HV Pilot

WP3 D1 Selected Algorithms Report

Version 1.0

29th September 2025

Prepared by:

Andrew Urquhart and Murray Thomson

CREST, Loughborough University

Prepared for:

David Penfold

National Grid Electricity Distribution

Confidential © Loughborough University

This report describes work in progress and must not be circulated beyond the Project Partners as set out in the Collaboration Agreement.

Contents

1. Introduction 3
1.1 Project overview 3
1.2 Document purpose 3
2. Phase identification approach 4
2.1 Single-phase connections 4
2.2 Voltage correlation 4
2.3 Phase labelling 5
2.4 Previous experience from the SMITN project 6
2.5 Correlation method 6
2.6 Confidence metrics 7
2.7 Time-of-day correlation 7
2.8 Differential voltage correlation 8
3. Sequential phase identification method 9
3.1 Preparing smart meter data 9
3.1.1 Spurious values 9
3.1.2 Clock offsets 9
3.1.3 Aligning samples to a common time series 10
3.1.4 Selecting smart meter to include 10
3.2 Validating substation connection records 10
3.3 Phase identification for three-phase substations 11
3.3.1 With substation monitoring 11
3.3.2 Without substation monitoring 12
3.3.3 Three-phase meters 12
3.4 Selecting voltages for HV phase identification 13
3.4.1 Three-phase reference substations 13
3.4.2 Single-phase substations 13
3.4.3 Voltage selections 13
3.5 Associating clusters from multiple three-phase substations 14
3.6 Determining phases of single-phase substations 14
3.7 Load drop compensation 14
3.8 Incremental phase identification 15
4. Feeder-scale phase identification method 15
5. Conclusions 16

1. Introduction

1.1 Project overview

The HV Pilot aims to identify the phases to which two-wire sections of high voltage (HV) overhead line are connected. This will enable more accurate power-flow modelling of the HV feeders and may ultimately lead to changes in the phase connections to minimise phase unbalance. Two techniques will be used. Smart meter voltage data will be used, building on the learning of the SMITN project. The other technique will use helicopter survey data to build a virtual model of the network to enable the connection of wires to be physically traced. These techniques will be developed and tested to determine accuracy and cost effectiveness compared to site inspections.

1.2 Document purpose

This document outlines the key algorithm steps that are expected to be required for phase identification of single-phase transformers and branches for the method based on smart meter data.

The development of the phase identification algorithms will be iterative, and the innovation work will respond to learning gained from using the smart meter data. The steps in this document are not prescriptive and not intended as a checklist for the future analysis. However, based on previous experience, there are a number of steps and techniques that we expect will be required and the planning and design of work package 3 will need to allow for this work.

The method described here can also be considered as a more detailed presentation of the phase identification approach that was initially summarised in the project proposal and then developed further in the project outline document.

This document only describes the phase identification method using smart meter data. A separate document in work package 4 will outline the methods to be developed for phase identification using machine vision.

The approach described in this document follows a rule-based process where an understanding of the network topology is applied at each stage and only the necessary data is carried forward to the next step. This is presented as a starting point for the analysis, and to provide an end-to-end definition of at least one method that could be used, but it may be preferable to combine some of these steps or to defer decision-making to a later stage. An alternative approach is described further in section 4.

2. Phase identification approach

2.1 Single-phase connections

This phase identification method is designed to detect the phases of individual single-phase transformer connections where these are connected to the three-wire HV feeder and the phases of two-wire branches.

Single-phase transformers are connected to two of the three line-conductors of an HV feeder. There is an expectation that many single-phase transformers on an HV feeder will be connected to the same two line-conductors, creating an unbalance of load currents. Although some records exist to show the conductor connections for individual single-phase transformers, these records are incomplete and not expected to be highly accurate.

Some side branches of the HV feeder may only have two of the three conductors, reducing the cost of the HV feeder line. Any transformers connected to these two-wire branches will necessarily be connected to the same two line-conductors. The connections of these two-wire branches to the three-wire line are also not well recorded.

There is also a possibility that the sequencing of the three line-conductors will change along the HV feeder, for example where overhead lines are swapped from left to right either side of a junction pole. Even when the phases of connections to the three-wire line are well recorded, the topology may therefore not be as expected, again leading to possible unbalance of load currents.

These scenarios are summarised in Figure 1 below.

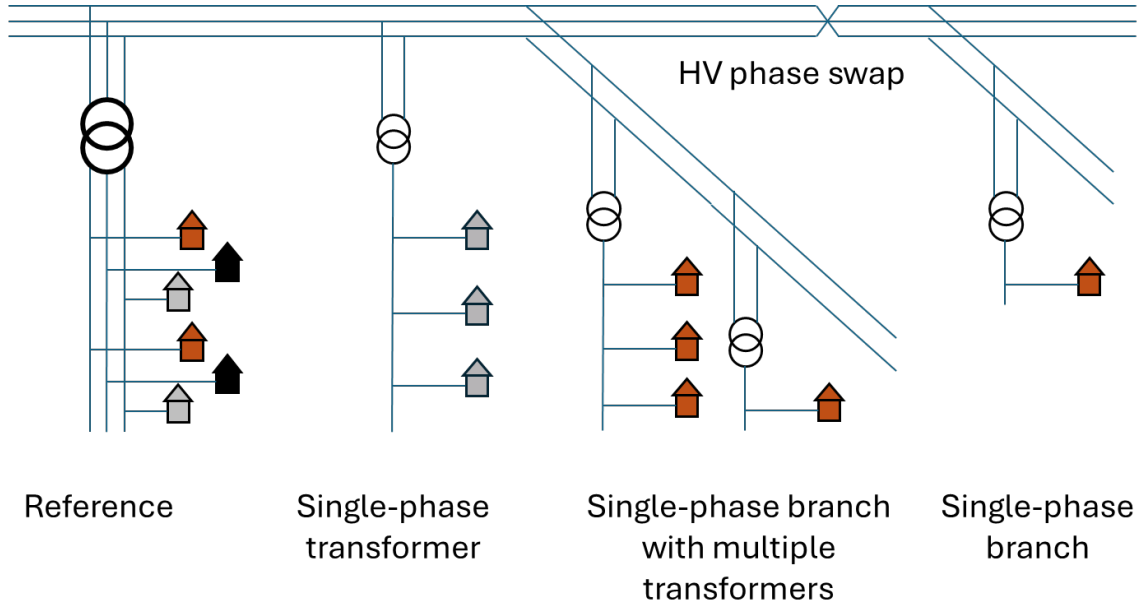


Figure 1: HV feeder with single-phase connections

2.2 Voltage correlation

The phase identification method relies on an assumption that voltages on the HV feeder will have variations over time that are different on each phase. These variations may be due in part to the combined load on each phase of the HV feeder, but there will likely also be voltage differences at the primary substation due to the load unbalance elsewhere on the distribution network.

Phases can be identified if the time-series voltage variations on a single-phase transformer are highly correlated with the time-series voltage variations of one of the phases at a three-phase reference transformer, as shown on the left in Figure 1. Analysis of HV feeder characteristics for the WP1 D1 Area Selection Report has shown that the very few HV feeders have no three-phase substations and so it is reasonable to assume that a three-phase transformer will be available.

Typically, there is no voltage monitoring directly on the HV feeder line, and measurement data will be obtained from sensors on the low voltage (LV) side of the transformers. Each of these line-to-neutral voltages is derived from a specific line-to-line connection on the HV feeder. In effect, the correlation method will identify pairs of HV lines for a single-phase transformer connection that are the same pair of lines forming one of the three phases at the in the delta-star wiring configuration of the reference transformer.

Since this method considers the electrical topology of the line conductors, there is no impact on the correlation whether the lines are on the right or left of a three-wire overhead line. The method is not affected by swaps in the sequence of line conductors but equally provides no information that would indicate how the lines what the sequence of the lines would be.

Split-phase transformers differ from single-phase transformers by providing a neutral point between two halves of the low voltage side transformer winding. This gives two phases on the LV side, with 0° and 180° phase angles. While almost all domestic loads will connect to one of the two phases with a 230 V line-to-neutral connection, older high-power loads can connect between the two lines, originally proving a 480 V supply. However, despite the differences in phase angle of these connections, any connected loads will be affected by the same time-series voltage variations of the two lines connected to the transformer. Both phases of the split-phase transformer can therefore be expected to be correlated with one of the line-to-line pairs, in the same way as if a single-phase transformer had been connected.

Measurements of the LV voltages are available from the smart meters of customers connected to each substation. This data is considered ‘operational’ (whereas the demand data is ‘personal’ unless aggregated) and so can be made available for network analysis purposes within NGED projects. Some substations are also equipped with monitoring, typically of voltage and demand data on each feeder and phase conductor.

2.3 Phase labelling

Voltage data from substation monitoring provides a reference to identify the specific phases, L1, L2, L3 of each conductor, as defined locally for each distribution substation. It is assumed that the voltage probes will be connected correctly, such that the measured data is an accurate representation of voltages on the labelled conductors.

However, these phases are not necessarily the same at each substation and will differ for substations that are on either side of any swaps in the sequencing of the HV conductors.

In the absence of substation monitoring, these may be no information available to label the phase clusters as L1, L2, L3. In this case the phase clusters may be labelled simply as A, B, C groups, and clusters from different substations that are on the same phases will mostly have different labels.

For the purposes of determining unbalance, it is only necessary to identify phase groups that are more heavily loaded, where a large proportion of the load is on the same phase, without the need to determine whether this is L1, L2, or L3.

If any intervention works are planned to address the unbalance, for example by moving a transformer from one phase to another, this would need to be preceded with by installing monitoring so that the correct physical conductors can be identified.

2.4 Previous experience from the SMITN project

In the SMITN NIA project, the voltage correlation method has been found to be successful for identification of phases within the group of customers connected to the same substation. This method can be used without requiring demand data, and either with or without knowledge of the voltages at the substation derived from monitoring data.

Phase identification accuracy can be improved by using substation monitoring data if this is available. Substation monitoring data provides a reference in which measurements for the three phases have already been separated and it is relatively straightforward to determine which of these phases has the highest correlation with the voltages from smart meter data for an individual customer. Where no substation monitoring is available, a clustering method is used to find three groups of customers for which the voltage variations are more highly correlated within the group than with the other groups.

Both techniques, either with or without substation monitoring data, can provide a high phase identification accuracy or 90% or above, but the accuracy of results where substation monitoring data is available are typically higher.

LV phase identification in SMITN used voltage data with 30-minute and then also with 1-minute resolution. High phase identification accuracy can be achieved with either resolution, although the measurement duration will need to be approximately 30 times longer using 30-minute data than with 1-minute data, such that there is a similar number of samples in the correlation calculation.

Phase identification in SMITN was also tested with synthesized 10-minute data, calculated by using 10-sample averages from the 1-minute data, with a corresponding increase in the required measurement duration for a similar accuracy of results.

Results with longer time resolutions are also degraded less accurate than those with higher resolution due to differences in the start time of the averaging periods between meters. Half-hourly voltage records from some meters did begin at 00 and 30 minutes after the hour, but started at other offsets, despite maintaining an approximately half-hourly resolution over time. The correlation method requires comparison of voltages between different meters assuming common half-hourly sampling periods and so the ±15-minute offsets for some meters reduces the accuracy of correlation results. The same effect may occur with higher time resolution but the range of possible time offsets is correspondingly reduced.

2.5 Correlation method

The phase identification method uses the Pearson correlation function

$$R_{xy} = \frac{C_{xy}}{\sqrt{C_{xx} \cdot C_{yy}}}$$

where C_{xy} is the covariance of variable x with variable y, C_{xx} is the variance of x, and where the covariance is defined as

$$C_{xy} = \frac{\sum_{i=1}^{N_t} (x_i - \bar{x})\,(y_i - \bar{y})}{N_t - 1}$$

where N_t is the number of time samples in the correlation calculation, and x_i and y_i are the time-series voltages from the two smart meters.

The correlation coefficient is unaffected by constant offsets on either data series since C_{xy} considers only variations above the mean. A constant scaling applied to either of the series also has no effect

since both the numerator and denominator in the calculation for R_{xy}, are proportional to series x and to series y.

Consequently, a voltage drop along the HV feeder that affects each phase equally will not compromise the correlation, although voltage drops that are variable in time and different between phases, which is a more probable scenario, will reduce the phase identification accuracy if these variations are not common to both the reference and the single-phase substation.

A correlation matrix is calculated containing the correlation of each smart meter with every other of the N_M smart meters at the substation. This provides an $N_M \times N_M$ correlation matrix where the values on the leading diagonal will all be unity (i.e. the correlation of voltages for each smart meter with themselves).

The distance between two voltage time-series is defined as $D_{xy} = 1 - C_{xy}$.

Distances can also be calculated between two clusters of meters, but since the correlation is only defined between individual meters, the correlation between two clusters is calculated as that between a particular pair of meters, one in each cluster. In SMITN, the maximum distance between any pair of meters in two clusters was taken as the distance between the two clusters. However, it is equally possible to define the distance between clusters as the minimum distance between any pair of meters in the two clusters. The SMITN project explored several linkage approaches for hierarchical clustering, including minimum (single-link), maximum (complete-link), average, centroid, and Ward's method. Although these methods differ in how they define the distance between clusters, in practice they produced very similar results for our dataset. As a result, the choice of linkage had no material effect on the clustering outcomes. Neither approach appears entirely robust when the number of time-series samples is smaller and where the correlations are lower.

The SMITN project tested both k-means and hierarchical clustering, with hierarchical clustering providing more reliable results.

2.6 Confidence metrics

The correlation coefficient inherently quantifies the extent to which one series of data can be associated with another, so this may form a confidence metric for the phase identification.

Other possible metrics could include a calculation of the difference between the correlation of the selected phase and the highest of the other phases, giving a measure of the distance between the selected phase and the others.

Where multiple smart meters are connected to a single-phase substation, the phase identification could be calculated for each of them and the confidence metric calculated based on the proportion that associate with the selected phase.

2.7 Time-of-day correlation

Voltage variations on all phases are subject to voltage drops that are common to all phases since the expected customer demand follows a diurnal cycle with peaks in the early evening and lower demand at night. This correlated behaviour may reduce the scope for phases of single-phase substations to be discriminated.

Two possible approaches to minimise this effect would be:

- Calculation of the phase correlations at times of low demand, when the load-related voltage variations are minimised. However, the demand on the distribution network in general will be

lower and so this period may also have less distinctive variations in the voltages at the primary substation.

- Correlation individually for each half-hour of the day. This reduces the impact of diurnal demand variations, since the correlation only compares demand variations within a half-hour period but not between half-hour periods. This approach reduces the volume of data available for each correlation by a factor of 48, which could reduce the accuracy, but provides 48 correlation results rather than just one. A confidence metric could be calculated based on the number of half-hours for which the indicated phase would be the selected phase.

  If the reduction of data series length by a factor of 48 affects the accuracy of results then lesser reductions, for example by correlating for each hour or 3-hour period, could be investigated

2.8 Differential voltage correlation

An alternative means of overcoming the inherent similarity between voltage drops on multiple phases is to use voltage differences. In this case a set of voltage differences $\Delta v_t = v_t - v_{t-1}$ is calculated for each smart meter and these values are used in place of the absolute voltages. In effect, the correlation is then based on the rate of change of voltage amplitude, rather than the voltage amplitude itself.

While the correlation results are independent of differences in the overall mean of either voltage time-series used in the calculation, they will be affected by any variations in the short-term means, such as might occur with the diurnal mean demand profile. The use of differential voltages with, for example, half-hourly voltage data, will in effect de-weight the impact of voltage differences that have increased over multiple half-hour samples, and increase the sensitivity to short-term changes.

Results from the SMITN project demonstrated that the use of differential voltages can provide increased accuracy for phase identification. This approach can also be tested for HV Pilot.

3. Sequential phase identification method

This section describes one possible approach to phase identification. This approach could be described as sequential as the phases are first identified with the context of an LV feeder or substation, and then phases from a reference substation are associated with phases at a single-phase substation.

3.1 Preparing smart meter data

3.1.1 Spurious values

The analysis will use voltage data from smart meters with either 30-minute or 10-minute resolution. Initially the meters will default to 30-minute resolution and so historical voltage data of up to 90 days duration can be collected.

It is then planned to switch the meters to 10-minute resolution, which will reset their historical record. A further set of 10-minute voltage data with 30 days duration can then be collected.

Based on experience from the SMITN project, there is an acknowledged risk that some meters will not be re-configured successfully and may either fail to report voltage data or retain the 30-minute resolution setting.

Voltage data has been found to include occasional extreme high or low values. In some cases these appear to relate to data logging faults as the values are orders of magnitude higher and these samples should be removed. In other cases, the high or low values may be genuine and caused by local fault conditions. These samples are mostly not representative of the voltage on the HV feeder and so should also be removed.

The exact thresholds for this data cleansing will need to be defined when the smart meter data has been received, but values outside of a range 210 V to 260 V are likely to require removal.

Interpolation to fill the missing samples is not necessarily required since it provides no additional information content. However, interpolation may be needed if the absence of samples from one meter would cause samples for the same timestamp from other meters to be excluded, such that further information would be lost from the analysis.

3.1.2 Clock offsets

Some smart meters have been found to have slight offsets in their clock times, such that measured voltages are recorded with a timestamp that is either ahead or behind the correct time. Voltages are correlated between meters on the basis of a common timestamp so these clock offsets reduce the accuracy of the correlation results.

A correction can be made for clock errors by finding calculating the correlation between voltages for any one meter with an average voltage derived from other smart meters. This correlation can be calculated with timing offsets of, for example plus or minus 10 sample periods. The optimum timing correction is then selected so that the correlation of the one smart meter being tested with the others is maximised.

Experience from SMITN showed that a small number of meters had clock offsets that were detectable with 1-minute resolution data, so it is unlikely that this will be a significant concern where the sample resolution is 10 minutes or higher.

3.1.3 Aligning samples to a common time series

As noted above in section 2.3, some smart meter voltage has been found to have time intervals that do not begin at integer multiples of the period length within the hour. For example, half-hourly samples may begin at 10 minutes and 40 minutes after the hour, rather than 00 and 30 minutes.

Either of two approaches can be taken to resolve this.

1. Samples with an offset are assigned new timestamps with the time amended to the nearest integer multiple of the averaging period. For examples, samples beginning at 10 minutes after the hour would be assigned 0 minutes, and samples beginning at 20 minutes after the hour would be assigned 30 minutes.

2. Half-hourly samples can be interpolated and filled to 1-minute resolution and then down-sampled to half-hourly resolution over the required 00 and 30 minute periods.

   It has been found that rounding the number of seconds for each timestamp to zero has minimal impact on the results and allow the interpolation process to use a period of minutes rather than seconds, allowing for faster processing.

3.1.4 Selecting smart meter to include

Ideally the data would include a sample from every smart meter MPAN at a substation in every timestamp period

In practice, there will be some MPANs with significant gaps in the available voltage data. If these smart meters are included in the set, there will be few samples for which there is voltage data for all, making the correlation results less reliable

Removing some smart meters from the set may increase the overall number of samples available for use in correlation. One possible process would be as follows:

- Collate data from all smart meters on the LV feeder or at the substation

- Find the minimum and maximum timestamp for any smart meter, and therefore the maximum possible number of samples from any smart meter in the data set

- Select only those smart meters that have data available for a defined percentage of the data set

Setting a very high data availability threshold may result in inaccurate results if there are many smart meters excluded. Equally, setting a very threshold may give inaccurate results if very few samples have data available for all the smart meters. A threshold of around 75% is suggested as a starting point.

3.2 Validating substation connection records

Customer records in CROWN do not always record the correct substation and LV feeder for each customer. This has two impacts on the phase identification method as used here:

1. The correlation clustering results for a selected LV feeder will be degraded if smart meters that are in practice connected to other substations or LV feeders are included.

   Ideally the analysis should exclude any smart meters that are recorded as being on the selected LV feeder but which are actually connected elsewhere. However, it is less important to include smart meters that are recorded as being connected elsewhere but that are actually

on the selected LV. Omitting these smart meters will reduce the number of smart meters available for the analysis, but the impact is no worse than if the customers did not have a smart meter installed.

2. If demand data is used for load drop compensation of the smart meter voltages, the aggregated demand data should only include customers from the correct LV feeder.

   In this case it is beneficial to exclude smart meters that are not on the selected LV feeder and also to include smart meters that are recorded as being connected elsewhere.

This concern was addressed in the SMITN project through the feeder identification method, again based on smart meter data. An alternative method has been developed where the distance between the customer location and the nearest point on the assigned LV feeder is calculated and compared with the distance to other LV feeders from nearby substations.

The assigned LV feeder may be assumed to be correct if the nearest distance to the customer location is below a defined threshold, for example 10 metres.

If the distance between the customer location is greater than this, and if another LV feeder is nearer and within an upper limit, for example 200 metres, the customer can be assumed to be connected to the other LV feeder.

For the purposes of clustering voltage data, it is only necessary to exclude smart meters that are unlikely to be on the selected LV feeder, and there is no need to determine where they are actually connected.

It is preferable to set thresholds that give a high confidence of excluding erroneous meter connections, accepting that this may risk removing some meters that are actually connected. Provided sufficient smart meters still remain, the degradation in results caused by including erroneous meter connections will be more significant than the loss in accuracy through having fewer meters to include. Smart meters with long service cables are also likely to be poor representatives of the voltage at the substation busbar.

If is expected that this reduction in the number of smart meters for use in the analysis will be less of a concern in future as the uptake of smart metering progresses. The SMITN BAU deployment should also reduce the risk of substation and feeder connections being incorrectly recorded.

3.3 Phase identification for three-phase substations

3.3.1 With substation monitoring

It cannot be assumed that substation monitoring data will be available for a business-as-usual application of the phase identification method. However, there will be some feeders where monitoring has been installed at one or more substations, and this data should be used where possible.

For this project HV feeders have been selected such that at least one substation will have monitoring. This allows for validation of the phase clusters that would be calculated based on smart meter data alone.

The method for each smart meter involves the following steps:

- Identify a set of samples for which data is available from both the smart meter and each phase of substation monitoring data
- Calculate the correlation between the smart meter voltages and each phase of the substation monitoring data

- Assign the smart meter to the phase with the highest correlation

3.3.2 Without substation monitoring

Phase identification without substation monitoring will likely follow similar methods to those developed in the SMITN project. The method involves the following steps:

- Calculation of the correlation between voltages for each smart meter with every other smart meter at the substation
- Initially, each smart meter is assigned an individual phase cluster
- Pairs of clusters are joined, where the correlation between them is highest.
- When clusters with multiple smart meters are joined, this may be on the basis of the highest minimum correlation between any pairing of a meter from one cluster with a meter from the other cluster.
- This process continues until only three cluster remain.

This method was developed based on substations in Milton Keynes, mostly a suburban area, where the LV feeders typically have many customers and therefore higher capacity feeder mains. In the V2G NIA project, the phase identification method was applied to smart meters in areas that are more rural, and more similar to those being considered for the HV Pilot project. A modified method was developed where the phase clusters were calculated separately for the customers on each LV feeder.

Once three phase clusters have been identified for each LV feeder, these clusters can be associated together to produce three clusters for the entire substation. One approach would be to use the same clustering method as used within the feeders, with two additional constraints:

- Clusters can only be associated with clusters from another LV feeder (i.e. it is assumed that the clusters from one feeder must be on different phases)
- Once a cluster has been associated with one cluster from an LV feeder, it cannot be associated with another cluster from the same LV feeder

3.3.3 Three-phase meters

Some LV feeders will have connections for three-phase smart meters. Ideally the voltage data from these meters could be used in the same way as for substation monitoring, assuming that the three voltage channels provide a correlation reference for which the time-series data is known to be on the three phases. However, experience from the V2G project showed that this is not always the case, and there can be three-phase meters on LV feeders where only one or two phases are present. The second or third phases may be either disconnected or linked to one of the other phases. Where three phases were clearly present, they were not found to correlate highly with any of the phases of other nearby meters. This may be due to high service cable voltage drops for customers with three-phase meters, who are also likely to have higher than average demand.

Despite these concerns, three-phase meters could potentially provide a useful phase reference in the absence of substation monitoring so further investigation within the HV Pilot project is recommended.

3.4 Selecting voltages for HV phase identification

3.4.1 Three-phase reference substations

As a minimum, the phase identification method requires one smart meter on each phase at the reference substation

If substation monitoring is available at the reference three-phase substation, the phases of the connected smart meters can be assigned by correlation with the monitoring data. The reference substation therefore needs to have only sufficient meters such that one will be found on each of the three phases.

If there is no monitoring at the reference substation then a greater number of smart meters will be needed so that phase clusters can be formed. A greater number of smart meters provides higher confidence that there is a high correlation between meters in the same phase cluster and a lower correlation between meters in different phase clusters. The HV feeder selection for the project has required that the reference three-phase substation should have 12 or more customers having a smart meter.

The selected HV feeders have many more substations than the minimum required, and the selected algorithms should be applied to all where possible.

Where more than one smart meter is available on each phase, or in each phase group if the phase labels are not known, there is a need to select which data will be used to represent the substation as a voltage reference. Ideally this should provide a reference that is as close as possible to the voltage at the substation busbar, minimising the impact of load-related voltage variations along the LV feeders that would not correlate highly with voltage variations at other substations on the same phase.

3.4.2 Single-phase substations

Since all the smart meters at a single-phase substation are on the same phase, a minimum of one smart meter is required. However, the selected HV feeders have at least three smart meters to increase the likelihood that data will be available to cover the same time period as for the voltages at the reference substation.

3.4.3 Voltage selections

A voltage time-series is required for each phase from the reference substation, and for the one phase at the single-phase substation for which the phase is to be determined.

The selected voltage reference could be voltages from:

- One smart meter that is closest, either in terms of the electrical path length, or with the lowest accumulated resistance, to the substation
- One smart meter that has the highest correlation with other smart meters in the phase group
- One smart meter that has the lowest annual load, as determined from the estimated annual consumption (EAC) from customer records in CROWN, such that voltage variations along the service cable are minimised
- A smart meter customer that does not have an electric vehicle (EV), such that voltage variations along the service cable are minimised
- The average voltage of all the smart meters in the phase group

Alternatively, rather than selecting just one smart meter to provide the voltage data for a single-phase substation, the correlation could be tested for all smart meters that are available. It would be expected that all would correlate with the same phase at the reference substation. If some smart meters at a single-phase substation correlate more highly with different phases at the reference substation, this may indicate that the substation has a three-phase transformer.

3.5 Associating clusters from multiple three-phase substations

There will mostly be more than one three-phase substation on an HV feeder, and on the project feeders there are at least two.

Phase identification results would ideally be the same regardless of which three-phase substation is used as a reference. However, phase identification results may be unreliable for three-phase substations that are very distant from a single-phase substation.

The correlation between phase groups of multiple three-phase substations should be tested. This will:

- Determine how phase groups should be associated, for example smart meters in phase group A at one substation may be on the same phase as group B at another substation
- Provide a confidence metric on the correlation of phase groups between substations. For example, if phase groups at two three-phase substations are highly correlated then it is reasonable to a high correlation with single-phase substations that are between the two. Otherwise, it may be that load-related voltage drops along the HV feeder have reduced the correlation and voltages at single-phase substations may only correlate highly with three-phase substations that are nearby.

3.6 Determining phases of single-phase substations

With the previous steps completed it is relatively straightforward to select the phases of single-phase substations. The selected voltages from the single-phase substation are correlated with voltages in all three phase groups from the reference substation and assigned to the phase group for which the correlation is highest.

3.7 Load drop compensation

If voltage drops within the LV feeder are found to degrade the results of the phase identification process it may be helpful to apply a compensation where the demand and impedances are known. Provided there are more than a minimum of 3 customers, the half-hourly aggregated demand for the feeder can be provided by smart meter data. This represents the demand in the feeder cable from the substation to the first junction or service cable connection from which there is more than one cable route to the aggregated set of smart meters. Although this may represent only a small section of the overall LV feeder length, it is likely to be a section that dominates the voltage drops since it carries the greatest aggregated current.

Customers with non-smart meters are not represented in this data. This might be addressed by synthesizing their demand data using Elexon profiles. Alternatively, compensating for only those customers for which the demand is known may be advantageous even if a more complete compensation for all customers is now possible.

This method of load drop compensation appears most applicable to the single-phase substation since the aggregated demand all occurs on the same phase conductor. It is less straightforward to apply this method at the reference substations as the demand is, by default, aggregated across all three phases such that the voltage drop in any one phase is not known.

Once the smart meters have been assigned to individual phases, it may be possible to obtain aggregated demand per feeder per phase, such that a load drop compensation would also be possible at the reference substation. However, this would require new aggregated demand data to be obtained. This may be viable for a BAU application but may not be possible within the timescales of this innovation project.

Provided that the number of customers is still sufficient to maintain privacy, custom aggregation groups could potentially be defined for customers on specific lateral branches of a single-phase substation or at the reference substation. This would increase the accuracy of the load drop compensation method.

3.8 Incremental phase identification

On long HV feeders it is possible that the voltage drops between the reference substation and remote single-phase substations will reduce the phase identification accuracy.

A possible method to address this could involve identifying the phases of nearby single-phase substations and then using those as references for further substations along the feeder.

Clearly this method depends on finding substations that are assigned to all three phases, unless it is found that a low correlation can be interpreted as indicating that a substation is not on the same phases as neighbouring substations, and the phase determined by process of elimination.

If the overall aim is to identify unbalance where many substations are connected to the same phase, it may not be necessary to determine the connections for substations that are on either of the two less heavily loaded phases.

4. Feeder-scale phase identification method

The previous section has outlined a method where correlations are calculated first within the group of smart meters on an LV feeder at the reference substation, then to associate the phases from one LV feeder with those of another, and then to compare with the voltages at a single-phase substation. The voltage data is reduced in volume at each stage with a single time-series being selected to represent each phase at a substation.

An alternative approach could consider the phase correlations on a wider scale, with the correlation being calculated between every pairing of smart meters on the HV feeder. The clustering could then be performed on this extended set of data. It would be expected that all smart meters at a single-phase substation, or on single-phase two-wire HV feeder branch, will be associated with the same cluster. Assuming that there is certainty that the substations are single-phase, the proportion of smart meters that correlate with the same phase cluster would provide a confidence metric for the results.

5. Conclusions

This report proposes a method of phase identification using voltage data from smart meters that could be adopted for the HV Pilot project.

Experience from previous projects is also included where this may be informative for the HV Pilot project.

It is expected that the development of the phase identification method will develop learning and new ideas as this progresses. The method here is a starting point and is intended to provide an agreed baseline method that can be developed initially and then improved based on the results obtained with real smart meter data from the selected HV feeders.

It is expected that all the techniques proposed here will be tested during the development of phase identification method, although there is no apparent need for every possible combination of these to be explored. The development will proceed using a simple approach where possible, adding in additional steps where this appears necessary to increase the confidence of accuracy of the results. For example, compensation for clock offsets or the timing drift between sample intervals may improve results, but it may be that these additional steps can be omitted without noticeable loss of accuracy.

However, from the perspective of a BAU implementation, it would be desirable to minimise the duration of data needed for the correlation calculations as this will reduce the volume of data that would need to be collected from smart meters. The development of the phase identification method should therefore aim to extract the maximum information content from the available data, rather than to minimise the amount of computation required once this data has been collected. For example, if it is found that using demand data for voltage drop compensation gives minimal improvement in accuracy, then it would be preferable to remove this from the method. Conversely, voltage drop compensation would be helpful if it allows for a shorter duration of voltage data to be used.

It is also possible that different techniques will give different phase identification results. In the absence of any reliable reference data where phases are known, it will be assumed that results with the greatest consistency are the most likely ‘true’ phase identification. Consistent results will provide the same phase identification when considering samples taken over different periods of time or using different times within a day. Comparison of results with those form work package 4 will also assist in determining which combination of these techniques give a more accurate result.