Title: **Deliverables 3b-4: Improved Model-Free Operating Envelopes and Other Considerations**

Synopsis: This report builds on the previous two reports (foundations, methodology, and extensive performance tests) and presents improvements and updates on the Project “Model-Free Operating Envelopes at NMI Level”. Specifically, this report presents improvements to the offline data pipeline previously defined, presents an alternative allocation technique to the one previously defined, and discusses how to cater for changes in voltage regulation devices as well as partial smart meter data availability.

Data Document ID: UoM-C4NET-Model-Free_OEs-D3b-4_v01

Date: 31st January 2023

Prepared For: Centre for New Energy Technologies (C4NET)
C/- SBC, Level 2, 520 Bourke St, Melbourne
VIC 3000, Australia

Prepared By: Vincenzo Bassi
Prof Luis(Nando) Ochoa
Department of Electrical and Electronic Engineering
The University of Melbourne

Revised By: Prof Luis(Nando) Ochoa
Prof Tansu Alpcan
Department of Electrical and Electronic Engineering
Prof Christopher Leckie
School of Computing and Information Systems (CIS)
The University of Melbourne

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

Executive Summary

The report at hand corresponds to the Deliverables 3b and 4, and additional studies under the title “Improved Model-Free Operating Envelopes and Other Considerations”, as part of the project “Model-Free Operating Envelopes at NMI level” funded by the Centre for New Energy Technologies (C4NET) and involving the Australian Distribution Network Service Providers (DNSPs) AusNet Services, Jemena, Citipower & Powercor, and United Energy.

Voltage calculations are key for DNSPs to adequately operate and plan its low voltage (LV) distribution networks in a context of high distributed energy resources (DERs) penetration. However, voltage calculations would normally require power flow analyses and, consequently, detailed three-phase LV network models, which, in practice, are not readily available for DNSPs. Taking advantage of smart meter data, available to all DNSPs in the State of Victoria, this project aims to demonstrate that is possible to capture the physics of three-phase LV circuits and create an electrical model-free approach to calculate voltages. These model-free voltage calculations can be then used to estimate operating envelopes (OEs), assess connection requests, or to carry out hosting capacity assessments, enabling DNSPs to bypass the costly, time-consuming, and error-prone process of producing and validating electrical models.

The report at hand builds on top of [1] (foundations and methodology), [2] (extensive performance tests) and presents improvements and updates on the project. Thus, presents improvements to the offline data pipeline previously defined, presents an alternative allocation technique to the one previously defined, determines the extent to which the proposed approach can cater for voltage regulation devices, and presents initial findings of cases with partial smart meter data availability.

Offline Data Pipeline Improvements
Improvements to the offline data pipeline are presented due to data issues involving customers with extremely low values of demand (zero or a few Watts) throughout most of the training data set. These customers caused a decrease in accuracy in a limited number of cases presented in [2] which, in the test data set, had normal values of demand (kilowatts). This is because the test values were very far from what the neural network (NN) has been trained with. To address this, an additional step to clear these issues is incorporated into the offline data pipeline. To demonstrate the effectiveness of the proposed improvements, worst-case scenarios from the analyses in [2] are considered. By implementing this new step, the proposed approach can produce more accurate voltage calculations for the customers with normal demand as this makes it possible for the NN to capture the underlying relationships.

Operating Envelopes Allocation Technique: Maximise Exports
A new allocation technique called *maximise exports* is presented. This allocation technique is different from the one presented previously in [2] (called equally distributed). Instead of allocating the same OE to everyone, this new technique aims to achieve larger overall/total exports from all active customers by limiting the exports of those customers that are more sensitive to voltage rise (i.e., those located further from the distribution transformer) and, thus, facilitating larger OEs to those less sensitive. OEs are calculated for the case of Jemena Substation B, where it is shown that the maximise exports allocation technique can produce larger overall exports compared to equally distributed OEs.

Voltage Regulation Devices
The business as usual (BaU) operation of voltage regulation devices (i.e., when specific settings are kept) is captured by the proposed approach, as demonstrated in [2]. However, if settings change after the NN was trained, the NN will continue to calculate voltages as before (as it was trained to cater for the previous setting). To overcome this issue, two options are proposed: to retrain the NN with the latest data that incorporates such a change or by revising calculated values considering the new settings. OEs are calculated for the case of United Energy Substation A, where it is shown that upstream changes can be directly incorporated into the proposed approach by revising calculated voltages.

Partial Smart Meter Data Availability
In Victoria, the smart meter data from some customers might not be available to DNSPs (e.g., commercial and industrial [C&I] customers whose data is managed by third parties). Preliminary analyses presented in this report show that it is possible to obtain accurate voltage calculations as well

as consistent operating envelopes when only 40-50% of the customers are considered (in this case, three-phase residential customers were removed as a proxy of C&I customers). Although preliminary results are promising, further analyses must be carried out to fully determine the extent to which the proposed approach can cater for partial smart meter data availability in the context of Victoria. This will be analysed in a later stage of the project.

Table of Contents

List of Figures

List of Tables

1 Introduction

The proliferation of residential, behind-the-meter, Distributed Energy Resources (DERs), such as solar photovoltaics (PV), batteries, and electric vehicles (EVs), is creating new electricity flows that were not considered when designing the very infrastructure they are connected to, i.e., the low voltage (LV) distribution networks. Specifically, coincident exports from PV systems and batteries can cause reverse power flows which can lead to overvoltage issues [3], [4]. Conversely, EV charging demands exacerbate power flows which can cause undervoltage issues [5]. Therefore, either from operation or planning perspectives, distribution companies, known as Distribution Network Service Provides (DNSPs) in Australia, need to accurately assess the impacts of these technologies over LV network voltages. Such assessment, to date, is normally carried out through electrical calculations using power flow analyses and, consequently, LV network electrical models.

However, DNSPs struggle to have accurate an updated electrical model of their LV networks, which means that the accurate calculations required to assess the impacts of DERs are not possible, creating a significant barrier for DNSPs to adequately operate and plan its distribution networks in a context of increasing DER penetration. Taking advantage of historical smart meter data, available to DNSPs in the State of Victoria, this project aims to demonstrate that is possible to capture the physics of three-phase LV circuits and create an electrical model-free approach to calculate voltages. These model-free voltage calculations can be then used to estimate operating envelopes (OEs), i.e., time-varying export or import limits for active customers (i.e., customers engaged with aggregators), assess connection requests, or to carry out hosting capacity assessments.

The foundations and methodology of the proposed electrical model-free voltage calculations along with the initial tests were presented in an interim report "Deliverable 0: Concept, Smart Meter Data, and Initial Findings" [1]. The second report, "Deliverables 1-2-3a: Model-Free Voltage Calculations and Operating Envelopes" [2], showed that the proposed approach can carry out accurate multi-LV circuit voltage calculations, achieving and average deviation below 1 V in many cases. It was found that the proposed approach can use, at any moment in time, as little as 3 weeks of historical smart meter data to produce a Neural Network (NN) that is adequate to calculate voltages for at least 10 months without requiring further updates. The model-free approach to calculate OEs was also presented, giving the foundations for its implementation, and presenting its initial tests and findings.

This report builds on top of [1], [2] and presents improvements to the offline data pipeline to overcome the data issues that were found to decrease the accuracy in a reduced number of cases in [2]. These data issues correspond to customers with extremely low values of demand (zero or a few Watts) throughout most of the training data set. Additionally, the report at hand presents an alternative allocation technique to calculate OEs, which, differently than the one defined in [2] (i.e., equally distributed OEs), aims to maximise the overall exports from active customers. Furthermore, this report discusses how to adapt the methodology if the settings of voltage regulation devices are changed. Finally, the report presents initial findings of cases with partial smart meter data availability.

For the next stage of the project, issues associated with reactive power demand data will be investigated. It was found that some customers show a fixed power factor which, in turn, brings challenges to the NN as the training data becomes less diverse. Furthermore, thermal constraints will be incorporated into the formulation of the model-free OEs calculation approach. Additionally, further analyses will be carried out to fully determine the extent to which the proposed approach can cater for partial smart meter data availability in the context of Victoria (no data from C&I customers). The case for other parts of Australia (no data from some residential customers) will also be analysed.

This report addresses Deliverables 3b and 4. Additional analyses related to partial smart meter data availability are also presented. The report is structured as follows: Chapter 2 presents improvements to the offline data pipeline. Chapter 3 presents the maximise exports allocation technique to calculate OEs. Chapter 4 discusses the necessary adaptations to cater for changes in the settings of voltage regulation devices. Chapter 5 presents the preliminary analysis considering partial smart meter data availability in the context of the Victoria.

2 Offline Data Pipeline Improvements

The analyses carried out in [2] shows that the proposed model-free voltage calculations approach can carry out accurate multi-LV circuit voltage calculations in most parts of the historical smart meter data available for each of the studied distribution transformers, achieving and average deviation below 1 V in many cases. However, an accuracy decrease is observed in a reduced number of cases. This is associated to data issues, specifically because of the presence of customers with extremely low values of demand (zero or a few Watts) throughout most of the training data. Given the nature of the data, i.e., real smart meter data, such issues are likely to be faced by DNSPs if they decide to implement the model-free approach. Therefore, for an adequate implementation of the proposed approach, it is critical to define a data pipeline that can be easily implemented by DNSPs. Although this was previously formulated in [1], [2], improvements are required to detect and overcome the issues identified in [2].

As presented in [1], the proposed approach is composed of two main stages, offline and deployment. In the offline stage, the most suitable Neural Network (NN) for the studied LV network is determined and trained using historical smart meter data. Then, the obtained NN is used to calculate voltages in deployment to, for instance, determine operating envelopes. The first step to define and train the most suitable NN for a given LV network is to build the corresponding training data set from the raw historical smart meter data. Based on the results in [2], it can be observed that as little as 3 weeks of historical smart meter data are required to produce a NN that is adequate to calculate voltages for at least 10 further months without requiring updates if no changes have occurred, e.g., incorporation of new customers, topological reconfigurations, conductor changes, etc. If such a change occurs, sufficient smart meter data for the new situation must be collected and used to update the corresponding NN.

It is a well-known issue that any machine learning model is as good as the data that is fed. Thus, for the NN to capture the physical relationships of the corresponding LV network, an adequate training data set comprised of 6,048 instances (equivalent to 3 weeks at 5 minutes resolution) of customers' active and reactive power (i.e., $\boldsymbol{P}$ and $\boldsymbol{Q}$) and its corresponding voltage magnitudes (i.e., $\boldsymbol{V}$) must be built from the historical smart meter data. In this context, even though pre-processing and filtering processes were defined in [1], [2], its effectiveness proved not to be sufficient for the challenges faced by DNSPs when dealing with real smart meter data. Specifically, it is shown that such processes do not cater for the presence of customers with extremely low values of demand (zero or a few Watts), which can have a high impact in the final voltage calculations. For further details please refer to [2].

After a rigorous analysis of the historical smart meter data provided by the DNSPs, it is observed that some customers can present active and reactive power values equal or close to 0 (e.g., less than 10 W) whilst having adequate voltage measurements for long periods or throughout the entire studied period. This could be due to smart meter malfunctions. For instance, from the 155 customers in the historical smart meter data provided for Jemena Substation B, 4 single-phase customers and 1 three-phase customer are found to have active and reactive power values equal or close to 0 throughout the entire data period (≈22 weeks). Similarly, one single-phase customer and one phase of a three-phase customer are found in the same situation but with random active and reactive power peaks of short duration that appears a reduced number of times within the studied period. Furthermore, one single-phase customer is found to have adequate active and reactive power values until certain point in the historical smart meter data, whereas other single-phase customer is found to have adequate active and reactive power values since certain point in the historical smart meter data. Even though the latter could be caused by smart meter malfunctions, it could also be associated to the disconnection or latter incorporation of those customers, or, alternatively, to extended periods when customers have disconnected all loads in their property (such as during extended periods of holidays).

For the proposed approach to adequately capture the physics of the studied LV network, it is critical to build the training data set with features (i.e., customers' active and reactive power) that provides meaningful information to the NN, i.e., with enough variety for the NN to accurately capture the underlying relationships. Essentially, any feature that does not provide meaningful information to the NN can create wrong voltage calculations when deploying the NN to calculate voltages. Specifically, if adequate values for such features are considered. Thus, all customers with all active and reactive power values equal or close to 0 must be removed from the training data set as their relationships are not represented. This also means that voltage calculations for such customers cannot be carried out in deployment.

In this context, building on top of the pre-processing and filtering processes previously defined in [1], [2] (hereinafter, Steps 1 to 3), this section presents a complete offline data pipeline to build the corresponding training data set that incorporates a new filtering step (Step 4) which aims to cater for the mentioned challenges. The complete offline data pipeline is presented in Table 1 and detailed below.

Table 1. Offline Data Pipeline

Step 1	Collect raw historical per-phase smart meter data
Step 2	Obtain customers' active and reactive power from raw historical smart meter data
Step 3	Remove invalid and unfeasible instances
Step 4	Remove customers whose relationships are not represented

- **Step 1:** Collect raw historical per-phase smart meter data
  Retrieve historical per-phase smart meter data from all customers in the studied LV network. As presented in [1], the format of the raw historical data is specific for each DNSP. It is important to note that, smart meters are designed to measure energy consumption in a wide range of formats, which means that the convention adopted for a specific smart meter needs to be properly defined to ensure the correct interpretation of the measurements.

- **Step 2:** Obtain customers' active and reactive power from raw historical smart meter
  Even though customers' voltages, $\boldsymbol{V}$, are directly measured by smart meters, customers' active and reactive power, $\boldsymbol{P}$ and $\boldsymbol{Q}$, must be derived using $\boldsymbol{V}$ and other available measurements. However, the format of the raw historical data is specific for each DNSP [1] and, thus, $\boldsymbol{P}$ and $\boldsymbol{Q}$ derivation will also be specific per DNSP, depending on the available measurements. For further details on $\boldsymbol{P}$ and $\boldsymbol{Q}$ derivation for each DNSP please refer to [1].

- **Step 3:** Remove invalid/unfeasible instances
  Invalid instances correspond to instances with missing (i.e., not recorded) values of active power, reactive power, or voltage magnitude, for at least one customer within the studied LV network. Similarly, unfeasible instances correspond to instances with unfeasible (i.e., beyond feasible limits) values of active power, reactive power, or voltage magnitude, for at least one customer. From the perspective of the model-free voltage calculations, it is essential that the training data set is comprised only of valid and feasible instances [1], [2]. Thus, if there are gaps or unfeasible values in the data for a single time instance (e.g., a missing voltage value for one customer) then all customers data for that time instance must be removed and deemed unusable as inputs to the proposed approach.

  Specifically, in this step all instances with missing measurements of active power, reactive power, or voltage magnitude, for at least one customer are deleted. Similarly, all instances with unfeasible measurements for at least one customer are deleted. Here, unfeasible measurements are defined as active or reactive power measurements beyond customers' connection point limits, or voltage measurements outside the range of [210 V, 280 V]. All the above will effectively reduce the size of the training data to be used.

  It is important to note that if a customer is found to have invalid (i.e., missing values) or unfeasible values (e.g., V = 0) throughout the entire data set, this customer must be removed from the historical data first and then invalid/unfeasible instances for the remaining customers must be removed. If not, such customer will lead to remove the complete data set.

- **Step 4:** Remove customers whose relationships are not represented
  This step is incorporated to the offline data pipeline to address the issues observed in [2] and remove all customers whose relationships are not represented from the training data set (and, consequently, from the calculation of voltages). Specifically, customers' active power values are

analysed at the end of Step 3. All customers with 99% of their corresponding active power values below 10 W are removed from the training data set.

After the offline data pipeline, the training data set is ready to be used to define and train the most suitable NN for the studied LV network. Once trained, the final NN requires measurements from all customers considered during the training process to calculate voltages in deployment. When the final NN is used for operation applications, it would require either operational data or forecast. If a complete, adequate set of data is not available (due to missing or incorrect data), a pre-processing process can be implemented to replace such input values by a previous value or even by 0. This, however, will impact the accuracy of the obtained voltage calculations as the interactions related to these data points will not be represented. For further details into partial smart meter data availability and its effects over the proposed model-free voltage calculations, please refer to Section 5. For planning applications, on the other hand, these issues are not expected given that the inputs are defined by the what-if scenario being investigated.

An alternative to the offline data pipeline defined in Table 1 would be to flag those customers and account for its normal operation when deploying the NN. Thus, a process should be implemented instead of Step 4 to detect and flag those customers whose operation should be considered when deploying the NN. Thus, for instance, if a given customer is found to have only P and Q values equal or close to 0 throughout the entire training data, the NN must be deployed considering scenarios that reflects this as other scenarios (for instance, the consideration of a PV system for this customer in a planning analysis) will lead the approach to calculate wrong voltage values.

2.1 Implementing the Offline Data Pipeline

To illustrate the importance of the offline data pipeline, and, particularly, of Step 4 (added to cater for the mentioned challenges), the worst-case scenarios for the analyses carried out in [2], i.e., considering up to Step 3 only, are used. Specifically, the worst-case scenario for Jemena Substation B and Jemena Substation A, which corresponds to the largest available distribution transformers, when using 3 weeks of historical data to produce the NN are analysed.

2.1.1 Jemena Substation B

Jemena Substation B corresponds to a LV network comprised of 4 LV circuits with a total of 155 customers (116 single-phase customers and 39 three-phase customers). As in [2], Step 1 to Step 3 of the data pipeline are applied to all the historical data. Here, 1 phase of a three-phase customer is removed as its data is found to be full of unfeasible values (voltages equal or close to 0). After Step 3, a total of 44,632 instances are considered for model-free voltage calculations, which is equivalent to $\approx$155 days (i.e., $\approx$22 weeks), spanning the period among 22-11-2020 and 30-09-2021. For further details on the corresponding historical smart meter data, please refer to [2].

Specifically, for Jemena Substation B, group 8 is considered. The training data set of group 8 spans the period between 24-05-2021 14:05:00 and 13-07-2021 21:05:00, which comprise a calendar period longer than 3 weeks (specifically, more than 7 weeks). This is because of the application of Step 1 to Step 3 of the data pipeline to the historical data. In other words, from the 14,496 instances that were recorded from 24-05-2021 14:05:00 to 13-07-2021 21:05:00, 8,437 instances were deemed unusable due to missing or incorrect values, leaving a total of 6,048 instances (i.e., an equivalent to 3 weeks at 5 minutes resolution).

Two NN are assessed, one using the offline data pipeline as in [2], i.e., up to Step 3, and other one using the complete offline data pipeline defined in Table 1. The performance of the NN is evaluated using the next 6,048 instances (i.e., an equivalent to 3 weeks at 5 minutes resolution) of the historical data (test data set). The obtained results are presented in Table 2, Figure 1, and Figure 2. Please note that the coloured regions in Figure 1 correspond to the difference between the maximum and the minimum value in each instance, whereas for the scatter plots presented in Figure 2 the x-axis corresponds to actual voltage values and the y-axis accounts for the values calculated by the NN.

Table 2. Model-Free Voltage Calculations Results

Group	Training Period	Training Initial Instance	Training Final Instance Test Initial Instance	Test Final Instance	Offline Data Pipeline	RMSE [V]	Av. Dev. [V]	Max. Dev. [V]
8	3 weeks	24-05-2021	13-07-2021	10-08-2021	As in [2]	9.95	8.65	49.25
					As in Table 1	0.71	0.54	7.17

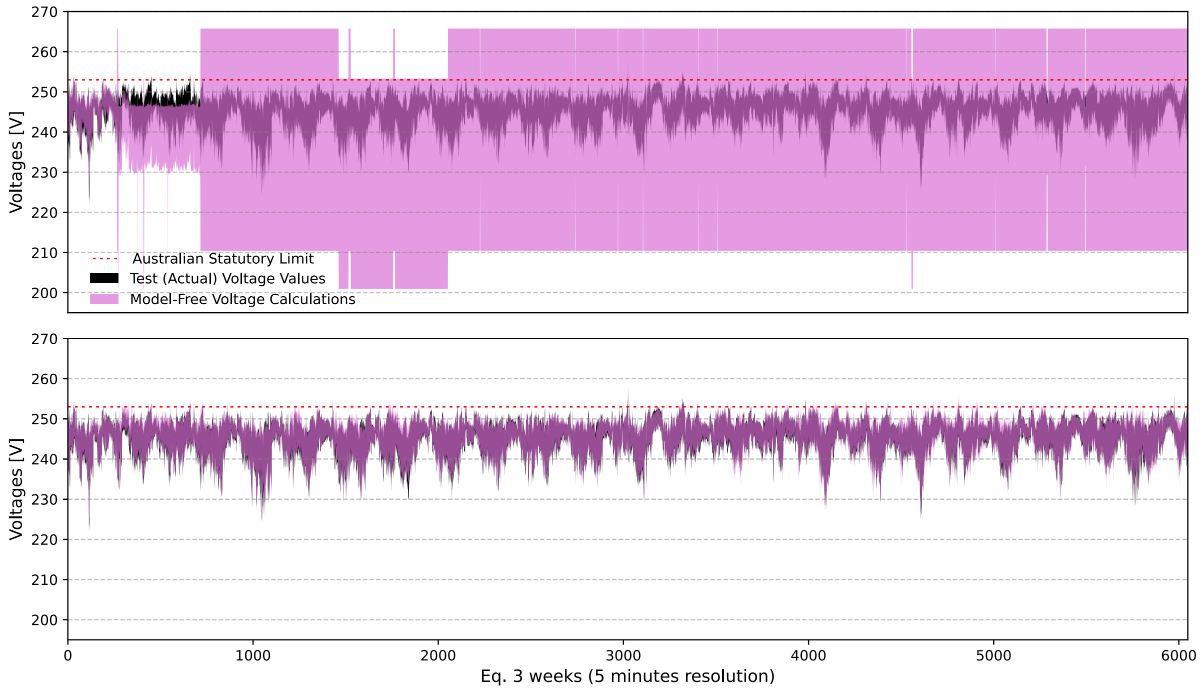

Figure 1. Time-Series Voltage Calculations, as in [2] (upper) vs Complete Offline Data Pipeline (lower)

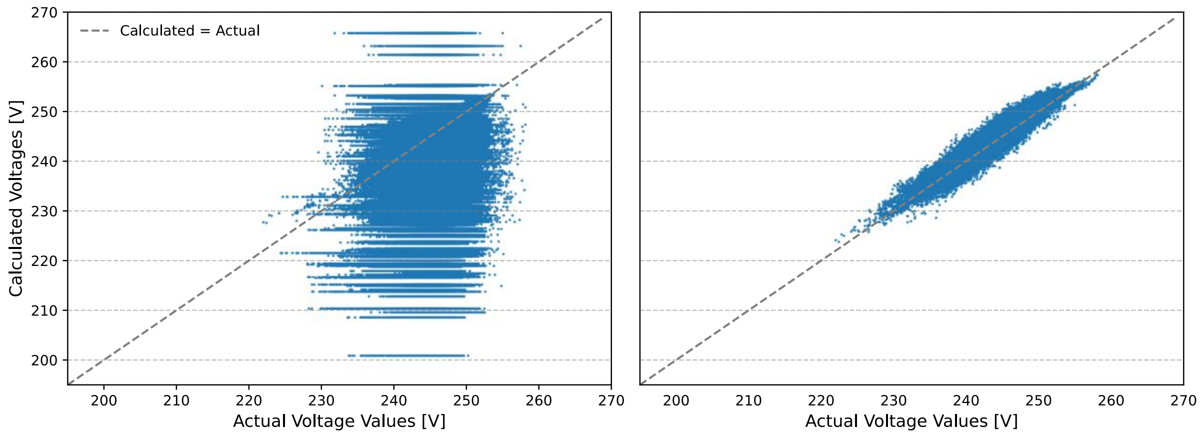

Figure 2. Calculated vs Actual Voltage Values, as in [2] (left) vs Complete Offline Data Pipeline (right)

First, it can be noted that the NN that considers the offline data pipeline as in [2] corresponds to the worst-case scenario in terms of Av. Dev. (8.65 V) when using 3 weeks for training in the analyses carried out in [2]. From Figure 1 (upper), and Figure 2 (left), it can be observed that the proposed approach calculates erroneous values through most of the test data set. Indeed, since instance 265, erroneous voltage calculations can be observed until the end of the test data set. For these instances, the NN calculates flat values per customer that can goes from 200 V to 265 V through several instances.

From Table 2, Figure 1, and Figure 2, it can be noted that by implementing the offline data pipeline as defined in Table 1, the proposed approach produces very accurate voltage calculations for those customers that are found to have meaningful information in their historical smart meter data. Overall, extremely accurate voltage calculations are obtained, achieving a RMSE of 0.71 V, an Av. Dev. of 0.54 V, and Max. Dev. of 7.17 V.

Specifically, in Step 4, 7 single-phase customers, 1 phase of a three-phase customer, and 1 three-phase customer are removed from the training data as its active power values are found to be 99% of the time below 10 W. This, in turns means that voltage calculations for these customers cannot be carried out in deployment. The data of these customers, for group 8, is presented below.

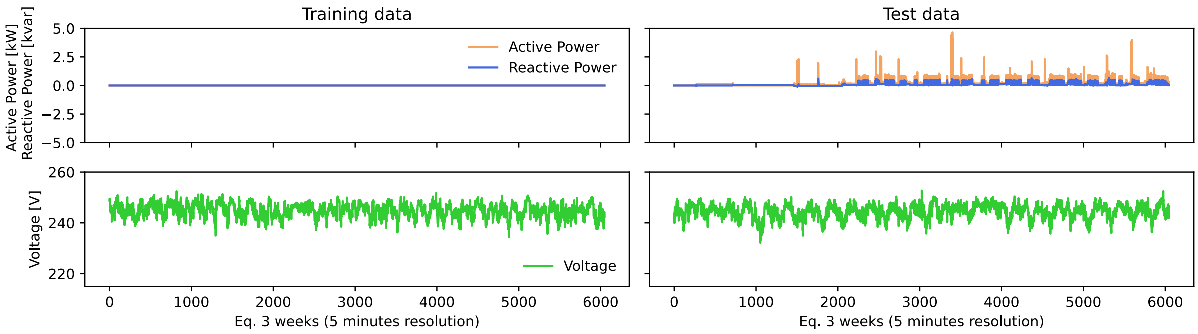

Figure 3. Customer 36 (*Supply_Data_SubB_Cust36*)

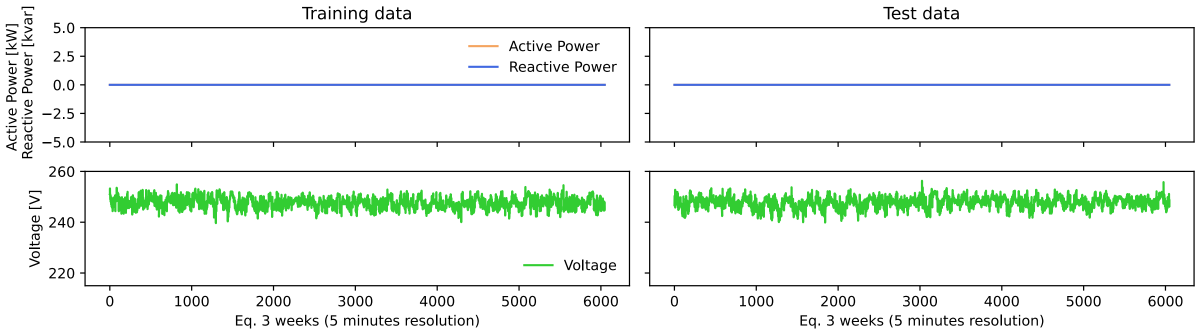

Figure 4. Customer 100 (*Supply_Data_SubB_Cust100*)

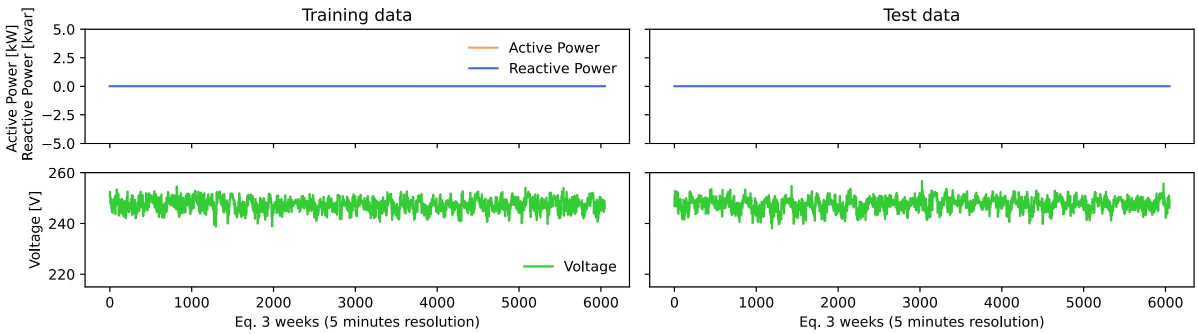

Figure 5. Customer 107 (*Supply_Data_SubB_Cust107*)

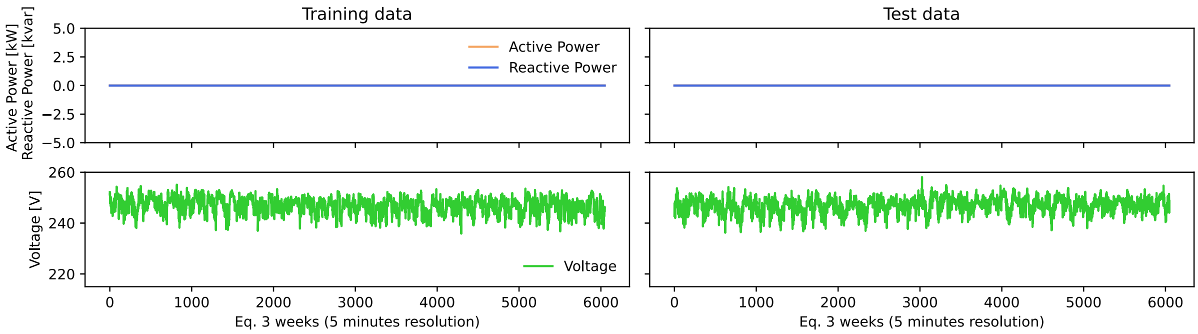

Figure 6. Customer 26 (*Supply_Data_SubB_Cust26*)

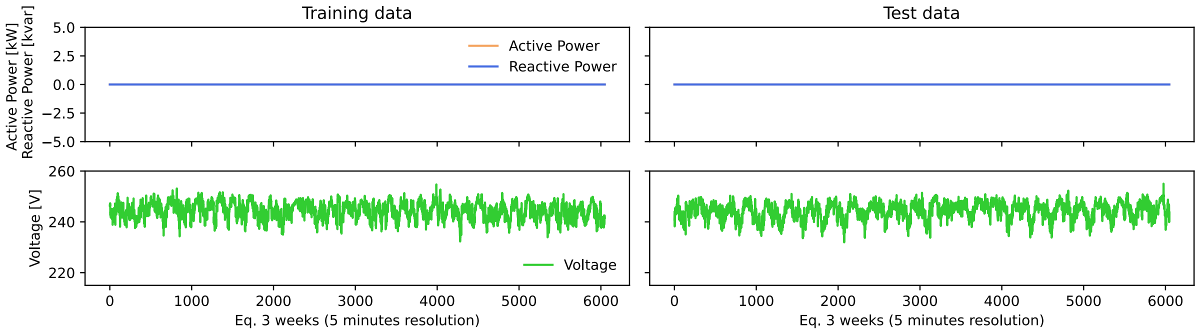

Figure 7. Customer 39 (*Supply_Data_SubB_Cust39*)

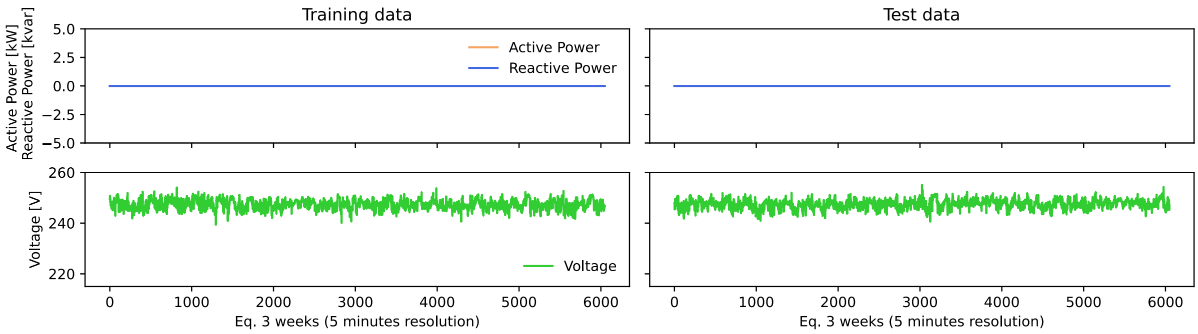

Figure 8. Customer 48 (*Supply_Data_SubB_Cust48*)

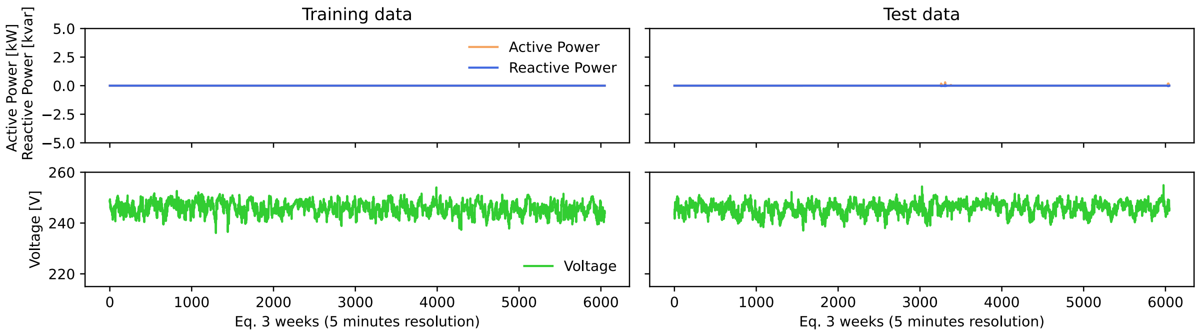

Figure 9. Customer 54 (*Supply_Data_SubB_Cust54*)

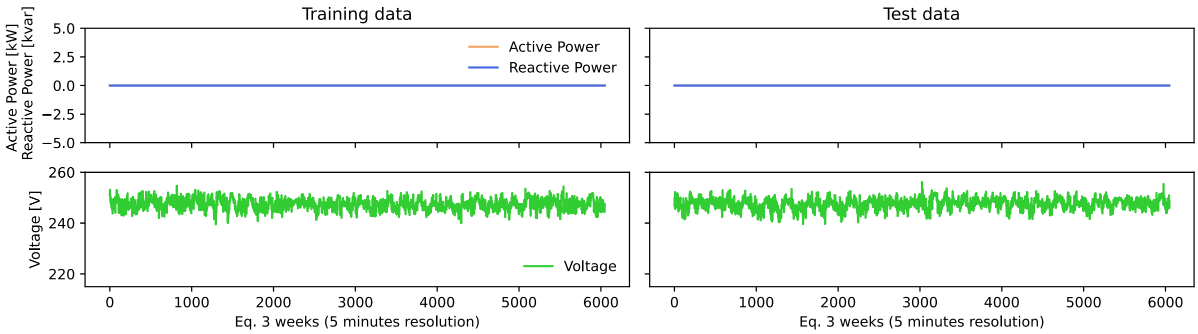

Figure 10. Customer 74 (*Supply_Data_SubB_Cust74*) - Phase 3

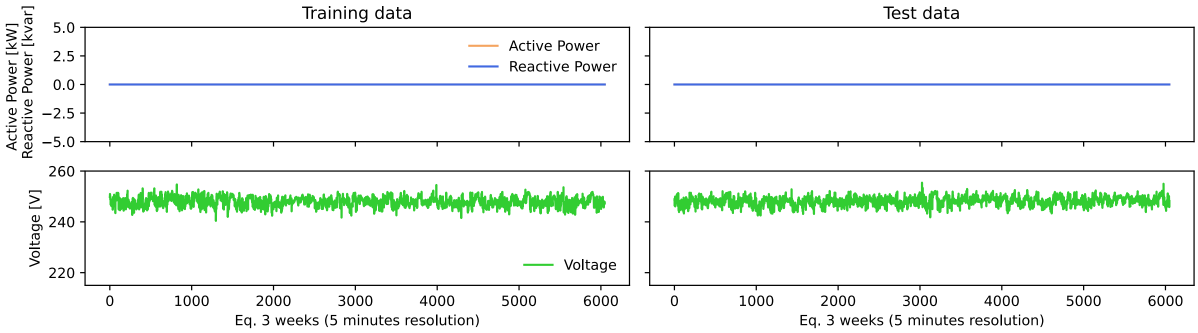

Figure 11. Customer 92_2 (*Supply_Data_SubB_Cust92_2*) - Phase 1

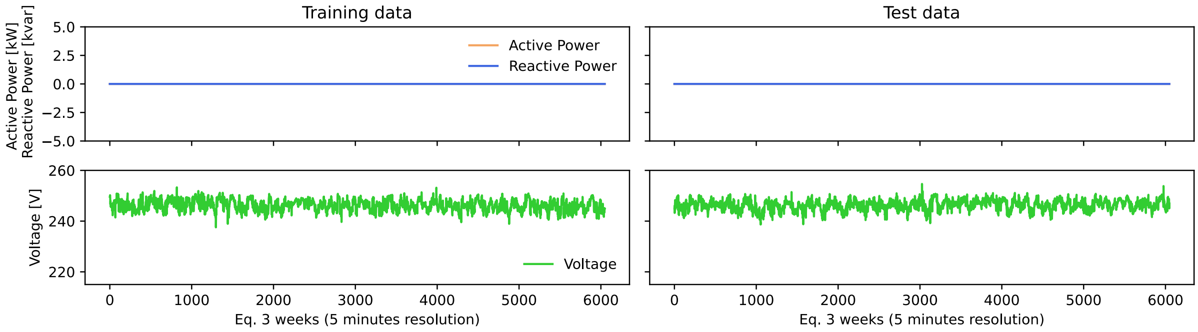

Figure 12. Customer 92_2 (*Supply_Data_SubB_Cust92_2*) - Phase 2

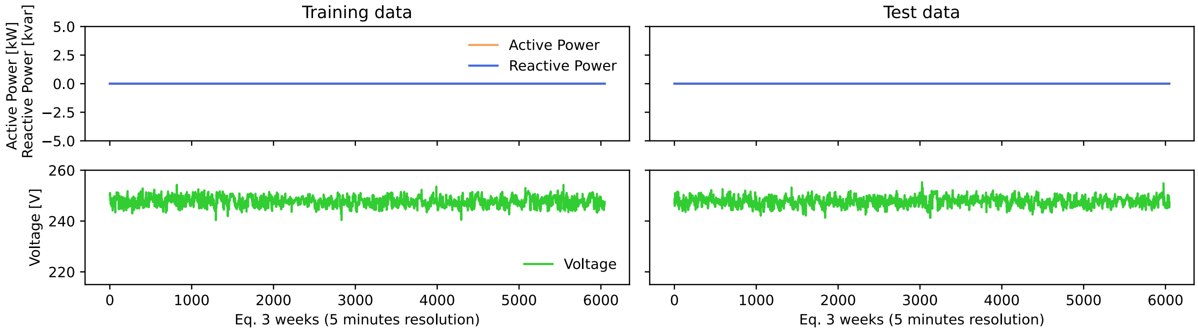

Figure 13. Customer 92_2 (*Supply_Data_SubB_Cust92_2*) - Phase 3

After a rigorous analysis of the complete historical smart meter data provided for Jemena Substation B, it is found that from these customers, only Customer 48 and 36 presents valid active and reactive power measurements at some point of the historical data, specifically, at the beginning and at the end, respectively. This might be explained by smart meter malfunctions, by the disconnection or later incorporation of such customers, or even to extended periods of holidays without loads in the property. All the remaining customers presents values equal or close to 0 throughout the entire studied period.

Furthermore, it is noted that the erroneous voltage calculations observed in Figure 1 (upper), and Figure 2 (left), can be explained mainly by Customer 36 (Figure 3) which has active and reactive power equal or close to 0 from the beginning of the historical data until 16-07-2021, i.e., until around instance 265 of the test set. This can be due to smart meter malfunctions or to the latter incorporation of Customer 36 and, as can be noted in Figure 1 (upper), and Figure 2 (left), has a big impact over the calculation of voltages. This is because the values for Customer 36 seen by the NN during training are equal or close to 0 (maximum and minimum active power value during training equal to 0.0005 kW and -0.0007 kW, respectively), i.e., its corresponding relationships are not represented. Thus, the weights of the NN were optimised during the training process of the NN to account for a situation in which all customers' voltage values observed in the training data set are related to active and reactive power values equal or close to 0 for Customer 36, and any change in this situation will create a big impact on the calculated voltages for all customers. This lead the NN to calculate erroneous voltage values as soon as adequate values of active and reactive power for Customer 36 are considered. This shows that essentially, a feature that does not provide meaningful information (i.e., with variety enough) to the NN to capture the underlying relationships during training can create issues when deploying the NN. Specifically, when the NN is expected to calculate voltages with adequate values for those features. Therefore, for an adequate implementation of the proposed approach, the offline data pipeline must be implemented to automatically detect and deal with such issues.

2.1.2 Jemena Substation A

Jemena Substation A corresponds to a LV network comprised of 3 LV circuits with a total of 170 customers (156 single-phase customers and 14 three-phase customers). As in [2], Step 1 to Step 3 of the data pipeline are applied to all the historical data. After Step 3, a total of 80,703 instances are considered for model-free voltage calculations, which is equivalent to ≈280 days (i.e., ≈40 weeks), spanning the period among 01-09-2020 and 29-08-2021. For further details on the corresponding historical smart meter data, please refer to [2].

Specifically, for Jemena Substation A, group 5 is considered. The training data set of group 5 spans the period between 28-10-2020 and 22-11-2020 (approximately, 3 calendar weeks). Two NN are assessed, one using the offline data pipeline as in [2], i.e., up to Step 3, and other one using the complete offline data pipeline defined in Table 1. The performance of the NN is evaluated using the next 6,048 instances (i.e., an equivalent to 3 weeks at 5 minutes resolution) of the historical data (test data set). The obtained results are presented in Table 3, Figure 14, and Figure 15.

Table 3. Model-Free Voltage Calculations Results

Group	Training Period	Training Initial Instance	Training Final Instance Test Initial Instance	Test Final Instance	Offline Data Pipeline	RMSE [V]	Av. Dev. [V]	Max. Dev. [V]
5	3 weeks	28-10-2020	22-11-2020	18-12-2020	As in [2]	11.90	9.78	44.27
					As in Table 1	1.39	1.11	9.74

First, it can be noted that the NN that considers the offline data pipeline as in [2] corresponds to the worst-case scenario in terms of Av. Dev. (9.78 V) when using 3 weeks for training in the analyses carried out in [2]. From Figure 14 (upper), and Figure 15 (left), it can be observed that the NN is calculating flat values per customer through several instances. As shown in Figure 15, such values can go below 210 V.

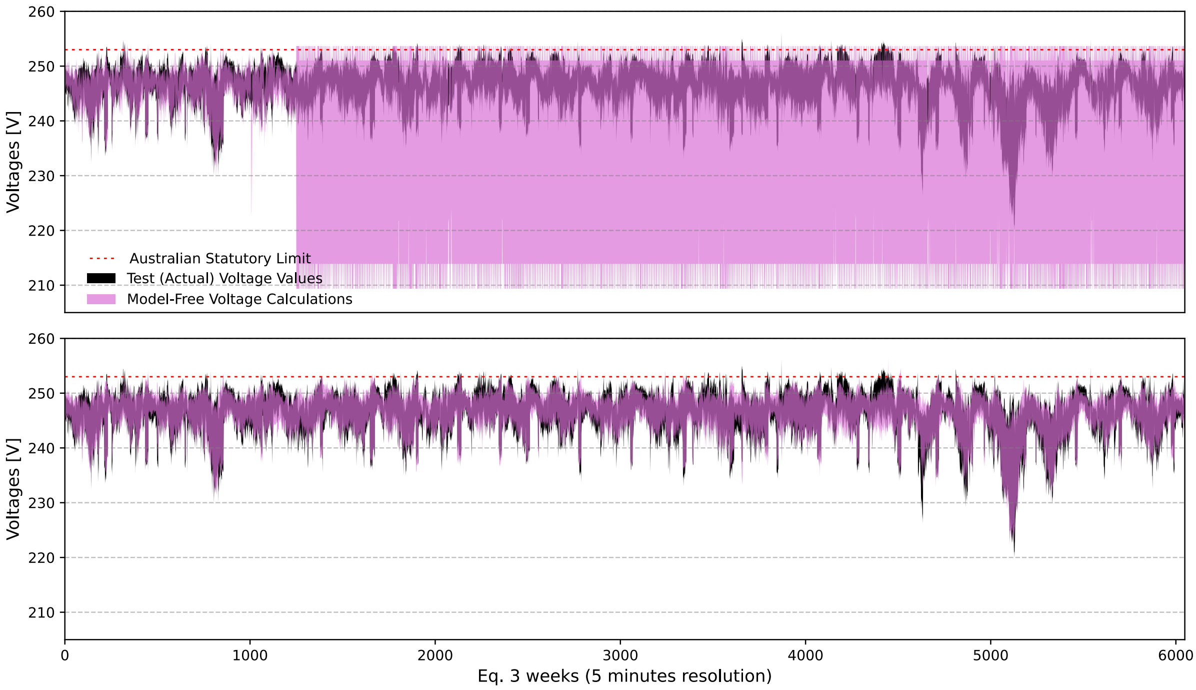

Figure 14. Time-Series Voltage Calculations, as in [2] (upper) vs Complete Offline Data Pipeline (lower)

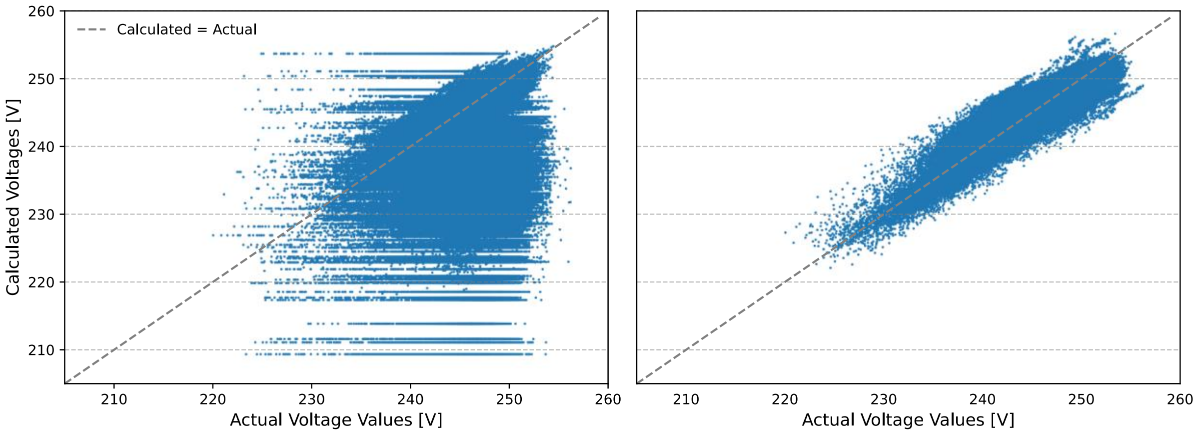

Figure 15. Calculated vs Actual Voltage Values, as in [2] (left) vs Complete Offline Data Pipeline (right)

From Table 3, Figure 14, and Figure 15, it can be noted that by implementing the offline data pipeline as defined in Table 1, the proposed approach produces very accurate voltage calculations for those customers that are found to have meaningful information in their historical smart meter data. Overall, accurate voltage calculations are obtained, achieving a RMSE of 1.39 V, an Av. Dev. of 1.11 V, and Max. Dev. of 9.74 V.

Specifically, in Step 4, 1 single-phase customer and 1 phase of 2 three-phase customers are removed from the training data as its active power values are found to be 99% of the time below 10 W. This, in turns means that voltage calculations for these customers cannot be carried out in deployment as their corresponding relationships are not properly represented.

In conclusion, the offline data pipeline defined in Table 1 allows DNSPs to build an adequate training data set from the raw historical smart meter data, preventing the proposed approach to consider customers whose relationships are not represented in the historical data as their inclusion might cause issues when the NN is deployed to calculate voltages considering adequate values for such customers.

2.2 Key Remarks

This section present improvements over the offline data pipeline presented in [1],[2] to overcome the accuracy decrease observed in a reduced number of cases in [2], which is found to be caused by data issues, specifically, customers with extremely low values of demand (zero or a few Watts) throughout most of the training data set. It is a well-known issue that any machine learning model is as good as the data that is fed. Thus, for the NN to accurately capture the physical relationships of the corresponding LV network, it is critical to build the training data set with features (i.e., customers' active and reactive power) that provides meaningful information to the NN, i.e., with enough variety for the NN to accurately capture the corresponding relationships. Essentially, any feature that does not provide meaningful information to the NN can cause wrong voltage calculations when deploying the NN to calculate voltages. This is key when the NN is deployed, for instance, using what-if scenarios.

Thus, the offline data pipeline is improved with Step 4 in Table 1, which aims to detect and remove customers whose relationships are not properly represented from the training data set. Specifically, customers with 99% of their corresponding active power values are below 10 W are removed. This in turns also means that voltage calculations for such customers cannot be carried out in deployment.

To illustrate the effectiveness of the offline data pipeline, the case of Jemena Substation B and Substation A are considered (worst-case scenarios in [2]). From the obtained results it can be observed that by implementing the offline data pipeline defined in Table 1, the proposed approach produces very accurate voltage calculations for those customers that are found to have meaningful information in their historical smart meter data, i.e., variety enough for the NN to capture their relationships, preventing the proposed approach to consider customers whose relationships are not represented in the historical data as their inclusion might cause wrong voltage calculations when the NN is deployed considering adequate values for such customers.

3 Operating Envelopes Allocation Technique: Maximise Exports

It is shown in [2] that the proposed model-free voltage calculations can be used with a heuristic algorithm to explore different import or export values for active customers (i.e., customers engaged with aggregators) and thus, define its corresponding operating envelopes (OEs), i.e., time-varying import or export limits that ensure network integrity. The proposed model-free OEs calculation approach is presented in Figure 16, for further details please refer to [2]. Please note that, to illustrate the concepts, this section considers export limits only. Nonetheless, it is important to note that for the case of import limits, the corresponding calculations mirrors the case of export limits but considering active customers importing active power.

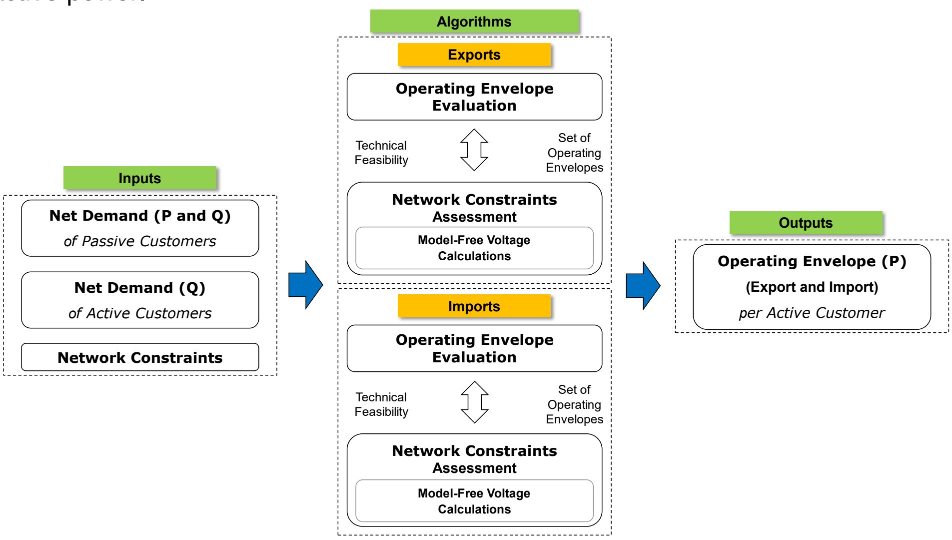

Figure 16. Model-Free Operating Envelopes Calculation Approach

As observed in Figure 16, the core of the model-free OEs calculation approach is the OEs algorithm which is composed of two main blocks, OE evaluation and Network Constraints Assessment. Both blocks interact in an iterative process used to determine the maximum exports according to the adopted allocation technique (i.e., how the OEs are allocated among customers). The OE algorithm defined in this project is as follows. In the first iteration, an initial set of OEs with P = 0 kW for the active customers is produced in the OE evaluation block. This set of OEs is then evaluated using the proposed model-free voltage calculations within the Network Constraints Assessment block, if constraints are not breached, a new set of OEs with increased values is produced according to the adopted allocation technique. This process is repeated until the network capacity is fully utilised.

Therefore, a key feature of the model-free OEs algorithm is the allocation technique to be considered. This will naturally impact the obtained OEs as define how OEs are allocated among the active customers. So far, the allocation technique considered throughout this project corresponds to equally distributed OEs, i.e., all customers are assigned with the same OE at each iteration. For further details please refer to [2]. Equally distributed OEs will be inherently limited by the active customers that are located further from the distribution transformer as those customers present higher sensitivity with respect to exports.

An alternative allocation technique is to maximise active customers' exports, which maximise overall exports by limiting the OE of those active customers that present higher sensitivity with respect to exports. Thus, in each iteration instead of increasing the OE of all active customers equally (as in the case of equally distributed OEs), the active customer with the lowest voltage is identified and only its corresponding OE is increased[1]. This allows the OE algorithm to prevent limitations due to those

[1] Here, for the case of imports, only the OE of the active customer with the highest voltage will be increased (i.e., consider larger imports).

customers that are located further from the distribution transformer and allocate larger OEs for those customers that are less sensitive with respect to exports, which, in turns, results in larger overall exports from the LV network. The general steps to calculate model-free OEs, for a specific instance, using both allocation techniques (i.e., equally distributed and maximise exports) are presented in Table 4.

Table 4. Model-Free Operating Envelopes Calculation Steps

Step 1	Produce the most suitable NN for the studied LV network
Step 2	Initialise input values
Step 3	Produce initial set of OEs
Step 4	Check initial set of OEs ▪ If OK? → Go to Step 5 ▪ If not → Active customers cannot export
Step 5	Increase OEs as per allocation technique
Step 6	Check the new set of OEs ▪ If OK? → Back to Step 5 ▪ If not → OE is defined by the previous iteration

- **Step 1:** Produce the most suitable NN for the studied LV network
  To do this, first, the offline data pipeline presented in Section 2 must be implemented to adequately define the corresponding training data set. Then, the methodology presented in [1] (and further updates presented in [6], details in Section 3.1) must be used to define and train the most suitable NN for the studied LV network.

- **Step 2:** Initialise input values
  Initialise input values to be considered. As presented in Figure 16, these values correspond to:
  - Net demand (P and Q) of passive customers
  - Net demand (Q) of active customers
  - Network constraints: voltage constraints, customers' connection point limitations, transformer limits, etc.

- **Step 3:** Produce initial set of OEs
  OE Evaluation block. In here, all active customers are initialised with its corresponding initial OE. In this study, in the first iteration, all active customers are assigned with P = 0 kW.

- **Step 4:** Check initial set of OEs
  Network Constraints Assessment block. The NN produced in Step 1 is used to assess if constraints are breached. If any considered constraint is breached, active customers cannot export. If not, continue to Step 5.

- **Step 5:** Increase OEs as per allocation technique
  OE Evaluation block. New set of OEs is defined based on the adopted allocation technique.
  - Equally distributed: OEs from all active customers are incremented equally
  - Maximise Exports: Only the OE from the customer with the lowest voltage in the previous iteration is incremented

  A heuristic process in which the corresponding OEs are incremented in 1 kW at each iteration is implemented.

- **Step 6:** Check new set of OEs
  Network Constraints Assessment block. The NN produced in Step 1 is used to assess if constraints are breached. If any considered constraint is breached, the corresponding OEs has been obtained and correspond to the set explored in the previous iteration. If not, go back to Step 5. This process is repeated until any considered constraint has been breached.

At the end of Step 6, a set of OEs for active customers (active power only) that ensures network integrity for the instance of interest is obtained. It can be noted that, Step 2 (Initialise input values) can be implemented using either operational smart meter data or forecast depending on how OEs are being produced, i.e., near real-time or in advance, respectively.

3.1 Improvements in Neural Network Structure

As noted, the calculation of OEs is all about pushing the limits of the studied LV network. Thus, the extrapolation capabilities of the proposed approach are critical. However, this cannot be analysed using historical smart meter data only, as an electrical model is required to run power flow simulations for different what-if scenarios and thus, assess the performance of the proposed approach in cases that goes well beyond what can be observed in the historical data. In this context, The University of Melbourne has carried out extensive studies using synthetic smart meter data and realistic electrical models. These findings have been recently accepted for publication on IEEE Transactions on Smart Grids and can be found at [6].

It is presented in [6] that the proposed approach must consider L2 regularisation to prevent overfitting on the historical data and to enhance its extrapolation capabilities. Thus, a penalty factor is added into the formulation of the corresponding NN. Such penalty factor is known as regularisation factor and must be determined. Hence, an additional hyperparameter (along with its respective search space) must be included into the K-fold cross validation process defined in [1], as in [6]. Besides, a NN structure adaptable to any kind of LV network based on its number of per-phase AMI devices (i.e., $|C|$) is defined for operation applications. This structure is presented in Figure 17 and its corresponding hyperparameters and settings are detailed in Table 5. Hereinafter, all analyses will consider this NN structure, hyperparameters, and settings, unless specified otherwise.

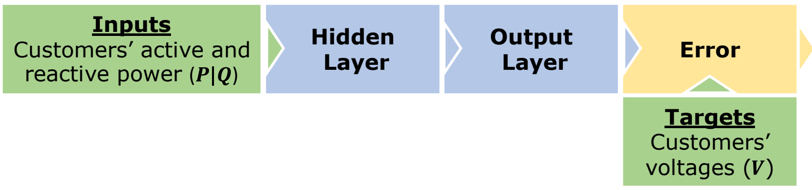

Figure 17. Neural Network Structure

Table 5. Hyperparameters and Settings

Hyperparameters and Settings	NN
Inputs	$2\|C\|$
Outputs	$\|C\|$
Output layer activation function	Linear
Error function	Mean Squared Error
Scaler	[0,1]
Optimiser	ADAM
Regularisation	L2
Number of Neurons	$5\|C\|$
Activation Function	Tanh
Learning rate	1e-4
Regularisation factor	1e-5
Batch Size	Eq. to 6 hours
Epochs	2,000

3.2 Case Study

To illustrate the calculation of OEs in the context of both allocation techniques, i.e., equally distributed and maximise exports, the case of Jemena Substation B is used. Jemena Substation B corresponds to a LV network comprised of 4 LV circuits with a total of 155 customers (116 single-phase customers and 39 three-phase customers). For further details on the historical data, please refer to [2]. Specifically, the case of group 1 when using 3 weeks for training is considered. The training data set of group 1 spans the period between 22-11-2021 and 07-01-2021. A total of 6 single-phase customers, 1 phase of 2 three-phase customers and a three-phase customer are removed from the historical data once the offline data pipeline defined in Section 2 is applied.

The first step is to define the most suitable NN for the studied LV network. Once trained, the performance of the NN is evaluated using the next 6,048 instances (i.e., an equivalent to 3 weeks at 5 minutes resolution) of the historical data (test data set). The model-free voltage calculations results obtained in the test data set are presented in Table 6, Figure 18, and Figure 19.

Table 6. Model-Free Voltage Calculations Results

Group	Training Period	Training Initial Instance	Training Final Instance Test Initial Instance	Test Final Instance	Season	RMSE [V]	Av. Dev. [V]	Max. Dev. [V]
1	3 weeks	22-11-2020	07-01-2021	13-03-2021	Summer	0.64	0.48	9.86

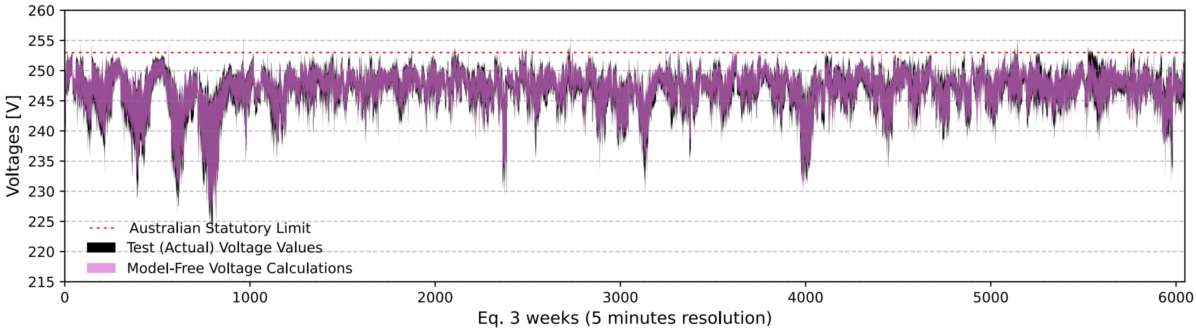

Figure 18. Time-Series Voltage Calculations

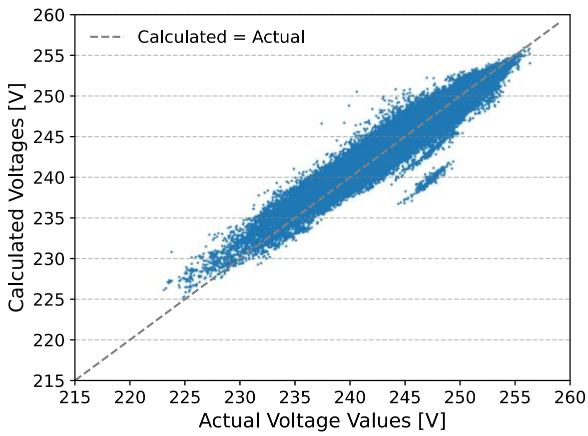

Figure 19. Calculated vs Actual Voltage Values

As observed, the proposed approach can capture the physics and produce accurate voltage calculations in the test data set, achieving an average deviation below 0.5 V. Once the NN is obtained (i.e., Step 1), is then used to calculate OEs for the first instance at noon of the test set. In this context, the next step (i.e., Step 2) is to initialise input values. The inputs considered for the present case study are listed below. It is important to note that half of the customers detected with PV systems are considered as active customers (i.e., a total of 6 active customers is considered), these customers are presented in Table 7.

- **Net Demand (P and Q) of Passive Customers:** The net demand of passive customers at the previous instance is considered to approximate the net demand of passive customers at the instance of interest.

- **Net Demand (Q) of Active Customers:** For simplicity, active customers are considered to operate with a unity power factor.
- **Network constraints:** For the present case study, only voltage statutory limits and customers' connection point limitations are assessed, i.e., voltages must be within 216 V and 253 V, and customers can export up to 14 kW.

Table 7. Active Customers

Customers with PV	Active Customer
Customer 97 (*Supply_Data_SubB_Cust97_1_P*)	✓
Customer 136 (*Supply_Data_SubB_Cust136_1_P*)	✓
Customer 137 (*Supply_Data_SubB_Cust137_1_P*)	
Customer 138 (*Supply_Data_SubB_Cust138_1_P*)	✓
Customer 140 (*Supply_Data_SubB_Cust140_1_P*)	✓
Customer 142 (*Supply_Data_SubB_Cust142_1_P*)	
Customer 145 (*Supply_Data_SubB_Cust145_1_P*)	
Customer 146 (*Supply_Data_SubB_Cust146_1_P*)	
Customer 149 (*Supply_Data_SubB_Cust149_1_P*)	✓
Customer 150 (*Supply_Data_SubB_Cust150_1_P*)	✓
Customer 151 (*Supply_Data_SubB_Cust151_1_P*)	

Using these input values along with an initial set of OEs equal to 0 kW for all active customers (Step 3), OEs are calculated using both allocation techniques, i.e., equally distributed and maximise exports, as in Table 4 (Steps 4 to 6). The obtained results are presented in Table 8, where it can be observed that a total of 66 kW of exports is obtained using equally distributed OEs, whereas, on the other hand, a total of 74 kW is obtained using maximise exports, the latter is achieved by constraining the OE of Customer 136 up to 4 kW. Thus, it can be concluded that the maximise exports allocation technique can produce larger overall exports by constraining the OE of those customers that have higher sensitivity to exports and, more importantly, that the NN can capture these relationships.

Table 8. Allocation Technique: Equally Distributed OEs vs Maximise Exports

Active Customer	Equally distributed OEs Exports [kW]	Maximise Exports Exports [kW]
Customer 97	11	**14** (↑)
Customer 136	11	**4** (↓)
Customer 138	11	**14** (↑)
Customer 140	11	**14** (↑)
Customer 149	11	**14** (↑)
Customer 150	11	**14** (↑)
Total Exports	**66**	**74** (↑)

3.3 Key Remarks

This section proposes an alternative allocation technique (maximise exports) to the one presented in [2] (equally distributed). This allocation technique aims to achieve larger overall exports from active customers by limiting the exports of those customers that are more sensitive to voltage rise (i.e., those located further from the distribution transformer). Additionally, the steps to calculate model-free OEs considering both allocation techniques are presented. Besides, improvements on the NN structure are introduced to enhance the extrapolation capabilities of the proposed approach, which are critical when deploying the NN to calculate, for example, OEs. Hereinafter, all NNs will consider the structure, hyperparameters, and settings presented in Section 3.1 unless specified otherwise.

OEs for a single instance are calculated for the case of Jemena Substation B group 1, where it is shown that the maximise exports allocation technique can produce larger overall exports with respect to equally distributed OEs by constraining the OE of those customers that have higher sensitivity to exports and, more importantly, that the NN can capture these relationships.

4 Voltage Regulation Devices

Voltage regulation devices, such as on-load tap changers (OLTCs) and capacitor banks, are normally used by DNSPs in the operation of distribution networks to help maintaining voltages within statutory limits. The operation of such devices has an effect over the relationships seek by the Neural Network (NN), and, therefore, must be studied. Usually, these devices are employed to maintain a specific setting or configuration (hereinafter, BaU[2] operation). For instance, for the case of the OLTC, to maintain a specific voltage reference (e.g., $V_{ref} = 1.00\ pu$) at the secondary side of the primary substation.

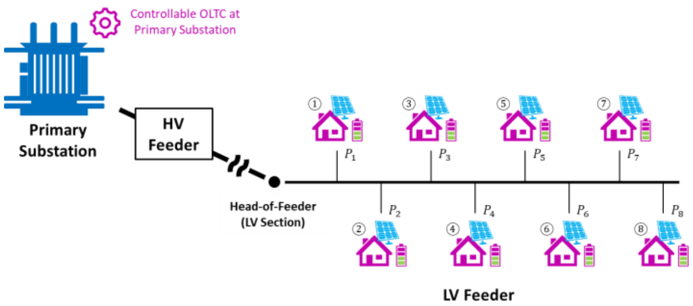

Figure 20. Schematic of an OLTC at the primary substation

In fact, BaU operation of voltage regulation devices can be observed in the historical smart meter data provided for this project. To illustrate this, the case of Jemena Substation B, group 1, is used. For further details on the corresponding historical smart meter data, please refer to [2]. Figure 21 presents a zoom into the test actual voltage values for Jemena Substation B group 1. Specifically, between instance 3,600 and 3,888, which is equivalent to 1 day at 5 minutes resolution, highlighting the BaU operation of a voltage regulation device in orange circles.

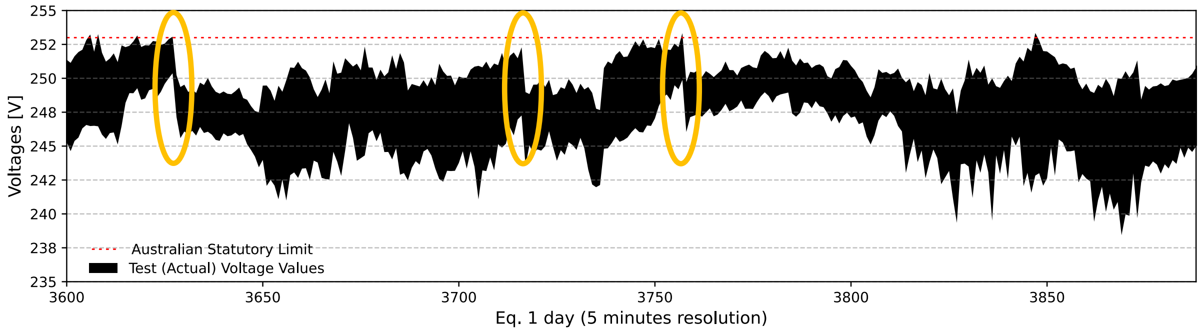

Figure 21. BaU Operation of Voltage Regulation Devices in the Historical Data

From the studies carried out in [2], it can be noted that BaU operation of voltage regulation devices is already being captured by the proposed approach. Hence, Figure 22 shows the voltage calculations obtained in [2] for the same period presented in Figure 21, when using the latest 3 weeks of historical data to produce the NN. It can be noted that the proposed approach can accurately calculate voltages when BaU operation of voltage regulation devices occur. This is because the effects of BaU operation of voltage regulation devices over the relationships seek by the NN are represented in the historical data used to train the NN and, thus, are being captured.

[2] Business-as-Usual

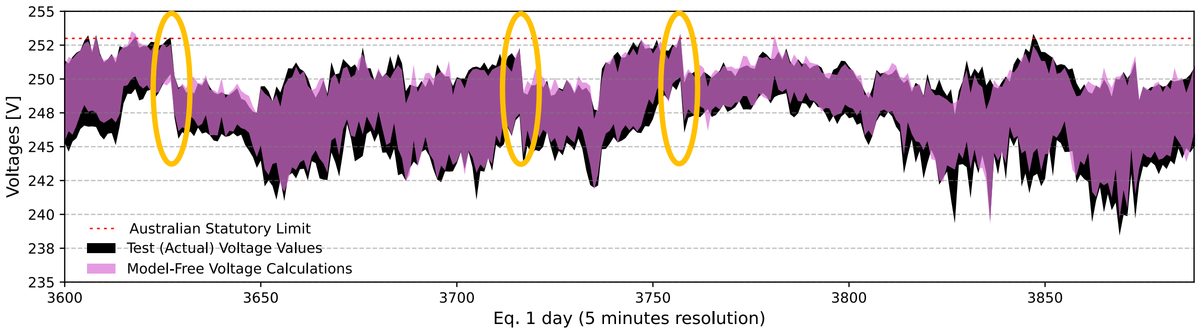

Figure 22. Model-Free Voltage Calculations under BaU Operation of Voltage Regulation Devices

However, under certain circumstances, it could be convenient for DNSPs to carry out temporal or permanent changes to the corresponding settings. For instance, for the case of the OLTC, to decrease the voltage reference at the secondary side of the primary substation (e.g., from $V_{ref} = 1.00\,pu$ to $V_{ref} = 0.98\,pu$) if customers' voltages are getting close to the upper statutory limit due to higher PV penetrations. Such new setting will naturally have an effect over the voltages at the head of the LV network and, consequently, over customers' voltages. However, as the NN was trained to cater for the previous setting, the NN will continue to calculate voltages as before. There are two options to capture new settings. These are detailed below.

1. (*Recommended*) Retrain the NN
   Once a change of setting occurs, the NN will continue to calculate voltages as before. This can be easily detected in a post-processing process as the NN will consistently calculate erroneous voltage values for all customers. If such a change is detected, sufficient historical smart meter data (i.e., an equivalent to 3 weeks at 5 minutes resolution) that represents the new setting must be collected. Once sufficient historical smart meter data is available, the NN can be updated by using the same NN structure as before and the latest historical data.

2. Use revised calculated voltages
   Estimate the effect of the new setting over head of the LV network voltages (hereinafter, ΔV) and incorporate it using revised calculated voltages as defined in (1).

$$\widehat{\boldsymbol{V}_{rev}} = \boldsymbol{V}_{calc} + \Delta\boldsymbol{V} \tag{1}$$

Changes in voltage regulation devices can be either temporal or permanent. If the change is expected to be of short duration (e.g., hours, days) option 2 would be recommended. However, if the change is expected to be of medium to long duration or even permanent, Option 2 should be implemented until sufficient historical smart meter data that represents the new setting is available. Once available, the NN should be updated using the same NN structure as before and the latest historical data as per Option 1.

Finally, DNSPs can implement other methods to help maintaining customers' voltages within statutory limits, which will naturally have an effect over the relationships seek by the NN. For instance, by using inverter functions such as volt-var. In fact, as shown in [1], the presence of the volt-var function can be observed in the historical data of some customers of Jemena and United Energy. Hence, as such interactions are being represented in the historical data, its effects are already being captured by the proposed approach (as per results in [2]). Furthermore, other control techniques that could be used by DNSPs to control customers' voltages are line drop compensation or voltage thresholds for certain critical customers (e.g., customers that are located further from the distribution transformer). In this context, the same principle applies, i.e., the effect of such voltage control techniques will be captured as long as the logic behind is represented in the historical data used to train the NN. Any change in such logic will change the underlying relationships of the corresponding LV network and, therefore, required to update the NN.

4.1 Revising Calculated Voltages

To illustrate the implementation of revised calculated voltages, the case of United Energy Substation B is used. United Energy Substation B corresponds to a LV network comprised of 2 LV circuits with a total of 77 customers (74 single-phase customers and 3 three-phase customers). For further details on the historical data, please refer to [2]. Specifically, the case of group 16 when using 3 weeks for training is considered. The training data set of group 16 spans the period between 17-12-2020 and 08-01-2021. A total of 2 single-phase customers are removed from the historical data once the data pipeline defined in Section 2 is applied.

Once trained, this NN is then used to calculate operating envelopes (OEs) for the first instance at noon of the corresponding test data set as in Section 3.2. In this context, equally distributed operating envelopes are considered, and half of the single-phase customers detected with PV systems are randomly allocated as active customers (i.e., 10 active customers). Active customers are considered to operate with a unity power factor. Finally, a change in the voltage reference of the OLTC at the primary substation with a $\Delta V = -5\ V$ is considered.

The first step is to define the most suitable NN for the studied LV network. This NN is defined as mentioned in Section 3.1. Once trained, the performance of the NN is evaluated using the next 6,048 instances (i.e., an equivalent to 3 weeks at 5 minutes resolution) of the historical data (test data set). The model-free voltage calculations results obtained in the test data set are presented in Table 9, Figure 23, and Figure 24.

Table 9. Model-Free Voltage Calculations Results

Group	Training Period	Training Initial Instance	Training Final Instance Test Initial Instance	Test Final Instance	Season	RMSE [V]	Av. Dev. [V]	Max. Dev. [V]
16	3 weeks	17-12-2020	08-01-2021	31-01-2021	Summer	1.16	0.90	8.14

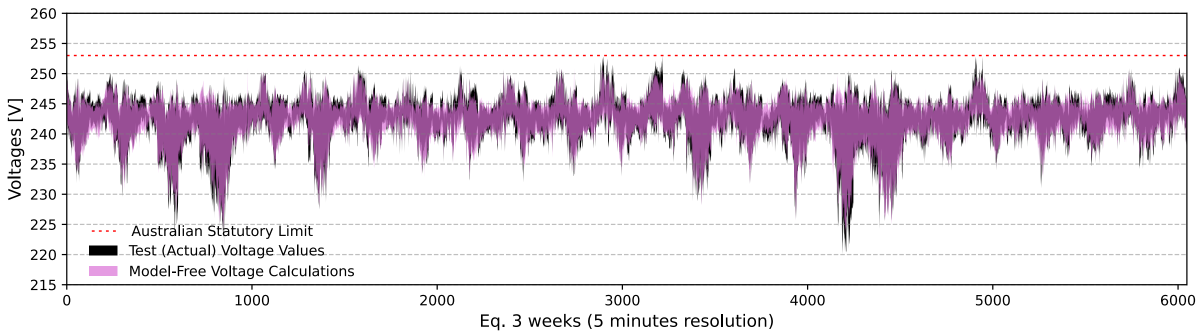

Figure 23. Time-Series Voltage Calculations

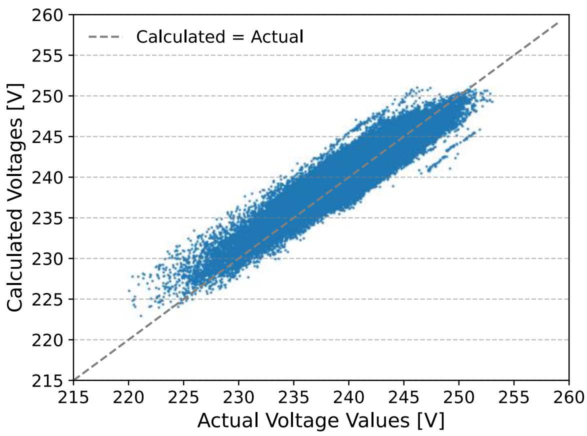

Figure 24. Calculated vs Actual Voltage Values

As observed, the proposed approach can capture the physics and produce accurate voltage calculations in the test data set, achieving an average deviation of 0.9 V. Once trained, this NN is used to calculate equally distributed OEs[3] for the first instance at noon of the test set following the steps presented in Table 4 (i.e., as in Section 3.2). OEs calculated for BaU operation of voltage regulation devices corresponds to those OEs calculated by the NN (which has not been updated to cater for the new setting with $\Delta V = -5\,V$), whereas, for the new setting, OEs are obtained using revised calculated voltages. Essentially, the voltages for the new setting will correspond to the application of equation (1) over the voltages calculated for BaU operation at each iteration. The obtained OEs are presented in Table 10, where maximum voltage among all customers for each export scenario are detailed. Please note that green values are to represent the respective OE, whereas red values are to represent the first export level that exceed voltage limits. As expected, OEs for the new setting are larger than those obtained for the previous situation, as the voltage reference has been decreased creating more room for exports.

Table 10. Operating Envelopes Calculations, BaU vs New Setting

OE Calculation (BaU)		OE Calculation (new setting)	
Exports [kW]	**Maximum Voltage [V]**	**Exports [kW]**	**Maximum Voltage [V]**
0	245.31	0	240.31
1	246.70	1	241.70
2	248.08	2	243.08
3	249.47	3	244.47
4	250.87	4	245.87
5	**252.27**	5	247.27
6	**253.68**	6	248.68
7	255.08	7	250.08
8	256.47	8	251.47
9	257.86	**9**	**252.86**
10	259.22	**10**	**254.22**

It is important to note that even though upstream setting changes can be directly incorporated in the proposed approach by using revised calculated voltages, it is always recommended to collect sufficient historical data for the new setting and update the corresponding NN as this will correspond to a more accurate representation of the actual situation (new setting).

[3] For simplicity, an equally distributed allocation technique is considered. However, it is important to note that the same logic applies when using maximise exports or any other allocation technique, i.e., the voltages for the new setting will correspond to the application of equation (1) over the voltages calculated for BaU operation of voltage regulation devices at each iteration.

4.2 Key Remarks

This section determines the extent to which the proposed approach can cater for voltage regulation devices. In this context, it is identified that BaU operation of voltage regulation devices (i.e., maintain specific setting) is already represented in the historical data used in this project and, thus, as per the results in [2], already being captured by the proposed approach. On the other hand, it is observed that when a change of setting occurs, the NN will continue to calculate voltages as before (as it was trained to cater for the previous setting). To overcome this issue, two options are proposed: 1) collect sufficient historical data representative of the new setting and then update the corresponding NN by using the same NN structure and the latest data; 2) incorporate the new setting by using revised calculated voltages (1). The implementation of revised calculated voltages in the determination of OEs is illustrated using the case of United Energy Substation A, where it can be observed that upstream changes can be directly incorporated in the calculation of OEs.

5 Partial Smart Meter Data Availability

So far, it has been considered that smart meter data from all customers connected to the studied distribution transformer is available for DNSPs. In this context, it has been shown that the proposed approach can carry out accurate multi-LV-circuit voltage calculations for each of the studied distribution transformers [2]. However, in Victoria, even though a full roll-out of smart meters is observed, the smart meter data from some customers might not be available for DNSPs. This is the case, for example, of some commercial and industrial (C&I) customers whose data is managed by third parties and, hence, unavailable for DNSPs. Therefore, it is important to assess if the proposed model-free voltage calculations can be used when smart meter data from all customers is not available as this could be the scenario faced by DNSPs when implementing the proposed approach in Victoria.

To illustrate this point, AusNet Site A is used. AusNet Site A corresponds to a LV network with a single LV circuit that supplies a total of 28 customers, 26 single-phase customers and 2 three-phase customers. Let us consider two scenarios, full and partial observability. For the scenario of full observability, it is considered that the smart meter data from all customers is available to the DNSP, whereas, on the other hand, for the scenario of partial observability, both three-phase customers (A1 and A24) are considered as C&I customers whose smart meter data is managed by a third party. A schematic of this is presented in Figure 25.

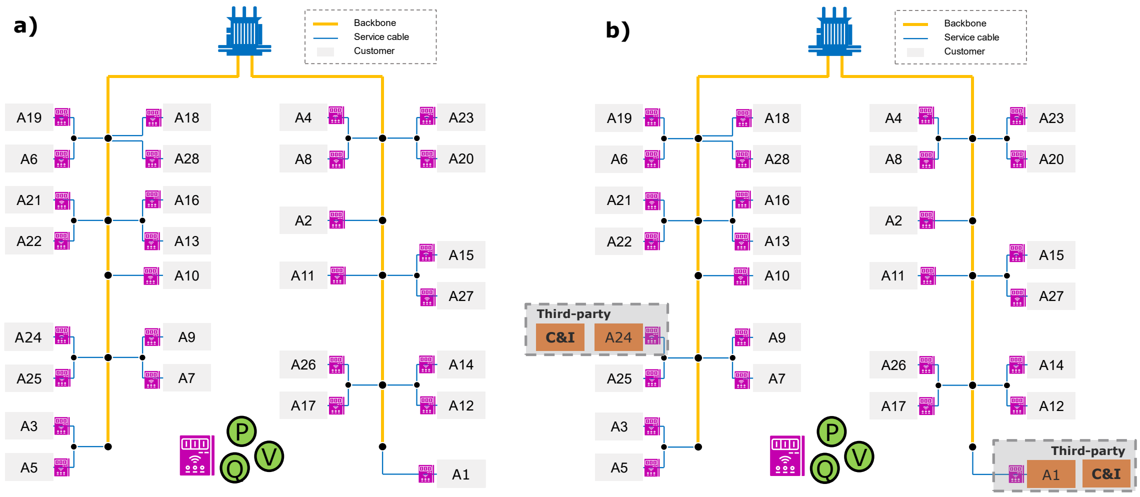

Figure 25. Schematic of Full (a) vs Partial (b) Observability

In this context, if smart meter data from all customers is available to the DNSP (i.e., full observability) the historical smart meter data from Site A will be comprised of 64 features (i.e., per-phase active and reactive power) and 32 targets (i.e., per-phase voltages). However, on the other hand, if the smart meter data from customers A1 and A2 is managed by a third party (partial observability), the historical smart meter data from Site A will be comprised 12 fewer features (i.e., per-phase active and reactive power of the three phases of A1 and A24) and 6 fewer targets (i.e., per-phase voltages of the three phases of A1 and A24). This in turns means that voltage calculations for customers A1 and A24 cannot be carried out as its corresponding relationships are not available for the DNSP and, therefore, not included in the historical data used to train the NN. NN inputs and outputs for both scenarios are detailed below.

- **Full Observability**
  - Inputs: $64 = 2 \times (26 + 2 \times 3)$
  - Outputs: $32 = 26 + 2 \times 3$
- **Partial Observability**
  - Inputs: $52 = 2 \times (26)$
  - Outputs: 26

It is important to note that an accuracy decrease is expected for the scenario of partial observability as customers A1 and A24 will have an effect over LV network voltages that is not being represented in the historical data used to train the NN. The magnitude of such accuracy decrease will be associated to the location and operation of such customers.

5.1 Preliminary Analyses

To illustrate the extent to which the proposed approach can cater for partial smart meter data availability, two scenarios are assessed, full and partial observability. For the scenario of full observability, data from all customers is available to the DNSP. For the scenario of partial observability, it is assumed that all three-phase customers are C&I customers whose data is managed by third parties and, consequently, unavailable to the DNSP. Hence, only the data from single-phase customers is available and, therefore, used to train and assess the corresponding NN. Specifically, the case of the two distribution transformers with the largest number of three-phase customers are studied, i.e., Jemena Substation B and United Energy Substation C.

It is important to highlight than even though these substations might represent worst-case scenarios from the perspective of unavailable data (i.e., larger number of three-phase customers which are assumed unavailable for the case of partial observability), the analyses presented in this section are only preliminary as the demands of the three-phase customers of these substations might not necessarily represent the demand of C&I customers. Therefore, further analyses must be carried out to fully determine the extent to which the proposed approach can cater for partial smart meter data availability in the context of Victoria. This will be analysed in a later stage of the project.

5.1.1 Jemena Substation B

Jemena Substation B corresponds to a LV network comprised of 4 LV circuits with a total of 155 customers (116 single-phase customers and 39 three-phase customers). For further details on the historical data, please refer to [2]. Specifically, the case of group 14 when using 3 weeks for training is considered. The training data set of group 14 spans the period between 10-08-2021 and 04-09-2021. A total of 6 single-phase customers, 1 phase of 2 three-phase customers and a three-phase customer are removed from the historical data once the offline data pipeline defined in Section 2 is applied.

Two scenarios are considered, full and partial observability. As mentioned above, for the partial observability scenario it is considered that only the smart meter data from single-phase customers is available. Thus, NN inputs and outputs are presented below.

- **Full Observability (FO)**
  - Inputs[4]: $444 = 2 \times (110 + 2 \times 2 + 36 \times 3)$
  - Outputs: $222 = 110 + 2 \times 2 + 36 \times 3$
  - Training data points: $4{,}027{,}968 = 6{,}048 \times (444 + 222)$
- **Partial Observability (PO)**
  - Inputs: $220 = 2 \times (110)$
  - Outputs: 110
  - Training data points: $1{,}995{,}840 = 6{,}048 \times (220 + 110)$

As observed, the scenario of partial observability considers unavailable around 50 % of the data points used in the scenario of full observability as this data assumed unavailable for the DNSP. The accuracy of both NNs is assessed using the test data of group 14 (i.e., the following eq. to 3 weeks in the historical data). Results for both scenarios are presented in Table 11, Figure 26 and Figure 27.

[4] Please note that after the offline data pipeline is applied, the data for Jemena Substation B group 14 is comprised of 110 single-phase customers, 2 two-phase customers, and 36 three-phase customers. The 2 two-phase customers appear as 1 phase from 2 different three-phase customers have been removed as part of the offline data pipeline.

Table 11. Model-Free Voltage Calculations Results

Group	Training Period	Training Initial Instance	Training Final Instance Test Initial Instance	Test Final Instance	Case	RMSE [V]	Av. Dev. [V]	Max. Dev. [V]
14	3 weeks	10-08-2021	04-09-2021	29-09-2021	FO	0.55	0.43	9.70
					PO	0.97	0.76	9.88

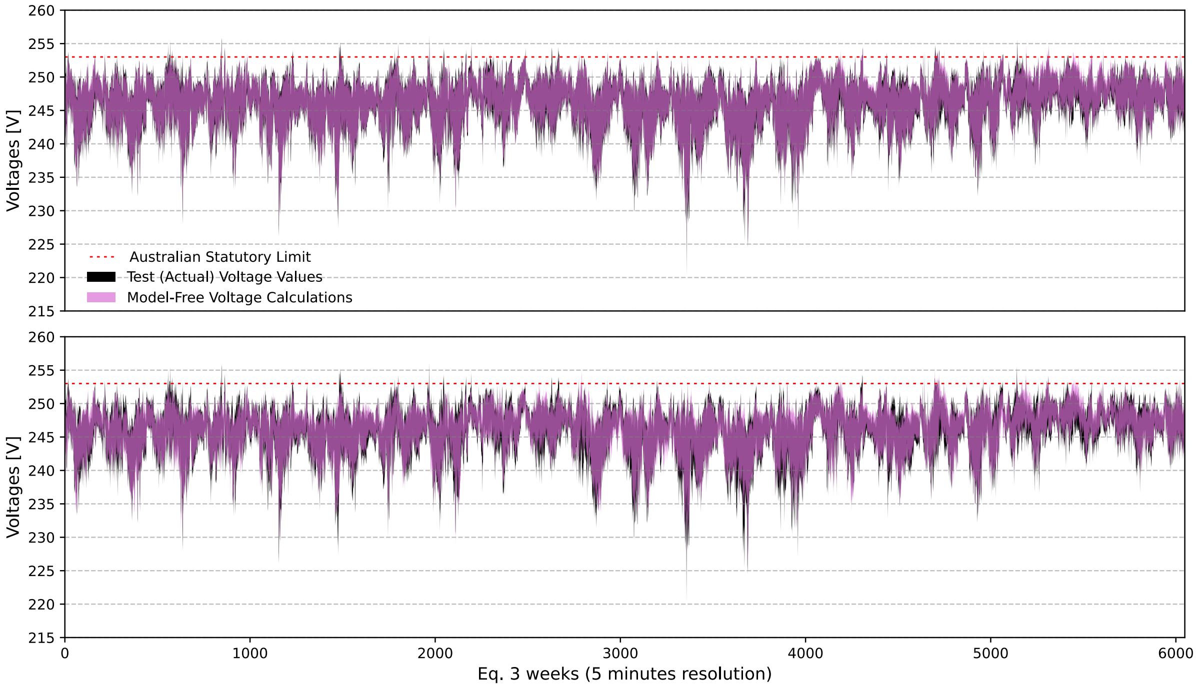

Figure 26. Time-Series Model-Free Voltage Calculations, Full (upper) vs Partial (lower) Observability

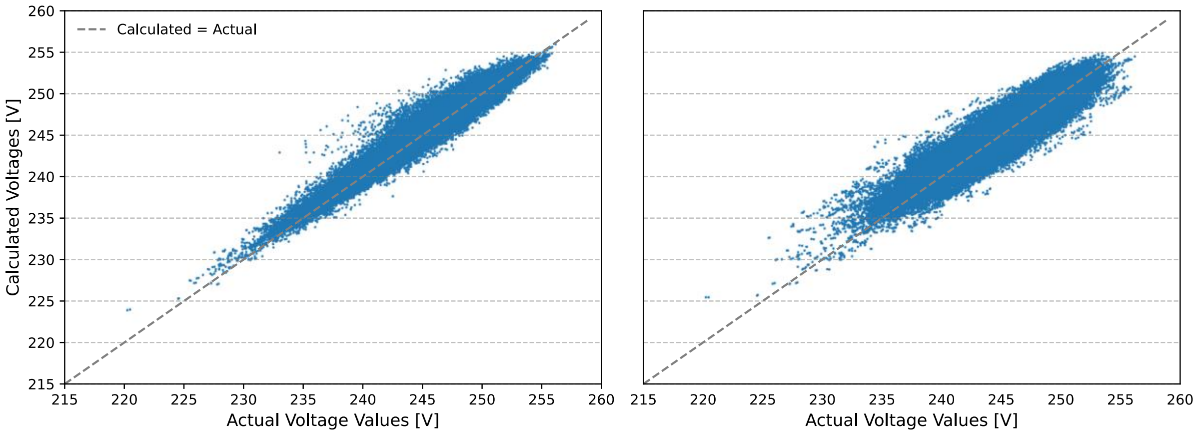

Figure 27. Calculated vs Actual Voltage Values, Full (left) vs Partial (right) Observability

As observed in Table 11, Figure 26 and Figure 27, an accuracy decrease is observed for partial observability, as expected. In fact, Av. Dev. decreases from 0.43 V to 0.76 V, i.e., approx. 77%. However, it can be noted that accurate voltage calculations are obtained in both scenarios, achieving an average deviation below 0.8 V. Thus, it can be noted that, for this case, the proposed approach can produce accurate voltage calculations even when around 50% of the data is not available (partial observability).

Even though accurate model-free voltage calculations can be obtained in both scenarios, the final application of the proposed model-free voltage calculations is to be used to calculate, for instance, operating envelopes (OEs). Thus, OEs for the first instance at noon of the test set are calculated as in Section 3.2 using both NNs. In this context, equally distributed OEs are considered, and half of the single-phase customers detected with PV systems are randomly allocated as active customers (i.e., 6 active customers). Active customers are considered to operate with a unity power factor. Results are presented in Table 12.

Table 12. Operating Envelopes Calculations, Full vs Partial Observability

Full Observability		Partial Observability	
Exports [kW]	Maximum Voltage [V]	Exports [kW]	Maximum Voltage [V]
0	250.14	0	249.21
1	250.51	1	249.72
2	250.88	2	250.33
3	251.25	3	250.92
4	251.62	4	251.50
5	251.99	5	252.08
6	252.36	6	252.65
7	252.72	7	253.21
8	253.08	8	253.77

As observed in Table 12, OE calculated for full observability is 7 kW, whereas OE calculated for partial observability is 6 kW. Such difference is expected as the NN trained for partial observability is less accurate and there will be differences in the calculations produced by both NNs. However, voltage calculations for both NNs for 6 kW and 7 kW of exports are presented in Figure 28, where it can be noted that the voltage calculations, and consequently, the OEs produced in both scenarios are consistent.

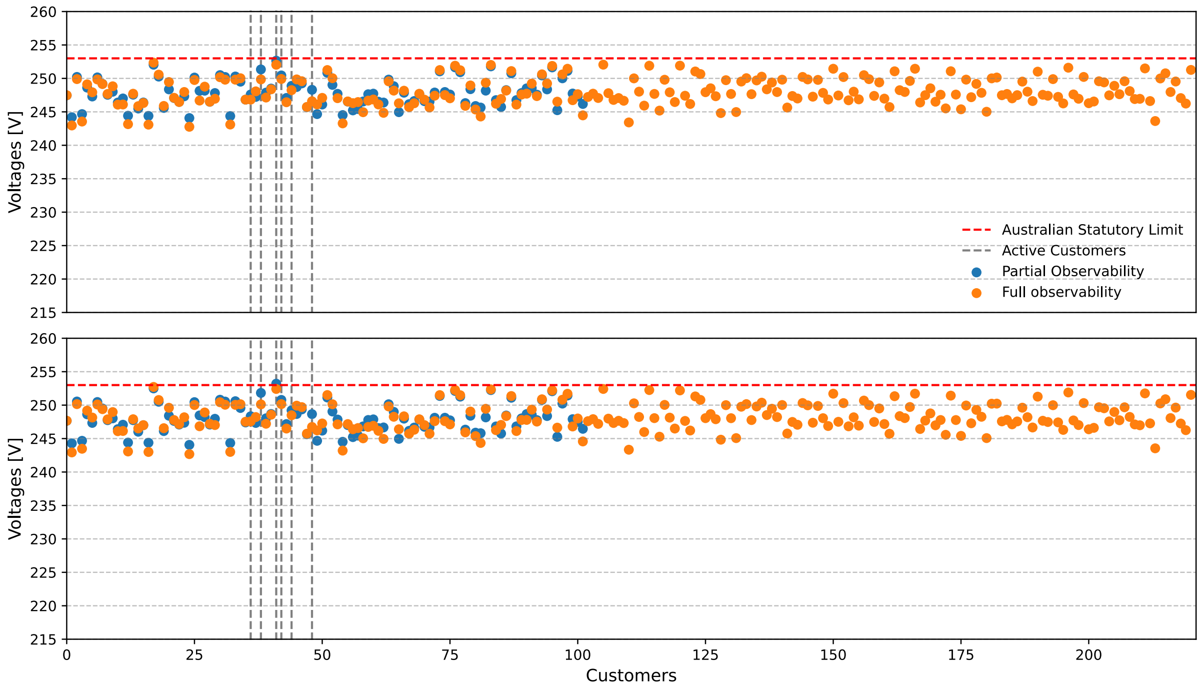

Figure 28. Model-Free Voltage Calculations per customer at 6 kW (upper) and 7 kW (lower) of exports, Full vs Partial Observability

5.1.2 United Energy Substation C

United Energy Substation C corresponds to a LV network comprised of 3 LV circuits with a total of 59 customers (40 single-phase customers and 19 three-phase customers). For further details on the historical data, please refer to [2]. Specifically, the case of group 9 when using 3 weeks for training is considered. The training data set of group 9 spans the period between 07-12-2020 and 13-01-2021. A total of 1 single-phase customer and 1 three-phase customer are removed from the historical data once the offline data pipeline defined in Section 2 is applied.

Two scenarios are considered, full and partial observability. For the partial observability scenario, it is considered that only the smart meter data from single-phase customers is available. Thus, NN inputs and outputs are presented below.

- **Full Observability (FO)**
  - Inputs: $186 = 2 \times (39 + 18 \times 3)$
  - Outputs: $93 = 39 + 18 \times 3$
  - Training data points: $1{,}687{,}392 = 6{,}048 \times (186 + 93)$
- **Partial Observability (PO)**
  - Inputs: $78 = 2 \times (39)$
  - Outputs: 39
  - Training data points: $707, 616 = 6{,}048 \times (78 + 39)$

As observed, the scenario of partial observability considers unavailable around 60 % of the data points used in the scenario of full observability as this data is assumed unavailable for the DNSP. The accuracy of both NNs is assessed using the test data of group 9 (i.e., the following eq. to 3 weeks in the historical data). Results for both scenarios are presented in Table 13, Figure 29 and Figure 30.

Table 13. Model-Free Voltage Calculations Results

Group	Training Period	Training Initial Instance	Training Final Instance Test Initial Instance	Test Final Instance	Case	RMSE [V]	Av. Dev. [V]	Max. Dev. [V]
9	3 weeks	07-12-2020	13-01-2021	18-02-2021	FO	1.56	1.19	11.81
					PO	1.63	1.22	13.08

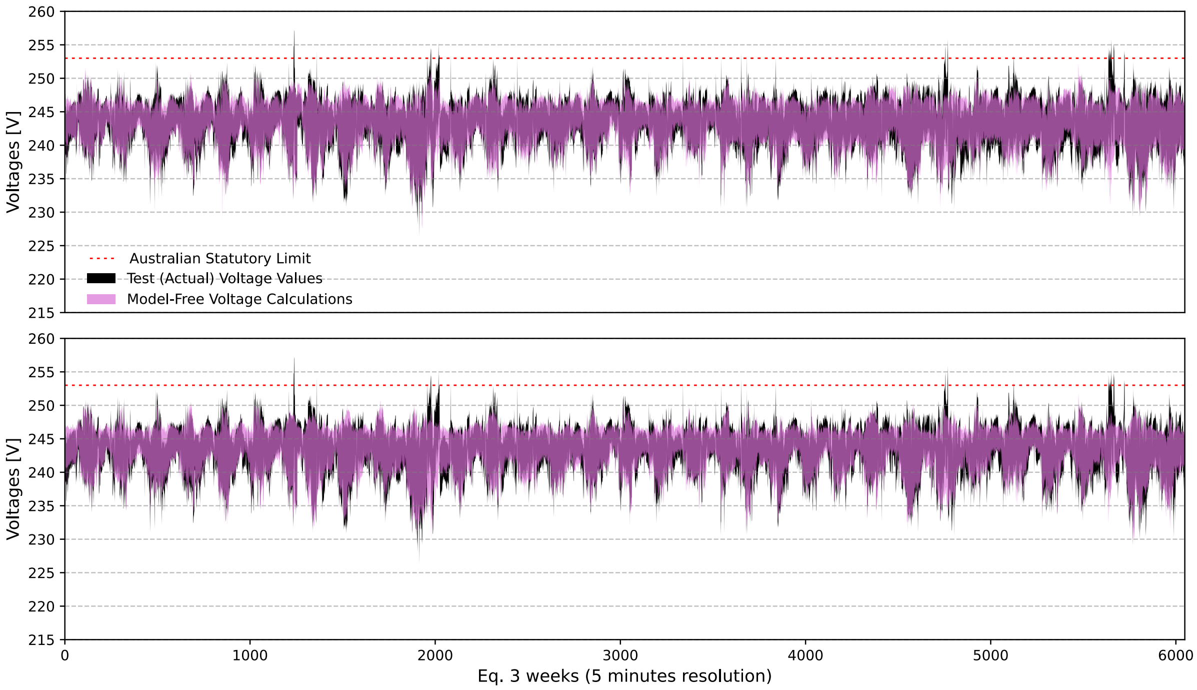

Figure 29. Time-Series Model-Free Voltage Calculations, Full (upper) vs Partial (lower) Observability

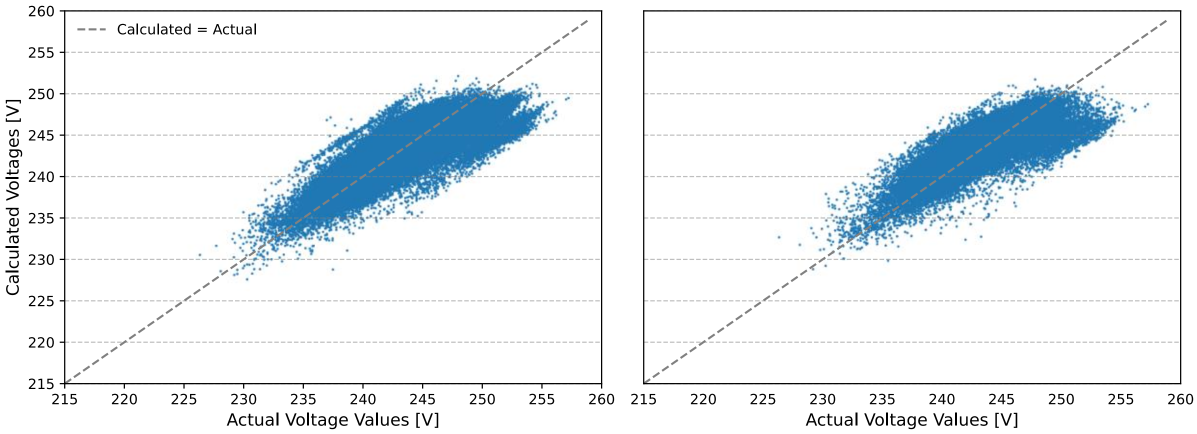

Figure 30. Calculated vs Actual Voltage Values, Full (left) vs Partial (right) Observability

As presented in Table 13, Figure 29 and Figure 30, an accuracy decrease is observed for partial observability, as expected. In fact, Av. Dev. decreases from 1.19 V to 1.22 V. However, it can be noted that accurate voltage calculations are obtained in both scenarios, achieving an average deviation below 1.3 V. Thus, it can be noted that, for this case, the proposed approach can produce accurate voltage calculations even when around 60% of the data is not available.

Using both NNs, Operating envelopes (OEs) for the first instance at noon of the test set are calculated as in Section 3.2. In this context, equally distributed OEs are considered, and half of the single-phase customers detected with PV systems are randomly allocated as active customers (i.e., 4 active customers). Active customers are considered to operate with a unity power factor. Results are presented in Table 14.

Table 14. Operating Envelopes Calculations, Full vs Partial Observability

Full Observability		Partial Observability	
Exports [kW]	**Maximum Voltage [V]**	**Exports [kW]**	**Maximum Voltage [V]**
0	246.07	0	245.86
1	246.17	1	245.91
2	246.26	2	246.17
3	246.61	3	246.66
4	247.16	4	247.14
5	247.89	5	247.89
6	248.61	6	248.75
7	249.32	7	249.60
8	250.03	8	250.44
9	250.75	9	251.26
10	251.47	10	252.06
11	252.17	**11**	**252.85**
12	**252.87**	**12**	**253.63**
13	**253.56**	13	254.39

As observed in Table 14, OE calculated for full observability is 12 kW, whereas OE calculated for partial observability is 11 kW. This is expected as the NN trained for partial observability is less accurate and there will be differences in the calculations produced by both NNs. However, voltage calculations for both NNs for 11 kW and 12 kW of exports are presented in Figure 31, where it can be noted that the voltage calculations, and consequently, the OEs produced in both scenarios are consistent.

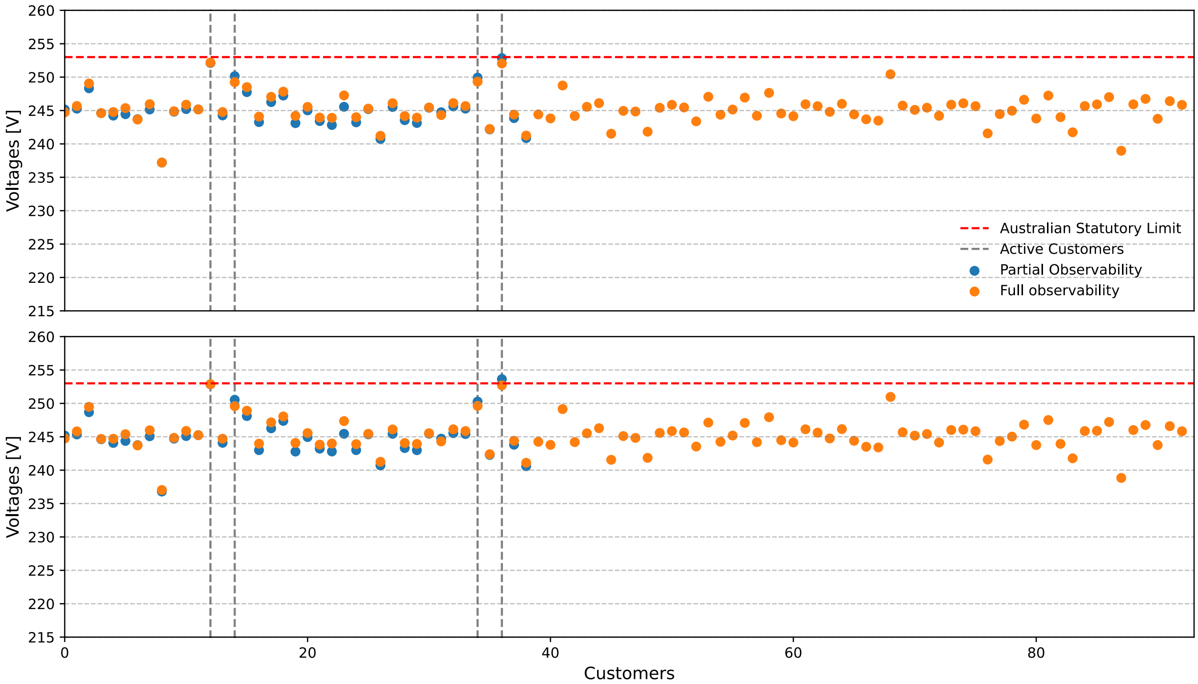

Figure 31. Model-Free Voltage Calculations per customer at 11 kW (upper) and 12 kW (lower) of exports, Full vs Partial Observability

5.2 Key Remarks

In Victoria, even though a full roll-out of smart meters is observed, the smart meter data from some customers might not be available for DNSPs (e.g., commercial and industrial [C&I] customers whose data is managed by third parties). Hence, it is important to assess if the proposed model-free voltage calculations can be used when smart meter data from all customers is not available as this could be the scenario faced by DNSPs when implementing the proposed approach in Victoria. Thus, preliminary analyses assess two scenarios: 1) full observability, i.e., data from all customers is available; 2) partial observability, i.e., all three-phase customers are assumed to be C&I customers whose data is managed by third parties and, thus, unavailable for the DNSP. Specifically, the case of United Energy Substation C and Jemena Substation B are analysed, where it can be observed that it is possible to obtain accurate voltage calculations as well as consistent operating envelopes when only 40-50% of the customers are considered (in this case, three-phase residential customers were removed as a proxy of C&I customers).

Although results in this section are promising, the analyses presented are only preliminary as the demand of three-phase residential customers of these substations might not necessarily represent the demand of C&I customers. Therefore, further analyses must be carried out to fully determine the extent to which the proposed approach can cater for partial smart meter data availability in the context of Victoria. This will be analysed in a later stage of the project.

6 Next Steps

The next steps for the project are as follows.

- An architecture for DNSPs to operationalise the proposed approach to calculate operating envelopes (OEs) will be created. This will provide the foundations for DNSPs in the case of further actions are taken to implement the proposed approach within its systems after the completion of this project.

- It has been observed in the historical data provided for this project that some customers can present values that correspond to a constant power factor (either 1 or different) or, simply, constant values throughout the entire training data set which can also create issues when the Neural Network (NN) is being deployed as the corresponding relationships are not being represented in the historical data. In other words, there is no variety enough in the data for the NN to capture those relationships and, thus, it produces wrong voltage calculations when deployed considering values that are different to those observed during training (e.g., when considering a what-if scenario). This is currently being investigated by The University of Melbourne.

- So far, the proposed model-free OEs calculation approach has considered only voltage constraints and the limitations of customers' connection points. However, further constraints, such as transformer or lines thermal ratings, could also limit the OEs of active customers and, thus, must be incorporated into the model-free OEs calculation approach.

- Further analyses will be carried out to fully determine the extent to which the proposed approach can cater for partial smart meter data availability in the context of Victoria, and the case of partial smart meter deployment (as in other parts of Australia, for example) will also be analysed.

7 Conclusions

Voltage calculations are key for Distribution Network Service Providers (DNSPs) to adequately operate and plan their low voltage (LV) networks, particularly in the presence of high DER penetrations. However, voltage calculations would normally require power flow analyses and, consequently, detailed three-phase LV network models, which, in practice, are not readily available for DNSPs. Taking advantage of smart meter data, available to all DNSPs in the State of Victoria, this project aims to demonstrate that is possible to capture the physics of three-phase LV circuits and create an electrical model-free approach to calculate voltages. These model-free voltage calculations can be then used to estimate operating envelopes (OEs), assess connection requests, or to carry out hosting capacity assessments, enabling DNSPs to bypass the costly, time-consuming, and error-prone process of producing and validating electrical models.

The report at hand builds on previous reports [1], [2] and presents improvements and updates on the project. Specifically, it presents improvements to the offline data pipeline to overcome the data issues that were found to decrease the accuracy in a reduced number of cases in [2]. Additionally, the report at hand presents an alternative allocation technique to calculate OEs, which, differently than the one defined in [2] (i.e., equally distributed OEs), aims to maximise the overall exports from active customers. Furthermore, this report discusses how to adapt the methodology if the settings of voltage regulation devices are changed. Finally, the report presents initial findings of cases with partial smart meter data availability. The conclusions of this report can be summarised as follows.

Offline Data Pipeline Improvements
Improvements to the offline data pipeline are presented due to data issues involving customers with extremely low values of demand (zero or a few Watts) throughout most of the training data set. These customers caused a decrease in accuracy in a limited number of cases presented in [2] which, in the test data set, had normal values of demand (kilowatts). This is because the test values were very far from what the NN has been trained with. Essentially, any feature that does not provide meaningful information to the NN (i.e., variety enough for the NN to capture the underlying relationships) can create wrong voltage calculations when deploying the NN to calculate voltages. Specifically, if adequate values for such features are considered. To address this, an additional step to clear these issues is incorporated into the offline data pipeline. It is shown that by implementing this new step, the proposed approach can produce more accurate voltage calculations for the customers with normal demand as this makes it possible for the NN to capture the underlying relationships, enabling the proposed approach to overcome the data issues found in [2].

Operating Envelopes Allocation Technique: Maximise Exports
A new allocation technique called *maximise exports* is presented. This allocation technique differs from the one presented previously in [2] (called equally distributed). Instead of allocating the same OE to everyone, this new technique aims to achieve larger overall/total exports from all active customers by limiting the exports of those customers that are more sensitive to voltage rise (i.e., those located further from the distribution transformer) and, thus, facilitating larger OEs to those less sensitive. It is shown that the maximise exports allocation technique can produce larger overall exports with respect to equally distributed OEs by constraining the OE of those customers that have higher sensitivity to exports and, more importantly, that the NN can capture these relationships.

Voltage Regulation Devices
The business as usual (BaU) operation of voltage regulation devices (i.e., when specific settings are kept) is captured by the proposed approach, as demonstrated in [2]. However, if settings change after the NN was trained, the NN will continue to calculate voltages as before (as it was trained to cater for the previous setting). To overcome this issue, two options are proposed: to retrain the NN with the latest data that incorporates such a change or by revising calculated values considering the new settings. It is shown that upstream changes can be directly incorporated into the proposed approach by revising calculated voltages.

It is important to note that even though upstream setting changes can be directly incorporated in the proposed approach by using revised calculated voltages, it is always recommended to collect sufficient historical data for the new setting and update the corresponding NN as this will correspond to a more accurate representation of the actual situation (new setting).

Partial Smart Meter Data Availability
In Victoria, even though all customers have smart meters, the smart meter data from some of them might not be available to DNSPs (e.g., commercial and industrial [C&I] customers whose data is managed by third parties). Preliminary analyses presented in this report show that it is possible to obtain accurate voltage calculations as well as consistent operating envelopes when only 40-50% of the customers are considered (in this case, three-phase residential customers were removed as a proxy of C&I customers).

Although results in this section are promising, the analyses presented are only preliminary as the demand of three-phase residential customers of these substations might not necessarily represent the demand of C&I customers. Therefore, further analyses must be carried out to fully determine the extent to which the proposed approach can cater for partial smart meter data availability in the context of Victoria. This will be analysed in a later stage of the project.

References

[1] V. Bassi, D. Jaglal, L. F. Ochoa, T. Alpcan, and C. Leckie, “Deliverable 0 – Concept, Smart Meter Data, and Initial Findings”, February 2022. PDF available on ResearchGate

[2] V. Bassi, D. Jaglal, L. F. Ochoa, T. Alpcan, and C. Leckie, “Deliverables 1-2-3a: Model-Free Voltage Calculations and Operating Envelopes”, July 2022. PDF available on ResearchGate

[3] A. Navarro-Espinosa and L. F. Ochoa, “Probabilistic Impact Assessment of Low Carbon Technologies in LV Distribution Systems,” in IEEE Transactions on Power Systems, vol. 31, no. 3, pp. 2192-2203, May 2016. PDF available on ResearchGate

[4] K. Petrou, A. T. Procopiou, L. F. Ochoa, T. Langstaff and J. Theunissen, “Impacts of Price-led Operation of Residential Storage on Distribution Networks: An Australian Case Study,” 2019 IEEE Milan PowerTech, 2019, pp. 1-6. PDF available on ResearchGate

[5] J. Quirós-Trotós, L. F. Ochoa, S. W. Alnaser and T. Butler, “Control of EV Charging Points for Thermal and Voltage Management of LV Networks,” in IEEE Transactions on Power Systems, vol. 31, no. 4, pp. 3028-3039, July 2016. PDF available on ResearchGate

[6] V. Bassi, L. F. Ochoa, T. Alpcan and C. Leckie, "Electrical Model-Free Voltage Calculations Using Neural Networks and Smart Meter Data," in IEEE Transactions on Smart Grid. PDF available on ResearchGate