experts | evolving | energy

AGGREGATED DATASETS – THIRD PARTY & SUSTAIN-H DATASETS ANALYSIS

A REPORT FOR WESTERN POWER DISTRIBUTION

CLIENT	WESTERN POWER DISTRIBUTION
DOCUMENT NO.	WESTERN002-AD-R-03
REVISION	E
ISSUE DATE	01 November 2021
STATUS	Final
PREPARED BY	Freya Espir
CHECKED BY	Bob Hodgetts
APPROVED BY	Felicity Jones, Nithin Rajavelu

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DISCLAIMER

This document has been prepared and is issued by Everoze Partners Limited to the named Client in accordance with the contract dated 9/12/2019 and subsequent Variation dated 21/05/2020. This governs how and by whom this report should be read and used. The associated NIA code is WPD_NIA_047.

ABBREVIATIONS

ASHP	Air source heat pump	DDF	Data Derating Factor
AUF	Accuracy uncertainty factor	EV	Electric Vehicle
BaU	Business as Usual	NIA	Network Innovation Allowance
CUF	Completeness uncertainty factor	RUF	Resolution uncertainty factor
DQF	Data Quality Factor	UF	Uncertainty factor

VERSION HISTORY

A	Phase 1 Initial draft	20 October 2020
B	Phase 1 Updated draft	23 October 2020
C	Phase 1 Final version	29 October 2020
D	Phase 2 Initial draft	18 October 2021
E	Phase 2 Final version	01 November 2021

1. EXECUTIVE SUMMARY

FutureFlex is a joint project delivered by Western Power Distribution (WPD), Everoze and SGC. It is funded by the Network Innovation Allowance (NIA). FutureFlex is trialling a new DSO service suitable for domestic flexibility, called Sustain-H. This service places data optionality at its heart.

The role of the *Aggregated Datasets* workstream is to quantify the impact of lower portfolio data quality for WPD's DSO services. This is needed because domestic portfolios have different data characteristics to large assets. The results of this workstream have informed payment mechanism design for Sustain-H, reflected in the Sustain-H Product Roadmap, with particular relevance for longer-term development. The Product Roadmap is part of WPD's commitment to translating the learnings from the Sustain-H Trial to Business as Usual (BaU).

The methodology of the *Aggregated Datasets* workstream focuses on data resolution, data completeness and data accuracy. The methodology is summarised in Figure 1 below: the approach is to independently analyse Resolution, Completeness and Accuracy Uncertainty Factors. These three Uncertainty Factors are then combined into a single Data Quality Factor for each technology type. This then informs any Data Derating Factor that may be introduced in the longer-term for Sustain-H remuneration, after the service is established.

The Aggregated Datasets analysis is conducted in two phases – of which this is the second. Phase 1 conducted analysis on third party datasets from two separate trials: Electric Nation (electric vehicles) and Freedom (heat pumps). Phase 2 refined the analysis using data from the Sustain-H service trial. Figure 1 captures the Methodology and how the phase 2 report adds to work completed within phase 1. Everoze prioritised the phase 2 analysis based on:

- <u>Data availability:</u> Sustain-H data was dominated by electric vehicles; and
- <u>Materiality:</u> Factors demonstrated to be of material significance to data quality.

The phase 2 analysis did not provide sufficient robust evidence to change the conclusions from the phase 1 analysis. Everoze conducted the following steps for the phase 2 analysis:

- <u>Review data resolution:</u> Everoze updated the Resolution Uncertainty Factor (RUF) using electric vehicle (EV) data from the Sustain-H trial. Only one dataset from the Sustain-H trial, a portfolio of 19 EV chargepoints, was suitable for this analysis. This dataset is significantly smaller than that used for phase 1 and so this limited the scope of the analysis that could be performed. The higher uncertainties inherent in small portfolios meant the associated findings from phase 2 do not provide conclusive evidence to challenge the conclusions drawn from the phase 1 analysis.
- <u>Review other factors:</u> The new Sustain-H datasets do not provide any new data to change the conclusions from phase 1 on the Completeness Uncertainty Factor (CUF) and the Accuracy Uncertainty Factor (AUF), and so further analysis was not performed on this during phase 2.
- <u>Update Data Quality Factor:</u> Everoze updated the Data Quality Factor using the updated RUF from the Sustain-H datasets, however as noted above, these findings should be treated with caution due to the high uncertainties inherent in small portfolios.
- <u>Assess implications:</u> Everoze teased out the implications for future Sustain-H service procurement and suggested areas for future investigation.

The analysis revealed a series of interesting results, summarised below:

1. **The limitations of available datasets pose a *substantial* challenge for reaching conclusions – emphasising the pressing need for WPD to gather more data in future**. Everoze's analysis was materially affected by the limitations of the Freedom, Electric Nation and Sustain-H trial datasets. Most notably, key issues included the short duration of the datasets used, the Electric Nation dataset not being a minutely dataset, and the lack of large portfolio datasets of minute granularity from the Sustain-H trial. As such, Everoze strongly advocates seizing future opportunities to secure further domestic flexibility data, including in innovation projects, to build on these learnings and appropriately factor in the meter data uncertainties for Sustain-H remuneration in the long term.

2. **Data resolution has the biggest impact on demand uncertainty, followed by Completeness, followed by Accuracy.** The standard deviation component of the Data Quality Factor is dominated by the Resolution Uncertainty Factor standard deviation (0.3), which is a magnitude of 10 greater than those calculated from the completeness analysis. In short, lower resolution half-hourly data *substantially* reduces the confidence WPD can

have in the ultimate peak demand compared to minutely resolution data. Meanwhile, completeness is of medium importance.

3. **Results vary significantly by dataset – hinting at a possible need for a technology-specific approach to analysing data resolution:** Consumption profiles vary dramatically between the Freedom data (for heat pumps), and the Electric Nation and Sustain-H datasets (for electric vehicles); this applies both across the day and within individual half hour settlement periods. This has implications for the Resolution Uncertainty Factor (RUF), suggesting that it may be more appropriate to derive a separate resolution factor per technology. It is further possible that other assets (such as batteries) will show different behaviours again. At this stage, it is difficult to make firm conclusions on the technology-specific attributes when the data analysed are from trials with different interventions which have different impact on demand patterns. Moreover, as noted below in point 6, any technology-specific attributes may diminish for large portfolio sizes. In any case, more data are required to repeat the resolution analysis, should appropriate data become available, to draw firm conclusions on the technology-specific impact on data resolution.

4. **For heat pumps at half-hourly resolution, the Data Quality Factor for a complete dataset was 1.43 for a 50% confidence level and 1.92 for a confidence level of 95%.** This means that we are 50% certain the peak demand recorded for a half-hourly metered heat pump portfolio is 1.43 times higher when assuming minutely resolution.

5. **For electric vehicles at half-hourly resolution, the Data Quality Factor (DQF) for a complete dataset (using Electric Nation data) was 1.18 for a confidence level of 50% and 1.53 for a confidence level of 95%.** This means we are 50% certain the peak demand recorded for a half-hourly metered electric vehicle is 1.18 times higher when assuming minutely resolution. For the Sustain-H dataset, the DQF is 1.91 and 4.36 for the 50% and 95% confidence levels, respectively, which is significantly higher than the findings using the Electric Nation dataset. The material difference in the findings using the two datasets is likely due to: (1) the constant power charging assumption made when processing the Electric Nation data, (2) the flexibility provider's interventions in the Sustain-H dataset where the EV chargers were responding to price signals on a minute-by-minute basis, and (3) low demand outside of 12am-7am in the Sustain-H dataset resulting in the peak-over-mean calculated to be statistically skewed due to the low half-hourly average demand during these periods.

6. **Analysis to date suggests that portfolio size strongly impacts the half-hourly Resolution UF, and consequently the Data Quality Factor.** Everoze repeated the resolution analysis on reduced population samples for both the Electric Nation and Freedom datasets. Due to the small portfolio size, the Sustain-H dataset was not included in this analysis. The smaller population datasets yielded larger RUFs. Interestingly, for a fixed portfolio size, the analysis yielded different results for the two datasets, which implies that technology type may also be a driver of variation in the Data Quality Factor for small portfolio sizes. The impact of portfolio size and technology type on the calculated Data Quality Factor (predominantly driven by the RUF), appears to diminish for large portfolios (> 100 assets). This convergence, or asymptotic behaviour, with increasingly large portfolios will have an impact in the design of the DQF/DDF where a simplified approach may be justified.

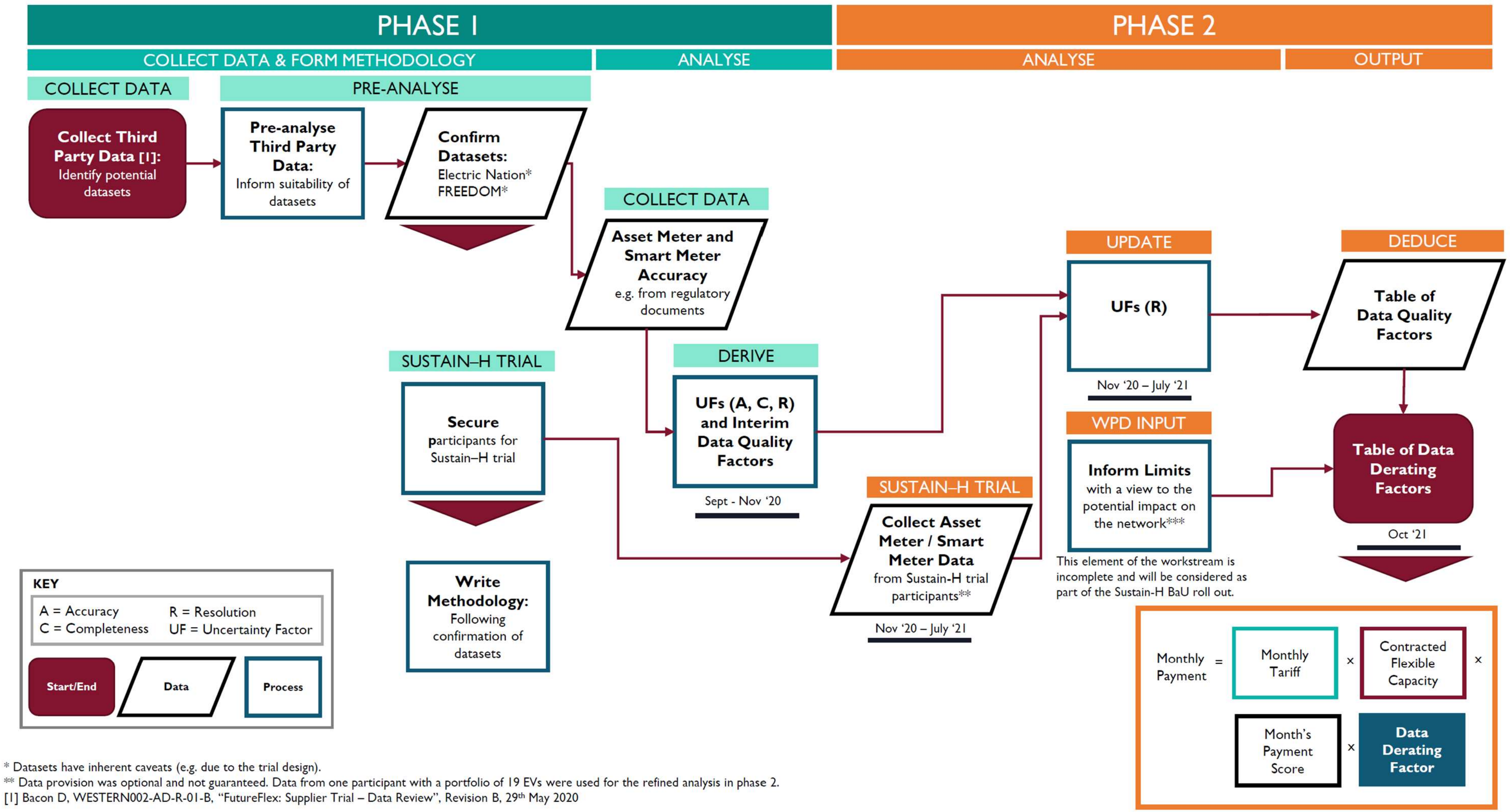
7. **The linear impact of data incompleteness means that WPD can take a pragmatic approach.** In Everoze's analysis, demand uncertainty increased in a broadly linear fashion as the dataset became more incomplete. For instance, an 80% complete dataset led to a 125% increase in demand uncertainty when considering a 50% confidence level. As a result, there is potential for WPD to adopt a pragmatic approach here; for instance, for a 50% confidence level an 80% complete dataset might result in a multiplier of 1.25 within the Data Quality Factor, with corresponding impact on payment. Everoze also investigated the impact portfolio size has on demand uncertainty for an 80% incomplete dataset, and found the impact is independent of portfolio size.

Based on the results above, Everoze has the following recommendations for commercial roll-out of Sustain-H:

- <u>Initially, do not introduce derating factor:</u> Everoze recommends that no derating factor for data resolution is used for initial BaU roll-out of the service. This is because the immediate priority is securing participation from domestic flex providers to mature the Sustain-H service. Given that low value has already been highlighted as a key risk for transitioning Sustain-H from trial to BaU procurement, we recommend that no derating is applied for BaU roll-out at least in the medium term until sufficient liquidity is established.
- <u>Analyse data resolution further, to improve understanding:</u> Everoze recommends that WPD conducts further analysis to refine the Resolution Uncertainty Factor analysis using high-resolution meter data across various

technologies, if/as new data becomes available in the future. This is needed to test the findings in this report, and to provide a robust evidence base for introducing any derating factors in the Sustain-H payment calculations in the future.

- Potentially introduce derating factor in future to incentivise higher resolution data: Everoze recommends considering introducing a derating factor in the future (based on the Data Quality Factor) to incentivize higher resolution meter data once sufficient liquidity for the Sustain-H service is established. As Sustain-H participation in the long term will likely be large-scale domestic portfolios, and as impact of small portfolios of homes and assets on WPD's network is expected to be minimal, a simplified approach focusing on the impact of the larger portfolios (> 100 homes) is justified.
- Consider introducing an availability mechanism in the longer-run: Everoze recommends that WPD considers introducing an availability mechanism in the performance assessment and payment calculations in the medium to long term to incentivize higher completeness in the meter data submitted.
- Remain flexible on metering solutions: Everoze recommends that WPD be flexible with metering solutions used for Sustain-H and accept MID-compliant metering solutions. If WPD accepts the prescribed requirements for domestic flexibility as set out in PAS 1878, we recommend WPD evaluates the Accuracy Uncertainty Factor for these metering solutions and its overall impact on the Data Quality Factor.



* Datasets have inherent caveats (e.g. due to the trial design).
** Data provision was optional and not guaranteed. Data from one participant with a portfolio of 19 EVs were used for the refined analysis in phase 2.
[1] Bacon D, WESTERN002-AD-R-01-B, "FutureFlex: Supplier Trial – Data Review", Revision B, 29th May 2020

FIGURE 1: AGGREGATED DATASETS WORKSTREAM PROCESS

2. INTRODUCTION

2.1 CONTEXT

FutureFlex is a joint project delivered by Western Power Distribution (WPD), Everoze and Smart Grid Consultancy (SGC), funded by the Network Innovation Allowance (NIA). FutureFlex is trialling a new DSO service suitable for domestic flexibility, called Sustain-H. This service places data optionality at its heart.

The role of the *Aggregated Datasets* workstream is to quantify the impact of lower portfolio data quality for WPD's DSO services, which will ultimately help to inform payment mechanisms for Sustain-H.

The first deliverable in Aggregated Datasets, "FutureFlex Supplier Trial – Data Review", examined available datasets to inform Everoze's analyses [1]. The second deliverable, "Aggregated Datasets – Methodology" [2] (the "Methodology"), outlined an approach to quantify the impact lower dataset resolution, completeness and metering accuracy has on demand prediction. The third deliverable followed on from the Methodology, presenting the results from the analysis conducted on third party datasets: data from the Electric Nation [3] and Freedom [4] trials. For the analysis, the Resolution, Completeness and Accuracy Uncertainty Factors (UFs) were analysed independently and then combined into a single Data Quality Factor (DQF). The calculated DQFs will inform the Data Derating Factor (DDF) for Sustain-H remuneration to reflect the commercial impact of the uncertainties quantified.

This report is an updated version of the third deliverable. The RUFs and subsequently DQFs calculated and presented in the third deliverable have been updated using asset-meter data from one of the Sustain-H trial participants who had the largest portfolio of EV chargepoints metered at the asset level. The results from the initial and refined analysis, as well as recommendations to WPD are presented in this report.

2.2 OBJECTIVES

The objective of the *Aggregated Datasets* workstream is to help quantify the impact of lower data quality, and thereby inform Sustain-H payments. Specifically, the workstream aims:

- To quantify to what extent aggregation might address data quality challenges at domestic level;
- To establish the implications for DSO service procurement; and
- To provide necessary information to, and integrate with, the Sustain-H service.

The output will be a method to estimate the Data Derating Factor (DDF) within the Sustain-H payment formula.

2.3 REPORT STRUCTURE

The report adopts the following structure:

- **Section 3 – Methodology:** This section recaps the Methodology outlined in [2], including outlining updates.
- **Section 4 – Results & Discussion:** This section presents the results of the analysis.
- **Section 5 – Conclusions:** This section reflects on key findings from the analysis, lists the limitations and recommendations.

3. METHODOLOGY

This section outlines the methodology that Everoze has used to analyse data from pre-existing datasets as presented in [1]. While analysing the data, a number of refinements to the Methodology were necessary.

This section includes the following:

- Summary of method;
- Method for analysing impact of Resolution on Quality as a flowchart;
- Method for analysing impact of Completeness on Quality as a flowchart;
- Method for analysing impact of Accuracy on Quality as a flow chart; and
- Method for calculating the DQF.

An extensive explanation of the methodology is written in Appendix 8, with any updates or changes to the original methodology highlighted in **red.**

3.1 SUMMARY OF METHOD

The Aggregated Datasets workstream aimed to quantify the impact that varying magnitudes of measurement uncertainty have on the measured aggregated portfolio demand. It was hypothesised that this impact will be inversely proportional to portfolio size, or, in other words, that the more the number of assets within the aggregation the more the various sources of uncertainty will tend to cancel out or diminish in terms of their proportional impact.

The method was based on the following logical sequence:

1. Any 30-minute mean demand value at an aggregate portfolio level, where completeness of the dataset is unknown and where demand is expected to vary over time (and hence within the 30-minute period), can be improved by an understanding of:
    a. Resolution – where correcting from 30-minute resolution (for example) to determine the peak 1-minute resolution demand during that period would increase the perceived demand.
    b. Completeness – where correcting for “incompleteness” would increase the perceived 30-minute average demand; and
    c. Measurement accuracy – where measurements of individual assets or smart meters are independent of each other, and normally distributed.
2. Because these factors are independent of each other, their respective impacts can be treated as such. This means they can be assessed independently, and combined using well-established statistical model(s) or simple arithmetic.
3. Therefore, Everoze produced a model of independent and ‘variable’ (therefore probabilistic in nature) Uncertainty Factors (UFs) for: Resolution using third party data from Electric Nation, Freedom and Sustain-H trial, Completeness from the Freedom trial, and Accuracy using data from asset meter and smart meter data sheets.
4. The Completeness and RUFs consist of a ‘multiplier’ of the 30 minute average, and the variability is represented by the standard deviation of the multiplier (from the available data).
5. For Accuracy, the multiplier has been set as 1. However, there is a standard deviation which Everoze has defined by accuracy measurement based on available information (e.g. from datasheets and Sustain-H trial participant data, and the number of measurements made).
6. The UFs have been combined into a single DQF for each combination of elements considered, with outputs in a tabulated form:
    a. Dataset for the pre-trial datasets (Electric Nation, Freedom) and Sustain-H trial dataset;
    b. Portfolio size;
    c. Completeness; and
    d. Resolution.
7. The DQFs will be used to establish the DDFs (commercial impact), although this will be undertaken by WPD outside of this sub-workstream.

3.2 DATA PRE-PROCESSING

The data was pre-processed following the logical steps captured in Figure 2 including a sample format for the output data files. The output files were fed into the technology-agnostic Completeness and Resolution code for the separate analyses. The pre-processing steps are described in Appendix 1 and a summary of the pre-processing results is set out below.

Freedom: The Freedom trial ran from 1st October 2017 to 30th April 2018. During this period, 53.2% of the 1-minute asset meter data (aggregated from 75 assets) was not flagged as 'null' data. A 61-day period from 1st March to 30th April 2018 displayed the highest consistent participant availability, and so this window was chosen to run the complete analysis and is referred to as the 'High Availability Window'. Everoze removed any whole-asset data if the asset was unavailable for more than 50% of this smaller sampling window. Following pre-processing of the dataset, 49 assets remained. The Resolution and Completeness analyses were performed on this High Availability Window of the full dataset: 49 assets from 1st March to 30th Apr 2018, with only 5.5% of the data being 'null' records.

Electric Nation: The Electric Nation trial ran from January 2017 to December 2018. Trial participants were either in 'Trial 1', 'Trial 2', 'Trial 3' or were not allocated to a particular trial. The different trials were subject to different interventions, therefore Everoze decided, if the impact to the population size used for the analysis was not significant, that the analysis should be performed using data from only one trial for consistency rather than combining the data from all trials. Following pre-processing, Trial 1 displayed the highest population availability (155 assets) over a period of approximately three months (18th April to 15th July 2018) . Therefore, Trial 1 was selected. The Resolution analysis was performed on this portion of the full dataset: 155 assets from Trial 1, from 18th April to 15th July 2018.

Sustain-H: The Sustain-H trial ran from November 2020 to July 2021. There were two seasons in the trial: winter (Nov – Feb) and summer (Mar – Jul). Trial participants entered portfolios into either or both seasons and provided a service to WPD by dropping their portfolio demand to-or-below a pre-agreed level during fixed delivery windows: Morning (8am – 12 pm) and / or Evening (4 pm – 8 pm). There were 5 participants in the winter and 5 participants in the summer seasons, 3/5 winter participants also participated in the summer season. Trial participants were asked to provide 24 hr asset meter data as either 1) minutely resolution data for the portfolio or individual meters or 2) the settlement 1 minute peak demand and settlement 30 minute average demand for the purposes of the Aggregated Dataset workstream on a voluntary basis. This data was additional to the portfolio settlement data required for the Sustain-H trial. Asset meter data were provided by 3/5 of the participants for each season. Data received is summarised in Table 1.

The Sustain-H datasets were relatively small, which limited the analysis which could be achieved for this phase 2 report. Despite receiving datasets from four participants across the two seasons, the majority of portfolios were too small to derive conclusive results (i.e. < 10 assets/portfolio). Only the winter season dataset from Participant 2 was deemed suitable for phase 2 of the analysis, with 23 EV chargepoints in the portfolio. Four of these were v2g chargepoints and 19 were smart chargers. Both types of charger operated on a minute-by-minute basis and were optimised against imbalance price forecasts, meaning there were instances when the chargers started charging or exporting power from the grid mid-way through or at the end of the settlement period as changes in price forecast indicate a new 'path' to charging up at the lowest possible cost. The v2g batteries exported power too, so behaved similarly to a battery. Therefore, Everoze conducted the resolution analysis using only the smart chargers so that the results could be compared to the Electric Nation dataset which effectively behaved as "dumb" chargers i.e. the charging was constant across a settlement period.

Participant	Winter season				Summer season			
	Households	EV Chargepoints	Batteries	Heat pumps	Households	EV Chargepoints	Batteries	Heat pumps
Participant 1	2	0	2	0	6	0	6	0
Participant 2	23	23 19 smart chargers and 4 v2g chargers	0	0	N/A	N/A	N/A	N/A
Participant 3	7	7	4	0	5	5	3	1
Participant 4	N/A	N/A	N/A	N/A	2	2	0	0

TABLE 1: ASSET METER DATA RECEIVED FROM SUSATIN-H TRIAL PARTICIPANTS

Phase 1 of the analysis found conclusive results from the completeness analysis using data from the Freedom trial. Therefore, following completion of the analysis in phase 1, it was decided that phase 2 would focus on refining the results for the RUF only. During phase 1, the Completeness analysis was only performed on the Freedom dataset, and not Electric Nation. This was due to the following:

1. **The minutely Electric Nation dataset was relatively incomplete compared to the Freedom dataset.** Following pre-processing, the Electric Nation dataset contained charging events for the cars 5% of the time. On average, 7 out of 155 cars were charging each minute. Therefore, if Everoze were to introduce null data randomly into the existing dataset on a per asset-asset day basis, (which included the times when the null data was 0kW), the probability that the null data would replace datapoints which are already 0kW would be high. Therefore, the impact on electricity demand would be minimal hence limiting the results.
2. **For the completeness analysis, the starting dataset must first be set up to be 100 % complete.** Therefore any null data must be replaced with realistic values based on the available data. However, due to the transactional nature of the Electric Nation dataset, there were no minutely 'null' datapoints, only null data on whole transaction level. The raw Electric Nation data was provided on a per-transaction basis only, displaying the 'Start' and 'End' charge times, battery charger power rating (either 3.6 kW or 7 kW), and the total charging duration for each charging event (termed transaction). To convert the per-transaction data into minutely data, Everoze assumed that the charging rate was constant for each transaction, at 3.6 kW or 7 kW. In reality, charging rate would have varied throughout the charging period.

3.3 RESOLUTION

Figure 3 captures the Methodology steps for the Resolution analysis. There were three key parts to the analysis which Everoze undertook:

- Determined the variance of the data at 1-minutely and for 30-minutely average data for the morning and evening Delivery Periods, and whole dataset (all settlement periods);
- Quantified the impact portfolio size has on uncertainty for the different averaging (for the Electric Nation and Freedom datasets only); and
- Estimated Resolution Uncertainty Factors, or "RUFs", for the derivation of the DQF.

3.4 COMPLETENESS

Figure 4 captures the Methodology steps for the Completeness analysis. This was only conducted on the Freedom dataset. There were three components to the Completeness analysis:

- Quantified the 'completeness' of the datasets for the portfolios considered;
- Determined the impact of different levels of data completeness on the uncertainty of measured portfolio demand representing the 'real' portfolio demand;
- Quantified the impact portfolio size has on uncertainty for a given completeness percentage; and
- Estimated Completeness Uncertainty Factors, or "CUFs", for the DQF equation.

3.5 ACCURACY

Figure 5 captures the Methodology steps for the Accuracy analysis. There were three components to the analysis:

- Quantified the 'Accuracy' of different meter types;
- Calculated the standard deviation of this uncertainty; and
- Estimated Accuracy Uncertainty Factors, or "AUFs", for the DQF equation.

3.6 DATA QUALITY FACTOR

The DQF was calculated by combining the outputs from the independent Resolution, Completeness and Accuracy analyses. The independent analyses have produced a series of 'Multipliers' and 'Standard Deviations' for different scenarios considered. Everoze has calculated the DQF for different levels of confidence, 50% and 95%, using Equation 1 and Equation 2 respectively. In these equations, M represents the 'Multiplier' and SD the standard deviation. Subscripts R, C and A denote Resolution, Completeness and Accuracy, respectively.

The DQFs have been calculated for each dataset independently, using data across all settlement periods rather than just the individual Delivery Periods. The DQF per dataset was first calculated in phase 1 and refined in phase 2 using data from the Sustain-H trial.

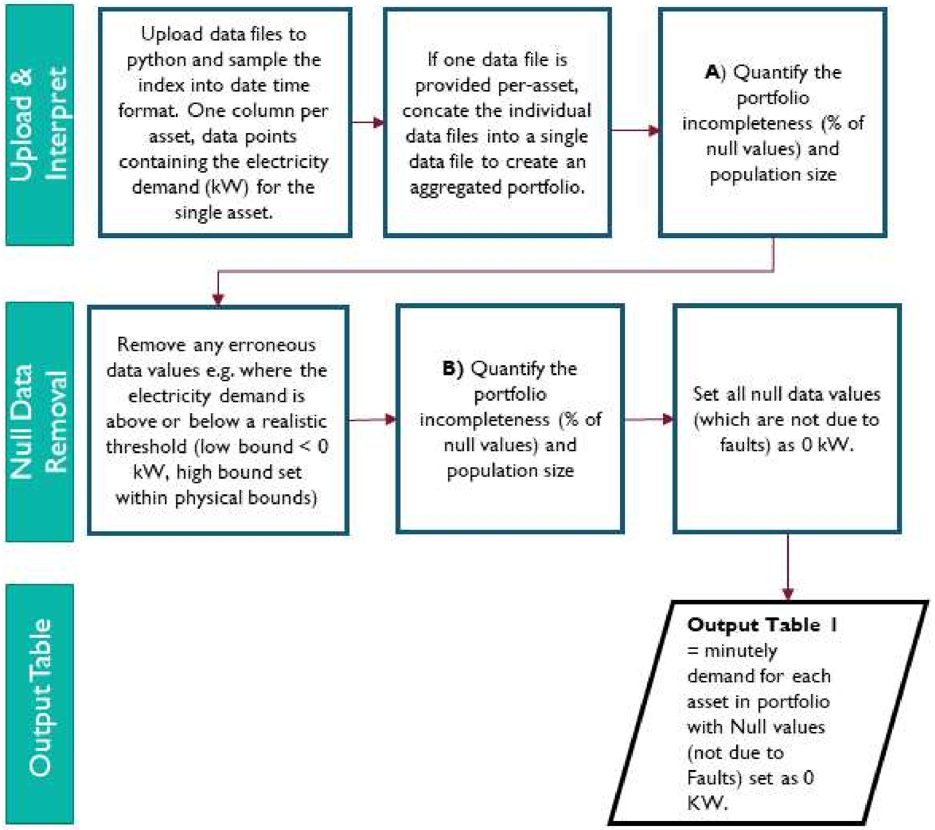
$$DQF = M_R \; x \; M_C \; x \; M_A$$

EQUATION 1 – DQF FOR A 50% CONFIDENCE LEVEL

$$DQF = M_R \; x \; M_C \; x \; M_A + (1.645 \; x \sqrt{(SD_R)^2 + (SD_C)^2 + (SD_A)^2})$$

EQUATION 2 – DQF FOR A 95% CONFIDENCE LEVEL

Pre-Processing

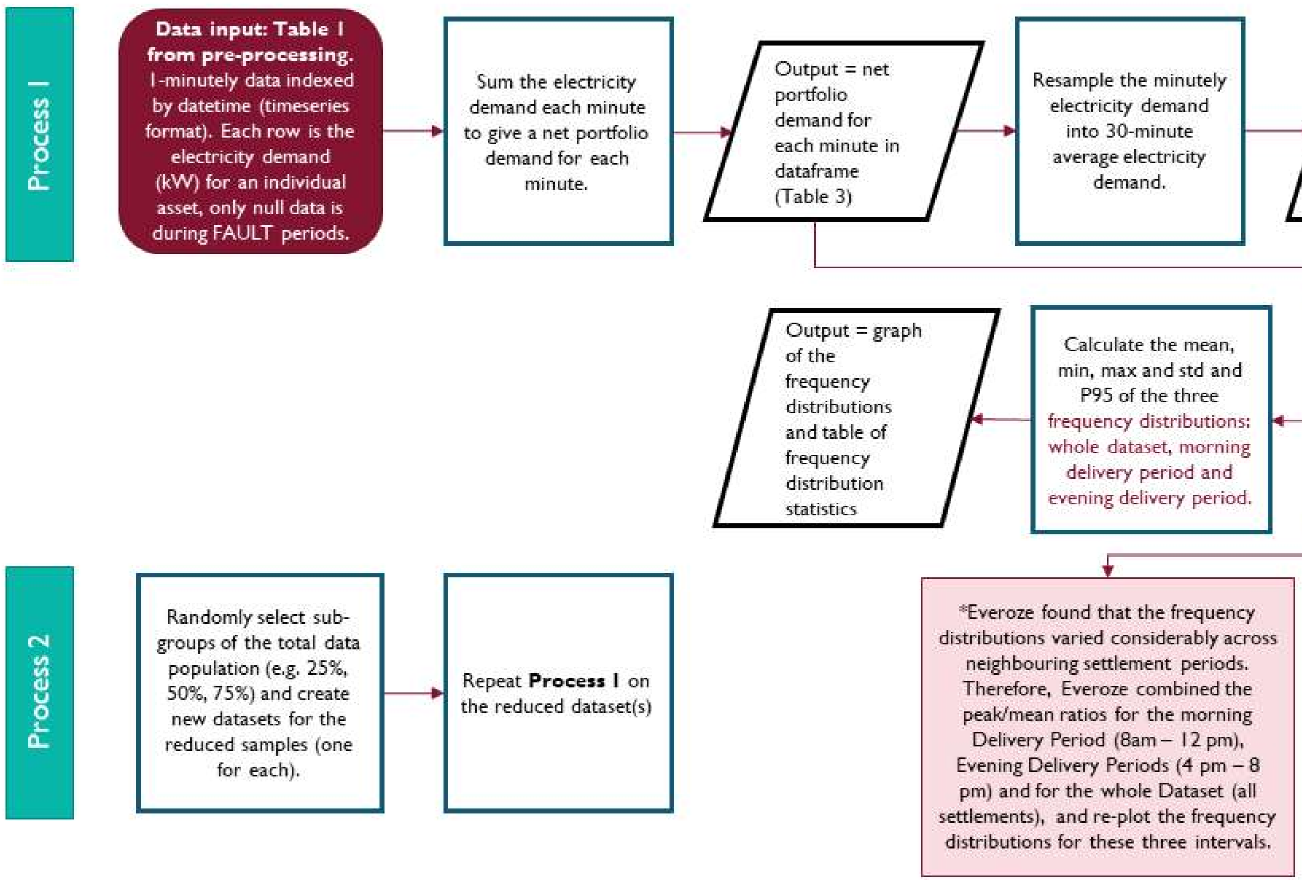


	Electricty Demand (kW)					
DateTime	Asset 1	Asset 2	Asset 3	Asset 4	Asset 5	Asset 6
05/03/2020 00:00	0.1	0	0.2	…		
05/03/2020 00:01	0.5	0.3	0			
05/03/2020 00:02	…					
05/03/2020 00:03						
05/03/2020 00:04						

Table 1

FIGURE 2: DATASET-AGNOSTIC PRE-PROCESSING STEPS

Resolution

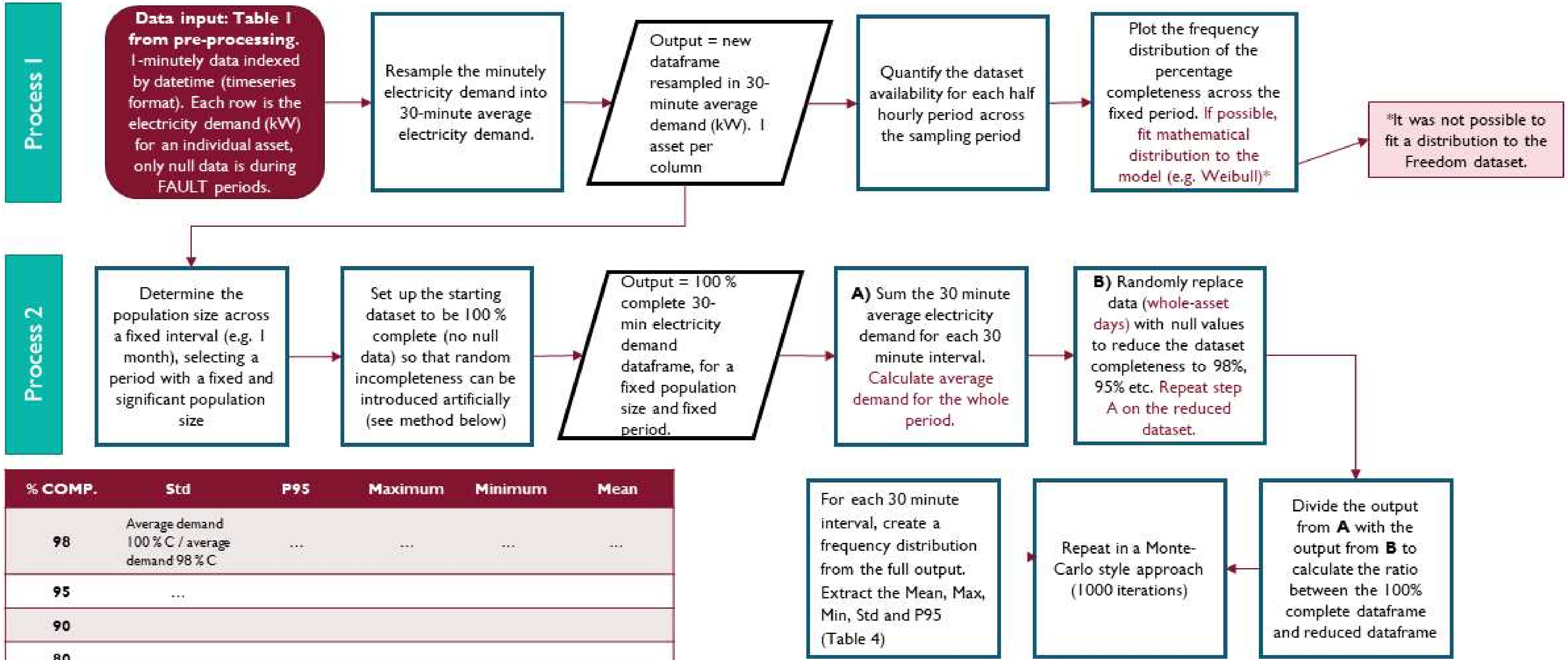


Timestamp	Portfolio Electricity Demand (kW)
05/03/2020 00:00	20
05/03/2020 00:01	21
05/03/2020 00:02	...
05/03/2020 00:03	
05/03/2020 00:04	

Table 2

FIGURE 3: DATASET-AGNOSTIC METHODOLOGY STEPS FOR THE RESOLUTION ANALYSIS

Completeness



% COMP.	Std	P95	Maximum	Minimum	Mean
98	Average demand 100 % C / average demand 98 % C	...	...	...	...
95	...				
90					
80					

Table 3

Method to make dataset 100 % complete, In order of priority:

a. Take the average for asset with identified invalid data across the same hour across the duration of the dataset, split between weekdays and weekends.
b. If the data available to perform the averaging process in item 1 is < 75%, then: take a through day average for that asset, split between weekdays and weekends
c. If the data available to perform the averaging process in item 2 is <75%, then either delete that asset all together, or take an average across all assets for that time of day.

FIGURE 4: DATASET-AGNOSTIC METHODOLOGY STEPS FOR THE COMPLETENESS ANALYSIS

Accuracy

Process 1

Data input: Smart meter and Asset meter accuracy measurements from regulatory documents and Sustain-H participants

Select an accuracy meter reading: if the Participant provides the meter accuracy, use that value. Otherwise, approximate accuracy based on regulatory data.

Using the equation below, calculate the accuracy measurements for the sample portfolio size.

$$AUF\ (\text{Standard Deviation}) = \frac{\text{Meter Accuracy for a single asset in Portfilio}}{\sqrt{\text{Number of Assets in Portfolio}}}$$

Output = dataset-specific standard deviation (a function of portfolio size and meter accuracy). Table 5 provides an example of the range of outputs*.

*It had been concluded that the multiplier for Accuracy is 1.0 as there is no expected bias in meter readings, only uncertainty from the meter accuracy

Population Size	Standard Deviation [%] for 1 % accuracy
1	1.000
5	0.447
25	0.200

Table 4

FIGURE 5: DATASET-AGNOSTIC METHODOLOGY STEPS FOR THE ACCURACY ANALYSIS

4. RESULTS & DISCUSSION

4.1 RESOLUTION

4.1.1 Determine the variance of the data at 1-minutely intervals and when averaged into half hourly intervals for a portfolio.

The intention of this exercise was to understand how much greater the demand can be on a short term (minutely) resolution when compared to a half-hourly resolution. This will enable WPD to understand how large the minutely peak demand might be in a given half hourly period when provided with only half hourly data.

To do this, Everoze has determined the 'peak/mean' ratio for the Electric Nation, Freedom and Sustain-H datasets following the analysis described in Section 3.3 and Appendix 1.

Results from this component of the resolution analysis have been refined following the initial analysis conducted in phase 1.

Everoze has produced the following outputs:

1. Line plot – peak/mean ratio: A line plot comparing the peak/mean ratio for the three datasets for each settlement period, Figure 6.
2. Daily consumption profile: To review the variation of the peak over mean ratio it is useful to also understand the daily variation in the power consumption of the three datasets, as shown in Figure 7.
3. Frequency distribution – peak/mean ratio: Frequency distribution plots of the peak/mean ratio for the Morning and Evening delivery periods, and all Settlement Periods, Figure 8, Figure 9 and Figure 10, and a table showing the Mean, Minimum, Maximum, Standard Deviation and P95 of those plots (Table 2).
4. Scatter plots – peak/mean ratio: Scatter plots capturing the variation of peak/mean ratio to the half hourly demand average (mean) across all settlement periods, Figure 11, Figure 12 and Figure 13.

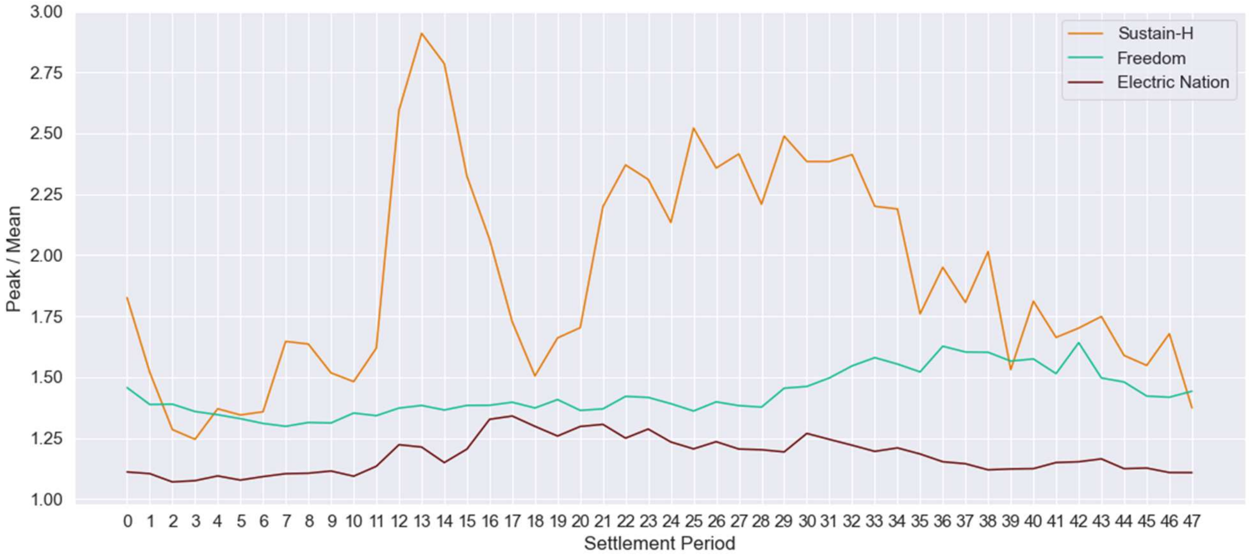


FIGURE 6: MEAN OF THE PEAK / MEAN RATIOS PER SETTLEMENT PERIOD

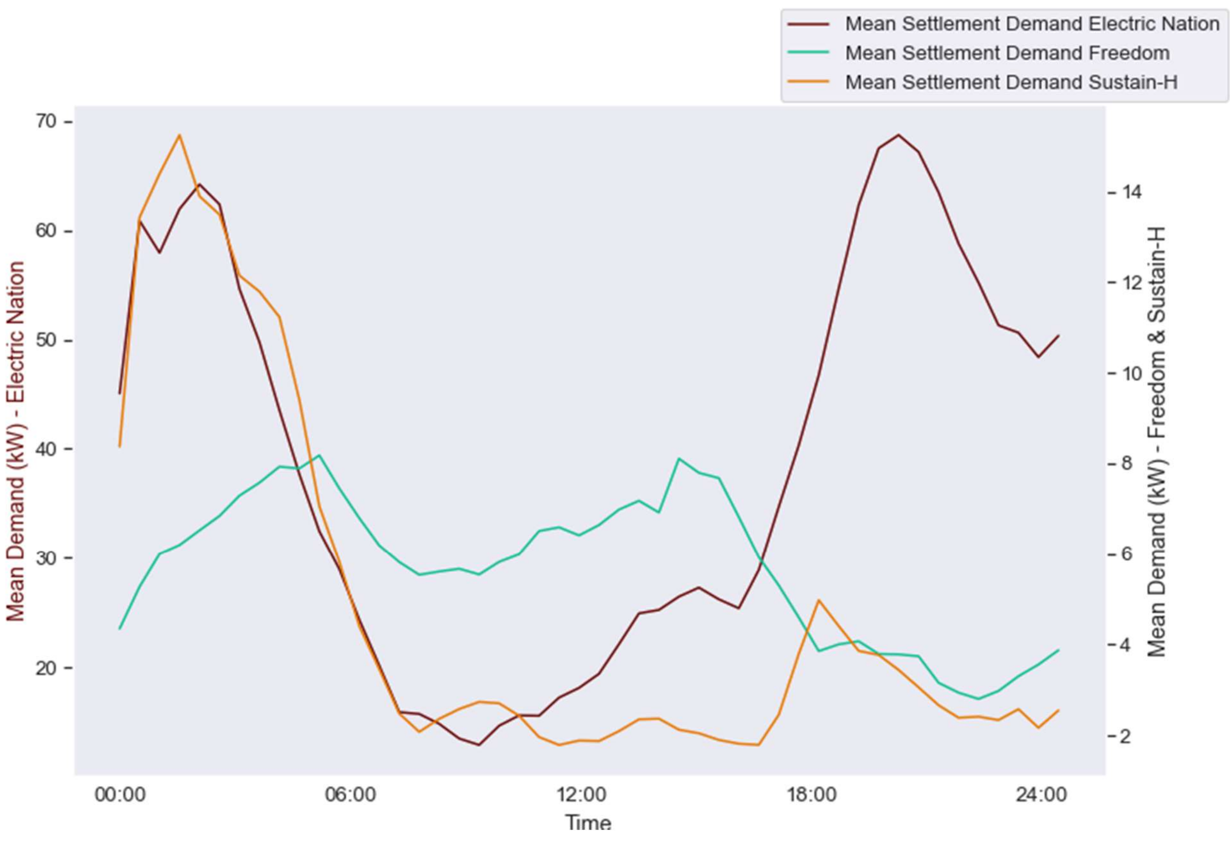


FIGURE 7: MEAN POWER DEMAND PER SETTLEMENT PERIOD

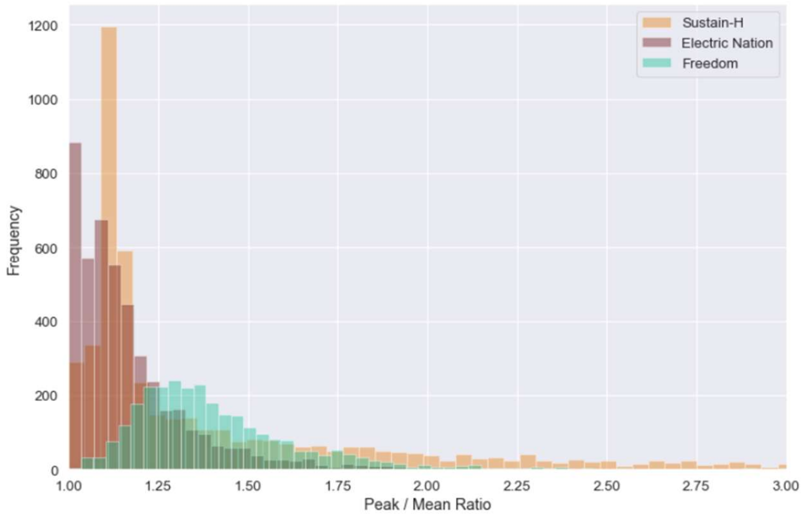


FIGURE 8: FREQUENCY DISTRIBUTIONS OF THE PEAK / MEAN RATIOS FOR ALL SETTLEMENT PERIODS (ALL DATA)

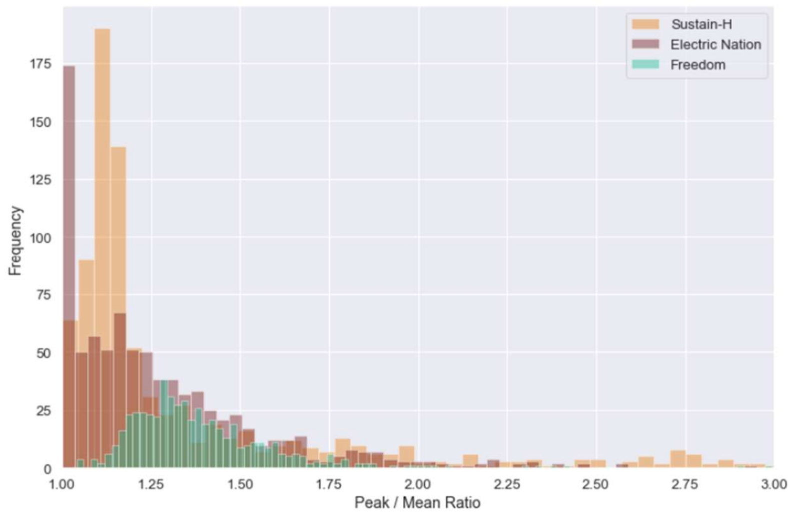


FIGURE 9: FREQUENCY DISTRIBUTIONS OF THE PEAK / MEAN RATIOS FOR THE MORNING DELIVERY PERIOD

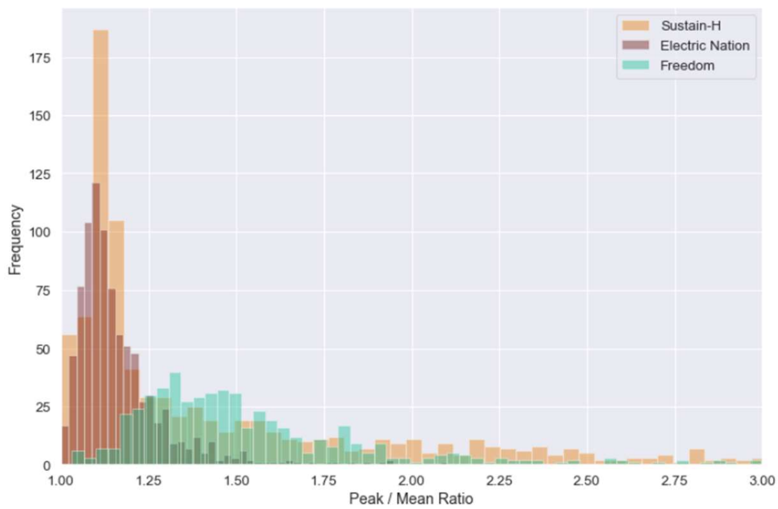


FIGURE 10: FREQUENCY DISTRIBUTIONS OF THE PEAK / MEAN RATIOS FOR THE EVENING DELIVERY PERIOD

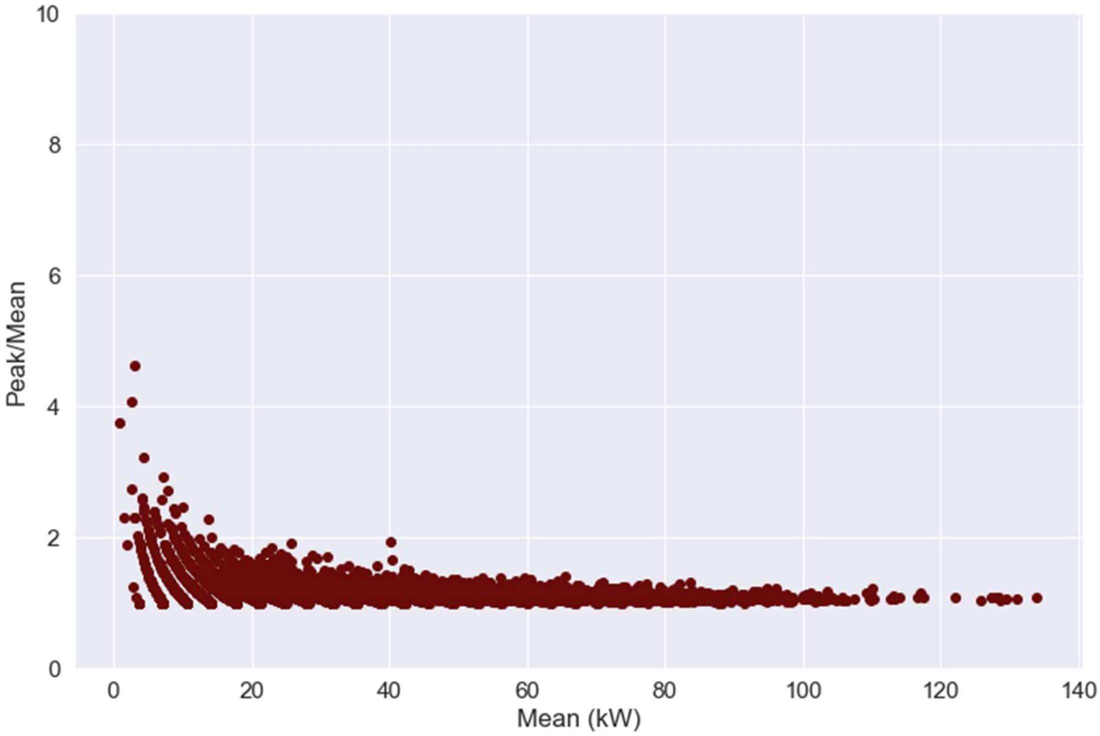


FIGURE 11: PEAK / MEAN AGAINST MEAN DEMAND – INCLUDES DATA FROM ALL SETTLEMENT PERIODS. DATASET: ELECTRIC NATION

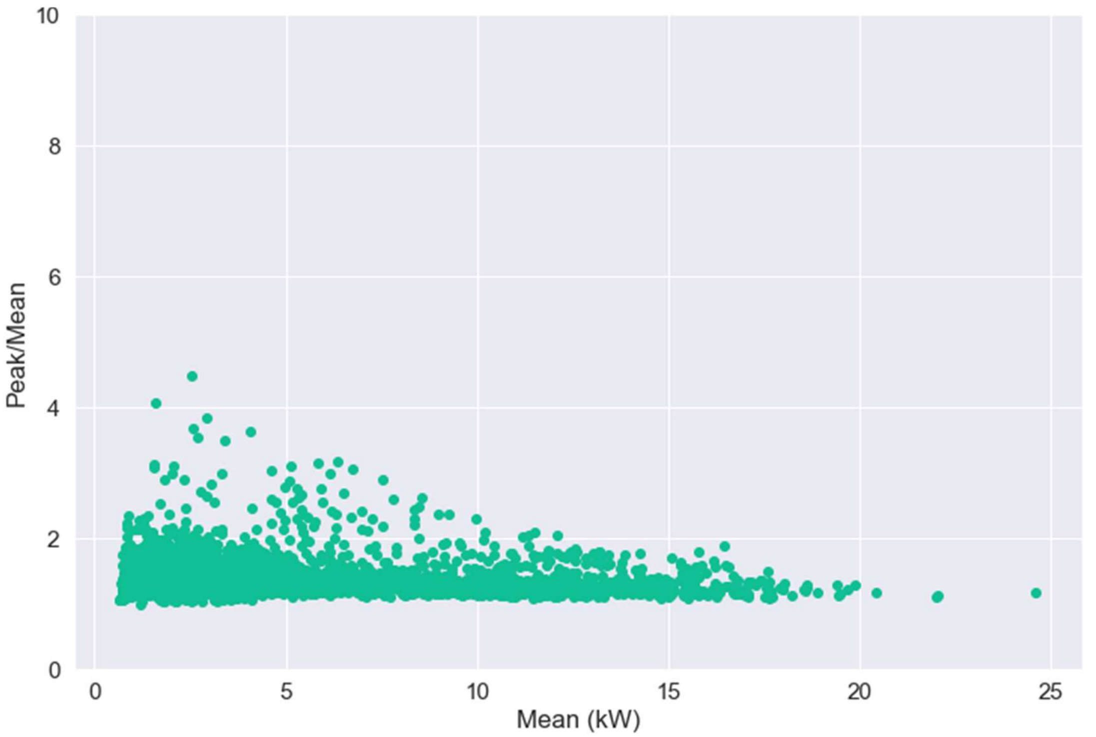


FIGURE 12: PEAK / MEAN AGAINST MEAN DEMAND – INCLUDES DATA FROM ALL SETTLEMENT PERIODS. DATASET: FREEDOM

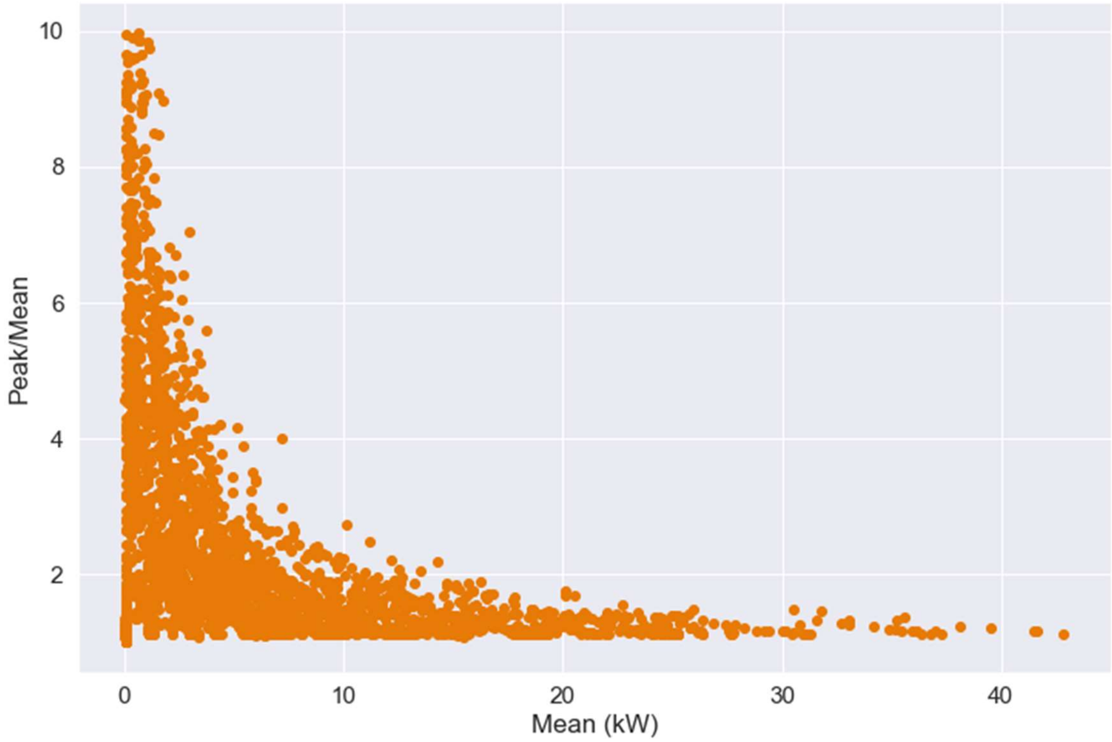


FIGURE 13: PEAK / MEAN AGAINST MEAN DEMAND – INCLUDES DATA FROM ALL SETTLEMENT PERIODS. DATASET: SUSTAIN-H

Delivery period	Dataset	Mean	Minimum	Maximum	Standard Deviation	P95
Morning	Freedom	1.39	1.04	2.99	0.22	1.76
	Electric Nation	1.29	1.00	4.62	0.34	1.88
	Sustain-H	1.96	1.00	9.98	1.61	5.79
Evening	Freedom	1.57	1.03	4.48	0.44	2.46
	Electric Nation	1.16	1.00	2.10	0.13	1.41
	Sustain-H	1.97	1.00	9.95	1.51	5.29
All settlement periods	Freedom	1.43	1.00	4.48	0.30	1.91
	Electric Nation	1.18	1.00	4.62	0.21	1.54
	Sustain-H	1.91	1.00	9.98	1.49	5.25

TABLE 2: KEY METRICS FROM THE PEAK / MEAN FREQUENCY DISTRIBUTIONS

Key observations

The trends observed from these three datasets are generally quite different. In particular:

- Consumption profiles vary dramatically between the EV datasets (Electric Nation, Sustain-H) and heat pump (Freedom) datasets. Figure 7 shows a plot of the mean power demand (kW) per settlement period for the Electric Nation, Freedom and Sustain-H datasets. The power demand for the Freedom dataset peaks at around 5 am and 2 pm, before the morning and evening electricity price peaks. This relationship is possibly seen because the heat pump systems in the Freedom project were optimised for least cost to customer, warming homes before peak energy price periods. In contrast, the Electric Nation dataset shows demand peaking at 2 am, reaching a low from 8 am when most people are using their cars, and steeply increasing from 4 pm to a peak around 8 pm as people plug in their cars when returning home from work. The Sustain-H dataset peaks in the morning, following a similar trend to the Electric Nation dataset. Conversely, the dataset shows only a small second peak in the evening around 6 pm, whereas the Electric Nation dataset peaks again at this time. This is likely a feature of the Sustain-H trial, which required participants to drop their power demand to a pre-agreed level during delivery windows. Participant 2, who's dataset was used for the RUF analysis, only participated in the evening delivery window (between 4 pm – 8 pm). Consequently, the power demand during this period is lower because charging was shifted outside of this the delivery window.
- The peak/mean ratio varied during the day for all three datasets – but the trends are not consistent with each other. This is shown in Figure 6. There are higher values during the evening for the Freedom dataset, whereas peak values for the Electric Nation dataset were during the morning. The Sustain-H peak/mean ratios were more than 50 % greater than the Electric Nation and Freedom datasets for 54 % and 33% of settlement periods. On 6 occasions (12.5% of settlement periods) the peak/mean ratios were over 100% greater than the Electric Nation dataset. Interestingly, the peak/mean ratios from the Freedom and Sustain-H datasets are similar during the night from approximately 8 pm until 6 am. In contrast, the variation through the day between the Freedom and Electric Nation datasets was not particularly large, with the highest and lowest settlement periods being within 0.5 of each other.
- The distributions of the peak/mean ratios shown for each delivery period and the whole day are very different. This is shown in Figure 8, Figure 9 and Figure 10 respectively. For the Freedom data these are characterised by Weibull distributions whereas they are exponentially decreasing for Electric Nation and Sustain-H. The differences can be explained by: 1) the typical behaviour of the types of asset within each dataset. By their very nature, heat pump power consumption can vary a reasonable amount over short time frames. Electric vehicle chargers are likely to be on for longer periods of time and are more binary in their operation. This means that for electric vehicles the peak for many settlement periods, may well be the same as the mean. 2) The asset's control system will cause power consumption variability, for example the Freedom hybrid heat pumps under full optimisation will be quite variable, whereas with less stringent control, the power demand profiles may not vary significantly. In the case of the Electric Nation trial, Everoze assumed a constant charging rate from start to end of the 'transaction' at 3.6 kW or 7 kW (depended on the car battery charger rating). Hence the mean and peak was often closer to 1 when compared to the other two datasets. This assumption was made because typically, when plugging in an EV for charging, the power demand profile was assumed to be relatively constant and flat until 80 % of the charge has been met, at this point the power demand decreased. However, the Sustain-H dataset behaved differently, which also consists purely of EV charepoints. The ratio rarely dropped below 1.5, even during the evening delivery window when all assets were required to drop their demand to a pre-agreed level. Portfolio size could explain some of the disparity between the Electric Nation and Sustain-H datasets, however the main driver is likely the different types of chargers used. The Sustain-H participants' chargers were 'smart' and optimised for imbalance price on a minute-by-minute basis, as explained in Section 3.2. However, the impact of portfolio size was investigated further and results are presented in Section 4.1.2.
- In addition, an interesting outcome was the similarity in the graphs of peak/mean plotted against the mean for all settlement periods, shown in Figure 11, Figure 12 and Figure 13. These graphs show there is a trend of increasing peak/mean ratio as consumption reduces. This is understandable as during a low consumption settlement period it requires less behaviour change to introduce a peak/mean ratio equal to or greater than 2.0. This outcome can also be demonstrated by the higher peak/mean ratio for the evening delivery period for the Freedom dataset, as observed by the low evening demand in Figure 7. The Sustain-H dataset showed a similar result. During the night / early hours of the morning when consumption was highest as shown in Figure 7, the peak / mean ratios were lowest, demonstrated by Figure 6.

4.1.2 Electric Vehicle Datasets: Comparing the variance of the data at 1-minutely intervals and when averaged into half hourly intervals for a portfolio.

Everoze investigated the variance in peak / mean values between the two electric vehicle datasets: Electric Nation and Sustain-H. To achieve this, Everoze:

1. Compared the mean and standard deviations of the peak/mean ratio of the full-population datasets.
2. Investigated the impact portfolio size had on the variation in peak/mean ratio. Everoze reduced the population size of the Electric Nation dataset from 155 down to 19, to be the same size as the Sustain-H dataset for direct comparison. Everoze re-calculated the peak/mean ratio for each settlement period on random sample of 19 assets from the full Electric Nation dataset 100 times, and calculated the peak/mean mean, min, max, std and P95 for all settlement periods, and the two delivery windows. The mean of 100 key metrics were calculated to give 1) values for the whole dataset 2) values for the morning delivery window and 3) values for the evening delivery window, as presented in Table 3.

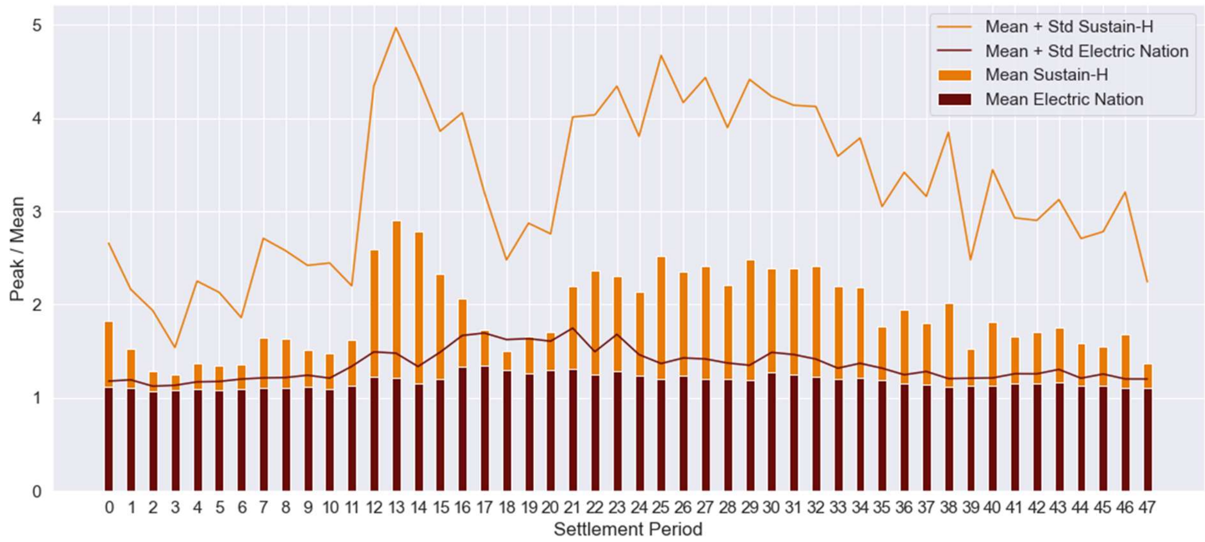


FIGURE 14: MEAN AND MEAN + STD OF THE PEAK / MEAN RATIOS PLOTTED PER SETTLEMENT

Delivery period	Dataset	Mean	Minimum	Maximum	Standard Deviation	P95
Morning	Electric Nation (n19)	1.45	1.00	4.89	0.77	3.26
	Electric Nation (n155)	1.29	1.00	4.62	0.34	1.88
	Sustain-H (n19)	1.96	1.00	9.98	1.61	5.79
Evening	Electric Nation (n19)	1.35	1.00	4.95	0.59	2.49
	Electric Nation (n155)	1.16	1.00	2.10	0.13	1.41
	Sustain-H (n19)	1.97	1.00	9.95	1.51	5.29
All settlement periods	Electric Nation (n19)	1.30	1.00	5.00	0.60	2.52
	Electric Nation (n155)	1.18	1.00	4.62	0.21	1.54
	Sustain-H (n19)	1.91	1.00	9.98	1.49	5.25

TABLE 3: KEY METRICS FROM THE PEAK / MEAN FREQUENCY DISTRIBUTION (N =19)

Key observations

1. Decreasing the population size of the Electric Nation dataset from 155 to 19 (to the same population size as the Sustain-H dataset) had the effect of increasing the mean peak / mean ratio by 11%, 16 % and 12 % for the morning, evening and all settlement periods respectively. The standard deviation increased by 226 %, 454% and 286 % (approximately an increase of 3x for a population decrease of 88%). This suggests that standard the deviation of the peak / mean ratio is strongly effected by portfolio size, however the mean of the peak / mean ratio variance is much smaller. Based on the effect seen by the Electric Nation dataset, increasing the Sustain-H dataset by 188% (to 155) could reduce the standard deviation to 1/3rd to 0.53, 0.50, and 0.50 for the morning, evening and all settlement periods respectively which is broadly similar to the Electric Nation dataset (with a population of 155).
2. Reducing the population size of the Electric Nation dataset to the same size as the Sustain-H dataset had the effect of increasing the peak / mean ratio. However, the average difference in the peak / mean ratio across datasets was 0.5 on average. The difference is likely to be explained by the way the chargers operate. Sustain-H chargers (Participant 2) were 'smart' whereas the Electric Nation chargers are assumed to be 'dumb'. The Sustain-H's chargers operated on a minute-by-minute basis and were optimised against imbalance price forecasts, meaning there were instances when the chargers start charging mid-way through or at the end of the settlement period, as changes in price forecast indicated a new 'path' to charging up at the lowest possible cost. This likely caused 'peakiness' in the data because the settlement peak power demand could be significantly above the average if the charger began to increase demand towards the end of a settlement period, ahead of the next settlement period. Whereas, the Electric Nation chargers had a constant charge across the whole charging transaction, causing the Electric Nation RUF to be much lower.

4.1.3 Quantify the impact portfolio size has on half hourly averaging uncertainty

Everoze investigated the impact of portfolio size on the half hour averaging uncertainty for the Freedom and Electric Nation datasets. The Sustain-H dataset population was considered too small to produce robust results and so this dataset was not considered for this analysis. The impact of portfolio size was calculated by removing whole-assets from the portfolios and repeating the analysis from Section 4.1.1. Everoze reduced the population of both datasets to 25%, 50%, and 75% of the initial size and used a Monte-Carlo approach to produce 1000 different combinations of a 25% reduced dataset, a 50% reduced dataset and a 75% reduced dataset. Everoze calculated the mean of the 'peak / mean' ratios across all settlement periods, and plotted the 1000 mean values as a frequency distribution in Figure 15 for the Electric Nation dataset and in Figure 16 for the Freedom dataset. Table 4 provides the mean values taken from the distributions in each figure.

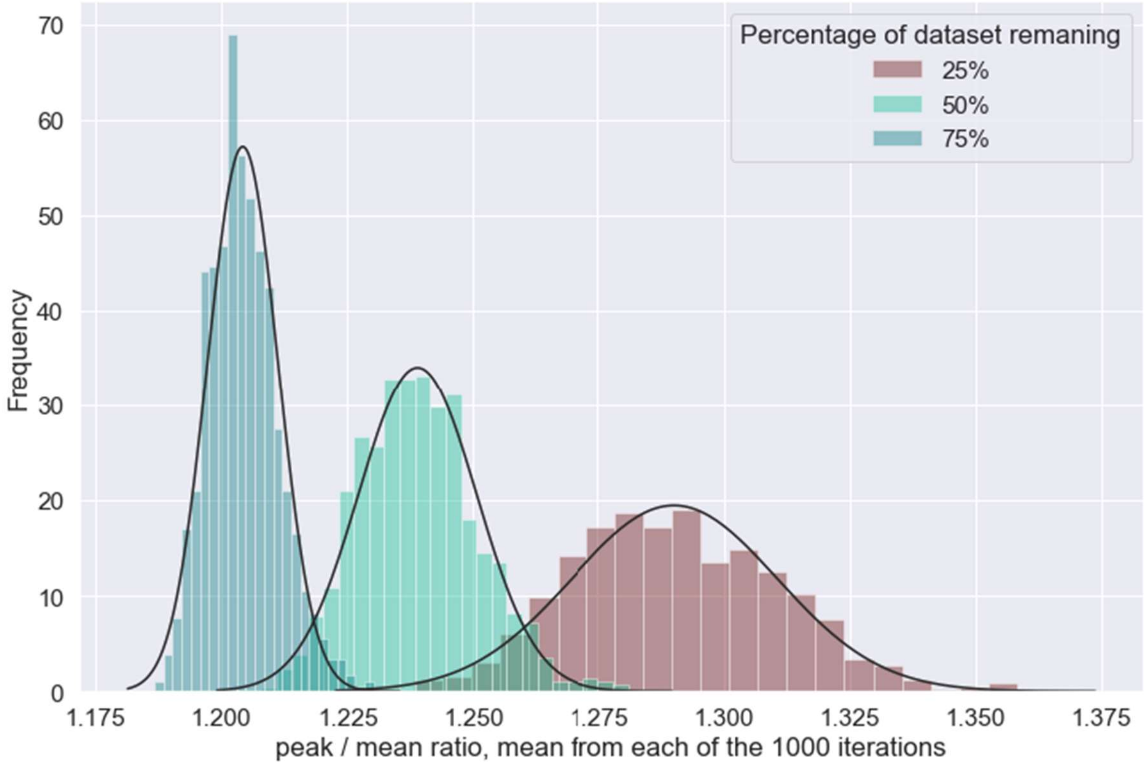


FIGURE 15: FREQUENCY DISTRIBUTIONS OF THE 'PEAK / MEAN' RATIO FOR THE REDUCED POPULATION ELECTRIC NATION DATASET

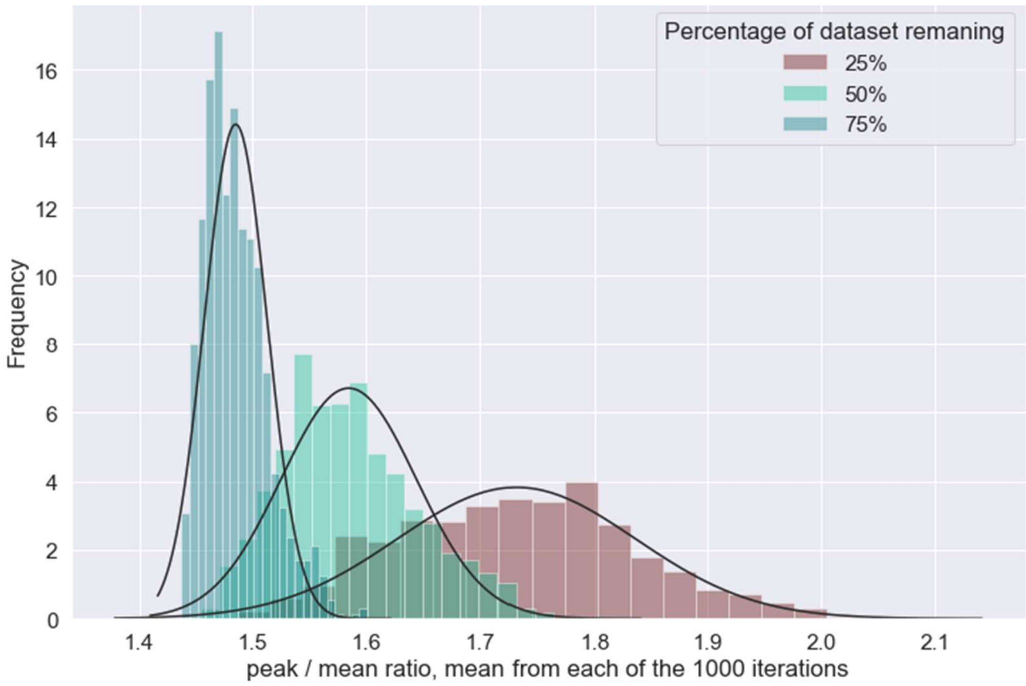


FIGURE 16: FREQUENCY DISTRIBUTIONS OF THE 'PEAK / MEAN' RATIO FOR THE REDUCED POPULATION FREEDOM DATASET

Dataset	25%	50%	75%	100%
Freedom	1.73	1.56	1.49	1.43
Electric Nation	1.29	1.24	1.20	1.18

TABLE 4: MEAN OF THE 'PEAK / MEAN' VALUES FOR THE 25 %, 50 % AND 75 % POPULATION SIZES

Everoze found that population size had a significant impact on the mean of the peak / mean ratios. The spread of the mean peak / mean values increased considerably as population size reduced. When the datasets were reduced to 25% of their initial population size, Electric Nation displayed a mean 'peak / mean' ratio almost 10% higher than the entire dataset. For the Freedom dataset there was a 20% increase for the 25% dataset.

The variance of the means of the reduced dataset 'peak / mean' distributions for the Electric Nation dataset (Figure 15) were consistently smaller than for the Freedom dataset (Figure 16). This could be attributed to either 1) the population size of the initial dataset; the population size of the Electric Nation dataset was ~3 times greater than the Freedom dataset or 2) the type of assets (heat pump or EV).

To investigate the population size further when comparing the two dataset Everoze reduced the Electric Nation dataset (which had a starting population of 155 assets) down to the equivalent number of assets in the Freedom dataset for each of its reduced datasets. So a 100% case (49 assets), and also 75% (37), 50% (25), 25% (13) were considered. Everoze investigated the hypothesis that irrespective of asset type, the mean 'peak / mean' ratio is driven entirely by population size. Table 5 compares the mean values for the two datasets for at the same population size.

Dataset	Assets = 13	Assets = 25	Assets = 37	Assets = 49
Freedom	1.73	1.56	1.49	1.43
Electric Nation	1.34	1.31	1.29	1.27
Sustain-H	1.91 (assets = 19)	-	-	-

TABLE 5: COMPARISON OF 'PEAK / MEAN' VALUES FOR THE TWO DATASETS

This was further explored by plotting the peak/mean as presented in Table 4 and Table 5 above against population size (Figure 17). A power law fit was applied and the trendlines are shown in the figure. The following observations were made:

The Electric Nation dataset shows a roughly linear trend across the population size, whereas the Freedom data shows a non-linear variation over the small population size considered in the analysis. The graph shows the trendlines for the peak/mean for both datasets converge for large population sizes. Based on these findings, Everoze hypothesises the following:

1. ***For sufficiently large populations (> 100 assets), the RUF's sensitivity on portfolio size is less pronounced and there is little variation across technology types.*** This suggests that there is reasonable grounds for taking a pragmatic approach when determining the DQF/DDF, such as considering a single common DQF value for populations above a certain size and not needing to distinguish this between technology types.
2. ***Variance in the RUF and therefore sensitivity to portfolio size, is only seen for small portfolios, and this sensitivity quickly diminishes for portfolios above a certain size.*** As smaller portfolios are unlikely to have a large impact on WPD's network constraint outcomes, if the aforementioned hypothesis of the variance in RUF being concentrated to small population portfolios only, a pragmatic and simplified approach may be considered where the same DQF/DDF as used for larger portfolios may also be used for small populations and avoid varying this for different population sizes and technology types.

Key observations:

1) Based on the available datasets, population size has a pronounced impact on the RUF. It is not possible to determine if the magnitude is characteristic to the different technologies because while the Electric Nation dataset appears to show smaller peak / mean ratios compared to the Freedom dataset with an equivalent portfolio size, the Sustain-H dataset (with a population size of 19 assets) showed non-conforming results.
2) Heat pump portfolios display greater demand uncertainty compared to EV portfolios (Electric Nation dataset with 'dumb chargers), and the uncertainty increases significantly as population size decreases. As noted above, the technology-specific characteristic appears to be diminished for large portfolios, and there is likely to be little difference in the demand uncertainty for the two datasets when considering large portfolios.

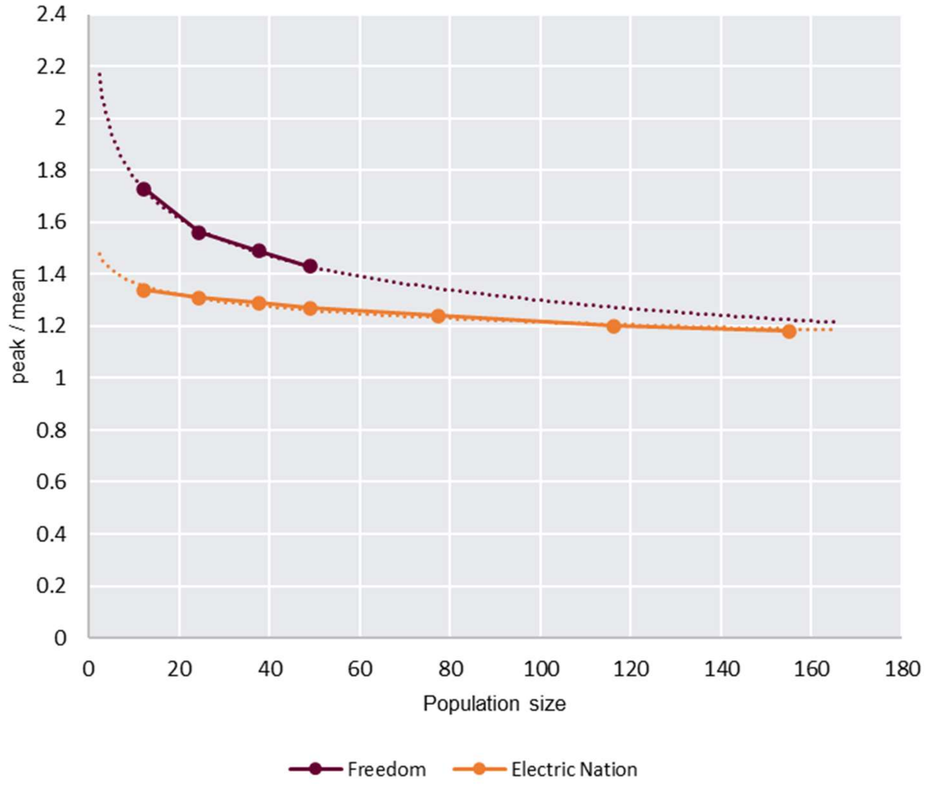


FIGURE 17: PEAK / MEAN VS POPULATION SIZE

4.1.4 Estimate value of RUFs

The RUFs are given in Table 6. Only the 'Whole Dataset' multipliers and standard deviations are used to calculate the DQF at this stage.

Delivery Period	Dataset	Multiplier	Standard Deviation
Morning	Freedom	1.39	0.22
	Electric Nation	1.29	0.34
	Sustain-H	1.96	1.61
Evening	Freedom	1.57	0.44
	Electric Nation	1.16	0.13
	Sustain-H	1.97	1.51
Whole Dataset	Freedom	1.43	0.30
	Electric Nation	1.18	0.21
	Sustain-H	1.91	1.49

TABLE 6: COMPONENTS OF THE RUFS

4.2 COMPLETENESS

The completeness analysis quantified the demand uncertainty introduced for different levels of incompleteness, when compared to a 100% complete dataset. Everoze removed whole-day assets randomly to reduce the 100% complete dataset, and calculated the ratio of the 100% complete whole-dataset power demand (kW) against the reduced dataset (kW) demand for the sampling period. As discussed in Section 3.2, only the Freedom dataset has been considered for the Completeness analysis.

Everoze randomly introduced null data, calculated the mean ratio of the 100% complete dataset demand (kW) against the reduced dataset demand (kW), and repeated this for 1000 or 10,000 iterations and plot the results from each iteration on a frequency distribution to capture the mean demand uncertainty and standard deviation for each distribution.

4.2.1 Quantify the 'completeness' of the dataset for the different portfolios considered

The initial data completeness of the Freedom dataset has already been discussed as part of the data pre-processing in the Section 3.2. While the dataset was reduced to a two month period for the analysis, this step also needed to consider the full trial dataset to provide more understanding of the dataset completeness and if there are any trends in the unavailability of data.

As an initial step Everoze produced two outputs from the analysis:

1. Percentage completeness of the datasets
2. Frequency distribution plot of the dataset availability for two sampling periods: a) full trial and b) the period used for the resolution and completeness analysis

Table 2 captures the percentage availability of the full trial period, and the two month High Availability Window from 1st March to 30th April 2018. It is clear there is a large difference in data availability and this is not surprising as these are trial data which improved in availability as the trial progressed.

Trial	Full-Trial Period	High Availability Window
Freedom	53.%	94.5%

TABLE 7: DATASET COMPLETENESS (NON-NULL DATA OVER SAMPLE PERIOD)

Figure 18 and Figure 19 display the frequency distribution plots for the full trial sample and High Availability Window respectively. Figure 18 does not show any particular trend in the unavailability of the data, while Figure 19 shows some possible tendency towards a reverse Weibull distribution for the much higher data availability case. Considering the following factors Everoze has concluded there is not sufficient evidence of a strong trend of unavailable data:

- The limited number of datasets for this analysis, only one;
- The reasonably random trend of unavailable data in the full-trial period;
- Where there is evidence of a possible trend, it is relatively weak and only defined by a two month period of data.

If Everoze had identified sufficient evidence of a trend, the null data would have been removed in the next steps of the analysis proportionally, to fit that trend. Based on the lack of a strong trend and evidence that assets are usually unavailable for an entire day within the Freedom dataset, Everoze introduced null data to reduce the 100% complete dataset on a whole-asset day basis, until 98, 95, 90 or 80 % of the data remained.

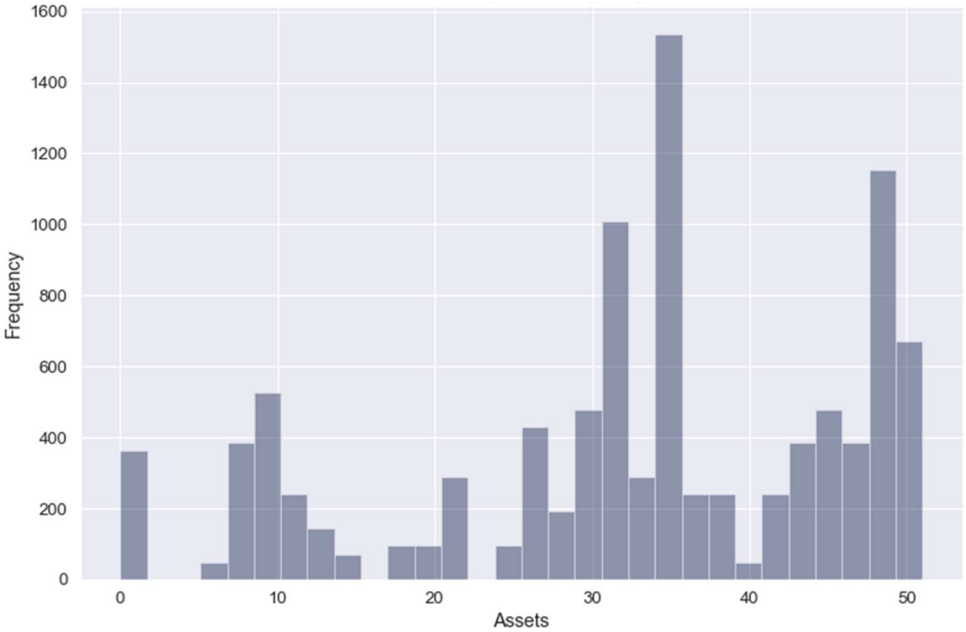


FIGURE 18: FREQUENCY DISTRIBUTION OF THE FREEDOM DATASET HALF-HOUR ASSET AVAILABILITY – FULL TRIAL PERIOD

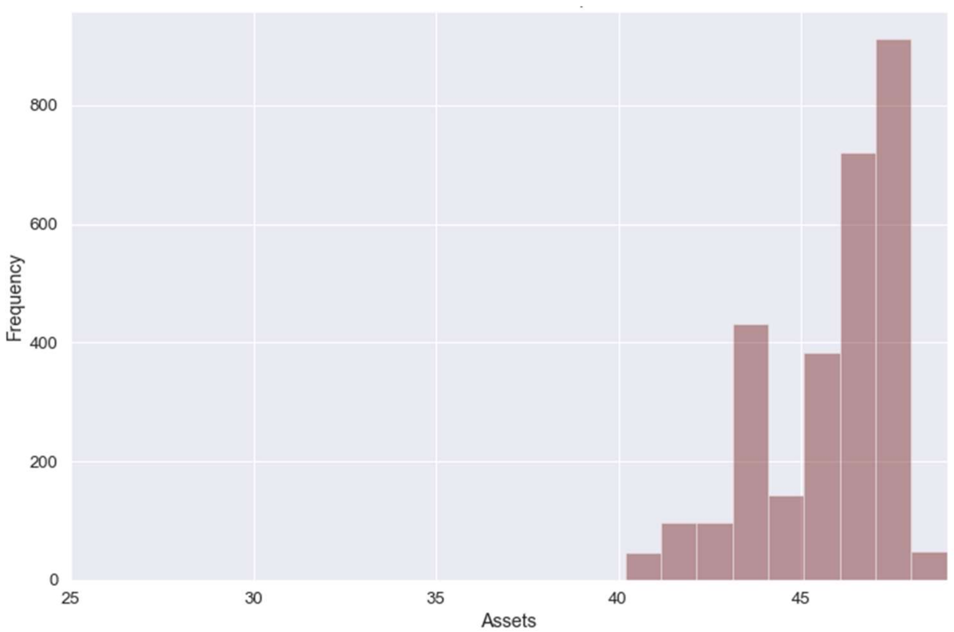


FIGURE 19: FREQUENCY DISTRIBUTION OF THE FREEDOM DATASET HALF-HOUR ASSET AVAILABILITY – SAMPLE SELECTED FOR ANALYSIS

4.2.2 For a fixed portfolio size, determine the impact 98, 95, 90 and 80 % portfolio data completeness has on measured demand uncertainty

Everoze has produced frequency distributions of the 1000 iterations of the 'Delta Demand' calculation (Equation 3) for the four different cases, as shown in Figure 20. Each iteration introduced null data randomly into the complete dataset, by removing whole-asset days, to reduce the dataset to the desired completeness percentage.

$$\text{Delta Demand} = \frac{100\ \%\ \text{complete dataset mean power demand (kW)}}{\text{x}\ \%\ \text{reduced dataset mean power demand (kW)}}$$

EQUATION 3

The reason for the range of different possible power demand levels for a given level of incompleteness, is due to the different power ratings of the individual assets, coupled with the assumption of no relationship between asset power rating and probability of data loss from that asset. This means that at one extreme, data incompleteness could occur solely in the assets of the lowest power rating, whilst at the other extreme, data incompleteness could occur in the assets solely in the assets of the largest power rating, and all variations in between. Therefore, the less complete the dataset, the wider the range of possible demand scenarios.

Each distribution appeared to be normally distributed, therefore a normal distribution fit were applied to each case and included in Figure 20. The key values that represent the normal distributions in are given in Table 8. For the 80% availability case Everoze also tested the outcome with 10,000 iterations with encouragingly similar results to the 1,000 iteration case, as shown in Table 8.

Based on the analysis of the Freedom dataset, the potential increase in true demand when compared to the demand calculated from an incomplete dataset, is shown numerically in Table 8 and graphically in terms of the mean, mean plus one standard deviation and the P95 level in Figure 21.

Completeness (%)	Iterations	Mean	Minimum	Maximum	Standard Deviation	P95
98	1000	1.0205	1.0100	1.0350	0.0032	1.0258
95	1000	1.0526	1.0388	1.0721	0.0053	1.0614
90	1000	1.1109	1.0899	1.1383	0.0077	1.1236
80	1000	1.2497	1.2141	1.3126	0.0131	1.2713
80	10,000	1.2503	1.2089	1.3089	0.0134	1.2723

TABLE 8: FREQUENCY DISTRIBUTION OUTPUTS FOR EACH COMPLETENESS CASE

Key observations:

1. <u>As might be expected, the mean of each frequency distribution is approximately equal to the ratio</u>: 100% divided by the reduced dataset completeness percentage (e.g. 100/80).
2. <u>Demand uncertainty increased as the dataset became more incomplete</u>, as reflected in the standard deviation values.
3. <u>The increase in uncertainty as unavailability increases is reasonably linear</u>.

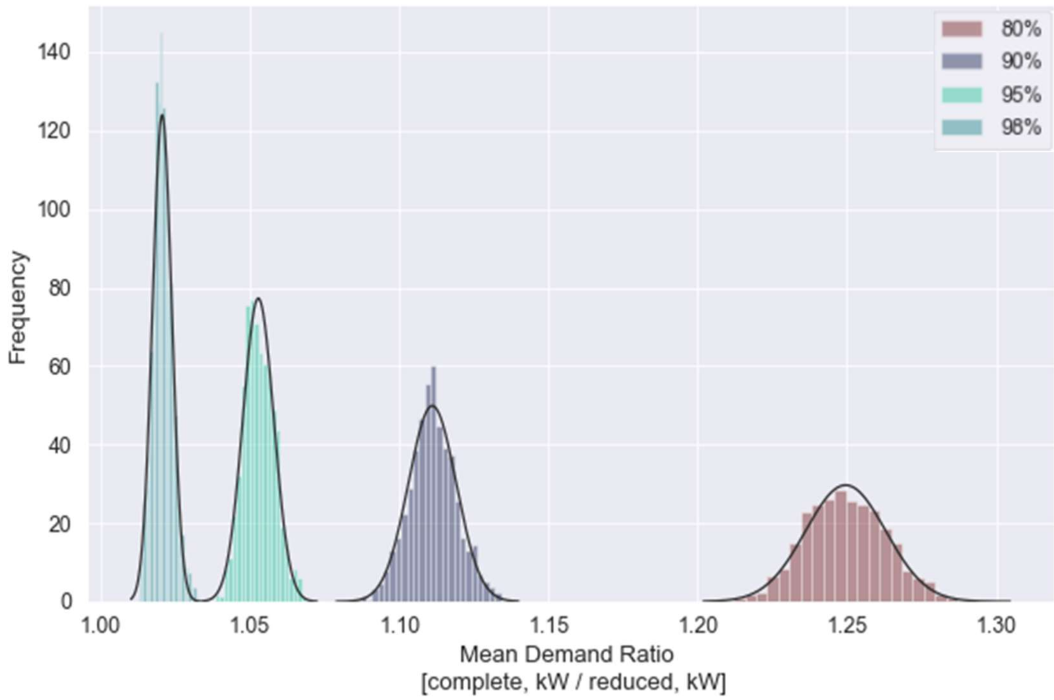


FIGURE 20: FREQUENCY DISTRIBUTIONS FOR EACH COMPLETENESS CASE FROM 1000 ITERATIONS

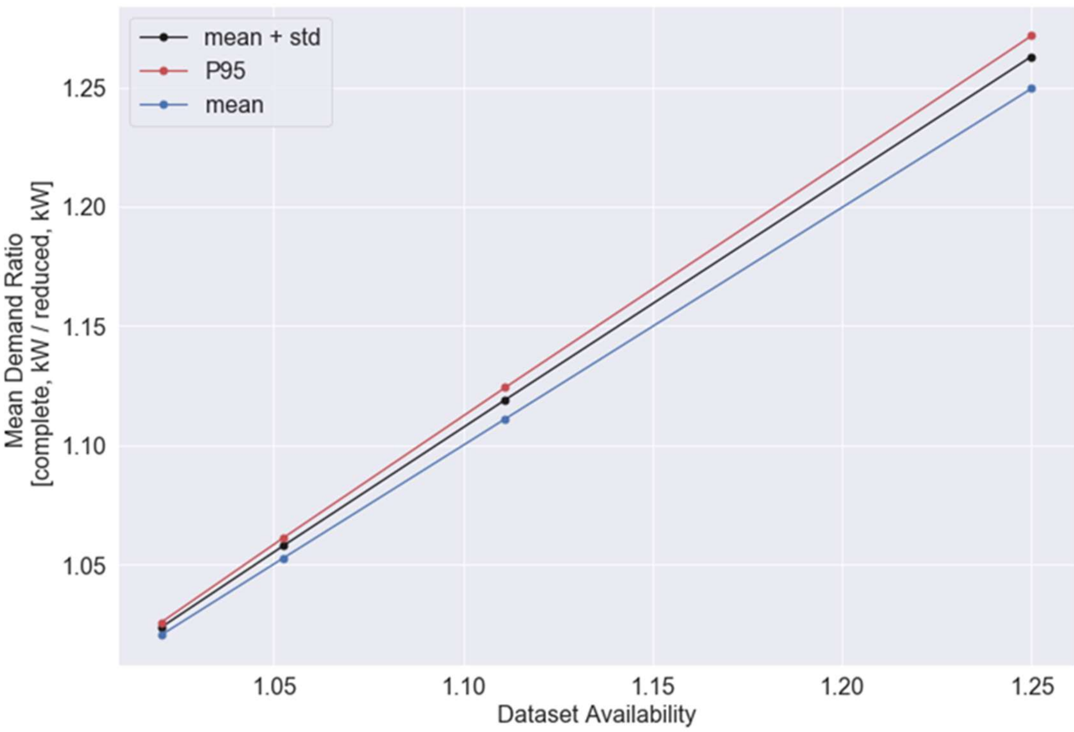


FIGURE 21: MEAN, MEAN PLUS ONE STANDARD DEVIATION, AND P95 VALUES FOR EACH COMPLETENESS CASE MODELLED.

4.2.3 Quantify the impact portfolio size has on measured demand uncertainty for the different completeness cases

Everoze investigated the impact portfolio size has on demand uncertainty for the 80% completeness case, by repeating the analysis from Section 4.2.1 on the Freedom dataset, using a 50% reduced population.

Everoze randomly removed 50% of the assets from the complete dataset, reducing the total number of assets from 49 to 24. The same process as described above in Section 4.2.2 was used to remove 20% of the remaining asset population using the same asset-day, reducing the completeness to 80%. This process was repeated for 100 iterations, and the mean of the 100% complete dataset demand (kW) / 80% complete dataset demand (kW) was calculated for each iteration. The mean of the 100 iterations equalled 1.25, which is the same as the mean calculated when the number of assets in the portfolio were 49. This demonstrates that portfolio size does not have a significant impact on the data completeness element of the demand uncertainty.

4.2.4 Estimate values of "CUFs"

Everoze calculated the CUFs for each completeness case. The multiplier is taken as the ratio of 100 / % completeness and the standard deviations are taken from the mean demand ratio (complete (kW) / reduced (kW)) frequency distribution plots.

Completeness (%)	Multiplier	Standard deviation
98	100 / 98 = 1.0204	0.0032
95	100 / 95 = 1.0526	0.0053
90	100 / 90 = 1.1111	0.0077
80	100 / 80 = 1.2500	0.0134

TABLE 9: COMPONENTS OF THE CUFS

4.3 ACCURACY

4.3.1 Quantifying the uncertainty introduced to the metered data by metering accuracy

It is necessary to understand the uncertainty introduced to the metered data attributable to the accuracy of the half hourly smart meters (SMETS1 and SMETS2 meters) and the asset level meters used by participants for the Sustain-H trial. It is expected that the impact of Accuracy on data quality will be relatively small compared to data Resolution and Completeness. Everoze developed an approach to quantify the impact of Accuracy on the measured demand as outlined in Section 3.5. It has already been concluded that the multiplier for Accuracy is 1.0 as there is no expected bias in meter readings, only uncertainty from the meter accuracy. The meter accuracy as defined by the accuracy measurement given in the available published information (from datasheets and Sustain-H trial participant data) has been taken as the standard deviation for this work.

Electricity meters are classified into a particular 'class', which indicates the maximum permissible errors (MPE) of the electricity meter, based on IST (In Service Testing) test criteria. IST is a national sampling scheme based on the British Standard BS 6002-1:1993. Class A meters have an MPE of +/- 2%, Class B meters have an MPE of +/- 1 %, and Class C meters have an MPE of +/- 0.5 %. These values are based on the test requirements in Table 4 of BS EN 50470-3:2006 for tests of accuracy at reference conditions [5]. Table 10 lists 5 accuracy measurements from MPE complaint meters, the first three are listed as UK nationally approved electricity meters [6], and the last two have been provided by Sustain-H participants.

Electricity Meter	Description	Class	Accuracy
ES-12B	Residential single phase elements SMETS 2 compliant EM	A and B	+/- 1 and 2 %
Liberty 101 - secure	Residential; single phase elements SMETS 2 compliant EM	Class B	+/- 1%
Aclara 1-201+	Residential; single phase elements SMETS 2 compliant EM	Class C	+/- 0.20%
Confidential	Participant 1: Sustain-H Participant asset meter	Class B	+/- 1%
Confidential	Participant 2: Sustain-H Participant asset meter	Class B	+/- 1%

TABLE 10: METER ACCURACY INFORMATION

4.3.2 Calculation of the standard deviation of the accuracy measurements

Assuming that the asset metering accuracy for each asset in a portfolio is the same, the standard deviation component of the AUF can be calculated using the following equation:

$$\text{Accuracy UF Standard Deviation Component} = \frac{\text{Meter Accuracy for a single asset in Portfilio}}{\sqrt{\text{Number of assets in Portfolio}}}$$

EQUATION 4

Based on the 1% accuracy of the meters provided by Sustain-H participants, the standard deviation component of the AUF for the Freedom dataset (49 assets) is calculated to be **0.0014 (0.14%),** and for the Electric Nation dataset (155 assets), the standard deviation component is **0.0008 (0.08%).** Table 11 presents the accuracy values for portfolio sizes ranging from 1 to 250, calculated for +/- 1 % and +/- 2 % meter accuracy values.

Portfolio size	Standard Deviation [%] for 1 % accuracy	Standard Deviation [%] for 2 % accuracy
1	1.000	2.000
5	0.447	0.894
25	0.200	0.400
50	0.141	0.283
100	0.100	0.200
150	0.082	0.163
200	0.071	0.141
250	0.063	0.126

TABLE 11: AUFS FOR 1% AND 2% ACCURACY

Key observations:

- Measurement accuracy has minimal impact on the measured demand, for a sample size of 100 and 1 % metering accuracy, the standard deviation calculated is 0.1 %.
- Residential smart meters and asset meters are typically accurate to +/- 1 %.

4.4 COMBINED DATA QUALITY FACTOR

When deciding on an approach to combine the independent uncertainty factors into a single DQF, Everoze concluded the following:

1. DQFs have been calculated for the three datasets separately. However, the completeness analysis was performed on the Freedom dataset only, and not the Electric Nation dataset or Sustain-H dataset for reasons described in Section 3.2. Therefore the CUF does not consider EV assets.
2. DQFs were calculated using data from all settlement periods, rather than for the two delivery periods. Results from the resolution analysis showed that there is some variation in peak/mean ratio throughout the day but there were no obvious trends for the morning or evening delivery periods. Therefore as there are limited datasets available for this work it is best to use the whole day values at this stage.
3. The DQFs calculated for the Electric Nation dataset combined the AUFs and RUFs only. Therefore the calculated DQF for EV assets is not finalised due to data limitations of the Electrical Nation and Sustain-H datasets.

Table 12 presents the DQFs for the Electric Nation, Freedom and Sustain-H datasets for both a central (P50) confidence level as well as a high (P95) confidence level. The root mean squared of the independent standard deviations calculated for the Freedom dataset varied marginally, from 0.3000 to 0.3003 as dataset completeness decreased. This is because the standard deviation component was dominated by the RUF standard deviation (0.300), which is a magnitude of 10 greater than those calculated from the completeness analysis, and a magnitude of 100 greater than those calculated from the accuracy analysis. The Multiplier component ranged from 1.43 to 1.79 for the Freedom dataset, and was fixed at 1.18 and 1.91 for the Electric Nation and Sustain-H datasets respectively.

Completeness (%)	Freedom		Electric Nation		Sustain-H	
Confidence	**P50**	**P95**	**P50**	**P95**	**P50**	**P95**
100	1.43	1.92	1.18	1.53	1.91	4.36
98	1.46	1.95	-	-		
95	1.51	2.00	-	-		
90	1.59	2.08	-	-		
80	1.79	2.28	-	-		

TABLE 12: TABLE OF DQFS FOR ELECTRIC NATION AND FREEDOM DATASETS

Each DQF in Table 12 provides a factor to multiply to the measured half hour period to provide the likely actual peak demand during that half hour period. Factors have been provided for different levels of confidence to take into account the uncertainty. Therefore for the Freedom dataset, with 100% data completeness, the 1 minute peak demand in a half hour settlement period was 1.43 bigger than the value for the half hour for a 50% confidence level. If a higher confidence level is required, such as 95%, then the factor increases to 1.92. For the Electric Nation dataset the values were lower, providing a 50% confidence level factor of 1.18 and a 95% confidence level of 1.53. The Sustain-H dataset values were substantially higher than the Freedom and Electric Nation results, which is likely driven by the small portfolio size.

5. CONCLUSIONS

The analysis revealed a series of interesting results, summarised below:

1. **The limitations of available datasets pose a *substantial* challenge for reaching conclusions – emphasising the pressing need for WPD to gather more data in future**. Everoze's analysis was materially affected by the limitations of the Freedom, Electric Nation and Sustain-H trial datasets. Most notably, key issues included the short duration of the datasets used, the Electric Nation dataset not being a minutely dataset, and the lack of large portfolio datasets of minute granularity from the Sustain-H trial. As such, Everoze strongly advocates seizing future opportunities to secure further domestic flexibility data, to build on these learnings and appropriately factor in the meter data uncertainties for Sustain-H remuneration in the long term.
2. **Data resolution has the biggest impact on demand uncertainty, followed by Completeness, followed by Accuracy.** The standard deviation component of the Data Quality Factor is dominated by the Resolution Uncertainty Factor standard deviation (0.3), which is a magnitude of 10 greater than those calculated from the completeness analysis, and a magnitude of 100 greater than those calculated from the accuracy analysis. In short, lower resolution half-hourly data *substantially* reduces the confidence WPD can have in the ultimate peak demand compared to minutely resolution data. Meanwhile, completeness is of medium importance.
3. **Results vary significantly by dataset – hinting at a possible need for a technology-specific approach to analysing data resolution:** Consumption profiles vary dramatically between the Freedom data (for heat pumps), and the Electric Nation and Sustain-H datasets (for electric vehicles); this applies both across the day and within individual half hour settlement periods. This has implications for the Resolution Uncertainty Factor (RUF), suggesting that it may be more appropriate to derive a separate resolution factor per technology. It is further possible that other assets (such as batteries) will show different behaviours again. At this stage, it is difficult to make firm conclusions on the technology-specific attributes when the data analysed are from trials with different interventions which have different impact on demand patterns. Moreover, as noted below in point 6, any technology-specific attributes may diminish for large portfolio sizes. In any case, more data are required to repeat the resolution analysis, should appropriate data become available, to draw firm conclusions on the technology-specific impact on data resolution.
4. **For heat pumps at half-hourly resolution, the Data Quality Factor for a complete dataset was 1.43 for a 50% confidence level and 1.92 for a confidence level of 95%.** This means that we are 50% certain the peak demand recorded for a half-hourly metered heat pump portfolio is 1.43 times higher when assuming minutely resolution.
5. **For electric vehicles at half-hourly resolution, the Data Quality Factor (DQF) for a complete dataset (using Electric Nation data) was 1.18 for a confidence level of 50% and 1.53 for a confidence level of 95%.** This means we are 50% certain the peak demand recorded for a half-hourly metered electric vehicle is 1.18 times higher when assuming minutely resolution. For the Sustain-H dataset, the DQF is 1.91 and 4.36 for the 50% and 95% confidence levels, respectively, which is significantly higher than the findings using the Electric Nation dataset. The material difference in the findings using the two datasets is likely due to: (1) the constant power charging assumption made when processing the Electric Nation data, (2) the flexibility provider's interventions in the Sustain-H dataset where the EV chargers were responding to price signals on a minute-by-minute basis, and (3) low demand outside of 12am-7am in the Sustain-H dataset resulting in the peak-over-mean calculated to be statistically skewed due to the low half-hourly average demand during these periods.
6. **Analysis to date suggests that portfolio size strongly impacts the half-hourly Resolution UF, and consequently the Data Quality Factor.** Everoze repeated the resolution analysis on reduced population samples for both the Electric Nation and Freedom datasets. Due to the small portfolio size, the Sustain-H dataset was not included in this analysis. The smaller population datasets yielded larger RUFs. Interestingly, for a fixed portfolio size, the analysis yielded different results for the two datasets, which implies that technology type may also be a driver of variation in the Data Quality Factor for small portfolio sizes. The impact of portfolio size and technology type on the calculated Data Quality Factor (predominantly driven by the RUF), appears to diminish for large portfolios (> 100 assets). This convergence, or asymptotic behaviour, with increasingly large portfolios will have an impact in the design of the DQF/DDF where a simplified approach may be justified.
7. **The linear impact of data incompleteness means that WPD can take a pragmatic approach.** In Everoze's analysis, demand uncertainty increased in a broadly linear fashion as the dataset became more incomplete. For instance, an 80% complete dataset led to a 125% increase in demand uncertainty when considering a 50% confidence level. As a result, there is potential for WPD to adopt a pragmatic approach here; for instance, for a 50% confidence level an 80% complete dataset might result in a multiplier of 1.25 within the Data Quality Factor, with corresponding impact on payment. Everoze also investigated the impact portfolio size has on demand uncertainty for an 80% incomplete dataset, and found the impact is independent of portfolio size.

5.1 CONCLUSIONS

Resolution

RECAP OF APPROACH: Everoze determined the peak/mean ratio for the Electric Nation, Freedom and Sustain-H datasets, in order to understand the extent to which peak demand exceeds half hourly average demand.

- <u>The 'typical' peak/mean ratio for resolution are around 1.2 to 1.6 – but generalisation depends on the required confidence level:</u> It was difficult to derive a single value for a "typical" peak/mean ratio, however this value is likely to be in the range of 1.2 to 1.6. If WPD need more certainty, such as a confidence level of 95%, then values of 1.8 and above need to be considered. In addition, note that as consumption is reduced, which is the behaviour WPD want to motivate and reward consumers to do, peak/mean ratios increase.
- <u>There is no trend for morning or evening delivery periods – suggesting that daily averages are appropriate:</u> While there is some variation in peak/mean ratio throughout the day there was no obvious trend for the morning or evening delivery periods. Therefore as there are limited datasets available for this work it is most appropriate to use whole day averages at this stage.
- <u>Population size does impact the peak/mean ratio of the dataset significantly.</u> Analysis to date suggests the magnitude of the impact is characteristic to the different technologies, however for population sizes above 100, the peak/mean ratio converges to 1.2/1.3 irrespective of technology type.
- <u>The distributions of the peak/mean ratios for the electric vehicle and heat pump datasets are very different – pointing to a need for a technology-specific approach to resolution factor and consideration of the asset's operating behaviour</u>. The Freedom data is characterised by Weibull distributions whereas it is exponentially decreasing for Electric Nation and Sustain-H datasets. The differences can be explained by the typical behaviour of the types of asset within each dataset. By their very nature, heat pump power consumption can vary a reasonable amount over short time frames. The Freedom hybrid heat pumps under full optimisation would have been quite variable, whereas with less stringent control, the power demand profiles may not vary significantly. The Electric Nation EV chargers were likely to have been on for longer periods of time and were more binary in their operation due to Everoze's processing assumption, that charging was constant during each charging 'transaction'. This means that for electric vehicles with dumb charging, the peak for many settlement periods, may well be the same as the mean. The Sustain-H dataset behaved differently again. The peak/mean ratio rarely dropped below 1.5, even during the evening delivery window when all assets were required to drop their demand to a pre-agreed level. Portfolio size could explain some of the disparity between the Electric Nation and Sustain-H datasets, however the main driver was likely the assumption made when processing the Electric Nation data which effectively assumed they were 'dumb' chargers. The Sustain-H participants' chargers were 'smart' and optimised for imbalance price on a minute-by-minute basis, causing the minutely peak demand to be significantly greater than the settlement mean.

Completeness

RECAP OF APPROACH: The completeness analyses quantified the demand uncertainty introduced for different levels of incompleteness, when compared to a 100 % complete dataset. Everoze removed whole-day assets randomly to reduce the 100 % complete dataset to 98, 95, 90 or 80 % completeness, and calculated the ratio of the 100 % complete whole-dataset power demand (kW) against the reduced dataset (kW) demand. This was iterated using a Monte-Carlo approach 1000 or 10,000 times, and the mean from each iteration was plot on a frequency distribution to capture the mean demand uncertainty and standard deviation for each completeness distribution.

The unavailability of data within a dataset is randomly distributed. For the Freedom dataset it was also apparent that when an asset is unavailable it was for the whole day.

- <u>Demand uncertainty increased in a broadly linear fashion as the dataset became more incomplete</u>. As might be expected, the mean of each frequency distribution plotted using half hourly data from each settlement period, is approximately equal to the ratio: 100% divided by the reduced dataset completeness percentage (e.g. 100/80 or 1.25).

Accuracy

RECAP OF APPROACH: The meter accuracy as defined by the accuracy measurement given in the available published information (from datasheets and Sustain-H trial participant data) has been taken as the standard deviation for this work.

- Measurement accuracy has minimal impact on the measured demand. For a sample size of 100 and 1 % metering accuracy, the standard deviation calculated is 0.1%. Residential smart meters and asset meters are typically accurate to +/- 1 %.
- The multiplier for the Accuracy Uncertainty Factor should be 1: It has been concluded that the multiplier for Accuracy is 1.0 as there is no expected bias in meter readings, only uncertainty from the meter accuracy.

5.2 LIMITATIONS

1. Dataset limitations: The Electric Nation dataset had inherent limitations because the dataset was not 1-minute resolution, rather each row in the dataset captured a single 'charging event', which included a start and stop time, and car battery charging rate (kW). Using the available data, Everoze re-created the dataset to be of minutely resolution, which likely introduced uncertainty in the demand. This, coupled with the comparatively low percentage of data over the 3 month sampling period which was not 0 kW (approximately 5 %), meant that the dataset was unlikely to produce valid results for the completeness analysis. Consequently, the completeness UF was only calculated using the Freedom dataset. The Freedom dataset had much higher availability, however the sampling period considered capable of producing valid results was only 2 months' worth of data. Sustain-H data received from the trial was not of the quantity expected. Participants who did provide asset meter data of minutely granularity, which was not a requirement but 'nice to have', were small in portfolio size. Portfolios > 10, and ideally > 100 were essential for the resolution analysis. Only one participant provided a dataset of EV chargepoints with 19 assets in the portfolio.

2. Absence of Completeness UF for Electric Nation and Sustain-H datasets: Due to the above limitation, the DQF calculated for the Electric Nation dataset and Sustain-H dataset does not include a completeness UF. Findings from the completeness UF calculated using the Freedom dataset were conclusive, and considering the very limited number of assets, Everoze did not use the Sustain-H trial datasets to re-calculate the completeness UF during phase 2 of the analysis.

3. Delivery period analysis trends lacking: Everoze found that, while there were some variation in peak/mean ratios throughout the day identified from the resolution analysis, there was no obvious trend for the morning or evening delivery periods. Data from the Sustain-H trial used to re-calculate the RUF were from one participant who provided a service during the winter evening delivery period only. Therefore as there were limited datasets available for this work, Everoze assumed it was best to use the whole day averages when quantifying the UFs at this stage.

4. Battery assets missing: Whilst comparison has been made between the results from the Freedom, Electric Nation and Sustain-H (EV only) datasets, other types of asset (such as domestic batteries) may show different behaviour again. Also, all the combinations of different assets will likely show different behaviours.

5. Portfolio effects: The calculated UFs at this stage have not considered the impact portfolio size might have on the factors. However, the Electric Nation dataset has approximately 3x the number of assets compared to the Freedom dataset, so the DQFs are likely driven by sample size. Sample sizes were going to be a key consideration in phase 2 of the analysis, however datasets received from the Sustain-H trial were not large enough for this to be a consideration in phase 2.

6. Flex actions within datasets: The Freedom assets experience multiple interventions during the trial duration, and these interventions (predominantly price signals) already introduce a 'flex action' which could drive uncertainty in the UF calculations. The Electric Nation assets also experienced interventions during the trial

which limited the charging rate for groups of cars in localised areas, to reduce network constraint. However the magnitude of the reduction was not known. Therefore Everoze corrected these interventions so removed the 'flex action' and re-calculated the demand as if not managed.

5.3 RECOMMENDATIONS FOR SUSTAIN-H SERVICE UNDER BAU

We recommend that no derating factor for data resolution is used for initial BaU roll-out. Any derating factor will have a substantial impact on remuneration for service providers. Given that low value has already been highlighted as a key risk for transitioning Sustain-H from trial to BaU procurement, we recommend no derating is applied for BaU roll-out at least in the medium term until sufficient liquidity is established.

However, in the long-run there are benefits to WPD in having higher resolution data. Higher resolution meter data means reduced uncertainties for the DSO in making network planning and flexibility decisions. Findings from this analysis highlighted that using half hourly meter data reduces confidence WPD can have in the ultimate peak demand compared to 1-minute demand data. The findings showed the 1-min peak demand can be as high as ~1.2 times the half hourly meter demand for a 50% confidence level for portfolios > 100 homes.

WPD should conduct further analysis to refine the Resolution UF analysis using high-resolution meter data across various technologies. The findings from the analysis showed the RUF tended to converge to 1.2 irrespective of technology type as portfolio sizes exceeded 100 homes. However, this is not a robust conclusion due to the limitations of the data used in the analysis. Should more data become available in the future, the analysis should be repeated for a range of technologies and larger datasets to verify the findings. Notably, the further analysis should test the following: (i) whether the asymptotic behaviour and convergence to a single RUF for large portfolios holds true across technology types and different datasets, and (ii) what value of RUF the results converge to. This will provide a robust evidence base for introducing derating factors in the Sustain-H payment calculations in the future.

As Sustain-H matures and further data is collated, WPD should consider introducing derating factors in future to incentivize higher resolution meter data. Given the benefits of higher resolution meter data, WPD should clearly signal its longer term direction of travel as one where it desires higher resolution meter data and encouraging service providers to provide higher resolution meter data where they are able to. WPD may wish to consider a suitable mechanism (using derating factors based on the Data Quality Factor) for the Sustain-H service to incentivize higher resolution data, and that any such mechanism is only introduced once sufficient liquidity is established. As Sustain-H participation in the long term will likely be large scale domestic portfolios, and as impact of small portfolios of homes and assets on WPD's network is expected to be minimal, a simplified approach focusing on the impact of the larger portfolios (> 100 homes) is justified.

We recommend a new availability metric is included in the Sustain-H performance calculations. Missing meter data from one or more assets reduces the certainty to WPD of the overall portfolio demand and in turn devalues the service to WPD. Findings from this analysis found that demand uncertainty increased in a broadly linear fashion as the dataset became more incomplete. An availability mechanism, where availability is calculated as a measure of completeness of the meter data provided over the contracted delivery periods, tied with service remuneration is one method to incentivize data completeness. For such a mechanism to work, service providers will need to provide individual home/asset meter data. To keep data requirements simple, Sustain-H BaU roll-out only requires service providers to submit aggregated portfolio data, and so WPD does not measure and reward asset availability as part of BaU roll-out. We recommend WPD considers introducing an availability mechanism in the performance assessment and payment calculations in the medium to long term to incentivize higher completeness in the meter data submitted.

Flexibility providers will need to use at least MID-compliant metering solutions. Findings from the phase 1 analysis shows that MID-compliant metering solutions (with up to 2% accuracy) have an acceptable level of uncertainty and so is suitable for Sustain-H. We recommend WPD be flexible with metering solutions used for Sustain-H and accept MID-compliant metering solutions. The recent PAS 1878 standard for energy smart appliances sets out metering accuracy to not exceed 10%. This is substantially less accurate than MID-compliant metering solutions. If WPD accepts the prescribed requirements for domestic flexibility as set out in PAS 1878, we recommend WPD evaluates the Accuracy UF for these metering solutions and its overall impact on the DQF.

6. REFERENCES

[1] Everoze, WESTERN002-AD-R-01-B, "FutureFlex: Supplier Trial – Data Review", Revision B, 29th May 2020

[2] Everoze, WESTERN002-AD-R-02-B, "Aggregated Datasets - Methodology", Revision B, 10th September 2020

[3] Passiv Systems "Freedom: Trial Data Description", September 2020

[4] Western Power Distribution, Electric Nation, "Electric Nation Customer Trial - CrowdCharge Transaction Data", 24th October 2019

[5] GOV UK, Office for Product Safety & Standards, "In-Service Testing (IST) Handbook", February 2020.

[6] GOV UK, "Schedule 4: UK nationally approved electricity meters", January 2020.

[7] PassivSystems, FREEDOM, "FREEDOM Trial Data", 18th April 2018

7. APPENDIX 1: PRE-PROCESSING

7.1 FREEDOM DATA

7.1.1 Pre-processing

Freedom Data Format Summary

Asset meter data from the Freedom trial was provided in minutely and half hourly format for each asset as independent data files. Three types of heat pumps were included in the Freedom trial: MasterTherm, Samsung and Dakin heat pumps. The minutely trial data description document [7] included the following columns:

- *heat pump power consumption*: measurements of instantaneous electrical power consumption in kW
- *heat pump reactive power consumption*: measurements of instantaneous (electrical) reactive power in kW
- *heat meter overall power*: measurements of instantaneous heat pump heat production in kW
- *heat pump heating control*: on/off events sent to MasterTherm/Samsung heat pumps (not used for Daikin)
- *1minute timestamps*

The half hourly power consumption data files were more extensive, providing information on the occurrence and types of faults present in the data. The Freedom trial datasheet indicated that **it is essential that the main data set (30 min resolution) is cross-referenced with the 1 minutely dataset to exclude fault periods.**

Freedom Pre-Processing:

1. Loaded file data paths to Python Pandas**.**
2. Read the 30 min resolution files and identified any errors when opening files. Only error was due to one 30 min dataframe which was empty and couldn't be opened. Removed the asset file, which reduced the sample by 1 (to 74)
3. Filtered through the opened 30 min res files to find which have 'System Faults'. Found that all remaining 30 min res files contain System Faults. Also found some datasets contained System Faults 100 % of the time. Removed files (assets) which were 100 % incomplete (64 / 75 of the 30 min data files remain).
4. Cross referenced the 1 min timestamps against the 30 min timestamps**.** This was done to identify when the 1 minute data (which did not have a column to indicate faults present) fell within a System Fault period in the 30 min dataset. Within the 30 min dataset, [7] had removed all data in the rows (including the timestamp) during fault periods. Therefore, if the 30 min timestamp from the 1 min datafile was also present in the 30 min datafile, then there was not a system fault. If the timestamp was not present in the 30 min datafile, then the 1-minute data electricity demand (kW) fell within a fault period, and these data points were converted to 'null'.
5. Calculated the % completeness of the 1 minute dataframe for each asset. Identified any files < 100 % complete following step 4.
6. For the 1 min processed data files, extracted the heat pump power consumption (kW), home ID and 1 minute timestamp.
7. Combined the files from step 6 into a single file, indexed by timestamp, 1 column per asset to create a portfolio of electricity demand data (kW)
8. Plot the availability of the datafile from step 7 against the 'date' to determine the time periods when the dataset was most available (i.e. low % of null datapoints and high participants). Found lots of Null data points during the start of the trial. Best data availability is from 1st Mar to 30th April.
9. Re-sampled the dataset to only include time stamps from 1st Mar to 30th Apr. Checked the availability of assets during this period, and removed any assets if the asset availability was < 50 %. This reduced the sample size to 49 assets. Saved the file ready to upload for the completeness and resolution analysis.

7.2 ELECTRIC NATION DATA

7.2.1 Pre-Processing

Electric Nation Data Format Summary

Each 'transaction' i.e. charging event is provided as a single row in the Electric Nation data file. The columns provide information on the start charging (ActiveCharging_Start) and end charging (EndCharge) time, and the total electrical demand during the charging event (Consumed kWh). Therefore the energy consumed during the charging period is a minutely average. Ideally, the data would be minutely 'actual' power consumption rather than an average demand during charging period.

Electric Nation Data Pre-Processing

Data pre-processing was undertaken in 8 stages:

1. Removed data points whereby the consumed kWh was greater than 100 kWh. This is a result of inaccuracies in the data. The largest battery capacity participating in the project was 100 kWh, therefore any charging events greater than this were an anomaly.
    a. Events where the ConsumedkWh < 0.5 were kept in despite the total charge being marginal, because this indicated that the car was plugged in and available for DSO services even if the car was charged.
2. Removed any data where Participant ID = Null. In these instances, the participant could not be linked to a charger which is required to quantify the total number of participants charging on any one date.
3. Removed instances where ActiveCharging_Start = Null. This was caused due to internet connectivity issues.
4. Removed instances when EndCharge = Null. This was always the case if ActiveCharging_Start = Null because the EndCharge time could not be approximated. Additionally this was caused if the internet was down during the end of the charging period.
5. Checked CrowdCharge note (incorrectly calculated EndCharge). CrowdCharge noted that there are instances when the EndCharge time was not calculated correctly in the datasheet. EndCharge was estimated (by CrowdCharge analysts) as the time when charging ended based on the meter value data (looking for sustained reductions in the current being drawn in the absence of demand management). Some of these calculations were inaccurate because the total consumed kWh over the charging duration was greater than technically possible given the charger capacity. For example, for a single transaction using a 7kW charger, if ActiveCharging_Start was 17:00 and EndCharge was 18:30, but 21kWh of energy was transferred then it's likely that EndCharge is inaccurate (because it isn't possible to add 21kWh in 1.5 hours using a 7kW charger).
    a. Everoze performed the following to check if the 'Charge time' was too short:
        i. ChargeTime = EndCharge – ActiveCharging_Start
        ii. ChargingRate = ConsumedkWh/ChargeTime
        iii. If ChargingRate>Battery Rating, then the end charging time is incorrect.
        iv. There are 379 transactions where this was the case, which is 0.53% of the data.

        To overcome this error, the EndCharge time was re-calculated for the rows where the above formula was true by the following:

        v. NewChargeDuration = ConsumedkWh/CarkW(rating of the participant's vehicle, either 3.6 kW of 7 kW which is provided per row).
        vi. NewEndCharge = ActiveChargingStart + NewChargeDuration

        Assumption: that the rate of charge was constant at 7kWh or 3.6 kWh throughout the charging period.
6. Corrected for those with managed charging, by re-calculating the charging duration as if not managed by assuming a constant charging rate equal to the maximum battery rating, because information was not available on the level of demand management during the trial / the total energy demand per minute when managed.

    The participants with chargers rated 3.6kW (16 A), who's chargers were controlled, control when the cars where charging regularly (because of their dual-fuel tariff) were rarely managed. Whereas those with 7 kW chargers (32A) and who used their car to charge during peaks times are more often than not managed. Any events where a start and end charge time cannot be determined due to an absence of meter values were not managed. Trials 1,2 were managed in some instances and the Trial 3 transactions were effectively 'managed' because of the CrowdCharge algorithm used alongside the ToU tariff, therefore there are no T3 Managed columns.

7. Removed any assets where the charging duration was longer than battery charging capabilities, e.g. 21 kW of charge was transferred in 1 minute (which was not possible because the chargers were rated at 3.6 or 7 kW). 3771 transactions were removed.
8. Set up the 'transaction' data remaining following step 7 into a aggregated1-minute timeseries portfolio, Everoze assumed a constant charging rate for each minute, for each charging event (asset). The charging rates were either 3.6 kW or 7 kW. Set all null data as 0kW (e.g. when the car was plugged in and not charging) and saved the file ready to upload for the resolution analysis.

Table 13 captures the number (and percentage) of transactions removed from the input Electric Nation data file during pre-processing stages 1 to 7. Stages 5 and 6 are not included because data was not removed from the input file, instead the EndCharge times were corrected and replaced.

Stage	Transactions Removed from Input File	Cumulative % Removed
1	21	0.03
2	3899	5.47
3	19707	27.65
4	20855	29.26
7	24626	34.60

TABLE 13: ELECTRIC NATION DATA PRE-PROCESSING: TRANSACTIONS REMOVED DURING EACH STAGE LISTED IN SECTION 7.2

7.2.2 Results from pre-processing

Factors which influenced the decision to use the Electric Nation (CrowdCharge) dataset included:

1. 65.4 % of the initial dataset (71,264 transaction) remained after correcting for the null data, meaning the data remaining was still substantive.
2. There were three trials conducted as part of the Electric Nation project. Trial 1 featured minimal interventions, meaning that the 'trial' data is close to a 'real-world' scenario and therefore the impact of trial interventions on the UF calculations are low if using Trial 1 data.
3. The RUF calculations, can be performed on the data remaining. The results might be slightly inaccurate because the trial data is not 'real-world', however, the UFs (and subsequently DQF) will be updated following retrieval of Sustain-H trial datasets.

7.3 SUSTAIN-H DATA

7.3.1 Pre-processing

Participant 2's winter dataset was used to conduct the refined analysis because this dataset was the only one received from trial participants with more than 10 assets.

Participant 2 provided 30min frequency data in .csv format for the winter months of the trial: November, December, January and February. Datasets were indexed by date and regular 30 min timestamps.

Columns included:

1. 'device type' (v2g or smart charger),
2. Average import power (kW) (i.e. power imported from the grid (+ve))
3. Average export power (kW) (power exported to the grid (-ve))
4. Average net imported power (i.e. import - export) (kW) and;
5. Max power (kW) (i.e. maximum power demand the device reached during each half hour period).

Everoze took the following steps to process the data for the resolution analysis:

1. Filtered the columns by device type: Everoze removed the four v2g chargers and retained 19 smart chargepoints. This was because the smart chargepoints only import power, whereas v2g chargers behave akin to batteries. Everoze removed the v2g chargers to allow for direct comparison with the Electric Nation dataset, which also used smart chargepoints.
2. For each settlement period, Everoze summed a) the max power demand to give the total peak demand during each half hour period (kW) and b) the average import power (kW) to give the total average imported power demand during each half hour period (kW).
3. For each settlement period, Everoze calculated the peak / mean using the output from 2a / 2b, and removed any values which were <1. This is because the ratio should always be at least equal to or greater than 1. This removed 8% of data points.

Portfolio size was a significant limitation and therefore it was not possible to investigate the portfolio scale impact of the RUF for the second phase of the analysis.

8. APPENDIX 2: UPDATED METHODOLOGY

8.1 RESOLUTION

8.1.1 Determine the variance of the data at 1-minutely intervals and when averaged into half hourly intervals for a portfolio.

Following pre-processing, it was found that the Electric Nation data could be analysed at a 1-minute resolution (with some pre-processing assumptions, as described in Section 7.2.2). The Freedom data was also made available at 30-minute and 1-minute resolution, provided by PassivSystems. Therefore Everoze computed the Resolution analysis using both Electric Nation, Sustain-H and Freedom datasets.

Analysis steps which have been conducted by Everoze on the trial datasets are presented below:

Step 1: For a fixed population size within a defined period (maximising the size and period of the dataset as appropriate), Everoze summed the minutely portfolio electricity demand. Everoze produced an output of the net portfolio demand (in kW) for each minute over the sampling period. The population size over the selected sampling duration was fixed for both datasets.

Step 2: Everoze quantified how the ratio of peak-1-minute / 30-minute mean varies for each HH settlement period (48 in total), for the dataset as a whole, and also for each Delivery Period in the Sustain-H trial. The following steps were taken to analyse the (peak-1-minute/30-minute mean):

a. Everoze calculated the mean (30-min) demand for each HH period within the time frame considered[1].
b. For each 30-min interval across the timeframe considered, Everoze calculated the (1-min peak / 30-min mean) ratio.
c. For each half hourly settlement period (48 in total), the frequency distribution of the (peak-1-minute/30-minute mean) ratios was derived using the output from b, to produce 48 frequency distributions. Everoze found that the frequency distributions varied considerably across neighbouring settlement periods. Therefore, Everoze combined the peak/mean ratios for the morning Delivery Period, evening Delivery Period and for the whole dataset, and re-plot the frequency distributions to detect dataset trends.
d. For each of the frequency distributions created from step 2.c, Everoze derived the following values and wrote the values in a look up table:
    1. Mean (peak-1-minute/30-minute mean) ratio;
    2. Maximum;
    3. Minimum; and
    4. Standard Deviation.
    5. P95 exceedance cases.

Step 3: Everoze extracted the "multiplier" (in this case the Mean (peak-1-minute/30-minute mean) ratio) and corresponding standard deviations from the whole dataset frequency distribution for the DQF calculation**.**

[1] Implicit assumption: for fixed population size, the variance in completeness of the dataset is small so that there is nedligible noise within the results. Using a dataset close to 100% completeness will improve the validity of this assumption.

8.1.2 Quantify the impact portfolio size has on half hourly averaging uncertainty

Step 1: Everoze randomly selected a sub-group of the data population from those analysed in Section 8.1.1 to create reduced population sizes (e.g. reduced 100 participants to 25 and 50 for further analysis).

Step 2: Repeat Steps 1 to 3 from Section 8.1.1 on the different population sizes to assess how the results vary.

Note – this was only calculated using the Electric Nation and Freedom datasets.

8.1.3 Estimate value of "RUFs"

Through analysis of the findings in the analysis described above, Everoze estimated the value of "RUFs" for each dataset for the morning Delivery Period, evening Delivery Period and whole dataset.

8.2 COMPLETENESS

8.2.1 Quantify the 'completeness' of the dataset for the different portfolios considered

Step 1: Everoze quantified the percentage completeness for the 1-minute dataset.

Step 2: Everoze created a 30 minute availability datasets from the 1-minute asset meter dataset. Half hour availability was calculated for each asset, by counting the number of non-null minutes within each half hour. If > 25 % of the minutes were null values for each asset, the availability for that asset-half hour was marked as '0'. All other half hour periods were marked as 1. Using this output, Everoze quantified the number of assets operating for each half hour period.

Step 3: Everoze plot the output from Step 2, the dataset availability, as a frequency distribution across the fixed period.

Step 4: Everoze visually interpreted the output from Step 3 with the intention of fitting a simple mathematical distribution to the probability distribution (e.g. a Weibull curve). After review of the plots, Everoze decided that the distribution did not conform to a typical distribution. This had directly implications on the Method for introducing null values to artificially introduce incompleteness into the dataset.

8.2.2 For a fixed portfolio size, determine the impact 98, 95, 90 and 80 % portfolio data completeness has on measured demand uncertainty

Potential invalid datapoints in a HH metered dataset (for example) may take the following forms:

1. An obviously invalid value (e.g. outside conceivable range)
2. Null value
3. Zero value (could or could not be valid)
4. A realistic value which may not capture the full duration of that time period (e.g. only 15 minutes of data in a 30 minute period)

In order to estimate the impact of Completeness on uncertainty, it is necessary to artificially introduce "incompleteness" into a complete dataset and quantify the difference in measured demand for varying levels of completeness. This requires the starting dataset to be 100% complete before introducing null (or back filled) values randomly. When aggregated over a portfolio, this could have the effect of artificially lowering the portfolio metered value for the affected periods, where the 'real' portfolio demand would be higher than this metered value.

Everoze conducted the completeness analysis using only the Freedom dataset. This decision was made due to reasons described in Section 3.2.

Steps 2 to 6 set out steps Everoze carried out to quantify the percentage uncertainty for different completeness and population size scenarios. Steps 2 to 6 introduced incompleteness as null values.

Step 1: Everoze set up the starting dataset to be 100 % complete so that random incompleteness can be introduced artificially. The completeness was introduced using the following approach, in order of priority:

a. Take the average for asset with identified invalid data across the same hour of either weekday or weekend across the duration of the dataset.

b. If the data available to perform the averaging process in item 1 is < 75%, then: Everoze took a through day average for that asset [split between weekdays and weekends]

c. If the data available to perform the averaging process in item 2 is <75%, then Everoze took an average across all assets for that time of day.

Step 2: For a fixed population size, Everoze summed the 30 minute average electricity demand for each 30 minute interval across the timeframe considered. The output would be the total demand for each 30 minute interval.

Step 3: Everoze randomly introduce an average completeness (e.g. 98%) into the dataset by allocating 2% of electricity demand readings as null values. In the previous deliverable, it was noted that the null values would be allocated using a skewed-normal (Weibull) distribution which had been derived from, and checked against, appropriate data. However, upon the analysis described in section 1, the distribution of the data did not followed a skewed-normal distribution, or any other distribution. Therefore, rather than applying a distribution to the data, whole-day assets were removed from the dataset, to reduce the overall availability to the desired completeness percentage e.g. to 98%.

Step 4: Everoze calculated the mean electricity demand (kW) output from step 2, and the mean electricity demand (kW) output from step 3. Everoze then divided the two to give a demand difference for the dataset as a whole (the numerator was always 100% complete 30-minute demand, therefore by dividing the 100 % complete 30 minute demand by the 98 % complete 30-minute demand, the ratio was than 1, or equal to 1, depending on the random allocation of incompleteness in the data).

Step 5: Everoze repeated Steps 3 to 4 for 1000 iterations, for different cases of random introduction of incompleteness, in the form of a Monte Carlo analysis. 10,000 iterations were performed for the 80 % completeness case for comparison.

Step 6: Everoze created frequency distributions from Step 5 and captured the following:

1. Mean;
2. Maximum;
3. Minimum; and
4. Standard Deviation.
5. P95

Step 7: Everoze extracted the “multiplier” (100 / incompleteness (e.g. 98)) and corresponding standard deviations for each of the frequency distributions from Step 6 for the DQF calculation**.**

8.2.3 Quantify the impact portfolio size has on measured demand uncertainty for the different completeness cases

Everoze investigated the impact portfolio size has on demand uncertainty for the 80% completeness case, by repeating the analysis from Section 4.2.1 on the Freedom dataset, using a 50 % reduced population.

To approach this, Everoze randomly removed 50 % of the assets from the complete dataset, reducing the total number of assets from 49 to 24. Next, 20 % of the ‘whole-day’ assets from the remaining asset population were removed, reducing the completeness to 80 %. This process was repeated for 100 iterations, and the mean of the 100 % complete dataset demand (kW) / 80 % complete dataset demand (kW) was calculated for each iteration.

8.2.4 Estimate values of “CUFs”

Through analysis of the findings in the analysis described above, Everoze estimated the “CUFs” for the Freedom dataset.

8.3 ACCURACY

There are three key parts to the analysis:

- Quantify the 'Accuracy' of different meter types;
- Calculate the standard deviation of this uncertainty; and
- Estimating "AUFs" for the Data Quality Factor equation.

These are each described below.

8.3.1 Quantify the uncertainty introduced to the metered data attributable to the metering accuracy of half hourly smart meters (SMETS1 and SMETS2 meters) and the asset level meters used by participants for the trial.

Everoze reviewed metering standards for residential electricity meters, and retrieved accuracy meter measurements from two Sustain-H participants, to determine appropriate parameters for metering accuracy.

Everoze undertook a broad sensitivity analysis to examine the consequences of varying meter accuracy on overall Data Quality for varying portfolio sizes.

8.3.2 Calculate the standard deviation of the accuracy measurements

The standard deviation of the combined accuracy measurements were assessed based on available information.

8.3.3 Estimate the values of "AUFs"

Through analysis of the findings in the assessment described above, Everoze defined the range of "AUFs" for a two accuracy meter values, and a range of portfolio sizes.