Title: **Advanced Planning of PV-Rich Distribution Networks - Deliverable 2: Innovative Analytical Techniques**

Synopsis: This document presents a smart meter-driven analytical technique proposed by The University of Melbourne to estimate PV hosting capacity in distribution networks. Two significantly different HV feeders, urban and rural, are modelled in detail with growing PV penetrations in a horizon of 5 years to create a large realistic smart meter data set. The analytical technique is then applied to this data set for different PV penetrations. The findings show that the proposed analytical technique provides adequate estimations of PV hosting capacity, making it a faster and simpler alternative to model-based approaches.

Document ID: UoM-AusNet-2018ARP135-Deliverable2_v02

Date: 22nd October 2019

Prepared For: Tom Langstaff
Manager Network Planning
AusNet Services, Australia
tom.langstaff@ausnetservices.com.au

Justin Harding
Manager Network Innovation
AusNet Services, Australia
justin.harding@ausnetservices.com.au

Prepared By: Dr Andreas T. Procopiou
Senior Research Fellow in Smart Grids
Department of Electrical and Electronic Engineering
The University of Melbourne

Revised By: Prof Luis(Nando) Ochoa
Professor of Smart Grids and Power Systems
Department of Electrical and Electronic Engineering
The University of Melbourne

Contact: Dr Andreas T. Procopiou
andreas.procopiou@unimelb.edu.au

Prof Luis(Nando) Ochoa
+61 3 9035 4570
luis.ochoa@unimelb.edu.au

This project received funding from the Australian Renewable Energy Agency (ARENA) under its Advancing Renewables Program. The views expressed herein are not necessarily the views of the Australian Government, and the Australian Government does not accept responsibility for any information or advice contained herein.

Executive Summary

This report corresponds to *"Deliverable 2: Innovative Analytical Techniques"* part of the project Advanced Planning of PV-Rich Distribution Networks with funding assistance by the Australian Renewable Energy Agency (ARENA) as part of ARENA's Advancing Renewables Program and led by the University of Melbourne in collaboration with AusNet Services. The project is established to develop analytical techniques to assess residential solar PV hosting capacity of electricity distribution networks by leveraging existing network and customer data. Additionally, planning recommendations will be produced to increase the hosting capacity using non-traditional solutions that exploit the capabilities of PV inverters, voltage regulation devices, and battery energy storage systems.

This document focuses on the methodology and assessment of a smart meter-driven analytical technique proposed by The University of Melbourne to estimate PV hosting capacity in distribution networks; two significantly different HV feeders, urban and rural, are considered. This document starts with some additional modelling aspects relevant to the modelling of HV-LV feeders. Then, the smart meter data provided (by AusNet Services) for the purposes of this project are detailed. Given the importance of using large volumes of smart meter data to explore and understand the relations between customer data and network state under different PV penetration levels, a methodology to produce realistic (hybrid) smart meter data for a horizon of 5 years is proposed. Lastly, this report presents the proposed analytical technique that makes use of smart meter data to construct a statistical regression model for each LV network, in a given HV feeder, and estimate its corresponding PV hosting capacity. The performance of the proposed hosting capacity estimation (HC estimation) methodology is assessed under three different PV system uptake trends, as well as considering the effects of network controllable elements such as the zone substation OLTC.

The key aspects of this report are summarised below.

Smart Meter Data

- Significant challenges related to data privacy and confidentiality issues prevented the facilitation of multiple days of historical smart meter data to The University of Melbourne. Nonetheless, efforts were made to provide 2 days' worth of 5-min resolution encrypted and anonymised data from ~3000 residential customers for the Urban HV Feeder U2 (CRE21).
- While the provided smart meter, data contain enough measurements (P, Q, V) from customers with PV system installations (20% of customers with PV systems), it was not possible to extract meaningful correlations between the PV penetration and its effects (i.e., voltage rise). This can only be captured by historical data that covers the evolution of PV penetration in time.

Hybrid Smart Meter Data

- Given the limited data availability, a methodology is proposed to produce a large volume of 30-min resolution hybrid smart meter data for a horizon of 5 years and with progressive PV penetrations. Actual anonymised demand (P, Q) and irradiance profiles from a previous project *"AusNet Mini Grid Clusters"* were used to run unbalanced, 30-min resolution, time-series, three-phase four-wire power flows for multiple days to extract customer voltages, V. In total, the database of the hybrid smart meter data (P, Q, V) produced for each HV-LV Feeder consists of more than 1 billion data points (>3Gb).
- Leveraging statistical techniques, the hybrid smart meter data were analysed to extract potential daily correlations and hint the direction towards the analytical approach to be adopted. For each LV network, a very strong linear correlation was found between the maximum voltage on a given day and the corresponding sum of all smart meter active powers (P, which can be negative due

to PV systems). These two features were used as inputs to the proposed Smart Meter-Driven PV Hosting Capacity Estimation methodology.

Smart Meter-Driven Hosting Capacity Estimation

- Based on a simple, yet practical, machine learning algorithm, a methodology is proposed to produce a regression model to estimate the PV hosting capacity in any given LV network using smart meter data. The main steps of the methodology, as if implemented by a DNSP, are presented below.
    - Smart meter database. For a given number of days (ideally covering most of the evolution of PV penetrations to date), the daily smart meter data (i.e., P, Q, V) from all customers in a given LV network are extracted from the smart meter database.
    - Data Processing. The smart meter data are analysed and cleaned from missing and inconsistent values. Then, the maximum voltage recorded for each day is identified and the corresponding (same timestamp) active powers are added up. Finally, a new dataset is produced containing the maximum voltage and the corresponding aggregated power for each day.
    - HC Estimation Model. The new dataset is used to train a supervised (i.e., gradient decent) univariate regression model which corresponds to the HC estimation model for the analysed LV network.
- **HC Estimation.** The model, in effect, estimates the aggregated active power (that can be negative due to PV systems) that can lead to voltages outside a pre-determined upper limit (e.g., 1.1 p.u.). This value, in turn, can be used to calculate the additional PV capacity that can be hosted by the LV network. Nonetheless, to understand what the HC Estimation might mean across customers, the estimated aggregated active power is presented as the diversified active power per customer. The latter also includes prediction limits to cater for uncertainties.
- **Accuracy**. It is important to highlight that the volume of smart meter data used to produce the HC estimation model plays an important role. More data helps to capture the variance of a larger sample of network conditions (i.e., voltage vs active power), thus increasing the model's estimation accuracy.

Case Studies

- The performance of the proposed HC estimation methodology is demonstrated and thoroughly assessed on two significantly different HV-LV feeders; urban and a rural.
- The assessment considers three different PV uptake trends through a horizon of 5 years.
    - Random PV Uptake. New PV system installations are randomly allocated to customers within the LV networks. A random allocation of PV systems represents a very realistic scenario which is currently seen in practice (i.e., residential PV systems are adopted by customers located at different locations within the network).
    - Head to End PV Uptake. New PV systems are allocated first to customers closer to the head of the LV feeders and then moving towards those at the far end. While unlikely, it represents one of the two extreme scenarios. It leads to the highest PV hosting capacity as the effect of voltage rise is in general lower for points closer to the head of the feeder (i.e., smaller impedance, hence smaller voltage drop/rise).
    - End to Head PV Uptake. New PV systems are allocated first to customers at the far end of the LV feeders and then moving towards those at the head. While also unlikely, it represents the other extreme scenario. It leads to the lowest PV hosting capacity as the effect of voltage rise is in general higher for points farther from the head of the feeder (i.e., larger impedance, hence larger voltage drop/rise).
- **Urban Feeder U2 (CRE21)**
    - Overall, it was found that the proposed methodology can provide meaningful and adequate HC estimations for this and similar urban feeders. In this case, such estimations were achieved from as early as 30% of PV penetration for the Random PV and End to Head PV uptake trends.

- For the Head to End uptake trend, it was found that early PV penetrations did not result in significant impacts, resulting in slight HC overestimations. This is primarily due the fact that customers expected to affect voltage rise the most (i.e., farthest customers) are the last installing a PV system; hence, the HC model cannot capture these effects until high PV penetration levels (i.e., >60%). Moreover, due to the relatively higher number of customers (>100) and feeders (>2) in urban LV networks, significant diversity can exist in terms of the LV feeders' length and number of customers in the same network. This makes the estimation of HC more challenging in such uptake trend. For example, a new PV installation at the end of a long feeder with many customers might have a completely different voltage rise effect compared to another with shorter length and lower number of customers.
- A further analysis using SCADA data from 2016 to represent the zone substation's OLTC actions was carried out. Although these voltage changes might not capture how the CRE21 OLTC would in reality act with the different PV penetrations, it was found that it can slightly reduce the accuracy of the HC estimations. Furthermore, because of a higher number of outliers (voltage spikes), the ability of the HC estimation model to include them in the prediction limits reduces.

- **Rural HV Feeder R1 (SMR8)**
  - Overall, it was found that the proposed methodology can have a much better performance in this and similar rural feeders as it is able to provide meaningful and adequate HC estimations from much earlier PV penetrations regardless the PV uptake trend. In this case, such estimations were achieved with as little as 20% PV penetration.
  - The higher accuracy of HC estimations at earlier PV penetrations can be explained due to the lower number of customers and feeders (up to 2) in rural LV networks. This means that the impacts of PV installations in a given LV network will evolve consistently, i.e., higher voltages will be seen with more PV installations (which is not the case in urban LV networks with multiple feeders due to the diversity in length and customer numbers). This consistency allows the HC estimation model to capture the effects more accurately and at earlier PV penetrations.

Table of Contents

List of Figures

List of Tables

1 Introduction

According to the Australian PV Institute, the aggregated installed capacity of solar PV in Australia is currently exceeding 6.5 GW, with many these installations being residential. The percentage of dwellings with solar PV varies from 12% in the Northern Territory to 30% in Queensland. This, combined with a growing number of commercial customers adopting the technology, will soon pose significant technical challenges on the very infrastructure they are connected to: the low voltage (LV) and high voltage (HV) distribution networks.

Due to the rapid uptake of the technology, many Distribution Network Service Providers (DNSPs) across the country have adopted the use of PV penetration limits based on the capacity of the distribution transformers feeding LV customers. Once this limit is reached, complex and time-consuming network analyses are often required to determine the need for any mitigating action due to asset congestion or voltage rise issues (e.g., network augmentation, use of off-load tap changers).

Whilst, in principle, the use of a PV penetration limit is a sensible approach to swiftly deal with many connection requests, the lack of advanced planning approaches has led DNSPs to adopt values that might under or over-estimate their actual hosting capacity, particularly due to voltage issues in LV networks and aggregated congestion issues in HV networks. Similarly, assessing the effectiveness of non-traditional solutions, such as actively controlling smart PV inverters or deploying distribution transformers fitted with on-load tap changers, becomes a task beyond typical planning studies carried out by DNSPs. All this, in turn, becomes a barrier for the widespread adoption of solar PV as it can create delays, increase cost, and could undermine the consumer attractiveness of the technology.

To help remove the aforementioned barriers and accelerate the adoption of solar PV in Distribution Networks, this project is established to develop analytical techniques to rapidly assess residential solar PV hosting capacity of electricity distribution networks by leveraging existing network and customer data. Additionally, planning recommendations will be produced to increase the hosting capacity using non-traditional solutions that exploit the capabilities of PV inverters, voltage regulation devices, and battery energy storage systems.

The report at hand is structured as follows: Chapter 2 provides some additional modelling aspects relevant to the modelling of HV-LV feeders. Then, in chapter 3 the smart meter data provided (by AusNet Services) for the purposes of this project are detailed. Given the importance of using large volumes of smart meter data to explore and understand the relations between customer data and network state under different PV penetration levels, a methodology to produce realistic (hybrid) smart meter data for a horizon of 5 years is also proposed. Chapter 4 presents the proposed analytical technique that makes use of smart meter data to construct a statistical regression model for each LV network, in a given HV feeder, and estimate its corresponding PV hosting capacity. The performance of the proposed hosting capacity estimation (HC estimation) methodology is assessed under three different PV system uptake trends, as well as considering the effects of network controllable elements such as the zone substation OLTC. Finally, conclusions and next steps are presented in Chapter 5 and 6 respectively.

2 HV-LV Feeders – Additional Modelling Aspects

This chapter presents additional modelling aspects and assumptions for the HV-LV feeders presented in the report "*Deliverable 1: HV-LV modelling of selected HV feeders*" [1]. These updates, which are based on updated information and data provided by AusNet Services, correspond to an updated model of the Urban HV Feeder U2 (CRE21) and the operation of the HV capacitors.

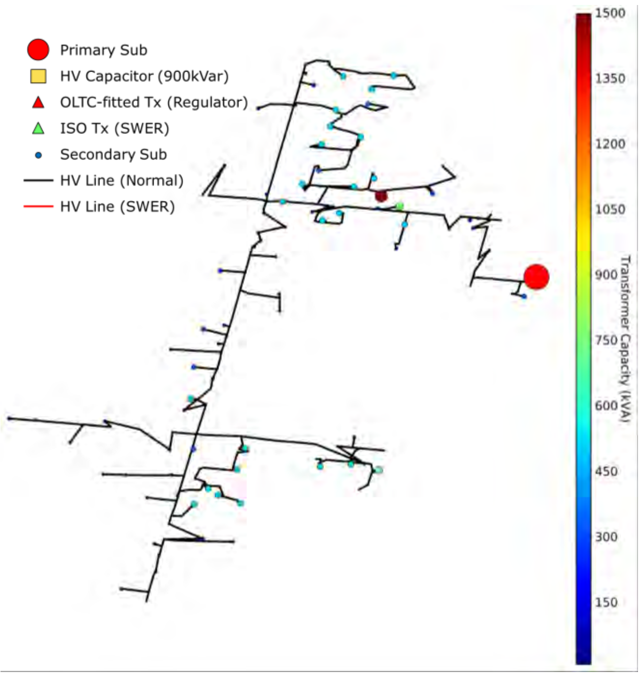
2.1 Urban HV Feeder U2 (CRE21) – Updated Model

This section presents the topology and general characteristics of the updated Feeder U2 model. Furthermore, for demonstration purposes, this section presents a time-series power flow analysis of the updated HV Feeder U2 considering a peak demand day. These analyses are presented to understand the time-series behaviour of the updated HV feeder.

2.1.1 Topology and General Characteristics

Figure 2-1 shows the topology of the updated Feeder U2, along with its general characteristics. The topology of the feeder model remains exactly the same as the one presented [1] and the updated information, highlighted in red, is listed below:

- The total number of residential LV networks is updated to 71 (79 in initial model).
- The total number of residential customers supplied is updated to 3,374 (4,626 in initial model).
- The total number of non-residential LV networks is updated to 9 (0 in initial model).
- The total number of residential customers supplied is updated to 9 (0 in initial model).



# of LV Residential Substations:	71
# of LV Residential Customers:	3,374
# of LV Non-Residential Substations:	9
# of LV Non-Residential Customers:	9
Distance of Farthest Transformer (km):	9
# of HV Capacitors (900kVar each):	0
# of SWER Transformers:	0
# of REGULATOR Transformers:	0
Total HV Conductors (km):	30
Total HV SWER Conductors (km)	0

Figure 2-1 Feeder U2 – Topology and General Characteristics

2.1.2 Business-as-Usual Analysis

Load profiles for each customer are randomly selected and allocated based on the procedure described in section 4.1.1 of deliverable [2] for the day 15 January 2014 (considered to be the highest demand day in the pool of smart meter data).

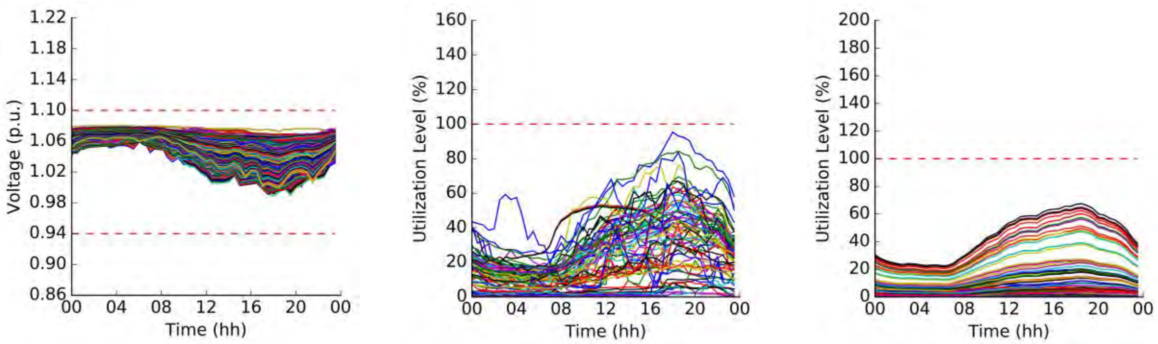
Figure 2-2 (a) shows the daily 30-min voltage profiles of all customers in U2. All profiles lie within the statutory voltage limits (i.e., 1.10 - 0.94p.u.). Given that the off-load tap position of all secondary distribution transformers is assumed to be at the nominal position (i.e., position 3), voltages during low demand (morning hours) are close to ~1.08p.u. However, during peak demand (late evening and night hours), when customers return home from work (and demand increases), voltages reduce to ~0.98p.u. While the new feeder model has a lower number of residential customers (i.e., 1252), a slightly lower voltage level is noticed when compared to the initial model. This can be explained given the fact that the new model considers 9 large non-residential loads (assumed to be ~2MWp commercial loads) which have a significant contribution to the voltage drop.

Similar to initial model, it is also important to highlight that this specific case (peak demand) allows quantifying the maximum level of voltage drop that can be faced in this network and hence understand the available tap position foot-room for each distribution transformer (i.e., ability to reduce off-load tap position without resulting in violation of the lower voltage limit). For this HV feeder, a foot-room of approximately 0.05p.u. exists which can be translated into 2 off-load taps positions (2.5% per step).

Considering the utilisation level of the assets (i.e., distribution transformers and HV lines), shown in Figure 2-2 (b) and (c), a similar behaviour to the initial model is also noticed. A higher utilisation is observed during peak demand (late evening and night hours) compared to the low demand during morning hours. All transformers (except a couple) and lines operate within their limits and the peak transformer and line utilisations are around ~67 and ~70%, respectively.

It should be highlighted that the corresponding behaviour is observed because the demand profiles used correspond to the peak demand day, hence the utilisation level of the assets is expected to be high as well as the voltage profiles low. However, it should also be highlighted that the load profiles used, correspond to customers located in different geographical area than the studied network. Hence the use of these load profiles (residential and non-residential) might be overestimating the behaviour.

To provide more understanding of the loading conditions of this case study, the monitored kVA power at the primary substation is presented in Figure 2-3 and the peak utilization level of the HV lines is demonstrated in Figure 2-4 considering a topology heatmap. In general, the updated model was found to have almost the same behaviour as with the initial model.



(a) (b) (c)

Figure 2-2 Feeder U1 BaU: (a) Voltage profiles, (b) LV Transformers and (c) HV lines utilisation

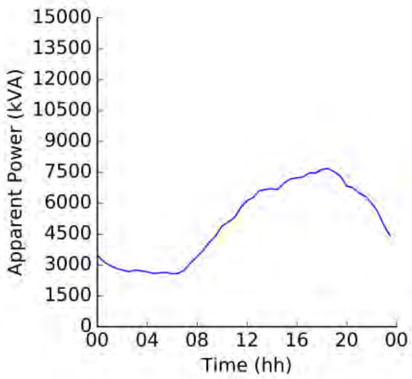


Figure 2-3 Feeder U2 BaU: Primary Substation Monitored Power (kVA)

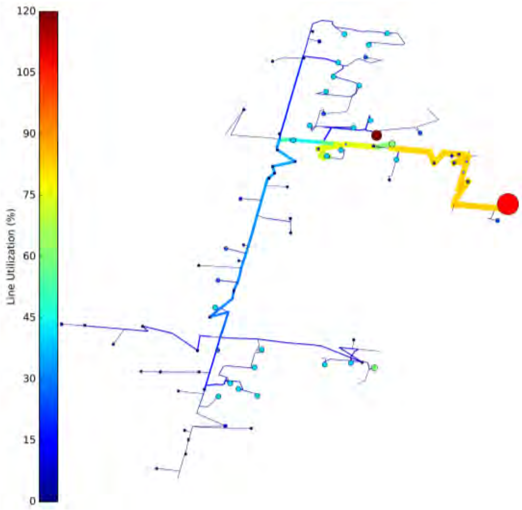


Figure 2-4 Feeder U2 BaU: Line Utilization Level Heatmap

2.2 HV Capacitors – Updated Operation

Based on updated information provided and discussions with by AusNet Services, all HV capacitors located in the Rural HV feeders (i.e., R1, R2) are now modelled with voltage-based operation (instead of time-based) so that their modelled operation is aligned with the current practice. In more details, the actions of any capacitor connected in an HV Feeder (operated by AusNet Services) are defined based on the voltage level measured at the connection point of the capacitor.

Considering the above information, the capacitors are assumed to:

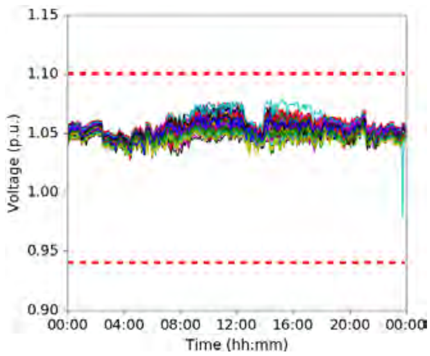
- Switch **on:** if the voltage at the point of connection is equal of lower than 0.96p.u. This is expected to boost the voltage at the connection point as well as downstream voltages closer to 1.0 p.u.
- Switch **off:** if the voltage at the point of connection is equal of higher than 1.08p.u. This is expected to lower the voltage at the connection point as well as downstream voltages closer to 1.0 p.u.

3 Smart Meter Data

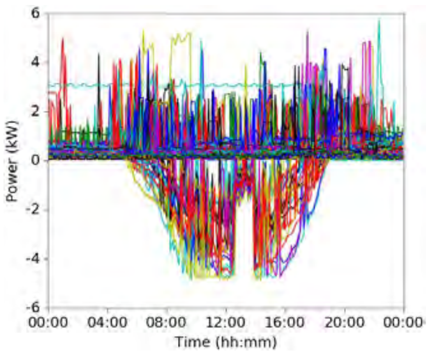
This chapter first presents a small volume (i.e., 2 days) of smart meter data (~3,000 residential customers), provided by AusNet Services, for one of the Urban HV Feeder U2 (CRE21). Then the importance of using large volumes of data in extracting correlations is highlighted along with the challenges facing to acquire additional larger sets of smart meter data. Then to tackle these challenges and help develop and validate those analytical techniques that can be used to estimate the Hosting Capacity of LV networks in a given HV feeder, a methodology is proposed to produce a large volume (i.e., 5-year, 30-min resolution) of smart meter data (referred here as hybrid smart meter data) for each customer in the modelled HV feeders. These smart meter data consider a realistic progressive adoption of PV systems (i.e., 0 to 100% of customers with PV, in a given HV feeder) through a horizon of 5-years. While hybrid smart meter data were produced for all modelled HV Feeders, for simplicity and demonstration purposes, the presentation of the produced hybrid smart meter data in this report, corresponds only to those of the Urban HV Feeder U2 (CRE21).

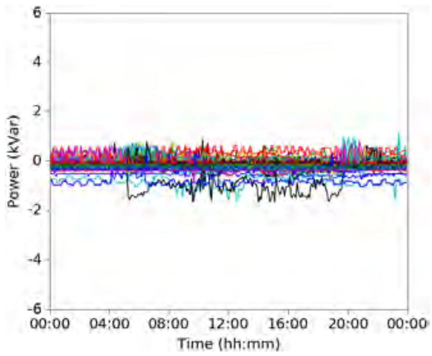
3.1 Provided Smart Meter Data

A small volume of smart meter data was facilitated by AusNet Services to the University of Melbourne. These correspond to encrypted and anonymised smart meter data of approximately 3,000 customers in the HV Feeder U2 (CRE21) [1] for two days: the peak (20 January 2019) and minimum (16 November 2018) demand days between the year 2018-2019. For each day 5-min resolution measurements of active (P), reactive (Q) power and voltage (V) are provided for each smart meter (i.e., 288 values per day for each parameter). These data were processed and cleaned from missing and inconsistent values and then grouped by LV substation. For demonstration purposes Figure 3-1 shows the daily (a) voltage, (b) active and (c) reactive power profiles of all customers (total of 108) in a sample LV network with almost 20% solar PV penetration (i.e., percentage of customers with PV).



(a) Voltage





(b) Active Power **(c) Reactive Power**

Figure 3-1 Smart Meter Data of all customers in a sample LV substation

While the provided data contain enough smart meter measurements from customers with PV system installations (20% of customers with PV system), extracting meaningful correlations between the evolution of PV generation, in this feeder, and the effects this might have (i.e., voltage rise) on the feeder was impossible to be achieved due to the limited number of days provided. This can only be captured by historical data that covers the evolution of PV penetration in time.

While smart meter data from the residential customers, within the area of AusNet Services, are in general available, significant challenges related to data privacy and confidentiality issues prevented the facilitation of additional data. Thus, acquiring smart meter data specific to HV feeders proved to be more challenging than expected.

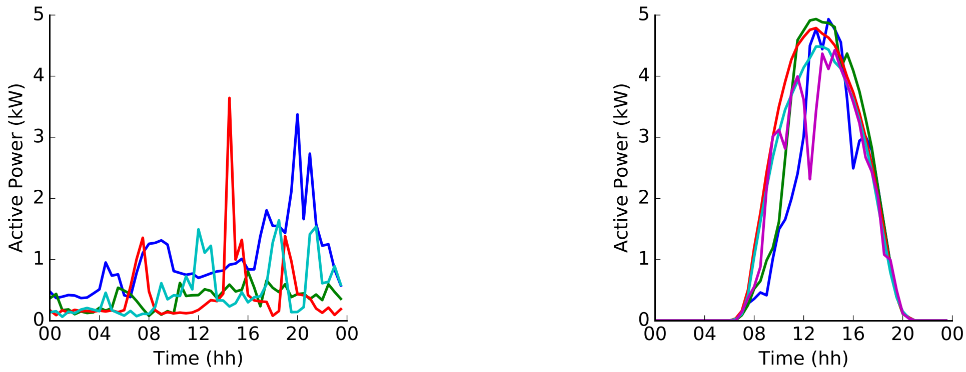
To define and validate different analytical techniques to assess the PV hosting capacity for a given HV feeder a large volume (i.e., years) of high-resolution (e.g., scale of minutes) smart meter data (i.e., P, Q, V) is required. Such data, which cover the evolution of PV penetration in time, will allow extract correlations between customer data and solar PV generation, important to define those analytical techniques that can help assess the PV hosting capacity.

3.2 Hybrid Smart Meter Data

To tackle the aforementioned challenges and help develop and validate those analytical techniques that can be used to estimate the Hosting Capacity of LV networks in a given HV feeder, a methodology is proposed to produce a large volume (i.e., 5-year, 30-min resolution) of smart meter data (referred here as hybrid smart meter data) for each customer in the modelled HV feeders. These smart meter data consider a realistic progressive adoption of PV systems (i.e., 0 to 100% of customers with PV, in a given HV feeder) through a horizon of 5-years.

For this purpose, a pool of 30-min resolution, year-long (i.e., 17,520 points), P and Q data from anonymized smart meters, collected from 342 individual residential customers in the year of 2014 is used. These data were facilitated to the University of Melbourne for the purposes of a previous projects with AusNet [2, 3]. Using this pool, the yearly demand and generation profiles were broken down in daily profiles, resulting in a pool of ~30,000 daily demand profiles and 90 daily generation profiles per season (total of ~12 million data points).

For demonstration purposes sample demand and generation profiles are presented in Figure 3-2.



(a) Customer Demand (b) PV generation

Figure 3-2 Sample demand and PV generation (assuming 5kWp system) profiles

This pool is then used to run unbalanced, 30-min resolution, time-series, three-phase four-wire power flows for multiple days (and years) covering various demand and generation scenarios for which the corresponding P, Q and V of each customer are collected to create a rich and precise realistic smart meter database for each of the HV feeders modelled in [1].

3.2.1 Methodology

To create a realistic smart meter database for each modelled HV feeder, power flow simulations were performed for a duration of 5 years where the penetration of solar PV (% of houses with PV) was progressively increasing from 0 to 100% throughout the years. This allows emulating the evolution of PV system integration and capturing their seasonal effects on the network, in a progressive manner.

The process to produce these smart meter data is descripted below:

1. For each customer, a random daily load profile is selected from the pool of daily load profiles corresponding and specific day of the year.

2. Depending on the PV penetration level (PV%) residential PV systems are randomly allocated to LV customers without a PV system. The size of the PV panels is based on Australian installation statistics where the proportion of PV installations with 2.5, 3.5, 5.5, and 8kWp is 8, 24, 52 and 16%, respectively. A realistic PV uptake is adopted by allowing uneven penetrations per LV networks and feeders as well as multiple PV installed capacities based on Australian PV installation statistics.

3. PV generation profiles are selected from the pool of daily PV generation profiles corresponding to a specific day of the year.

4. Unbalanced, 30-min resolution, time-series, three-phase four-wire power flows over a 24-hour period are carried out with OpenDSS.

5. The P, Q, and V of each customer are collected for each 30-min point of the 24-hours.

The process described above is repeated for 1,825 times each one representing a day of the 5-year horizon (i.e., 365 days x 5 years). The solar PV penetration growth (number of new PV installations) for each day is defined as the total number of customers in the HV feeder divided by the total number of simulation days (i.e., 1,825). In terms of the demand, no load growth is considered.

3.3 Case Study – Urban HV Feeder U2 (CRE21)

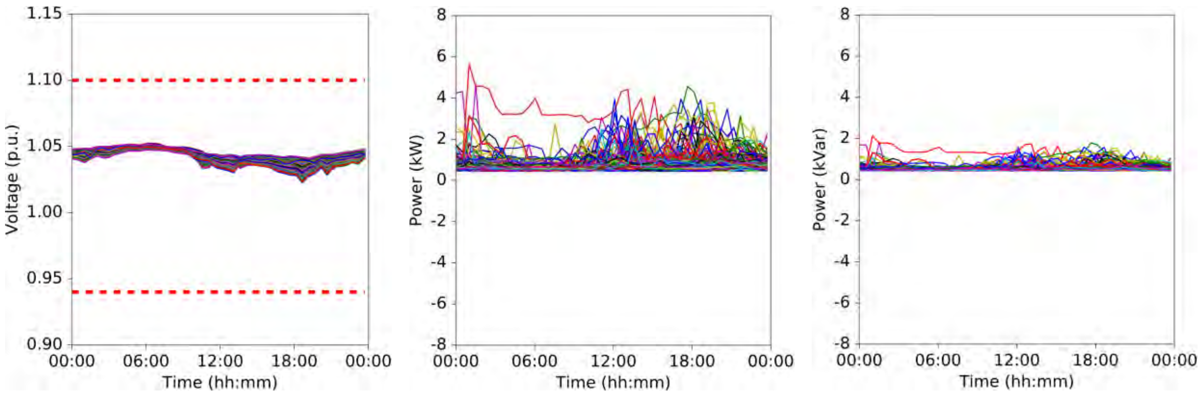
The methodology presented in the previous section was adopted to produce hybrid smart meter data for all 3,374 residential customers in the HV Feeder U2 (CRE21). For simplicity, the voltage at the head of the HV feeder is considered to be constant at 22 kV (1.0 pu) which corresponds to the voltage target setting used by the on-load tap changers (OLTCs) at the substation. Given the available voltage foot-room shown in 2.1.2, and to match the voltage level (~1.05 early morning hours) as shown in the provided smart meter data (see section 3.1, Figure 3-1a), all LV transformer off-load taps are reduced by one step (i.e., set to position 2 effectively transforming to 420 V), leading to ~1.05p.u. of voltage on the secondary side of the LV transformer assuming no-load conditions.

For this case study, the daily PV penetration growth for this feeder is defined as 0.06% and corresponds to approximately 2 new PV system installations per day. For example, a 20% penetration of solar PV is reached approximately on the 337th day (total of 1,825) of the analyses.

Once the simulation process is finished, a set of 87,600 P, Q and V points is created for each residential customer, representing 5-years' (i.e., 30-min resolution) worth of smart meter data measurements (P, Q and V). For simplicity and ease of referencing, these data were timestamped in the form of YYYY-MM-DD HH:MM:SS and given a range between 2018-01-01 00:00:00 to 2022-12-30 23:30:00. In total, the database of the hybrid smart meter data produced for the rural HV Feeder U2 (CRE21) consists of almost 1 billion data points (>3Gb). These will be used to produce analytical techniques to help estimate the corresponding PV Hosting Capacity in a given HV feeder (and LV substations).

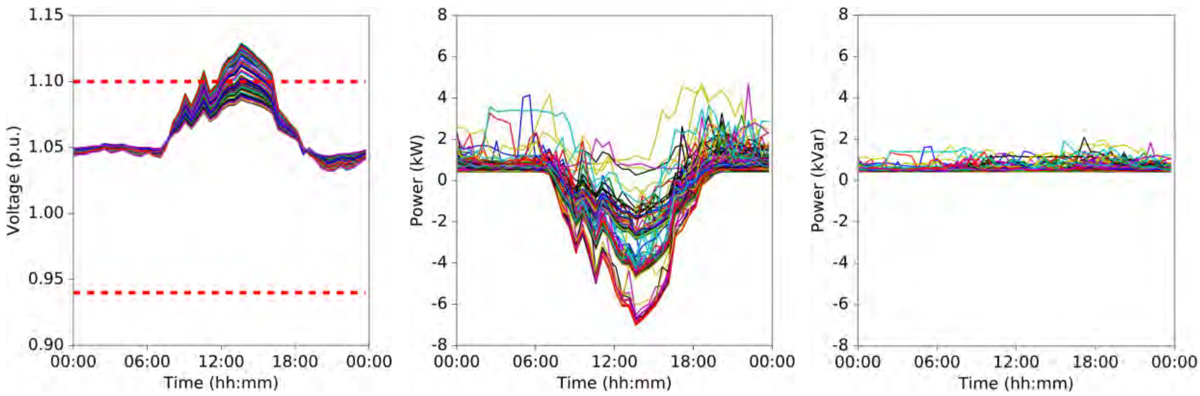
3.3.1 Visualization of Hybrid Smart Meter Data

For demonstration purposes the smart meter data (P, Q, V) of all residential customers (total of 138) connected to one of the LV networks connected (number 24 in the database) are presented in Figure 3-3 and Figure 3-4 for the days 2018-01-01 (0% PV) and 2022-10-30 (100% PV). As it can be observed, during the first simulation day, Figure 3-3, none of the customers have a PV system (all P values are positive – demand only) and all voltages are within the statutory limits. On the other hand, Figure 3-4 shows a case where all customers have PV systems (i.e., negative P during midday – PV generation) while some of the customers experience voltages that go beyond the maximum statutory limit. Considering the reactive power, no significant differences are observed for either of the two cases. It is also important to highlight that the since the PV systems are assumed to operate at a unity power factor, reactive power is only affected by the load. These profiles, and behaviour, are aligned with the measurements of the provided smart meter data (see Figure 3-1).



(a) Voltage **(b) Active Power** **(c) Reactive Power**

Figure 3-3 LV 24: Sample Hybrid Smart Data (138 customers): day 2018-01-01 (0% PV)



(a) Voltage **(b) Active Power** **(c) Reactive Power**

Figure 3-4 LV 24: Sample Hybrid Smart Data (138 customers): day 2022-10-30 (100% PV)

To provide more understanding of the produced smart meter database for each of the LV substations figures Figure 3-5 , Figure 3-6 and Figure 3-7 are provided to help visualize the evolution of data due to the solar PV penetration through the simulation years.

Figure 3-5, presents the maximum and minimum measured voltage (considering all customer voltages) for each 30-min point of the 5-year analysis. This allows visualizing the effect of increased PV penetration through the years which essentially pushes both the maximum and minimum voltages to higher levels as the time passes (i.e., higher number of PV systems; hence more generation).

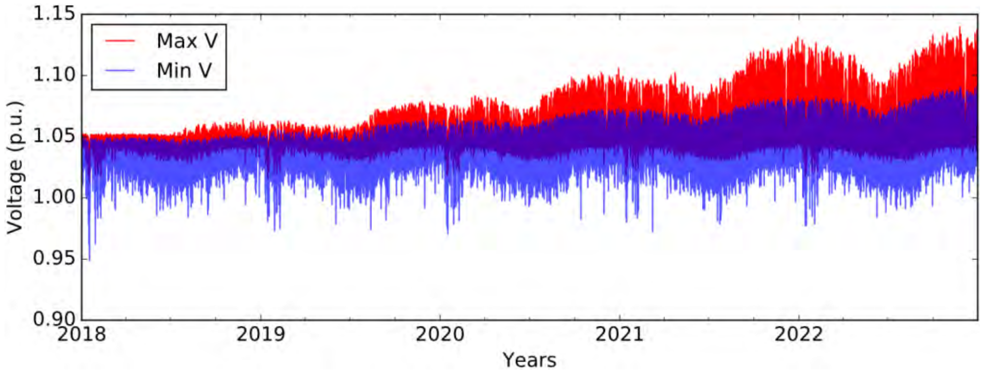


Figure 3-5 LV 24: Evolution of Maximum and Minimum Voltage

Indeed, as shown in Figure 3-6, which presents the aggregated active power of all customers in the corresponding substation, for each 30-min point of the 5-year analysis, a similar trend is noticed as with the voltages. The magnitude of the aggregated kW is progressively increasing through the years as a result of the increasing PV penetration.

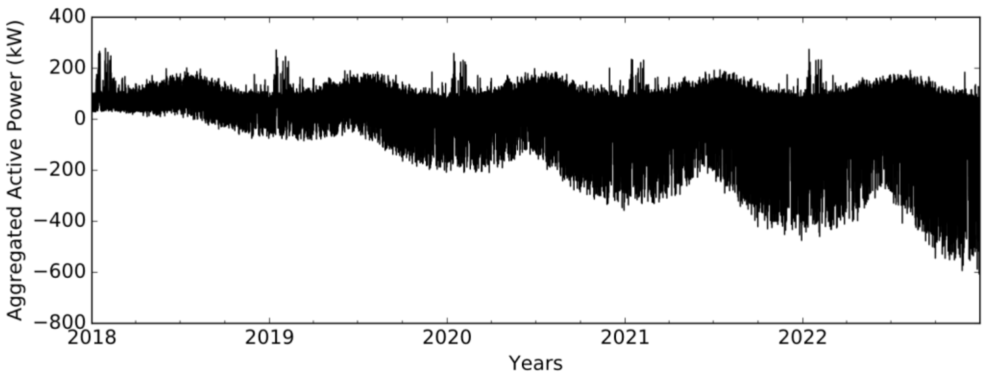


Figure 3-6 LV 24: Evolution of Aggregated Active Power (kW)

While voltages and aggregated power are progressively increasing in magnitude through the years, the aggregated reactive power (Figure 3-7) is kept at similar magnitude levels throughout the whole 5 years. This can be explained since the PV systems are assumed to operate at a unity power factor which is also aligned with the reactive power measurements presented in the provided smart meter data (see Figure 3-1).

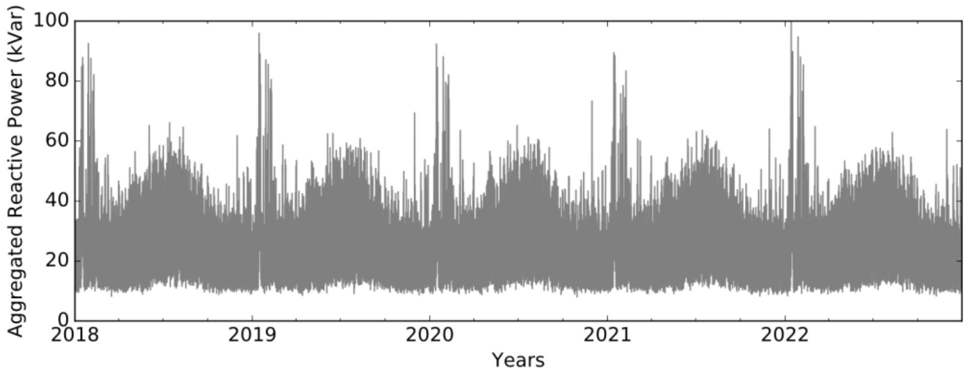


Figure 3-7 LV 24: Evolution of Aggregated Reactive Power (kVar)

3.3.2 Correlation of Hybrid Smart Meter Data

A key component in defining the most suitable analytical technique, is to understand the relation and potentially correlation of the corresponding parameters of the smart meter data. As briefly highlighted in the previous subsection a similar trend was observed on the maximum and minimum voltage evolution (Figure 3-5) with the aggregated active power evolution (Figure 3-6). However, the extent to which these datasets are related, needs to be statistically defined to hint the direction towards the analytical approach to be adopted.

For this, a widely used statistical method, the Pearson product-moment correlation (PPMC) coefficient, was adopted to measure of any linear correlation between the smart meter data parameters. Given that the PV Hosting Capacity in a given LV network is defined by the maximum customer voltage, this section tries to show the correlation of the maximum voltage (considering all customers in an LV network) with the aggregated active and reactive power. The minimum voltage is also considered for completeness.

The PPMCC, as defined in (1), is the covariance of any two sets of variables (i.e., X and Y) divided by the product of their standard deviations (σ_X and σ_Y) and ranges between +1 and −1, where:

- -1 indicates that the two variables are perfectly negatively linearly related.
- 0 means that two variables don't have any linear relation.
- 1 means that two variables are perfectly positively linearly related.

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \cdot \sigma_Y} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2} \times \sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \quad (1)$$

The PPMC coefficient of the smart meter data was calculated for all LV substations in the HV Feeder U2 (CRE21) [1] and the average coefficient values of the statistical analysis are provided in Figure 3-8, in a heatmap form. As observed, a very strong negative correlation (i.e., -0.94 and -0.92) exists between the aggregated active power and the corresponding maximum and minimum voltages. In other words, these parameters can be described by a negative linear relationship; the higher the PV generation (i.e., negative aggregated power) the higher the maximum voltage. On the other hand, the reactive power has a very weak correlation with the voltage.

Given the strong correlation between voltages (i.e., maximum and minimum) and the aggregated active power of all smart meters, these will be used in the next chapter as inputs to a proposed Smart Meter-Driven PV Hosting Capacity Estimation methodology.

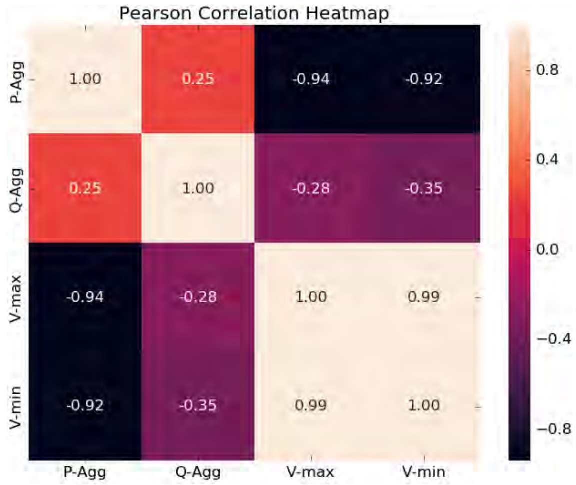


Figure 3-8 HV Feeder U2 (CRE21): Pearson Correlation of Smart Meter Data

4 Smart Meter-Driven Hosting Capacity Estimation

This chapter presents a proposed methodology to estimate the solar PV hosting capacity in any given LV network using smart meter data. Based on a simple -yet practical- machine learning algorithm with supervised learning, a univariate regression model is produced to estimate the maximum voltage level in a given LV network based on its state (i.e., aggregated active power). Then to demonstrate the proposed methodology an example is presented along with several performance related metrics and important aspects that should be taken into consideration when the methodology is adopted. Finally, two thorough case studies are presented where the proposed methodology is adopted on two very different feeders, the Urban HV Feeder U2 and the Rural HV Feeder R1. In both case studies, the methodology is adopted in all LV networks connected to each of the feeder and results are presented only for one of the LV networks. The overall performance is discussed considering all LV networks.

4.1 Methodology

Figure 4-1 presents the flow chart of the proposed methodology and the main steps are detailed below:

1. **Smart meter database.** In this step, the daily smart meter data (i.e., P, Q, V) from all customers (indexed i) in a given LV network are extracted from the smart meter database.
2. **Data Processing.** In this step, the smart meter data are analysed and cleaned from missing and inconsistent values. Then, the maximum voltage recorded for each day is identified and the corresponding (same timestamp) active powers are added up. Finally, a new dataset is produced containing the maximum voltage and the corresponding aggregated power for each day.
3. **Model Production.** The new dataset is used to train a supervised (i.e., gradient decent) univariate regression model i.e., $y = ax + c$, where y is the dependant variable (i.e., maximum voltage), x is the dependant variable (i.e., aggregated active power) and c is the y-axis intercept.
    a. First the dataset is spitted in two subsets: the training (75% of dataset) and test (15% of dataset) subsets.
    b. The training subset is used to train the model by adopting a gradient decent optimization algorithm to find the minimum of the loss function (i.e., mean square error of each predicted value of y and x).
    c. Once the training is finished, the test subset is used to evaluate the R^2 performance of the trained model. R^2 is a statistical measure that describes the percentage (0 to 100%) of the dependent variable variation been explained (correctly predicted) by the trained model. In general, the higher the R^2, the better the model can predict the dependant variable (i.e., maximum voltage) given an independent variable (i.e., aggregated active power).
4. **HC Estimation Model.** At this step the final HC estimation model can be used to estimate the hosting capacity of the corresponding LV network. An example of the HC Estimation model is given in section 4.2.

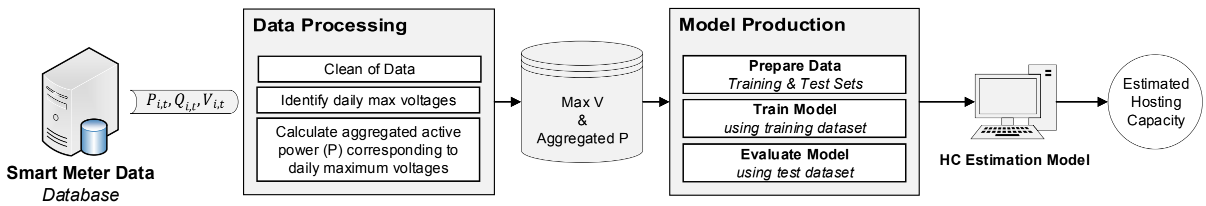


Figure 4-1 Smart Meter-Driven Hosting Capacity Estimation Flow chart

4.2 HC Estimation Model Demonstration

This section provides a demonstration example of the HC Estimation model and how this can be used to estimate the corresponding HC in an LV network. An example of the HC Estimation Model is presented in Figure 4-2 which for simplicity and demonstration purposes the model is trained using limited volume of smart meter data. First, the most important elements of the model are described. Then the process to estimate the corresponding HC is detailed along with its estimation performance. Lastly, the importance of the training data volume has on the model accuracy is discussed and demonstration examples are provided.

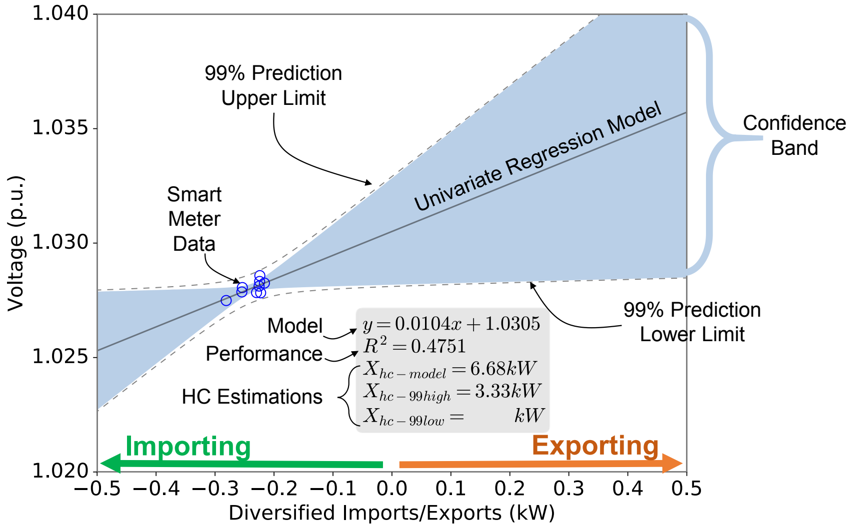


Figure 4-2 HC Estimation Model demonstration example

4.2.1 Parameter Description

- **Y-axis.** Corresponds to the daily maximum voltage (p.u.) of all customers in the LV network.
- **X-axis.** Corresponds to the diversified imports/exports (kW) in the corresponding LV network. The diversified imports/exports (kW) correspond to the LV network aggregated active power (i.e., net demand) divided by the total number of customers connected to the LV network. While the model, in effect, estimates the aggregated active power, this is presented as a diversified value (per customer) in order to understand what the HC estimation might mean across customers. It is important to highlight that a positive value here represents power exports (i.e., demand lower than PV generation hence reverse power flow) and a negative value represents power imports (i.e., demand higher than PV generation).
- **Smart Meter Data.** Corresponds to the actual smart meter measurements and are denoted with a blue circle marker.
- **Model.** Corresponds to the trained regression fit model ($y = ax + c$), that allows estimating the maximum voltage (y) in the LV network for a given value of diversified active power (x).
- **Performance.** Corresponds to the R^2 performance of the trained model. In general, the higher the R^2, the better the model can predict a dependant variable (i.e., maximum voltage) given an independent variable (i.e., aggregated active power). While this metric provides an overall value of the model performance, other metrics (i.e., prediction limits, confidence band) should also be considered to understand the level of estimation accuracy.
- **Prediction limits.** Corresponds to the 99% upper and lower prediction limits of the trained regression model and are denoted with a dashed grey line. In other words, the estimated maximum voltage has a 99% probability to lie between the upper and lower prediction limits.
- **Confidence band.** Corresponds to the 99% confidence level of the estimated maximum voltages. In other words, this band shows there is a 99% confidence that, in average, the maximum voltage considering different scenarios (generation, demand) will be within this band. It is important to note that the smaller the band, the more accurate is the trained model.

- **HC Estimations.** Corresponds to the estimated Hosting Capacity (i.e., diversified active power that leads to a maximum voltage of 1.1p.u). Three values are provided and correspond to:
  - $X_{hc-model}$ The HC directly estimated by the model.
  - $X_{hc-99high}$The HC considering the upper 99% prediction limit.
  - $X_{hc-99low}$ The HC considering the lower 99% prediction limit.

  The $X_{hc-99high}$ and $X_{hc-99low}$ can be considered as the range of possible HC estimations with the former representing a conservative estimation while the latter a more optimistic one. The $X_{hc-model}$, lies between the two.

4.2.2 HC Estimation

For a given LV network the HC estimation model can be used to estimate the hosting capacity for any LV network. Using the demonstration example shown in Figure 4-2, the HC (i.e., $X_{hc-model}$) of this example LV network (i.e., 138 customers) is estimated to be 6.68kW (diversified exports) per customer which corresponds to 921kW (6.68 x 138 customers) of aggregated active power exports. In other words, this LV network is estimated to host up to 921kW of solar PV installed capacity.

From the planning perspective, this number can be used along with available information regarding the already installed solar PV capacity to speed up new PV connection requests. For example, using the estimated HC and the already installed solar PV capacity, the additional cumulative solar PV capacity that can be accommodated in an LV network can be estimated and used to quickly assess new PV installation requests.

4.2.3 Accuracy

While any model can be used to estimate the corresponding HC in an LV network it is important to take into account its estimation accuracy. Considering the demonstration example in Figure 4-2, which is trained using a significantly low number of data points (10 days), it is observed, and as expected, that the R^2 performance is extremely low (47%) showing a very poor estimation accuracy.

While the R^2 provides an overall value of the corresponding performance it is important to look at the model prediction limits and confidence level band, as these metrics allow understanding the level of accuracy of the corresponding estimations. For the same example, the significantly large distance between the prediction limits as well as the large confidence level band show that the estimation accuracy is reducing for values lying further than the smart meter data measurements (i.e., used for the model training). Indeed, the large distance between the model prediction limits is leading to an extremely large (and technically infeasible) HC estimation range i.e., $X_{hc-95high} = 3.33kW$ and $X_{hc-95low} = infeasible\ kW$.

As briefly mentioned in the previous sections, this poor performance is due to the low number of data points used to train the model and hence the confidence level of any prediction is low. Thus, the volume of training data plays an important role to the model's performance (see section below).

4.2.4 Volume of Training Data

To understand the importance the volume of training data has on the model's performance, Figure 4-3 is provided which compares the demonstration example model (i.e., trained on 10 days' worth of data) with two other models trained on 20 and 100 days' worth of data, respectively.

As observed, increasing the volume of training data can increase the overall performance (R^2) of estimation model. Considering the model trained using 20 days' worth of data, the R^2 performance increased from 0.47 to 0.75, while the same number increased to 0.78 when 100 days are used for the corresponding training. While a smaller improvement in R^2 performance is observed between the 20- and 100-days cases, it's important to highlight the performance improvement in terms of the prediction limits and the confidence level band.

For example, considering the prediction limits, a larger volume of training data is helping to reduce the overall distance between the two limits; hence the HC Estimation range (i.e., $X_{hc-95high}$ and $X_{hc-95low}$). To be more exact, and for this case, the use of 100 days' worth of data shows that the distance of between the $X_{hc-95high}$ and $X_{hc-95low}$, reduced by a four-fold compared to the case where 20 days' worth of data are used to train the model. Moreover, as observed, a larger volume of training data can help keep the distance between the two prediction limits almost constant for estimations that lie outside or further to the training data.

More importantly, as Figure 4-3 shows, a larger volume of training data also improves the model's estimation confidence level. While the confidence level of any model is expected to be high (i.e., small band) for estimations that are statistically closer to the training data (i.e., smart meter data), this reduces (i.e., larger band) for any estimations that are outside or further to the training data. The use of larger volume of training data, helps capturing the variance of a larger sample of network conditions (i.e., voltage vs active power), thus increasing the confidence of the model's estimation accuracy.

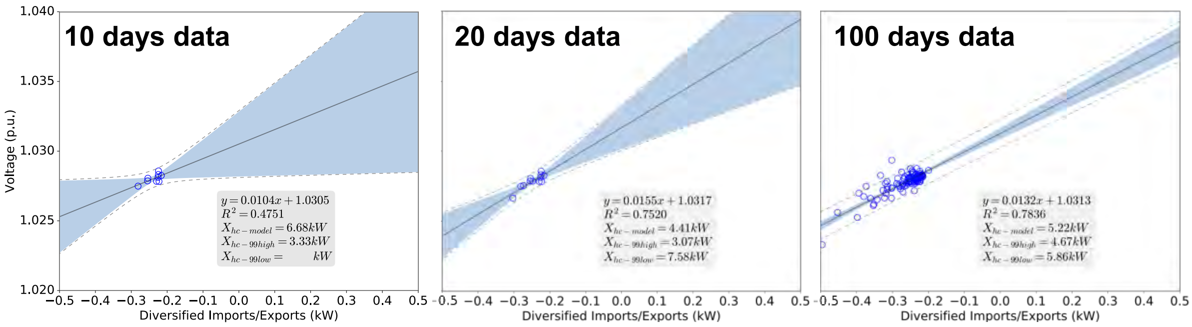


Figure 4-3 HC Estimation Model: Importance of training data volume

4.3 Case Study 1 – Urban HV Feeder U2 (CRE21)

This section presents a case study performed on the urban HV Feeder U2 (CRE21) considering the proposed Smart Meter-Driven hosting capacity estimation methodology. The performance and ability to estimate the HC capacity of LV networks with the proposed methodology is assessed using 5-year hybrid smart meter data produced using the approach detailed in chapter 2. To thoroughly assess the performance of the proposed HC estimation three hybrid smart meter datasets are produced considering three different PV uptake trends through the horizon of 5-years. These are listed below:

a) **Random PV Uptake.** New PV system installations are randomly allocated to customers within the LV networks. A random allocation of PV systems represents a very realistic scenario which is currently seen in practice (i.e., residential PV systems are adopted by customers located at different locations within the network).
b) **Head to End PV Uptake.** New PV systems are allocated first to customers closer to the head of the LV feeders and then moving towards those at the far end. While unlikely, it represents one of the two extreme scenarios. It leads to the highest PV hosting capacity as the effect of voltage rise is in general lower for points closer to the head of the feeder (i.e., smaller impedance, hence smaller voltage drop/rise).

c) **End to Head PV Uptake.** New PV systems are allocated first to customers at the far end of the LV feeders and then moving towards those at the head. While also unlikely, it represents the other extreme scenario. It leads to the lowest PV hosting capacity as the effect of voltage rise is in general higher for points farther from the head of the feeder (i.e., larger impedance, hence larger voltage drop/rise).

Lastly, the effects of controllable elements such as the Zone Substation OLTC are also incorporated to assess the performance of the proposed HC estimation methodology.

While the case study considered all 71 residential LV networks connected in the HV Feeder U2 (CRE21), for demonstration purposes, detailed results are provided only for the LV Network 24, which supplies 138 residential customers through four feeders connected on a 500kVA transformer. The overall performance considering all 71 LV networks is given in Table 4-3.

4.3.1 Random PV Uptake

To assess the performance of the proposed methodology, the estimation model is constructed at different penetration levels of solar PV and the corresponding HC (i.e., amount of aggregated/diversified kW leading to 1.1 of maximum voltage) of the LV network is estimated to understand the extent to which penetration level the proposed HC estimation methodology can provide meaningful results.

Considering all smart meter data for this LV network, on average, the HC found to be 320kW of exports (2.32kW diversified), and this value is used to assess the performance of the HC estimation model. Additionally, for comparison purposes, the aggregated power of the fist voltage violation (i.e., 313kW-aggregated/2.27kW-diversified exports) is considered understand the extent to which the model's prediction limits cover potential outliers (i.e., worst case scenario). These are also shown in Table 4-1.

Table 4-1 LV 24 – Random PV Uptake: HC based on Smart Meter Data

	Diversified kW	Aggregated kW	Penetration (%)
Average HC	2.32	320	60
First Voltage Violation (worst case)	2.27	313	57

Figure 4-4, presents the HC estimation models constructed at different penetration levels (10 to 100%, steps of 10%). For example, for a given X% penetration level, the model is constructed with all smart meter data from day 1 (0%) until the day were X% of customers have a solar PV. This process is trying to emulate real-life case scenarios where the user (e.g., DNSP) can use all available smart meter data (in a given LV network) until the current day (i.e., representing an X% penetration level) to construct the Smart Meter-Driven HC Estimation Model. For clarity and additional information all smart meter data points, shown in Figure 4-4, are colour-coded based on their distance from the LV transformer. This allows understanding where each maximum voltage point is located within the network. The numerical results and performance of each model is given in Table 4-2.

As it can be observed, the magnitude of the daily maximum voltage is increasing with the penetration levels and around 60% of PV penetration level, the total aggregated power is leading to voltages that violate the upper statutory limit (i.e., 1.1 p.u.). Considering the location (i.e., colour of data points) of the maximum voltage , Figure 4-4 shows that this is varying through the days as new PV system are installed in different locations within the network (i.e., random uptake). For example, while at a given day the maximum voltage location could be located at a customer with a PV system in the middle of an LV feeder, a new PV system installation of the same size located at the end of the feeder might lead to larger voltage rise due to the larger electrical distance (i.e., higher impedance between the head of the feeder and the customer). On the other hand, a new larger PV system installation closer to the head of another LV the feeder (connected to the same LV transformer) might lead to higher voltage rise. This effect is in fact more obvious at low penetration levels (below 50%) where the number of customers with PV systems is low, hence the location and size of every new PV system installation can affect the location of the maximum voltage. At larger penetration levels (above 50%), where more and more PV systems are likely to be installed further to the LV transformer, the location of the maximum voltage is shown to be always located closer to the end of the feeders (i.e., red colour).

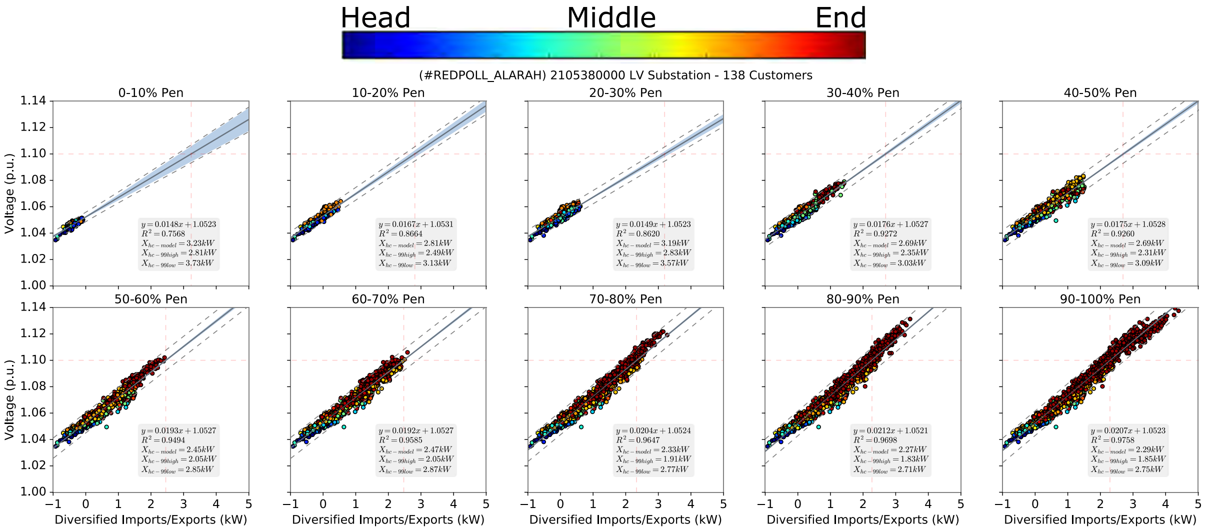


Figure 4-4 LV 24 – Random PV Uptake: HC Estimation at different PV penetrations

Considering the performance, Figure 4-4 and Table 4-2, show that the accuracy error (i.e., percentage difference of estimation from actual) of the models up to 30% of penetration level is high, highlighting that the estimated HC is not very accurate. This low accuracy is partly because the number of PV systems at this penetration level might still not be enough to affect the state of the network (i.e., reverse power flows), hence these interactions not captured in the corresponding model. Nonetheless, while these errors do not go beyond 42%, from 30% onwards, the performance of the models increases significantly, and the accuracy error drops below 18% for all penetrations. This can be explained as the effect of larger number of PV systems in the network is now captured by the models and the estimations are closer to the actual average HC.

Table 4-2 LV 24 – Random PV Uptake: Performance of HC Estimation Model

Penetration Level	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Estimated HC ($X_{hc-model}$) Diversified P (kW)	3.23	2.81	3.19	2.69	2.69	2.45	2.47	2.33	2.27	2.29
Estimation HC Range $[X_{hc-99high}, X_{hc-99low}]$	[2.69-3.91]	[2.39-3.23]	[2.73-3.67]	[2.25-3.13]	[2.19-3.21]	[1.93-2.97]	[1.93-2.99]	[1.79-2.89]	[1.69-2.83]	[1.71-2.87]
Average Actual HC within HC Estimation range?	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES
First voltage violation within HC Estimation range?	NO	NO	NO	YES	YES	YES	YES	YES	YES	YES
Accuracy Error (%) Estimated HC vs Actual HC	42	24	40	18	18	8	9	2	0	1
Estimated HC ($X_{hc-model}$) Aggregated P (kW)	446	388	440	371	371	338	341	322	313	316

It is important also to highlight, that while the model is providing an estimated HC, $X_{hc-model}$, this is accompanied by an estimation range, $[X_{hc-99high}, X_{hc-99low}]$, aiming at capturing potential estimation errors. For example, considering the first voltage violation (i.e., 1.107p.u.) recorded in by the smart meters (occurred at around 57% of PV penetration), the corresponding aggregated power for this case (313kW of exports) was successfully captured within the models' HC estimation range as early as 30% of PV penetration level. Considering the aforementioned, $X_{hc-99high}$ could be considered as a more conservative HC estimation that can be used to cater for the corresponding estimation errors.

To understand the overall performance of the proposed methodology, the average accuracy error of the estimated HC is calculated for each penetration level and for all LV networks that experienced voltage violations (total of 25). Moreover, for each penetration level, the percentage of LV networks for which the HC estimation range successfully covered the actual average HC (calculated using smart meter data) and the first voltage violation, are provided. In general, the average accuracy error is aligned with the results shown in Table 4-2, where the error appears to be high for low penetration levels and reduces with higher penetrations. Overall, it was found that the proposed methodology can provide meaningful estimations from 30% of PV penetration and onwards. In more details, at 30% penetration, the actual HC for 65% of the networks was included in the proposed methodology estimation range, while the same number increased to 78% and 83% for 40 and 50% PV penetrations, respectively. More importantly, a very similar performance is observed for the case of first voltage violation, which shows that the proposed methodology can cover potential outliers (i.e., worst case scenario).

Table 4-3 All LV Networks – Random PV Uptake: Overall Performance

Penetration Level	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Average Accuracy Error (%)	39	40	34	21	18	11	8	5	4	4
Percentage (%) of LV networks where Average HC was included in the HC Estimation range	48	30	65	78	83	96	100	100	100	100
Percentage (%) of LV networks where the first voltage violation was included in the HC Estimation range	44	22	60	72	80	96	100	100	100	100

4.3.2 Head to End PV Uptake

Similar to the section 4.3.1, the analysis was performed considering a head to end PV uptake trend where the new PV systems are allocated first to customers located closer to the head of the LV feeders and lastly to the customers located at the far ends. Given that the effect of voltage rise is in general lower for points closer to the head of the feeder (i.e., smaller impedance, hence smaller voltage drop/rise), the HC capacity of the LV network was expected to be larger than the random PV uptake case. Indeed, as Table 4-4 shows, the average HC capacity was found to be to be 362kW of exports (2.63kW diversified), while the aggregated power of the fist voltage violation was 349kW (2.53kW diversified) exports. In general, such PV uptake trend, compared to the previous, led to almost 20% more customers with a PV system. These numbers of HC are then used to assess the performance of the proposed methodology in the unlikely scenario of such PV uptake trend.

Table 4-4 LV 24 – Head to End PV Uptake: HC based on Smart Meter Data

	Diversified kW	**Aggregated kW**	**Penetration (%)**
Average HC	2.63	362	80

First Voltage Violation (worst case)	2.53	349	73

Like the previous section, Figure 4-5, presents the HC Estimation models constructed at different penetration levels (10 to 100%, steps of 10%) with smart meter data points being colour-coded according to their distance from the LV transformer. The numerical results and performance of each model is given in Table 4-5.

While the magnitude of the daily maximum voltage, in this case, shows to have the same behaviour as the previous case (increasing with the penetration levels), the voltage rise effect from penetration to penetration is significantly smaller. As a result of this, the PV penetration level leading to voltages that violate the upper statutory limit (i.e., 1.1 p.u.) is shifted to 80% (i.e., 20% more than previous case). Considering the location (i.e., data points colour) of the maximum voltage, Figure 4-5 shows that this is starting from a blue colour and slowly transitioning to a red colour at higher penetration levels. This colour transitioning (i.e., maximum voltage location) is the effect of new PV installations happening systematically to those customers located closer to the head of the feeders and slowly moving to the next customers until a 100% penetration is achieved. For example, considering a 0% penetration level, a new PV system installation will happen at the first customer located closest to the head of the feeder. Given that no other customers further to this one has a PV system, the maximum voltage (due to generation) will most likely recorded at that corresponding location (i.e., head of the feeder - blue colour), until the next PV installation, further to this customer, will happen; larger electrical distance from the head, hence voltage rise. Thus, considering this, the location of the maximum voltage is moving from the head (blue) to the end (red) of the feeder as the penetration of PV systems increases.

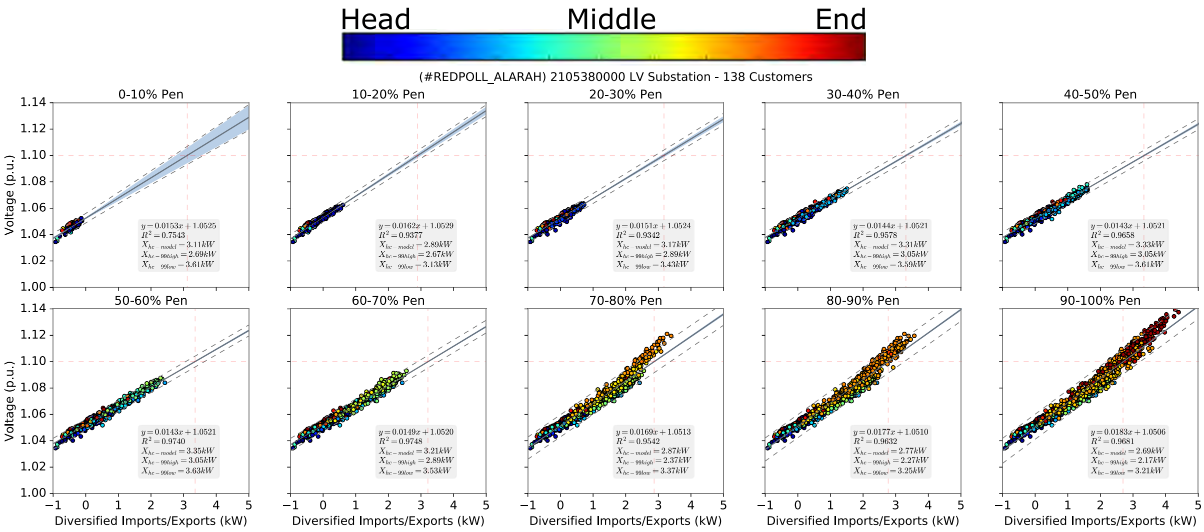


Figure 4-5 LV 24 – Head to End PV Uptake: HC Estimation at different PV penetrations

Considering the performance, Figure 4-5 and Table 4-5, show that the proposed method is overestimating the corresponding HC with an average accuracy error of ~20% (i.e., percentage difference of estimation from actual) until very high penetration levels (>70%) which drops. While a low performance is observed in this case, it should be noted that this mainly because of this unlikely uptake trend scenario where the customers with the largest effect in voltage rise (farthest), are assumed to install a PV system last; hence the HC model cannot capture these effects until high PV penetration levels. For example, up to 70% of penetration level, the smart meter data show a very strong linear behaviour which is perfectly captured by the corresponding HC Estimation models. Up to this penetration level, where ~96 customers (out of 138) are allocated a PV system, the last 42 customers are mostly located at remote areas with significantly larger electrical distance than the other customers already having a PV system. Thus, additional PV system installations at these customers have a larger effect on the voltage rise; hence the corresponding voltage jump from 70% penetration onwards.

Table 4-5 LV 24 – Head to End PV Uptake: Performance of HC Estimation Model

Penetration Level	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Estimated HC ($X_{hc-model}$) Diversified P (kW)	3.11	2.89	3.17	3.31	3.33	3.35	3.21	2.87	2.77	2.69
Estimation HC Range [$X_{hc-99high}$, $X_{hc-99low}$]	[2.59-3.77]	[2.61-3.21]	[2.83-3.51]	[2.97-3.67]	[2.97-3.71]	[2.97-3.71]	[2.81-3.63]	[2.23-3.51]	[2.13-3.41]	[2.03-3.37]
Actual HC within HC Estimation range?	YES	YES	NO	NO	NO	NO	NO	YES	YES	YES
First voltage violation within HC Estimation range?	NO	NO	NO	NO	NO	NO	NO	YES	YES	YES
Accuracy Error (%) Estimated HC vs Actual HC	18	9	20	25	26	27	22	9	5	2
Estimated HC ($X_{hc-model}$) Aggregated P (kW)	429	399	437	457	460	462	443	396	382	371

In terms of the overall performance of the proposed methodology for this particular case, Table 4-6, shows that indeed the average accuracy error, considering all LV networks that experience voltage violations (total of 25), was fairly high (i.e., >30%) until relatively high penetration levels (i.e., 50%). Moreover, the percentage of the LV networks where their average HC and first voltage violation was included in the proposed methodology estimation range was found to be low (less than 30% of networks) for all penetrations until 50%. While a low performance is observed for this case, it should be highlighted that this is based on a PV uptake trend scenario that is unlikely to be seen in practice. Moreover, although an overestimation of the HC is observed, it is important to note that the actual HC capacity of these LV networks was in average ~70%, where at this penetration level the performance of all estimation models was good enough to provide meaningful estimations. It should also be noted that this effect is also due to the relatively higher number of customers (>100) and feeders (>2) in urban LV networks which can lead to significant diversity in terms of the LV feeders' length and number of customers in the same network. This makes the estimation of HC more challenging in such uptake trend. For example, a new PV installation at the end of a long feeder with large number of customers might have a completely different voltage rise effect compared to another with shorter length and lower number of customers. Additional discussion of this effect is provided in section 4.4 which considers a Rural HV Feeder supplying mostly LV networks with low number of customers (<100) and feeders (<2).

Considering the aforementioned, care should be taken when the proposed approach is adopted in such cases.

Table 4-6 All LV Networks – Head to End PV Uptake: Overall Performance

Penetration Level	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Average Accuracy Error (%)	35	37	37	39	34	26	20	11	7	5
Percentage (%) of LV networks where Average HC was included in the HC Estimation range	41	30	22	19	30	56	63	96	100	100
Percentage (%) of LV networks where the first voltage violation was included in the HC Estimation range	26	15	11	11	22	52	70	96	100	100

4.3.3 End to Head PV Uptake

Similar to the section 4.3.1 and 4.3.2, the analysis was performed considering an end to head PV uptake where the new PV systems are allocated first to customers located at the far end of the LV feeders and lastly to the customers located at the head. Given that the effect of voltage rise is in general higher for points further to the head of the feeders (i.e., larger impedance, hence larger voltage drop/rise), the HC capacity of the LV network was expected to be lower than the previously shown PV uptake cases (i.e., random and head to end). Indeed, as Table 4-7 shows, the average HC capacity was found to be to be 248kW of exports (1.8kW diversified), while the aggregated power of the fist voltage violation was 237kW (1.7kW diversified) exports. In general, such PV uptake trend led to almost 10% and 30% less customers with a PV system compared to the random and head to end uptake cases, respectively. These numbers of HC are then used to assess the performance of the proposed methodology in the extreme scenario of such PV uptake trend.

Table 4-7 LV 24 – End to Head PV Uptake: HC based on Smart Meter Data

	Diversified kW	Aggregated kW	Penetration (%)
Average HC	1.8	248	50
First Voltage Violation (worst case)	1.7	237	47

As performed also in the previous sections (4.3.1 and 4.3.2), Figure 4-6, presents the HC Estimation models constructed at different penetration levels (10 to 100%, steps of 10%) with smart meter data points being colour-coded according to their distance from the LV transformer. The numerical results and performance of each model is given in Table 4-8.

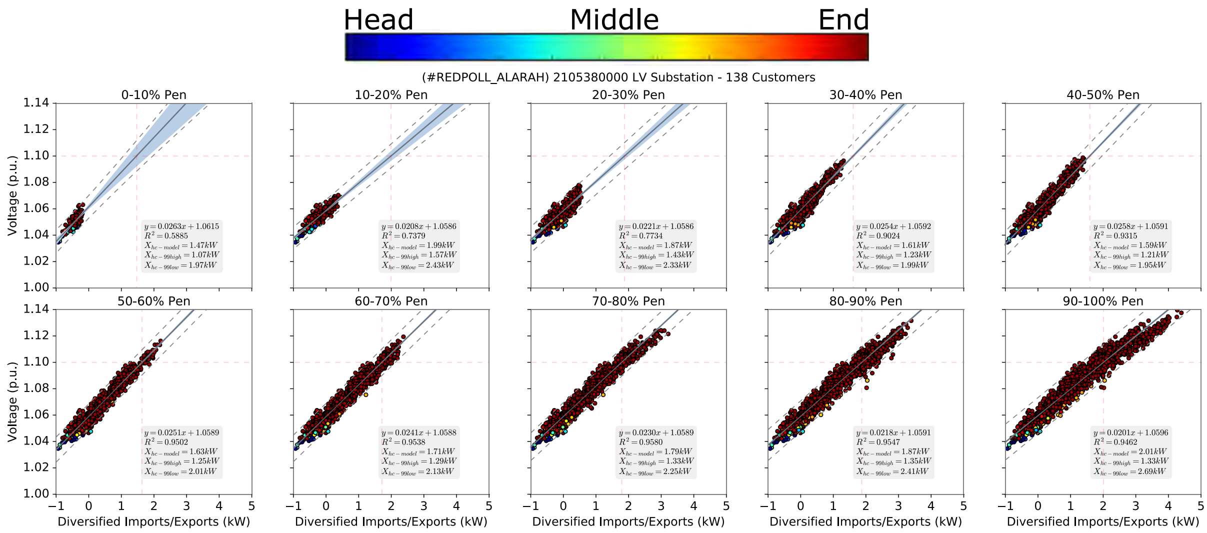


Figure 4-6 LV 24 – End to Head PV Uptake: HC Estimation at different PV penetrations

While the magnitude of the daily maximum voltage, in this case, shows to have the same behaviour as the previous cases (increasing with the penetration levels), the voltage rise effect from penetration to penetration is significantly larger. As a result of this, the PV penetration level leading to voltages that violate the upper statutory limit (i.e., 1.1 p.u.) is reached as early as 50%. Considering the location (i.e., data points colour) of the maximum voltage, Figure 4-6 shows that this is always located at the end of the feeders (red colour) regardless the penetration level. This is indeed expected given that the farthest customers, which are more likely to experience a higher voltage rise than any other customer in the network, are those that first install a PV system. Thus, since all other new PV system installations will always have a smaller electrical distance, than the ones already installed (at the farthest points), the location of the maximum voltage point is expected to be the same regardless the new installations.

Considering the performance, Figure 4-6 and Table 4-8, show that the proposed method achieves a low error regardless the penetration level. To be more exact, the error is always kept below 17% and significantly reduces to 10% as early as 30% of penetration level and drops even further for higher penetrations. This performance can be explained since the HC estimation model is capturing from very early penetration levels, the effects of those customers that are contributing the most to the voltage rise. Moreover, it should be noted that the first voltage violation (i.e., 1.104p.u.) recorded in by the smart meters (occurred at around 47% of PV penetration), the corresponding aggregated power for this case (237kW of exports) was successfully captured within the models' HC estimation range as early as 10% of PV penetration level.

Table 4-8 LV 24 – End to Head PV Uptake: Performance of HC Estimation Model

Penetration Level	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Estimated HC ($X_{hc-model}$) Diversified P (kW)	1.47	1.99	1.87	1.61	1.59	1.63	1.71	1.79	1.87	2.01
Estimation HC Range $[X_{hc-99high}, X_{hc-99low}]$	[0.95-2.15]	[1.45-2.59]	[1.29-2.49]	[1.11-2.11]	[1.11-2.07]	[1.15-2.13]	[1.17-2.25]	[1.19-2.37]	[1.19-2.57]	[1.13-2.89]
Actual HC within HC Estimation range?	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
First voltage violation within HC Estimation range?	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Accuracy Error (%) Estimated HC vs Actual HC	14	17	10	5	6	4	1	5	10	18
Estimated HC ($X_{hc-model}$) Aggregated P (kW)	203	275	258	222	219	225	236	247	258	277

Considering the overall performance, the average accuracy error is aligned with the results shown in section 4.3.1, where the error appears to be higher for low penetration levels and reduces with higher penetrations. This, as previously discussed, is because a larger number of PV systems in the network helps capture their effects in the estimation model; hence increase the corresponding accuracy. Overall, it was found that the proposed methodology can provide meaningful estimations from 30% of PV penetration and onwards. In more details, at 30% penetration, the actual average HC for 81% of the LV networks was included in the proposed methodology estimation range, while the same number increased to 96% and 100% for 40 and 50% PV penetrations, respectively. More importantly, a very similar performance is observed for the case of first voltage violation, which shows that the proposed methodology can cover potential outliers (i.e., worst case scenario).

Table 4-9 All LV Networks – End to Head PV Uptake: Overall Performance

Penetration Level	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Average Accuracy Error (%)	38	33	27	19	14	10	7	6	5	8
Percentage (%) of LV networks where Average HC was included in the HC Estimation range	52	63	81	96	100	100	100	100	100	100

Percentage (%) of LV networks where the first voltage violation was included in the HC Estimation range	63	59	67	89	100	100	100	100	100	100

4.3.4 Effects of Zone Substation OLTC

The last three sections assess the performance of the proposed methodology under three different PV uptake cases, one realistic (i.e., random) and two extreme (i.e., head-to-end and end-to-head). All cases, however, were performed considering the voltage at the head of the HV feeder to be constant at 22 kV (1.0 pu) which corresponds to the voltage target setting used by the on-load tap changer (OLTC) at the zone substation. While the latter is considered a valid and realistic assumption, it is worth understanding the effect that the OLTC actions at the zone substation might have in the performance of the proposed methodology.

Thus, in this section, the case study presented in section 4.3.1 (random PV uptake) is repeated while also incorporating the corresponding actions of the zone substation OLTC. For this, 30-min resolution voltage measurements recorded for the year 2016 at the head of the corresponding HV feeder where provided by AusNet Services. These were then used in the power flow analyses to incorporate the corresponding OLTC effects by varying the voltage at the head of the HV feeder according to the provided data. While this assumption is adopted, it should be noted that is used here for demonstration purposes only. In practice, the OLTC zone substation actions are based on the loading conditions of multiple connected HV feeders (i.e., 6 in total), which are not modelled in this report; only one of these feeders (Urban HV Feeder U2) is considered. Also, the corresponding provided measurements do not consider the effects of simultaneously high penetration levels of PV systems (in all connected HV feeders) might have on the operation of the zone substation OLTC (i.e., reverse power flows, hence higher voltage at the HV head forcing the OLTC to reduce tap positions).

Table 4-10 shows, the average HC capacity is slightly reduced (i.e., 289kW) compared to the case shown in section 4.3.1 (i.e., 320kW) and this is due to the fact that the OLTC actions are in general pushing the voltage slightly higher to cater for voltage drop issues. Interestingly, it is important to highlight that because of this, such actions might lead to cases where the voltage is leading to voltage violations at even lower penetration levels. This is indeed, shown in Table 4-10 where first voltage violation for this case was recorded at 18% of PV penetration level; a comparably lower value compared to all previous sections. This can also be visualised in Figure 4-7 where the maximum voltage in some of the days was significantly higher than the rest.

Table 4-10 LV 24 – Random PV Uptake with OLTC: HC based on Smart Meter Data

	Diversified kW	Aggregated kW	Penetration (%)
Average HC	2.10	289	58
First Voltage Violation (worst case)	0.60	82	18

In general, Figure 4-7 shows that the daily maximum voltage behaviour is aligned with the one shown in section 4.3.1 however, the effect of OLTC is resulting to a larger spread in terms of maximum voltages for the same imports/exports. Due to the latter, the confidence level of the HC estimation models shows to be low for smaller penetration level and progressively increasing for larger. In terms of prediction limits, the distance was found to be considerably larger when compared with section 4.3.1, regardless the penetration level. This means, that the effect of the OLTC is affecting the performance of the proposed methodology.

Considering the numerical results provided in Table 4-11Table 4-8, a higher accuracy error is noticed and the estimation range is larger for all penetration levels. While the performance of the HC estimation models is reduced because of the OLTC effects it is important to note that the average HC is successfully captured by the HC estimation range from as early as 10% PV penetration level. On the other hand, the proposed methodology fails to capture the first voltage violation within the models' HC estimation range, regardless the penetration level.

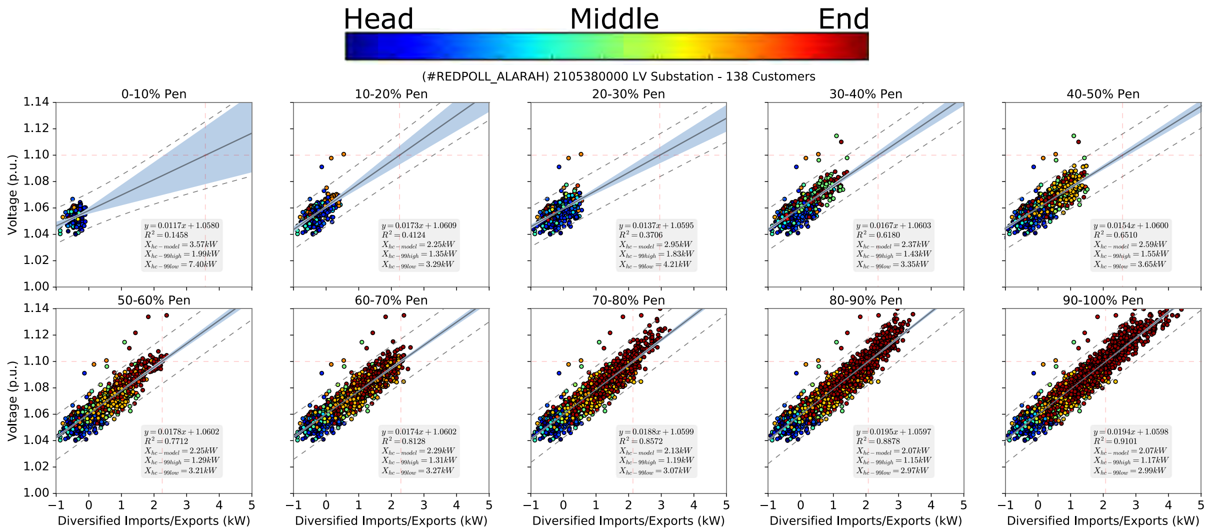


Figure 4-7 LV 24 – Random PV Uptake with OLTC: HC Estimation at different PV penetrations

Table 4-11 LV 24 – Random PV Uptake with OLTC: Performance of HC Estimation Model

Penetration Level	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Estimated HC ($X_{hc-model}$) Diversified P (kW)	3.57	2.25	2.95	2.37	2.59	2.25	2.29	2.13	2.07	2.07
Estimation HC Range [$X_{hc-99high}$, $X_{hc-99low}$]	[1.65-10.0]	[1.11-3.63]	[1.49-4.63]	[1.15-3.63]	[1.25-3.97]	[1.01-3.49]	[1.03-3.55]	[0.91-3.35]	[0.89-3.25]	[0.89-3.25]
Actual HC within HC Estimation range?	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
First voltage violation within HC Estimation range?	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
Accuracy Error (%) Estimated HC vs Actual HC	70	7	40	13	23	7	9	1	1	1
Estimated HC ($X_{hc-model}$) Aggregated P (kW)	493	310	407	327	357	310	316	294	286	286

In terms of the overall performance, while the average accuracy errors were found to be higher than section 4.3.1, the proposed methodology can provide meaningful estimations from 10% of PV penetration and onwards. In more details, at 10% penetration, the actual HC for 79% of the LV networks was included in the proposed methodology estimation range, while the same number increased to 86% and 96% for 20 and 40% PV penetrations, respectively. However, the proposed methodology's estimation range ability to cover potential outliers (i.e., worst case scenario) is significantly reduced by the OLTC control actions.

Once again, it is important to highlight that while this analysis shows that the zone substation OLTC actions are expected to affect the performance of the proposed methodology, this was based on data that might not represent the actual control actions; hence, it might underestimate or overestimate the corresponding performance effects.

Table 4-12 All LV Networks – Random PV Uptake with OLTC: Overall Performance

Penetration Level	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Average Accuracy Error (%)	69	36	39	26	25	18	16	12	10	10
Percentage (%) of LV networks where Average HC was included in the HC Estimation range	79	86	86	96	96	96	96	100	100	100
Percentage (%) of LV networks where the first voltage violation was included in the HC Estimation range	11	25	18	25	25	32	32	32	32	36

4.4 Case Study 2 – Rural HV Feeder R1 (SMR8)

This section presents a case study performed on the rural HV Feeder R1 (SMR8) considering the proposed Smart Meter-Driven hosting capacity estimation methodology. It should be noted that given the nature of rural HV feeders (i.e., long, high number of LV networks), the corresponding connected LV networks are considerably smaller compared to those in urban HV feeders consisting of lower number of customers (usually less than 100) and lower number of LV feeders (up to 2). Such analysis will help understand the performance of the proposed HC capacity estimation methodology on feeders with significantly different characteristics. As such, the performance and ability to estimate the HC capacity of LV networks with the proposed methodology is assessed using 5-year hybrid smart meter data produced using the approach detailed in chapter 2. To thoroughly assess the performance of the proposed HC estimation methodology, three hybrid smart meter datasets are produced considering three different PV uptake trends through the duration of 5-years analysis. These are listed below:

d) **Random PV Uptake.** New PV system installations are randomly allocated to customers within the LV networks. A random allocation of PV systems represents a very realistic scenario which is currently seen in practice (i.e., residential PV systems are adopted by customers located at different locations within the network).
e) **Head to End PV Uptake.** New PV systems are allocated first to customers closer to the head of the LV feeders and then moving towards those at the far end. While unlikely, it represents one of the two extreme scenarios. It leads to the highest PV hosting capacity as the effect of voltage rise is in general lower for points closer to the head of the feeder (i.e., smaller impedance, hence smaller voltage drop/rise).
f) **End to Head PV Uptake.** New PV systems are allocated first to customers at the far end of the LV feeders and then moving towards those at the head. While also unlikely, it represents the other extreme scenario. It leads to the lowest PV hosting capacity as the effect of voltage rise is in general higher for points farther from the head of the feeder (i.e., larger impedance, hence larger voltage drop/rise).

While the case study considered all 705 residential LV networks connected in the HV Feeder R2 (SMR8), for demonstration purposes, detailed results are provided only for the LV Network 236, which supplies 76 residential customers (connected on two feeders) through a 200kVA transformer. The overall performance considering all 705 LV networks is given in Table 4-3.

4.4.1 Random PV Uptake

Considering all smart meter data for this LV network, on average, the HC was found to be 109kW of exports (1.44kW diversified), and this value is used to assess the performance of the HC estimation model. For comparison purposes, the aggregated power of the fist voltage violation (i.e., 105kW-aggregated/1.39kW-diversified exports) is considered to also understand the extent to which the model's prediction limits cover potential outliers (i.e., worst case scenario). These are also shown in Table 4-13.

Table 4-13 LV 236 – Random PV Uptake: HC based on Smart Meter Data

	Diversified kW	Aggregated kW	Penetration (%)
Average HC	1.44	109	36
First Voltage Violation (worst case)	1.39	105	35

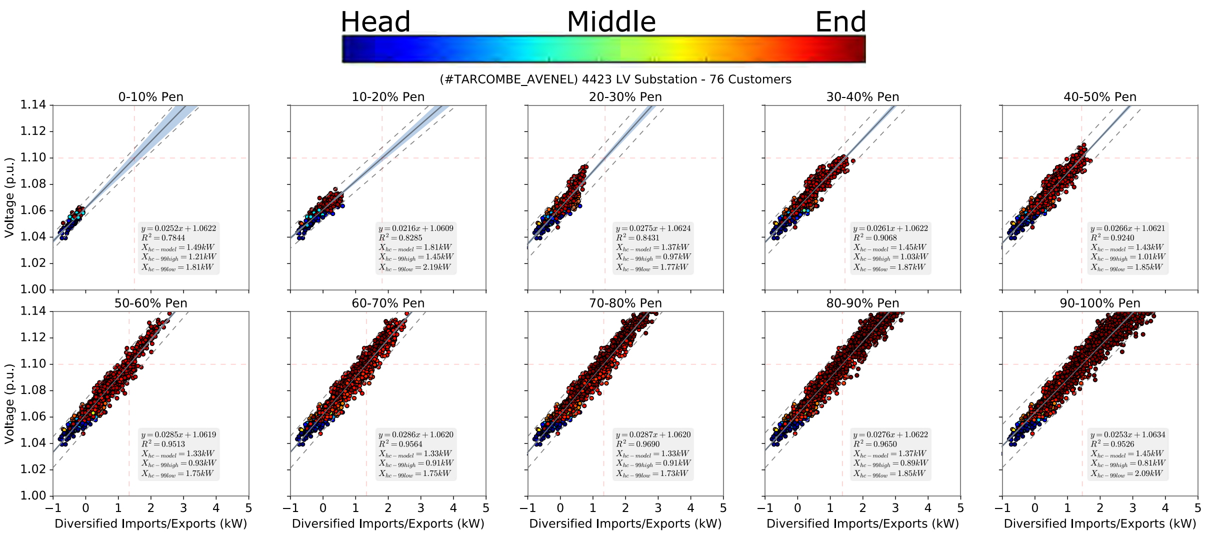


Figure 4-8 LV 236 – Random PV Uptake: HC Estimation at different PV penetrations

Considering the performance, Figure 4-8 and Table 4-14, show that the accuracy error (i.e., percentage difference of estimation from actual) of the models, regardless the penetration level, is very low (<7%), highlighting that the estimated HC has a much better accuracy compared to the cases shown in the previous case study (urban feeder). The higher accuracy of HC estimations at earlier PV penetrations can be explained due to the lower number of customers and feeders (up to 2) in rural LV networks. This means that the impacts of PV installations in a given LV network will evolve consistently, i.e., higher voltages will be seen with more PV installations (which is not the case in urban LV networks with multiple feeders due to the diversity in length and customer numbers). This consistency allows the HC estimation model to capture the effects more accurately and at earlier PV penetrations. Moreover, considering the first voltage violation recorded in by the smart meters (occurred at around 35% of PV penetration), the corresponding aggregated power for this case (105kW of exports) was successfully captured within the models' HC estimation range as early as 10% of PV penetration.

Table 4-14 LV 236 – Random PV Uptake: Performance of HC Estimation Model

Penetration Level	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Estimated HC ($X_{hc-model}$) Diversified P (kW)	1.49	1.81	1.37	1.45	1.43	1.33	1.33	1.33	1.37	1.45

Estimation HC Range $[X_{hc-99high}, X_{hc-99low}]$	[1.13-1.91]	[1.35-2.31]	[0.85-1.91]	[0.89-1.99]	[0.87-1.97]	[0.81-1.87]	[0.79-1.87]	[0.79-1.85]	[0.75-1.99]	[0.61-2.29]
Actual HC within HC Estimation range?	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
First voltage violation within HC Estimation range?	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Accuracy Error (%) Estimated HC vs Actual HC	3	25	4	0	0	7	7	7	4	0
Estimated HC ($X_{hc-model}$) *Aggregated P (kW)*	113	138	104	110	109	101	101	101	104	110

To understand the overall performance of the proposed methodology, the average accuracy error of the estimated HC is calculated for each penetration level and for all LV networks (total of 41) that experienced voltage violations. Moreover, for each penetration level, the percentage of LV networks for which the HC estimation range successfully covered the average HC (calculated using smart meter data) and the first voltage violation, are provided. In general, the average accuracy error appears to be higher for low penetration levels and reduces with higher penetrations. Overall, it was found that the proposed methodology can provide meaningful estimations from as early as 30% of PV penetration. In more details, at 30% penetration, the actual HC for 81% of the LV networks was included in the proposed methodology estimation range, while the same number increased 83% and 86% for 40 and 50% PV penetrations, respectively. More importantly, a very similar performance is observed for the case of first voltage violation, which shows that the proposed methodology can cover potential outliers (i.e., worst case scenario).

Table 4-15 All LV Networks – Random PV Uptake: Overall Performance

Penetration Level	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
Average Accuracy Error (%)	43	44	35	26	21	17	12	9	7	7
Percentage (%) of LV networks where Average HC was included in the HC Estimation range	88	69	76	83	86	95	100	100	100	100
Percentage (%) of LV networks where the first voltage violation was included in the HC Estimation range	86	71	76	86	93	95	100	100	100	100

4.4.2 Head to End PV Uptake

Similar to the section 4.4.1, the analysis was performed considering a head to end PV uptake where the new PV systems are allocated first to the customers located closer to the head of the LV feeders and lastly to the customers located at the far ends. Given that the effect of voltage rise is in general lower for points closer to the head of the feeder (i.e., smaller impedance, hence smaller voltage drop/rise), the HC capacity of the LV network was expected to be larger than the random PV uptake case. Indeed, as Table 4-16 shows, the average HC capacity was found to be to be 196kW of exports (2.58kW diversified), while the aggregated power of the fist voltage violation was 177kW (2.33kW diversified) exports. In general, such PV uptake trend, compared to the previous, led to almost 30% more customers with a PV system. These numbers of HC are then used to assess the performance of the proposed methodology in the extreme scenario of such PV uptake trend.

Table 4-16 LV 236 – Head to End PV Uptake: HC based on Smart Meter Data

	Diversified kW	**Aggregated kW**	**Penetration (%)**
Average HC	2.58	196	66

First Voltage Violation (worst case)	2.33	177	63

While the magnitude of the daily maximum voltage, in this case, shows to have the same behaviour as the previous case (increasing with the penetration levels), the voltage rise effect from penetration to penetration is, as expected, significantly smaller. As a result of this, the PV penetration level leading to voltages that violate the upper statutory limit (i.e., 1.1 p.u.) is shifted to 66% (i.e., 30% more than previous case). Considering the location (i.e., data points colour) of the maximum voltage, shows that this is starting from the head of the feeder (blue data points) and slowly transitioning to the end of the feeder (red data points) at higher penetration levels. This as explained in the previous case study is the effect of new PV installations happening systematically to those customers located closer to the head of the feeders and slowly moving to the next customers until a 100% penetration is achieved.

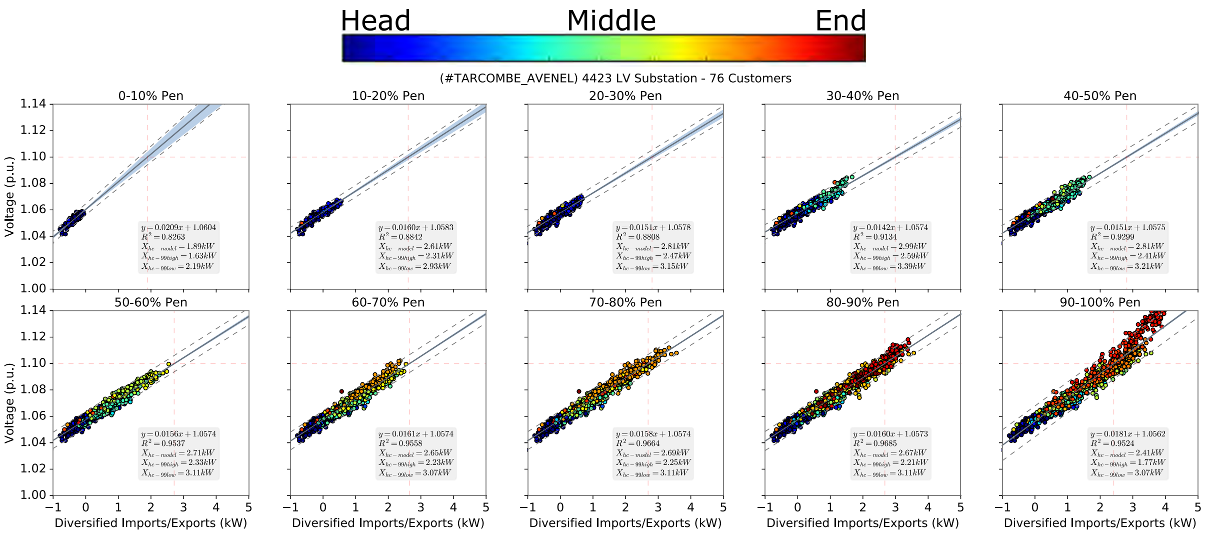


Figure 4-9 LV 236 – Head to End PV Uptake: HC Estimation at different PV penetrations

Considering the performance, results show that the proposed method in this case is providing a very good performance compared to the one in section 4.3.2. In more details, the accuracy error is considerably lower for all penetration levels (average of ~8%). Moreover, it is shown that the proposed methodology estimation range can provide accurate and meaningful results from as low as 20% of PV penetration level; the average HC is successfully included in the estimation range.

Table 4-17 LV 236 – Head to End PV Uptake: Performance of HC Estimation Model

Penetration Level	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Estimated HC ($X_{hc-model}$) Diversified P (kW)	1.89	2.61	2.81	2.99	2.81	2.71	2.65	2.69	2.67	2.41
Estimation HC Range $[X_{hc-99high}, X_{hc-99low}]$	[1.55-2.29]	[2.23-3.01]	[2.37-3.25]	[2.47-3.51]	[2.31-3.33]	[2.21-3.23]	[2.11-3.19]	[2.13-3.25]	[2.09-3.25]	[1.59-3.25]
Actual HC within HC Estimation range?	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES
First voltage violation within HC Estimation range?	NO	YES	NO	NO	YES	YES	YES	YES	YES	YES
Accuracy Error (%) Estimated HC vs Actual HC	26	1	9	16	9	5	2	4	3	6
Estimated HC ($X_{hc-model}$) Aggregated P (kW)	144	198	214	227	214	206	201	204	203	183

In terms of the overall performance of the proposed methodology for this particular case, Table 4-18, shows that the average accuracy error, considering all LV networks that experience voltage violations (total of 41), was very similar with the one found in section 4.4.1. Moreover, the percentage of the LV networks where their average HC and first voltage violation was included in the proposed methodology estimation range was found to be high (average of 86%) for all penetration levels. More importantly, a very similar performance is observed for the case of first voltage violation, which shows that the proposed methodology can cover potential outliers (i.e., worst case scenarios).

Compared to the section 4.3.2, the comparably higher performance found in this case can be explained due to the relatively lower number of LV feeders (up to 2) and customers (usually less than 100) found in rural LV networks. This reduces the effects of having different lengths of feeders and number of customers; hence leading to different voltage rise effect compared to other feeders in the same network.

Table 4-18 All LV Networks – Head to End PV Uptake: Overall Performance

Penetration Level	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Average Accuracy Error (%)	45	57	35	24	22	21	21	17	13	8
Percentage (%) of LV networks where Average HC was included in the HC Estimation range	86	62	76	81	86	88	88	95	100	100
Percentage (%) of LV networks where the first voltage violation was included in the HC Estimation range	79	64	71	81	86	88	95	100	100	100

4.4.3 End to Head PV Uptake

Considering an end to head PV uptake tend, Table 4-7, shows that the average HC capacity of the LV network, was 75kW of exports (0.99kW diversified) with the fist voltage violation recorded at 22kW (0.29kW diversified) of exports. These results show that such uptake trend can potentially lead to 16% and 46% less customers with a PV system compared to the random and head to end uptake trends, respectively. These numbers of HC are then used to assess the performance of the proposed methodology in the extreme scenario of such PV uptake trend.

Table 4-19 LV 236 – End to Head PV Uptake: HC based on Smart Meter Data

	Diversified kW	Aggregated kW	Penetration (%)
Average HC	0.99	75	20
First Voltage Violation (worst case)	0.29	22	15

While the magnitude of the daily maximum voltage, in this case, shows to have the same behaviour as the previous cases (increasing with the penetration levels), the voltage rise effect from penetration to penetration is significantly larger. As a result of this, the PV penetration level leading to voltages that violate the upper statutory limit (i.e., 1.1 p.u.) is reached as early as 20%.

Considering the performance, Figure 4-10 and Table 4-21, show that the error can be relatively high (average of 36%) for low penetration levels and this reduces with higher penetrations. While errors exist, it is important to highlight that the actual average HC is successfully captured in the model's estimation range as early as 10% penetration. More importantly, the first voltage violation (i.e., 1.104p.u.) recorded by the smart meters (occurred at around 15% of penetration), was also successfully captured within the models' HC estimation range as early as 10% penetration. These results highlight the effectiveness of the model to provide meaningful HC estimations from early penetration levels.

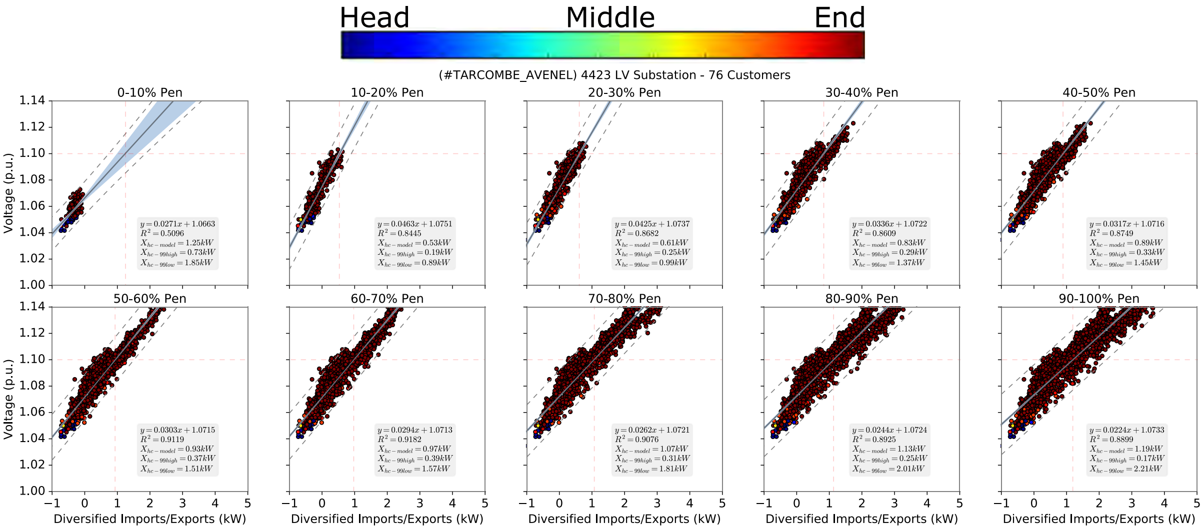


Figure 4-10 LV 236 – End to Head PV Uptake: HC Estimation at different PV penetrations

Table 4-20 LV 236 – End to Head PV Uptake: Performance of HC Estimation Model

Penetration Level	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Estimated HC ($X_{hc-model}$) Diversified P (kW)	1.25	0.53	0.61	0.83	0.89	0.93	0.97	1.07	1.13	1.19
Estimation HC Range $[X_{hc-99high}, X_{hc-99low}]$	[0.61-2.09]	[0.07-1.01]	[0.13-1.11]	[0.13-1.53]	[0.17-1.63]	[0.19-1.67]	[0.21-1.75]	[0.09-2.03]	[0.01-2.27]	[0.13-2.51]
Actual HC within HC Estimation range?	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
First voltage violation within HC Estimation range?	NO	YES	YES	YES	YES	YES	YES	YES	YES	YES
Accuracy Error (%) Estimated HC vs Actual HC	26	46	38	15	9	5	1	8	14	20
Estimated HC ($X_{hc-model}$) Aggregated P (kW)	95	40	46	63	68	71	74	81	86	90

Considering the overall performance, the average accuracy error is aligned with the results shown in previous sections, where the error appears to be higher for low penetration levels and reduces with higher penetrations. Overall, it was found that the proposed methodology can provide meaningful estimations from 10% of PV penetration and onwards. In more details, at 10% penetration, the actual HC for 62% of the LV networks was included in the proposed methodology estimation range, while the same number increased to 81% and 83% for 20 and 30% PV penetrations, respectively. More importantly, a very similar performance is observed for the case of first voltage violation, which shows that the proposed methodology can cover potential outliers (i.e., worst case scenario).

Table 4-21 All LV Networks – End to Head PV Uptake: Overall Performance

Penetration Level	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Average Accuracy Error (%)	79	30	30	24	21	15	11	7	7	9

Percentage (%) of LV networks where Average HC was included in the HC Estimation range	62	81	83	85	88	95	100	100	98	100
Percentage (%) of LV networks where the first voltage violation was included in the HC Estimation range	45	88	86	90	90	98	100	100	98	100

5 Conclusions

This document corresponds to *"Deliverable 2: Innovative Analytical Techniques"* part of the project Advanced Planning of PV-Rich Distribution Networks funded by the Australian Renewable Energy Agency (ARENA) and led by the University of Melbourne in collaboration with AusNet Services. It focuses on the methodology and assessment of a smart meter-driven analytical technique proposed by The University of Melbourne to estimate PV hosting capacity in distribution networks; two significantly different HV feeders, urban and rural, are considered.

Chapter 1 introduced the current state of residential PV system installations in Australia and the technical and operation challenges the widespread adoption of these might bring to the Distribution Network Service Providers (DNSPs).

Chapter 2 provided some additional modelling aspects and assumptions are considered for the HV-LV feeders presented in Deliverable 1. These were based on updated information and data provided by AusNet Services and corresponded to:

- An updated model of the Urban HV Feeder U2 (CRE21). The updates focus on the total number of residential and non-residential LV networks as well as the total number of residential and non-residential customers connected in each LV network.
- An updated operation of the HV capacitors. All HV capacitors located in the modelled HV feeders are now modelled with voltage-based operation (instead of time-based) so that their operation is aligned with the current practice.

Chapter 3 presented and described a small volume of smart meter data provided by AusNet Services and correspond to 2 days' worth of 5-min resolution encrypted and anonymised data from ~3000 residential customers for the Urban HV Feeder U2 (CRE21). While the provided smart meter, data contained enough measurements (P, Q, V) from customers with PV system installations (20% of customers with PV systems), it was not possible to extract meaningful correlations between the PV penetration and its effects (i.e., voltage rise). This can only be captured by historical data that covers the evolution of PV penetration in time, however, given the significant challenges related to data privacy and confidentiality issues the facilitation of additional multiple days of historical smart meter data to The University of Melbourne was not possible.

To tackle the aforementioned challenges and help develop and validate those analytical techniques that can be used to estimate the Hosting Capacity of LV networks in a given HV feeder, a methodology is proposed in Chapter 3 to produce a large volume of smart meter data (referred here as hybrid smart meter data) for each customer in the modelled HV feeders. These smart meter data consider a realistic progressive adoption of PV systems (i.e., 0 to 100% of customers with PV, in a given HV feeder) through a horizon of 5-years. For this actual anonymised demand (P, Q) and irradiance profiles from a previous project "AusNet Mini Grid Clusters" were used to run unbalanced, 30-min resolution, time-series, three-phase four-wire power flows for multiple days to extract customer voltages, V. In total, the database of the hybrid smart meter data (P, Q, V) produced for each HV-LV Feeder consists of more than 1 billion data points (>3Gb).

With the corresponding hybrid smart meter data at hand and leveraging statistical techniques, daily correlations between the data were identified that hinted the direction towards the analytical approach to be adopted. For each LV network, a very strong linear correlation was found between the maximum voltage on a given day and the corresponding sum of all smart meter active powers (P, which can be negative due to PV systems). These two features were used as inputs to the proposed Smart Meter-Driven PV Hosting Capacity Estimation methodology.

Chapter 4 presented the proposed analytical technique that makes use of smart meter data to construct a statistical regression model for each LV network, in a given HV feeder, and estimate its corresponding PV hosting capacity. The performance of the proposed hosting capacity estimation (HC estimation) methodology was assessed under three different PV system uptake trends, as well as considering the effects of network controllable elements such as the zone substation OLTC.

The proposed Smart Meter-Driven Hosting Capacity Estimation is based on a simple, yet practical, machine learning algorithm, a methodology is proposed to produce a regression model to estimate the PV hosting capacity in any given LV network using smart meter data. The main steps of the methodology, as if implemented by a DNSP, are presented below.

- Smart meter database. For a given number of days (ideally covering most of the evolution of PV penetrations to date), the daily smart meter data (i.e., P, Q, V) from all customers in a given LV network are extracted from the smart meter database.
- Data Processing. The smart meter data are analysed and cleaned from missing and inconsistent values. Then, the maximum voltage recorded for each day is identified and the corresponding (same timestamp) active powers are added up. Finally, a new dataset is produced containing the maximum voltage and the corresponding aggregated power for each day.
- HC Estimation Model. The new dataset is used to train a supervised (i.e., gradient decent) univariate regression model which corresponds to the HC estimation model for the analysed LV network.

The proposed Smart Meter-Driven Hosting Capacity Estimation model, in effect, estimates the aggregated active power (that can be negative due to PV systems) that can lead to voltages outside a pre-determined upper limit (e.g., 1.1 p.u.). This value, in turn, can be used to calculate the additional PV capacity that can be hosted by the LV network. Nonetheless, to understand what the HC Estimation might mean across customers, the estimated aggregated active power is presented as the diversified active power per customer. The latter also includes prediction limits to cater for uncertainties.

In terms of the accuracy of the model, it is important to highlight that the volume of smart meter data used to produce the HC estimation model plays an important role. More data helps to capture the variance of a larger sample of network conditions (i.e., voltage vs active power), thus increasing the model's estimation accuracy.

The performance of the proposed HC estimation methodology was demonstrated and thoroughly assessed on two significantly different HV-LV feeders (urban and a rural) and the assessment considered the following three different PV uptake trends through a horizon of 5 years.

- Random PV Uptake. New PV system installations are randomly allocated to customers within the LV networks. A random allocation of PV systems represents a very realistic scenario which is currently seen in practice (i.e., residential PV systems are adopted by customers located at different locations within the network).
- Head to End PV Uptake. New PV systems are allocated first to customers closer to the head of the LV feeders and then moving towards those at the far end. While unlikely, it represents one of the two extreme scenarios. It leads to the highest PV hosting capacity as the effect of voltage rise is in general lower for points closer to the head of the feeder (i.e., smaller impedance, hence smaller voltage drop/rise).
- End to Head PV Uptake. New PV systems are allocated first to customers at the far end of the LV feeders and then moving towards those at the head. While also unlikely, it represents the other extreme scenario. It leads to the lowest PV hosting capacity as the effect of voltage rise is in general higher for points farther from the head of the feeder (i.e., larger impedance, hence larger voltage drop/rise).

Considering the Urban Feeder U2 (CRE21), it was found that the proposed methodology can provide meaningful and adequate HC estimations for this and similar urban feeders. In this case, such estimations were achieved from as early as 30% of PV penetration for the Random PV and End to Head PV uptake trends.

For the Head to End uptake trend, it was found that early PV penetrations did not result in significant impacts, resulting in slight HC overestimations. This is primarily due the fact that customers expected to affect voltage rise the most (i.e., farthest customers) are the last installing a PV system; hence, the HC model cannot capture these effects until high PV penetration levels (i.e., >60%). Moreover, due to the relatively higher number of customers (>100) and feeders (>2) in urban LV networks, significant diversity can exist in terms of the LV feeders' length and number of customers in the same network. This makes the estimation of HC more challenging in such uptake trend. For example, a new PV installation at the end of a long feeder with many customers might have a completely different voltage rise effect compared to another with shorter length and lower number of customers.

A further analysis using SCADA data from 2016 to represent the zone substation's OLTC actions was carried out. Although these voltage changes might not capture how the CRE21 OLTC would in reality act with the different PV penetrations, it was found that it can slightly reduce the accuracy of the HC estimations. Furthermore, because of a higher number of outliers (voltage spikes), the ability of the HC estimation model to include them in the prediction limits reduces.

As for the Rural HV Feeder R1 (SMR8), it was found that the proposed methodology can have a much better performance in this and similar rural feeders as it is able to provide meaningful and adequate HC estimations from much earlier PV penetrations regardless the PV uptake trend. In this case, such estimations were achieved with as little as 10% PV penetration.

The higher accuracy of HC estimations at earlier PV penetrations can be explained due to the lower number of customers and feeders (up to 2) in rural LV networks. This means that the impacts of PV installations in a given LV network will evolve consistently, i.e., higher voltages will be seen with more PV installations (which is not the case in urban LV networks with multiple feeders due to the diversity in length and customer numbers). This consistency of allows the HC estimation model to capture the effects more accurately and at earlier PV penetrations.

6 Next Steps

The next steps to be carried out by The University of Melbourne for the "Advanced Planning of PV-Rich Distribution Networks" project include:

Task 3 Traditional Solutions

This task will investigate the use of traditional solutions such as network augmentation and/or change of off-load tap changer positions to increase hosting capacity. An advanced stochastic augmentation assessment methodology will be developed that significantly departs from traditional augmentation analyses (commonly based on deterministic, worst-case scenarios, and often ignoring aggregated congestion issues in HV networks due to solar PV in LV networks). The methodology will be demonstrated by quantifying the level of network augmentation required to meet PV penetrations expected in the horizon(s) of interest (e.g., 5 years).

Deliverable 4: Traditional Solutions (Delivery Date: 10th February 2020)
Synopsis: A technical report presenting the methodology and initial findings corresponding to the use of traditional solutions to increase PV hosting capacity.

Deliverable 5: Workshop (Delivery Date: February 2020)
Synopsis: A workshop presenting the key findings from the first year of the project.

Deliverable 6: International Conference (Delivery Date: February 2020)
Synopsis: Findings of from the first year of the project are expected to be presented as academic papers to the international community at top-class conferences.

7 References

[1] A. T. Procopiou and L. F. Ochoa, "Deliverable 1: HV-LV modelling of selected HV feeders," *Advanced Planning of PV-Rich Distribution Networks*: AusNet Services, 2019, p. 36.

[2] K. Petrou and L. F. Ochoa, *AusNet Mini Grid Clusters - Deliverable 1 "HV and LV Network Modelling"*, Melbourne: University of Melbourne and AusNet Services, 2017.

[3] A. T. Procopiou and L. F. Ochoa, *Solar PV Penetration and HV-LV Network Impacts*, Melbourne, VIC, Australia: The University of Melbourne & AusNet Services, 2018.