



UNC Workgroup 0754R

June 2022

Contents

- Workgroup Meeting 6 Recap (22nd March 2022)
 - Key Discussion Points
- Introduction to Area 2: Improve validation processes
- Update on Area 1: Trial Alternative approaches to deriving SND_t
- Conclusions and Next Steps
- Timeline



Glossary

Useful Links

- [Uniform Network Code Section H](#)
- [Demand Estimation Methodology](#)
- [Demand Modelling Approach \(2021 version\)](#)
- [UIG Task Force Findings](#)
- [NDM Algorithm Consultation Material](#)
- [UNC Request for 0754R Workgroup](#)

For those not familiar with all the industry abbreviations please find full name of those used in various presentation material shared with this UNC Workgroup, below:

- ALP: Annual Load Profile
- AUGÉ: Allocation of Unidentified Gas Expert
- CDSP: Central Data Services Provider
- CWV: Composite Weather Variable
- DAF: Daily Adjustment Factor
- DESC: Demand Estimation Sub Committee
- DM: Daily Metered
- DOW: Day of Week
- EUC: End User Category
- ILF: Indicative Load Factor
- LDZ: Local Distribution Zone
- MAPE: Mean Absolute Percentage Error
- MPE: Mean Percentage Error
- MPR: Meter Point Reference Number
- NDM: Non-Daily Metered
- PLF: Peak Load Factor
- R²: R Squared Statistical Measure
- SNCWV: Seasonal Normal Composite Weather Variable
- SND: Seasonal Normal Demand
- UIG: Unidentified Gas
- UNC: Uniform Network Code
- WAR: Winter Annual Ratio
- WCF: Weather Correction Factor
- WSENS: Weather Sensitivity

Workgroup 0754R

Background

- UIG Task Force produced a number of recommendations to help reduce temporary UIG levels/volatility. This included findings associated with the modelling error within the NDM Algorithm
- DESC is responsible for the NDM Algorithm (UNC Section H) and has an obligation to review it every 3 years (UNC H 2.2.2)
- Prior to moving forward with the above a consultation was performed during Q4 of 2020 to assess the levels of support for making improvements to the NDM Algorithm
- A more detailed view of the background to this Workgroup and current state overview is provided in the March meeting papers [here](#)

Rationale for workgroup

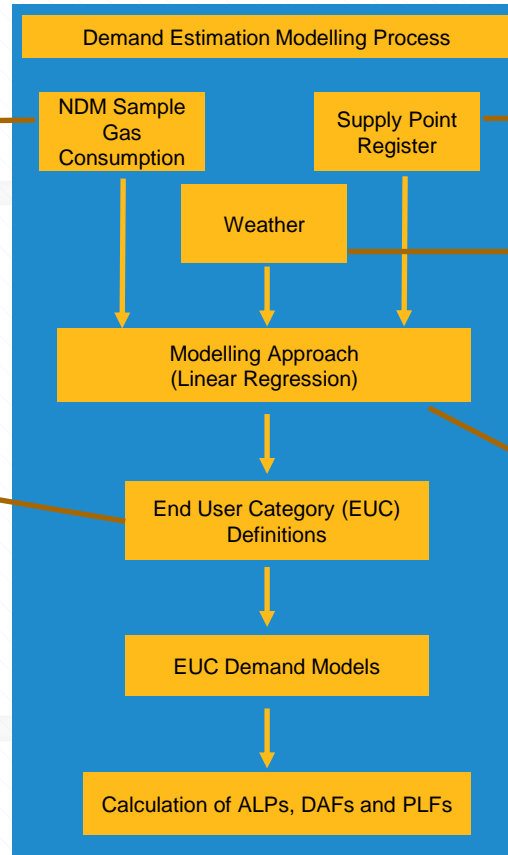
- Supports DESC's UNC obligation to review the NDM Algorithm
- UIG Task Force findings will be explored and progressed
- Clear industry support for investigating advanced analytical approaches
- A Workgroup maintains focus and increases visibility across the industry
- Improved NDM Allocation will result in a reduction in UIG volatility and subsequent Meter Point reconciliation/UIG volumes (temporary)

Workgroup 0754R: Investigation Areas

Investigate use of advanced analytics to...

1. Improve the validation processes of NDM sample gas consumption to identify erroneous supply meter points ([UIG TF 13.2.8](#))
2. Improve the 'infilling' of missing data

Investigate use of advanced analytics to review the appropriateness of the existing EUC definitions



Investigate use of additional data items on Supply Point Register for use in derivation of EUC definitions and production of Demand Models

Investigate use of advanced analytics to enhance further the weather vs demand relationship ([UIG TF 13.2.5](#))

Investigate use of advanced analytics to trial alternative approaches for producing more accurate EUC demand models ([UIG TF 13.2.6](#) & [13.2.7](#))



Meeting 6 Re-cap

(22nd March 2022)

Meeting 6 Key Discussion Points

The main headlines from meeting 6 of 754R were...

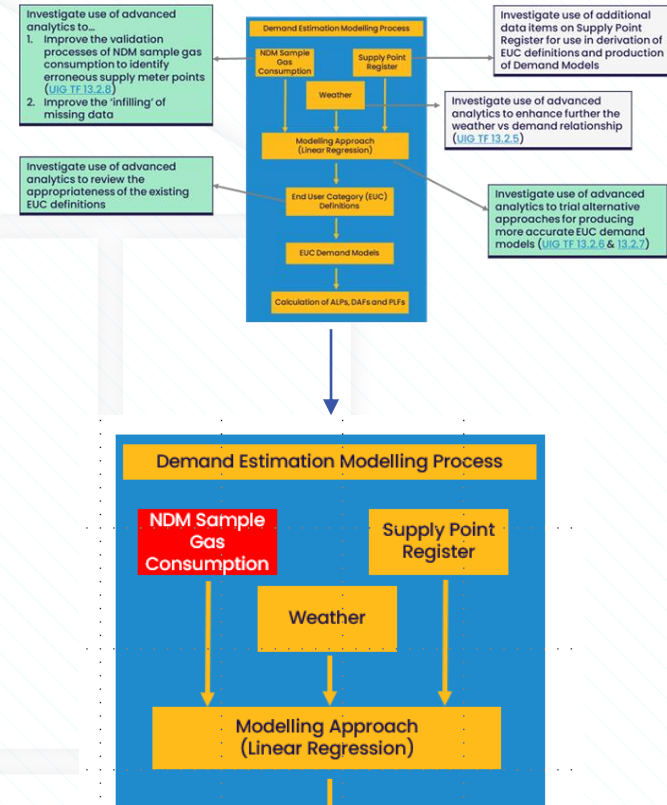
- Area 1: Provided an overview of the Indicative Load Factors calculation and the results from applying it to the new approaches
 - The method produced comparable results to current process
 - Results from the Gradient Boosting method requires further investigation – focus will be on the Peak Demand calculation.
- Area 1: Time was spent on understanding the models and their characteristics
 - Results provided for all the test EUCs for Monthly, Day Of the Week and Holiday Code trends
 - More investigation required into Day Of the Week trends for 02BNI and 05B datasets.
- Area 2: Focusing on using Advanced Analytics on validation datasets was introduced.
 - The objective and background was presented
 - The proposed areas / methods that will initially be investigated was introduced



Area 2: Improve Validation Processes

Modelling Process Flow

- In this process flow the key inputs are:
 - NDM Daily Gas consumption data submitted by multiple providers
 - Weather data
 - Supply Point Register (relies on shipper updates to UKLink)
- A key part of the demand modelling process is to ensure the input data is validated and cleansed
- Why do we validate?
Poor data leads to poor models and interpretation



Area 2: Objective

- Explore the use of Advanced Analytics techniques to develop and improve validation process prior to modelling
- Identify potential weakness, development opportunities and make recommendations which link to evidence of reduction in NDM modelling error.

| Potential Approach | Data | Systems |
|---|---|--|
| <p>Investigate latest data cleansing techniques/scripts</p> <p>'Uncertainty Estimation' UIG TF 13.2.8</p> <p>Compare current post validation results to revised methods - both 'infill' approach and ability to identify suspicious demand patterns</p> | <p>Daily Gas Consumption</p> <p>Daily Weather</p> <p>Supply Point Attributes (AQ, MSC, Correction Factor)</p> | <p>SAS Demand Estimation Modelling</p> |

Sample Data - Intended plan

- Utilise Advanced Analytics to enhance our existing validation routines (see table)
 - To help identify suspicious demand patterns in assessing sample MPRs
- Infilling:
 - Assess whether Advanced Analytics can assist with infilling of missing data
- We are going to investigate:
 - Techniques for identifying demand patterns and difference including:
 - Uncertainty estimation (as suggested by the UIG Task Force)
 - Others to be determined

Appendix 2 –Daily Gas Consumption Data Validation

The following provides the proposed validation criteria for use against the Daily Gas Consumption Data in the 2021 Gas Demand EUC Modelling. Section 1 of the NDM Algorithms Booklet will contain further details of the validation process and outcomes

Small NDM: 0 to 2,196 MWh p.a.

| Source | EUC Bands | Missing Days | | Consecutive Zeros | | Spike Ratios | |
|---|-----------|--------------|------------|-------------------|------------|--------------|--------|
| | | Summer | Winter | Summer | Winter | Summer | Winter |
| Xoserve Managed sample (and any third party data) | 01 and 02 | 15 or more | 15 or more | N/A | 33 or more | 15:01 | 08:01 |
| Network Managed sample (and any third party data) | 02, 03 | 28 or more | 28 or more | N/A | 20 or more | 13:01 | 05:01 |

Large NDM: >2,196 MWh p.a.

| Source | EUC Bands | Missing Days | | Consecutive Zeros | | Spike Ratios | |
|---|-------------------|--------------|------------|-------------------|------------|--------------|--------|
| | | Annual | Winter | Annual | Winter | Annual | Winter |
| Network Managed sample (and any third party data) | 05, 06, 07 and 08 | 40 or more | 20 or more | N/A | 20 or more | 08:01 | N/A |

Where:

Summer period is defined as 1st April 2020 to 30th September 2020.

Winter period is defined as 1st October 2020 to 31st March 2021.

Annual period is defined as 1st April 2020 to 31st March 2021

Current Validation Process

- Current validation process attempts to filter out all errors in data that has been submitted. This is while also ensuring we maintain sufficient sample numbers for modelling.
- The current validation rules were originally designed when there were smaller datasets and technology / computers were less capable than they are now
- We have moved from having two or three suppliers of data to over 20 which on the positive side has provided extra data to sample from but the challenge is with data quality and inconsistency between providers.
- The following slide lists the validation outcomes from the recent sample data collection

Spring 2022 Modelling Validation Outcomes

TABLE A.1 - VALIDATION SUMMARY OF DAILY GAS CONSUMPTION DATA

| Analysis Period | 01/03/21 to 31/03/22 |
|--|----------------------|
| START: MPRs with at least 330 'Daily Gas Consumption Data' records present within the analysis period | 58,508 |
| REMOVED: Exceeds Missing Read Tolerance | 15,019 |
| REMOVED: Missing Correction Factor | 244 |
| REMOVED: Exceeds Consecutive Zero Consumption Tolerance | 5,142 |
| REMOVED: AQ Ratio Failures | 2,543 |
| REMOVED: Winter vs Summer Consumption Ratio Tolerance (Band 09B sites) | 31 |
| REMOVED: Winter Annual Ratio (WAR) Outside of Tolerance | 239 |
| REMOVED: Twin Stream Site | 11 |
| REMOVED: Excluded MPRN (i.e. Hy Deploy or FlexGen site) | 15 |
| REMOVED: Excessive Number of Records Infilled | 871 |
| REMOVED: Multiple Data Providers | 182 |
| REMOVED: Exceeds Missing Read Tolerance Post Validation | 6,415 |
| REMOVED: Other (i.e. Scottish Independent LDZs; No LDZ on IS-U) | 335 |
| REMOVED: TOTAL | 25,098 |
| PASSED VALIDATION (Pre Stratification) | 33,410 |
| REMOVED: MPRs not required following Stratification | 8,417 |
| PASSED VALIDATION (Post Stratification) | 24,993 |

Proposed Advanced Analytics Validation

- The validation can be split as follows:
 - Firstly, identification and finding data errors (focus of the current arrangements)
 - Secondly, it can identify patterns of demand which raise questions as to the appropriateness of including these sites in the analysis for a particular EUC
 - e.g. Consumption patterns suggests an I&C usage however it has been categorised on UKLink with a Domestic Market Sector Code (i.e. how it is allocated)
- The techniques on the following slides cover both categories with the objective of improving the quality of the sample data being used in the training and reviewing of demand models
- We have used the latest sample data collected as the input to this analysis, in order to identify additional data errors which may not have been identified by the current rules.

Proposed Advanced Analytics Validation: Focus Areas

- The following approaches have been identified:
 - **Uncertainty Estimator:** UIG Task Force validation suggestions ([UIG TF 1328](#))
 - **Individual MPR Regression:** Use of computer processing capability to target individual MPRs patterns – regression test Monday to Thursday for each MPR for high level pattern
 - **Winter Zero Consumption:** Analysis of the number of zero consumption in the winter period, December to March) (i.e. targeting questionable patterns)
- Each of these have been taken and an initial approach investigated and developed
- Note: Any approaches that are successful can be implemented into the existing validation process without requiring any modifications.



Area 2: Uncertainty Estimator

UIG Task Force techniques

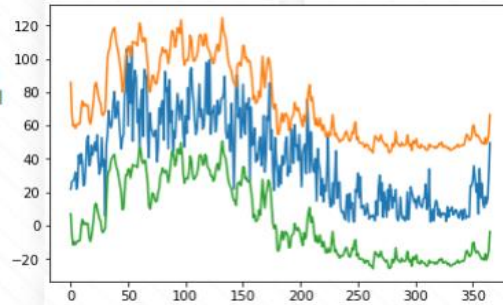
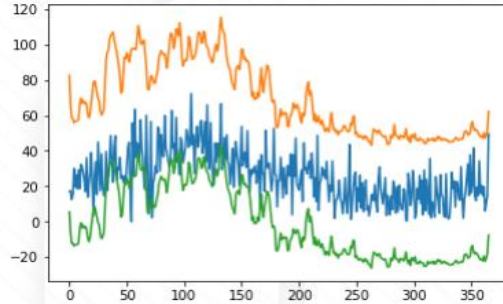
- The UIG Task Force suggested several techniques which could be used in assessing sample meter points.
- One of their methods looks to assess how an individual meter point demand pattern compares with a pattern of the wider group i.e. EUC.
- The pattern for each EUC group is developed by:
 - Looking at the observations daily
 - Acknowledging the range of demands (from the highest to zero demand at the lowest level)
 - From a statistical perspective focusing on the core 95% of demand values thereby creating
 - An upper boundary at the 97.5% level
 - A lower boundary at the 2.5% level
 - Using these boundaries to evaluate an MPR's demand pattern.
- An excessive number of data points that are outside these boundaries could suggest the MPR may not be representative of those in the group
- An MPR will be further investigated where [25%] of data points fall outside boundaries.

Example Output from the UIG Task Force

The orange series represents the upper boundary where the top 2.5% of consuming sites fall outside that specific boundary on a given day

The blue series represents the actual consumption

The green series represents the lower boundary where the bottom 2.5% of consuming sites fall outside that specific boundary on a given day

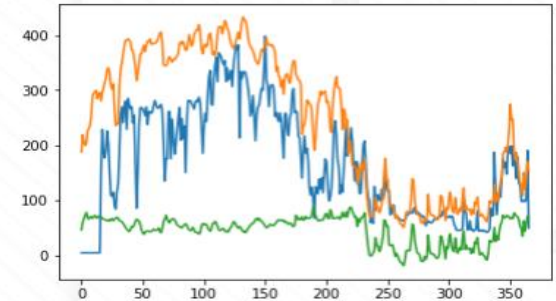
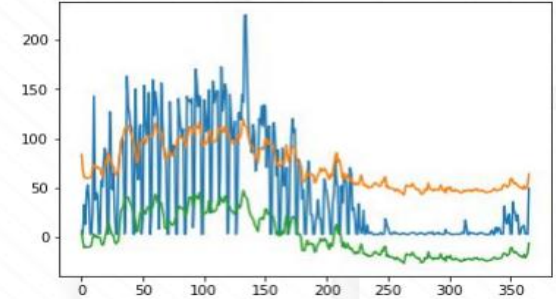


MPRs within the boundary

The orange series represents the upper boundary where the top 2.5% of consuming sites fall outside that specific boundary on a given day

The blue series represents the actual consumption

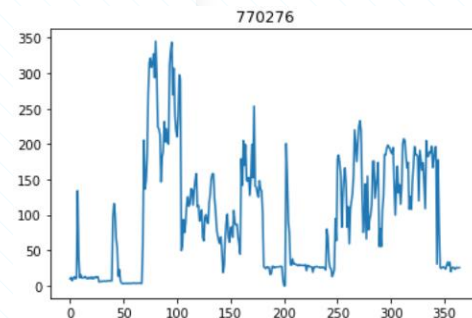
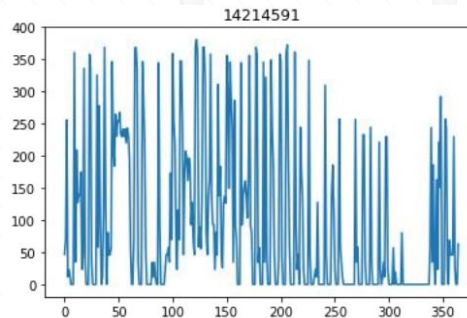
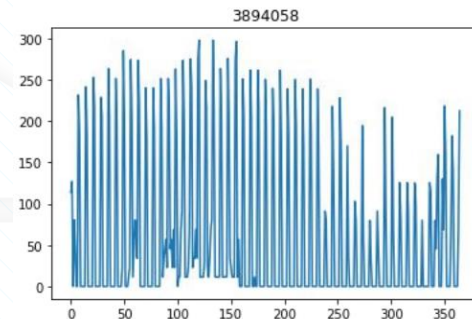
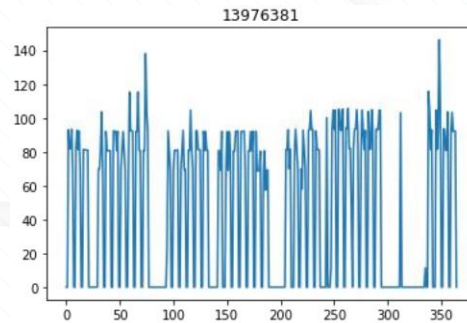
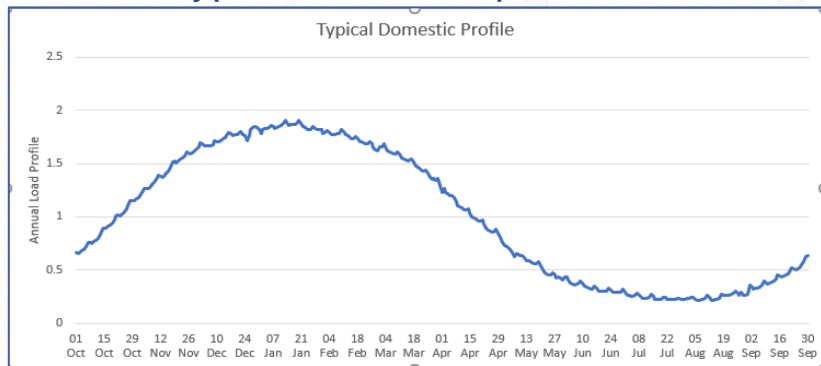
The green series represents the lower boundary where the bottom 2.5% of consuming sites fall outside that specific boundary on a given day



MPRs outside the boundary
(require investigation)

Sample data - Task Force Findings

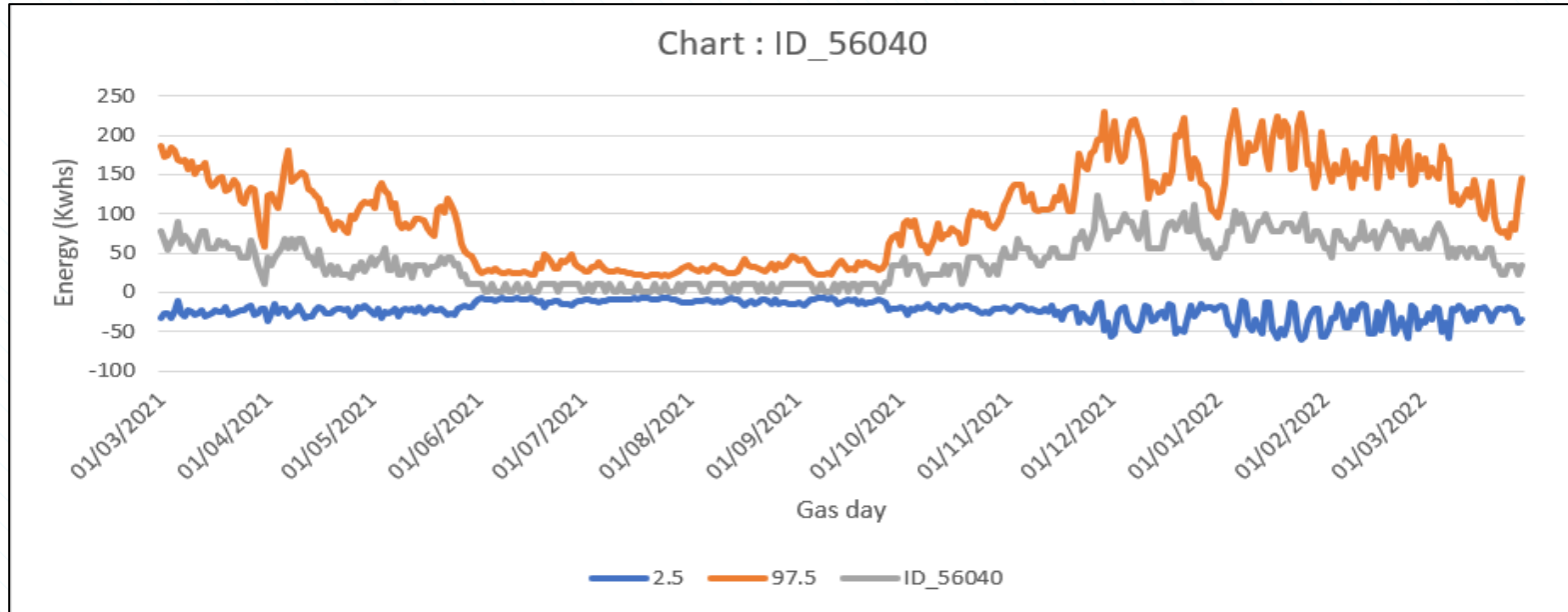
- The UIG Task Force produced an assessment of Sample MPRs
- These charts are a selection of demand patterns that:
 - Passed validation
 - Were believed to be domestic
- Visually they do not seem to match a typical domestic profile.



Note the reference numbers above are anonymized IDs and not MPRs

2022 Sample Test Results (typical profile)

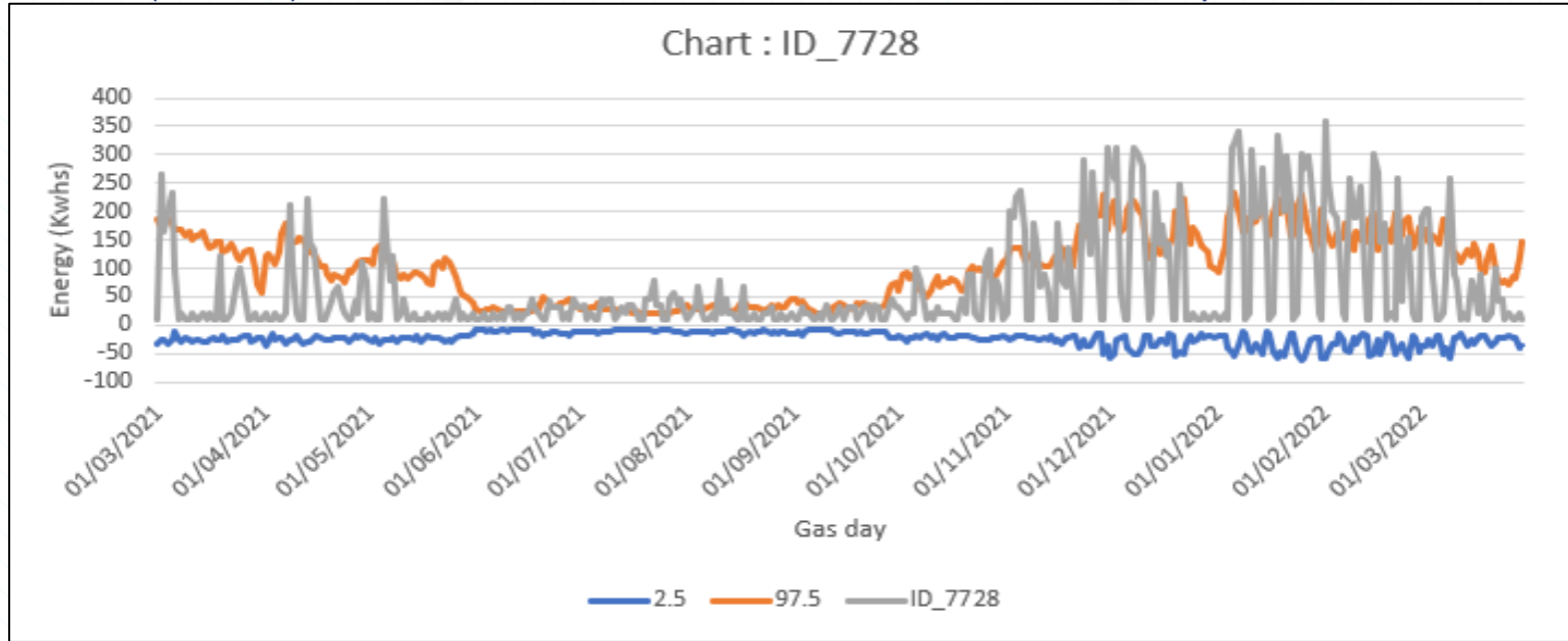
- The chart shows the result for MPR (id 56040) in EUC 01BND in LDZ WM (AQ 14000 KWhs).



- The blue and orange line sets the 2.5 and 97.5 boundaries.
- The demand is comfortably between the two boundaries and as such does not give any cause for concern.

2022 Sample Test Results (Unusual Profile)

- The MPR (id 7728) is also in EUC 01BND, in LDZ WM and has a sample AQ of 24,500 KWhs



- This chart shows a demand pattern that could be a cause for concern. It has a significant number of data points outside the boundaries and would need further investigation.
- That investigation showed that the demand dropped away at weekends, which raises a question if this is genuinely a domestic site

Uncertainty Estimator – Conclusion

- This looks like a useful technique we can employ in our validation to identify unusual demand patterns and or issues with data held on the supply point register.
- The process still needs some refinement to produce similar output to UIG Task Force e.g. the distribution is not correct yet and needs some additional work
- Other EUCs would be tested to see if this technique can be used for them.



Area 2: Individual MPR Regression

Individual MPR Regression

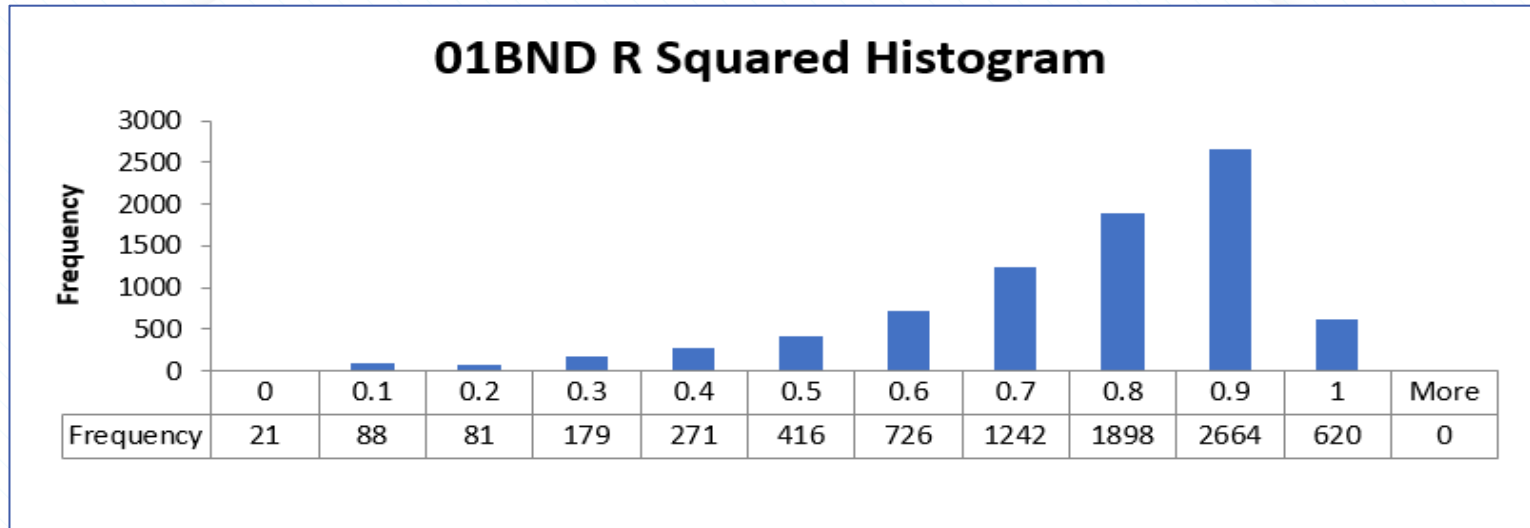
- Advanced Analytics applications and processing power capability has allowed us to investigate the underlying patterns for individual MPRs in large volumes.
- Taking the core Monday to Thursday data points, these can be used to look at the underlying demand vs weather pattern for each MPR
- EUC 01BND has been assessed with the principle assumption that the data points should show an inverse relationship between weather and demand.
 - Is this true in the case of cooking only loads etc?
- For all EUCs, as knowledge and analysis builds up, traits for specific EUCs could become evident
 - This could also help us in identifying potential new EUC groupings (based on these traits).

Individual MPR Regression Assessment

- The approach / hypothesis is that for 01BND models, they will generally have a strong relationship to weather
- The chosen data items to check and compare were:
 - R^2 - a statistical measure for identifying 'Goodness of fit'. Its value range from 0 to 1 (sometimes expressed as 0 to 100% with 100% indicating a perfect fit / direct relationship)
 - CWV intercept – this is a measure used in modelling and it represents the CWV at which demand is zero. $\text{CWV intercept} = (- \text{model intercept} / \text{model gradient})$
 - EUC 01BND typically has an intercept in the range 14-18 degrees.
 - The higher the CWV intercept the less weather sensitive the MPR is.
- When reviewing the output from the regressions we are looking for the following:
 - A low R^2 would suggest a poor demand to weather relationship
 - The CWV intercept should be positive (>0)

Individual MPR Regression: 01BND Results – R²

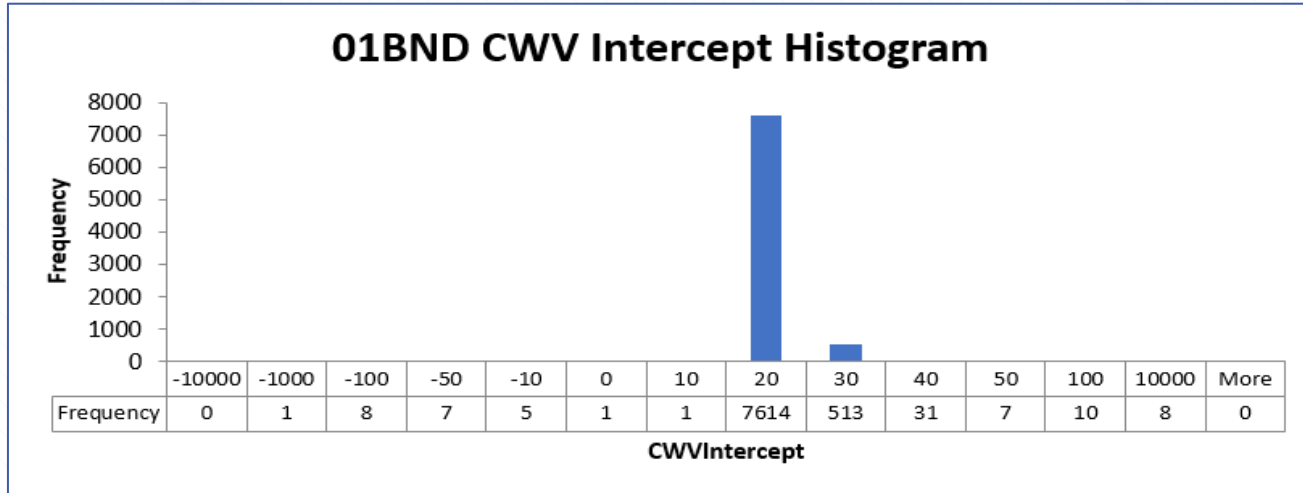
- A histogram of the R² values is shown below. There were a reassuring number of sites which showed a strong demand to weather relationship.



- There were over 369 sites with an R² of less than 0.3.
- Instances with an R² of less than 0.1 should be investigated for errors and to consider if they are genuine and/or representative.

Individual MPR Regression: 01BND Results - CWV intercept

- The second check looked at the CWV intercepts with a histogram of the results shown below.

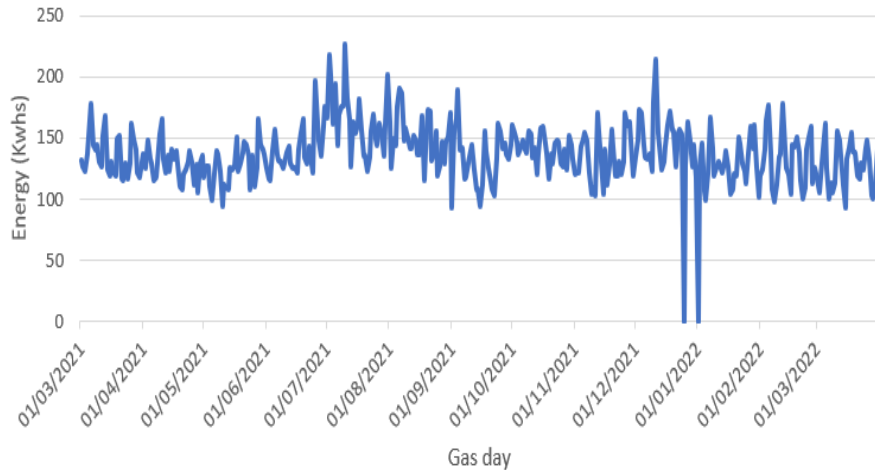


- There were a reassuring number of sites which showed a positive CWV intercept with the bin of 20 (7614 sites) representing 93.7% of the sites passing validation.
- Any value less than 0 suggests a site that is not weather sensitive and potentially increases demand as CWV increases

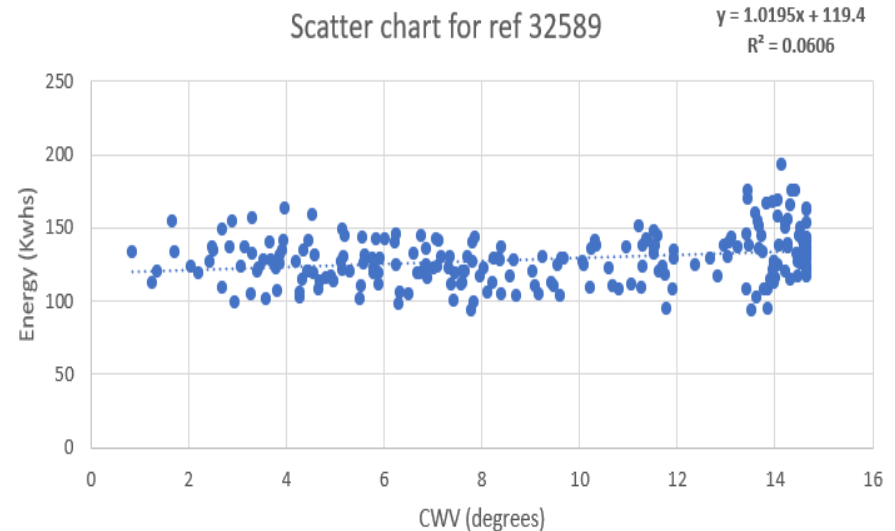
Individual MPR investigation (1)

- Several of the sites with the suspicious (and negative CWV intercepts) were investigated and found to also have very low R^2 (suggesting the two methods complement each other)
- Example 1: R^2 of 0.0606 and CWV Intercept of -114
This site is currently flagged on UK Link as 'Domestic' and is attracting a domestic profile. The demand pattern and scatter would suggest it is not representative of a domestic MPR.

Timeseries Chart for ref 32589



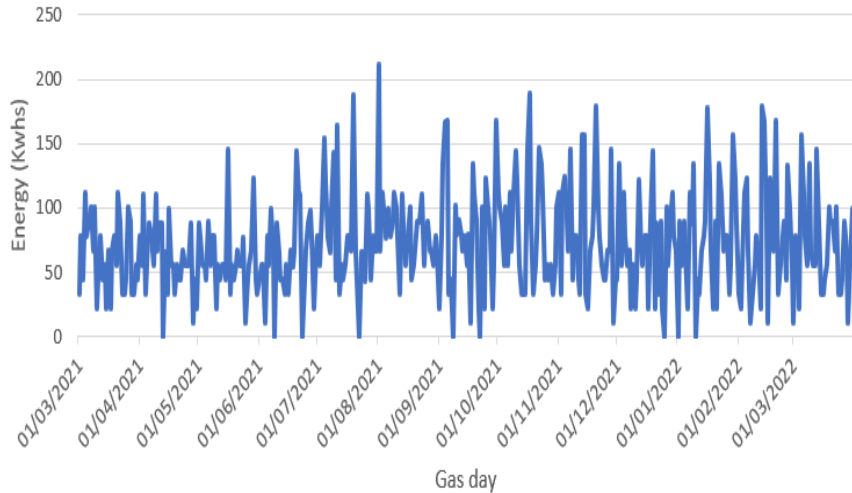
Scatter chart for ref 32589



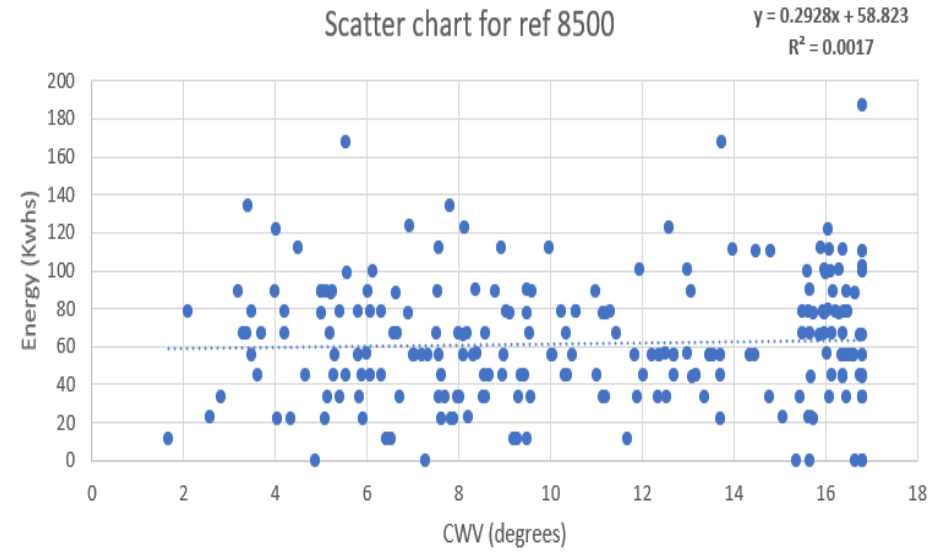
Individual MPR investigation (2)

- Example 2: R^2 of 0.0017 and CWV Intercept of -5902
This site is currently flagged on UK Link as 'Domestic' and is attracting a domestic profile.
The demand pattern and scatter would suggest it is not representative of a domestic MPR.

Timeseries Chart for ref 8500



Scatter chart for ref 8500



Individual MPR Regression - Conclusion

- This test highlighted instances of MPRs that although these had passed initial validation required some further investigation.
- We were able to conduct these checks soon after this years' sample data had been validated
 - Of the sites that had a negative CWV intercept, 17 that we investigated remained in the modelling datasets.
 - These were removed prior to running the modelling.
- This test could also be used to highlight where a different EUC sub band profile exists, or incorrect records held on supply point register
- The test may be improved to focus on other areas e.g., weekend effects



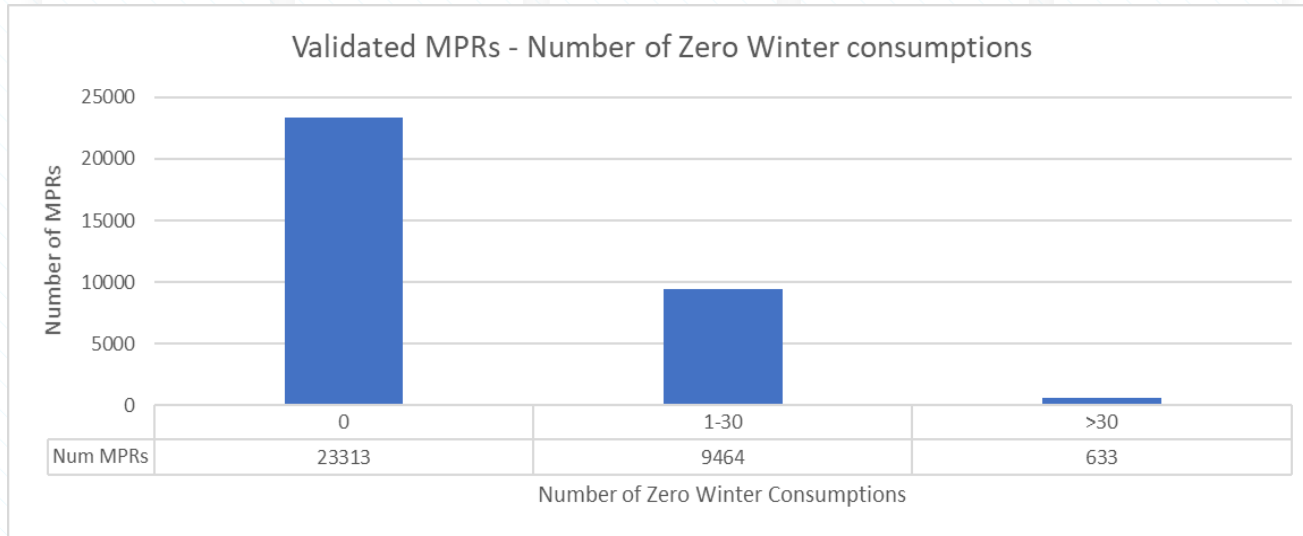
Area 2: Winter Zero Consumption

Winter Zero Consumption (1)

- While investigating techniques supporting validation, several of the MPRs identified in the previous two tests had a high amount of zero consumptions.
- While not strictly 'Advanced Analytics', looking at level of zero consumptions in winter could provide additional insight into unrepresentative demand patterns.
- Existing validation looks for consecutive zeros only. We investigated the hypothesis that there are a significant number of MPRs with zero consumption days over winter which are passing validation, but the demand pattern may not be desirable.
- This test is looking to assess the number of zeros, where winter has been treated as the months of December to March at the end of the sample period
- This test would highlight MPRs for further investigation

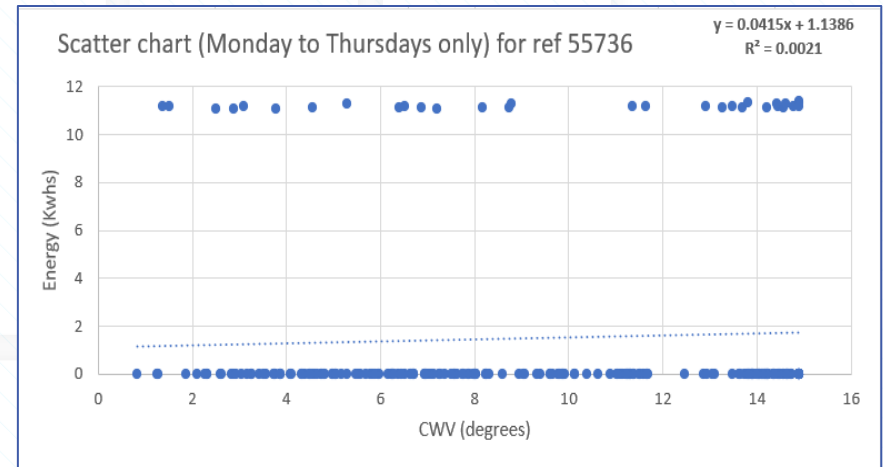
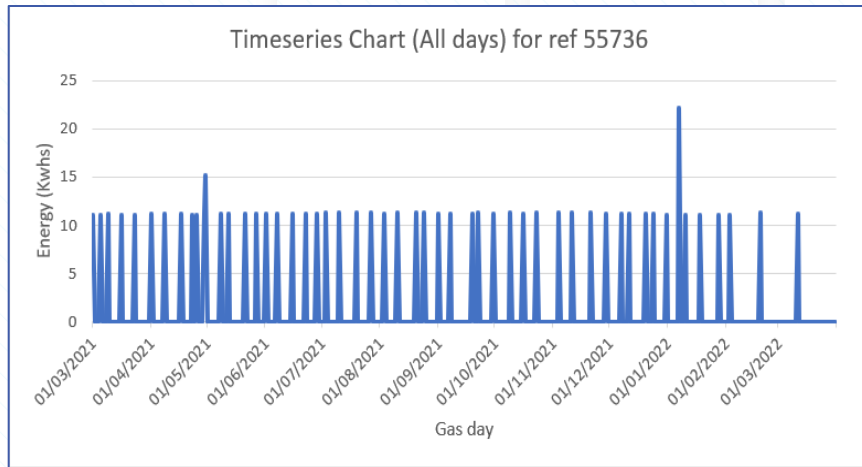
Winter Zero Consumption - Results

- This test was applied to all EUCs and MPRs that passed validation in the latest sample dataset.
- This histogram shows the MPRs with 'zero', '1-30' and 'over 30' zero consumptions.
- There were 633 MPRs which had over 30 days of zero consumption in the winter period.
- The largest was an 01BND MPR which had 83 zeros in the winter period (example 1)



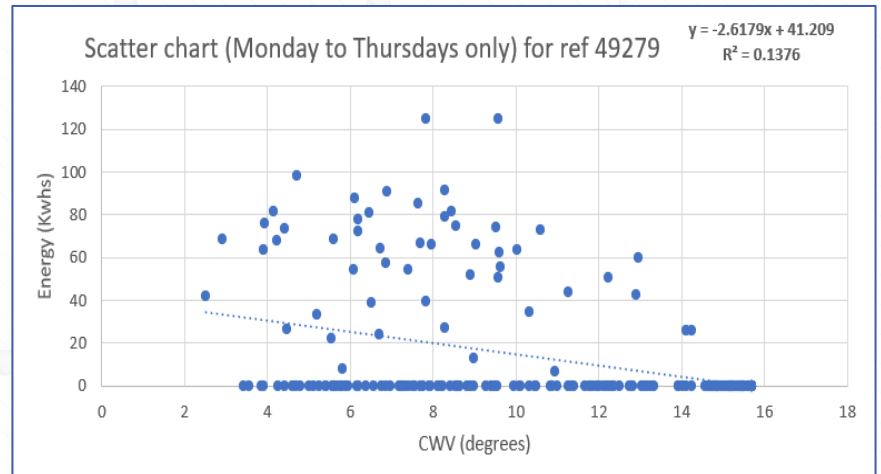
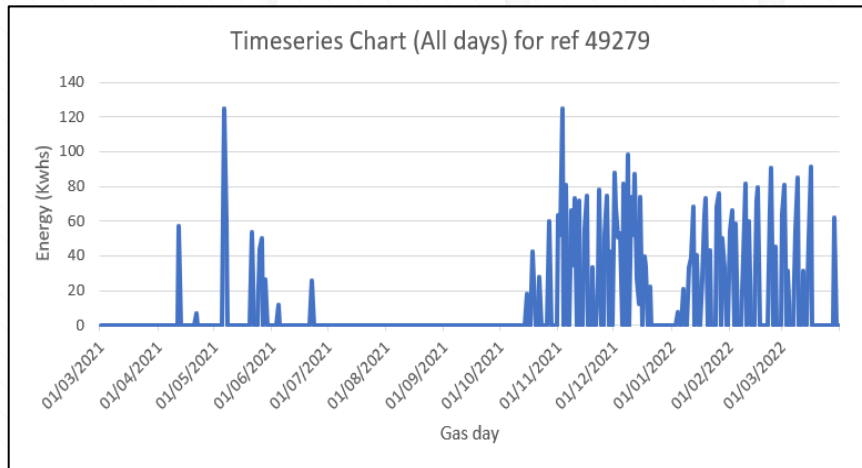
Winter Zero Consumption – Example 1

- This MPR 55736 is in EUC 01BND with a sample AQ of c550 KWhs and had 83 zeros.
- Individual MPR regressions had an R^2 of 0.0021 and CWV Intercept of -76.63
- Further investigation required into this demand pattern.
 - Registering demand only one day a week
 - Is it maybe a cooking load (except you would expect demand on more than one day a week) ?



Winter Zero Consumption – Example 2

- This MPR 49279 is in EUC 01BND with a sample AQ of c3924 KWhs and 56 zeros.
- Individual MPR regressions had an R^2 of 0.176 and CWV Intercept of 15.43
- This passed all validation with a low R^2 in the Individual MPR regressions suggesting it may not be a suitable demand pattern.



Area 2: Summary and Next Steps

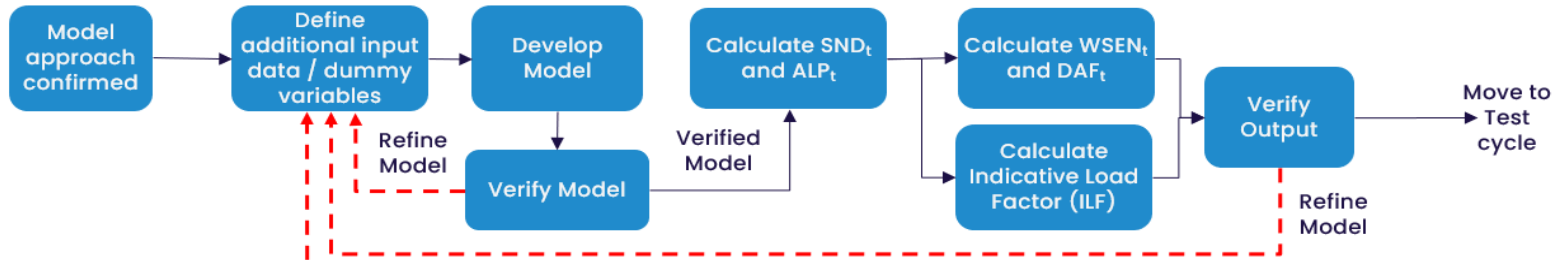
- The analysis shared here should help gain insight on which MPRs seem unrepresentative of their EUC and so improve the EUC demand model performance. In addition the checks could also potentially reveal different 'EUC groupings' outside of the current definitions used by the industry.
- Each Summer, DESC confirm its 'Adhoc Workplan' which is effectively a list of items it would like to investigate further ahead of the next year's modelling cycle. One of the proposed items for the forthcoming year is to look at the current validation rules and techniques
- Pending the results of the 'Adhoc Workplan' we would expect to use the techniques discussed today in future validation runs and also in any analysis investigating the appropriateness of EUC groupings
- We welcome any insight from the Workgroup on validation / cleansing routines used within your own organisations that you feel would benefit / be suitable to the Demand Estimation modelling process



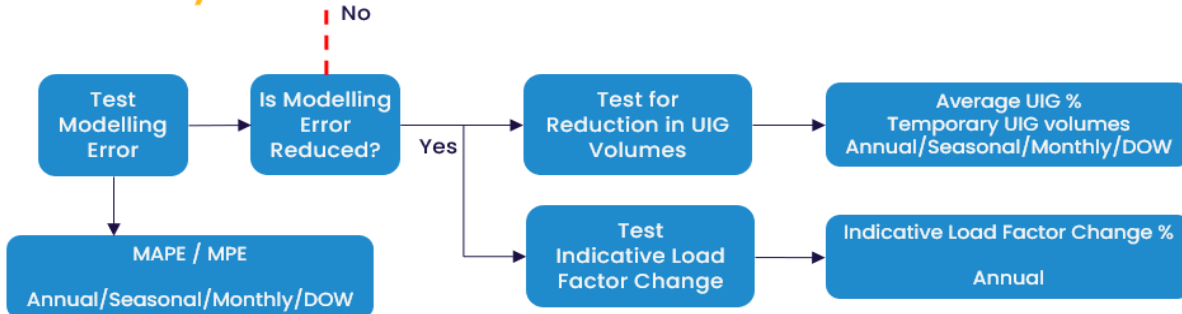
Update on Area 1: Trial Alternative approaches to deriving SND_t

Workgroup 0754R: Area 1 Development Approach

Development Cycle



Test Cycle



Evidence

For each model collate:

- Approach details,
- Modelling Error,
- UIG outcomes,
- ILF changes
- Process observations

Area 1 Update

- Investigation into ILF calculation (o/s action from previous meeting)
 - From the Gradient Boosting approach, the lack of ‘cold data points’ to train the model, means it is struggling to predict extreme demands.
 - The current linear model approach gives it an advantage in predicting these extreme demands
 - This is a clear issue that would need to be resolved before moving away from the current methodology
- Key takeaways from Area 1 analysis to date:
 - Reminder: Industry consultation carried out prior to Workgroup being established confirmed that access to demand models and parameters (i.e. ALPs and DAFs) ahead of the Gas Year was still a key requirement
 - DE team’s understanding of alternative Advanced Analytic options available for Demand Modelling has been enhanced e.g. the shortlisted approaches of ‘Neural Network’ and ‘Gradient Boosting’ have been used for analysis presented to this Workgroup
 - Results from these alternative approaches have revealed areas where the current modelling approach could be improved e.g. treatment of days of the week
 - One of our key tests (as described on previous slide) is ‘Has the modelling error reduced’ ? and although in some categories we have observed improvements, overall there has been no ‘breakthrough’ when compared with the current approach

Conclusion and Next Steps

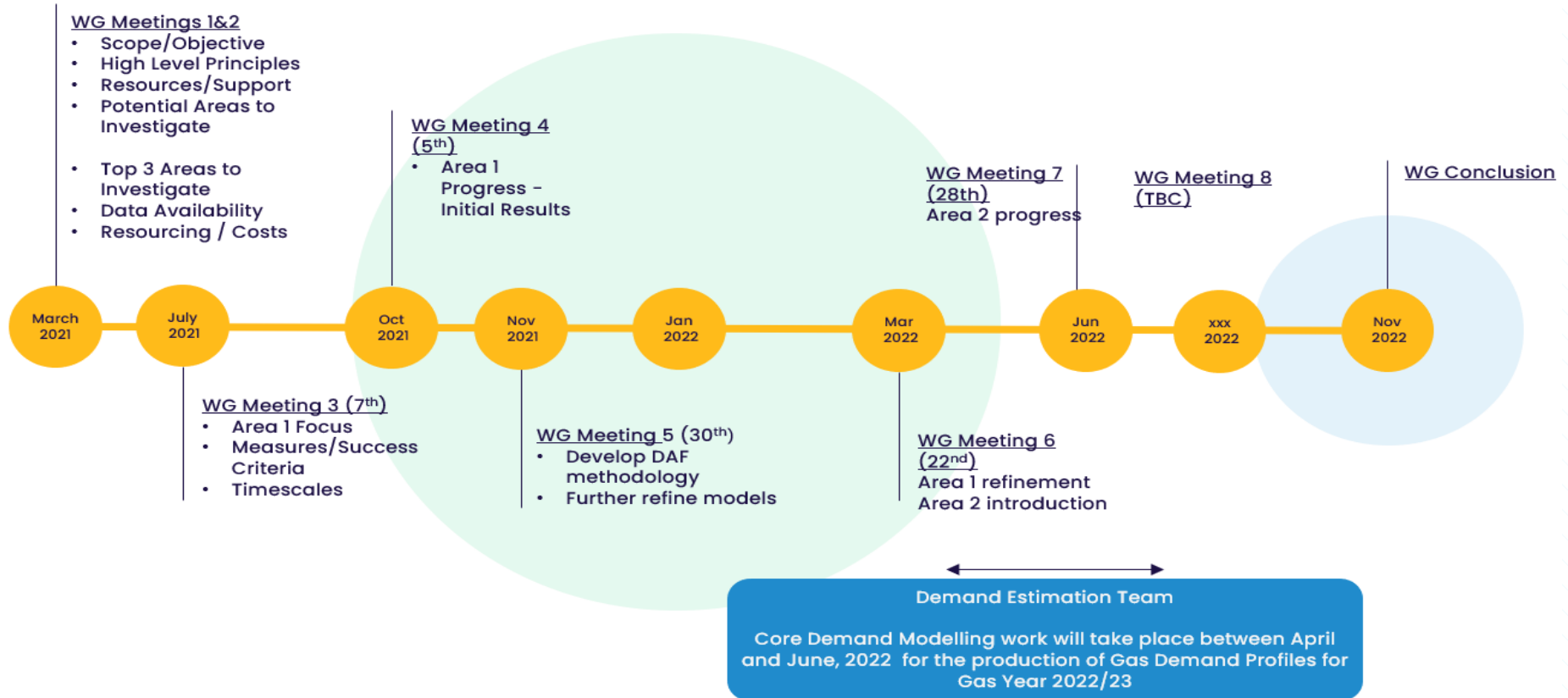
Conclusion:

- DESC's 'Adhoc Workplan' will naturally pick up 'Areas 2 and 3' within its annual cycle
- This year we shall be proposing a review of current validation rules and how we can use techniques shared as part of this Workgroup to improve EUC demand modelling
- In addition we shall also be proposing a review of EUC definitions which again can be achieved using some of the Advanced Analytic techniques learned as part of this Workgroup
- Area 1 analysis, although providing useful insight, has not revealed a clear alternative approach to the existing demand modelling process, given the requirement to maintain ALPs, DAFs and PLFs

Next Steps:

- Is the workgroup expecting any further analysis?
- Workgroup end date was extended to Nov 2022
- Workgroup Report needs to be prepared over the summer?

Workgroup 0754R: Timeline





Thank you