor is composed mainly of undetected theft, and this provides the basis for how it is split. The following sections describe the logic and calculation steps for this process.

7.9.1 Background

The purpose of this analysis is to use detected theft records to create a set of factors that can be used to split the Balancing Factor between the 36 combinations of EUC/Product Class. This approach is based on the assumption that is made throughout the Unidentified Gas analysis: that the Balancing Factor is composed mostly of undetected theft and should therefore be split based on our best estimate of the relative incidence of such theft.

The issue with using detected theft records to split undetected theft is that detected theft patterns are not necessarily consistent with wider (undetected) theft. Theft will only be detected where it is looked for, and so detected theft rates are heavily influenced by the investigation and detection activity that each Supplier chooses to carry out. There is a lot of detected theft information available and it is a potentially valuable resource for deriving information about undetected theft, but the challenge of this process is to remove the bias caused by the targeting of the Suppliers' the investigation and detection activities to produce unbiased factors that will be indicative of overall theft.

The process of detecting theft has three stages, as follows:

- Lead
  Suspicious meter read pattern, observation of physical signs of tampering at the meter, or tip-off. Meter reads can be identified as suspect via the Supplier's own analysis or TRAS outlier analysis. Notification of physical evidence of theft can comes from the Meter Asset Manager (MAM), Meter Reading Agent (MRA) or a Gas Transporter (GT).
- Investigation
  The Supplier decides which leads, if any, to investigate further. Each Supplier uses their own criteria and investigation rates vary widely – from investigating nothing to investigating everything. On average, around 35% of leads are investigated (2016 SPAA Theft of Gas report).
- Detection
  The proportion of investigations that lead to a detected theft again varies between Suppliers, from 0% to 40% (2016 SPAA Theft of Gas report). On average, approximately 20% of investigations result in a detection of theft.

7.9.2 Principles

In order to produce objective factors that can be applied to undetected theft (and hence the Balancing Factor), the bias needs to be removed from two of the above stages:

1. The bias caused by what detection activity (or lack of it) the Supplier chooses to carry out.
2. The bias caused by the Suppliers' selection of which leads to investigate.

Given the data currently available, the relationship between investigation and detection (as recorded in the raw data) can be assumed to be constant and representative for any given EUC/Product Class category, although Investigation → Detection rates will vary from category to category. This assumption of single Investigation → Detection rates for each category may not be strictly true, because if the Suppliers' selection criteria are effective there is likely to be a "diminishing returns" effect – i.e. the highest ranked (and therefore first) investigations are more likely to relate to a confirmed theft, whilst the more investigations are carried out the lower the quality of the lead being followed and the lower the probability of it relating to an actual theft. The data currently available does not allow this effect to the quantified, however, and very detailed data from Suppliers covering the ranking of leads and the sequence of investigations would be required in order to do this.

Therefore, the Investigation → Detection relationship for each EUC/Product Class category is assumed to be constant. This assumption will remain valid whatever "diminishing returns" effect really does exist as long as Supplier detection behaviour remains consistent (i.e. they maintain a similar strategy over time).

As with all elements of the Unidentified Gas analysis, the undetected theft factors must be split by EUC and Product Class as described above, and in this case by Meter Type (traditional/Smart/AMR) as well. The need to split by Meter Type only creates three more categories because the remaining 33 all have mandatory requirements for the meter: Smart Meter or AMR for anything except Product Class 4, and AMR for EUC 04B and above. This leaves just Product Class 4 EUCs 01B-03B that can take either a Smart or traditional meter.

The general principle is therefore to base our method on existing leads and investigations (split by EUC/Product Class/Meter Type), but to adjust the numbers of these to what they would have been if the investigations had been carried out equitably based only on:

1. Population
2. Propensity to suspicious activity

Once these figures have been derived, category-by-category Investigation → Detection rates can be used to convert them to unbiased detection numbers. The relative size of these adjusted detection figures provides the basis for undetected theft factors which can then be used to split the Balancing Factor.

7.9.3 Data

The data specified below represents a subset of the contents of the TRAS Outcome files that are provided by Suppliers. This data is supplied to the AUGE via the CDSP, and the most recent file contains theft records up to the end of September 2019.

The removal of bias from the "leads" data and the production of category-by-category Investigation → Detection rates both require record by record theft data from the TRAS Outcome files. This must cover **all** leads (not just those that were investigated or led to a detection) for the full analysis described here to be carried out. This includes both ETTOS leads that were not investigated and TRAS Qualified Outliers that were not investigated.

The minimum information each record must contain is as follows. Items in italics are provided by the CDSP based on the data supplied by SPAA.

- (Dummy) MPRN
  Real MPRN to be supplied to the CDSP, who will convert to dummy MPRNs consistent with other datasets.
- Meter Serial Number
  Provided to CDSP only, not to the AUGE.
- *EUC*
  To be provided by the CDSP, queried using MPRN.

- *Product Class*
  To be provided by the CDSP, queried using MPRN.
- *Meter Type (traditional/Smart/AMR)*
  To be provided by the CDSP. This will be queried from the asset data using MPRN, and also calculated using the Meter Serial Number based rule set provided by the AUGE. If either source returns AMR or Smart, this is the assigned value.
- *Meter installation date*
  To be provided by the CDSP, queried using MPRN.
- Source of lead (MAM, MRA, GT, TRAS, own analysis, tip-off)
  From TRAS Outcome files.
- Lead investigated? (Yes/No)
  From TRAS Outcome files.
- Theft detected? (Yes/No)
  From TRAS Outcome files.
- Assessed Loss
  From TRAS Outcome files.
- Tamper Code
  From TRAS Outcome files.

Data is therefore initially supplied to the CDSP, and then from the CDSP to the AUGE once it has been anonymised and the additional fields added.

7.9.4 Calculation

The first step of the analysis is to produce a set of unbiased leads, split by the 39 population categories (i.e. the original 36-way EUC/Product Class split plus 3 for meter type). Leads from the following sources can be considered free from Supplier bias:

- TRAS
- MRA
- Tip-off
- Field Agent
- Police

Leads from all of these sources either come from the whole population where any given theft is equally likely to be flagged (MRA, tip-off, Field Agent and Police) or are the result of dedicated analysis that is applied to the whole population without Supplier specific targeting (TRAS). Leads from other areas (own analysis, MAM, GT) *may* have inherent targeting that could skew the number of leads coming from these sources and create bias across the population categories. These are therefore discounted in the "unbiased leads" calculation.

It is vital that the leads considered in this analysis (and the subsequent investigations and detections) consist only of those associated with thefts that contribute to UIG, i.e. fiscal theft needs to be removed. The removal of leads associated with fiscal theft requires a process of estimation, for two reasons:

1. The lack of a dedicated Fiscal Theft code in TRAS means that the number of fiscal thefts must be estimated using other information.
2. Only a confirmed theft can only be classified as fiscal. Therefore, the number of unbiased leads associated with suspected fiscal theft must be estimated from the proportion of confirmed fiscal thefts.

The method for estimating the number of fiscal thefts and removing them from the dataset is described in Section 7.9.5 below.

The unbiased leads derived in this way (as a subset of the overall leads) and split by EUC/Product Class/Meter Type therefore reflects a combination of both the population of each category and the propensity to suspicious activity. At this point the option exists to scale the "unbiased leads" total to the overall leads total, which would result in estimates for each population category of what the number of leads would have been without any targeting. The final output from the overall detected theft analysis is a set of factors that are used to split the Balancing Factor, which therefore operate on the basis of their relative level rather than their absolute level. As such, scaling the leads in this way does not result in any tangible difference in the output, but it nevertheless ensures that the leads total remains the same - this may aid industry parties in understanding the process. This is therefore strictly speaking an optional step, but one which is applied for this reason.

These "unbiased leads" figures must now be converted first into "unbiased investigations" and from there into "unbiased detections".

Whilst the method for the Investigation → Detection step has already been defined in Section 7.9.2 above, and appropriate rates for each population category can be calculated from the record-by-record theft data, the category-by-category Lead → Investigation step has not yet been defined.

Whilst we now have unbiased figures for leads, as described above, we cannot use category-by-category Lead → Investigation rates calculated from the raw data because the decision to investigate certain leads but not others still lies with the Supplier and hence may still contain an element of targeting. If such Lead → Investigation rates were calculated from the raw data (and aggregated across Suppliers), any targeting effect would manifest itself as a deviation from uniform values: if there was no targeting, all the rates would be the same.

At this point an assumption needs to be made that the quality of leads (which can be regarded as the likelihood of any given lead meriting further investigation) does not vary between population categories: so, for example, leads from Population Category A will have a similar probability of meriting further investigation as leads from Population Category B. This is a reasonable assumption because whilst the data granularity (i.e. meter read frequency) will vary between population categories, this will affect only the speed with which a lead can be identified rather than the quality of the lead itself.

Therefore, based on this assumption, any deviation from uniformity in the Lead → Investigation rates across population categories reflects different Supplier behaviour in following up these leads. If, for example, a large domestic Supplier rarely acts on any leads but a large Supplier for small commercials follows up almost all leads, this will knock on into differences between the Lead → Investigation rates for EUCs 01B-03B (all relevant Product Classes).

With no differences in Supplier behaviour there would be a constant Lead → Investigation rate across all population categories, and therefore this needs to be the basis for the step from unbiased leads to unbiased investigations in the theft analysis. A single rate calculated as the ratio of all leads to all investigations (both aggregated across all categories and all Suppliers) should therefore be used as a constant value to convert unbiased leads to unbiased investigations.

As described in Section 7.9.2 above, category-by-category Investigation → Detection rates derived directly from the raw data can now be applied to the unbiased investigations figures. These rates can be calculated from the raw data without the need for further manipulation because it is the identification of leads and the decision to investigate that are affected by the different theft regimes of different Suppliers – once an investigation is under way, its likelihood of resulting in a detection of theft is unaffected by the decision process that led to the investigation.

The result of this stage of the calculation is a set of estimated unbiased theft detections for the time period covered by the raw theft dataset, split by the 39 population categories (36 EUC/Product Class plus 3 for meter type). In the final steps of the process these must first be converted to kWh of theft rather than the number of thefts, and then projected forward to the forecast year. At this point, the dual figures for Product Class 4 EUCs 01B-03B are combined into single figures for each, and these final figures are converted to factors. This final output is used to split the Balancing Factor. This process is carried out as follows:

1. For each EUC/Product Class/Meter Type category, calculate the average kWh stolen per theft. These figures will reflect not only the higher quantities of gas consumed by larger sites, but also any effects caused by meter read frequency and data granularity affecting theft duration.
2. Multiply the number of thefts by the average kWh to give the unbiased total stolen energy from detected thefts.
3. Calculate the change in population for each population category from the theft dataset year to the forecast year, as a percentage Pn% for each category.
4. Scale each unbiased stolen energy figure by each Pn – this is the best estimate of unbiased total stolen energy from detected thefts for the forecast year.
5. Add the individual component unbiased stolen energy figures for PC4 01B, PC4 02B and PC4 03B to give single estimates for each of these EUC/Product Class categories.
6. Convert these raw figures to proportions for each EUC/Product Class category.
7. Apply these proportions to split the Balancing Factor estimate for the forecast year.

7.9.5 Fiscal Theft Adjustment

In order to restrict the unbiased leads, unbiased investigations and unbiased detections data to only cover thefts that contribute to UIG, fiscal theft must be removed at the "unbiased leads" stage. This requires a process of estimation due to the fact that only confirmed thefts have a tamper code associated with them, and so there is no record of leads that are associated with suspected theft of a particular type. In addition, the fiscal theft tamper code has not yet been implemented in TRAS, and therefore even the confirmed fiscal thefts must be estimated using other information.

The fiscal theft adjustment is carried out according to the following procedure:

1. The total number of confirmed thefts from ETMs is calculated.
2. It is assumed that all valve tampers on ETMs are fiscal thefts. The proportion of all ETM confirmed thefts that are assumed to be fiscal thefts (according to this definition) is calculated.

3. Unbiased leads, calculated as described in Section 7.9.4 above, are split into those associated with credit meters and those associated with ETMs.
4. The proportion of fiscal thefts calculated in Step 2 is applied to the population of unbiased leads associated with ETMs calculated in Step 3.
5. These records are removed from the unbiased leads figures and play no further part in calculations. The calculation process goes on to the "unbiased investigations" and "unbiased detections" phases based on the adjusted unbiased leads with (estimated) fiscal theft removed.

7.9.6 Smart Meter Theft Adjustment

The above sections describe the proposed theft calculation method in full. In addition to this, the following extension to the methodology is proposed, for implementation when sufficient data exists to implement it.

The Smart Meter population is young, and existing theft work shows that there is an approximate lead time of 8 years until all thefts that are going to be detected have been detected. This timescale may be reduced for Smart Meters due to the more detailed information that comes from them, but this is yet to be proven.

This phenomenon will not affect the Investigation → Detection rates for Smart Meter population categories, but it will affect the number of leads, i.e. where "young" thefts have not yet produced enough suspicious meter readings to generate a lead, and so they will not yet be investigated and detected. Therefore, an adjustment for this effect will be possible when there is enough data to support it, which would use the meter installation date record for each detected theft in the training data year with the following logic applied:

- For Smart Meters up to 1 year old, only P_1% that will generate a lead have yet done so.
- For Smart Meters 1-2 years old, P_2% that will generate a lead have done so.
- ⋮
- For Smart Meters 7-8 years old, P_8% that will generate a lead have done so.
- For Smart Meters over 8 years old, all that will generate a lead will have done so.

where $P_8 > P_7 > \ldots > P_1$

The installation date field in the detected theft data allows the total number of Smart Meter leads from the raw data (for any given population category) to be further stratified by meter age. The factors P_1, P_2, …, P_n as defined above will then be applied to these stratified figures to scale the number of (raw, untargeted) leads to what they would have been if the population was mature, i.e. an estimate of the actual number of Smart Meter sites with suspicious meter reads rather than just those it has been possible to identify at this early stage. Raw Smart Meter leads from all sources will first be scaled in this manner to give a set of revised targeted leads before the processes detailed above are applied to remove bias and output the sets of unbiased leads, investigations and detections.

This calculation will only be possible when sufficient data exists to support it and allow theft detection rate curves (similar to those used in the existing Detected Theft calculation for the whole population) to be generated specifically for Smart Meters. In the data currently supplied, there are 523 confirmed thefts from Smart Meters out of a total of 12,644 confirmed thefts. This is not currently sufficient to implement the Smart Meter adjustment, but this will be assessed again during factor calculation in the future.

If the reasonable assumption is made that propensity to steal is steady over time, the Smart Meter population is the only one that requires this pre-adjustment. The AMR and traditional meter populations are mature and so in these cases the issue will not occur.

7.9.7 Process Summary

Figure 25 below shows a simplified graphical representation of the theft analysis process steps. The start point for the process is the dataset containing all leads, with the final output being the split of the Balancing Factor in line with proportions of unbiased theft (from each EUC/Product Class category) for the forecast year. The additional potential steps of the Smart Meter population adjustment, which will be carried out when sufficient data is available, are shown in grey.

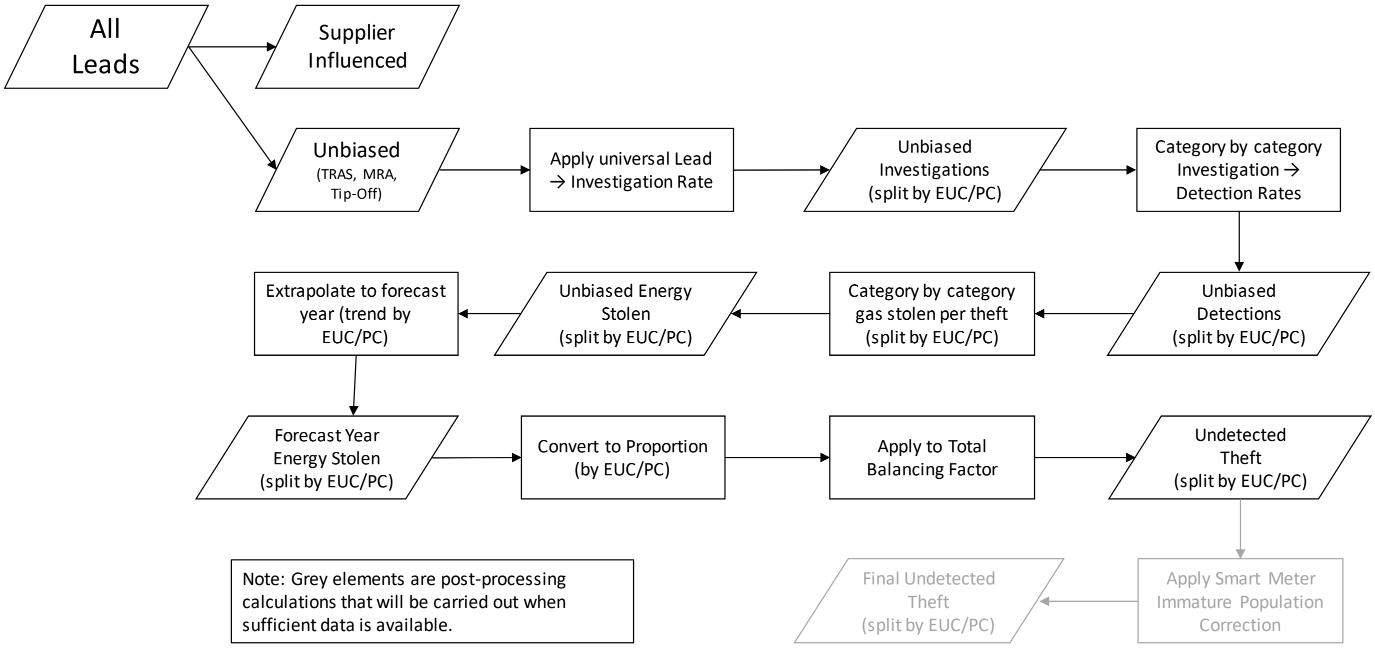


Figure 25: Theft Analysis Process Steps

7.9.8 Data Limitations

The theft analysis described above requires detailed line-by-line data, not only for confirmed thefts but also for all investigations carried out (regardless of whether a theft was detected), and all leads (including ETTOS leads and TRAS qualified outliers) regardless of whether they were investigated. The additional line-by-line information on leads and investigations allows the effect of Supplier theft targeting to be estimated and removed.

The release of the full data, including all ETTOS leads and TRAS qualified outliers that were not investigated, has to be authorised by the Theft Issues Group (TIG). This data has not yet been released due to TIG's concerns about the completeness of the data. At a meeting held on 5/12/2019, TIG put forward their position that it is currently the responsibility of the Supplier to enter data for ETTOS leads and TRAS outliers, and just as with confirmed thefts in CMS, only a subset of Suppliers comply with this. This results in the potential for skewing of the data towards the market sectors that the compliant

Suppliers work in. TIG raised concerns that using this data as a basis for removing bias may have little effect, or even make the bias worse due to the natural skewing of the data.

It was therefore agreed that full data would be supplied for the purpose of assessing its fitness for purpose, but it would not feed into the factor calculation process at this time. In the meantime, the same partial implementation of the theft method that was used in the 2019/20 AUGS has been used for the factor calculations.

This data issue means that whilst all requested fields are present for the theft records supplied, only records involving actual investigations are included, and hence the following are missing:

- ETTOS leads that were not investigated
- TRAS Qualified Outliers that were not investigated

The loss of this data does not prevent the new theft analysis from being carried out, but it does limit its ability to remove the effects of Supplier theft detection targeting from the data. There are two steps in the theft detection process where bias is introduced by the Supplier's chosen theft strategy:

1. Generation of leads
   Leads from some sources result from a standard process applied equally across the whole population of meter point and hence are unbiased: they treat each meter point in the same manner and flag suspicious behaviour. The best example of this type of lead is the TRAS qualified outlier. Other leads, such as those generated by the Supplier's own analysis, are likely to be targeted towards certain elements of the population because of the composition of the Supplier's portfolio and any specific theft detection strategy they are following. These leads are biased and would skew the undetected theft split.

2. Investigation of leads
   Each Supplier chooses what leads to follow up, and each has their own strategy. Some generate large numbers of their own leads and base their investigations on these, whilst others do relatively little of their own analysis and base their investigations on TRAS leads. The choices Suppliers make in what leads to investigate can again lead to bias towards certain market sectors.

The data as supplied (i.e. investigations only) allows the first type of bias to be removed but not the second. This is because we have the lead source for each and so those from unbiased sources (e.g. TRAS) can be included and those from biased sources (e.g. own analysis) can be discounted. The fact that we only have data from investigations and not the full set of leads they came from means that we have no information about the Suppliers' strategy for the selection of investigations (i.e. the Lead→Investigation step) and cannot remove bias from this source.

Work will continue to obtain the full data and carry out the assessment of whether it is fit for purpose. The decision of whether to implement the full or partial theft method for the next draft of the AUG Statement will be made based on the results of this work.

7.9.9 Balancing Factor Split

Based on the methods described above, the split of the Balancing Factor between EUCs and Product Classes is as shown in Table 12 below.

	01B	02B	03B	04B	05B	06B	07B	08B	09B
Product 1	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Product 2	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Product 3	11.40%	0.23%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Product 4	81.85%	6.45%	0.07%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Table 12: Balancing Factor Split

7.10 Product Class Population and Aggregate AQ

The total population of meter points (either in terms of meter point count or aggregate AQ), split by EUC and Product Class, is used in a number of parts of the Unidentified Gas calculations. In all cases the forecasts used are for April 2021, and these are based on a number of instances of the asset data provided by the CDSP with figures then extrapolated to the midpoint of the forecast year. This method is described in Section 5.12 above. The areas these figures are used in are as follows:

- The principles for splitting IGT CSEPs and Shipperless/Unregistered Unidentified Gas between EUCs and Product Classes are based on meter point counts.
- The split of the Balancing Factor is based on the projection of unbiased theft to the forecast year, with the projection based on population change.
- The aggregate AQ by category is used as the denominator in the UIG factor calculations.

The instances of the asset data available begin at Nexus go-live in June 2017. The most recent data available is for November 2019. They allow trends to be established for each EUC and Product Class combination for extrapolation to the forecast year, and they also allow step changes to be detected and accounted for in the trends (such as the compulsory transfer of non-mandatory DM sites from PC1 to PC2).

CSEPs are included in all populations. The current AUG year is the second year for which these populations have been available on a site-by-site basis for direct inclusion in the figures: in previous years, an estimation process had to be applied to aggregate data (at the EUC level) for CSEPs.

Table 13 to Table 15 below show both the count of meters and aggregate AQ split by Product Class and EUC for Jun 2017 (actual data, Nexus go-live), for November 2019 (latest data), and for the forecast year (extrapolated data).

Number of Sites

	01B	02B	03B	04B	05B	06B	07B	08B	09B	Total
Product 1	21	15	18	44	59	203	189	207	267	1,023
Product 2	0	1	0	0	2	0	2	0	0	5
Product 3	53,920	12	1	0	1	0	0	0	0	53,934
Product 4	23,625,140	192,795	45,803	19,449	4,818	1,478	505	187	6	23,890,181
Total	23,679,081	192,824	45,822	19,493	4,881	1,681	696	394	273	23,945,143

Aggregate AQ (GWh)

	01B	02B	03B	04B	05B	06B	07B	08B	09B	Total
Product 1	0	3	9	65	251	2,108	3,895	8,899	39,309	54,539
Product 2	0	0	0	0	11	0	48	0	0	59
Product 3	771	1	1	0	3	0	0	0	0	776
Product 4	315,171	26,569	20,772	23,314	16,387	13,333	10,146	7,435	436	433,563
Total	315,943	26,573	20,782	23,379	16,652	15,440	14,089	16,334	39,745	488,937

Table 13: Meter Point Population and Aggregate AQ, June 2017

Number of Sites

	01B	02B	03B	04B	05B	06B	07B	08B	09B	Total
Product 1	0	0	0	0	0	0	0	0	311	311
Product 2	22	9	19	44	53	189	171	211	0	718
Product 3	2,144,181	44,442	14,165	4,301	877	149	61	28	0	2,208,204
Product 4	22,005,982	163,112	31,720	14,865	3,905	1,355	474	218	0	22,221,631
Total	24,150,185	207,563	45,904	19,210	4,835	1,693	706	457	311	24,430,864

Aggregate AQ (GWh)

	01B	02B	03B	04B	05B	06B	07B	08B	09B	Total
Product 1	0	0	0	0	0	0	0	0	39,764	39,764
Product 2	0	2	10	61	220	1,981	3,589	8,867	0	14,731
Product 3	28,469	6,664	6,257	5,027	2,881	1,355	1,226	1,058	0	52,937
Product 4	295,316	21,533	14,267	17,893	13,316	12,166	9,589	8,750	0	392,830
Total	323,785	28,199	20,534	22,981	16,417	15,502	14,404	18,675	39,764	500,262

Table 14: Meter Point Population and Aggregate AQ, November 2019

Number of Sites

	01B	02B	03B	04B	05B	06B	07B	08B	09B	Total
Product 1	0	0	0	0	0	0	0	0	311	311
Product 2	31	7	23	52	49	192	157	212	0	723
Product 3	3,983,947	60,707	16,083	5,501	1,165	131	65	28	0	4,067,627
Product 4	20,449,463	158,633	29,904	13,468	3,572	1,379	481	250	0	20,657,150
Total	24,433,441	219,347	46,010	19,021	4,786	1,702	703	490	311	24,725,811

Aggregate AQ (GWh)

	01B	02B	03B	04B	05B	06B	07B	08B	09B	Total
Product 1	0	0	0	0	0	0	0	0	39,698	39,698
Product 2	0	1	11	72	193	2,024	3,348	8,957	0	14,607
Product 3	52,761	8,936	7,091	6,443	3,797	1,194	1,301	1,015	0	82,538
Product 4	277,580	20,622	13,283	16,245	12,209	12,381	9,736	9,763	0	371,817
Total	330,341	29,559	20,385	22,760	16,199	15,600	14,385	19,734	39,698	508,660

Table 15: Projected Meter Point Population and Aggregate AQ, April 2021

7.11 Extrapolation to Forecast Period

The training period for the Unidentified Gas calculations consists of a maximum of six years' data, running from 2011/12 to 2016/17. Data for the Consumption Method for calculating total Unidentified Gas is available for this length of time. Training data for individual directly-calculated elements of Unidentified Gas is limited by the time period for which snapshots are available: for these, the snapshot files typically do not start as early as 2011/12, but run beyond 2016/17, with Shipperless/Unregistered snapshots available until September 2019.

The analysis of Unidentified Gas is carried out on an LDZ by LDZ basis, due to the fact that it is dependent on identifying and extrapolating trends in each Unidentified Gas component and these trends differ across LDZs. The final UIG factors are based on the combined national total Unidentified Gas, however, and so figures are aggregated across all LDZs once they have been calculated to achieve this.

The total Unidentified Gas is first calculated for the training period using the Consumption Method, which is described in detail in Section 7.2, and the permanent Unidentified Gas total for each year is obtained by subtracting temporary Unidentified Gas from it. For the forecast year, the total permanent Unidentified Gas is obtained by extrapolating permanent Unidentified Gas from each category (both the Balancing Factor and the directly calculated elements) forward to the relevant year and summing the resultant values. These Unidentified Gas energy values are then converted into factors using the method described in Section 7.12 below. There is no requirement for an additional independent estimation of the permanent Unidentified Gas total for the forecast year.

The method of extrapolation differs depending on the Unidentified Gas category. This section contains a description of the process for each category to clarify how permanent Unidentified Gas is calculated for the forecast year.

7.11.1 Unregistered and Shipperless

The principles of extrapolating this element of Unidentified Gas are based around creating piecewise trends for each Unregistered/Shipperless category and LDZ/EUC/Product Class combination and using these to make estimates for the training years and the forecast year. Piecewise fits are required because of the effects of Mods 0410A, 0424 and 0425, all of which affect Unregistered or Shipperless Unidentified Gas. The implementation of these Mods during the training period results in the Unidentified Gas pattern changing at specific points (e.g. the implementation dates). As many individual time segments as necessary are used to accurately model the Unidentified Gas pattern for each combination. Many segments may be required for an individual sub-category, as described and illustrated in Section 7.3.

The fit for the most recent snapshots is used to extrapolate forward to the forecast year in each case with the maximum degree of accuracy. In order to capture all Unregistered/Shipperless categories, LDZs, EUCs and Product Classes, a total of 1872 piecewise trends are required and hence it is not possible to present them in this document. Excel files containing all of this information are available to code parties on request.

The only exception to the above method is for the "Awaiting GSR Visit" category, which uses different input data. For this category, it is assumed that for the forecast year sites will continue to become Shipperless at the present rate. Within this context, the effects of Mods 0424 and 0425 are then modelled as follows:

- For the PTS element of "Awaiting GSR Visit" Unidentified Gas, all Unidentified Gas from this source is temporary from 2014/15 onwards.
- For the SSrP element of "Awaiting GSR Visit" Unidentified Gas, following Nexus go-live, any site without a record of the installation date of the new meter will contribute permanent Unidentified Gas until its GSR visit regardless of its Isolation date. It is assumed that new meters are installed at a steady rate for these sites and as such the average period of the year for which a meter was present and capable of flowing gas is 6 months (i.e. 50% of the year). The following calculation steps are applied:
  - Multiply the Shipperless SSrP Unidentified Gas element by the "% of Year with Meter" factor. This gives the total (permanent plus temporary) Unidentified Gas for the Shipperless SSrP element of "Awaiting GSR Visit".
  - Multiply this total by the "% of Meters without Install Date" factor. This gives the Permanent element of the Unidentified Gas total from this source. The remainder of the Unidentified Gas for the Shipperless SSrP element of "Awaiting GSR Visit" is temporary.

7.11.2 IGT CSEPs

The principle of defining trends for the training period, as described above for Unregistered/Shipperless sites, is also used for IGT CSEPs. In this case, 468 individual trends are required to cover each LDZ/Product Class/EUC combination. Whilst this approach provides a method of extrapolating to the forecast period, which was used in the Unidentified Gas analyses for pre-Nexus years, this is not necessary for this Unidentified Gas category post-Nexus and the trends are only used for the training period. This is due to the fact that only permanent Unidentified Gas is calculated for the forecast period and under Nexus all IGT CSEPS Unidentified Gas is temporary.

For the training period, only Unidentified Gas arising from EUC 01B is permanent, whilst the remainder is temporary. It is not possible to present all of the trends in this document, but Excel files containing this information are available on request.

7.11.3 Consumer Metering Errors

For previous AUG years, this process has been assumed to be steady, and the figures calculated from the latest dataset have been applied across all training and forecast years. A number of instances of this dataset are now available, however, which allows a trend to be established for each EUC/Product Class combination. These trends are used to extrapolate to April 2021, this midpoint of the forecast year, and these figures are used in the forecast Unidentified Gas calculations.

7.11.4 Detected Theft

Detected theft is mostly temporary in nature and hence is used only in the calculations for the training period. Any permanent Unidentified Gas arising from theft detected after the line-in-the-sand is taken forward to the forecast year via the Balancing Factor.

7.11.5 CSEP Shrinkage

Shrinkage estimates for the forecast year are available in the annual statements from the Gas Transporters [19], and these are used as the basis for the forecast year calculations. The best estimate for the percentage CSEP Shrinkage is applied to these and then split by EUC/Product as described in Section 7.7 above.

For the CSEP Shrinkage calculation, for the training period only a single total for Unidentified Gas from this source is required. For the forecast year a split by EUC/Product is required, and this is done on the basis of CSEP throughput (as opposed to overall throughput). The latest available figures for CSEP throughput split by EUC and Product (as described in Section 7.4.3 above) are used for this calculation.

7.11.6 Volume Conversion Error

The volume conversion error relating to atmospheric pressure and gas temperature are calculated individually for each gas year of the training period. The average from the six years of the training period (2011/12 to 2016/17) has been used as the value to take forward to the forecast year. This value is 610.7GWh.

For the forecast year, a split by EUC/Product is required. This is based on the aggregate AQ of meters without volume conversion by EUC and product Class

The estimate of volume conversion error resulting from using the standard CF rather than a site specific CF for meters in EUC04 and above is estimated as 31.7GWh. This is split by EUC and product class based on the ratio of aggregate AQ for meters using the standard CF in EUC04 and above without volume conversion to the aggregate AQ for meters in EUC04 and above without volume conversion.

The final step is to take the aggregate volume conversion error for pressure, temperature and standard CF by PC/EUC and rescale this based on the aggregate AQ by PC/EUC projected to the forecast year. The total UIG from volume conversion remains the same but the split by PC/EUC will change based on the predicted Product Class populations for April 2021.

7.11.7 Balancing Factor

The Balancing Factor is calculated individually for each gas year of the training period by subtracting temporary Unidentified Gas plus all directly estimated permanent Unidentified Gas from the total Unidentified Gas derived using the Consumption Method. It should be noted, however, that the partitioning into gas years is arbitrary and relies on the allocation algorithm to assign consumption to individual gas years. It is therefore important to use multiple gas years in the calculation of the Balancing Factor to correct for any misallocation of consumption between gas years. The average from the six years of the training period (2011/12 to 2016/17) has been used as the value to take forward to the forecast year. This value is 7,413 GWh.

7.11.8 Total Permanent Unidentified Gas

The total permanent Unidentified Gas for the forecast year, split by EUC/Product Class, is the sum of the Balancing Factor and each directly calculated category of Unidentified Gas. Each of these has been estimated for the forecast year as described above.

The forecast of total permanent Unidentified Gas for 2020/21 is 8,097 GWh.

7.12 UIG Factor Calculation

The final step in the calculation process is the production of the UIG factors. These are a fundamental link between the population of the EUC/Product Class combination and the Unidentified Gas from it.

They are therefore calculated using the detailed estimates of the value of Unidentified Gas (for the year in which the factors will be in force) described in Section 7.11 above. Once the Unidentified Gas for each EUC/Product combination for the forecast year has been estimated, this is converted into a factor by dividing it by the aggregate AQ for that EUC/Product (i.e. the best estimate of the AQ for that EUC/Product combination for the forecast year). This AQ data has been supplied to the AUG Expert by the CDSP.

Therefore, each factor is calculated as follows:

$$\textit{UIG Factor}_{PRODUCT,EUC} = \textit{Unidentified Gas (GWh)}_{PRODUCT,EUC} \,/\, \textit{Aggregate AQ (TWh)}_{PRODUCT,EUC} \times 50$$

The Unidentified Gas term and the Aggregate AQ term have different units in this calculation (GWh and TWh respectively) and the raw output is also multiplied by a factor of 50. These steps are to ensure that the resulting factors give sufficient precision when expressed to 2 decimal places, which is a requirement for the output.

The final step of the factor calculation is to apply a smoothing process as described in Section 4.3.5 above. This is achieved using individual cubic fits for each Product Class and ensures a smooth transition of the factors from EUC to EUC, removing any random variation that has arisen from the statistical modelling process. The final factors presented in Section 8 below represent the output from the smoothing calculations.

8 UNIDENTIFIED GAS FACTORS

The final Unidentified Gas factors, calculated as described in this document, to apply to the 2020/21 gas year are as follows:

Supply Meter Point Classification	Class 1	Class 2	Class 3	Class 4
EUC Band 1	0.20	4.91	50.76	122.53
EUC Band 2	0.20	4.91	13.20	133.14
EUC Band 3	0.20	4.54	8.89	15.16
EUC Band 4	0.20	3.59	8.89	11.67
EUC Band 5	0.20	2.36	8.30	8.00
EUC Band 6	0.20	1.16	5.98	4.79
EUC Band 7	0.20	0.31	5.17	2.47
EUC Band 8	0.20	0.12	0.28	1.48
EUC Band 9	0.20	0.20	0.20	0.20

Table 16: Final AUG Table

These factors are based on updated consumption estimates calculated in 2019 using the most recent meter asset data and meter read data.

The most striking feature of these figures is that, as for the 2019/20 AUG year, the factors for EUCs 01B and 02B in Product Class 4 are much higher than those for any other category. Their relative magnitude has changed, however: in 2019/20, the PC4 01B factor was higher, whilst for 2020/21 it is the PC4 02B factor that is slightly higher. In addition, for 2020/21 the factor for PC3 01B has risen. Other factors have remained at a level broadly comparable to 2019/20.

All of these effects are due to the mass migration of sites from PC4 to PC3. The detailed reasons are as follows:

- The population projection for April 2021 in Section 7.10 above shows that approximately 4m sites are expected to transfer to PC3 01B by this time. Whilst theft from Smart Meters is lower than that from traditional meters, this inevitably includes sites where theft is taking place, and as such the number of confirmed thefts from PC3 has risen. This has been extrapolated to the expected population level for April 2021 and accounts for the rise in the factor for PC3 01B compared to the previous year. The fact that this factor is considerably lower than the one for PC4 01B reflects the relative level of theft from Smart Meters and traditional meters. Theft from commercial sites with Smart Meter or AMR is extremely rare, with only 7 confirmed thefts from such sites since TRAS began in 2016. Whilst the migration of sites to PC3 is less pronounced for EUC 02B, it is still present and the number of PC3 EUC 02B sites is expected to double by April 2021. The very low level of theft from 02B sites with Smart Meter or AMR means that the vast majority of theft from 02B remains in PC4, but with a reduced overall population. This has led to the factor for this category increasing so that it is now the highest individual factor.

Table 17 below shows the relative contribution to the permanent UIG(f) total for the forecast year from each Unidentified Gas source.

Unidentified Gas Source	Contribution (%)
Total Directly Measured	**9.48%**
- CSEP Shrinkage	*0.16%*
- Shipperless/Unregistered	*1.04%*
- Consumer Meter Errors	*0.43%*
- Volume Conversion	*7.84%*
Balancing Factor	**90.52%**

Table 17: Contribution to Unidentified Gas by Source

9 CONSULTATION QUESTIONS AND ANSWERS

This section captures a history of the questions raised by industry bodies during consultation periods and the AUG Expert's responses. These currently relate to the 2017/18, 2018/19 and 2019/20 AUG Statements. The questions have been assessed against the AUGE Guidelines [1] and responses provided as appropriate. All questions and answers have also been published on the Joint Office website.

Due to the in-depth nature of the questions raised and the detailed responses required, it is not appropriate to publish full transcripts in this document. Instead, this section contains a summary of the organisations that provided questions. The questions themselves and their associated responses can be found in the following external documents:

- "AUGS 2017-18 BG Response - DNVGL Comments" [20]
- "AUGS 2017-18 ICoSS Response - DNVGL Comments" [21]
- "AUGS 2017-18 E.ON Response - DNVGL Comments" [22]
- "AUGS 2018-19 British Gas - DNV GL Comments" [46]
- "AUGS 2018-19 Corona Response - DNV GL Comments" [47]
- "AUGS 2018-19 Eon Response - DNV GL Comments" [48]
- "AUGS 2018-19 ICOSS Response - DNV GL Comments" [49]
- "AUGS 2018-19 General Query - DNV GL Comments" [50]
- "AUGS 2019-20 British Gas - DNV GL Comments" [52]
- "AUGS 2019-20 ICoSS - DNV GL Comments" [53]
- "AUGS 2019-20 Scottish Power - DNV GL Comments" [54]
- "AUGS 2019-20 Shipper1 - DNV GL Comments" [55]

Note that all responses contained in these documents relate to the Unidentified Gas calculations at the time they were written, rather than reflecting the process as it currently stands. Therefore, wherever information differs between the responses and the latest AUGS, this is because the Unidentified Gas analysis has evolved and information in the response documents has been superseded. The information supplied in the latest version of the AUGS is always the most up-to-date.

Organisation Name	Date of Communication
British Gas	15/03/2017
ICoSS	15/03/2017
E.ON	15/03/2017
British Gas	09/04/2018
Corona	09/04/2018
E.ON	09/04/2018
ICoSS	09/04/2018
Undisclosed	18/04/2018
British Gas	21/01/2019

Organisation Name	Date of Communication
ICoSS	22/01/2019
Scottish Power	04/02/2019
Undisclosed	21/01/2019

Table 18: Responses to the Proposed 2017/18, 2018/19 and 2019/20 AUG Statements

10 CONTACT DETAILS

Questions can be raised with the AUG Expert at AUGE.software@dnvgl.com

11 REFERENCES

[1] Framework for the Appointment of an Allocation of Unidentified Gas Expert V7.0, 20th April 2017

[2] Uniform Network Code – Transportation Principal Document Version 5.04, 7th June 2017

[3] Mod 0432: Project Nexus – Gas Demand Estimation, Allocation, Settlement and Reconciliation reform, 17 January 2014

[4] Mod 0473/0473A: Project Nexus – Allocation of Unidentified Gas, 20 November 2014

[5] Standard Conditions of Gas Supply Licence (Gas Act 1986)

[6] Review of Shrinkage Calculations in Shrinkage Leakage Model, DNV GL, June 2017

[7] Mod 0424: Re-establishment of Supply Meter Points – prospective measures to address Shipperless sites, 20 December 2012

[8] Mod 0425V: Re-establishment of Supply Meter Points – Shipperless sites, 7 January 2014

[9] Mod 0410/0410A: Responsibility for gas off-taken at Unregistered Sites following New Network Connections, 12 July 2013

[10] 2011 Allocation of Unidentified Gas Statement for 2012/13

[11] Mod 0594R: Meter Reading Submission for Advanced & Smart Metering, 15 September 2016

[12] Mod 0572: Amendment to the definition of AUG Year within UNC TPD Section E

[13] IGT 053: Introduction of annual updates to the AQ values within the CSEP NExA table, 1 October 2014

[14] Mod 0398: Limitation on Retrospective Invoicing and Invoice Correction (3 to 4 year solution), 1 April 2014

[15] Joint Gas Distribution Network (GDN) Response to the Energy UK Gas Retail Group Study into the effect of shrinkage on domestic customers, November 2016

[16] Energy UK Gas Retail Group Study into the effect of shrinkage on domestic customers, October 2015

[17] Shrinkage Leakage Model Review, December 2017

[18] Department for Business, Energy & Industrial Strategy: Smart Meters – Quarterly Report to end September 2018, 29 November 2018

[19] LDZ Shrinkage Quantity Proposals, 2009/10 to 2017/18, http://www.gasgovernance.co.uk/Shrinkage/Shrinkage-Quantity-Proposals

[20] AUGS 2017-18 BG Response - DNVGL Comments, April 2017

[21] AUGS 2017-18 ICoSS Response - DNVGL Comments, April 2017

[22] AUGS 2017-18 E.ON Response - DNVGL Comments, April 2017

[23] Allocation of Unidentified Gas Statement for 2017/18, DNV GL, 28/04/2017

[24] Gas Energy Measurement in Consumer Billing, D.F. Lander, 23 July 2014

[25] Accuracy of CV Determination Systems for Calculation of FWACV, D.F. Lander, 22 May 2012

[26] The Gas (Calculation of Thermal Energy) Regulations 1996

[27] Code of Practice for Gas Meter Asset Managers, Edition 10, 3 Nov 2017

[28] https://www.getthedata.com/open-postcode-elevation

[29] Correction Factor Application, UK Link Programme Change Request Form, UKLP CRDBI065, 9 Jan 2015, http://www.gasgovernance.co.uk/sites/default/files/ggf/UKLP065%20-%20Correction%20Factor%20application%20CR.pdf

[30] Written evidence submitted by Ray Cope to the Energy and Climate Change Committee, January 2013, https://publications.parliament.uk/pa/cm201314/cmselect/cmenergy/108/108vw02.htm

[31] Modification NC0396, Proposal to allow for RbD Adjustments to be processed consequent to the recalculation of LDZ Specific Shrinkage Factors for revised temperature data for the period from the implementation of RbD to the start of current Supply Year 1999/2000, Transco, 4 August 2000, http://www.gasgovernance.co.uk/index.php/NCMP/400

[32] Final Proposed LDZ Shrinkage Factors North East and Northern LDZ Gas Year 2005/06, Northern gas Networks, 30 August 2005

[33] Suspension of the Validation between Meter Index and Unconverted Converter Index, DSC Change Proposal XRN4621, 23 February 2018

[34] Gas energy measurement, A consultation document, Ofgem, November 2000

[35] Summary Report for Domestic Gas Temperature Survey for Shippers (2000), BG Technology

[36] The derivation of LDZ gas temperatures for the period 1996-2000, L M Wallis & D Gregory, Advantica R5181, February 2002

[37] Examination of environmental factors affecting gas metering accuracy, L M Wallis, BG Technology, 1998

[38] The domestic meter temperature survey, BG Technology, 6 July 2000

[39] Framework for the Appointment of an Allocation of Unidentified Gas Expert, V8.1, 8 June 2018

[40] Electronic gas volume conversion systems, IGEM/GM/5 Edition 3

[41] Gas meters – Conversion Devices, Part 1: Volume conversion, BS EN 12405-1:2005+A2:2010

[42] Misreading of Conventional Diaphragm Domestic Gas Meter, Rene Hermkens, KIWA Technology, 2012 Kuala Lumpur World Gas Conference

[43] Framework for the Appointment of an Allocation of Unidentified Gas Expert V8.1, 8 June 2018

[44] Open letter: Gas energy measurement in consumer billing, Ofgem, 21 August 2014

[45] UIG Task Force Update V1.1, Xoserve, 13 Nov 2018, https://www.xoserve.com/wp-content/uploads/UIG_Nov_Sian_Update-v1.1.pdf

[46] AUGS 2018-19 British Gas - DNV GL Comments, April 2018

[47] AUGS 2018-19 Corona Response - DNV GL Comments, April 2018

[48] AUGS 2018-19 Eon Response - DNV GL Comments, April 2018

[49] AUGS 2018-19 ICOSS Response - DNV GL Comments, April 2018

[50] AUGS 2018-19 General Query - DNV GL Comments, April 2018

[51] UNC642 and 642A: 'Changes to settlement regime to address Unidentified Gas issues'; and UNC643: 'Changes to settlement regime to address Unidentified Gas issues including retrospective correction'. OFGEM, 7th June 2018

[52] AUGS 2019-20 British Gas - DNV GL Comments, March 2019

[53] AUGS 2019-20 ICOSS Response - DNV GL Comments, March 2019

[54] AUGS 2019-20 Scottish Power Response - DNV GL Comments, March 2019

[55] AUGS 2019-20 Shipper1 Response - DNV GL Comments, March 2019

[56] Gesche M. Huebner, Megan McMichael, David Shipworth, Michelle Shipworth, Mathieu Durand-Daubin & Alex J. Summerfield (2015) The shape of warmth: temperature profiles in living rooms, Building Research & Information, 43:2, 185-196, DOI:10.1080/09613218.2014.922339

[57] UK shallow ground temperatures for ground coupled heat exchangers, Jon Busby, British Geological Survey

[58] 'Final Allocation of Unidentified Gas Statement for 2019/20', DNV GL, 27 March 2019.

[59] 'Domestic Meter Temperature Monitoring Exercise analysis of returned data', M. C. Gaskell, BG Technology, June 1999

[60] 'Summary Report of the Domestic Gas Temperature Survey (2000)', M.C. Gaskell; M.R. Ranzetta, BG Technology, August 2000

[61] 'The reality of English living rooms – A comparison of internal temperatures against common model assumptions', Gesche M. Huebnera, Megan McMichaela, David Shipworth, Michelle Shipworth, Mathieu Durand-Daubin, Alex Summerfield, Energy and Buildings 2013

[62] VADODARIA, K., LOVEDAY, D.L. and HAINES, V., 2014. Measured winter and spring-time indoor temperatures in UK homes over the period 1969-2010: A review and synthesis. Energy Policy, 64, pp. 252 – 262

GLOSSARY

AGI	Above Ground Installation
ALP	Annual Load Profile (deeming algorithm parameter)
AMR	Automated Meter Reading
AQ	Annual Quantity. An estimate of annual consumption under seasonal normal conditions
AUG	Allocation of Unidentified Gas
AUGE	Allocation of Unidentified Gas Expert
AUGS	Allocation of Unidentified Gas Statement
Balancing Factor	An aggregate of the combined Unidentified Gas of various items calculated by subtraction. This includes theft, errors in the shrinkage estimate, open bypass valves, meters "Passing Unregistered Gas", unknown sites, and additional common cause variation.
BEIS	The Department for Business, Energy and Industrial Strategy
CMA	Competition & Markets Authority
Consumption Method	Unidentified Gas methodology using meter reads and metered volumes
CSEP	Connected System Exit Point
CV	Calorific Value
CWAALP	Cumulative Weather Adjusted Annual Load Profile
CWV	Composite Weather Variable
DAF	Daily Adjustment Factor (deeming algorithm parameter)
DM	Daily Metered
DME	Daily Metered Elective. A site below the DM mandatory threshold of 58,600,000 kWh which the shipper has elected to be DM. The meter read equipment is provided by the shipper.
DMM	Daily Metered Mandatory. A site with an AQ above the DM mandatory threshold of 58,600,000 kWh.
DMTS	Domestic Meter Temperature Survey
DMV	Daily Metered Voluntary. A site below the DM mandatory threshold of 58,600,000 kWh which is voluntarily DM. The meter read equipment is provided by the transporter.
ECV	Emergency Control Valve
EMIB	Energy Market Issues for Biomethane Projects
ETM	Electronic Token Meter

EUC	End User Category
EWCF	Estimated Weather Correction Factor (deeming algorithm output - alternative to WCF based on CWV rather than demand)
Found Meter	A meter being supplied by a Shipper but for which the CDSP have no record
FWA	Flow Weighted Average
FWAAT	Flow Weighted Average Air Temperature
FWAGT	Flow Weighted Average Gas Temperature
FWAP	Flow Weighted Average Pressure
GFD	Gas Flow Day
GSR	Gas Safety Regulations
ICoSS	The I&C Shippers and Suppliers Group
ICTS	Industrial & Commercial Temperature Study
IGT	Independent Gas Transporter
Isolated Meter	A meter that has been disabled (through capping or clamping) and hence is no longer capable of flowing gas, and this information has been conveyed to the CDSP and recorded on their system.
LDZ	Local Distribution Zone
LSP	Larger Supply Point
MAM	Meter Asset Manager
Model Error	The statistical error associated with any modelling or estimation process. It an inherent part of any statistical model and does not imply that the model itself is inadequate or incorrect.
MPRN	Meter Point Reference Number
MSL	Mean Sea Level
NDM	Non-Daily Metered
ODR	OFGEM Data Request
OUG	Own Use Gas
PC	Product Class
PSND	Pseudo Seasonal Normal Demand, calculated using AQ values rather than being based on historic metered demands
PTS	Passed To Shipper
RbD	Reconciliation by Difference
RPD Meter	Rotary Positive Displacement meter
SAP	System Average Price

SF	Scaling Factor (deeming algorithm output)
SNCWV	Seasonal Normal Composite Weather Variable
SND	Seasonal Normal Demand
SPAA	The Supply Point Administration Agreement
SPC	Supply Point Component
SSP	Smaller Supply Point
SSrP	Shipper Specific rePort
TPD	Transportation Principal Document (of UNC)
TRAS	Theft Risk Assessment Service
UIP	Utility Infrastructure Provider
UNC	Uniform Network Code
UIG	UNC definition of total Unidentified Gas calculated on a daily basis as part of the settlement process
WAALP	Weather Adjusted Annual Load Profile
WCF	Weather Correction Factor (deeming algorithm output)
WSENS	Weather Sensitivity (deeming algorithm parameter used in EWCF definition, reflecting the sensitivity of an EUC to difference in CWV from seasonal normal)

APPENDIX A
Raw Data

This appendix describes the raw data provided by the CDSP for the Consumption Method.

ALLOCATIONS

This data contains all allocations including CSEPs from 01/10/2008 onwards.

Name	Description
GAS_DAY	Date - Gas day for which allocation applies
LDZ	Char[2] - LDZ identifier e.g. EA
EUC	Char[11] - Full EUC Code e.g. WM:E0708W02
ALLOCATED_ENERGY	Number - Final allocated energy value (kWh). Includes CSEPs

ANNUAL_QUANTITY

This data includes all meter points active at any point from 01/10/2008 onwards, not just those currently live. It includes all within gas year updates, appeals etc.

Name	Description
MPR_ID	Number – Unique dummy ID for meter point which is used consistently throughout the data
AQ_EFFECTIVE_DATE	Date - Date on which AQ becomes effective
EUC	Char[11] - Full EUC Code e.g. WM:E0708W02
AQ	Number - Annual Quantity to apply from effective date (kWh)
SITE_TYPE_FLAG	Char[1] - Indicator ="N" for NDM meter point, "D" for DM meter point or "U" for Unique site
CLASS	Number – Meter point product class

CSEPS

This data contains information for gas year 2008 onwards.

Name	Description
GAS_YEAR	Date - Gas year for which CSEP AQ/Numbers apply
EUC*	Char[11] - Full EUC Code e.g. WM:E0708W02
TOTAL_AQ	Number - Aggregate CSEP AQ at start of gas year
COUNT_OF_SUPPLY_POINTS	Number - Count of supply points at start of gas year

* Note that the EUC classification for CSEPs is based on a nominal maximum AQ

The CDSP also provide the success rate for the AQ calculation process. This is used in the process to adjust CSEP consumption for AQ bias.

FACTORS

This data is provided from 01/10/2008

Name	Description
LDZ	Char[2] - LDZ identifier e.g. EA
EUC	Char[11] - Full EUC Code e.g. WM:E0708W02
GAS_DAY	Date - Gas day for which factors applies
ALP	Number - Annual Load Profile (6 d.p.)
DAF	Number - Daily Adjustment Factor (6 d.p.)
EWCF	Number - Estimated Weather Correction Factor (8 d.p.)
CV	Number - Calorific Value (1 d.p.)

METER_ERRORS

All meter error adjustments from 01/04/2008 onwards. In addition to these resolved meter errors, any open errors will be taken from the JoT website. Where appropriate, the AUG Expert will clarify these with the CDSP.

Name	Description
BILLING MONTH	Month and year where billing correction was applied for the given meter error
LDZ	Char[2] - LDZ identifier e.g. EA
AGGREGATE ENERGY	Total correction for period of error in kWh
REASON	Reason for adjustment
ADJUSTMENT	Value of adjustment in kWh over billing period. Positive value represents an over-read

METER_READS

This data includes all meter reads from 01/10/2008 onwards. Multiple records for a meter point with the same date are filtered by the CDSP using the following methodology.

Where there is an A (Actual) Read Type and an E (Estimate) Read Type the CDSP remove the E and retain the A Read. Where there are Read Types of R (Replacement) the CDSP retain this read and remove the original read type that it replaced. Where there are multiple R Reads they are ranked by number e.g. R01 and R02 and the highest number is the latest replacement read that is retained.

Name	**Description**
MPR_ID	Number - Unique dummy ID for meter point which is used consistently throughout the data
START_READ_DATE	Date – Start date of metered period
METER_READ_DATE	Date - Date of meter read
IMP_IND	Char[1] - Indicator ="Y" for imperial meter read, else "N"
METER_READ_VAL	Number - Value of meter read
METERED_VOL	Number - volume of gas since previous meter read in units appropriate for meter (imperial or metric)
ROUND_THE_CLOCK_ IND	Number – Number of times the meter index has passed zero since the last read.
AQ	Number – Prevailing Annual Quantity at time of meter read (kWh)
METER_READ_FREQ	Char[1] - Indicator for frequency of meter reads (A-Annual, 6-6 monthly, M-monthly)
SSP_LSP	Char[3] - "SSP" or "LSP"
EUC	Char[11] - Full EUC Code e.g. WM:E0708W02
READ_TYPE_CODE	Char[4] - Code for type of meter read

METER_INFO

This data includes all available meter asset data.

Name	**Description**
MPR_ID	Number - Unique ID for meter used across ALL data
LDZ	Char[2] - LDZ identifier e.g. EA
NUM_DIALS	Number - Number of meter dials
IMP_IND	Char[1] - Indicator ="Y" for imperial meter read, else "N"

METER_FITTED_DATE	Date - Date meter was fitted
UNITS	Number - Multiplier for meter read units (1, 10, 100 etc.)
CORRECTION_FACTOR	Number – Volume correction factor
STATUS_UPDATE_DATE	Date – Date of record
METER_STATUS	Char[27] – e.g. Live, Removed
READ_FACTOR	Number
METER_MECHANISM	Char[30]
CSEP_FLAG	Char[1] - Indicator ="Y" if meter point is on a CSEP, else "N"
DERIVED_METER_TYPE	Char[30]
ELEVATION	Number

NEW_LOST_SITES

This data contains all meter points with a first confirmation date or an end date from 01/10/2008 onwards.

Name	Description
MPR_ID	Number - Unique dummy ID for meter point which is used consistently throughout the data
START_DATE	Date - First confirmation date for meter point
END_DATE	Date - Date meter point was excluded from allocations process

PRIMES

This data includes details of all prime meter points active at any time since 01/10/2008.

Name	Description
MPR_ID	Number - Unique dummy ID for meter point which is used consistently throughout the data
LDZ	Char[2] - LDZ identifier e.g. EA

SUBS

This data includes details of all sub-prime meter points active at any time since 01/10/2008.

Name	Description
MPR_ID	Number - Unique dummy ID for meter point which is used consistently throughout the data
PRIME_MPR_ID	Number – Dummy ID for the prime meter

A list of re-confirmation dates has also been provided for meter points which were previously in a prime/sub configuration but are no longer.

APPENDIX B
Data Pre-processing

This appendix describes the various processes used to derive information from the raw data prior to carrying out the Consumption Method calculation.

LDZ Movers

Some meters have records associated with more than one LDZ. These meters are identified and their MPR_ID is recorded in a separate table which also includes the latest LDZ they are recorded against and any previous LDZs. Any data associated with a meter which is not for the latest LDZ is then copied to the appropriate LDZ. This is required as processing is done on an LDZ by LDZ basis with data held in separate LDZ specific tables. This copying of data ensures that when the meter consumption is calculated in the latest LDZ, all of the relevant data is present. To ensure no double counting occurs, the SITE_LIST tables are populated to ensure that each meter is only present in one LDZ (its latest LDZ).

Identification of Bad Meter Reads

The algorithm for flagging bad reads is as follows:

Given sequential meter reads mr_1, mr_2, mr_3 and mr_4 calculate:

$$con_1 = mr_2 - mr_1$$

$$con_2 = mr_3 - mr_2$$

$$con_3 = mr_4 - mr_3$$

If any of these are negative, we check for meter index rollover

If the meter was replaced, we leave the consumption null

Then if the meter was not replaced during the period we check

- If ($con_3 > 0$) and ($con_2 < 0$) and ($con_1 > 0$) then we have a bad reading
- If $con_1 > \text{abs}(con_2)$ then mr_2 is bad
- Else if $con_3 > \text{abs}(con_2)$ then mr_3 is bad

Identification of Meter Read Gaps and Overlaps

For each meter read the number of days between the start read date for a record and the meter read date from previous record is calculated.

- A positive value indicates a gap
- A negative value indicates an overlap

Note, start read dates are only available for the records provided from 2013 onwards.

NDM DM Changes

Using the site type flag from the AQ records meter points are identified which are recorded as both DM and NDM. The start and end date (where relevant) of the meter point's status as NDM are then recorded.

APPENDIX C
Gas Temperature Data From DMTS & ICTS

The AUG Expert has used 2 key sources of gas temperature data. A summary of the gas temperatures reported by these is provided in the following sections

The Domestic Meter Temperature Survey (DMTS)

Although the AUG Expert does not have access to the raw data and results of the DMTS, a summary of the findings are provided in the final report[60]. The FWAGT data for the 1999-2000 Common Dataset are reproduced here (Table 19) for reference including standard deviation, sample size and the error based on 95% confidence intervals. All values are as reported in the DMTS report and are not calculated by the AUGE.

LDZ	Meter Location	FWAGT	Standard Deviation	Sample Size	Error (±)
EA	External	9.37	1.31	170	0.2
	Internal	15.12	4.1	34	1.38
EM	External	9.11	1.57	236	0.2
	Internal	13.7	3.29	112	0.61
NE	External	8.79	1.27	97	0.25
	Internal	13.47	2.64	54	0.7
NO	External	8.5	1.33	97	0.26
	Internal	13.19	2.98	27	1.13
NT	External	10.13	1.62	173	0.24
	Internal	16.43	3.48	236	0.44
NW	External	9.01	1.23	296	0.14
	Internal	13.07	2.87	230	0.37
SC	External	7.95	1.12	215	0.15
	Internal	16.92	3.98	58	1.02
SE	External	10.16	1.39	164	0.21
	Internal	16.1	4.6	92	0.94
SO	External	9.74	1.5	162	0.23
	Internal	15.42	3.09	55	0.82
SW	External	9.53	1.31	134	0.22
	Internal	13.56	4.23	27	1.6
WM	External	9.26	1.34	145	0.22
	Internal	12.86	2.87	86	0.61
WN	External	9.33	1.13	91	0.23
	Internal	12.6	2.07	7	1.54
WS	External	9.86	0.88	81	0.19
	Internal	14.66	1.45	11	1.42

Table 19: Summary of Measured FWAGT from DMTS for 1999-2000 Common Set

The Industrial & Commercial Temperature Survey (ICTS)

The study into gas temperatures carried out in 2002[36] provides FWAGTs for 4 categories of supply point. The FWAGTs are reproduced here for reference in Table 20.

LDZ	Domestic (derived)	Small I&C	Large I&C	DM
EA	9.40	9.60	10.10	11.10
EM	10.10	10.10	10.90	12.10
NE	9.40	9.30	9.90	11.20
NO	9.00	8.80	9.40	10.50
NT	12.80	13.30	13.40	14.80
NW	9.70	9.70	10.40	11.40
SC	8.30	8.40	8.80	9.90
SE	10.70	11.20	11.50	13.00
SO	9.70	9.70	10.60	11.80
SW	10.10	10.10	11.00	12.10
WM	8.90	8.90	10.00	10.70
WN	9.00	9.00	9.90	10.70
WS	10.60	10.40	11.30	12.60

Table 20: Summary of Measured FWAGTs from ICTS

About DNV GL

Driven by our purpose of safeguarding life, property and the environment, DNV GL enables organizations to advance the safety and sustainability of their business. We provide classification and technical assurance along with software and independent expert advisory services to the maritime, oil & gas and energy industries. We also provide certification services to customers across a wide range of industries. Operating in more than 100 countries, our professionals are dedicated to helping our customers make the world safer, smarter and greener.