WPD INNOVATION

VM Data

NIA_WPD_045

Six Monthly Progress Report

Reporting Period: Nov 2019 – Mar 2020

WESTERN POWER DISTRIBUTION

Serving the Midlands, South West and Wales

Version Control

Issue	Date
1.0	17/04/2020

Publication Control

Name	Role
Arpita Nair (ElectraLink) & Max Hudson (IBM)	Author
Ricky Duke (WPD)	Reviewer
Jon Berry (WPD)	Approver

Contact Details

Email

wpdinnovation@westernpower.co.uk

Postal

Innovation Team
Western Power Distribution
Pegasus Business Park
Herald Way
Castle Donington
Derbyshire
DE74 2TU

Disclaimer

Contents

1 Executive Summary 4

2 Project Manager's Report 9

3 Progress against Budget 22

4 Progress towards Success Criteria 23

5 Learning Outcomes 24

6 Risk Management 26

7 Intellectual Property Rights 29

8 Consistency with Project Registration Document 30

9 Accuracy Assurance Statement 31

Glossary 32

1 Executive Summary

The Virtual Monitoring (VM) Data project, funded by WPD's Network Innovation Allowance, will deliver a virtual monitoring capability across the WPD Low Voltage (LV) network. This will reduce need for physical monitoring and improve our knowledge of asset loading against time, thus avoiding the costs associated with physical monitoring and demonstrating RIIO-ED2 cost savings and transition to Distribution System Operator (DSO).

Current lack of access to Half-Hourly (HH) data about household power flows on WPD's network inhibits the understanding of LV network load flows, and of where Electric Vehicles (EVs) and Low Carbon Technologies (LCTs) are connected at LV level. With the acceleration of LCT take up, this could result in clustering on the network which then creates a need to install physical monitoring at substations to monitor the loading of the network. The VM Data project will also create half-hourly load profiles for WPD's customers, including those with EV/LCT that will be aggregated up into a Virtual Monitoring tool for the LV networks.

The VM Data project is the follow on from the LCT Detection project. It focuses on the ability to harness structured and unstructured data from ElectraLink, combine it with cognitive analytics, and use this to develop a cost-effective virtual monitoring capability across the network. There are three principal ways it will deliver value:

- Enhancement of the proof of concept models under the LCT Detection project, together with other existing toolsets to solve visibility issues on the network and other Distribution Network Operators (DNOs), predominantly focusing on lack of visibility of EVs and LCTs;
- By using this visibility and existing datasets, a set of consumption profiles will be generated and used to develop a mechanism to provide virtual monitoring capabilities in the absence of smart meters and access to smart data; and
- By mapping virtual monitoring at Meter Point Administration Number (MPAN) level to the LV connectivity model, we will enable half-hourly loading profiles on feeders and distribution transformers to be analysed.

The first work package of this project will gather data on the network required for development of a Virtual Monitoring tool. All relevant existing datasets needed to develop the tool will be brought together, including our existing datasets (such as substation locations or the Embedded Generation dataset) and ElectraLink's Data Transfer Service (DTS) dataset. Other NIA project datasets will be analysed for model integration, e.g. Electric Nation, OpenLV, LV Connect and Manage. We will use surveys to supplement this data and confirm the LCT status of MPANs where status is unknown in our samples.

We will strengthen the LCT detection model by creating a negative data set from MPANs known not to have LCTs installed. We will also validate the LCT detection model results by selecting MPANs labelled as PV and EV, and then confirm whether this is correct. Validating the MPANs labelled as not having LCTs is more difficult since there is a fairly high probability that an MPAN selected purely at random will not have LCT, so we will use a larger sampling size and check their LCT status.

The updated and validated model will be used to fill in the visibility gaps for LV-connected customers, including those with EVs and LCTs on WPD's networks. Access to this data will be enabled through ElectraLink's Energy Market Data Hub and the modelling will use meter flow messaging data to advance our understanding of LCT locations. We expect to significantly enhance the LCT modelling to predict the presence of LCTs by analysing the following data:

- Consumption data;
- Historic weather data;
- Known LCT connections;
- Notified installations (e.g. G83);
- Demographic data;
- Known LCT-free connections; and
- Freeform text comments made by engineers.

The model's output will be analysed against actual substation monitoring data to verify the accuracy of using profiles as a proxy for direct access to smart meter data for the purposes of network monitoring. Existing substation monitoring will be used to validate the virtual monitoring.

To further improve the validation process and refine the model algorithms, we will purchase HH data from a Supplier / Meter Services company for customers who are known to have smart meters, have given their permission for the electricity consumption data to be used and have known household / socio-economic characteristics.

The final stage of the project is to take the output and learning from the first two work packages – understanding what is on the network and understanding the impact this will have on network power flows – and create a model for the WPD network. This model can then be used to understand network constraints and identify any requirement for more granular physical monitoring.

The benefits to forming a more accurate picture of the LV network are:

- Improved DNO power flow models using virtual monitoring at half-hourly intervals to identify e.g. reverse power flows, distribution transformer and feeder heating characteristics, and power quality issues;
- Improved potential for demand management / curtailment options; and
- Improved asset reinforcement/ investment decision making process.

The project will deliver the basis for a BAU-adaptable technology platform that will drive benefits for WPD's network and its customers through digitalisation of processes to enable facilitation of EV and LCT uptake.

1.1 Business Case

The objective of the Virtual Monitoring project is to provide many of the same benefits that would be obtained by fitting feeder-level physical monitoring to substations, but without the capital cost involved. It is unlikely that the outputs of this project could negate the need for monitoring altogether, but even replacing 50% of physical monitors with virtual monitors would lead to significant savings in the order of £52m.

In addition to providing a similar level of information as physical monitoring installed in a substation, the VM project would enable "what-if" analysis to be performed, allowing investment decisions to be made using not just the current load profiles but also possible future load profiles.

Another benefit would be improved knowledge as to where on the network LCTs are installed so that this can also be used in decision making. For example, a high proportion of PV in an area could indicate a risk of sudden surges in demand due to changing weather conditions.
For customers the benefits of the VM project would be:

- Financial savings leading to lower bills as a result of deferred reinforcement and reduction in LV monitoring investment;
- A more resilient network with lower risk of outages, leading to reduced disruption due to overloads being flagged up before a fault occurs; and

- The estimated cost benefit of virtual monitoring is £52M in total in our licence areas. This is based on:
  - 76,114k LV ground mounted substations across our network, with very few having physical real-time monitoring, of which 52,566 are forecast to be loaded upwards of 90% of their total capacity by 2030 (taken from the network assessment tool); and
  - An average £2k install cost per monitor.

If monitoring requirements of this group are reduced by 50% as detailed in the learning objectives, through virtual monitoring, this could potentially defer £52m of monitoring spend.

The benefits to forming a more accurate picture of the LV network are:

- Improved DNO power flow models using virtual monitoring at half-hourly intervals to identify e.g. reverse power flows, distribution transformer and feeder heating characteristics, and power quality issues;
- Improved potential for demand management / curtailment options; and
- Improved asset reinforcement/ investment decision making process.

The project will deliver the basis for a BAU-adaptable technology platform that will drive benefits for WPD's network and its customers through digitalisation of processes to enable facilitation of EV and LCT uptake.

1.2 Project Progress

This report covers project progress for the period November 2019 (Project Start) to March 2020.
Project activities in this period have focussed on discovery and design thinking activities, including site visits and workshops. The analytics phase of the project is composed of 9 sprints of one-month duration. A sprint report has been produced for each sprint detailing the analysis and other activities performed during the sprint and the results obtained.

1.3 Project Delivery Structure

1.3.1 Project Review Group

The project has not yet held its first project review group. The project review group was due to meet in April, but this had now been suspended until June due to a pause in the project.

The VM Data Project Managers Group meets on a monthly basis. The Project Managers are responsible for (among other things):

- The day-to-day liaison between the Parties;
- Co-ordinating the provision of the Project Activities, including overseeing the performance and quality thereof, from ElectraLink and the provision of the Input Requirements from WPD and the Project Partners;
- Arranging and attending (personally or by representative), at each Party's own cost, progress meetings and other meetings, at intervals and locations as agreed between the Parties from time to time, to discuss developments and seek to resolve any issues arising;
- Identifying and agreeing in writing on behalf of the Parties any Background IPR used, or to be used, in the course of the Project and the owner of the same, prior to or as soon as reasonably practicable following its disclosure in the course of the Project;
- Identifying and agreeing in writing on behalf of the Parties any Foreground IPR created or developed, or to be created or developed, in the course of the Project and the owner of the same, prior to or as soon as reasonably practicable following creation or development of the same in the course of the Project;

- Resolving any disputes;
- Preparing and agreeing the Close Down Report; and
- Such other matters as may be agreed between the Parties from time to time.

1.3.2 Project Resource

ElectraLink:
ElectraLink's primary role in the project is as follows:

- **EMDH data provision:** ElectraLink will take industry data from the Energy Market Data Hub (EMDH) data set and supply it to IBM for the purposes of LCT detection model enhancement and validation;
- **Locate LCTs:** ElectraLink will subcontract a company with meter reading capability to locate and identify known LCT installations for the purposes of LCT Detection validation and customer profiling;
- **Obtain HH demand profiles:** ElectraLink will subcontract the acquisition of half-hourly customer demand profiles of customers with smart meters and known LCT installation status to support 1) LCT detection model development, 2) Virtual LV network monitoring; and
- **Enable Virtual LV monitoring:** ElectraLink will support WPD to match EMDH consumption data with the customer profiles created in Activity 3. ElectraLink will also support WPD to aggregate: 1) The LV network connectivity model with addresses that the LCT connectivity model identified as having LCTs; 2) Half hourly customer profiles from Activity 3 against addresses, grouped by LV feeder; and 3) EMDH consumption data scaled by the half-hourly customer profiles from Activity 3.

IBM:
IBM's primary role in the project is to establish a data analytics platform and carry out a data analytics programme based on data provide by WPD and ElectraLink. The project will be carried out in three phases:

- **Discovery**: A business analysis phase incorporating design thinking activities and setting out the work plan for subsequent phases;
- **Execution**: Five one month long "sprints" of data analytics activity covering two workstreams – LCT Detection and Profile Modelling; and
- **Consolidation**: Bringing the two workstreams together to produce a final report and model which will enable WPD to predict load patterns on the LV network.

1.4 Procurement

Table 1-1 details the current status of procurement for this project.

Table 1-1: Procurement Details

Provider	Data Services	Area of project applicable to	Delivery Timeline
Hildebrand (DCC Other User – third party)	Anonymous electricity consumption data for approximately 2,200 households, including the first two letters of the MPAN to identify the distribution network and the outer postcode to identify the geographical region. Anonymous electricity consumption	Profile Modelling	Delivered – March 2020

	data for approximately 700 households where the MPAN is unknown		
Drax - B2B Energy Supplier	Pseudonymised 2,000 HH profiles for SME customers in the WPD area	Profile Modelling	Delivered – March 2020
Morrisons Data Services	LCT Model Validation site surveys	LCT Detection	On pause due to COVID-19 H&S restrictions

1.5 Project Risks

A proactive approach is taken to ensure effective risk management for the VM Data project. A RAID (Risks, Assumptions, Issues, and Dependencies) log is maintained, examined and updated by WPD, ElectraLink and IBM. This activity ensures that risks are frequently reviewed, examining: whether risks still exist, whether new risks have arisen, whether the likelihood and impact of risks have changed, for reporting of significant changes that will affect risk priorities, and to deliver assurance of the effectiveness of control.

Risks are reported to WPD within each monthly report. At each monthly meeting, the RAID log is reviewed and updated by the project delivery team. ElectraLink provides a critical overseeing role within the meeting to ensure that all risks are being effectively captured and managed.

Contained within Section 6.1 of this report are the current top risks associated with successfully delivering VM Data as captured in the RAID log. Section 6 provides an update on the most prominent risks identified.

1.6 Project Learning and Dissemination

A Project Learning Log is maintained. Project lessons learned and what worked well are captured throughout the project lifecycle. These are captured through a series of on-going reviews with stakeholders and project team members. These are reported in Section 5 of this report.

Project Dissemination Activities during this period:

- A VM Data introductory brochure was produced for LCNI and general dissemination to stakeholders;
- The VM Data brochure was circulated widely on social media (Twitter and LinkedIn);
- ElectraLink secured a speaking slot at Solar & Storage Live (September 2020), for WPD to co-present at this industry-wide event on the project, with ElectraLink; and
- ElectraLink submitted the LCT detection project for Best Data Project at the annual Network Awards – the project has been selected as a finalist. This will serve as a springboard for a communications campaign in the lead up to the Awards ceremony in July, for the VM Data project.

2 Project Manager's Report

2.1 Project Background

The energy market is complex and evolving, particularly with growing smart technologies and embedded, renewable generation. For DNOs like us, the increasing number of 'invisible' changes to customer demand connected to the LV networks (growth of EV's and Embedded Generation etc.) challenge existing network practices to the extent that the status quo is no longer possible.

At present, technology change is outpacing changes in modelling and forecasting of consumer uptake of 'smart', Distributed Energy Resources (DER) or EV technology; therefore, it is difficult to monitor or understand the change in requirements on the LV network, under existing arrangements, without monitoring EV and DER impacts directly at source (or substation level). While smart meters will improve the visibility of network load and generation in the longer term, there is a need for a solution that can identify unregistered equipment and enable visibility of power flows on the LV networks. Whilst Smart meters will one day provide this functionality, LCT take-up is out-pacing Smart Meter installations. Therefore there is a risk to the network from widespread take-up of these technologies and a need to monitor the situation in the most cost effective manner.

We need to have more data available to understand network conditions and to better understand the evolution of the market as we see proliferation of DER and EV. This project will build on the LCT Detection project by initially validating and improving on LCT detection techniques using data from customer surveys, data from other innovation projects and customer HH load profiles.

The project will then create a virtual monitoring tool for our LV networks by obtaining HH data for customer types and extrapolating them to profiled customer consumption data from Data Transfer Service (DTS).

2.2 Project Overview

The LCT Detection project, which was successfully completed in 2019, used IBM's cutting-edge AI and cognitive analytics capability combined with ElectraLink's DTS dataset to demonstrate that it was possible to detect hitherto unknown LCTs on the LV network. This project will build upon this work, validating the existing model and looking to enhance the existing LCT detection model through the introduction of additional data sets. Furthermore, this project plans to use half hourly meter readings to construct a set of consumption profiles that can be allocated to MPANs. Given the impact of LCT on consumption, the enhanced LCT Detection model will be a critical enabler for the generation of increasingly precise MPAN level load profiles.

The aggregation of these consumption profiles will allow the load profile at substations, and other network assets to be estimated, through an aggregation engine, which will be developed with our data tool / platform teams as part of this project.

2.3 Project Progress

At this stage in the project we have completed the Discovery phase and three of the month long analytics sprints, as well as the 2 initial sprints in the Discovery Phase. This section summarises the key activities undertaken, and results obtained.

2.3.1 Discovery Phase

The VM Data project started with a 2-month discovery phase, covering Sprints 1 & 2 of the Sprint Delivery Plan. The purpose of this phase was for the project team to gain a better understanding of the business users and processes that will be impacted by the project deliverables, in order to ensure that the outputs of the analysis will add business value to WPD.

Two key activities took place during the Discovery Phase:

- Business Analysis - a series of visits to WPD sites to engage with the anticipated users of the outputs of the project; and
- A Design Thinking workshop involving attendees from WPD, ElectraLink and IBM.

Detailed reports have been written for each of these activities, here we summarise the work undertaken.

Business Analysis

The project team made a series of visits to WPD sites. For each visit, we made a short presentation to invited employees, and then spent some time sitting alongside the employees talking about their daily work, the pain points in current processes and how better information may be able to help them.
In carrying out these discussions, the team did not use a scripted interview or checklist of questions, instead preferring to ask open questions so as to not restrict the information shared. It was found that employees were generally very happy to talk through what they were doing and showed a great deal of interest in the work being undertaken.
Using this approach, the team were able to:

- Confirm our understanding that the outputs the project is aiming to produce will be valuable to the business;
- Brief WPD employees to begin engagement, acknowledging that a successful project will likely need further engagement as it progresses; and
- Identify areas that may hinder the future implementation of the project outputs (e.g. problems with the mapping of MPAN to feeder/substation) so that these can be addressed.

The sites visited were Aston, Birmingham and Avonbank, Bristol. These site visits in part acted to confirm the potential business value of the VM Data project, as well as providing an opportunity to hear employee opinions on the work and raising any issues that may hinder implementation further down the line. Overall, the visits were deemed extremely successful, with very positive reactions and a great deal of interest shown by a wide variety of WPD employees.

Key Findings from Business Analysis

- **Data Quality:** Almost everybody we spoke to mentioned the unreliability of the Crown data set and how it could not be relied upon to be accurate or up to date;
- **Network Planning:** The focus is on peak demand, or worst case scenario. Even if they knew where PV generation might reduce demand it would not be relied on for planning decisions;
- **Current Process**: Currently when the planners are not confident in the data available at feeder or substation level, they will clip a load monitor to get a more accurate picture. This is time and resource consuming;
- **Profile Development:** New demand profiles will need to take into account factors such as geography and seasonality. Profiles that take into account modern heating methods will be useful in planning for pockets of new development; and
- **Substation Level:** There is currently a good understanding of load at the primary substation level, however information at distribution substation level is limited. A simplifying assumption that each connected substation contributes equally to demand at the primary substation is often made. A more accurate picture of this would be useful to the business.

Design Thinking Workshop

The workshop was designed as a user-centred approach to reach a mutual understanding of the issues being addressed by the VM Data Project, analysing the current processes we believe will be beneficiaries of the project outputs and identifying the potential benefits of virtual monitoring. There were around 20 attendees from WPD, ElectraLink and IBM plus 4 IBM facilitators.

Morning Sessions
The morning sessions focussed on the potential users of the VM Data project outputs:

User profiles were developed for:

- LV Network Planners;
- HV Network Planers;
- Policy Developers; and
- Asset / Investment Strategy.

These profiles show what the users are Doing, Thinking, Feeling and Saying and help us to understand the impact that the project could have on their working day.

A number of Scenarios were considered. The Scenarios were typical tasks that users might undertake. Once again, the focus was on Doing, Thinking, Feeling and Saying. The scenarios considered were:

- Conversion of a Victorian house into 6 flats;
- The local authority wants to install solar panels on a row of terraced houses; and
- Planning a back feed to allow repair of a network fault.

Analysis of these scenarios allowed us to understand where the outputs of the VM Data project could be of value.

In the final session of the morning a set of User Needs statements were prepared in the format:
As a…
I need…
So that…

Afternoon Session
The afternoon sessions focused on the analytics workstreams planned for the next phase of the VM Data project, namely LCT Detection and Consumption Profiles. LCT Detection was the focus of a similar workshop held at the start of the LCT Detection project and so less time was dedicated to this workstream.

Hypothesis generation was used to create a set of hypotheses related to

- The factors that influence the amount and time of day when energy is consumed (profiles); and
- The presence or absence of LCT.

The data necessary to prove or disprove each hypothesis was defined and finally a selection of the hypotheses was evaluated for feasibility v business value.

Findings from Workshop
Using this approach, the team were able to:

- Confirm Business Understanding of the challenges faced by WPD with respect to understanding the load on the LV Network, and the ways in which the planned outputs of the VM Data project could help to address those challenges;
- Generate Hypotheses to be tested during the next Evaluation phase of the project; and
- Identify Data Sources that could be used for the analytics.

Detailed findings from the Design Thinking workshop are documented in the Design Thinking report.

2.3.2 Sprint 3

Sprint Plan

We originally planned to validate the original LCT Detection model during this sprint, however the data required was not available. Instead the analysis focussed on the initial steps of profile development, as well as getting the fundamental requirements for analytics in place.

Prerequisites for Analysis

The analytics environment was commissioned towards the end of the Discovery Phase. In sprint 2 the environment was tested and optimised with real scripts and data.

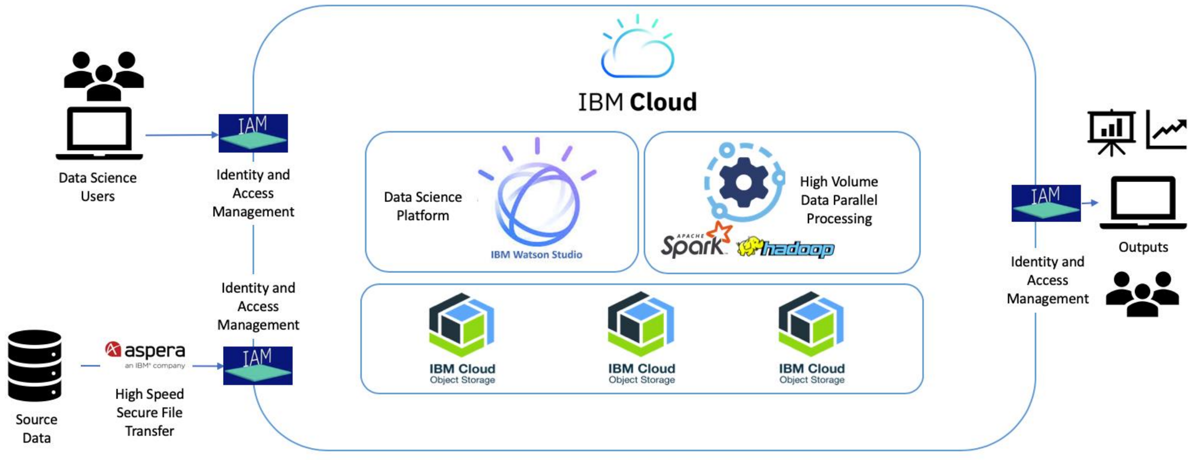


Figure 2-1: VM Data Reference Architecture

In addition to setting up the environment, the initial data sets were loaded to Cloud Object Storage, from where they were examined and documented in a data catalogue.

LCT Analysis

For this sprint we had no data additional to that used for the original LCT Detection model which meant that further refinement of the model was not possible. The model was however refreshed to run in the new environment.

Profiles Analysis

Initial pieces of work that will contribute to the profiles analysis were performed in this sprint:

- A table mapping MPAN to feeder to Distribution substation was prepared. This will be used to:
  - Understand the properties of the MPANs connected to each feeder and distribution substation so that different profiles observed at substation level can be analysed to determine drivers of the differences; and
  - Provide the structure for aggregating MPAN level profiles to feeder and substation level.
- The Average Daily Consumption (ADC) in summer, in winter and the ratio between the two was calculated for those MPANs where the interval between meter readings allowed this to be done. Where MPANs have differences in the ADC Summer: Winter ratio it indicates that they consume electricity in different ways and so this can be used to help identify groups of MPANs that may have different patterns of energy consumption. As expected, the majority of MPANs have lower consumption in summer to winter.

- The ADC attributes were investigated for relationships with the Rural/Urban category[1], with the deprivation index[2]. It was found that the most rural properties had higher energy consumption in both summer and winter, and also the biggest relative difference between the two. Conversely no relationship between deprivation index and ADC was observed.

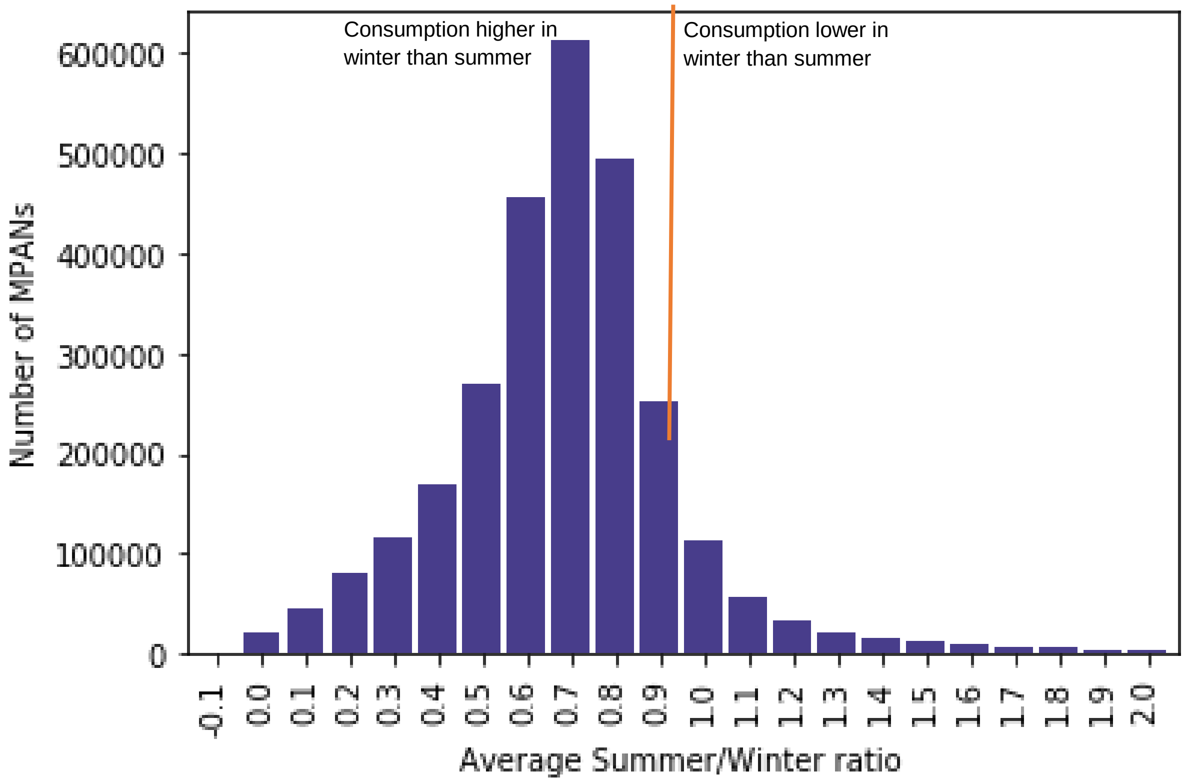


Figure 2-2: Average Summer/Winter consumption

1 https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/239478/RUC11user_guide_28_Aug.pdf

2 https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019

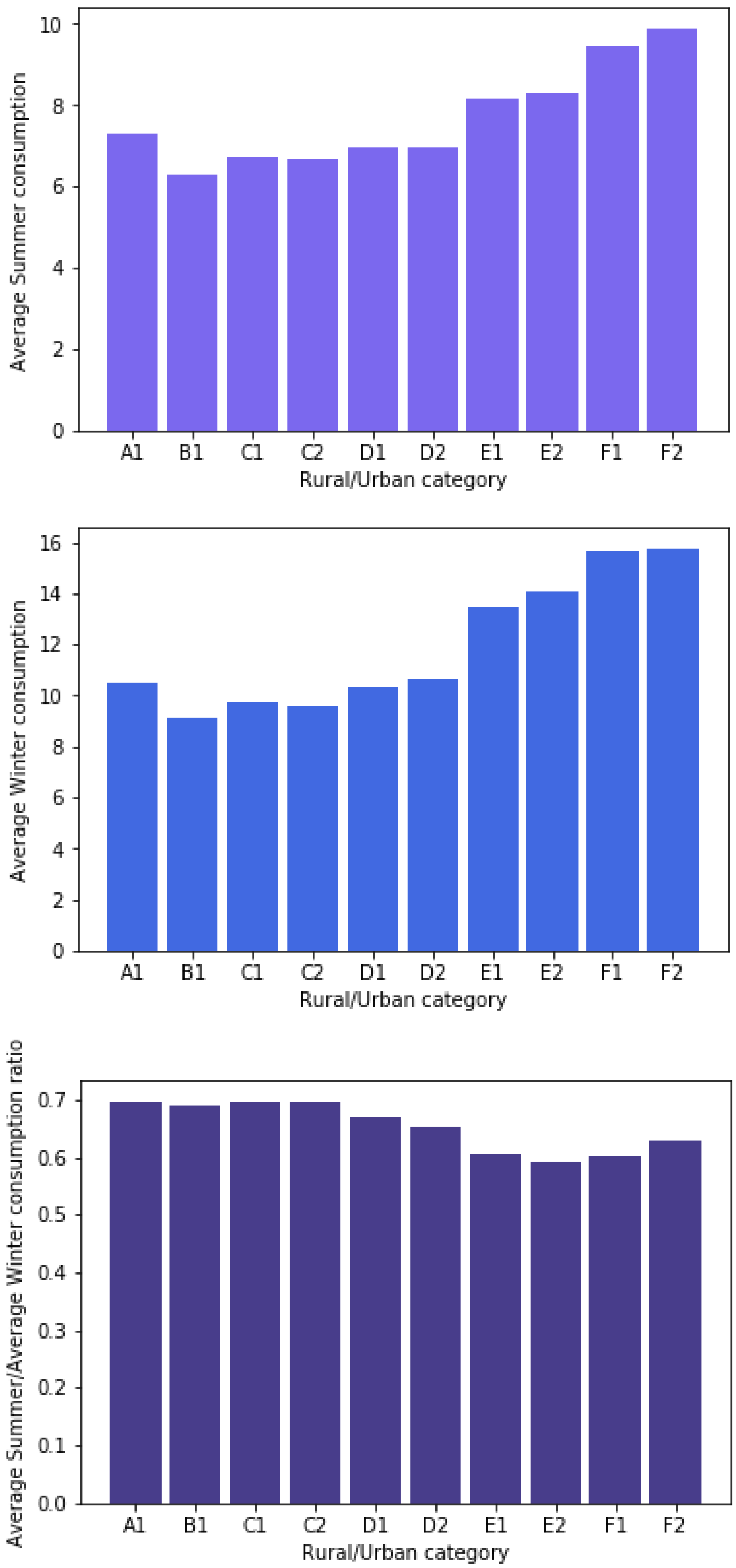


Figure 2-3: ADC by Rural / Urban category

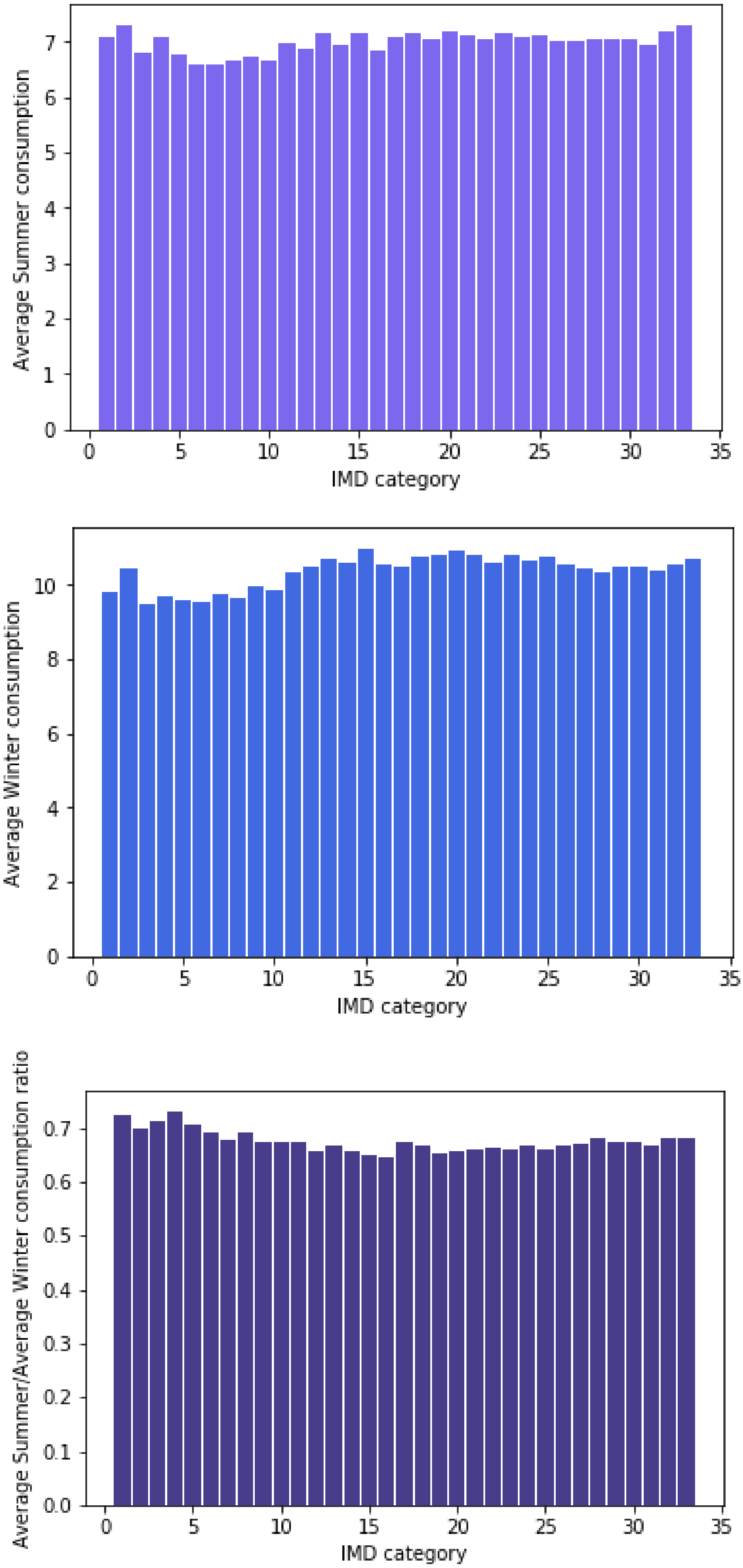


Figure 2-4: ADC by Deprivation Index

Average summer consumption and Average Winter consumption are kWh per day.

The deprivation index was grouped into 33 categories from 1 (most deprived) to 33 (least deprived) (the IMD categories are defined as IMD score = 0 – 1000 is category 1, IMD score = 1001 – 2000 is category 2 etc).

2.3.3 Sprint 4

Sprint Plan

In sprint 4 neither of the additional LCT data sets (validation data and negative data set), nor the half hourly MPAN level data was available. This sprint focussed on analysis using data that was available to inform both the LCT detection and profiles work streams.

Data Preparation

Two data sets were acquired during this sprint that required substantial manipulation before they could be used in the analysis:

- **Energy Performance Certificate** (EPC) Data – an EPC is required when a property is built, sold, rented or when LCT grants are applied for. This data set is freely available and contains a wealth of data regarding the energy consumption characteristics of the property. The data set contains the address and postcode, but this is in a different format to the address and postcode for MPANs. An algorithm was coded to match the two sets of data resulting in 44% of WPD MPANs being matched to EPC data; and
- **Distribution Substation data** – Energy data for 79 distribution substations was obtained from EA Technology. This data was manipulated to produce half hourly consumption data for each feeder.

LCT Analysis

In the previous LCT Detection project we worked with pseudonymised data with no address information. For this project we have the protocols in place to work with PI and consequently have real MPANs and address information. This has opened up new data sets to be included in the analysis. We also received Supplier data at MPAN level from ElectraLink.

Analysis of the Supplier data showed that customers of the Big 6 suppliers, with the exception of Scottish Power were less likely than the average to have PV, and with the exception of Scottish Power and SSE were less likely to have EV. It has been noted that Scottish Power have an offering alongside Nissan associated with their Leaf electric vehicle.

In the charts below the “Difference from Expected EV” is the percentage difference between the number of EV or PV we would expect for a supplier if we assumed the population average across all suppliers and that actually observed. For example, if a Supplier has 1,000,000 customers and across population 1% of MPANs are known to have EV we would expect the Supplier to have 10,000 customers known to have EV. If they actually have 8,000 known EV then the Difference from Expected = -0.2. If they actually have 15,000 then the Difference from Expected = 0.5.

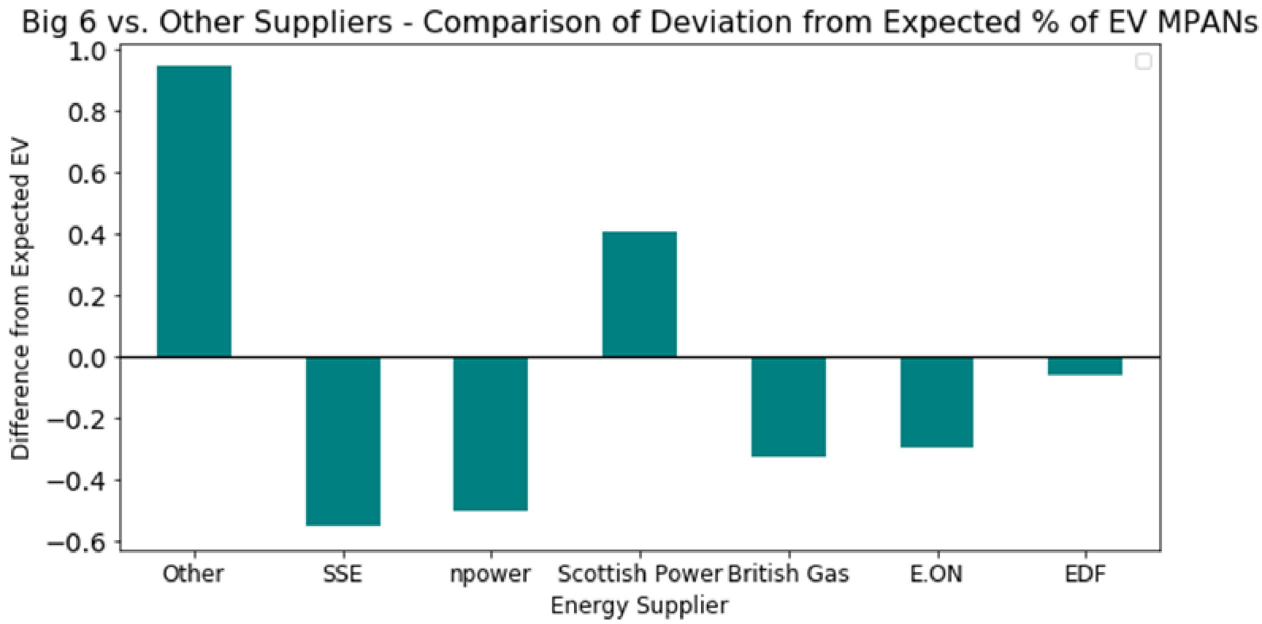


Figure 2-5: Chart showing deviation actual from expected EV for Big 6 v Other suppliers

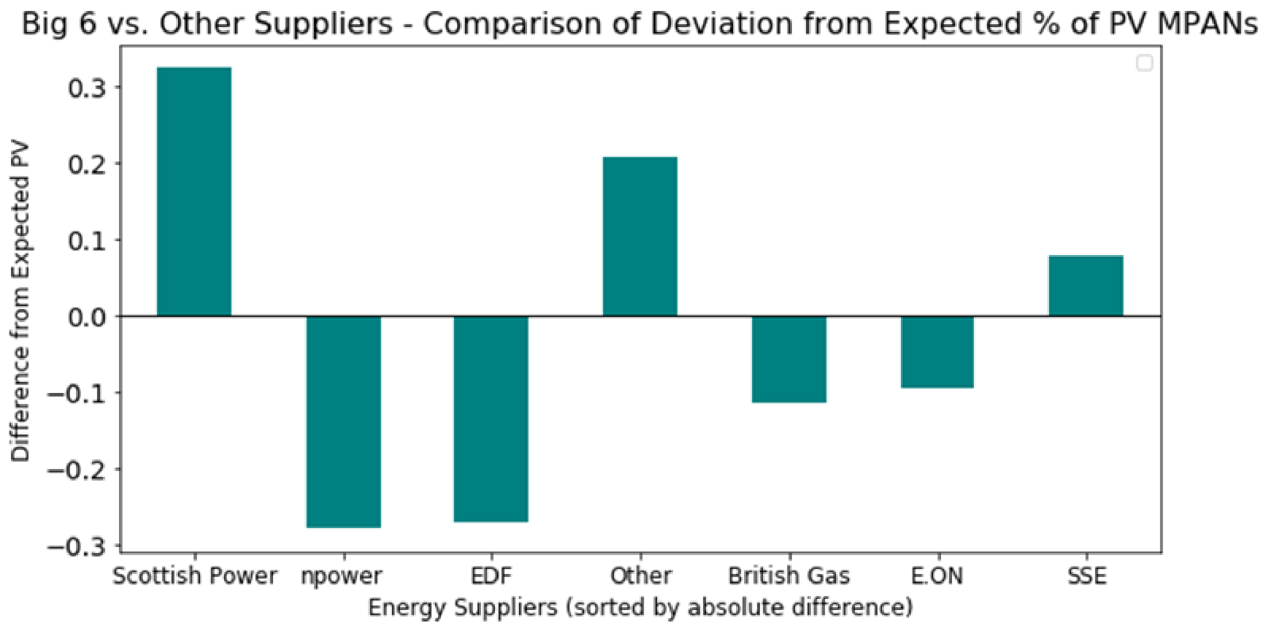


Figure 2-6 Chart showing deviation actual from expected PV for Big 6 v Other suppliers

We investigated the relationship between LCT and various attributes included in the EPC data set.

Table 2-1: Relationship between LCT and property type

Attribute	EV	PV
Overall Energy Efficiency	No observable difference	Higher energy efficiency
Build Form	Detached > 2x more likely to have EV	Detached more likely to have PV
Property Type	Houses are more likely to have EV; Flats are less likely to have EV	Houses and bungalows are more likely to have PV; Flats are less likely to have PV
Roof Type	Little difference	Pitched roofs are more likely to have PV

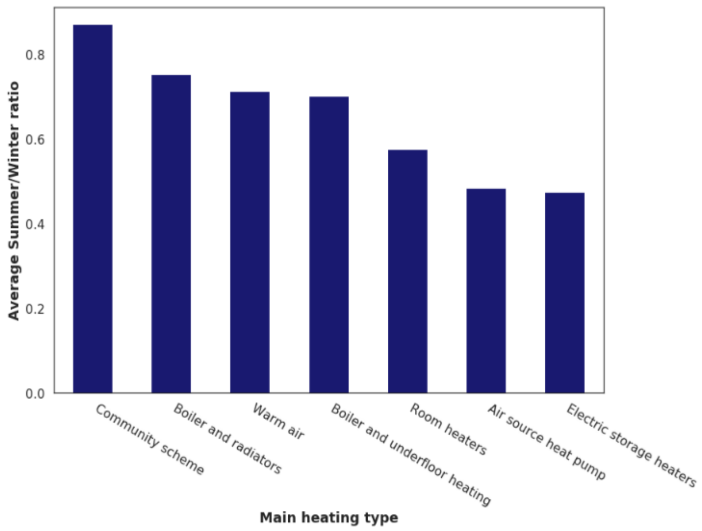
We also showed that properties with PV had a lower ADC Summer:Winter ratio than the population average, i.e. they consumed relatively less electricity in summer months.

Profiles Analysis

We used the ADC ratio calculated in the previous sprint as a proxy for differences in energy consumption behaviour and investigated how this related to the various attributes obtained from the EPC data set.

Table 2-2: ADC ratio to attributes in the EPC data set

Attribute	Relationship
Overall Energy Rating	The most and the least efficient categories (A and G) had the lowest ADC Summer:Winter ratio. This is likely driven by low summer consumption due to LCT in the most energy efficient properties, and high winter consumption due to inefficient heating in the least efficient properties
Main Heating Energy Rating	More efficient properties have higher ADC ratio (low winter consumption)
Hot Water Energy Rating	More efficient properties have higher ADC ratio (low winter consumption)
Walls Energy Rating	No observable relationship
Windows Energy Rating	No observable relationship
Lighting Energy Rating	No observable relationship
Property Type	Park Homes have a low ADC ratio, but the number of these is small
Built Form	Enclosed end and mid terrace have slightly lower ADC ratio.
Mains Gas	Properties without mains gas have lower ADC ratios likely driven by high winter consumption due to using electricity for heating
Number of habitable or heated rooms	The ADC ratio has no relationship with the number of habitable or heated rooms in a property although the absolute consumption increases as the number of rooms increases
Main Heating Type	There is a strong relationship between heating type and ADC ratio. Those properties with community heating have the highest values and those with electric storage heaters have the lowest values
Transaction Type	The properties where the EPC was acquired due to application for a grant associated with LCT have the lowest ADC ratios.



There appears to be a significant difference in ADC ratio between different primary heating types.

Figure 2-7: Relationship between ADC ratio and Main Heating Type

This chart shows the relationship between ADC Summer / Winter ratio and the main heating type. Properties with Electric storage heaters have the lowest ADC Summer / Winter ratio – i.e. their ADC in Winter is higher than ADC in summer and the difference between the two is the highest for all of the different heating types. Conversely properties with communal heating have similar consumption in winter as to summer – ADC ratio is close to 1. The ADC ratio does not give any indication of the absolute consumption, just the difference between consumption in summer and consumption in winter.

In Sprint 4 we also performed our first analysis using unsupervised machine learning to identify consumption profiles using half-hourly data. Since we did not have MPAN level data we trialled the method using distribution substation data and found that we could identify profiles and also determine the drivers for the differences, in this case seasonality appeared to be one driver.

In the chart below each daily profile has been normalised with respect to the maximum demand. This means that at the time of maximum demand the value = 1 and the value at each of the other time points is the proportion of the maximum. Normalising in this way means that we can look at clusters of patterns in demand, rather than the absolute value of demand.

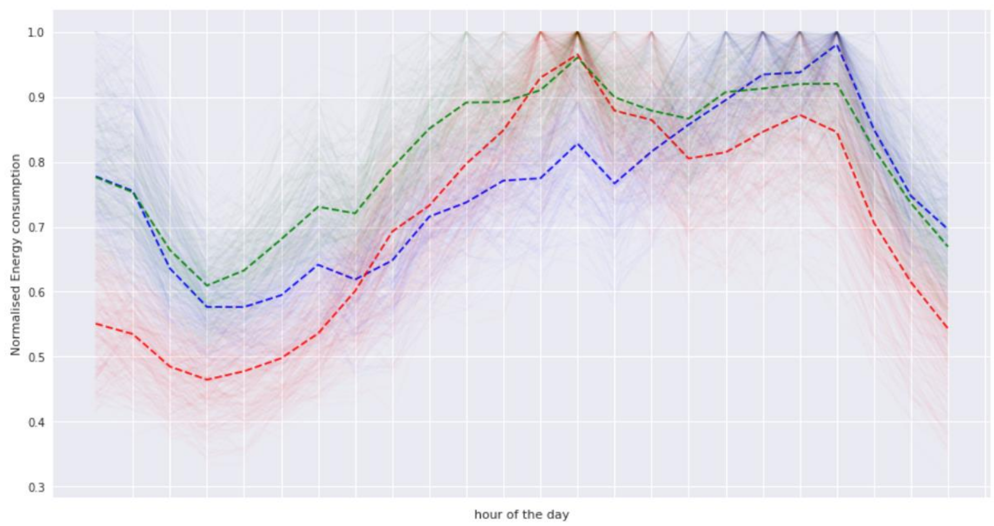


Figure 2-8: Normalised Consumption profiles showing 3 clusters for each day at substation 873105

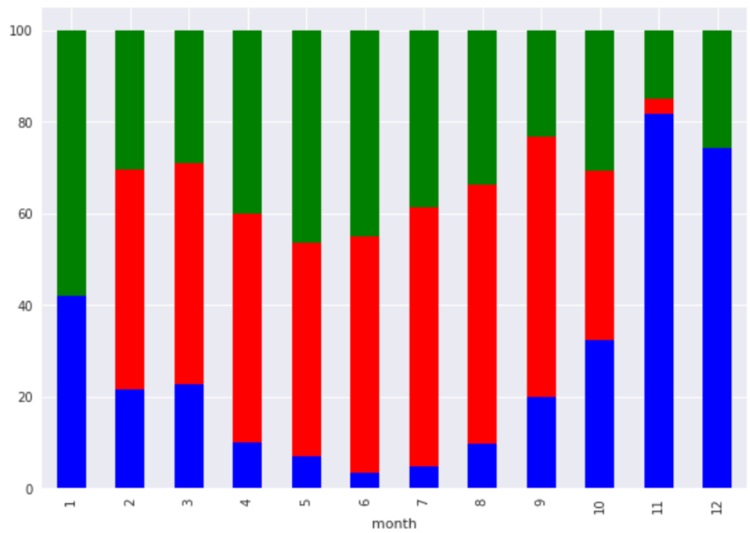


Figure 2-9: The percentage of days in each month within each of the identified clusters

2.3.4 Sprint 5

Sprint Plan

The majority of the analysis undertaken in sprint 5 was a detailed assessment of different unsupervised learning algorithms to determine the best candidates to be used with half hourly consumption data. The analysis was designed to test the method and not to identify consumption profiles.

Method and Results

We assessed 5 different algorithms, each with a range of input parameters, and 3 normalisation methods.

The method we used was as follows:

- We created a control test data set using data from feeders with only domestic MPANs connected and used September 2019 weekdays to control for seasonality and for weekday / weekend differences.
- The data was normalised in three different ways:
  - With respect to the maximum consumption;
  - As a percentage of total daily consumption; and
  - As a rate of change.
- We tested five different algorithms with each data set and a range of input parameters (specific to the algorithm). The algorithms tested were:
  - K-means;
  - Mean Shift;
  - Hierarchical;
  - DBSCAN; and
  - Gaussian mixtures.
- We used four different quantification metrics to score the performance of the algorithms.

We found that having controlled for seasonality and weekday / weekend the major difference between profiles was an overnight peak in one profile versus a morning and evening peak in a second – i.e. as for the existing ELEXON profiles. Each of the algorithms was able to distinguish these two profiles but the hierarchical method scored consistently better than the others. In a scenario where we were looking for previously unidentified profiles, we would drill down further into the data set to determine more subtle differences; however that was not the purpose of the analysis in this sprint.

We concluded that the hierarchical method was the best candidate to take forward for future analysis, and that 2 out of the 3 normalisation methods (with respect to the maximum, and as a percentage of the total) gave good results and we would continue to use both.

2.3.5 Conclusion

After 5 months of the VM Data project we have:

- Confirmed the business value that can be delivered by the project;
- Set up an analytics environment;
- Created data sets and attributes to be used in the analysis;
- Identified attributes associated with the presence of LCT and related different patterns of energy consumption;
- Shown that unsupervised machine learning techniques can be used to find consumption profiles from half hourly data and that we uncover the reasons for the differences; and
- Evaluated a range of different unsupervised machine learning methods and identified the best candidate to use for MPAN level consumption profiles.

3 Progress against Budget

Table 3-1 summarises the details of the progress that has been made with respect to the project budget.

Table 3-1: Project progress against budget

Sanction Description	Sanction Value	Planned Spend to Date	Spend to date	Explanation & Measures
ElectraLink Contract	£2,316,250	£1,060,857	£1,060,857	On track as per budget
WPD Project Management	£85,372	£42,160	£27,379	Less PM time has been required than planned for.
Equipment Costs	£86,000	£86,000	£0	Delayed due to analysis of HH data.
WPD Network services	£11,247	£11,247	£0	Delayed due to analysis of HH data.
TOTAL	**£2,498,869**	**£1,200,264**	**£1,088,236**	

4 Progress towards Success Criteria

Table 4-1 presents the progress the project has made towards the Success Criteria.

Table 4-1: Progress towards success criteria

Success Criteria	Progress
An improved version of the LCT Detection model that can identify LCT on the LV network with a measurable improvement in accuracy (precision).	We have identified additional attributes that can be included in a future version of the LCT detection model and these will be built into the model and tested when the additional LCT data sets are available.
Validation of positive LCT detection results and negative LCT detection results showing, with a high level of confidence, the level of accuracy of the improved LCT Detection model. This will include validation against half-hourly feeder-level readings from a set of our substations with LV substation monitoring.	We are awaiting the validation data set to complete this exercise.
A set of 15+ customer profiles, which will enable the extrapolation of ElectraLink's monthly consumption data to be extrapolated to half hour load profiles on each LV feeder	We have used Distribution substation and Feeder level half hourly data to show that we can create load profiles using this data type and we have refined our methodology. Next steps are to use the method with MPAN level half-hourly data.
Validation results showing the level of accuracy of the half-hourly feeder loads predicted by the VM data model.	Not anticipated at this stage in the project
Delivery of an LCT detection and Virtual Monitoring approach that can be transferred into the business as a BAU approach.	Not anticipated at this stage in the project
Incorporation of the Virtual Monitoring data outputs into one of our tools or ADMS platform. The project will provide load data at feeder level, with improved accuracy, in a format that is compatible with our systems. The data will be integrated into the existing capabilities/applications that are currently being used by planners/control engineers (e.g. GE PowerON, Network Assessment Tool or WinDebut)	Not anticipated at this stage in the project

5 Learning Outcomes

The project maintains a comprehensive learning log. The lessons learned during this period have primarily been in the following areas:

- HH data acquisition;
- Technical file transfer requirements;
- GDPR restrictions;
- Unsupervised learning outcomes; and
- Load profile clustering methods.

Details of the learning log entries created in the last 6-month period are provided in the Table 5-1 below.

Table 5-1: Learning Log entries created in the past 6 months

Workstream	Learning Detail (event, effect, trigger, early warning indicators, recommendations)	Internal Outcomes
Pre-mobilisation	Learning generated in terms of the approach used by IBM to manage initial uncertainty around data availability during the project stage whilst project planning.	Build flexibility into the Sprint process to be able to meet budget and schedule under different half-hourly data availability scenarios
Pre-mobilisation	It has been beneficial to involve the data scientists in the approach definition planning meetings well as project managers, to facilitate cross-team and partnership understanding of the Sprint process	Ensures expedient and smooth project mobilisation into delivery phase
Discovery Phase	It has been useful to have a detailed early handover and familiarisation of the project to the new personnel in the team	Ensures smooth transition of the project in the early stages of the project mobilisation into delivery phase
Execution Phase	There are approximately 7.65m MPANs included in the mapping data received from WPD, of these around 90% are domestic MPANs (profile classes 1 and 2).	Confirmation of the WPD asset structure (Primary/ Secondary Sub-stations and consumer endpoints)
Execution Phase	There are 184,281 Distribution substations in the WPD network. On average there are 41 MPANs per Distribution substation, with the highest number of MPANs at a substation being 1269. 47,741 Distribution substations with only one MPAN of which roughly half are domestic properties. These are distributed across WPD's distribution licence areas, but concentration is highest in rural areas such as West Wales.	High number of single MPAN transformers on domestic profiles aligns with the rural nature of WPD's distribution area. Potentially useful hypotheses around these network nodes?
Execution Phase	Aspera file transfer is blocked from WPD locations.	When the file upload was initially tested - test file sizes were small and did not trigger the Aspera high speed transfer client. On attempting to load actual data files - team hit a blocker. WPD has

		raised a helpdesk ticket to resolve; in the interim - data was transported to and uploaded by ElectraLink.
Execution Phase	Acquisition of HH consumption data for domestic properties - There are challenges in this area due to the sensitivities around GDPR and access to MPAN level (personally identifiable) domestic HH meter consumption data.	Project plan needs to be re-evaluated and contingency measures to be put in place
Execution Phase	GDPR requirements for HH consumption data - WPD, ElectraLink and IBM need to put in place a robust process that meets all data protection requirements for suppliers of HH data including segmentation rules, anonymisation and aggregation.	Processes and measures need to be put in place for data protection requirements of suppliers
Execution Phase	Sprint report shows how we used unsupervised learning to determine load profiles on a single distribution sub.	We have shown that unsupervised machine learning is a potentially useful route for developing consumption profiles.
Execution Phase	The optimal method for determining load profile clusters from HH data at feeder level is the "Hierarchical Clustering" Method	Hierarchical Clustering Method will be the primary method used for profile cluster creation with data at MPAN level
Execution Phase	Optimum objective evaluation methods cannot stand alone when determining profile clusters - manual intervention is required using business/ industry knowledge	To fully enable VM-Data as a planning tool - low frequency clusters need to be investigated by manual intervention and the model configured accordingly - e.g., without manual intervention - an EV charging profile could be smoothed and incorporated into a more generic profile cluster. A cycle of manual analysis and reconfiguration would accommodate the current "niche" profile which could proliferate in the future.

6 Risk Management

Our risk management objectives are to:

- Ensure that risk management is clearly and consistently integrated into the project management activities and evidenced through the project documentation;
- Comply with WPD's risk management processes and any governance requirements; and
- Anticipate and respond to changing project requirements.

These objectives will be achieved by:

- Defining the roles, responsibilities and reporting lines within the Project Delivery Team for risk management;
- Including risk management issues when writing reports and considering decisions;
  - Maintaining a risk register;
- Communicating risks and ensuring suitable training and supervision is provided;
- Preparing mitigation action plans;
- Preparing contingency action plans; and
- Monitoring and updating of risks and the risk controls.

6.1 Current Risks

The VM Data risk register is a live document and is updated regularly. There are currently 19 live project related risks and 4 risks which have been escalated to an issue status. Mitigation action plans are identified when raising a risk and the appropriate steps then taken to ensure risks do not become issues wherever possible. In the below table, the details of the project's top five current risks, which have not been escalated to issue, by category, are given. For each of these risks, a mitigation action plan has been identified and the progress of these are tracked and reported.

Table 6-1: Top five current risks (by rating)

Details of the Risk	Risk Rating	Mitigation Action Plan	Progress
The HH profile data is not available in time for analysis	125 Severe	Continuous review of the project plan to ensure sufficient backlog until data becomes available	ElectraLink is reporting weekly to WPD against a data acquisition plan and engaging with IBM to plan backlog accordingly
Customer LCT data obtained insufficient for model validation resulting in reduction in confidence in model results and negative dataset being compromised	100 Severe	ElectraLink is engaged with data collection organisations. Early data from site visits showed a reasonable level of access rates	Initial survey results have been shared with the project team. Site visits currently on hold due to COVID-19 restrictions
Milestones may be missed due to constraints, e.g. staff leaving project and not being replaced by people with the same skill level, or insufficient staff resources available to meet the deliverables to	80 Severe	Due to COVID-19, there is a risk of widespread sickness. However, a separate COVID-19 risk log is being maintained and reviewed every 24 hours to identify any early risks	Continuous monitoring of COVID-19 RAID log

time and budget due to poor understanding of deliverables			
Resources become unavailable during the lifetime of the project and suitable replacements cannot be supplied within an acceptable timeframe	80 Severe	Due to COVID-19, all workforce is currently at risk. However, a separate COVID-19 risk log is being maintained and reviewed every 24 hours to identify any early risks	Continuous monitoring of COVID-19 RAID log
The consumption profiles generated do not cover enough of the population to be useful	60 Severe	Validate model as early as possible and iterate as necessary to ensure we have adequate coverage	Progress to be monitored in the Profiles Modelling stage of the project

Figure 6-1 below provides a snapshot of the risk register, detailed graphically, to provide an on-going understanding of the projects' risks.

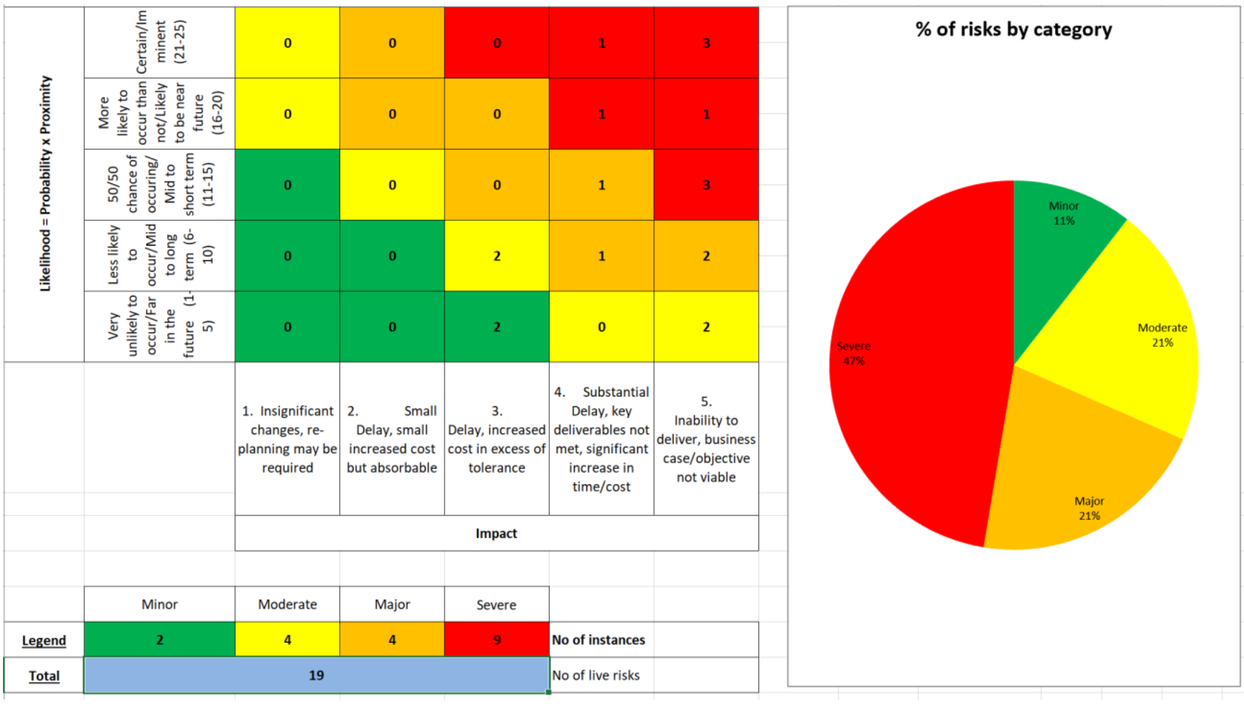


Figure 6-1: Risk Register Graph

Update for risks previously identified

Details of the Risk	Previous Risk Rating	Current Risk Rating	Mitigation Action Plan	Progress
Data required by IBM cannot be supplied to them due to data protection/GDPR regulations	Moderate	Closed	Data Protection Strategy	Risk Closed as IBM now have processes in place to handle Personal Information
Risk that the site visits are not completed in time for the reports to be written for Sprint 2 deliverables	Severe	Raised as Issue	Early planning of logistics	Raised as an issue due to COVID-19
The current data protection does not allow IBM to process unanonymised data	Major	Closed	Meeting with legal team to clarify data approach	Complete, IBM can now process Personal Information
Model validation proves that the datasets used cannot adequately predict LCT uptake	Major	Closed	Include a negative dataset in the model. Include comprehensive HH datasets to train the model. Validate the results with a comprehensive dataset. Include a diverse set of data including information from ONS, and other publicly available data sets that relate to demographics	Not relevant, The team is now collecting data for a negative data set
There are too few records of known Heat Pump installations in WPD records to support the development of a model which can identify existing installations and predict likely areas for future uptake. (estimate only 30% of heat pumps installed are known to WPD);	Moderate	Closed	Adjust scope of project.	Heat pumps have been omitted at the project scoping stage
The model development will be delayed due to illness of core IBM team members	Moderate	Closed	In the event of prolonged illness or absence of a member of the IBM team, IBM will take steps to find a replacement consultant with a similar skill set. The IBM team is documenting findings throughout, commenting on code and storing code in a repository that a new team member can be given access to.	Duplicate risk of more wider team illness
The data protection strategy is not signed off.	Moderate	Closed	Early Engagement	Complete – Risk is no longer valid

7 Intellectual Property Rights

No foreground IP entries have been made to the IPR register during the project's tenure.

8 Consistency with Project Registration Document

The scale and cost of the project has remained consistent with the registration document, a copy of which can be found here:
https://www.westernpower.co.uk/innovation/projects/virtual-monitoring-data-vm-data

9 Accuracy Assurance Statement

This report has been written and compiled by the Arpita Nair (Project Manager – ElectraLink) and Max Hudson (Project Manager – IBM).
This report has been checked by Ricky Duke (Project Manager - WPD)
This report has reviewed and approved by Jon Berry (Innovation Manager – WPD)

All efforts have been made to ensure that the information contained within this report is accurate. WPD confirms that this report has been produced, reviewed and approved following our quality assurance process for external documents and reports.

Glossary

Abbreviation	Term
ADC	Average Daily Consumption
DBSCAN	Density-based spatial clustering of applications with noise
DER	Distributed Energy Resources
DNO	Distribution Network Operators
DSO	Distribution System Operator
DTS	Data Transfer Service
EAC	Estimated Annual Consumption
EMDH	Energy Market Data Hub
EPC	Energy Performance Certificate
EV	Electric Vehicles
GDPR	General Data Protection Regulation
HH	Half Hourly Meter
HV	High Voltage
LCT	Low Carbon Technology
LV	Low Voltage
MPAN	Meter Point Administration Number
PV	Photovoltaic
RIIO	Revenue=Incentives + Innovation + Outputs
SSE	Scottish and Southern Energy
VM	Virtual Monitoring
WPD	Western Power Distribution

Western Power Distribution (East Midlands) plc, No2366923
Western Power Distribution (West Midlands) plc, No3600574
Western Power Distribution (South West) plc, No2366894
Western Power Distribution (South Wales) plc, No2366985
Registered in England and Wales
Registered Office: Avonbank, Feeder Road, Bristol BS2 0TB

wpdinnovation@westernpower.co.uk
www.westernpower.co.uk/innovation

@wpduk