Electricity
Distribution

Smart Meter Innovations and Test Network (SMITN)

EARLY LEARNING WORKSHOP

29/03/2023

10:00 – 11:30

nationalgrid

Contents

1

Overview

Smart Meter Innovations and Test Network

NIA funded

Establishes a validated test network

Four use cases:

- **Phase identification**
- **Planning profiles**
- **Feeder allocation**
- **LCT detection**

Smart meter data offers new opportunities to DNOs.

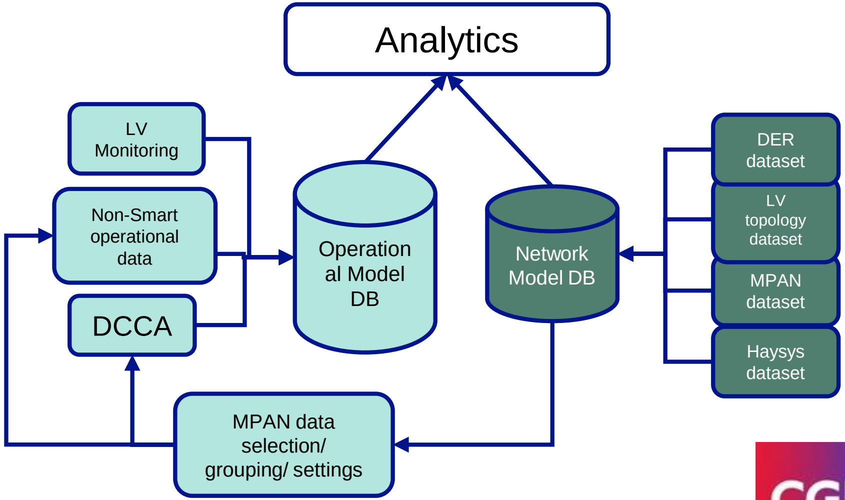
The SMITN project aim is to extract additional value from smart meter data by creating a validated LV network against which algorithms using smart meter data can be tested.

The project has also tested a new method to identify customer's LV feeder from outside their premises.

Partner Roles

CGI

Set up the data processing infrastructure, implemented the selected algorithms and evaluated the performance.



HAYSYS

Developed the new Feeder Finder and carried out surveys.

Loughborough University

Selected the Test Network, advised on the algorithms to be used and review the work produced by CGI and HAYSYS.

GHD

Provided additional quality validation

Data Sources

Smart meter data

- 1-minute and 30-minute voltages
- Monthly demand per MPAN
- Half-hourly demand per feeder
- Half-hourly exports per MPAN
- Max demands and alarms

Test Network

GridKey substation monitoring

- 1-minute voltage, current, power

Network and asset data

Electric Office - LV network GIS

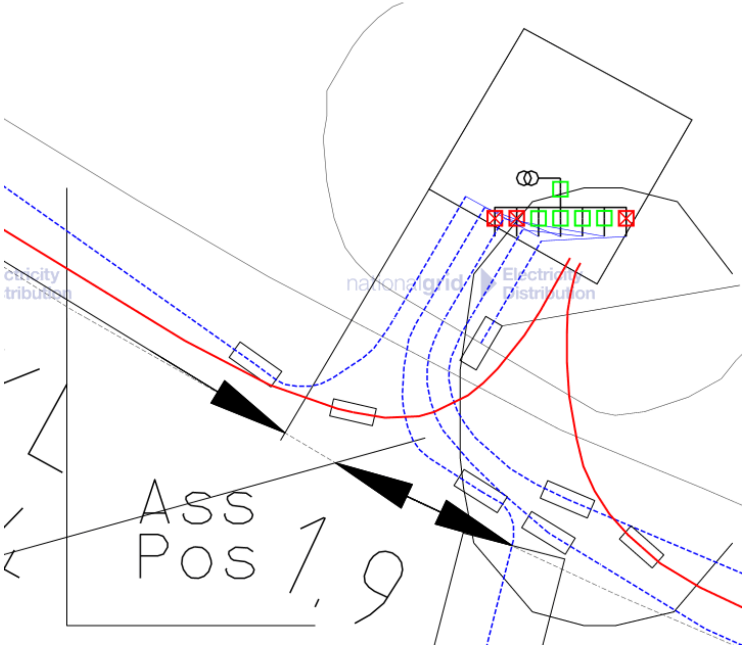
- Feeder mains
- Service cables and phases if known
- Customer locations

CROWN - asset and customer data

- Estimated Annual Consumption (EAC)
- Substation and feeder circuit ids
- Profile classes, time pattern regimes

Connectivity Uncertainties

- Feeder mains are well characterised in Electric Office GIS
- Service cables – some shown in Electric Office
- Customer phase connections – some shown in Electric Office but many unknown
- Customer feeder connections – recorded in CROWN asset data but some records are implausible given the locations of the feeder mains
- Numbering differences between EO and CROWN



Electricity Distribution

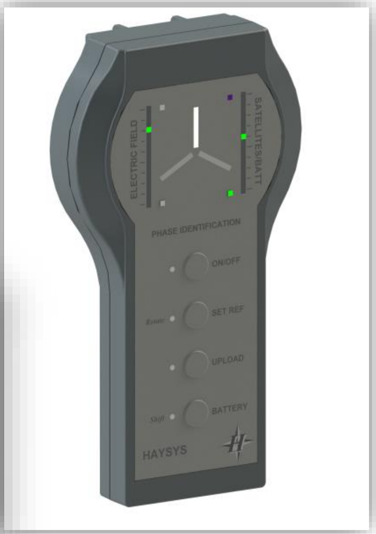
2

Test Network Validation

nationalgrid

HAYSYS Phase survey

- Survey of over 8,000 properties within the test network using the HAYSYS Phase Finder tool
- Photograph all of the meters seen at each of the properties
- Report on any LCT devices seen
- Undertaken in two sets
  - 1st July 2022 – 4th August 2022
  - 27th October 2022 – 3rd January 2023
- Provide phase dataset for validation



HAYSYS Feeder Finder

Development

- Development of a new tool that can identify the substation feeders
- Two parts
  - Signal Injector
  - Signal Detector
  - Feeder Finder HMI (Tablet)
- Injecting unique codes on 5 MHz carrier - to be detected outside properties
- Up to four codes injected on four feeders simultaneously

Survey

- Four unique codes injected into four feeders
- Walk the cable path with the detector showing the feeder codes
- Provide feeder dataset for validation

Electricity Distribution

3

Phase Identification

nationalgrid

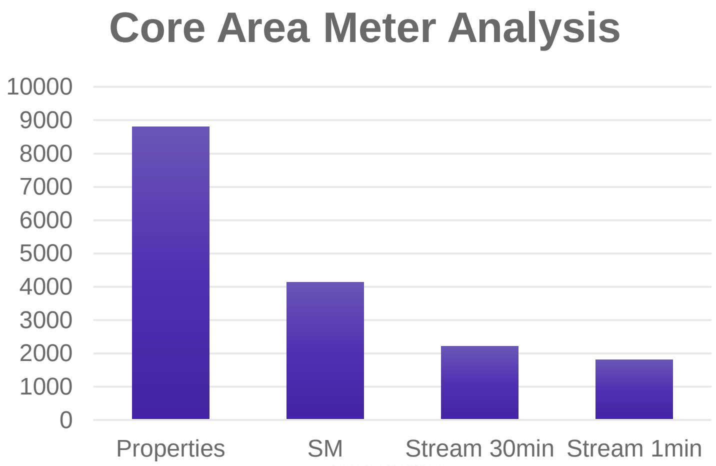
Phase Identification - Early Learning

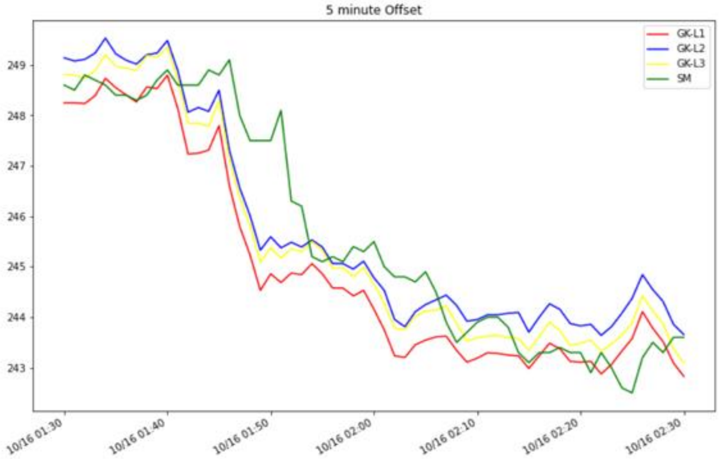
Test Network

- 47% have installed a SM
- 57% of SMs successfully responded to half-hourly voltage requests
- 82% was managed to re-configured to 1-minute

Exploratory Data Analysis – Data Issues

- Data anomalies
- Clock offsets
- Measurement interval synchronisation
  - Timestamp rounding
  - Linear interpolation





Approach A: Voltage Correlation between SMs and SS - Results

The correlation used time-series voltage data from a single-phase SM paired with SS monitoring data. For each single-phase SM, three correlation results are obtained, one for each phase measurement at the substation. The assigned phase was the phase with the highest correlation.

Experiments:

1. Time reference rounding vs. linear interpolation

- To address the clock synchronisation issue, two approaches have been developed, rounding and linear interpolation.
- No major differences were found between the two approaches.

2. Time-series voltage data vs time-series voltage step changes

- The use of time-series magnitude voltage data (time-series voltage) and the time-series of the voltage step changes (step changes) are compared.
- The use of step changes have a small positive impact in the correlation results.

Approach	Percentage of Substations (%)	Accuracy Score
A	60%	> 0.8
A	31%	0.5-0.8
A	9%	< 0.5

1116 out of 1347 devices labelled correctly based on HAYSYS validation data

Accuracy 0.83

Approach B: Clustering SMs into 3 groups

In the absence of monitoring data in the SS, the algorithm groups the SMs into three groups using unsupervised machine learning.

The limitation of this approach is the absence of the actual phase reference (L1, L2, L3)

The algorithm is outlined as follows:

1. Calculation of voltage correlation between each pair of SMs. This is a matrix of dimensions NxN, where N is the number of SMs in the analysis.
2. Clustering the SMs into 3 groups
3. Repeat step 1,2 using multiple variations of the input data and assess their performance.

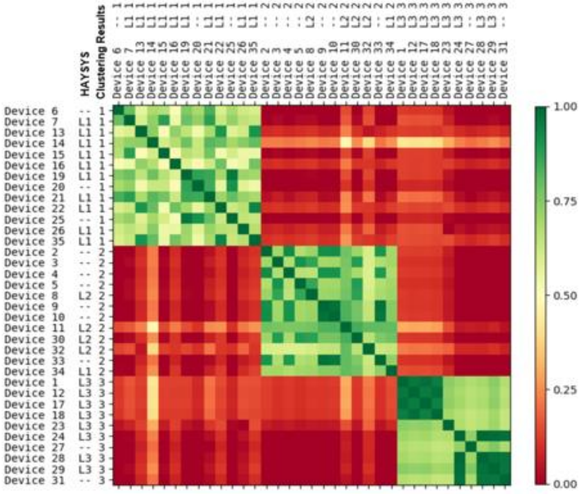
The label for each cluster was determined using the HAYSYS validation data.

Accuracy 0.80

1076 out of 1347 devices

Experiments:

1. K-means vs Hierarchical Clustering
   - **Hierarchical Clustering** with Complete linkage function was proved to perform better than k-means
2. Voltage vs Step changes
   - The use of **step changes** improved the algorithm's performance
3. Fisher Z transformation: It did not improve the results.



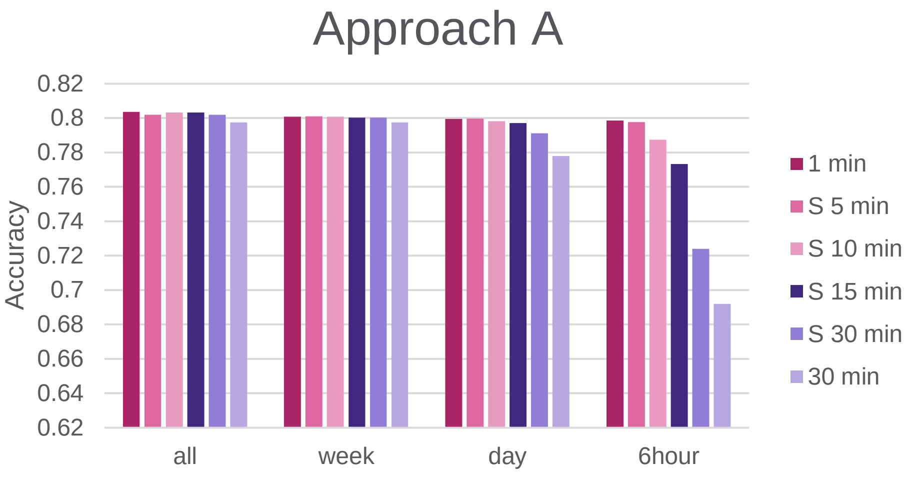
1-minute vs Half-hourly

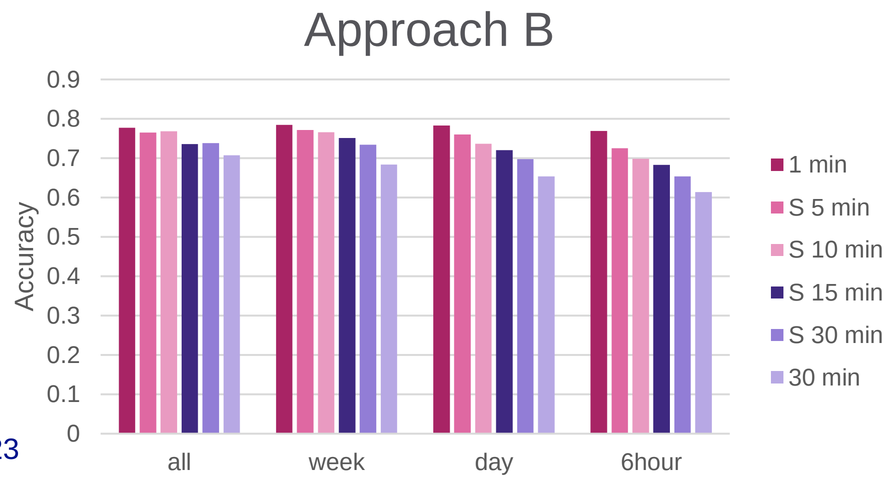
Assumptions:

- The algorithm works better with higher resolution data, because higher resolution data captures more information.
- The measurement interval synchronisation and clock offset in HH data could impact the HH results.

We resampled the 1-minute data to synthesize 5, 10, 15 minute and half-hourly voltage data.

- The accuracy drops as we decrease the sample size of the data, especially when the sample size is reduced below one week.
- The accuracy decreases significantly with half-hourly data, with further decrease in real half-hourly data due to measurement interval offsets.





Approach A and B agreement

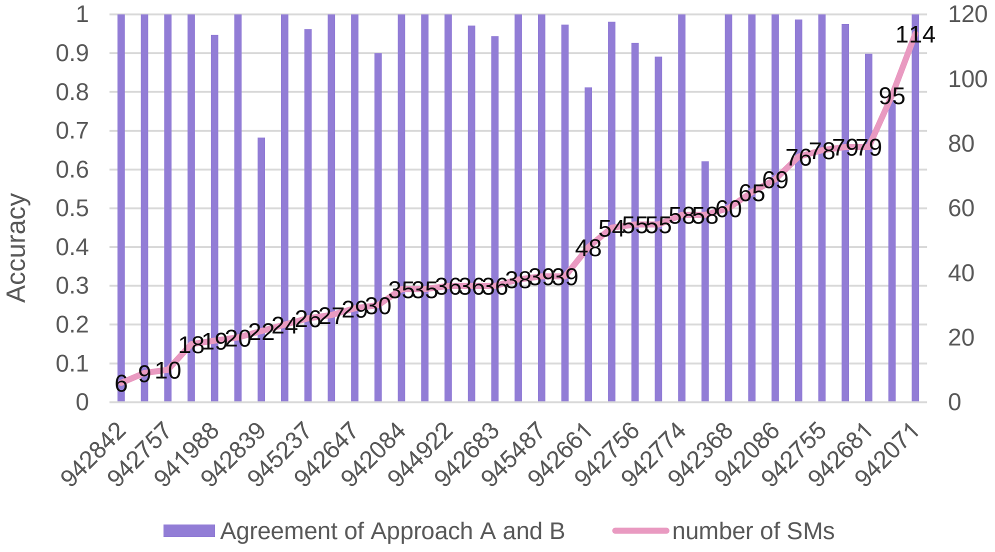
The purpose of this section was to assess if the results from Approach A and B consistently agree.

Approach A will work well regardless of the number of SMs per SS.

Approach B could be compromised if there are insufficient neighbours to form a cluster group.

The accuracy of the Approach B is calculated using the results from the Approach A as validation set.

- The results from the Approaches A and B are consistent in more than 90% of the SS.
- Both methods provide good agreement, even where there are low numbers of smart meters.



Phase Identification - Learning Points

- ✓ Improvements to smart meter voltage data capture process would significantly improve visibility of the customer connections.
- ✓ Phase identification using voltage data works well, either with substation monitoring as a reference, or using a clustering approach.
- ✓ Clustering techniques are more dependent on having sufficient smart meters on each phase.
- ✓ Clusters still need to be identified as L1, L2, L3, using phase data where available.
- ✓ Higher resolution voltage data can give more accurate results, especially when used as an input into the clustering algorithms. The analysis of using synthetic 5-minute and 10-minute average data shows that it could be used as an optimum balance between accuracy and data management effort.
- ✓ Ongoing work to understand the remaining differences to the survey
- ✓ Phase identification depends on MPAN records for substation allocations so that the correct set of smart meters are included in the clustering
- ✓ Methods developed to detect substation allocation errors, linking to the load profile and feeder identification work

Electricity Distribution

4

Planning Profiles

nationalgrid

Planning Profiles - Approach

Aims

A. Create half hour profiles for LV Feeders and for Secondary substations for use in Network Planning

Approach

1. A Bottom-up approach is used i.e. create an estimate for individual MPANs then use information about the connectivity of the MPAN to calculate the LV Feeder / substation consumption.
2. Unmetered supplies (e.g. street furniture) is not estimated
3. LV Network losses are not estimated
4. Reactive Power is not estimated

Validation

Metering on each LV Feeder at the substations used for validation to compare the following criteria:

A. Daily Average kW
B. Daily Peak kW
C. Annual Peak kW
D. Daily HH Profile

Planning Profiles – MPANs and Methods

MPAN Types

1. **Half Hour metered MPANs** (Measured HH data used)
2. **Smart Meter MPANs** (Measured HH data for group of MPANs used)
3. **Non-Half Hour (NHH) metered MPANs** (Use of one of the estimation methods below)

Estimation Methods

1. **NHH1 -** Use of Elexon Daily profile coefficients and MPAN Estimated Annual Consumption
2. **NHH2 -** Use of Seasonal Day Type profiles and MPAN Estimated Annual Consumption
3. **NHH3 -** Smart Meter Data as a proxy for NHH MPANs
4. **NHH4 -** Smart Meter to create profile coefficients . Apply to NHH using Estimated Annual Consumption

Connectivity

1. **Crown Connectivity**
2. **Electric Office (EO) Connectivity**
   - Where the service connection was shown in EO
   - Where the service connection is not in EO using a proximity algorithm to identify the closest feeder to the MPAN

Planning Profiles – Data Issues

Smart Meter data - Spurious values

1. A half-hour period value > 500kW
2. Average kW for a day < -10kW
3. Average kW for a day > 200kW

HH Meter Data

1. Missing data

LV Feeder Measurements

1. Missing data
2. Spurious values

Planning Profiles – Initial Assessment

As the quality of the MPAN connectivity was important an initial analysis was done to categorise substations/feeders so that poor connectivity did not obscure the comparison between the methods.

Determine Connectivity (Crown or EO)

Using the minimum, maximum and average values of the following criteria

- Daily Average kW
- Daily Peak kW
- Annual Peak kW
- Half Hour Profile

Select Feeder/Substation to include

Selection based on the minimum Daily Average demand across the 4 estimation methods (NHH1, NHH2, NHH3, NHH4)

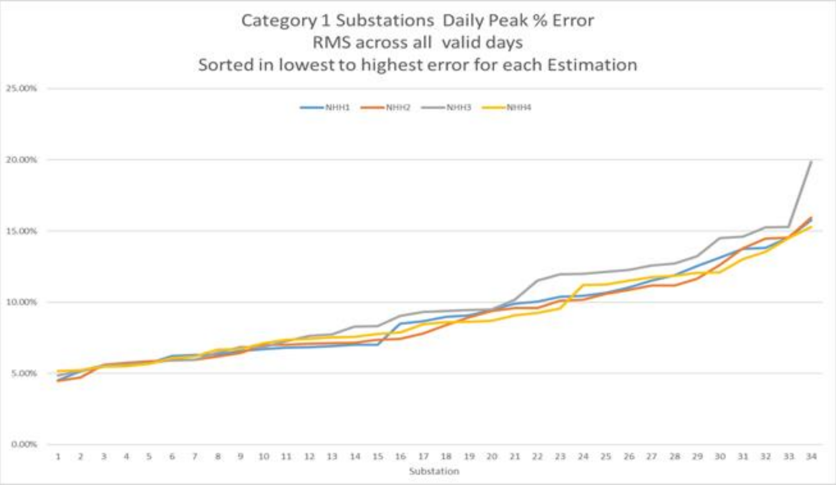
- Substations: < 10%
- Feeders: < 15%.

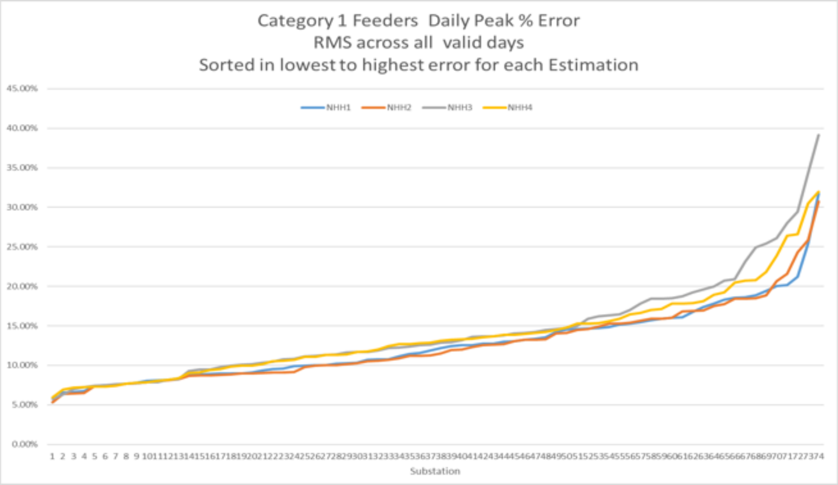
Based on these values 34 (out of 45) substations and 74 (out of 115) feeders were included for the comparison between the estimation methods.

Planning Profiles – Daily Peak Error

These graphs show the RMS of daily peak errors for substations and feeders. All the estimation techniques have a similar performance against this criteria.

Is a 5 -15% error for substation and 5 - 20% error for feeders sufficient?



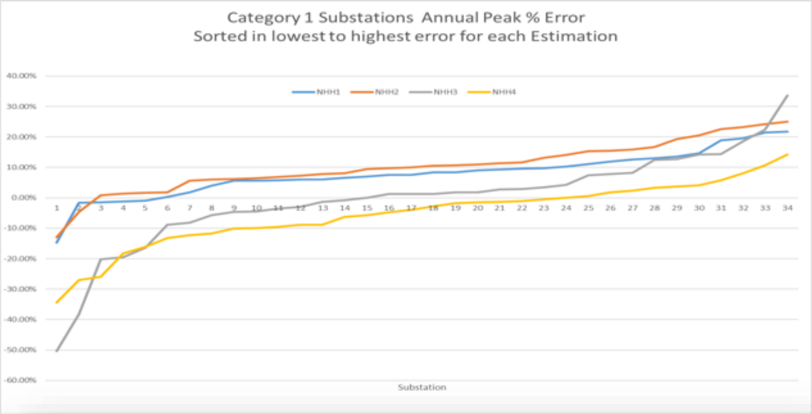


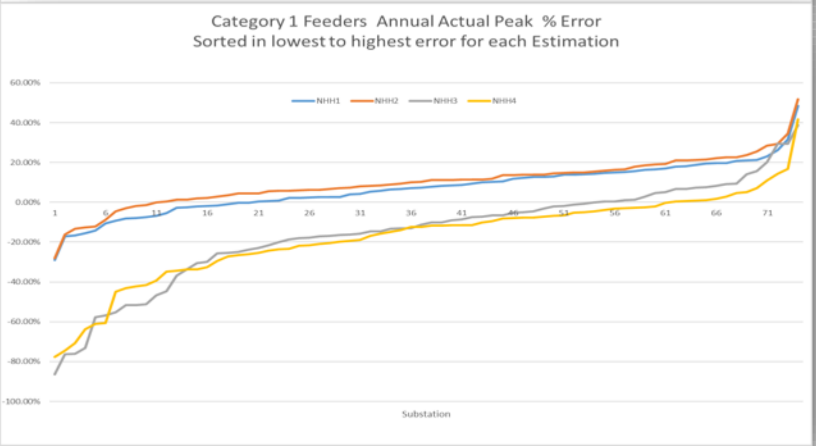
Planning Profiles – Annual Peak Error

These graphs show the annual peak % error. A positive value means an under estimate and a negative value is an overestimate

The two smart meter methods (NHH3, NHH4) tend to overestimate whist the Profile coefficient methods (NHH1, NHH2) tend to underestimate.

Overall NHH1 appears to provide the least error. About 66% of substations are with 50% of feeders have less than 10% error.

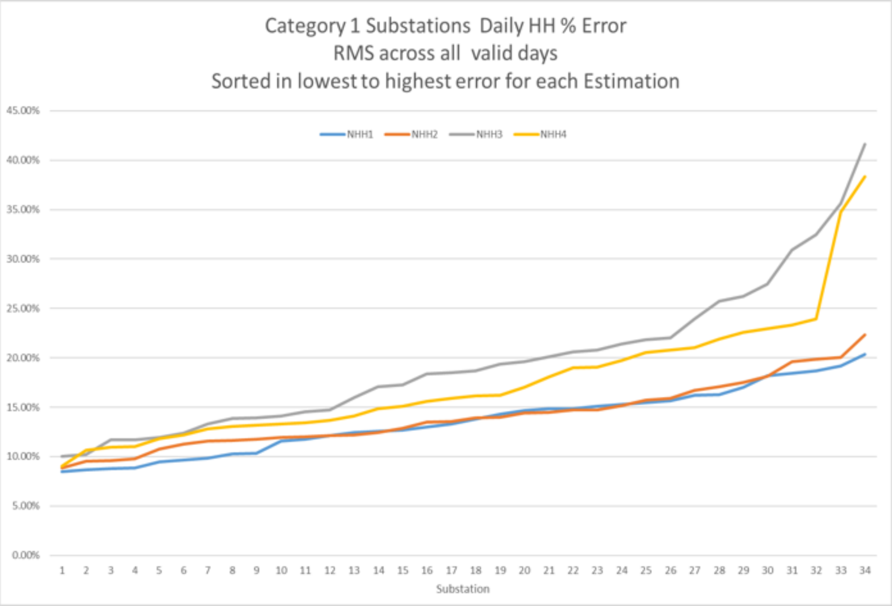


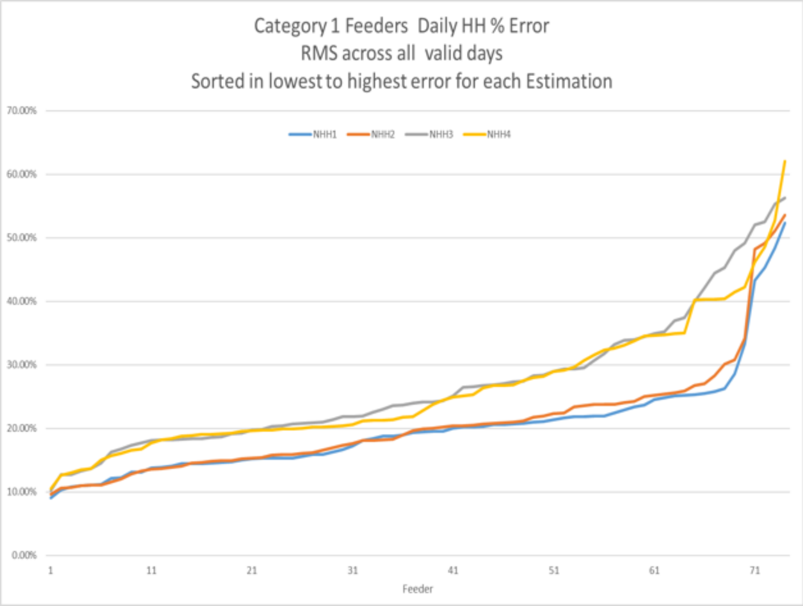


Planning Profiles – HH Profile Error

HH Profile error is a measure of the profile fit.

The two profile coefficient methods (NHH1, NHH2) have the least errors.

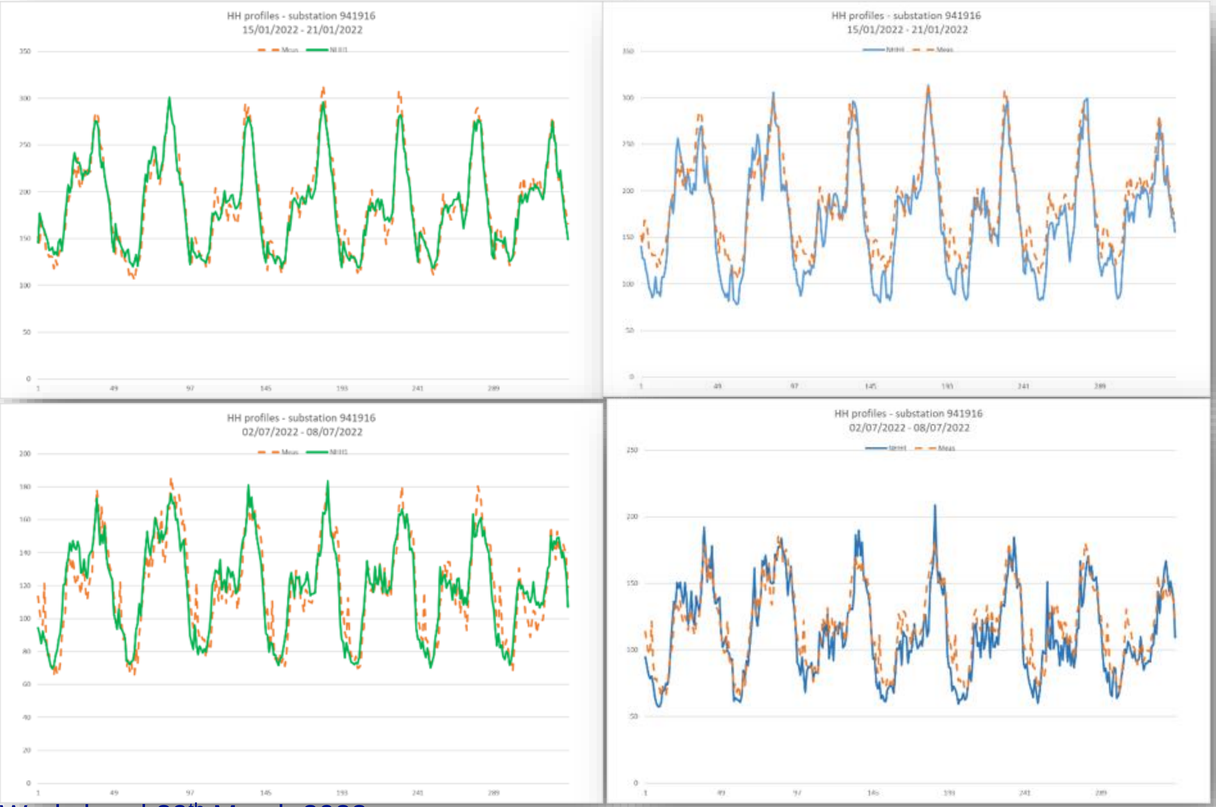




Planning Profiles – Substation HH – Daily Error 2%

These graphs show HH profiles for 1 summer week and 1 winter week for estimation methods NHH1 and NHH4 for a substation with a low daily average error (2%).

Both visually look close with the NHH1 method marginally better than the NHH4 for both winter and summer weeks.



Planning Profiles – Learning Points

- ✓ Focus on creating accurate profiles for a known set of customers per feeder
  - ➢ Builds on phase and feeder identification to resolve and confirm the connectivity
- ✓ Synthesized profiles compared against substation monitoring
- ✓ RMS daily peak accuracy 5%-20% for feeders, 5%-15% for substations
  - ➢ Better with more aggregation
- ✓ Annual peaks mostly within ±20% but some feeders with significant error
  - ➢ Scaling generic profiles (NHH1 and NHH2) tends to under-estimate
  - ➢ Scaling smart meter data (NHH3 and NHH4) tends to over-estimate
- ✓ Synthesized half-hourly profiles better by scaling generic profiles (NHH1 and NHH2)
- ✓ Errors generally similar throughout days of the week and months of the year

Electricity Distribution

5

Feeder Allocation

nationalgrid

Feeder Allocation- Approaches

01

Discrepancies between EO and CROWN

02

Outliers from Phase Identification

03

Correlation with neighbouring SS

04

Clustering SMs into Feeder groups

05

Feeder Profiles

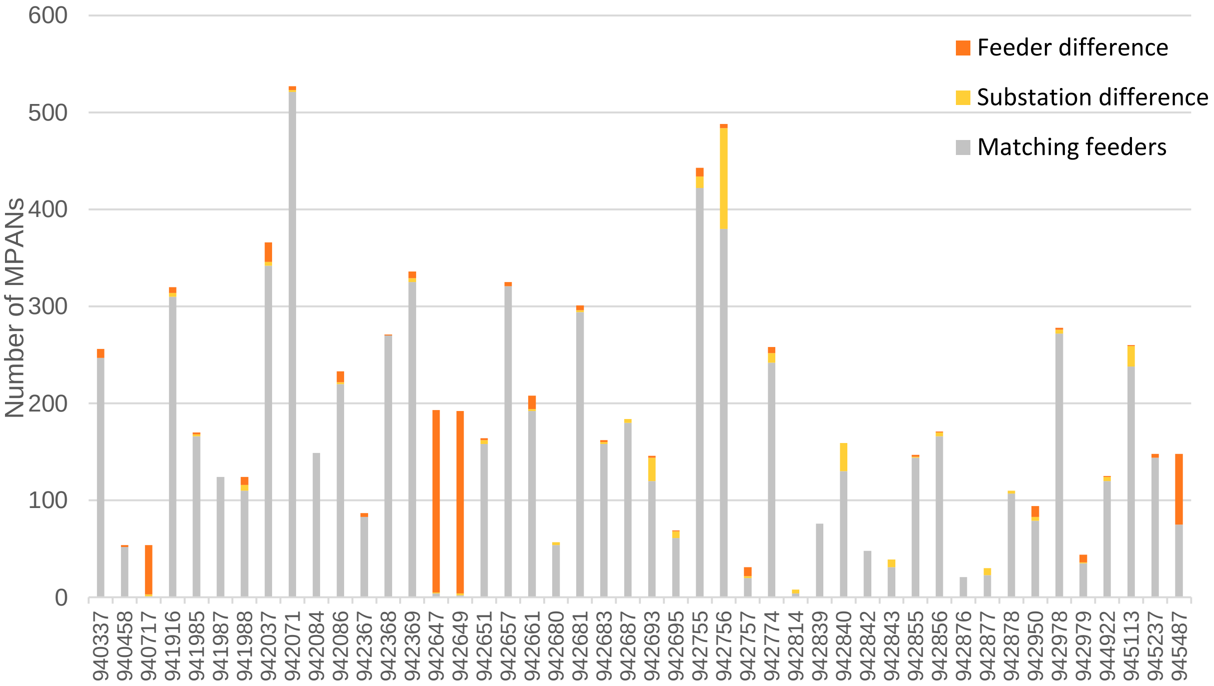
Feeder Allocation – Discrepancies between EO and CROWN

CGI's Integrated Network Model (INM) was used to store the EO and CROWN data.

To identify the feeder for each MPAN in EO, two methods were used:

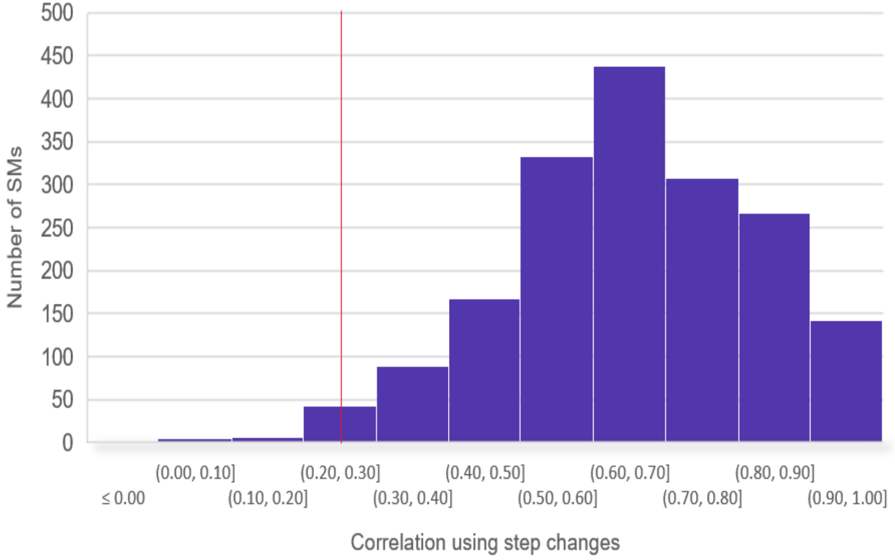
- Where the service connection and premise details exist in EO, the feeder information was obtained from the EO LV feeders with the "connectivity" approach.
- Where the service connection and MPAN information does NOT exist in EO, a proximity algorithm is used to identify the closest LV feeder for each MPAN using the coordinates obtained from CROWN.

The algorithm compares the information from CROWN and EO and flags the properties where the LV connectivity is different in the two systems.



Feeder Allocation - Outliers from Phase Identification

- This method initially uses the results from Phase Identification use case.
- Identifies properties that have been assigned to the wrong Secondary Substation (SS).
  - Properties with very low correlation are pinpointed as outliers.
  - Properties that the phase result is not stable if we use different time intervals are pinpointed as outliers.
- It cannot identify properties that have been assigned to the wrong feeder
- Requires SS monitoring voltage data.

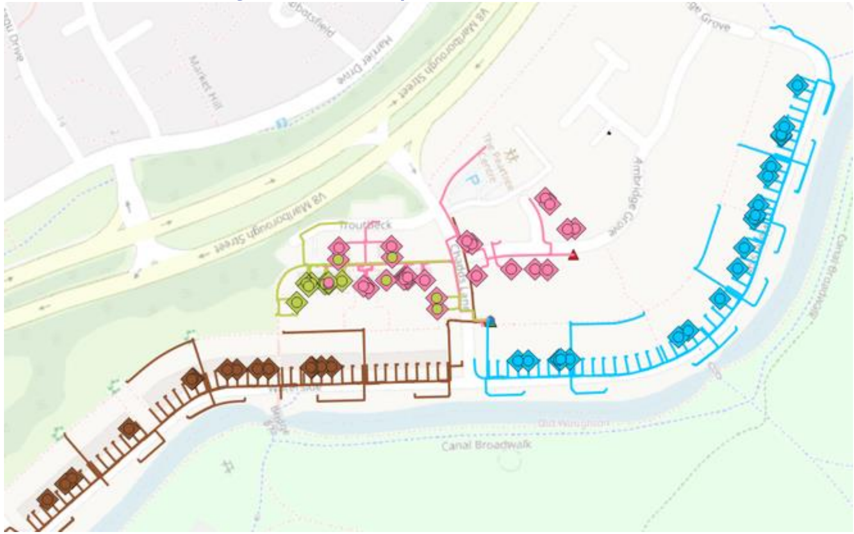


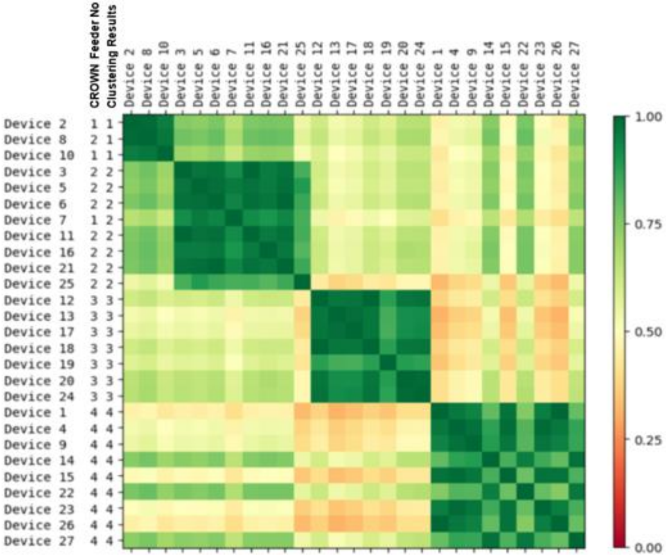
Feeder Allocation – Correlation with neighbouring SS

- In the case of where the examined secondary substations and the neighbouring secondary substations are fed from the same HV feeder, the correlation could be high even if the property has been assigned to the wrong substation.
- The algorithm recalculates the correlation using voltages from the neighbouring SS on the same HV feeder. In absence of monitoring data the algorithm could work using SM as reference from both substations.
- If the correlation result increases when compared to neighbouring SS, the algorithm identifies these properties as outliers and highlights them for review.

Feeder Allocation – Clustering SMs into Feeder Groups

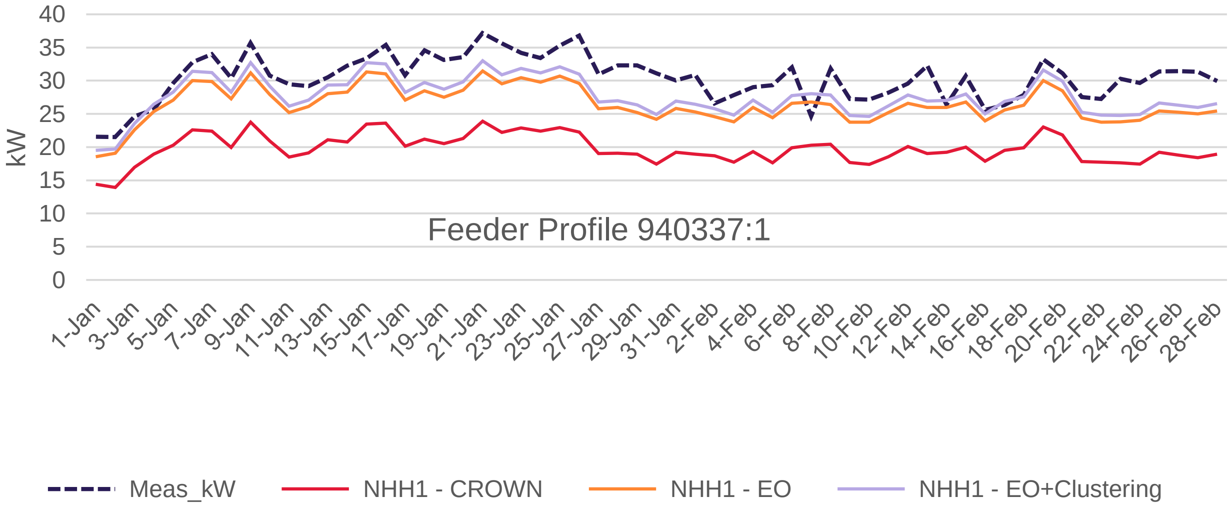
- This approach is focused on identifying SMs that have been assigned to the wrong feeder.
- It assumes that the SMs have been assigned to the correct SS.
- The algorithm calculates the voltage correlation between each pair of SMs. This is a matrix of dimensions NxN, where N is the number of SMs in the analysis.
- For each phase, the algorithm clusters the SMs into M groups using clustering techniques, where M is the number of feeders in each substation.
- This approach requires accurate phase information. The results from “Phase Identification” use case were used.
- No SS monitoring data is required.





Feeder Allocation – Results from Feeder Profiles

- The algorithms used in “Planning Profiles” use case are sensitive to customer to feeder connectivity.
- The “Planning Profile” results can give another source of truth between the EO and CROWN discrepancies.
- The algorithm identifies the feeders where the daily MAPE is very low when we use the data from the one source, but the error increases significantly when the other source has been used. Then, the properties where the LV connectivity is different in the two systems are highlighted for review.
- Feeder load profiles are re-calculated using the MPAN to feeder connectivity from EO, CROWN and the Clustering.



Feeder Allocation – Learning Points and Next Steps

- ✓ Feeder allocation errors in the database occur on most feeders
  - ➢ On some feeders, there are significant numbers of errors due to numbering swaps
- ✓ Feeder identification is less straightforward than phase identification
  - ➢ Many approaches requires that phase identification is already correct
  - ➢ Voltages for connections near the substations are similar on each feeder
- ✓ Good results can be obtained by voltage correlation methods
  - ➢ Depends on reasonable starting point with correct substation allocation
- ✓ Validation using the HAYSYS Feeder Finder survey is in progress.
- ✓ Investigation of the integrated use of a combination of algorithm types to increase the probability of data accuracy.

Electricity Distribution

6

LCT Detection

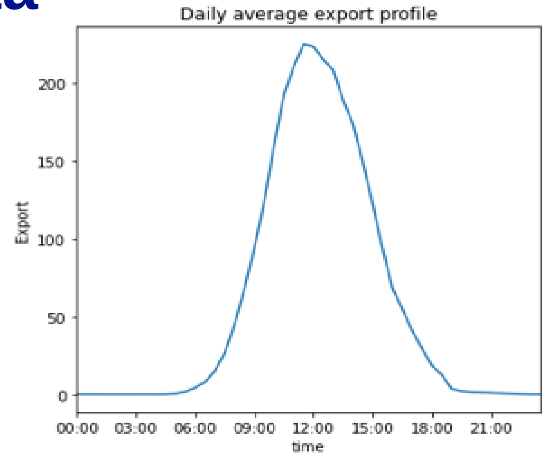
nationalgrid

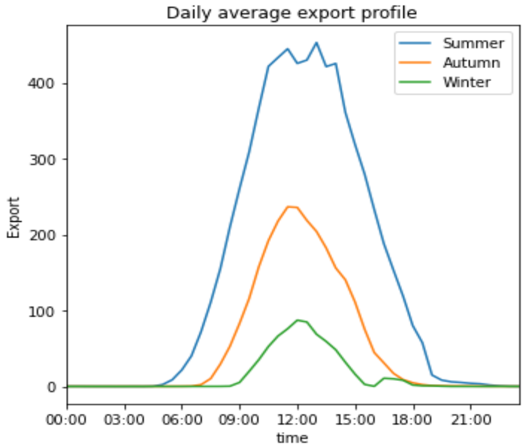
LCT Detection – Half-hourly Export Data

- Half-hourly export data for individual MPANs can help DNOs identify embedded generation which has not been recorded in their systems.
- We are focusing solely on the identification of PV installations using HH export data, as there is a low presence of energy storage or other types of generation in the core area.
- The algorithm compares the daily average maximum export of the devices with “unknown” PV installation with the daily average maximum export during the night.
  - ➢ 3.62% of the properties have been recorded as having PV in CROWN in the core area.
  - ➢ 94.1% of the properties with PVs have average export higher than 0.
  - ➢ There are 27 properties that have export during the day but there is no LCT information, 20 of them have been verified as having a PV using Google Maps.

Next Step

- An ML algorithm can be created using HH AE data and other relevant datasets such as weather data to identify the different types of LCTs present at a property.
  - ➢ Require export data from other types of LCTs such as batteries and wind turbines.





LCT Detection – Aggregated Monthly Consumption

❑ Datasets:

- Aggregated monthly consumption data per device for 12 months.
- Maximum consumption data in any half-hour period within a month for each device and the time that have been recorded. Available for 12 months.

❑ Assumptions:

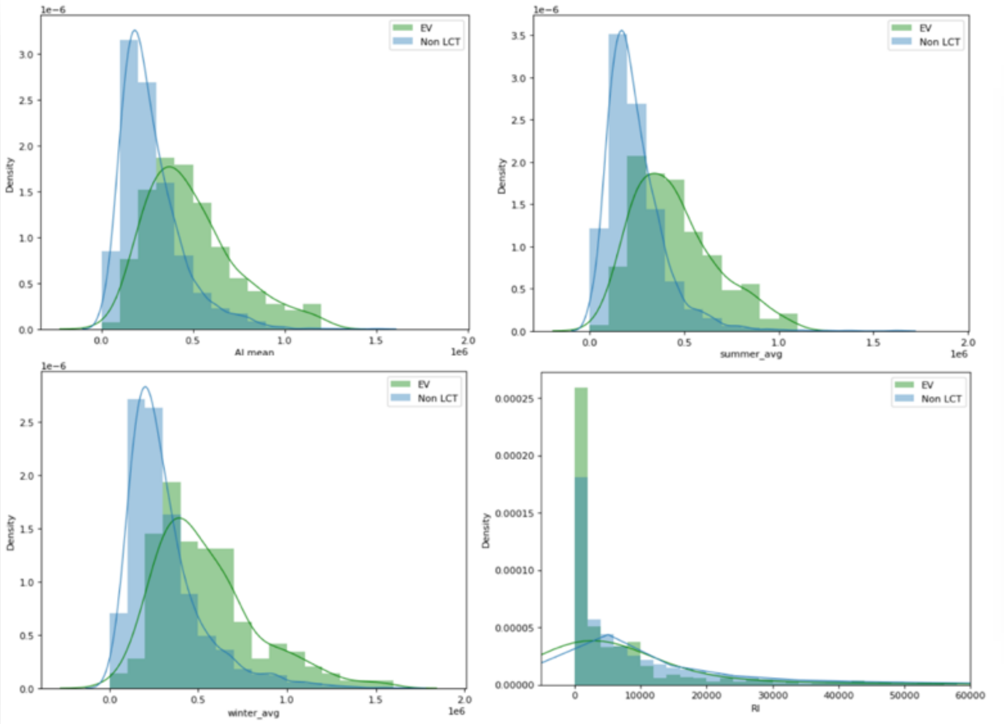
- Having an EV, a property may have higher energy demand compared to a property without an EV charger.
- Unusual spikes in energy demand during certain months can be indicative of an EV charger.
- The monthly demand could be more constant and not vary significantly with an EV.
- The maximum consumption could be higher in properties with EVs.
- The time that the maximum consumption have been recorded could be during the night as people tend to charge their car during the night.

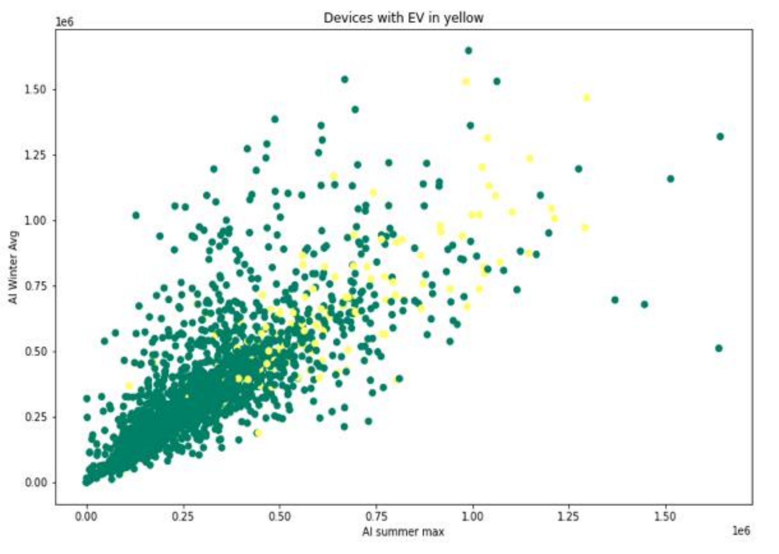
❑ Limitations:

- Limited data points: Because we have only 12 data points for each MPAN, it is difficult to identify any consumption pattern changes from the installation of an LCT within the 12 months.
- No negative set: The training data consists of positive and unlabelled properties.
- Limited Data for properties with heat pumps and more than one type of LCT installation.

LCT Detection – Exploratory Data Analysis for EVs

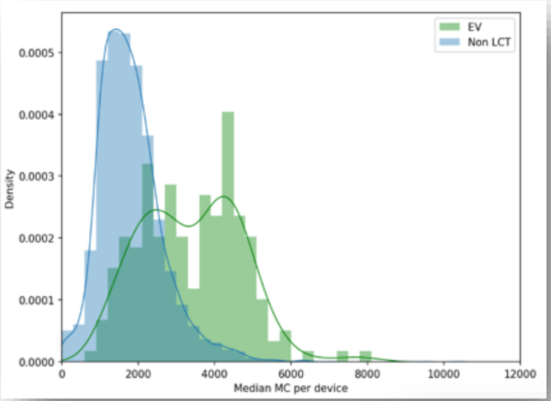
- Monthly aggregated consumption data per device

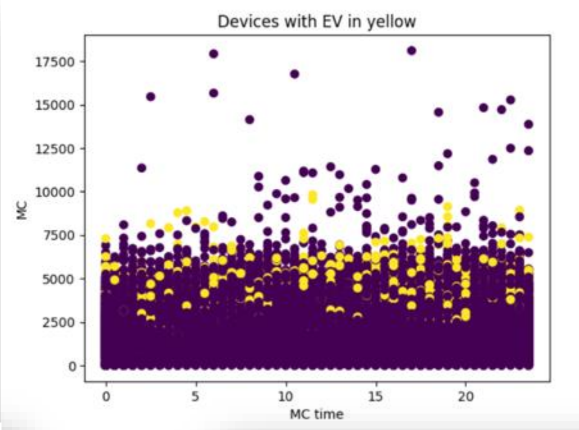


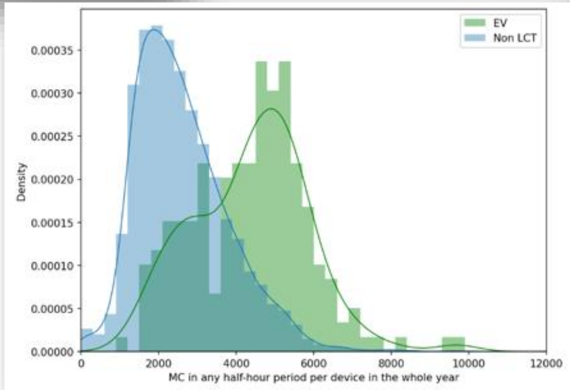


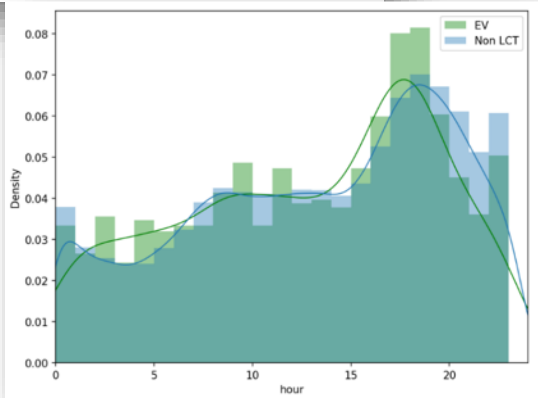
LCT Detection – Exploratory Data Analysis for EVs

- Maximum consumption in any half-hour period per month per device









LCT Detection – Modelling Approach

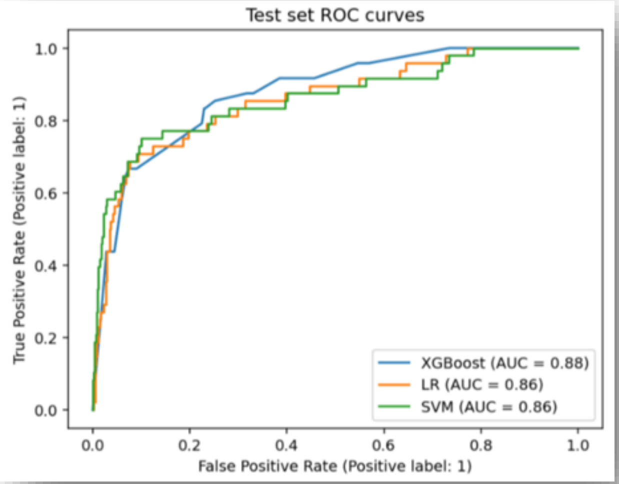
Experiments:

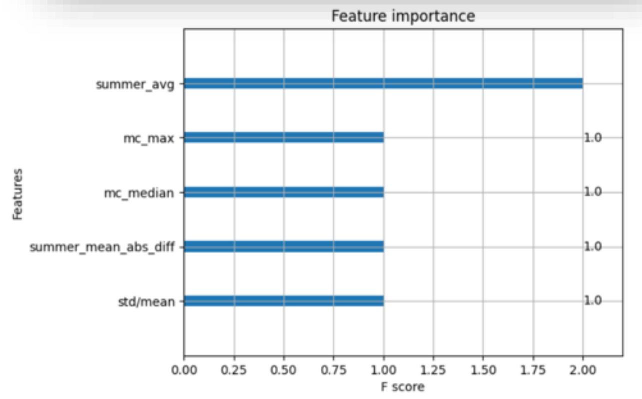
2. B. Binary Classification and PU (Positive-Unlabelled) learning

- We converted the time series data into statistics that we can train the model with.
- The positive examples are assigned a weight of 1, and the unlabelled examples are assigned a weight that reflects their estimated probability of being positive.
- The weighted LR model is then trained using the weighted data, where the positive examples are given higher importance than the unlabelled examples.

  - **Weighted Logistic Regression/Weighted XGBoost**
  - **Weighted Support Vector Machine (SVM)**
    - 70% of properties with EV chargers in CROWN have been classified as EVs.
    - 8% of the "unlabelled" properties have been classified as EVs.

- The summer consumption and the maximum consumption features have higher impact on the models.





LCT Detection – Identifying EVs using Voltage Data

- 1-minute voltage data was used to identify trends in the data that could reveal the existence of EVs.
- Voltage data are available only for the core area where the existence of EVs are low (only 39) devices.
- Devices with known EVs have occasional drops in voltage for several hours at a time.
- We can subtract the voltage data from another device on the same phase and feeder to try and cancel out the general voltage trends seen across all the devices on that phase.

LCT Detection - Learning Points

- ✓ HH export data for individual MPANs has good potential for detection of PV and other types of generation.
- ✓ Monthly demand per MPAN can be used to detect EVs based on closeness to a median curve. This has a high recall rate but also many false detections. It is difficult to separate properties with a high consumption from those with a high consumption due to EV charging.
- ✓ Weighted ML algorithms are very promising for identifying EVs (PU learning).
- ✓ The summer consumption and the maximum consumption features have higher impact on the models.
- ✓ Voltage data looks promising to identify EVs due to distinct periods with large drops while charging.

Next Steps

- ✓ Creation of a pattern recognition algorithm that can identify the voltage drops caused by EVs.

7

Questions and Summary


Any
Questions

Next Steps

Final Dissemination Webinar

Mid May

Closedown report Published

Mid July