Smart Meter Innovations & Test Network (SMITN)

Closedown Report

May 2023

nationalgrid

o

Version Control

Issue	Date
0.1	12/12/2022
0.2	06/04/2023
0.3	16/05/2023
0.4	24/05/2023
1.0	21/06/2023

Publication Control

Name	Role
Andrew Urquhart, Loughborough University Iro Psarra, CGI Nadim Al-Hariri, CGI Joe Hayden, HAYSYS	Authors
Liza Troshka, Steve Pinkerton-Clark	Reviewers
Jenny Woodruff	Approver

Contact Details

For further information please contact:

nged.innovation@nationalgrid.co.uk

Postal

Innovation Team
National Grid Electricity Distribution
Pegasus Business Park
Herald Way
Castle Donington
Derbyshire DE74 2TU

Disclaimer

Contents

1 Executive Summary

Smart meter data can be used in new ways to benefit Distribution Network Operators (DNOs). Improving load estimates for planning use has been a long-term goal, but additionally the use for phase identification, feeder confirmation and identifying unregistered low carbon technologies will help improve the accuracy of our Low Voltage (LV) network model and in turn the accuracy of our LV network analysis. Smart Meter Innovations & Test Network project (SMITN) explored these use cases with a test network area that had been validated for phase and feeder connectivity so that the accuracy of the algorithms could be calculated with confidence. This validation involved the use of the HAYSYS Phase Finder unit and the development of a new HAYSYS Feeder Finder tool. CGI were the main project partner with support from Loughborough University and quality control checks performed by GHD. Project costs were approximately £915k with benefits estimated at £740k per annum with the majority of the benefits relating to reducing the number of distribution substations requiring monitoring. The project lasted for 16 months from March 2022 to July 2023.

Project SMITN has utilised findings of the previous work by Loughborough University to determine the impact of low overall smart meter coverage on the accuracy of simple load estimates. SMITN has also extended our use of the Cassandra system which is used to manage aggregated smart meter data provided via the Data and Communications Company (DCC) adaptor. To ensure General Data Protection Regulations (GDPR) compliance half-hourly load data values were aggregated over at least three customers. However, SMITN made use of previously unused data items recorded in smart meters, which were average voltage (in Volts) as measured within the smart meter and exported power (in kWh) i.e. where power flows from the customer's premises onto the DNO's network. As these data items are not considered to be Personal Data they could be associated with individual premises without needing to be aggregated which makes them more useful in the various algorithms. The average voltage data resolution was improved by setting the smart meters to record the average voltage over 1 minute time intervals rather than the default time interval of thirty minutes.

Algorithms for the use cases were discussed with the wider industry before assessing the effectiveness of different approaches. One of the major issues for the project was obtaining data in timescales that were suitable for the project, due to the time and resources required to set up new data extraction routine, and creating a process to ensure secure transmission of the data to CGI. Similarly, the Data Privacy Plan has been amended to allow for aggregation of half-hourly load data for sections of LV feeder and grouping customers with similar profiles together.

Data cleansing activities that were most useful were checking for offsets between the clock times used by different meters and testing for suspected errors in the CROWN data for the LV feeder to which each customer is connected.

Very good results were seen for phase identification, especially where the substations are monitored (83% mean accuracy) so that phases could be positively confirmed for the customer groups created.

Where monitoring is not present, mathematical clustering methods can still create groups of customers believed to be connected to the same phase for later phase confirmation and these were seen to have mean accuracy levels over 80%. There was a degree of overlap between the use cases as errors in how customers were recorded as being connected to a particular substation or LV feeder would impact the phase determination method and also result in lower accuracy for load profiles. This suggests that the business as usual process will need to be iterative and will improve as underlying data quality improves.

For load profiles, it was found that adapting the Elexon imbalance settlement process used to estimate half hourly consumption for different customer types[1] was more accurate than trying to create local representative profiles from neighbour's smart meter data.

It was possible to identify premises suspected of having unregistered domestic Photo Voltaic (PV – Solar Panels) and confirm the presence of rooftop solar panels using publicly available satellite and street view images. The model to identify Electric Vehicle (EV) chargers faced a greater

[1] https://www.elexon.co.uk/knowledgebase/what-is-the-imbalance-settlement-process/

challenge as other devices have a similar impact on voltage, but further refinement of the model is likely to improve results.

The roll out into Business-as-Usual (BaU) will need to involve several teams within National Grid Electricity Distribution (NGED) and will need to determine the desired timeframes to cover all of the LV networks within each licence area in order to determine the processing power required and the issues around scaling up and automating the requests for smart meter data. The long-standing issue of how to store algorithm results and to keep them distinct from legacy or field validated data also needs to be resolved.

2 Project Background

Data quality and complete representation of our LV networks is increasingly important to support LV planning and operation as large volumes of Low Carbon Technologies (LCTs) are connected. Our Business Plan for the price control period covering 2023 – 2028, known as RIIO-ED2, includes a new commitment to provide a same-day connection response for low carbon technologies via an online service. This automated system will not allow time for network data to be checked using local knowledge from the planning team and therefore relies on accurate and complete datasets. We will also need to be able to support wider data sharing with third parties.

Previous Network Innovation Allowance (NIA) projects, such as the Losses Investigation[2], have highlighted the issues with our data such as customers being associated with the wrong LV feeder or sometimes the wrong distribution substation. Similarly, phase information, while now being captured for new substations, is often missing entirely from our network connectivity records though it is sometimes captured as text items on the plans. These labels are readable for humans but not easily associated with the corresponding identifier for a meter, the Meter Point Administration Numbers (MPANs) and incorporated into network modelling. Recent experience in analysing smart meter data to determine demand profiles has demonstrated that, even in new-build developments, there are accuracy issues for the LV feeder records in the CROWN asset database where planned connectivity has been used and not updated if final connections were different. This leads to potential complications when dealing with outages and maintaining supplies to vulnerable customers. It also introduces errors to the demand estimation using smart meter data as errors in the records cause the incorrect set of customers to be aggregated when calculating the half-hourly profile on each feeder.

Without understanding how customers are connected to different phases we cannot determine the impact of phase unbalance, which results in increased losses and reduced thermal capacity. In responding to a new LCT connection request the associated modelling will need to make some allowance for potential unbalance. There may be different numbers of customers on each phase, and it is highly likely that the phases of existing LCTs will not be evenly distributed. Some allowance will need to be made for this potential unbalance. However, if the phase connections are known, potential risks of having many LCTs on the same phase can be detected. New connections can also be permitted when the phase connections are known to be more favourable, whereas in the absence of this information, they may have been unnecessarily constrained.

There are existing tools that can be connected to the distribution substation to inject a signal onto the network which can then be used to detect the phase or feeder to which a customer is connected. However, this often requires access to the inside of customer premises in order to test for the signal which results in issues when customers are not at home or do not wish to grant access to their premises. Explaining the requirement to customers can also be time consuming compared to a method that did not require access.

While progress has been made to reduce the number of low carbon technology installations that are not recorded on our systems, this is still an issue. Previous projects have highlighted the issues in assessing the accuracy of predicted characteristics, such as the customer phase, when comparing against a reference dataset that is also known to include inaccuracies. Having a validated network allows predicted values from algorithms to be assessed as correct or incorrect with confidence.

Smart meter data provides new possibilities to understand network loads in order to use realistic values for network planning. Our current LV planning tool, Connect LV, uses an ACE 49 approach to determining peak network conditions which uses estimated consumptions for customers rather than incorporating smart meter data. While substations profiles can be generated for the distribution substations using the Network Investment Forecasting Tool (NIFT) this does not incorporate smart meter data and would therefore not be as accurate as a method that used smart meter data where available.

[2] Losses Investigation project https://www.nationalgrid.co.uk/projects/losses-investigation

Smart meter data coverage is still relatively low with a UK average of around 45% of customers[3], however previous analysis suggests we can improve our planning data by incorporating the smart meter readings with our legacy data for customer type, estimated annual consumption etc. We have not previously explored how to optimise this application for different types of substations, or for different levels of smart meter coverage, so that we can implement new methods as coverage levels increase.

For the development and testing of the algorithms in this project, the identification of phases and feeders were largely based on voltage data from smart meters. This data can be requested by a DNO, with each request providing 4320 records from each smart meter, equivalent to 3 days duration with 1-minute resolution, or approximately one month with 10-minute resolution. The impact of time resolution for use with the algorithms has been determined as part of this work. This voltage data does not describe the demand profile for individual customers and so is categorised as operational data.

Conversely, the development of load profiles requires customer demand data. This is categorised as personal data and must be aggregated to anonymise the consumption of any one customer so that it no longer is categorised as personal data. We have recently started capturing half-hourly data aggregated at LV feeder level within the Cassandra system alongside monthly consumption data for individual MPANs. This data forms part of the data inputs to the algorithms, combined with techniques to estimate demands from customers without smart meters.

Following our privacy policy, customer demand data is aggregated to include at least three customers to allow a safety margin over the minimum two customers required to be aggregated together to achieve anonymisation. This data has half-hourly resolution with data being available for a time history of three months or more.

The focus of this project was largely on substations serving domestic customers where new connection requests for LCTs can be expected. The algorithms were therefore developed and optimised for substations with domestic customers, including those with a mix of domestic and non-domestic customers. Only substations that overwhelmingly feed non-domestic customers were excluded.

The benefits of the project were expected to be:

- The capacity to establish customer feeder and phase with a high confidence using smart meter data.
- The capacity to create estimated load profiles for LV feeders and distribution substations at a level of accuracy that can be used to support LV planning activities and that can be recalculated easily as smart meters and low carbon technologies are added to the network.
- The capacity to identify unregistered low carbon technologies associated with segments of LV feeders for further investigation and validation.
- The ability to validate customers' feeder association without requiring access within the premises.
- The establishment of a validated test network which can be used to measure the performance of other algorithms or test equipment in the future.

These have generally been achieved and are reflected in the project objectives and success criteria which are assessed in section 6.

Having an accurate LV network model and load estimates is a prerequisite for enabling self-serve connection applications. It also enables identification of the best phase for new connections to existing networks to reduce rather than increase phase unbalance.

3 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1099629/Q2_2022_Smart_Meters_Statistics_Report.pdf

3 Scope and Objectives

The scope and objectives for project SMITN are given in the table below.

Table 3-1: Status of project objectives

Objective	Status
To determine a representative test network of selected distribution substations and validate the key features of this network by carrying out surveys	✓
To capture smart meter data using new aggregation groups	✓
To develop algorithms using smart meter data for the following use cases: a) Customer to Phase connection prediction b) Customer to Feeder connection prediction c) LCT identification of potential locations and types d) Provision of LV feeder and Distribution substation planning profiles for use in network planning	✓
To apply the algorithms for data relating to the test network	✓
To assess the performance of the algorithms and where possible identify the factors that affect accuracy	✓
To capture the learning from the project and disseminate this to interested parties	✓

4 Success Criteria

The success criteria for project SMITN are given in the table below.

Table 4-1: Status of project success criteria

Success Criteria	Status
A set of phase validated networks and a subset where customer-to-feeder association has also been validated that can be used to test future innovations beyond this project.	✓
An understanding of how well the smart meter data can be used to support phase identification and what factors affect the accuracy of the algorithms used.	✓
An understanding of whether smart meter data can be used to support the identification of unregistered LCTs and what factors affect the accuracy of the algorithms used	✓
An understanding of whether smart meter data can be used to support the validation of customer-to-LV feeder associations and what factors affect the accuracy of the algorithms used	✓
An understanding of how well aggregated smart meter data can be combined with other data available to DNOs to create estimated LV feeder load profiles and what factors affect the accuracy of these estimates.	✓
An understanding of the practical issues for using the newly developed feature for validating customer-to-LV feeder association without requiring connection to the customer's wiring and confirmation that the new feature is fit-for-purpose	✓
An increased insight of the phase unbalance on the test networks	✓
A view of the value of disaggregated data from smart meters rather than the default position of aggregation at the LV feeder level	✓

5 Details of the Work Carried Out

The project involved:

- Creating a validated test network against which to evaluate algorithms by carrying out surveys to confirm customer phase, feeder and where possible LCT information.
- Developing a new Feeder Finder tool, that allows feeder identification from outside customers' premises.
- Selecting algorithms to apply that use smart meter data; and
- Evaluating the performance of the algorithms.

There were a number of project partners with roles as given below.

Company	Role
WPD / NGED	The project was registered by Western Power Distribution (WPD) but during the project the company name was changed to National Grid Electricity Distribution (NGED) and that name is used in this report. Project management and provision of network data, smart meter data etc. And support for the phase survey via the Network Services part of the business.
CGI	Algorithm selection, providing and configuring the data processing environment, setting up the data processing for the selected algorithms, accuracy measurement.
Loughborough University	Area selection, support of algorithm selection and input of local knowledge from the losses report and other recent work on planning profiles, advising on the Go/ No Go decision to reflect data availability and quality.
HAYSYS	Development of the Feeder Finder, provision of phase and feeder surveys.
GHD	Independent validation of the data processing to ensure that the selected algorithms and accuracy measurements are correctly configured in the data processing environment.
Smart Grid Consultancy(SGC)	SGC were brought into the project to provide project management support after changes in the NGED team structure limited the available resources.
Power Projects Ltd	Power Projects Ltd. were brought into the project at a later stage to provide additional support for the feeder survey

This project was achieved through a series of work packages that are described, along with the deliverables for each work package, below.

5.1 WP1 – Network Selection & Validation, Algorithm review, Data provision.

This work package contained several deliverables. The first was delivered by Loughborough University (LU) which was to confirm the selection of the substations within the test network, considering existing substation monitoring data. It aimed to ensure we gathered data from the best mix of substations that will reflect the variations between substations in terms of size, locations, customers etc. for analysis. The results were documented in WP1 - D1 Area Selection report.

This was found to be more complex than originally anticipated as the selection was trying to take many factors into account simultaneously. CGI introduced a preference for selecting substations where the majority of a High Voltage (HV) feeder was monitored so that the load estimates generated could be compared to the total for the HV feeder. However, it also became apparent that some meters had known configuration issues and would not be suitable to support 1-minute average voltage readings, so the selection needed to avoid areas where those meters made up the majority of the smart meters that were installed. There was also a preference for including areas where there were high volumes of smart meters as this would ensure that data was already being gathered for aggregated LV feeder half hourly data. In the end two areas were selected, firstly a core area in Milton Keynes being the focus for phase and feeder identification, and this area was complemented by a second non-contiguous ‘extended area’ where VisNet monitoring units were installed which was used to the support the LCT detection and planning profile use cases. While the location of installed VisNet monitoring units was known, which units were fully operational was not known until later, when iHost (system used to access substation monitoring data) access was granted. This resulted in yet more revisions to the selected sites.

This work package also included an assessment of the available methods and algorithms for the SMITN use cases and selection of algorithms to be applied. This evaluation included a workshop with representatives from other DNOs to ensure that the widest range of methods was considered, and that any practical experience was reflected in the selection.

It was originally planned that, having selected the algorithms, the work to determine the hardware and software interfaces required to support the data processing would be a discrete activity that would be completed before populating the data stores. However, work was carried out on a more fluid basis as data became available. This meant that data processing for one use case was well underway before the data processing environment for the next use case was set up. This was a more practical use of the time available although this approach did limit the extent to which the results from one use case could be used to enhance the data quality for the others, for example by including the demand for detected LCTs within the load profile estimates.

The data provision from NGED was required to meet the strict cybersecurity requirements. No direct interfaces were created to the underlying systems providing the data but rather a Secure File Transfer Protocol site (SFTP) was used for relaying data between NGED and the other partners and CGI.

CGI then processed this data within a Microsoft Azure environment using two databases, one containing the relatively static network data such as mappings between distribution substations and HV feeders and another containing all the time series and operational datasets i.e. smart meter data, distribution substation monitoring data from the VisNet and GridKey units, primary substation monitoring data etc.

This work package also included an assessment of the available data and its quality in relation to required data for the selected algorithms. This Go/No Go decision provided a break point at which the project could be discontinued if it became apparent that the data would not be sufficient to provide worthwhile results. Fortunately, the assessment was that while there were some data quality issues, it was worth continuing with the project. The key concern related to the data extraction of 1-minute voltage data from the smart meters, which at the time had not been demonstrated, but was found to be work well for most devices. However, this voltage data could not be obtained from all meters as they did not respond to the instruction to change the timeframe over which the average voltage was calculated, with a corresponding impact on the analysis results.

Network validation involved confirming the phase of all customers connected to the core test network in Milton Keynes. This was necessary to provide confident phase data against which to assess the accuracy of the phase predicting algorithms. The survey was carried out by Haysys using their Phase Finder device. It was known that the Phase Finder works best where the service connections for properties are spaced out well and easily accessible, so the area selection work tried to avoid areas where there were lots of blocks of flats. Where LCTs were visible to the surveyor they were noted to help ensure the records that were used to assess the success of LCT detection algorithms were as up to date as possible. The option of trying to contact the customers to confirm any connected LCTs was not considered to be cost effective. However, the survey work was enhanced by taking images within the meter box where an external meter box was easily accessible. It was hoped that this could be used to resolve queries over installed meters at a particular site but might also open the door for future Artificial Intelligence (AI) image recognition projects, e.g.to identify the customer’s meter type, by providing a stock of images.

Given that the outputs of the project were mostly related to the data processing carried out by CGI, it was vital that they correctly applied the selected algorithms to the data and that the accuracy measurement was carried out correctly. GHD were employed to provide quality assurance by providing third party validation of the CGI data processing work. Originally there was concern that the data processing might be difficult to assess, but the use of Jupyter workbooks allowed good visibility of the code used to apply the algorithms with helpful annotation.

Deliverables

Deliverable	Purpose
WP1 - D1 Area Selection report	To determine the best selection of distribution substations, considering the need to represent different types of customers, available monitored sites, ease of survey and data quality issues.
WP1 - D2 Design & Build Trials Data Processing Infrastructure	Determine the required hardware, software and interfaces required to obtain data, process it to support the various use cases and analyse the accuracy of the results. This will include determining custom aggregation groups for LV feeders
WP1 - D3 Populated Data Stores	To provide data to the analysis engine which will be used to test the algorithms for the various use cases.
WP1 - D4 Selected Algorithms Report	To describe which algorithms were considered for the use cases and confirm which were ultimately selected and why. These algorithms, along with the evaluation metrics and calculations, are documented within the issued reports to allow for independent quality checking and to support integration into BAU by NGED. The independent quality check it was carried out by GHD, and they confirm that the report meets their needs as part of the approval process.
WP1 - D5 Data Quality Assessment / Go - No Go decision.	This was a check of the data using samples available at the time and with an understanding of the likely data requirements of the algorithms selected for WP1 D4. This decision point was ahead of starting work on the activity to validate the network using the HAYSYS equipment so that had the project been unlikely to be able to achieve the required outcomes due to data quality then it could have been changed or closed down so that further expense would have been limited.
WP1 – D6 Validate data processing architecture	This was an independent quality check to ensure that the algorithms set up in the Microsoft Azure environment were operating as expected and that the processes to evaluate the accuracy of the results from the various algorithms was also working correctly. The documented algorithms and processes in WP1 – D4 will provide the specification of how the algorithms and processes were expected to operate.
WP1 – D7 Validated test network Report & Dataset	Following on from a positive decision in WP1 D5, this gives confirmation of the phase to which customers in the test network are connected so that the values suggested by the phase identification algorithm can be evaluated for accuracy. This report also captures practical issues around the validation activity that can be used to improve the feeder validation activity later in the project.

5.2 WP2 – Phase identification

Identifying the phase to which LV customers are connected enables a better understanding of phase unbalance. LV networks can be modelled more accurately using knowledge of the actual phase unbalance rather by than assuming that the demand is fully balanced. Accurate phase records allow us to plan new connections to minimise additional losses and to report single phase faults more accurately for IIS (Interruptions Incentive Scheme) purposes.

This work package involved the use of voltage data from smart meters to identify customers connected to the same phase. The techniques involved correlation and clustering methods and were tested using several variations of the methods:

- Including voltage data from existing monitoring at the distribution substation where available (Please note - even with plans to monitor additional sites in ED2, monitored sites will still be a small minority of distribution substations due to the costs.)

- Considering either absolute voltage data or differential changes in voltage from one sample to the next
- Using voltage values that were averaged over different time periods i.e., 30-minute average and 1-minute average.
- Using different sample sizes of data to determine how sensitive a particular method was to when the sample was taken or the number of data points available.

Deliverable

Deliverable	Purpose
WP2 D1 - Phase Identification Report	This documents the learning from applying the selected algorithms to the data for the phase identification use case. It includes an assessment of overall accuracy and the factors that affect accuracy. This report also shows how the voltage data can best be configured to enable phase identification. E.g., what time resolution, duration of data required, whether samples should be contiguous or cover a wide range of dates, what is the impact of smart meter time interval synchronisation errors

5.3 WP3 – Combination load profiles for planning

If all customers on an LV feeder had a smart meter and if currently unmetered loads such as street lighting were also metered, then we would be able to aggregate meter readings to get a good estimate for the whole LV feeder load in any half hour.

This work package tackled the issue of how to estimate the half hourly load for customers without smart meters so that an aggregated value can be provided. By determining the most accurate approach, this can then be used to generate profiles for different LV feeders and substations as if they had monitoring installed. Loughborough University had carried out some work previously in this area by scaling the aggregated smart meter data up to a total value based on the number of customers with and without smart meters. They also found that scaling the data by the proportion of annual demand covered by the smart meters gives better results than scaling simply by the number of customers.

Aggregating similar customers in the same locality but on different LV feeders or substations required an amendment to our data privacy plan, which previously relied on aggregation of either an entire LV feeder or sub-sections of an LV feeder.

However, there were still further opportunities to refine this approach e.g.

- By scaling the demand from customers on the same feeder for each profile class separately.
- By adding demand based on Elexon profiles, scaled by the Estimated Annual Consumption for each customer.
- By creating a wider set of profiles for customers that include LCTs based on customers with smart meter in the local area, but not necessarily at the same substation.

The first of these approaches is likely to be appropriate if all the houses on a feeder and with the same profile class are similar (in construction, as well as in occupancy and uptake of LCTs). The available smart meter data is then considered as a randomised sub-sample of the customers on the feeder and the aggregated smart meter demand is then scaled accordingly. This approach is less useful where there is a mix of customer types and/or property types.

In all cases, the load estimates also needed to make an allowance for the output from embedded generation and load from any traditionally half-hourly metered customers, i.e., part of the original 100 kW market before smart meter roll out, that are LV connected. Any behind-the-meter demands for LCTs connected to the LV network are included in the aggregated smart meter data, although not in the Elexon profile data representing the demand for non-half-hourly metered customers.

After reviewing previous work, the contribution from unmetered supplies and losses was deemed to be below the level of estimation accuracy and so these aspects were omitted.

Monitoring data was available for each of the distribution substations in the test network such that this could be used as a reference for validation of the estimated load profiles and to determine their accuracy. The scope of this work package was focussed on determining which approach gave the most accurate results but also included giving a high-level recommendation on how the load profiles generated can be converted for use in BAU LV planning software such as ConnectLV.

Deliverable

Deliverable	Purpose
WP3 D1 – Combination load profiles for planning Report	This documents the learning from applying the selected algorithm to the data for the use case to create combination load profiles for distribution sub-stations and feeders. It includes an assessment of overall accuracy and the factors that affect accuracy. It also compares the results to the pro-rating approach previously applied by Loughborough University.

5.4 WP4 – Voltage detection of unregistered equipment

Different methods were found to be appropriate for the detection of electric vehicle (EV) chargers and for solar PV. The number of heat pumps and battery storage systems in the test network area is still very low and so the LCT detection algorithms were focused on EV and PV installations.

The relationships between voltage alarms and LCTs were initially investigated to determine whether this can be used to identify unregistered equipment. While some rudimentary patterns were observed this seemed to indicate that voltage alarms were produced by combinations of circumstances rather than LCTs alone, therefore investigations then switched to consider demand and time-series voltage data for each MPAN rather than alarms that are triggered only on limited occasions.

Export data from each MPAN is considered to be operational data since it largely relates to the solar irradiance rather than the personal demand and is available at half-hourly resolution without aggregation over the customers on each LV feeder. This data has been used to indicate the presence of PV installations when exports occur during daylight hours and with greater frequency during the summer months.

Monthly demand data for each MPAN with a smart meter is considered to be sufficiently aggregated (over time) that it can be used by DNOs for operational purposes. This provides a slightly greater level of demand visibility than the Estimated Annual Consumption (EAC) data for each customer provided by energy suppliers.

The dataset available to the project was not of sufficient duration to enable long-term detection of step changes in demand as LCTs are installed so the analysis aimed to identify features of the monthly demand data that are characteristic of MPANs with LCTs connected. The algorithms compared a set of features for the set of customers known to have electric vehicles (EVs) with the wider set of customers not known to have EVs. Strictly this second set of data must be considered as having unknown LCT connections as some customers may have unregistered EV chargers.

Clearly it is not straightforward to distinguish MPANs with EVs from those without based on monthly demand data alone as some customers may have an EV charger but use it very rarely, and other customers without EVs may also have very high demands. Any detection algorithm therefore makes a compromise between having an under-sensitivity where actual EV installations are not detected, versus an over-sensitivity with a high proportion of false detections. Three machine learning algorithms were tested (weighted logistic regression, weighted XGBoost, and weighted support vector machine), all using a set of 21 features derived from the monthly demand data.

Further work has investigated whether the voltage drops to each customer can be used to identify EV charging sessions. Substation monitoring can be used as a reference where this is available, but otherwise the analysis can consider the voltage at each MPAN relative to neighbours on the same feeder and phase.

The voltage drops between an MPAN, and its neighbours are smoothed to remove short-term variations. An EV charger is detected if the algorithm find voltage drops that are greater than a defined threshold and with a sufficient length of time.

Future investigations could consider an algorithm including both the demand and voltage data in a combined method.

Deliverable

Deliverable	Purpose
WP4 D1 – Voltage validation of unregistered equipment Report	This documents the learning from applying the selected algorithm to the data for the use case to identify unregistered equipment e.g., EV charge points, heat pumps, battery storage or PV generation. It includes an assessment of overall accuracy and the factors that affect accuracy. It also compares the results to those from the LCT detection NIA project.

5.5 WP5 – Customer to Feeder association validation

For recent LV feeder installations, the service cables between customers' premises and the mains cable are now generally included in the network connectivity GIS data. However, for older feeder the service cables are mostly not recorded. Instead, the substation and LV feeder for each customer is recorded in the CROWN asset database. Although this data is rarely missing, it has been found that some proportion of this data is incorrect. This was a finding of the Losses Investigation where it was necessary to scrutinise LV connectivity models carefully to ensure that the measurements of power in and out of LV networks was calculated accurately. Where network records do not include services, and there are multiple LV cables running in parallel down a street, it can be particularly difficult to determine which feeder each customer is supplied from. Errors in these records can cause problems during fault finding, and lead to the incorrect set of smart meter MPANs being aggregated together to represent the demand of each distribution substation or LV feeder.

Having determined that the validation of customer-feeder association should be one of the SMITN use cases, there was a need to provide accurate data against which to evaluate algorithm results.

This work package involved the development and testing of new equipment by HAYSYS to confirm customer to LV feeder mapping without requiring access to connections within the customer's property. The Feeder Finder injects signals onto the LV feeder at the distribution substation which are then detected on the service cable leading to the customer's property. Setting up the Feeder Finder does not require an interruption to customer supplies but does require access close to the LV cables.

The Feeder Finder tool was validated by comparing results at sites at sites both in Cardiff and on the Milton Keynes test network, where the customer-feeder data was known to be correct. The tool was then used to survey selected sites from the test area as recommended by Loughborough University and CGI.

This data then allowed CGI to carry out their analysis of algorithms to predict the feeder that a customer is connected to, in order to identify potentially incorrect values in our data. This use case had some overlaps with the phase identification work as incorrect feeder allocation could result in an additional cluster appearing when using clustering to determine phase groups, although it was found that feeder identification is less straightforward than phase identification.

CGI has tested a number of algorithms including:

- "Outliers from Phase Identification", where meters with very low correlation result are pinpointed for review,
- "Correlation with neighbouring substation", a technique that re-calculates the correlations between the meters from two neighbouring substations,
- "Clustering smart meters into feeder groups", an algorithm that uses clustering techniques to create groups of meters for each feeder,
- "Voltage correlation with aggregated feeder demand", an algorithm that correlates the voltage differences with the aggregated load on each feeder.

- “Discrepancies between Electric Office (EO) and CROWN”, where the derived feeder or substation from EO differs from the data recorded in CROWN,
- “Feeder Profile Outliers”, an algorithm that identifies feeder allocation errors by comparing the difference between predicted demand and the measured demand, where substation monitoring is available.

The analysis by CGI found feeder allocation errors in the database occur on most feeders, with a significant number of errors due to numbering swaps.

The use of voltage data is promising in the identification of outliers and misallocations, whereas voltages for connections near the substations were similar on each feeder meaning the clustering approach used couldn’t easily separate them.

In the future it would be interesting to investigate the integrated use of a combination of algorithm types (ensemble modelling) to increase the probability of data accuracy.

Deliverables

Deliverable	Purpose
WP5 D1 Feeder Finder tool development & validation report	This report documents the development work by HAYSYS to create a Feeder Finder tool. This outlines the design of the tool and how its accuracy has been validated.
WP5 D2 – MPAN to Feeder Association Validation	The report documents the results of the work by HAYSYS to validate the association between customers and LV feeders for selected locations to provide a known value against which the accuracy of the selected algorithm can be compared. The report includes the results dataset and any learning relating to the use of the tool in practice.
WP5 D3 - MPAN to Feeder association analysis	This documents the learning from applying the selected algorithm to the data for the use case to identify customer to feeder mappings. It includes an assessment of overall accuracy and the factors that affect accuracy.

5.6 WP6 – Reporting and Dissemination

The project concluded by distilling and disseminating the learning from the project. Dissemination took place in May via webinar. The webinar slides are available on the SMITN project website.[4]

Deliverables

Deliverable	Purpose
WP6 D1 Closedown Report	This document brings together the learning from the other reports and provides the information required for the project closedown report.
WP6 D2 - Dissemination complete	This includes the preparation of dissemination materials and the delivery of dissemination events to share the learning from the project. This is to ensure value-for-money of innovation projects by enabling the benefits of this project to be replicated elsewhere and preventing duplication of effort.

[4] https://www.nationalgrid.co.uk/innovation/projects/smart-meter-innovations-and-test-network-smitn

6 Performance Compared to Original Aims, Objectives, and Success Criteria

6.1 Objectives

Objective	Status	Performance
To determine a representative test network of selected distribution substations and validate the key features of this network by carrying out surveys	Complete	These were completed as part of the project deliverables. The selection of the test area was updated during the project as required. The phase survey work was delayed, but provided the data required and added value by capturing images for future use.
To capture smart meter data using new aggregation groups	Complete	The algorithms demonstrated new uses of disaggregated export data for detection of connections with PV generation. A process was also demonstrated for obtaining disaggregated demand data for commercial smart meter (other those for traditionally half-hourly metered customers).
To develop algorithms using smart meter data for the following use cases: a) Customer to Phase connection prediction b) Customer to Feeder connection prediction c) LCT identification of potential locations and types d) Provision of LV feeder and Distribution substation planning profiles for use in network planning	Complete	The algorithms were documented in the algorithm selection report following the workshop with external stakeholders. This report was then reviewed by GHD to ensure this was documented clearly enough to support their validation work.
To apply the algorithms for data relating to the test network	Complete	The data was used for all the use cases as documented in the selected algorithm report. Additional data for MPANs outside of the test network area has been used for the LCT detection work to broaden the sampling of connections with known LCTs.
To assess the performance of the algorithms and where possible identify the factors that affect accuracy	Complete	This has been achieved. The results have been captured in the individual use case reports.
To capture the learning from the project and disseminate this to interested parties	Complete	The learning from the project has been summarised in this closedown report as well as being shared via the dissemination webinar. It will also be presented at CIRED.

6.2 Success Criteria

Success Criteria	Achieved	Performance
A set of phase-validated networks and a subset where customer-to-feeder association has also been validated that can be used to test future innovations beyond this project.	✓	The HAYSYS phase survey provided phase data for 8,790 customers. Additionally, the feeder survey confirmed the feeders to which 125 customers were connected. This has been supplemented with additional information where LCT was visible for the customers being surveyed and also photographs within external meter boxes which may be useful for future development of AI image recognition tools.
An understanding of how well the smart meter data can be used to support phase identification and what factors affect the accuracy of the algorithms used.	✓	The analysis has been successful at showing the results from two different approaches and showing the impact on accuracy of time-series length. Recommendations to maximise accuracy are included in the Phase Identification report.
An understanding of whether smart meter data can be used to support the identification of unregistered LCTs and what factors affect the accuracy of the algorithms used	✓	This has been achieved with the results being documented in the LCT detection report. The analysis found that PV installations can reliably be detected from the active export energy data which can be utilised without aggregation. Electric vehicle chargers can also be identified but with a lower confidence level. The trial area had insufficient numbers of customers with heat pumps to allow for these to be distinguished from more conventional electric heating.
An understanding of whether smart meter data can be used to support the validation of customer-to-LV feeder associations and what factors affect the accuracy of the algorithms used	✓	This has been successfully delivered with the results recorded in the Feeder Validation report. Incorrect feeder data was found using a combination of techniques. e.g., distance from MPAN to nearest feeder, correlation, clustering of voltage data with other customers, voltage profile analysis etc. This also found a surprising number of customers associated with the wrong substation
An understanding of how well aggregated smart meter data can be combined with other data available to DNOs to create estimated LV feeder load profiles and what factors affect the accuracy of these estimates.	✓	This analysis successfully compared the use of Elexon profiles to local profiles based on similar customers with smart meters. The key requirement is to estimate the demand for customers with conventional non-half-hourly meters. It was found that these estimates are more accurate if based on generic profiles, scaled by EACs, rather than by assuming that it can be calculated by scaling the demand from smart meters on the same feeder or substation.
An understanding of the practical issues for using the newly developed feature for validating customer-to-LV feeder association without requiring connection to the customer's wiring and confirmation that the new feature is fit-for-purpose	✓	This was achieved and the project discovered the practical issue around lampposts acting as earthing spikes and reducing the strength of the signal and generated a solution via signal boosting. The feeder finder was validated using two substations and seen to be operating correctly in most cases. There are some locations where further investigation of the cable routing may be needed to confirm the results.

Success Criteria	Achieved	Performance
An increased insight of the phase unbalance on the test networks	✓	This has been achieved by the provision of the phase data from the survey. Results show that the phase identification algorithm and the phase survey agree very closely, provided that the substation reference phase for the survey has been configured correctly.
A view of the value of disaggregated data from smart meters rather than the default position of aggregation at the LV feeder level	✓	This has been achieved by using different aggregation groups within the analysis. The NGED customer privacy plan has been updated to include a wider range of permissible customer aggregations. Disaggregated half-hourly export data has been essential in identifying the locations of PV installations. Disaggregated half-hourly import data is not available. If access to this data had been possible, the accuracy of the feeder identification and load profiling algorithms could be improved. Without this disaggregated data it is not possible to re-combine the demands of individual smart meters to correct the aggregation groups, which can only take place through a much longer process in which the asset database is changed, followed by repeat collection of the smart meter data. The associated timescales prevent this from being used in a closed-loop algorithm.

7 Required Modifications to the Planned Approach during the Course of the Project

7.1 Trial area selection

As outlined in section 5 Details of the Work Carried Out, the selected area was updated to focus on HV feeders that had significant monitoring installed, but also to reflect the technical limitations such as meters that were not well suited to providing the required data and substation monitoring that was not reporting back data correctly.

7.2 Substation monitoring data

The availability of substation monitoring data was generally good, although some site works were needed to reset loggers that had stopped reporting to the data centre. Monitoring data was downloaded from the data centre by accessing an Application Program Interface (API) using custom software developed by Loughborough University for this project. Owing to issues with the data centre indexing, this software needed to request multiple extracts that were then combined to complete the dataset. In general, the substation monitoring has performed well and provided reliable reference data for the SMITN test network.

7.3 Phase survey/ Feeder survey

The phase survey underwent changes reflecting staffing availability. It was difficult to provide the staffing required to support the phase and feeder survey due to network services focussing on delivering our work plan before the end of the price control period that completed in March 2023. Eventually support was sought from external contractors, however there were delays in procuring suitable staff and ensuring they had the right authorisation for the work.

7.4 Phase anomalies

During the work on phase identification, some unusual results were detected that called into question the accuracy of the results of the phase survey. Given the importance of the accuracy of this data, site visits were arranged to determine whether the anomalies were caused by other factors. A group of several substations were found to have reversed phase rotation, thought to be due to a conductor cross in the 11 kV overhead lines. As a result of this crossing, the constellation of voltage phase angles on the LV side of the transformer has a 60° offset from the phase angles with the conventional rotation sequence. The phase survey results from these substations were therefore unreliable. This highlights the importance of the phase survey having a reference phase angle that is set according to the same rotation as for the MPAN connections being tested. In practice, setting the reference on the Phase Finder tool at each substation to be surveyed, could help to reduce the risk of a cross on the HV network impacting the survey results.

7.5 Data processing infrastructure – build, data population and validation.

Infrastructure build and data population phases were less discrete and more fluid than in the original plan. This meant that not all of the infrastructure build work was complete before the data population started and that this work was approached in a more agile manner to reflect the availability of the required data. The alteration of the timing had no negative impact on the quality or cost of key deliverables.

7.6 Smart Meter data provision

The process to establish new data requests from the smart meters was more complex than anticipated, and some requests required managerial sign off. While this slowed progress, this was necessary to ensure that the requests did not overload the system.

7.7 Other data provision

It was originally expected that the provision of some of the key datasets could be automated, however this was not the case. This resulted in more of time being required for NGED Project Management to carry out the data extractions and upload data onto the SFTP site.

Additionally, there were some minor changes to project plan dates that were handled during the project using the normal change control process.

8 Project Costs

Table 8-1: Project Spend

	Budget (£)	Actual (£)	Variance (£) (+ underspend, - overspend)
Main Contractors – CGI, Haysys and Loughborough	713,532	713,532	0
Quality Assurance Contractor GHD.	25,000	23,456	1,544
Third party data provision	25,000	14,434	10,566
NGED Network Services + Power Projects Ltd Support for the HAYSYS surveys	37,710	11,253	26,457
NGED project management	30,368	54,140	(23,772)
Smart Grid Consultancy	0	11,224	(11,224)
Total Before Contingency	831,610	828,039	3,571
Contingency @10%	83,161	0	83,161
Project Total	**914,771**	**828,039**	**86,732**

The project is within budget overall with some areas of overspend balanced by other areas of underspend.

Project management is overspent due to additional project management time being required during the project, especially in relation to data provision, which became a routine monthly task.

The difficulty in securing Network Services staff to support the feeder survey resulted in the recruitment via Power Projects Ltd. These costs replaced those that would have been incurred by Network Services therefore this has not triggered contingency spend, but has actually still come in under the expected budget.

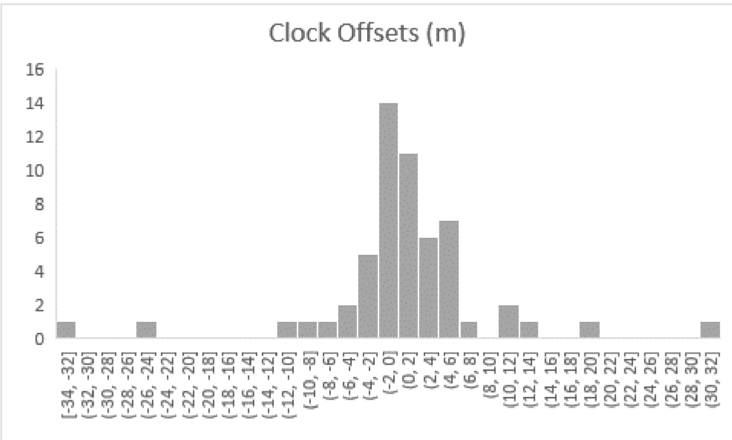
The costs for data provision were less than the expected budget. This has helped to balance out the project management overspend so that the project is marginally underspent overall before contingency. When contingency is included the project is approximately £87k under budget.

9 Lessons Learnt for Future Projects and outcomes.

A summary of the key learning points is given below. Further details of the learning were included in the dissemination event. The slides and video for that event can be viewed on the SMITN webpage.[5]

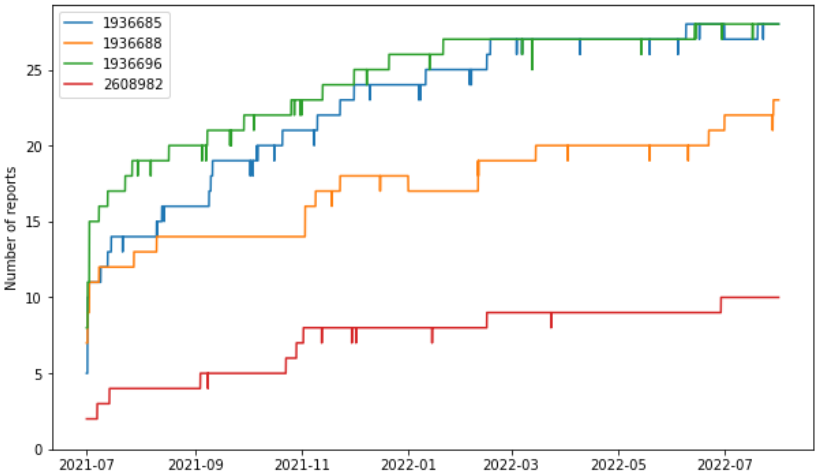
9.1 Data Quality and Availability

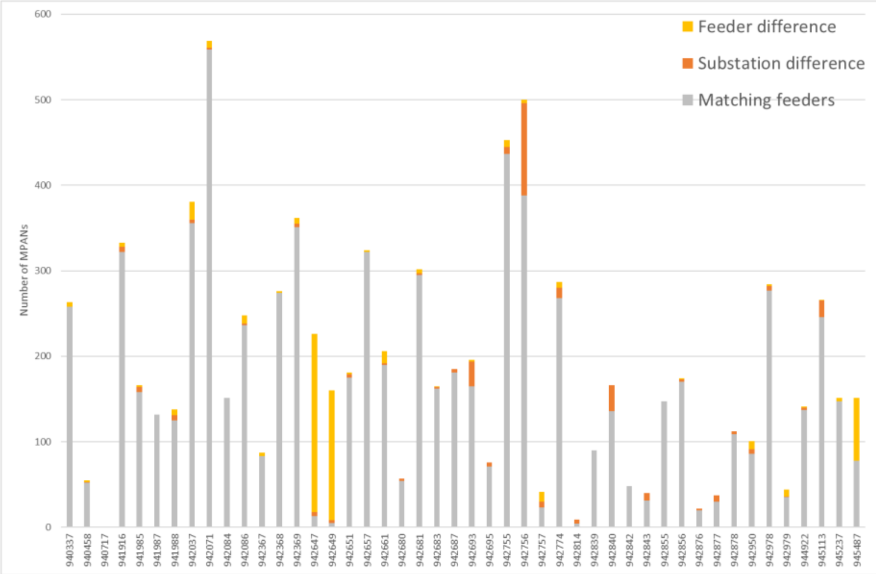
As SMITN was one of the first projects making substantive use of smart meter data, some limitations were discovered around the retrieving the data. The project also uncovered some previously unknown issues with NGED's legacy data.

Item	Learning Detail
Smart meter MPAN coverage	• Around 47% of the 8,809 MPANs in the trial area had smart meters installed.
Half-hourly voltage data availability	• The smart meters were initially configured to record voltage data with half-hourly time resolution. • Initially, half-hourly voltage data was requested from the devices for the period 1st May 2022 until 31st August. Around 2,200 devices gave a response out of a population of 4,140, which gives a success rate of 57%.
1 minute voltage data availability	• When the 4,140 smart meters were sent a message to configure them to record 1-minute voltage data, only 1,800 gave a positive response, i.e. under 50%
Voltage data MPAN coverage	• Combining these two issues, 1-minute voltage data was only available for 43% of the full set of customers with smart meters, and so only around 20% of the full set of customers. While the coverage of smart meters is expected to increase as the roll-out progresses, further action is needed to resolve the voltage reporting issues and this would significantly improve the utility of the SMITN phase and feeder identification algorithms.
Clock synchronisation	 • The timing of the smart meter voltage data was generally well-aligned to the substation monitoring data. Only 56 SMs out of 1810 found to have a clock synchronisation issue relative to substation logger. It is unclear from this data if the offset is introduced from the SM or from the substation logger. These time offsets could be corrected by correlation method relative to the substation monitoring data.

[5] https://www.nationalgrid.co.uk/innovation/projects/smart-meter-innovations-and-test-network-smitn

Item	Learning Detail
Voltage measurement interval synchronisation	• Separately from the issue of clock synchronisation, many of the smart meters used randomised starting times for the voltage measurement intervals. Half-hourly voltage data could be recorded starting at any time within a clock half-hour, rather than at 00 and 30 minutes past the hour. Successive measurement offsets were consistent, such that the data still provided valid time-series data, but each meter would have a different time offset. Similarly, 1-minute data samples could begin at any time within the minute and not at 00 seconds. • Reducing the measurement averaging intervals from half-hour to 1-minute therefore provided both an improved time resolution and an improved alignment of the measurements between the smart meters. • This pattern was observed on meters from multiple vendors, and it is unclear why this approach has been adopted for the voltage data. Demand data measurements are consistently recorded for half-hour periods starting at 00 and 30 minutes.
Data provision timescales	• Even though the project had anticipated that getting hold of data would take some time, the complexity of data extraction from smart meters had been underestimated. The necessity for some extracts to be approved internally due to their possible impact on other work areas suggests that the roll out into BAU will be similarly complex due to the potential impact of introducing large scale data retrievals.
Smart meter uptake vs LCT uptake	• Customers with LCTs are more likely to have a smart meter than customers who do not. Including the PV, 4% of MPANs in the trial area have an LCT. The probability of having a smart meter, given that there is an LCT, is 50%, compared to 42% for customers in general, so the conclusion that smart meters are more likely at customers with an LCT still holds.
Clock Change	• We investigated using monitoring data from primary substations that would provide HV feeder currents and primary substation busbar voltages. This data could be useful as a voltage reference where distribution substation monitoring is not available. However, this data was not used in later development of the algorithm. Half hourly primary substation data is available as Historical Analogue (HISTAN) files. Time Series Data Store (TSDS) provides another way to access primary substation monitoring data and is often available at higher resolution than half-hourly. The HISTAN data appears to have an issue around clock change where it is going the wrong way at the wrong time based on the March data for 2022 and 2021 which is looping back between midnight and 23:00 rather than jumping forward an hour between 02:00 and 03:00
HISTAN data reliability	• The average of TSDS raw data over 30 minutes are more reliable than the HISTAN half-hourly data in terms of agreement with GridKey units at primaries. HISTAN data was abandoned in favour of data from TSDS.
Visibility of tap change impacts	• The correlation between busbar voltage and tap changer position is not as clear as expected and while it had been hoped that tap change operations could be used as a way to filter out related voltage changes on smart meters making the picture of local voltage impacts clearer, this may not be the case.
Monthly consumption data outliers	• The monthly consumption data being provided by the smart meters for individual customers showed a number of customers having unusual values. Therefore, some data cleansing and sense checking was required for this data item.
Seasonal consumption variation	• The monthly customer demand data also provided a means to compare seasonality in the data. It is clear that some customers consumption increases significantly over the winter months but other profiles are relatively flat.

Item	Learning Detail
1-minute average voltage data retrieval	• The data processing and telecommunications burden of retrieving 1-minute voltage data is significant and this may suggest that roll out of BAU covers one area at a time rather than trying to scale up to cover retrieval of all voltage data at once. Once the algorithm has been applied then the smart meters can be reconfigured back to their default state.
Increase in the numbers of smart meters over time	 • The aggregated consumption data for four different LV feeders is given above. All the totals trend upwards which reflects additional smart meters being added in to the aggregated totals over time. In addition to this trend, there are also short-term variations for some half-hours when one or more of the meter do not provide demand data. These variations have been taken into account in the profile estimation methods where smart meter demands are scaled appropriately for the number of meters reporting in each half-hour period..
VisNet and GridKey unit availability	• The System Overview reports for iHost suggest that there are large numbers of monitoring sites for both VisNet and GridKey monitoring units that are not functioning correctly.
Nearest feeder filter	• Data from Electric Office was used to suggest the nearest LV feeder to MPANs. This was then compared to the distance to the second nearest LV feeder. Where there was a large difference between the nearest and the second nearest LV feeder there could be more confidence that the nearest feeder would be the one that supplied the MPAN. If this disagreed with the CROWN data then it could be substituted. • Where the second nearest feeder was not very much further away the same substitution could not be made as it may be that the slightly further away feeder was the one to which the customer was connected.
CROWN feeder data	• Analysis based on the proximity of customer locations to feeder mains suggested that incorrect feeder association (yellow on the chart) is common and that customers being associated with the wrong substation (red on the chart) is not uncommon.

Item	Learning Detail
	• 

9.2 Trial area selection

Item	Learning Detail
Site requirements do not coincide	• The data requirements for each use case were different and finding sites that were suitable to support all the use cases was not possible. This resulted in the selection of a core area supplemented by a periphery trial area that was used to provide the sites with the right features that were not within the core area. This was an iterative process that involved several revisions as further data became available.

9.3 Phase/ Feeder Surveys and Feeder Finder Development

Item	Learning Detail
HV phase cross	• There were a number of customers where the Phase Finder appeared to give incorrect results. On investigation this was due to a conductor cross on the HV network which resulted in the phase reference taken for the survey being incorrect. As a result of this crossing, the phase angles on the LV side of the transformer have a 60° offset from the phase angles with the conventional rotation sequence. The phase survey results from these substations were therefore unreliable. Furthermore, since the phase angles are rotated by 60° rather than a multiple of 120°, there is no simple mapping between the recorded phase and the actual phase than can be applied retrospectively to correct the data. • As a simple working method, the phase reference should be set afresh for each new substation in the phase survey. • In practice the set of customers connected to a substation is not always known correctly in advance, and the surveyors will tend to work systematically along streets without necessarily knowing that some addresses are served from a different substation. • Experience from the test network suggests that the substations with phase reversals are in isolated groups where there is no possibility of interconnection via a link-box to another substation with differing phase rotation. It is therefore sufficient for the phase reference to be reset whenever the survey moves into a new section of LV network

Item	Learning Detail
	where the phase rotation could potentially be different. However, the method of setting the reference for each new substation is a more easily applied working practice.
Signal attenuation around earthing spikes	• For the HAYSYS Feeder Finder, initially steel lampposts were acting like earth spikes and damping the signal so that the range of the detector from the injector (at the substation), was reduced.
Cable detection	• The Feeder Finder would detect the LV cables while the user was walking between test locations, acting as if it were a cable detector. This functionality may be a useful addition even though it was not expected. • The Feeder finder was able to distinguish between four circuits that were connected in parallel. • It was difficult for the Feeder Finder to determine which feeder was connected to a property where the service cable was very short, i.e., where the property front was on the pavement with no front garden and there were multiple feeders buried beneath the pavement.

9.4 Phase identification

Item	Learning Detail
Viability of the technique	• Voltage data from smart meters can be used to effectively identify the phase to which each property is connected, with accuracy of up to 100% of smart meters for some feeders when validated against the phase survey. • Errors in the phase identification method can arise when the group of smart meters is poorly defined, for example where CROWN records include MPANs from other substations. This leads to smart meters being added to the incorrect clustering groups or being correlated with the incorrect substation voltages. • Some errors have also occurred in the validation data where the phase reference was not set correctly for each substation.
Comparison of approaches	• The approach where voltage data from smart meters are correlated with voltages existing substation monitoring had 83% mean accuracy, slightly higher than the 80% mean accuracy using clustering approaches that were applied where substation metering is not installed. • A limitation of the clustering technique is that the actual phase label is unknown and other data sources need to be used to label the resulted classes with the corresponding phases. • Correlations using the magnitude of the step changes of voltages (i.e., the difference of each sample from the previous sample) works more efficiently than using the time-series voltage data as input in the algorithms
Use of 1-minute vs 30-minute average voltage data	• 1-minute data provided better results for both approaches and was less sensitive to the duration of the data sample. i.e., good results could be gained with just a week's data compared to several months of 30-minute resolution data. • Analysis 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, and this is an area for further investigation.
Managing clock synchronisation issues	• Both linear interpolation and rounding gave similarly effective methods of resolving issues around clock synchronisation.

Item	Learning Detail
Accuracy assessment	• As well as an accuracy metric, the Rand index was found to be useful where there were issues with label accuracy.
Best correlation approach	• Using 1-minute step change data correlated with the substation metering data gave the best results with 56% of the substations having an accuracy more than 80% (i.e. more than 80% of the customers for the substation were identified correctly) • The Fisher-Z transform, applied to the correlation results has been tested to determine whether clustering is improved. Results suggest not.
Best Clustering approach	• Hierarchical clustering worked better than K-means. • 1-minute resolution step change voltage data worked better than the actual voltage values themselves. • Where the best approach was used 51% of the substations had an accuracy over 80% (i.e. more than 80% of the customers for the substation were identified correctly)
Looped services	• This provided an additional way to validate the phase information as both properties on a looped service would be connected to the same phase. It also allowed phase data to be ascertained where only one customer in the pair had a smart meter.
Existing phase records	• Where phases are already recorded in the Electric Office database, they were found to be only 50% accurate compared to phases determined from the correlation methods or from the HAYSYS survey.
Impact of number of smart meters	• Results did not show any clear trend in relation to the number of smart meters per feeder, with good accuracy even for small groups of only 6 meters However, the clustering method does require at least one smart meter in each phase such that three groups can be formed.

9.5 Feeder identification

Item	Learning Detail
Comparison of approaches	• Substation and feeder identification methods have generally required that the phases are already known correctly. • This highlights the need for an iterative approach as phase identification also requires that the substation and feeders of each MPAN connection are known. • Feeder identification is less straightforward than phase identification as the differences, identified through voltage correlation, relate only to the loads on the LV feeder itself, and smart meters on all feeders that are close to the substation busbar will be very similar.
Phase correlation outliers	• Smart meters that are allocated in CROWN to the incorrect substation may be identified as clustering outliers in the phase identification method. However, the correlation of smart meter voltages with those on neighbouring substations can be high where both substations are on the same HV feeder. • The proposed method therefore selects allocation to the substation that has the higher phase correlation, rather than using a particular threshold to categorise outliers. • This method can be applied either with or without substation monitoring. In the absence of monitoring data, the comparison uses smart meters from the two substations.

Item	Learning Detail
Feeder correlation methods	• A voltage correlation and clustering method has been developed to assign smart meters to feeders. • Results mostly agree with those from the feeder survey, and differences may relate to survey issues rather than errors in the correlation method. • A second method using correlation between smart meter voltage drops and feeder load data gives results in 100% agreement with the voltage clustering method. This second method uses substation monitoring data for the feeder loads so cannot be deployed on all substations but gives confidence that the voltage correlation results are accurate.
Comparison of load profiles	• Load profiles can be estimated for LV feeders using either the set of MPANs listed in CROWN or those with closest proximity to the mains feeder in the Electric Office network diagram. Feeder allocation errors can be detected where the load profile based on the network diagram has a lower error relative to substation monitoring than the load profile based on CROWN feeder allocations. • Although this method cannot identify individual MPANs with incorrect allocation, it is possible to identify where groups of MPANs have the incorrect feeder.
Use of 1-minute vs 30-minute voltage data	• Given that feeder identification is more challenging than phase identification, the methods have used 1-minute voltage data throughout, minimising any issues due to measurement interval synchronisation and increasing the sensitivity to short-term voltage differences that are expected to affect all meters simultaneously.

9.6 LCT detection

Item	Learning Detail
Smart meter take up	• Customers with LCT are more likely to have a smart meter than customers who do not. Including the PV, 4% of MPANs in the Losses area have an LCT. The probability of having a smart meter, given that there is an LCT, is 50%, compared to 42% for customers in general, so the conclusion that smart meters are more likely at customers with an LCT still holds.
Voltage alarms	• These were assessed for their potential to indicate unregistered LCT. It was found that while customers with LCT were more likely to have voltage alarms the data was not able to provide conclusive relationships. • Alarms were seen to happen more often on or around the start of an hour suggesting they were being triggered by the automatic operation of other devices on a timed schedule e.g., a customer with an EV charger might start charging at any time of the evening but the low voltage alarm was generated at 1am, triggered by the switching on of a concurrent load for an electric storage heating system.
PV detection	• Half-hourly active export readings are a good indicator of PV installations. In the test network area 94% of known PV installations can be identified by having non-zero export data. • The remaining 6% of MPANs with known PV are not detected by this method but are also of lesser consequence to the network operation, by definition since there are no periods of net export. It is also possible that some of these undetected installations are no longer operational. • As expected, MPANs with PV have net exports peaking during daylight hours. A small number have net exports throughout the night, with one of these known to have energy storage. However, there are still too few locations with energy storage to enable these to be well characterised in terms of exports.

Item	Learning Detail
EV detection from monthly demand data	• Although the existing LCT data can be used to create a set of MPANs with known LCTs, there is no equivalent negative set of customers without LCTs. The remaining customers must therefore be classed as being either positive or unlabelled. • Customers with EVs tend to have higher monthly demand than those without, but the probability distributions are overlapping and so detection of EVs based on a threshold will inevitably have some proportion of false positive and false negative classifications. • A simple technique has been tested by comparing the monthly demand for each MPAN to the median monthly demand for customers with EVs and those unlabelled. This has a very high rate of false positives, 24% of the total number of MPANs. • Better results can be obtained using machine learning algorithms for binary classification. These are trained using multiple features of the 12 months of demand data, aiming to achieve a higher detection rate without unduly increasing the number of false positives. • A Weighted Support Vector Machine method was found to give the best detection rate while maintaining a false positive rate below 10%. The summer consumption and the maximum consumption features have higher impact on the models.
EV detection from voltage data	• Voltage data looks promising to identify EVs due to big drops they caused in the voltage. • Many properties that have not been recorded as EVs have been identified by the algorithm to have voltage drops with high duration. This can be explained by the fact that other large loads with long duration could cause the same voltage impacts as EVs. However, we could use this approach to validate and reduce the number of the unlabelled properties that have been predicted as having an EV from the "EV detection using monthly demand data" approach. • A Machine Learning model could identify which properties have EVs based on frequency of drops, average drop lengths, and percentage of other Smart Meters on the same feeder that the algorithm detects drops relative to.
Heat pump detection	• There was insufficient representation of MPANs with heat pump in the sample data to support development of a detection algorithm.

9.7 Planning profiles

Item	Learning Detail
Feeder / substation association accuracy	• Problems with customers being incorrectly associated with the wrong LV feeder or substation can introduce inaccuracy into the estimates for LV or substation profiles. In one example a half-hourly metered customer was attributed to the wrong LV feeder which resulted in inaccurately high and low estimates on the real and assumed LV feeder.
Profile methods	• Given that the aggregated demand for smart meters and for traditionally half-hourly meters MPANs is known, the accuracy of the estimated load profiles depends on the process used to model conventional non-half-hourly meter demands. • Load profiles for Non Half-Hourly NHH meters calculated considering daily temperature compensation and Time Pattern Regimes (TPR) have a higher accuracy than those using the set of 15 generic Elexon profiles. • An alternative approach to estimating the demand of NHH meters is to scale the demand from smart meters located on the same feeder, treating the domestic smart meters as a proxy for the demand of the neighbouring domestic NHH meters.

Item	Learning Detail
	• Load profiles estimated by scaling smart meter data were more accurate if scaled according to the total annual demand of the full set of domestic meters on the feeder, rather than according to just the total number of meters. • Both the approaches using load profiles calculated from generic profile data were more accurate than the two approaches using scaled smart meter data to represent the NHH meters. Results were obtained by modelling the NHH meters using scaled average profiles, such as the Elexon profiles, then by scaling the demand of smart meters on the same feeder.
Daily peak estimation	• The profile estimation methods for LV feeders calculated daily peaks within 5% to 20% of measured values. This range indicates the RMS over the test duration (approx. 9 months) of the daily errors. • The RMS daily error for substation demands was between 5% and 15%, a lower range than for the LV feeders due to the higher level of aggregation
Annual peak estimation	• The annual peak LV feeder demand was mostly estimated within ±20% of measured values but there were a few feeders with much greater error. • The approaches using load profiles calculated from generic profile data tend to under-represent the peaks. This is expected as the generic profiles represent the average demand over a large number of customers and do not account for individual variations either between customers or over shorter timescales. • The approaches using scaled smart meter tend to over-estimate the annual peaks. Scaling up the smart meter data to account for the NHH meters takes no account of the diversity between those customers, and so does not account for the probability that their peak demands would not occur at the same time.

9.8 Quality Validation

The project outputs were dependent on the data handling, algorithm processing and accuracy assessments being correctly set up by CGI. Given the importance on this being correct, third-party review was seen as essential. While originally it was expected that these datasets and processes might be difficult for a third party to access, the use of a shared Secure FTP site and early involvement in algorithm selection and development helped to ensure that the third-party review could take place.

Item	Learning Detail
Use of Jupyter workbooks	• These were very helpful in providing access to the code that was being used to perform the calculations alongside context information. It also made data sources clear which made the process of third party review much simpler.

9.9 Project process

Item	Learning Detail
Multi-party team	• There were useful synergies generated from bringing together multiple parties with different backgrounds.
Secure FTP site and use of MoveIT	• While a secure website was originally tested for use to support SMITN, the different access to technologies of the different parties made this difficult to use. However, an alternative secure FTP (SFTP) site was setup and used successfully.
Data Protection Impact Assessment	• The second stage Data Protection Impact Assessment used within NGED is very time consuming to populate, even when there are reasonable mitigations in place. More time needs to be allowed for this process in future projects.

10 The Outcomes of the Project

The project has resulted in the following outcomes.

- We have a better understanding of smart meter data availability and the processes required to gather data.
- We have a better understanding of the response levels from smart meters to being reconfigured to and from 1-minute voltage data measurement.
- We have gained an understanding of the general level of inaccuracy in the CROWN data for feeder and substation association and the phase data within EO.
- We have improved our ability to gather more targeted aggregated data as a result of revising the data privacy agreement.
- We have established the actual phase connectivity for 8,790 customers linked to the 46 distribution substations within the test network in the Milton Keynes area.
- We have established the actual feeder connectivity for 125 customers from the feeder survey.
- We have captured data about LCTs that were visible from the roadside which can be used to assist with LCT validation.
- We have captured photographs of the inside of customer meter boxes which can be used in future projects e.g., AI image processing.
- We have developed and validated methods for phase and feeder identification with indicative accuracy metrics. The way in which these can be set up for BAU is currently under discussion.
- We have established the best of four proposed methods for estimating load profiles for LV feeders and for distribution substations.
- We have identified potential methods for determining the locations of LCTs. The locations of PV installations can be determined with high confidence using export demand data.
- Visual validation for domestic PV has been attempted using free to access on-line resources.
- The learning from the project has been shared with stakeholders and published on the NGED website.

11 Data Access Details

New data was captured as part of the project as follows;

- Phase survey and feeder survey records
- GridKey substation monitoring data at 1-minute resolution
- Smart meter voltage data with 1-minute and 30-minute resolution
- Aggregated half-hourly smart meter demand data for LV feeders and substations.
- Monthly smart meter demand data for individual MPANs
- Half-hourly smart meter export data for individual MPANs
- Half-hourly smart meter data for individual MPANs of commercial customers
- Alert and alarm messages for smart meters

Please contact the innovation team to request access to data.

12 Foreground IPR

The algorithms used to analyse the data for the various use cases are included in Jupyter workbooks and as Structured Query Language (SQL). These are available on request and can be provided alongside a data dictionary to provide context.

13 Planned Implementation

All four of the SMITN use cases are planned to be implemented into business as usual. i.e.

- Phase data
- Feeder / Substation association
- LCT detection
- Substation profiles

Some preliminary work needs to take place to determine where algorithm results will be stored and the degree of data required to indicate the provenance of the data. (e.g. date of calculation, which algorithm where more than one option exists, version of algorithm / model etc.)

Roll out into BAU will require considerable automation of processes that were carried out manually for SMITN, notably data extraction.

The interdependence between the use cases suggests that rather than rolling out one use case at a time it will be more likely to take an area at a time and apply the use cases so that where corrections to the phase, feeder and substation connectivity are found these can be fed back into the algorithm for calculating substation profiles. The optimum scale of an area to be assessed at once needs to be determined. Similarly the optimum ordering of areas will need to be established to ensure benefits are captured by substitution of the process in place of substation monitoring.

14 Contact

Further details on this project can be made available from the following points of contact:

nged.innovation@**nationalgrid.co.uk**

Innovation Team

National Grid
Pegasus Business Park,
Herald Way,
Castle Donington,
Derbyshire
DE74 2TU

15 Glossary

Abbreviation/Term	Meaning
AI	Artificial Intelligence
API	Application Program Interface
CROWN	NGED's enterprise asset management system
DCC	Data and Communications Company
DNO	Distribution Network Operator
EO	Electric Office – NGED's Geographical Information System
EV	Electric Vehicle
GDPR	General Data Protection Regulations (data protection)
GridKey	Substation monitoring units developed by Lucy Electric
GIS	Geographic Information System
GUI	Graphical User Interface
HH	Half Hourly / customers with Half Hourly electricity metering
HISTAN	Historical Analogue
HP	Heat Pump
HV	High Voltage (6.6 and 11 kV)
iHost	Software system to access substation monitoring data
INM	Integrated Network Model
LCT	Low Carbon Technology
LV	Low Voltage (0.4 kV)
Move It	Synchronisation software used to automatically transfer data from NGED to the SFTP
MPAN	Meter Point Administration Number
NGED	National Grid Electricity Distribution
NHH	Non Half-hourly
NIA	Network Innovation Allowance
NIFT	Network Investment Forecasting Tool
NOP	Normally Open Point
PC	Profile Class as used by Elexon to categorise customers

Abbreviation/Term	Meaning
PV	Photovoltaic
RMS	Root Mean Square
SFTP	Secure File Transfer Protocol
SQL	Structured Query Language
TSDS	Time Series Data Store – a historian for real-time network monitoring data
VisNet	Substation monitoring unit supported by EA Technology
WP	Work Package
WPD	Western Power Distribution (Former name for NGED)

Registered Office: Avonbank, Feeder Road, Bristol BS2 0TB
nationalgrid.co.uk

nationalgrid