The future

Architecture of Tools for Load Scenarios (ATLAS)

Data Processing Methodology

3 October 2016

CONTENTS

VERSION HISTORY

Version	Date	Author	Status	Comments
V1	01/08/2016	Christos Kaloudas	Completed	
V2	09/09/2016	Rita Shaw	Completed	Shared with internal Steering Group and Element Energy.
V3	03/10/2016	Christos Kaloudas	Completed	Updated based on feedback. For website.
				Future updates to reflect changes in Q data processing and revised estimation of non-monitored generation output.

EXECUTIVE SUMMARY

This report corresponds to the Deliverable “Data Processing Methodology” of the Architecture of Tools for Load Scenarios (ATLAS) project at Electricity North West Ltd., funded under the Network Innovation Allowance (NIA) scheme. A key part of the scope of the ATLAS project is to propose a demand forecasting methodology covering how to forecast active (P) and reactive (Q) power and related metrics in a set of scenarios for all Electricity North West Grid Supply Points (GSPs), Bulk Supply Points (BSPs) and primary substations.

This report presents the Data Processing methodology that is used in ATLAS to adjust monitoring time-series data of P demand that might be erroneous, missing or not representative of the actual demand of a substation or group of interconnected substations.

The key characteristics of the proposed Data Processing methodology are that:

- it involves both half-hourly and daily analyses to deal with monitoring data issues including step changes from switching operations and network reconfigurations, missing and/or erroneous data (e.g., half-hourly spikes and/or dips);
- it is generic and can be applied even without any extra network connectivity data;
- it allows minimum recalibration when applied to GSPs, BSPs or primary substations; and
- it has a modular structure that provides flexibility in its use (e.g., selection of smoothing of demand oscillations and/or purely half-hourly analyses).

The proposed methodology is tailored to the monitored component of true P demand data (i.e., the combination of measured demand with metered export from local generators). A tool has been successfully developed and used to process and improve a 5-year period of half-hourly monitored P demand data of over 70 BSP and 380 primary substations at Electricity North West Ltd. An adapted version of the methodology is also being used to process Q demand data.

A practical implementation flowchart is provided, which demonstrates how the proposed methodology could be used to analyse a past year’s monitoring data. This highlights the approach to sense-check the results of the automatic processing compared to the previous year’s peak. In limited cases where the available monitored P and Q demand data is significantly incomplete, this process uses the half-hourly processed demand data together with more detailed manually obtained data (e.g., measurements of HV feeders, connectivity data of DG units) to obtain half-hourly demand profiles that can be used for further studies.

Given that both transmission and distribution operators are interested in more accurate assessments of the underlying true P demand, this report also presents an initial approach to estimate half-hourly the outputs of non-monitored distributed generation. The approach is implemented for over 30,000 DG units, which are connected downstream primary substations across Electricity North West’s region and for which no monitoring data is available to the DNO.

The aggregated latent demand of these DG units is combined with processed P demand data (i.e., monitored component of true demand) across all primaries of Electricity North West to provide a more accurate estimation of the true demand (i.e., half-hourly aggregated primary demand from financial year 2012 to 2016).

The developed prototype tool of the proposed Data Processing methodology will be further used in the ATLAS project to:

- support the identification of historical trends of P demand, during periods of peak and minimum load using different periodicities (e.g., seasonal trends or individually per month); and,
- provide a representative set of daily half-hourly demand profiles across all (or most) GSPs, BSPs and primary substations of Electricity North West Ltd, as the basis for scenario development and network modelling.

1 INTRODUCTION

This report corresponds to the Deliverable "Data Processing Methodology" of the Architecture of Tools for Load Scenarios (ATLAS) project at Electricity North West Ltd., funded under the Network Innovation Allowance (NIA) scheme.

The ATLAS project has been established to achieve four primary objectives:

- enable well-justified Strategic Planning of network capacity;
- provide historical load analysis and forecast scenarios to demonstrate efficient load-related expenditure in Ofgem's RIIO-ED1 control and build a well-justified business plan for RIIO-ED2;
- more accurately and credibly meet annual reporting obligations to National Grid and Ofgem; and,
- provide better information to stakeholders and customers on future loading, enhancing connection customer experience.

To achieve these objectives the ATLAS project proposes a demand forecasting methodology for active (P) and reactive (Q) power and related metrics in a set of scenarios for all Electricity North West Grid Supply Points (GSPs), Bulk Supply Points (BSPs) and primary substations.

This report presents the Data Processing methodology, which can be used to process historical time-series monitoring data of P demand that might be erroneous, missing or not representative of the actual demand of a substation or group of interconnected substations.

The demand definitions, as well as the available half-hourly monitoring demand data of Electricity North West and previous approaches to process this data, are presented in section 1. In section 2 the proposed data processing methodology is explained in detail and demonstrated using monitoring P demand data of BSP substations of Electricity North West.

Practical insights regarding the estimation of non-monitored latent demand from distributed generation that suppresses P demand of substations are described in section 3. In section 4 practical insights and recommendations for the application of the proposed methodology, as well as future work are presented.

Finally, conclusions are drawn in section 5.

1.1 Demand Definitions

Body text and instead definitions as in the Engineering Recommendation (EREC) Engineering Technical Report 130 (ETR 130) [1] will be adopted.

According to EREC ETR 130:

- **Measured demand** is the summated demand measured at the normal (network) infeed points to the network for which Group Demand is being assessed. In practice this is historically the measured value from SCADA (i.e. FLA system of Electricity North West Ltd). It should be noted that for generation-dominated networks, the measured demand may actually be negative, but the lack of a consistent sign-convention for metering data may make this unclear.

- **Latent demand** is the demand that would appear as an increase in measured demand if the distributed generation (DG) within the network (for which the Group Demand is being assessed) was not producing any output. Latent demand is thus the combination of the output of both half-hourly metered export generators (known in theory from metering data and processed by Electricity North West's Demand and Generation dashboard) and other generators suppressing demand at customer premises (described and estimated in ATLAS).

- **Group (or true) Demand** is a DNO's estimate of the maximum demand of the group being assessed for ER P2/6 compliance with appropriate allowance for diversity [2]. True or group demand, measured demand and latent demand are related in accordance with EREC ETR 130 as shown in Fig. 1. The demand (i.e., true demand or group demand) is equal to the measured demand plus generation (i.e., latent demand).

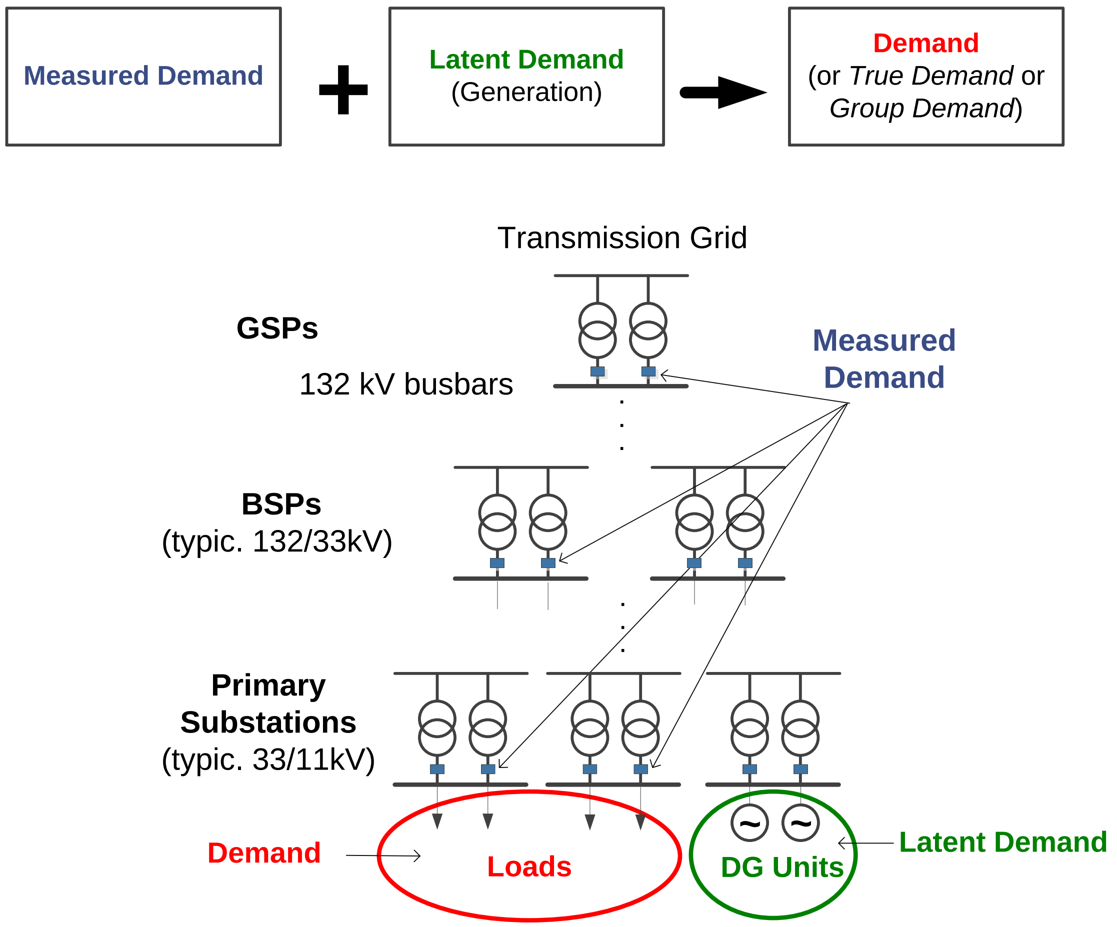

Fig. 1. Demand, measured demand and latent demand.

The Group Demand at GSPs must be consistent with the demand data submitted to a transmission company under the terms of the GB Grid Code. In practice this means that Group Demand needs to be corrected to system-normal and then adjusted to weather-correct outturn values to Average Cold Spell (ACS) conditions for weeks 44 to 12, and to average weather for the rest of the year. However, weather correction approaches are not considered in this report.

The proposed data processing methodology in this report is tailored to be used for time-series of *demand* data (true demand).

1.2 Monitoring Data at Electricity North West Ltd

Electricity North West Ltd follows a similar approach to other British Distribution Network Operators (DNOs) in storing half-hourly averages of measured and latent demand data that are further analysed for network planning studies. The derivation of monitoring time-series demand data and the associated data issues are further described in this subsection.

Derivation of Monitored Demand Data

Monitoring sensors for voltage (V), current (I), active (P) and reactive (Q) power are used to get half-hourly averages of P and Q measured demand data at GSPs, BSPs and primary substations, as well as latent demand for most large-sized DG units.

In Electricity North West's Demand and Generation Dashboard system, half-hourly V and I data from the SCADA system are first used to assess half-hourly apparent power (S) data. It should be noted that the V and I data are accurate and according to [3]:

- at 5% of the current transformer (CT) rating, the largest error was found 0.68%;
- at 100% of the CT rating, the largest error was found 0.26%; and,
- at 120% of the CT rating, the largest error was found 0.13%.

Currently, P and Q measurements are suspected to be of lower accuracy, so this monitoring data is instead used to produce half-hourly estimates of power factor (PF). These PF estimates are combined with the apparent power (S) assessed using the V and I measurements to produce the P and Q demand data that appears as **measured demand** in Electricity North West's Dashboard of monitoring data. This ensures that P, Q and S values are consistent. *Given the desire for higher accuracy of both P and Q monitoring It should be highlighted that this process for deriving measured P and Q is under review, as is the underlying accuracy and calibration of the monitoring units.*

The corresponding **true demand** values on the Dashboard of monitoring data are produced by extracting the generation P and Q data that are associated with metered generators connected to every GSP, BSP and primary substation, and then combining this with the measured demand as described in section 1.1 above. Since a drop in measured demand could occur due to a rise in generation (latent demand), the underlying time-series of true demand is expected to be more stable (i.e., less oscillatory) than the time-series of measured demand.

This true demand data from Electricity North West's Dashboard has been used as the input (i.e., unprocessed data) of the proposed data processing methodology in this report.

Data Issues

Monitoring data might in many cases be erroneous or not representative of the actual demand of a substation or group of interconnected substations. In general, the type of issues related with monitoring demand data can be grouped as:

- spikes – i.e., half-hourly extreme positive or negative values compared to the adjacent half-hourly demand values, which are obviously spurious;
- flat demand periods (i.e., consecutive half-hourly points with the same demand) and/or oscillations around a constant value – i.e., due to telecommunication issues or missing data (e.g., zero values); and,
- step changes in demand, due to temporary switching operations or more permanent network reconfigurations.

It should be noted that, unlike instances seen in other DNOs, the stored half-hourly data in the Demand and Generation Dashboard of Electricity North West Ltd do not exhibit linear interpolation issues in cases of missing data (i.e., seen as a constantly increasing/decreasing demand between two known demand points).

A previous internal work at Electricity North West Ltd [4] highlighted the above mentioned data issues and a potential processing methodology. That methodology was never implemented, largely because it proposed a semi-automatic approach to data cleansing involving manual processes, which was difficult to practically deliver.

The next section describes the methodology proposed in the ATLAS project that can be used for the automatic processing of time-series monitoring data of demand.

2 DATA PROCESSING – METHODOLOGY

Distribution network planning has been traditionally focusing on periods of peak demand (e.g, in planning network reinforcements). Nonetheless, there are currently a number of significant issues that may require network planners to use more detailed time-series (e.g., half-hourly) data of demand. Such issues include:

- the effects from different mixes of demand and generation, taking into account the particular profiles of historical true demand, low carbon technologies (e.g., e-vehicles and heat pumps) and DG types (e.g., irradiance based PV and stochastic wind profiles);
- the transmission-distribution interactions during periods of both peak and minimum demand considering the individual and combined demand profiles of substations; and,
- the improved understanding of demand characteristics and trends on substations with different characteristics (e.g., domestic vs non-domestic customer dominated substations).

This section describes the proposed methodology, which uses time-series (e.g., half-hourly) monitoring data of demand to identify the half-hourly substation demand in a system-normal configuration. The processed demand data is the output of the suggested implementation. This data can be used as the basis for scenario development in demand forecasting and network modelling.

The following subsections describe all processes of the proposed methodology. Five years historical half-hourly monitored BSP P demand data of Electricity North West are used to demonstrate its application.

2.1 Methodology Overview

An overview of the proposed data processing methodology is shown in Fig. 2. The suggested approach consists of 4 stages:

- **Stage 1**: the time-series monitored demand data is first *split into blocks*. A block corresponds to the time-series data between two time instants that exhibit a significant step change in demand. Each block will be further treated in next stage (Stage 2) as a potential period of demand that needs to be shifted up or down.

- **Stage 2**: a control algorithm that includes 3 modules (i.e., Controls #1 to #3) is implemented using *half-hourly analyses* and the following logics:
    - Control #1: eliminate extreme demand values, then identify and uplift blocks with zero, oscillating around zero and negative demand;
    - Control #2: identify and shift blocks (up or down) where step changes are due to switching operations and/or network reconfigurations; and,
    - Control #3: curtail peaks and minima that have derived from the shifted blocks in controls #2 and #3;

- **Stage 3**: further processing of the time-series demand data is carried out using *daily analyses*. The grouping of demand data in blocks defined in Stage 1 is no longer taken into account. Any zero (or oscillating around zero) P demand data that have been uplifted in Stage 2 are given typical time-series behaviour (i.e., normalised profiles and daily peaks/min). The time-series demand data are further processed in Stage 3:
    - to be first corrected for switching operations and/or network reconfigurations; and,
    - to then be further corrected using a demand smoothing process.

- **Stage 4**: Although the demand profiles are expected to be representative of the actual demand after the application of Stage 3, there can be a limited number of substations where unreasonable step changes remain in demand data. Therefore, in Stage 4 the processes of Stages 1 and 2 are repeated, using the processed data of Stage 3 as input, and different settings for the control coefficients relative to Stage 2. Outputs of Stage 4 are expected to correct the remaining step changes in demand after Stage 3.

The following subsections describe in more detail each stage of the developed data processing methodology.

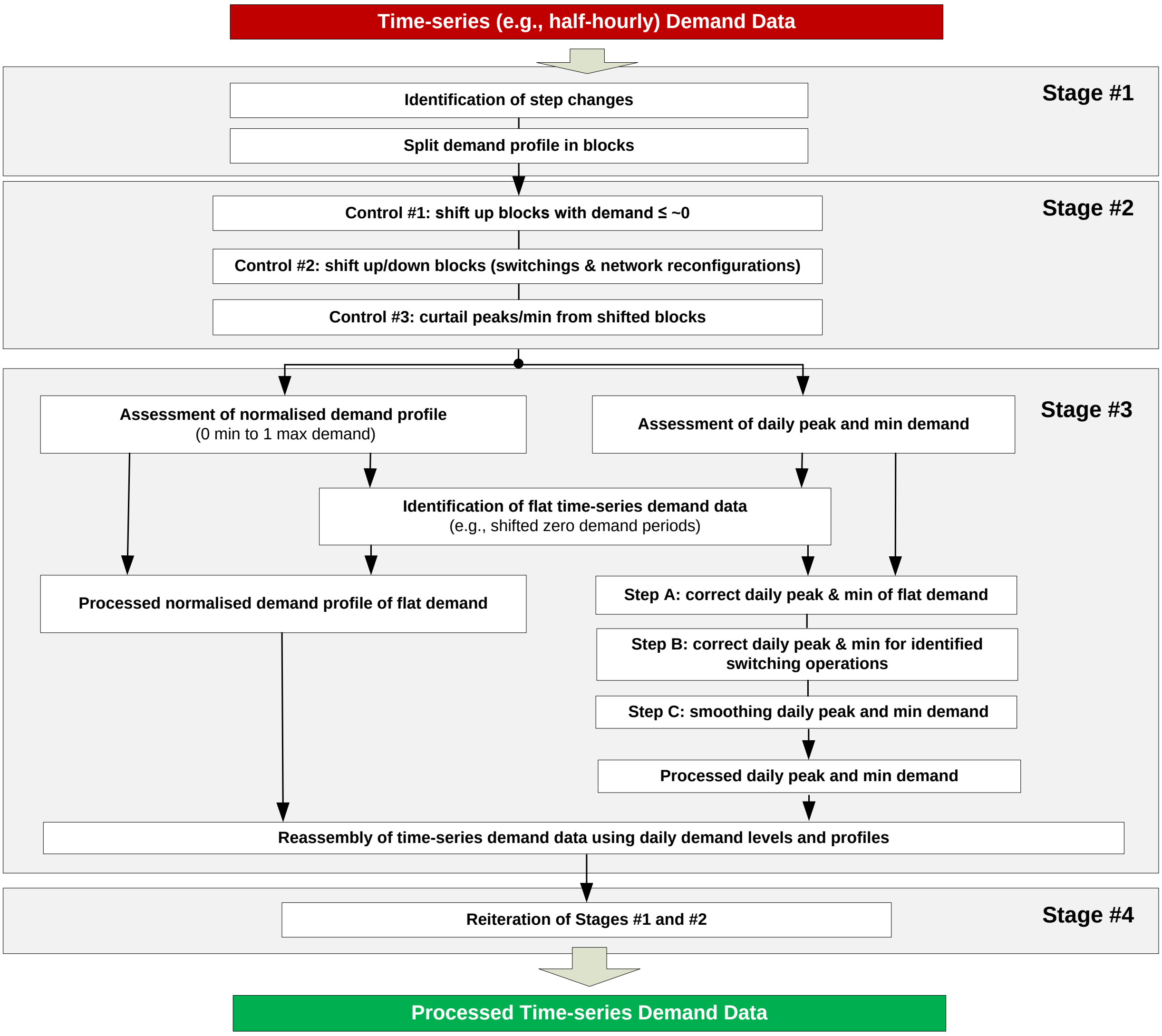

Fig. 2. Overview of the data processing methodology.

2.2 Stage 1: Identification of Half-hourly Step Changes

The aim of Stage 1 of the proposed data processing methodology is to identify half-hourly step changes in the time-series monitoring demand data.

A step change can be caused by many possible factors (i.e., errors in telecommunication system, half-hourly 'spike' errors, switching operations and/or network reconfigurations). *It can be identified as an abnormal change of demand value between two consecutive half-hourly time instants*.

Time-series monitoring data can be seen as a series of demand data. Considering P demand data, the vector **P** contains the whole series of P_i demand data for each i-th time-series point (e.g., half-hourly) within a historical period of N points (e.g., N=17,520 half-hourly points for 1 year of historical data).

To identify the step changes, the d**P**/dt vector is assessed using the **P** vector, where dt is equal to the time step of the monitoring data (e.g., half-hour). Each element dP_i/dt (for i=1 to N-1) of the new vector represents the difference of demand values between two consecutive half-hourly (hh) points. The dP_i/dt elements with high absolute values could correspond to a potential step change. The selection of a proper threshold for the absolute dP_i/dt values is not an easy task, since the monitoring data might include long periods of zero time-series demand data or deviation of this data could be different across examined substations.

Given that the deviation in demand between two consecutive half-hourly points (e.g., P demand at 5 am and 5.30 am in the morning) cannot be as significant as within a day (e.g., P demand at 5 am and 3 pm of the same day), the adoption of a percentage of the standard deviation (SD) of the **P** vector – SD(**P**) can be a reasonable threshold. Iterative calculations using the 2011 to 2016 historical half-hourly P demand data for over 70 BSPs and 380 primary substations of Electricity North West Ltd have revealed that a ±80%·SD(**P**) threshold can adequately identify most of the step changes.

Fig. 3a and Fig. 3b show the half-hourly monitoring **P** demand data and the corresponding d**P**/dt data, respectively, of an Electricity North West BSP. In Stage 1 of the data processing methodology, the P demand profile of Fig. 3a is split into blocks. Each block is defined between two consecutive dP_i/dt that exceed the ±80%·SD(**P**) threshold. Fig. 3a shows the mean value of each block (green coloured curve), where it is evident that many of these blocks correspond to spikes, zero demand values (or oscillations around zero demand) and switching operations.

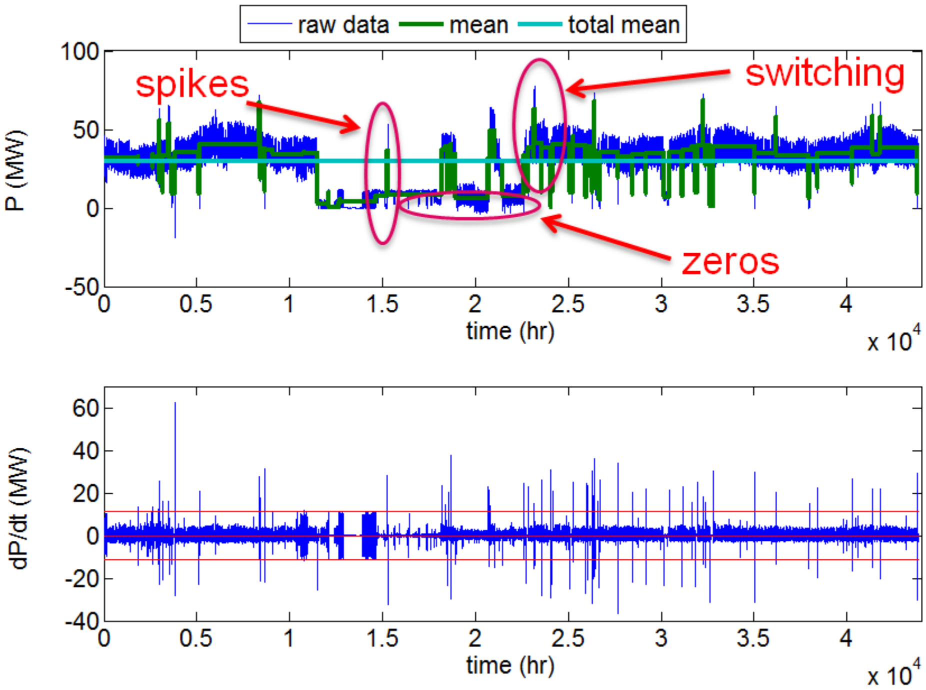

Fig. 3. Half-hourly profiles of a) P demand and b)dP/dt demand data.

Having identified potential step changes and segmented the time-series demand data in blocks, the methodology continues with Stage 2, which is described in the next subsection.

2.3 Stage 2: Half-hourly Processing Using Control Modules

The segmentation of the time-series demand data in blocks (Stage 1) is used in Stage 2 to shift, if and when needed, the time-series demand data of one or more blocks. The decision on which blocks need to be shifted is taken following a control algorithm that consists of 3 modules, i.e. Controls #1 to #3, which are described in the following subsections.

Control #1: Dealing with erroneous or missing data

The first control module (Control #1) of Stage 2 is used to uplift the blocks that exhibit zero, oscillating around zero or negative demand. It is noteworthy that errors in the telecommunication system or in general loss of monitoring data might lead to short or longer periods with zero demand or oscillations around zero demand. Additionally, active power (P) true demand data cannot exhibit negative values and therefore it is highly possible that negative monitoring data values correspond to erroneous data.

To deal with values which are zero, around zero or negative, the first process in Control #1 sets these values equal to zero. More specifically, any demand value above 250MW for BSPs and primaries, as well as below 2MW at a BSP or 0.1MW at a primary substation, respectively, are first set equal to zero.

Using the blocks identified in Stage 1, we next assess the statistical mean of the demand of each j-th block. Control #1 applies the following inequality to compare the mean of the block to the mean of the whole time series:

$$\mathrm{mean}\left(\mathbf{P}_{\mathbf{block-j}}\right) < (1 - \mathrm{c}_1) \cdot 100 \cdot \mathrm{mean}(\mathbf{P}) \tag{2}$$

where c_1 determines how far the statistical mean of the j-th block needs to be reduced in % compared to the statistical mean of the whole series of half-hourly demand data (**P**). Demand data will be shifted upwards if the inequality is true. Iterative calculations using the whole range of historical half-hourly P demand data from financial year 2012 to 2016 have revealed for over 70 BSPs and 380 primary substations of Electricity North West Ltd that a c_1=50% threshold can adequately identify most blocks with unreasonably zero or negative demand.

If inequality (2) is satisfied, the examined j-th block can be shifted upwards by:

$$\mathrm{Shift}_1 = \left|\mathrm{mean}\left(\mathbf{P}_{\mathbf{block-j}}\right) - \mathrm{mean}(\mathbf{P})\right| \tag{3}$$

After the potential shifting of one or more blocks of demand, the Control #1 module implements a final check to eliminate any potential half-hourly spike. More specifically, any P_i value that exhibits a value that is simultaneously larger than P_{i-1} + SD(**P**) and P_{i+1} + SD(**P**), i.e. more than a standard deviation from the preceding and following values, gets a new value. This value is the average of P_{i-1} and P_{i+1}. It should be noted that the same check is applied as part of the Control #3 module.

Fig. 4a shows the half-hourly monitored demand data of an Electricity North West BSP, corresponding to financial years 2011 to 2015 (5 years). It is evident that this data includes spikes, long periods of zero (or oscillating around zero) demand and potentially demand step changes due to switching operations. The application of Control #1 module leads to the updated time-series data (vector $\mathbf{P}^{(1)}$) of Fig. 4b. Zero demand values and blocks with particularly low mean values have been shifted up in this case.

The derived $\mathbf{P}^{(1)}$ demand data after the implementation of the Control #1 module can be then used as the input to the Control #2 module, as presented in the next subsection.

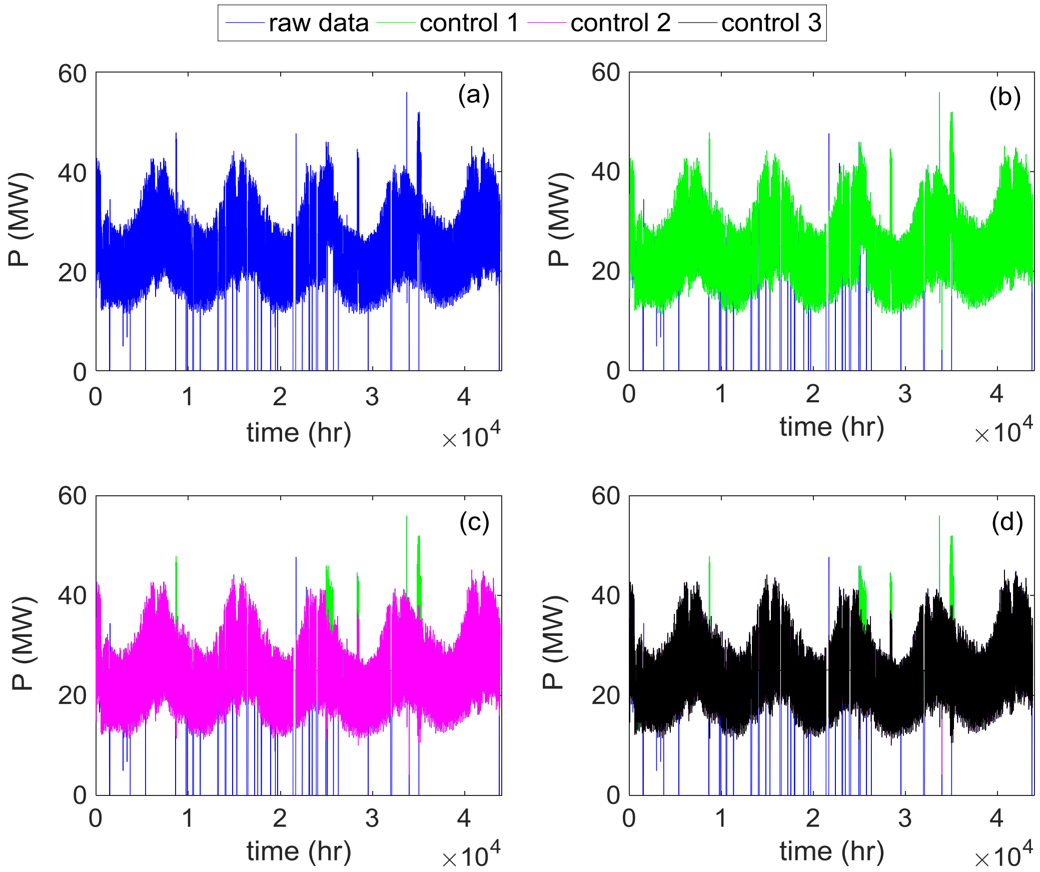

Fig. 4. Application of Stage 2 of the data processing methodology. Results using 5 years raw monitoring half-hourly BSP demand data.

Control #2: Switching operations and network reconfigurations

The shifting of demand blocks in Control #1 was principally related to zero values and does not primarily focus on switching operations or temporary network reconfigurations. Such operations can in some cases lead to significant step changes in the examined series of demand data. Using $\mathbf{P}^{(1)}$ demand data after Control #1, the Control #2 module first checks whether the demand of one or more blocks needs to be shifted upwards or downwards and then shifts it to towards the mean of $\mathbf{P}^{(1)}$.

Similarly to the Control #1 logic, the Control #2 module checks the following inequality for each block j (i.e., vectors $\mathbf{P}_{\mathbf{block\text{-}j}}^{(1)}$):

$$\left|\frac{\mathrm{mean}(\mathbf{P}^{(1)})-\mathrm{mean}\left(\mathbf{P}_{\mathbf{block-j}}^{(1)}\right)}{\mathrm{mean}(\mathbf{P}^{(1)})}\right|\cdot 100 > c_2 \tag{4}$$

where c_2 determines how far the statistical mean of the j-th block needs to be smaller or larger in % compared to the statistical mean of the whole series of half-hourly demand data – **P** – as the criterion for whether the block should be shifted.

In practice, the setting of c_2 needs to be able to identify blocks of demand that correspond to significant step changes (e.g., from temporary network reconfigurations lasting several days or switching operations lasting just a few hours). Iterative tests using monitoring data for BSP and primary substations have shown that a $c_2 = 30\%$ value can indeed correct significant step changes in demand without affecting the normal behaviour of demand. Any step changes in demand that could not be identified using this value can be corrected in Stage 3 using a different rationale that considers the daily characteristics of demand.

If inequality (4) is satisfied, the examined j-th block can be shifted towards the mean value of the whole time series by:

$$\mathrm{Shift}_2 = \mathrm{mean}\left(\mathbf{P}^{(1)}\right) - \mathrm{mean}\left(\mathbf{P}_{\mathbf{block-j}}^{(1)}\right) \tag{5}$$

Fig. 4c shows the updated time-series demand data (vector $\mathbf{P}^{(2)}$) after the application of the Control #2 module. It is evident that $\mathbf{P}^{(2)}$ data does not exhibit the significant spikes or short periods of particularly high demand of the $\mathbf{P}^{(1)}$ data (Control #1). The $\mathbf{P}^{(2)}$ demand data can then be used to apply the Control #3 module, as presented in the next subsection.

Control #3: Curtailment of peaks in shifted demand

The demand shifting of one or more blocks in Controls #1 and #2 can potentially lead to unreasonably high or low (and even negative) demand values *within the shifted blocks*. In order to avoid the introduction of artificial (and not representative of the actual demand) peak and/or minimum demand values in shifted blocks, the Control #3 module is used to curtail any demand value of the shifted blocks that exceeds a pre-defined bandwidth.

The Control #3 module uses two coefficients, i.e. $c_{3\text{-min}}$ and $c_{3\text{-max}}$ that are used to determine this bandwidth around the statistical mean of the whole range of time-series data $\mathbf{P}^{(2)}$. More specifically, if the i-th half-hourly demand value - $P_i^{(2)}$ (output of Control #2) belongs to a block where demand has been previously shifted (i.e., in Controls #1 and/or #2) and one of the inequalities (6a) or (6b) is satisfied:

$$\left|\mathrm{mean}\left(\mathbf{P}^{(2)}\right) - c_{3-\min}\right| \geq \left|P_i^{(2)}\right| \tag{6a}$$

$$\left|\mathrm{mean}\left(\mathbf{P}^{(2)}\right) + c_{3-\max}\right| \leq \left|P_i^{(2)}\right| \tag{6b}$$

then this $P_i^{(2)}$ value is curtailed to obtain the updated $P_i^{(3)}$. The $P_i^{(3)}$ value will be set equal to $\left|\mathrm{mean}\left(\mathbf{P}^{(2)}\right) - c_{3-\min}\right|$ or $\left|\mathrm{mean}\left(\mathbf{P}^{(2)}\right) + c_{3-\max}\right|$ depending on whether inequality (6a) or (6b) is satisfied, respectively.

Attention should be paid to the selection of $c_{3\text{-min}}$ and $c_{3\text{-max}}$ that determine the bandwidth in (6a) and (6b), so as not to result in shorter or longer periods where time-series data exhibit flat demand (i.e., consecutive half-hourly demand data of constant value). The selection of a particularly wide bandwidth could lead to the assessment of seasonal or annual peak and minimum values based on demand data of the shifted blocks. On the other hand, the selection of a particularly narrow bandwidth could lead to long periods of time-series data with flat demand, which would not contain any information regarding the actual demand profile characteristics.

A reasonable bandwidth in Control #3 that could not lead to extreme peak and/or minimum values, as well as periods of flat demand, is to set both $c_{3\text{-min}}$ and $c_{3\text{-max}}$ equal to the standard deviation of $\mathbf{P}^{(2)}$.

The derived $\mathbf{P}^{(3)}$ demand data after the application of the Control #3 module can then be further processed in Stage 3, which is presented in the next subsection.

2.4 Stage 3: Daily Analysis

The application of Stages 1 and 2 can identify and then correct any significant step changes caused by erroneous or missing data, switching operations and temporary network reconfigurations. However, further data processing is required to deal with:

- periods of flat time-series demand data; and,
- the remaining switching operations or other spikes/step changes that could not be identified.

The periods of flat time-series demand data that remain after Stage 2 practically correspond to initial zero demand values (e.g., lack of data) or oscillations around zero demand (e.g., errors in monitoring system, telecommunication system). There can also be limited cases of demand from shifted blocks that has led to particularly high demand values that have been curtailed (see Control #3 module of Stage 2).

Depending on the amount of erroneous and/or missing raw monitoring data, there can be cases where switching operations could not be identified in Stage 1 and corrected in Stage 2. A typical example could be the temporary connection of a primary substation to a different BSP. Such an operation could lead to a simultaneous shifting of both daily peaks and minimum demand levels of the associated BSPs. Thus, an additional process is needed after Stage 2 to correct most of the remaining issues.

Similarly to the previous stages, Stage 3 of the proposed methodology considers and corrects the whole range of half-hourly (or hourly etc) demand data (i.e., $\mathbf{P}^{(3)}$ data). All processes in Stages 1 and 2 focus on the behaviour of demand using half-hourly resolution in analyses, e.g. identification of half-hourly demand step changes, demand curtailments enduring one or more half-hourly time steps etc. By contrast, all processes in Stage 3 involve the analysis of daily demand characteristics, i.e. daily peaks, minima and normalised demand profiles (0 minimum to 1 maximum daily demand).

The analysis of daily characteristics of time-series demand data can be in practice useful to:

- identify cases of switching operations and temporary network reconfigurations that could not be identified as significant step changes in demand in Stage 1 and corrected in Stage 2. More specifically, the analysis in Stage 3 considers smaller but simultaneous changes in daily peak and minimum values of demand;
- the provision of a typical daily demand profile shape to days with flat time-series demand (e.g., periods of zero raw monitoring demand data that have been shifted in Stage 2); and,
- the independent treatment of the demand levels of different week days (e.g., treating Mondays differently from Fridays or Sundays).

The following subsections first explain the way that the periods of flat time-series demand are identified. Next, the analyses carried out to correct daily and minimum demand values, as well as to provide typical demand profile for periods of flat time-series demand are presented. The corrected normalised demand profiles, daily peaks and daily minima demand values can then be used in a reassembly process to get the final processed time-series demand data, as described in the end of this section.

Identification of periods of flat demand

To identify the days of flat demand, the following values need to be first assessed for *each day* the of the $\mathbf{P}^{(3)}$ demand data:

- daily peak demand;
- daily minimum demand; and,
- normalised daily demand profile, i.e. time-series profiles between 0 for daily minimum and 1 for daily maximum.
- standard deviation of the normalised daily profiles.

The flat demand time-series data – $\mathbf{P}_{\mathbf{flat}}$ – consists of the days that exhibit:

- standard deviation of the normalised daily demand profiles below 0.1 (i.e., days that have limited non-zero half-hourly demand data); *or*,
- daily peak demand > +1% of the corresponding minimum demand (i.e., days with practically same demand throughout the day due to erroneous or missing monitoring data).

Correction of normalised demand profiles

Having identified the flat demand periods (i.e., groups of days), a representative normalised profile needs to be assigned to these days. Although more sophisticated approaches could be that this assignment considers seasonal and/or yearly characteristics, a simple approach can be taken into account to keep the methodology generic.

The chosen simple approach is to first identify the day that exhibits the median standard deviation of its normalised profile amongst all examined days with non-flat demand. The normalised profile of this day can then be assigned to all days of flat demand.

It should be highlighted that the normalised demand profiles of the periods of non-flat demand are not further processed at this stage in order to be as representative as possible of the actual demand characteristics.

Correction of daily peak and minimum demand

In order to correct the daily peak and minimum demand values, three steps need to be followed:

- **Step A**: correction of daily peak and minimum demand for the periods of identified flat time-series demand;

- **Step B**: correction of daily peak and minimum demand for switching operations; and,
- **Step C**: smoothing daily peak and minimum demand.

All the above mentioned steps involve daily analyses individually for each day of the week.

Step A
For periods of flat demand, the daily peak and minimum demand values are set equal to the average respective peak and minimum values of the day of the week before and after the examined period.

This could be better described with an example. Assuming there is flat daily demand data for a BSP substation on Thursday 12/04/2016, the corresponding daily peaks and minimum demand values of this day will be set equal to the average of the Thursday 05/04/2016 and Thursday 19/04/2016 values.

Step B
Given that *significant* step changes in demand caused by switching operations and/or temporary network reconfigurations have previously been identified and corrected in Stages 1 and 2 (zero values and any large step changes), a different approach needs to be taken in Stage 3 to identify the remaining switching operations that have not been identified and corrected.

Such a case could be the connection of a primary substation to a different BSP, which could lead for example to a circa 20 % change in daily peak demand and 10 % change in daily minimum demand of the associated BSPs. This step change could not be identified and corrected in Stages 1 and 2 (e.g., a c_2=0.3 in Control #2 of Stage 2 would identify step changes of over ±30%).

Reasonable thresholds that have successfully identified such cases of switching operations using historical half-hourly demand data of BSPs and primary substations of Electricity North West Ltd are the simultaneous:

- ±10% change in daily peak demand; and,
- ±5% change in daily minimum demand.

Step C
Although all processes up to now are expected to have successfully dealt with all possible monitoring data issues, there can still be unwanted deviations or any other remaining deficiencies in monitoring time-series data, requiring the smoothing of daily peak and minimum demand values by a moving average assessment.

The combined effect of Steps A to C in Stage 3 are best shown by the example in Figures 5 and 6 overleaf, using 5 years (from 2011 to 2015) half-hourly monitoring P data of a BSP substation of Electricity North West Ltd. It should be noted that for the examined substation, there were particular issues regarding the available 2012 monitoring data. These issues with data availability have now been resolved, but the initial development of the processing methodology used this data set with missing values, which can be an interesting example to check and demonstrate the performance of the developed processing methodology.

After the application of Steps A and B in Stage 3, Fig. 5 shows the daily peak and minimum demand values for the same day of the week (i.e., day G corresponds to Saturdays using original monitoring data (end of Stage 2) and the corresponding updated values (after Step B of Stage 3).

Focusing on year 2012 (week 53 to week 104), there are periods of flat demand, i.e. consecutive days with the same peak and minimum value (old values in Fig. 5). Using step A of Stage 3 it is evident that the corresponding updated peak and minimum demand values are the average of the time-series data before and after these flat demand blocks. The application of step B on the same case (after step A) has led to the identification and correction of simultaneous >10% and >5% step changes in peak and minimum daily demand, respectively, from week 246 to week 247.

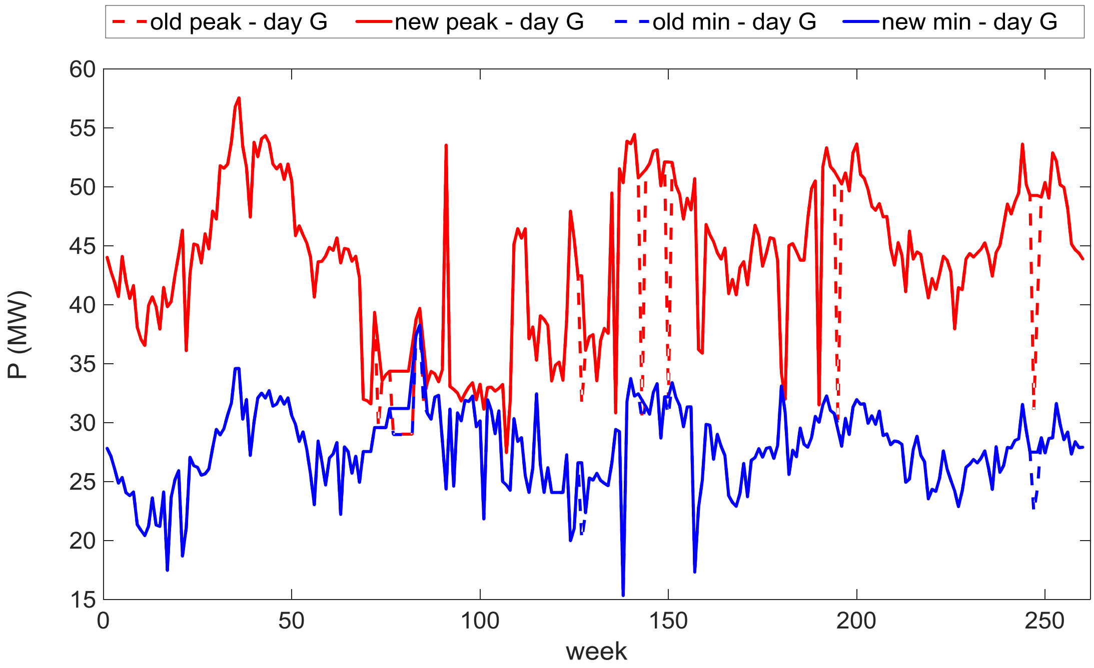

Fig. 5. Daily peak and minimum values of an Electricity North West's BSP before and after the application of steps A and B of Stage 3 of the proposed methodology.

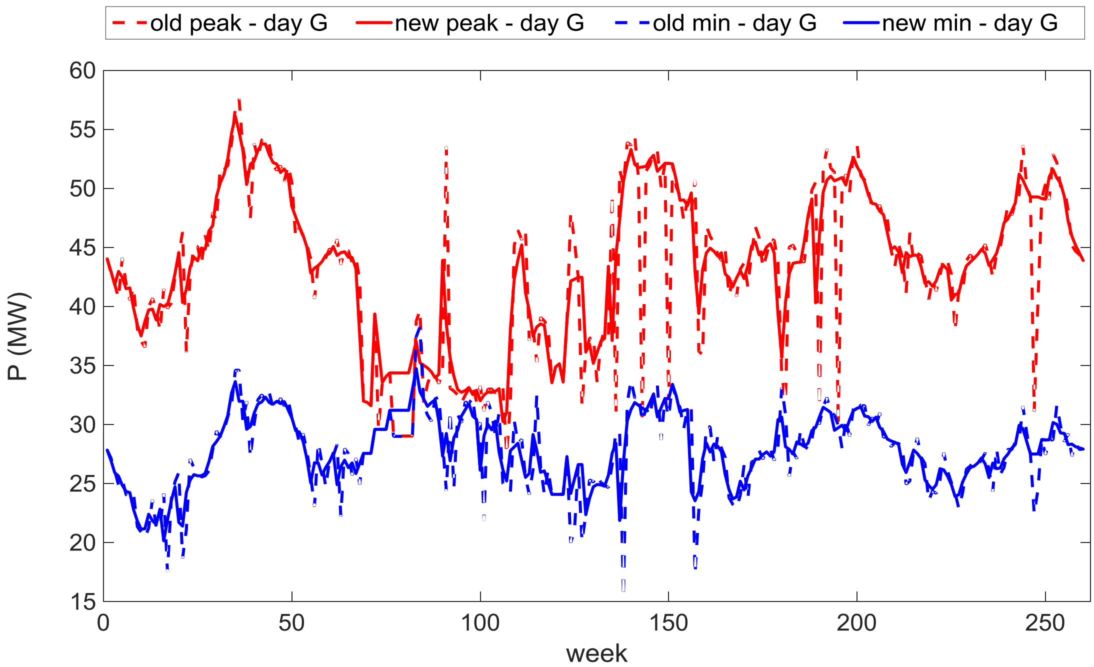

Fig. 6. Daily peak and minimum values of an Electricity North West's BSP before and after the application of steps A to C of Stage 3 of the proposed methodology.

Nonetheless, there are still remaining issues in daily peak and minimum demand values, such as simultaneous sags in daily peak and minimum demand (e.g., around week 140), wider than usual daily peak to minimum range (e.g., around week 90) and narrower than usual daily peak and minimum range (e.g., almost equal peak and minimum values around week 180). The smoothing process of step C can further correct these issues.

Fig. 6 shows the updated daily peak and minimum demand values after the application of step C for the same case with Fig. 5. It is evident that the updated (new) daily peak and minimum demand values after step C do not exhibit at all or at the same extent after step B the above mentioned remaining issues.

The corrected daily peak and minimum demand values after the application of steps A to C can then be used together with the corrected normalised demand profiles (process presented in next subsection) to reassemble the final processed time-series demand data.

Reassembly of time-series demand data

The corrected daily peaks, minima and normalised demand profile data from the daily analyses of Stage 3 of the methodology are finally used to reassemble the time-series demand data. More specifically, the daily normalised demand profiles are multiplied by the daily peak-to-minimum demand values to get the processed demand data.

A case of processing time-series data of flat demand periods is demonstrated in Fig. 7. The missing monitoring P demand data of a BSP for a period of 150 hr (i.e., zero demand values of raw data from time=16,340hr to 16,490hr in Fig. 7) are first uplifted in Stage 2 from zero to 25MW constant demand throughout the whole period. Next, in Stage 3 the data are further corrected to obtain demand profiles that are aligned with the demand behaviour before and after this period of missing monitoring data.

Other advantages of the use of Stage 3 to further correct time-series demand data include the correction of data anomalies caused in Stage 2 and the smoothing of demand step changes and/or spikes that have not been corrected in Stage 2. Fig. 8 shows how these corrections take place considering the processing of time-series P demand data of a BSP.

The data processing of Stage 3 results in smoother demand profiles that do not include periods of flat demand. This can be seen in Fig. 7 and Fig. 8, but also in Fig. 9 for a 5-year period of time-series P demand data of a BSP at Electricity North West. Although most data issues have been properly corrected in Stage 2, any spiky demand levels are further corrected in Stage 3. The smooth demand profiles of the output of Stage 3 maintain the demand characteristics and trends of the examined substation and can be further used for post-processing analyses (e.g., identification of historical demand trends, inputs/loads in time-series power flow simulations).

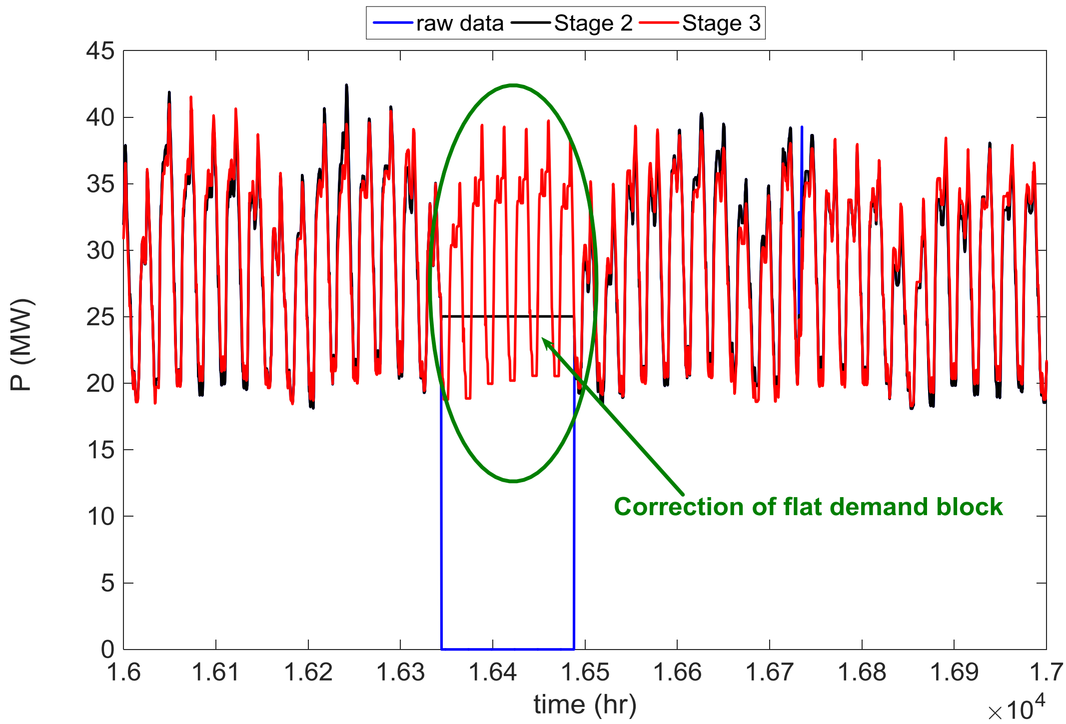

Fig. 7. Correction of flat demand block using half-hourly P demand data of a BSP substation of Electricity North West Ltd.

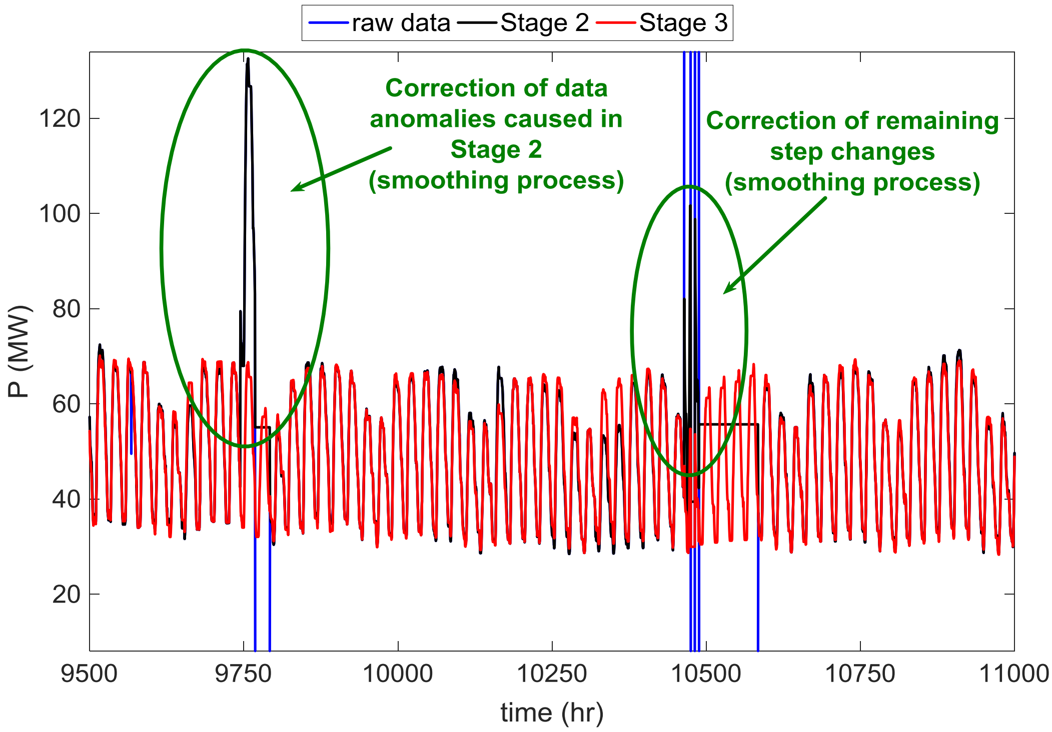

Fig. 8. **Smoothing demand data using half-hourly P demand data at a BSP substation of Electricity North West Ltd.**

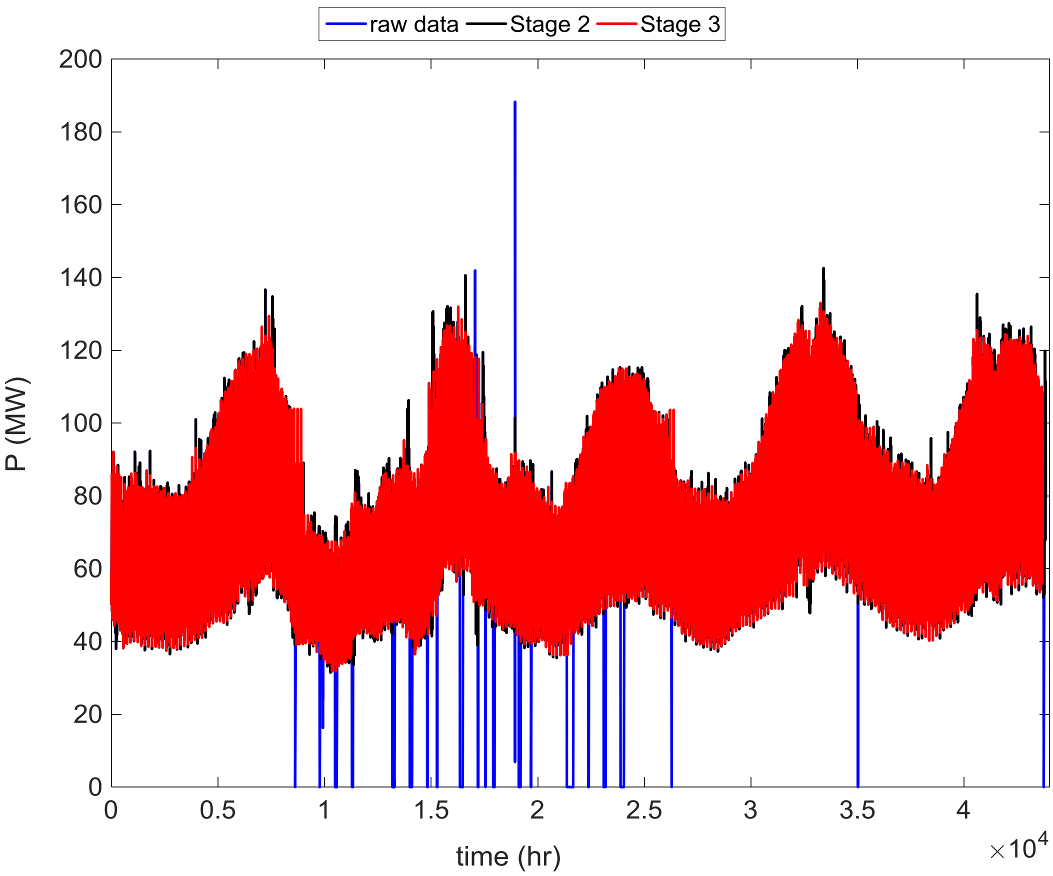

Fig. 9. **Processed half-hourly P demand data (years 2012 to 2016) of a BSP substation of Electricity North West Ltd.**

2.5 Stage 4: Reiteration of Half-hourly Analysis

Although for the majority of substations the processed data after Stage 3 can be considered as a sufficient representation of the actual time-series variation of true P demand, there can be a limited number of cases where additional processing is required. Such a case can be seen in Fig. 10, where any step change that could not be identified and corrected in Stage 2 (i.e., time t=16,600 to 16,800 hr) resulted in unreasonable demand amplifications in Stage 3.

Stage 4 is in practice a re-iteration of Stages 1 and 2 using as input the processed data of Stage 3. Nonetheless, given that this input does not anymore exhibit the same data issues with the raw monitoring data, the coefficients of the control modules (Stage 2) need to be set properly. More specifically, these settings can be as follows:

- Control #1: this module can be neglected at this stage;
- Control #2: c_2 control coefficients similar to Control #2 of Stage 2 can be used; and,
- Control #3: given that Stage 3 output data are expected to exhibit lower deviations, $c_{3\text{-min}}$ and $c_{3\text{-min}}$ should be set at higher values than in Stage 2.

In order to allow a better insight into the way that the different processes of Stages 2 to 4 modify the raw monitoring data, Appendix A presents the corresponding results following the proposed methodology using 5 years of half-hourly P demand data of 10 BSPs and 10 primary substations at Electricity North West Ltd.

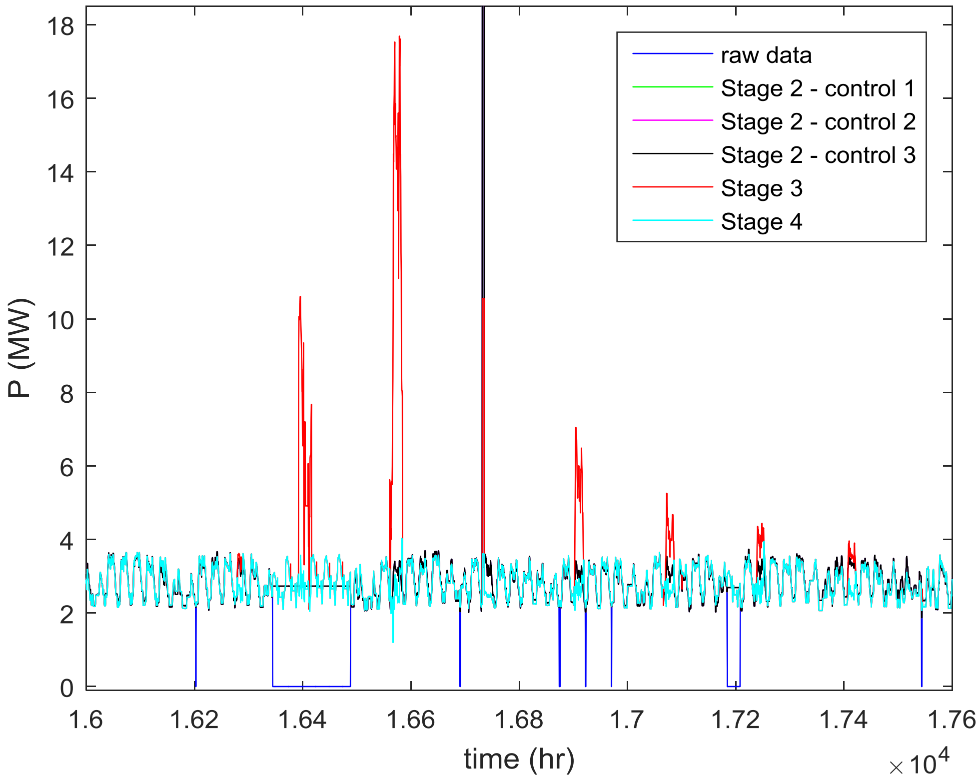

Fig. 10. Raw and processed data from using all stages of the proposed data processing methodology.

2.6 Recommendations for Methodology Settings

As highlighted in the previous subsections, the developed data processing methodology is generic and allows the fully automated processing of time-series P demand data. Nonetheless, proper settings are needed at the different stages (i.e., Stages 1 to 4) when processing data at BSPs and primary substations.

This subsection summarizes the recommendations on these settings that have been presented in the previous subsections 2.1 to 2.5 based on the application of the proposed methodology on half-hourly monitoring data of the P demand of over 70 BSP and 380 primary substations belonging to Electricity North West Ltd.

<u>Stage 1</u>

- Identification of step changes: dP_i/dt values that exceed a bandwidth of ±80% of the standard deviation of the examined time-series P data (i.e., SD(**P**)) are identified as potential step changes in demand;

Stage 2:

- *Control #1:* Demand values above 250MW and below 2 and 0.1MW for BSPs and primary substations, respectively, are first treated as irreversibly erroneous and set equal to zero. Next the control logic leading to demand shifting can be applied using a c_1=50% (see subsection 0) coefficient for both BSPs and primaries.

- *Control #2:* Adequate results have been derived using c_2=30 and 40% (see subsection 2.3) for BSPs and primary substations, respectively. Nonetheless, an additional setting of c_2=150% is worth being used for particularly small primary substations with mean demand value below 3MW.

- *Control #3:* Adequate results have been obtained for both BSP and primary substations considering that every demand value of shifted blocks in Controls #1 and #2 must lie within a bandwidth defined by the ± of the standard deviation value of the time-series data (**P** - after Control #2 / Stage 2 around the corresponding mean value of **P**, i.e. using $c_{3\text{-min}}$= $c_{3\text{-max}}$=SD(**P**) in equations (6a) and (6b) (see subsection 2.3).

Stage 3

- Switching operations and temporary network reconfigurations are identified for both BSP and primary substations considering a simultaneous ±10% and ±5% change in daily peak and minimum demand, respectively.

Stage 4:

- *Control #1:* This control module is not used in Stage 4.
- *Control #2:* The same settings as in *Control #2* of Stage 2 are used.
- *Control #3:* The logic is the same as for the corresponding control module in Stage 2, but a larger bandwidth using $c_{3\text{-min}}$= $c_{3\text{-max}}$=3·SD(**P**) (where SD(**P**) the standard deviation of the time-series data after Control #2 / Stage 4).

2.7 Summary

This section has presented the overview and further details of the proposed data processing methodology in the ATLAS project. The key points of the proposed approach are:

- it is tailored to time-series P demand data, but can be extended to other types of demand data (e.g., Q or S demand data);
- it has a modular structure of 4 stages running in series one after the other;
- it involves half-hourly and daily analyses to deal with missing data, spikes in demand and step changes derived from any potential source (e.g., switching operations, telecommunication problems);
- it is generic and fully automated requiring minimum number of changes when processing BSP and primary demand data; and,
- it has been successfully tested using 5 years of half-hourly P demand data of over 70 BSPs and 380 primary substations belonging to Electricity North West Ltd.

3 ESTIMATION OF NON-MONITORED GENERATION FOR THE ASSESSMENT OF TRUE DEMAND

This section discusses the challenges for DNOs to make proper estimates of the latent demand of non-monitored distributed generation (DG). The non-monitored DG refers to generation with no export metering data provided to the DNO – DNOs do not routinely receive any data on the output of a generator.

As shown in Fig. 11, the underlying true demand can be in practice assessed as the sum of the following three components:

- Monitored Component of True Demand: as shown in Fig. 11, this is the sum of the measured demand and the MW exports of all distributed generation (DG) sites with available monitoring exports. This data has been obtained Electricity North West's Demand and Generation Dashboard, and processed with the methodology described in the preceding sections.
- Non-monitored DG: this corresponds to the estimated half-hourly latent demand (generation) of all DG sites without monitoring of export. The estimate is made by combining the capacity, DG type and a generation profile for the appropriate DG type. We do not have any data on the output of these generators, so we estimate based on a profile.
- Effects of DG with export metering on reducing customer demand: given that many DG units are installed on sites with underlying demand (e.g., factory with installed CHP unit), in many cases the actual generation profiles cannot be derived using the available monitoring export and import data. This component will estimate the effects of these generators on masking the underlying true customer demand. The methodology to estimate this effect is planned to be developed at a later stage in ATLAS project.

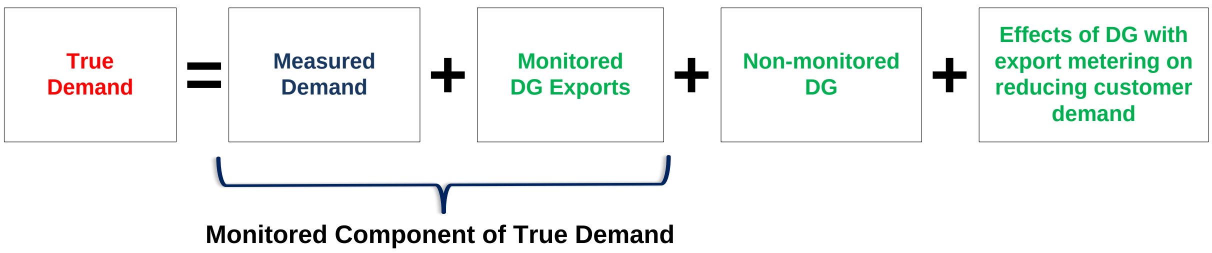

Fig. 11. Assessment of true demand using available monitoring data and estimates of non-monitored generation

Therefore, an estimate of the non-monitored DG is necessary to achieve a more accurate assessment of the true demand.

3.1 Practical Challenges for DNOs and National Grid

Increased penetrations of small and medium sized DG units (including domestic and non-domestic photovoltaics on buildings) lead to suppression of the measured demand at substations. Given the limited visibility of small and medium sized DG, the DNOs are not in many cases in position to tell if fluctuations in monitored GSP, BSP and/or primary demand are due to changes in demand or generation. This uncertainty affects both operational and network planning processes, including identifying the hosting capacity of Low Carbon Technologies (LCTs) and scheduling of network planning interventions, such as traditional network reinforcements and/or Demand Side Response (DSR) services.

The deployment of Active Network Management (ANM) systems, such as the Network Management System of Electricity North West, is expected to enhance the capabilities of DNOs to control DG units. This will necessitate the estimation of non-monitored generation to assess the true underlying demand. The latter will be also a main challenge for National Grid [5], given that the transmission operator will not be in position to tell if fluctuations in GSP demand are due to changes in demand or ANM activities.

For the above mentioned reasons, estimating the impact of small and medium sized non-monitored DG units is relevant to network management in the transition from distribution network to distribution system operators.

The following subsection discusses some practical approaches to estimate latent demand of non-monitored DG units and associated practical insights per DG type. There are limitations on the basic data around the capacity of installed non-monitored generation, by location and installation date; based on that estimates can be made from profile data of the output associated with those generators.

3.2 Estimation of Non-monitored Distributed Generation

There is a variety of different types of DG units ranging from domestic photovoltaics (PV) to large combined heat and power (CHP) plants in oil and chemical industries. Nonetheless, it should be noted that the vast majority of non-monitored DG units mainly consists of small and medium sized renewable energy sources (i.e., PV, onshore wind generation and CHP).

Photovoltaics

Photovoltaics (PV) are among the fastest growing renewable energy sources worldwide. The UK Government's Feed-in Tariff scheme has incentivized PV generation leading by the end of 2015 to over 8GW of aggregated installed PV capacity [6]. Currently, most of the total PV capacity corresponds to small- and medium-scale PV plants (≤1MW).

Electricity North West is aware that the amount of PV notified directly to it by installers (as required for example under the terms of Engineering Recommendation G83) is inconsistent and in some local authorities can be significantly lower than the amount suggested by Ofgem's public feed-in-tariff reports on its website. We are taking steps to obtain the detailed data set from Ofgem, but in the meantime the analysis in this report reflects only our own records of generation connected in parallel to our network, which is likely to understate the total effect of non-monitored PV on suppressing network demand.

Although a significant amount of PV is not monitored by DNOs, databases with detailed irradiance data per region can be used to estimate the corresponding generation profiles. Fig. 12 shows the PV profiles for all days of an average year for the Electricity North West region. These are half-hourly profiles that have been generated using minute by minute irradiance data derived from a tool developed by the University of Loughborough [7]. It should be noted that this is a relatively simple input, and does not yet reflect the historic variations in weather and irradiance over past years.

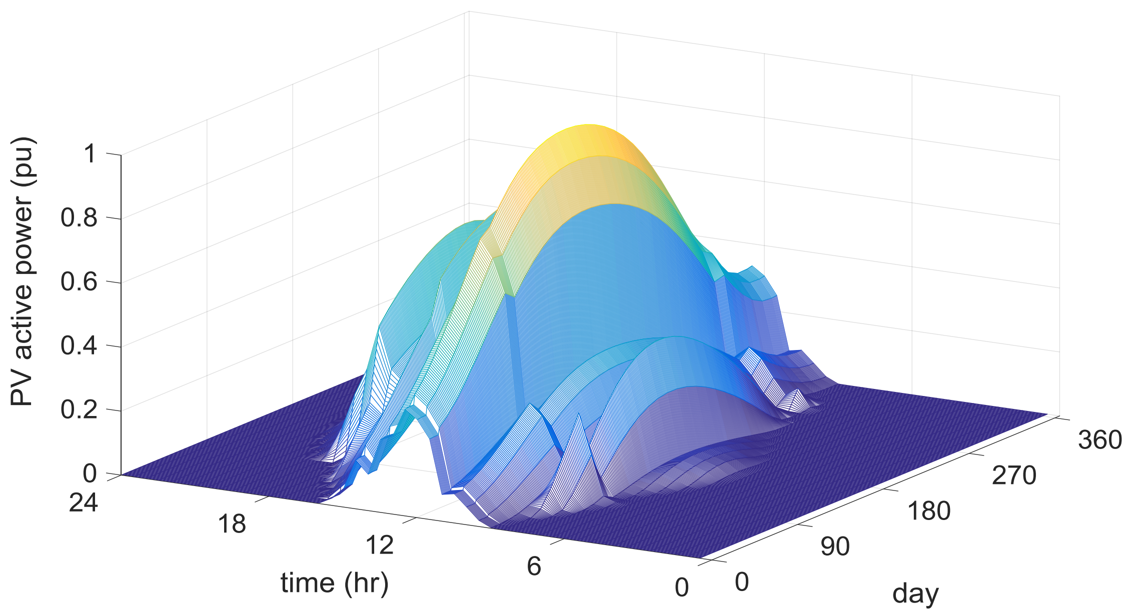

Fig. 12. Average half-hourly PV profiles of all days of the year of the Electricity North West's region.

Wind Generation

The estimation of time-series generation profiles of wind farms is a far more challenging task, due to the stochastic nature of wind and the effects of landscape etc. Nonetheless, monitoring data for the exported generation of most medium and large sized wind farms are available to DNOs.

This data can be used to assess normalised half-hourly generation profiles. These profiles can then be combined with the installed capacity of small wind farms in the same area for which export monitoring data are not available.

Combined Heat and Power

There is a large variety of Combined Heat and Power (CHP) units (i.e., used for different economic sectors, running on different fuels), some of which have monitored export output (i.e. latent P and Q demand). Nonetheless, in the case that such a DG unit is not exporting to network, there is a need to not under- or overestimate its contribution to suppressing customer demand.

Available load factor (LF) and installed capacity data can be used to produce estimated generation profile of CHPs without available monitoring data. According to the Digest of UK Energy Statistics (DUKES) 2016 [8], there has been a 12 % fall in the LF of CHPs across UK from 2011 to 2015, mainly driven by the chemicals and refineries sectors. In 2015 the average LF value of CHPs was 51.3% with CHPs operating in average for 9 hours per day.

Regarding micro CHPs, heat load data could be also used in future to have better estimates of the generation profiles [9]-[10]. Nonetheless, it should be noted that in the initial estimation of non-monitored latent demand for the ATLAS project, the DUKES 2016 load factor data for CHPs have been also used for micro CHPs.

As a next step in the ATLAS project, the available monitoring data of CHPs will be used to estimate the effects of non-monitored CHPs following a similar approach to the one described for wind generators. Increased accuracy in the estimation of generation profiles of CHPs could be achieved in future using any further available data regarding the associated industrial/ commercial sector, the on-site underlying true demand, as well as any other financial incentives and tariffs.

3.3 Estimation of Aggregated Non-monitored Primary Latent Demand

Using the profile assumptions described in subsection 3.2 for PV, wind generation, CHP and other DG types, the individual and aggregated half-hourly latent demand (i.e., generation) of primary substations can be assessed considering:

- the installed capacity per DG type per primary substation;
- whether a DG unit is operational or not; and,
- the date of connection of each DG unit.

Fig. 13 shows an estimation of the aggregated latent P demand of all primary substations in the Electricity North West’s region. The estimated profile of the aggregated latent demand uses half-hourly resolution and considers the historical data of connections of over 30,000 DG units in the Electricity North West’s region within the last 5 financial years (i.e., time t = 0 corresponds to 1st April 2011).

The PV and wind generation profiles have been obtained following the assumptions described in subsections 3.2.1 and 3.2.2, respectively. For the CHPs and other generators fuelled by natural gas, biogas, waste, biomass, hydro and biodiesel, the corresponding load factors defined in DUKES [8] have been considered. More specifically, all DG units apart from PV and wind generators have been considered to be operating at a constant output defined by their capacity and the associated DUKES 2016 load factor data.

The peak values per year in Fig. 13 (i.e., 5 consecutive “hills” with increasing values) corresponding to the maximum annual generation of non-monitored sites are seeing an increase mainly driven by the fast installation of PV units from 2012 to 2016. The corresponding minimum generation values are not affected by PV units (i.e., zero PV output overnight), but mostly by the increasing installation of CHPs within the same period.

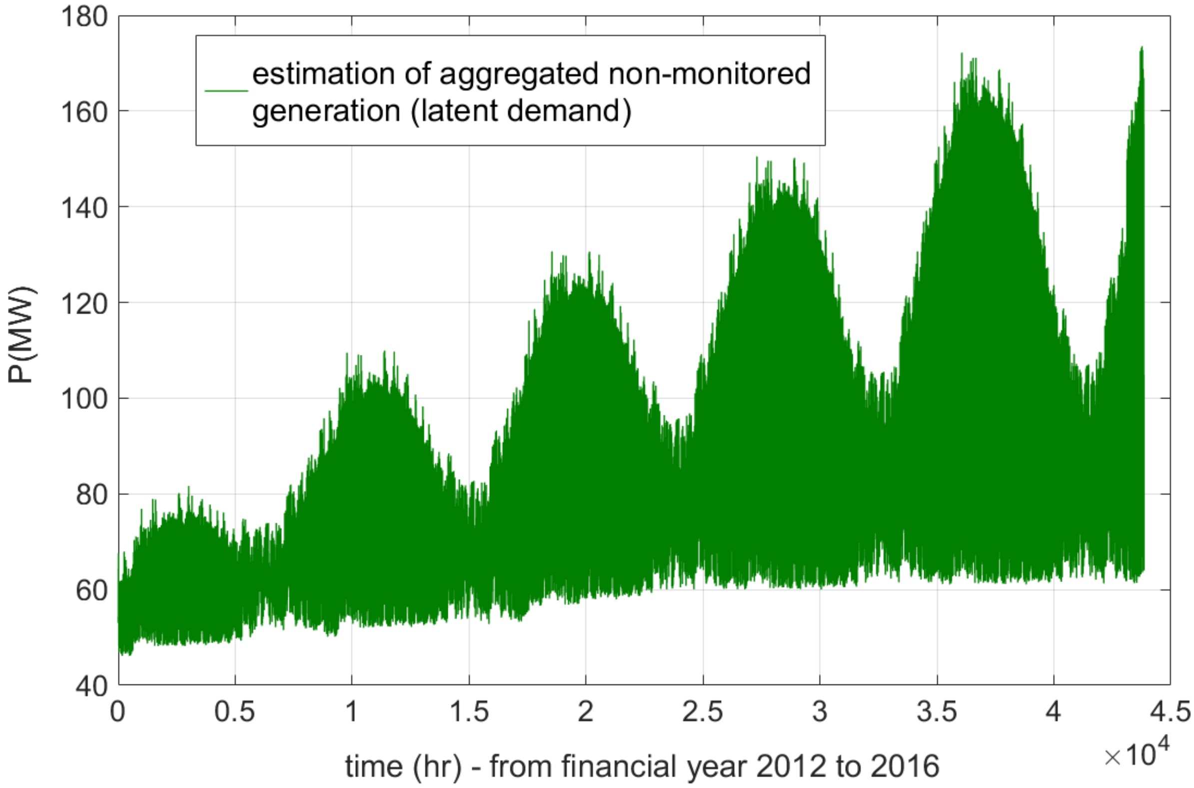

Fig. 13. Aggregated non-monitored half-hourly primary latent P demand of the Electricity North West's region.

It should be noted that the obtained generation profile in Fig. 13 is expected to over- or underestimate the output of non-monitored DG units connected downstream primary substations across Electricity North West's region. This is mainly due to the following three reasons:

- For PV and wind generation that correspond to the most dominant DG types in terms of installed capacity, the installed capacity data have been first combined with monitoring data from large PV and wind farms to estimate a ratio of actual output to installed capacity. For example, this assessment has given a 77% peak PV output in average across Electricity North West's region (e.g., a 100kW PV is expected to exhibit a maximum annual output of 77kW). This assessment is expected to over- or underestimate the actual output of DG units due to uncertainties on installed capacity data (i.e., overestimated values); and,
- This assessment has been based on manually obtained DG data of over 30,000 DG units, based on DG applications and connections notified to the DNO and recorded on its systems. It is expected that a less or more significant number of DG units is actually operational, but has not been considered in this assessment. This fact highlights the necessity for more coordinated actions between DNOs and Ofgem (e.g., comparison with FITS data) so as to estimate the effects of all DG units on monitored demand; and,
- The profile data that have been used for the different types of DG units are expected to introduce lower or higher errors to the estimated overall generation. For example, CHP units operating at different industrial or non-industrial customers are expected to exhibit significant differences regarding their half-hourly generation.

3.4 Assessment of True Demand Using the Non-monitored DG Estimation

Having estimated the non-monitored DG (i.e., latent demand), a more accurate assessment of the underlying true demand can be achieved. Fig. 14 and 15 show the daily profiles of aggregated true P demand of primary substations in Electricity North West's region for a winter and a summer day, respectively.

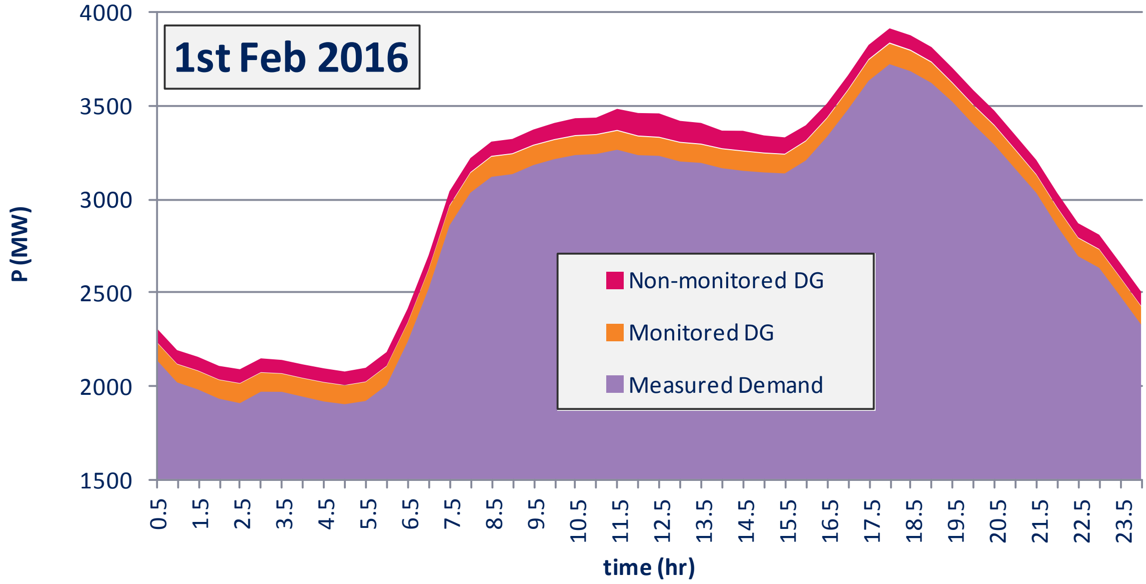

Fig. 14. Winter day profile of the aggregated demand of all primary substations in the Electricity North West's region.

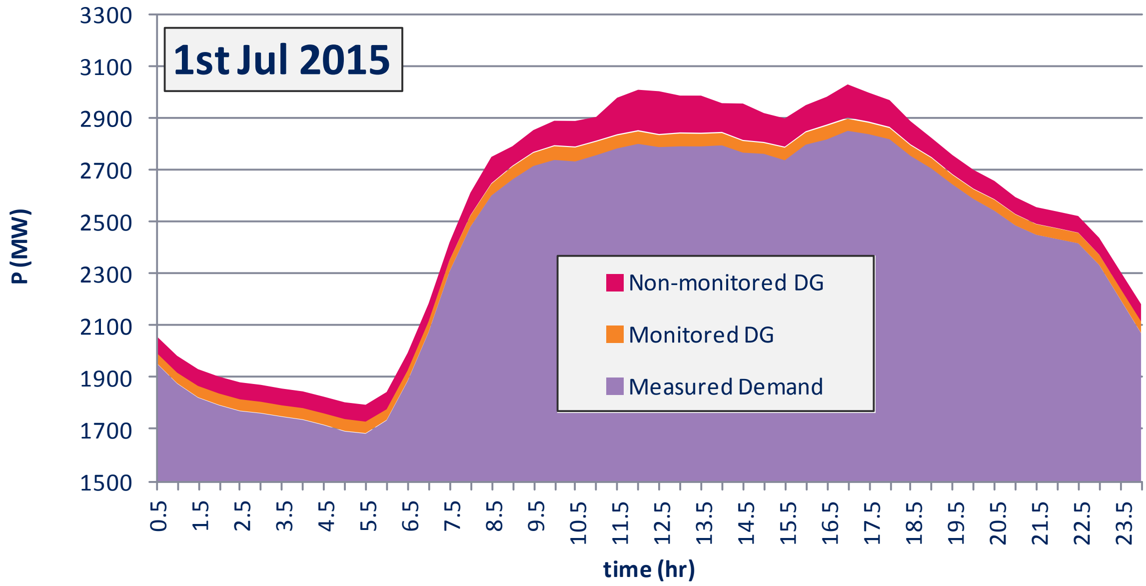

Fig. 15. Summer day profile of the aggregated demand of all primary substations in the Electricity North West's region.

It can be seen in both figures that apart from the monitored DG, the non-monitored DG units can also suppress significantly the aggregated primary demand. Comparing summer and winter, PV suppresses measured demand more significantly in summer due to longer day time and higher sun irradiance values.

It should be highlighted that more accurate assessment of the underlying true demand can be achieved by:

- using more accurate data regarding the installed capacity and number of DG units per primary substation;
- using more informed assumptions (possibly per DG site) for the profiles of CHP units; and,
- having available data or estimations on the effects of DG units that are installed on sites with underlying demand (e.g., factory with installed CHP unit).

Future work should further consider these factors to improve the accuracy of estimated half-hourly non-monitored generation profiles.

3.5 Summary

This section highlighted the challenge for both transmission and distribution operators to assess non-monitored latent demand from DG units. This process is necessary to have better estimates of the true demand of substations. Practical insights regarding the estimation of non-monitored latent demand of PV, wind power plants and CHPs have been also presented in this section.

The suggested assumptions are used in an initial estimation of the half-hourly generation profiles of over 30,000 non-monitored DG units installed in Electricity North West's region, taking into account the installed DG capacity per DG type per primary substation and the date of connection. The half-hourly aggregated latent demand of primary substations is presented from financial year 2012 to 2016.

Finally, half-hourly processed demand data corresponding to the monitored component of true demand across all primary substations belonging to Electricity North West are used together with the obtained non-monitored DG profiles to produce a more informed estimate of the underlying true demand. The obtained profiles highlight the significant effects of PV on continuously suppressing the monitored demand of primary substations from 2012 to 2016.

4 PRACTICAL INSIGHTS AND RECOMMENDATIONS

This section presents practical insights from the application of the proposed methodology to process time-series monitoring data of P demand at Electricity North West Ltd, as well as recommendations for applying the methodology to Q demand data and future work.

4.1 Implementation Using a Prototype Tool

As part of the ATLAS project, a prototype tool has been developed in Matlab [11] software using the proposed data processing methodology. This tool has been tested using 5 years of half-hourly P monitoring data of over 70 BSPs and 380 primary substations belonging to Electricity North West Ltd. This prototype also informs a future business-as-usual approach which will be defined by the end of 2017.

The application of the proposed data processing methodology to these substations has shown that:

- the approach is generic and it can be used even in the absence of network connectivity data (e.g., primary substations per BSP);
- the methodology uses a modular structure allowing the user of the corresponding tool to apply or not particular processes (e.g., smoothing process in Stage 3); and,
- the computational time for the processing of a year's half-hourly demand data per substation can be less than 3.5 s using Matlab software on a personal computer (pc) with 2.8GHz CPU running on a 32-bit operating system (e.g., the data processing of 5 years half-hourly data for the full range of 70 BSPs requires less than 25 min). This reflects the standard available pc available to planning engineers, but the computational time could be significantly reduced using a 64-bit operating system (i.e., formulating matrices of data instead of vectors).

In the wider context of using the proposed data processing methodology, the importance of visual sense checks of the processed demand data should be highlighted. In practice, the comparison between the monitoring and processed half-hourly data can be in many cases the most proper way to validate the successful use of the proposed methodology.

Nonetheless, a semi-automatic approach for the practical implementation of the proposed data processing methodology for the GSP, BSP and primary substations of Electricity North West Ltd to process a recent year of demand data is shown in Fig. 16. The inputs of this implementation flowchart are:

- the true demand data obtained from the available original V, I, P and Q measurements of substations and DG units (i.e., Demand and Generation Dashboard data); and,
- the validated peak P demand of the previous year for GSPs, BSPs and primary substations (i.e., value that may have been obtained using further manually-identified data relating to HV feeder loading, and DG connectivity and loading of major customers)

The following steps should be followed per GSP, BSP and primary substation using the above inputs to obtain the final half-hourly profiles and peak demand values that can be used for further studies **for a new year of data**:

- Step #1: the proposed ATLAS data processing methodology is applied using the available Dashboard half-hourly true P demand data. The processed data are then used to identify the peak P demand values. In practice, two general cases are expected after Stage 2 of the proposed methodology:
    1. i. the available monitoring data have an accurate yearly peak value and the smoothing process (Stage 3) slightly curtails this value; and,
    2. ii. the available monitoring data have a yearly peak value that does not correspond to the actual demand of the substation. In this case, Stage 3 is expected to provide a more accurate peak value.

  In order to keep a generic approach, both peak values from Stages 2 and 3 are obtained and further used in the following step (Step #2).
- Step #2: if the identified peak demand value (i.e., either from Stage 2 or 3) exhibits a <5% difference from the last year's validated peak value, then the identified peak value and demand profiles using the proposed methodology can be used in further studies as they are. Otherwise, the implementation continues with Step #3.

- <u>Step #3:</u> the peak demand value of the current year is obtained using more data (i.e., not only Dashboard data). More specifically, monitoring data of HV feeders, connectivity of DG units etc are used to provide a validated peak demand value of the current year. Additionally, any missing Dashboard data are identified (e.g., in cases of asset replacements and/or new substations) and replaced in the Dashboard. A visual sense check of the processed demand data may also trigger requirements for additional and/or recalibrated data.

- <u>Step #4:</u> if the validated peak demand value of Step #3 exhibits a <20% difference from the corresponding last year's value, then the processed half-hourly data are automatically scaled taking into account the validated peak demand value. Otherwise, the half-hourly demand profile of the examined substation is obtained from the manual allocation of demand data from adjacent substations.

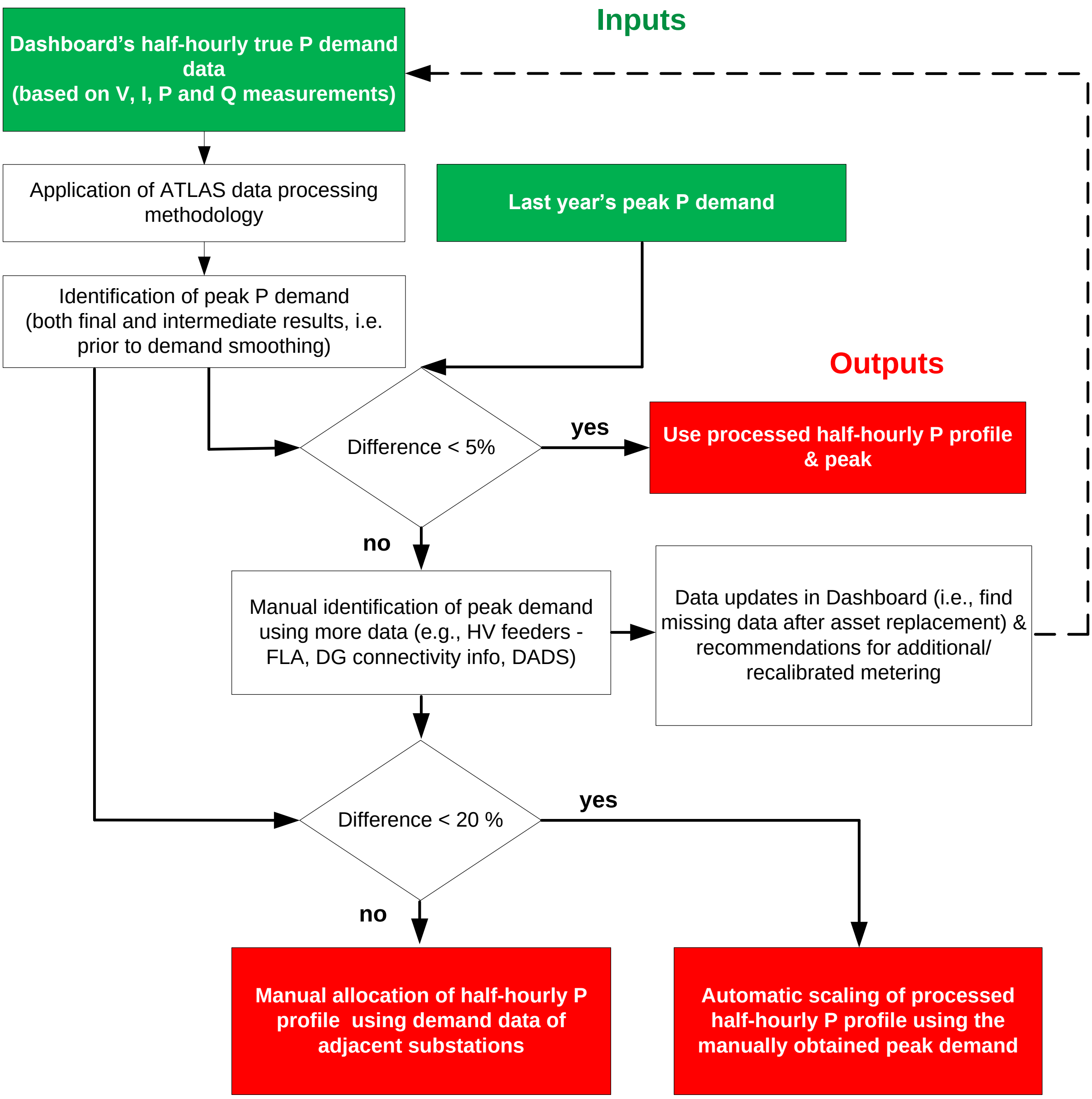

Fig. 16. The practical implementation of the proposed data processing methodology.

The use of different thresholds – i.e. 5% and 20% – is aligned with the following rationales:

- a ±5% difference (Step 2) between the peak demand values of consecutive years can be in practice considered as an extreme change, given this could lead to a change in 'load index' category;
- a ±20% difference (Step 2) practically means that the half-hourly profile of the processed data cannot be representative of the actual demand of the substation under normal running conditions.

Regarding this practical implementation of the proposed data processing methodology, it should be highlighted that:

- a major objective is to have a profile (i.e., half-hourly P demand data) that the DNO's strategic planning engineers would validate;
- if initial data (i.e., the Demand and Generation Dashboard data of Electricity North West) are complete, then the automatic ATLAS data processing methodology is expected to produce adequate demand profiles – in practice there will always be some cases where data is not available in the Dashboard system;
- the business as usual implementation is under review and could possibly use a different software platform (e.g., Python); and,
- for P demand data, one major objective is to have a profile (i.e., half-hourly data) that the DNO engineers would validate. If initial data (i.e., the Demand and Generation Dashboard data for Electricity North West) are complete, then the methodology can produce representative profiles of the actual demand of substations.

4.2 Processing Reactive Power Data

The proposed data processing methodology has been tailored for the processing of P demand (true demand) data. Nonetheless, the application of the methodology to process Q demand data is possible with proper modification, and this is now being addressed in the ATLAS project.

Unlike true P demand, Q demand can exhibit either positive or negative values, as shown in Fig. 17. Negative Q demand at GSPs and/or BSPs could be practically found during periods of minimum load, when significant Q gains are produced from 132 and/or 33kV lines (particularly cables), which are loaded below their natural loading [11].

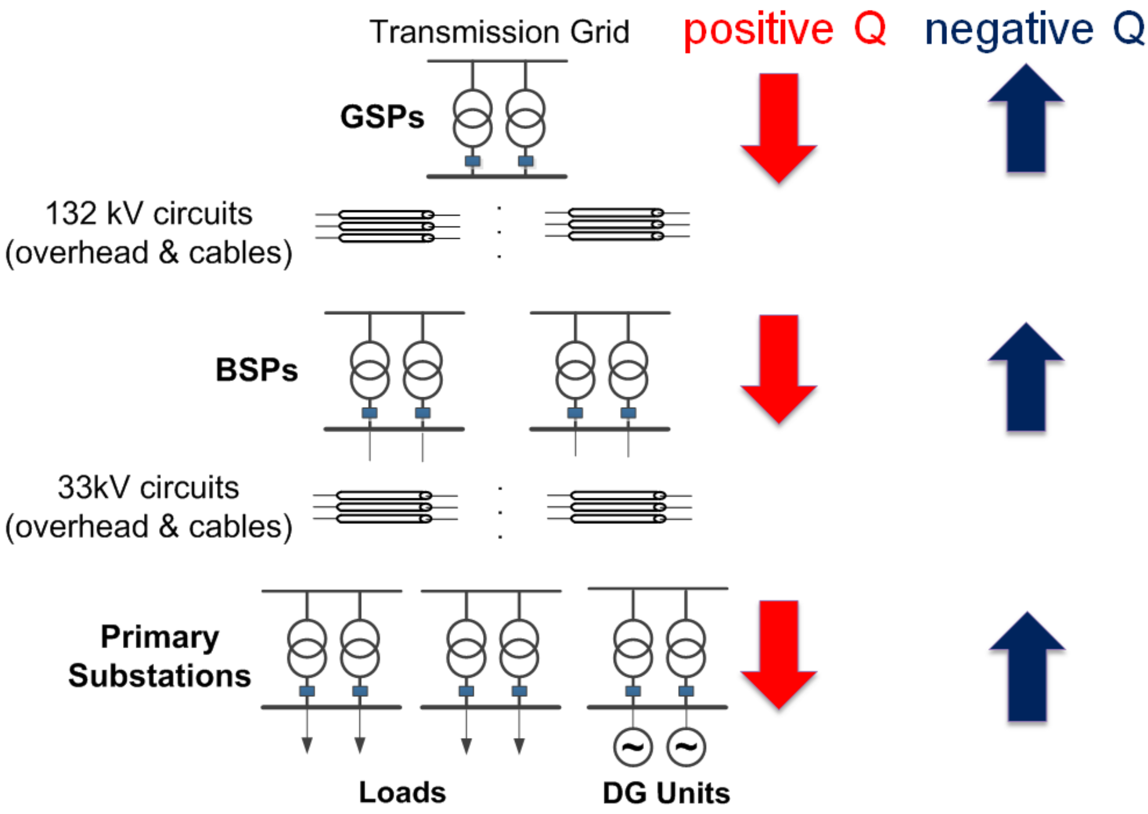

Fig. 17. Flow of reactive power in a typical UK distribution network.

The bidirectional flows of reactive power at substations is the main factor that requires differentiated settings for the coefficients of Stages 1 to 4 of the proposed data processing methodology

<u>Stage 1</u>

- Identification of step changes: similarly to the processing of P demand data, the dQ_i/dt values that exceed a bandwidth of ±80% of the standard deviation of the examined time-series Q data (i.e., SD(**Q**)) are identified as potential step changes in Q demand ;

Stage 2:

- *Control #1:* Given that Q demand can exhibit both positive and negative values (particularly for GSPs and BSPs), there is no initial treatment of negative values as zero values. The control logic leading to demand shifting can be applied similar to the processing of P data using a c_1=50% coefficient for both BSPs and primaries. After this shifting process of Q demand, any values above 150MVAr and below -20MVAr for BSPs are treated as erroneous and set equal to zero.

- *Control #2:* Similarly to the processing of P data, the same settings of c_2=30 and 40% for BSPs and primary substations, respectively, can be used for the Q data. Nonetheless, it should be highlighted that the control logic needs to be applied to the normalised time-series Q data (i.e., 0 min to max 1 pu demand) to avoid erroneous results derived from processing positive and negative demand values at the same time.

- *Control #3:* Similarly to Control #2, normalised Q demand values need to be considered again. The proper selection of $c_{3\text{-min}}$ and $c_{3\text{-max}}$ coefficients for BSPs and primary substations requires further investigations

Stage 3

- Similarly to the processing of P demand data, switching operations and temporary network reconfigurations are identified for both BSP and primary substations considering a simultaneous ±10% and ±5% change in daily peak and minimum Q demand, respectively.

Stage 4

- This stage is tailored for the processing of P demand data and should in general not be considered for the processing of Q demand data.

Data issues

The successful creation of Q profiles using the proposed processing methodology for Q demand data is of course limited by the raw Q measurements, and in particular issues such as:

- the ability of the monitoring and SCADA system to accurately identify and capture the direction of the Q flow;
- the ability of the pre-processing system (i.e., Demand and Generation Dashboard) to estimate meaningful Q demand data in the cases of false measurements of "negative" Q demand; and,
- the overall accuracy of raw P and Q monitoring data (under review, as discussed in section 1.2).

The above mentioned issues are more critical for Q than for P, and particularly at primary substations, which can exhibit particularly low Q demand values. There can be cases where the processed Q demand data are not representative of the actual Q demand of a substation, and current work is investigating whether it could practically useful to apply the processing methodology on the *aggregated* demand of a *group* of primary substations rather than to individual primaries (e.g., per BSP substation).

It should be highlighted that a visual sense check in some cases can be used to select intermediate results of the methodology, particularly for small sized primaries where the quality of Q measurements is poorer.

4.3 Recommendations for Future Work

This subsection provides recommendations for updates and additional work aiming to enhance the use of the proposed methodology.

Use of additional data beyond the Demand and Generation Dashboard

Currently the proposed methodology uses as inputs the available data of the Demand and Generation Dashboard of Electricity North West Ltd (i.e., half-hourly demand data of GSP, BSP and primary substations). The preference is to correct and enhance this data source to make it representative of the normal running demand of substations e.g. historic data missing after system changes. However where this is not possible, future updates of the proposed methodology could consider the use of additional data that could potentially enhance the performance of the processing methodology, such as:

- monitoring data of HV feeders and the associated network connectivity per primary substation; and,
- network connectivity of DG units in EHV and HV networks.

Updates for 64-bit operating system

The prototype tool that applies the proposed methodology has been tested on personal computers running on 32-bit operating systems. This fact combined with the large volumes of analysed data has posed certain limitations in modelling. More specifically, the prototype tool has been formulated accordingly so as to assign vectors containing the whole range of time-series demand data for each substation.

Using a 64-bit operating system the corresponding formulation could involve matrices that contain the data of a large number of substations (e.g., one matrix for 5 years half-hourly demand data of all BSPs). Given that linear algebraic computations are used in the proposed methodology, the use of such matrices could lead to significantly reduced computational times.

Post-processing of time-series demand data

Future work should also focus on the post-processing of the outputs of the prototype data processing tool, including:

a) the application of weather corrections;
b) seasonal regression and error analyses to identify historical demand trends; and,
c) the extraction of daily demand profiles of substations to be used in time-series power flow simulations and other studies that require network modelling.

Weather correction of P demand data can be easily implemented by multiplying demand data with weather correction coefficients. Therefore, the main challenge is to have meaningful time-series weather correction coefficients that reflect the actual effects from weather variations on P demand during both periods of peak and minimum load.

Demand forecasting requires the identification of the historical demand trends using the processed data of the proposed methodology. These trends can be used for the recalibration of demand forecasting models, as well as components of these models. In the ATLAS project the half-hourly P and Q demand data will be further analysed using different fitting approximations and associated error analyses to identify seasonal historical trends of P and Q demand, as well as the corresponding Q/P ratios during both periods of peak and minimum load.

The penetration of different DG types that exhibit a large variety of daily profiles in distribution networks has led to an increased interest in network studies that involve time-series simulations. Depending on the preferred approach, daily demand profiles of substations could be obtained a) for typical dates, b) as average monthly/seasonal profiles or c) as a typical profile considering meaningful volatility / probabilistic variations. Therefore, future work should focus on the further processing of the outputs of the proposed methodology to derive proper daily demand profiles for network studies.

4.4 Summary

This section first presented practical insights from the application of the proposed methodology to process time-series monitoring demand data of Electricity North West Ltd. The benefits from the modular structure, as well as the generic and fully automated application of the methodology were described. The associated computational times using the developed prototype tool in order to process a year’s long half-hourly demand data per substation was found less than 3.5 s in most of the cases.

Next, recommendations were presented for the application of the methodology for P demand data, as well as suggestions for the extension to Q demand data. More specifically, proper settings of the different control modules of the data processing tool were suggested for P demand data of BSPs and primary substations.

Recommendations for future work were also described. Potential benefits from the development of a processing tool using a 64-bits system were discussed, whereas the post-processing of the outputs of the prototype data processing tool could include a) the application of weather corrections, b) seasonal regression and error analyses to identify historical demand trends; and, c) the identification of proper daily demand profiles to be used in time-series power flow simulations.

Finally, a semi-automatic approach for the practical implementation of the proposed methodology to produce half-hourly profiles for multiple years for the GSP, BSP and primary substations of Electricity North West Ltd was presented. Nonetheless, the importance of visual sense checks of the obtained processed demand data was also highlighted.

5 CONCLUSIONS

In this report the Data Processing methodology that will be further used in the Architecture of Tools for Load Scenarios (ATLAS) project has been presented. The methodology is used in the ATLAS project to adjust monitoring time-series P demand data that might be erroneous, missing or not representative of the actual demand of a substation or group of interconnected substations.

The key characteristics of the proposed Data Processing methodology are that:

- it involves both half-hourly and daily analyses to deal with monitoring data issues including step changes from switching operations and network reconfigurations, missing and/or erroneous data (e.g., half-hourly spikes and/or dips),
- it is generic and can be applied even without any extra network connectivity data;
- it allows minimum recalibration when applied to GSPs, BSPs or primary substations; and
- it has a modular structure that provides flexibility in its use (e.g., selection of smoothing of demand oscillations and/or purely half-hourly analyses).

The proposed methodology is tailored to the monitored component of true P demand data (i.e., the combination of measured demand with metered export from local generators). A tool has been successfully developed and used to process and improve a 5-year period of half-hourly monitored P demand data of over 70 BSP and 380 primary substations at Electricity North West Ltd. An adapted version of the methodology is also being used to process Q demand data.

A practical implementation flowchart is provided, which demonstrates how the proposed methodology could be used to analyse a past year's monitoring data. This highlights the approach to sense-check the results of the automatic processing compared to the previous year's peak. In limited cases where the available monitored P and Q demand data is significantly incomplete, this process uses the half-hourly processed demand data together with more detailed manually obtained data (e.g., measurements of HV feeders, connectivity data of DG units) to obtain half-hourly demand profiles that can be used for further studies.

Given that both transmission and distribution operators are interested in more accurate assessments of the underlying true P demand, this report also presents an initial approach to estimate half-hourly the outputs of non-monitored distributed generation. The approach is implemented for over 30,000 DG units, which are connected downstream primary substations across Electricity North West's region and for which no monitoring data is available to the DNO.

The aggregated latent demand of these DG units is combined with processed P demand data (i.e., monitored component of true demand) across all primaries of Electricity North West to provide a more accurate estimation of the true demand (i.e., half-hourly aggregated primary demand from financial year 2012 to 2016).

The developed prototype tool of the proposed Data Processing methodology will be further used in the ATLAS project to:

- support the identification of historical trends of P demand, during periods of peak and minimum load using different periodicities (e.g., seasonal trends or individually per month); and,
- provide a representative set of daily half-hourly demand profiles across all (or most of) GSPs, BSPs and primary substations of Electricity North West Ltd, as the basis for scenario development and network modelling.

6 REFERENCES

[1] Engineering Directorate of Energy Networks Association (ENA), “Application Guide for Assessing the Capacity of Networks Containing Distributed Generation”, Engineering Technical Report 130, Jul. 2006.

[2] Electricity North West Ltd, “Code of Practice 260 – P2/6 Compliance Register”, issue 1, Jan. 2010.

[3] J. Crabtree, A. McHarrie, “CT Accuracy Measurements”, EA Technology.

[4] Electricity Engineering & Planning Department, “Computation Engine for Demand Forecasting”, Expansion Planning and Pricing (EPP) project, United Utilities Electricity Ltd.

[5] National Grid, “System Operability Framework 2015”, November 2015. [Online]. Available: http://www2.nationalgrid.com/UK/Industry-information/Future-of-Energy/System-Operability-Framework/

[6] Department of Energy and Climate Change. “Solar Photovoltaics Deployment in the UK”, December 2015. [Online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/488141/Solar_photovoltaics_deployment_December_2015.xlsx

[7] I. Richardson, M. Thomson, D. Infield, and C. Clifford, “Domestic electricity use: A high-resolution energy demand model,” Energy Build, vol.42, pp. 1878-1887, Oct. 2010.

[8] Department for Business, Energy and Industrial Strategy, “Digest of United Kingdom Energy Statistics 2016”, 2016. [Online]. Available: https://www.gov.uk/government/statistics/digest-of-united-kingdom-energy-statistics-dukes-2016-main-chapters-and-annexes

[9] J. A. Jardini, C. M. V. Tahan, M. R. Gouvea, F. M. Figueiredo, "Daily load profiles for residential, commercial and industrial low voltage consumers", IEEE Transactions on Power Delivery, vol 15(1), Jan 2000.

[10] L. Pedersen, R. Ulseth, "Method for load modelling of heat and electricity demand", 10th International Symposium on District Heating and Cooling, September 2006.

[11] MATLAB 15a, The MathWorks Inc., Natick, MA, 2015.

[12] C. G. Kaloudas, L. F.Ochoa, B. Marshall, S. Majithia, “Deliverable 4. Second Year Six Month Report - REACT Project”. Apr. 2015 [Online]. Available: http://www.smarternetworks.org/Files/REACT_151111105435.pdf