Outcomes Report

PROJECT SHIELD

Project SHIELD

Synchronising Heterogenous Information to Evaluate Limits for DNSPs

APRIL 2022

Contents

Disclaimer and Acknowledgements

This project received funding from ARENA as part of ARENA's Advancing Renewables Program.

Definitions

AEMC – Australian Energy Market Commission

AEMO – Australian Energy Market Operator

AER – Australian Energy Regulator

AMI – Advanced Metering Infrastructure – an integrated system of smart meters, data management systems and communication networks that enable 2-way communication between the utilities and the customers

ARENA – Australian Renewable Energy Agency

BTM – Behind the meter (customer level)

CEFC – Clean Energy Finance Corporation

DEIP – Distributed Energy Integration Program – an initiative that has resulted in the energy industry working in unison to enhance the potential of consumer-owned energy resources. Collaboration of energy peak bodies, market authorities, industry associations and consumer associations

DER – Distributed Energy Resources – the name given to renewable-energy units or systems that are commonly located at houses or businesses to provide them with power. Includes rooftop solar PV units, battery storage, thermal energy storage, electric vehicles and chargers, smart meters and home energy management technologies.

DNSP – Distributed Network Service Providers – e.g. Energex, Ergon Energy Network, Essential Energy

DOE – A dynamic operating envelope is a principled allocation that maximises additional export (generation) and import (load) the network can accept at any one time without breaching any of the specified technical and operational limits.

DPV – Distributed Photovoltaics is a type of DER that converts the sun's rays to electricity. Includes all grid-connected solar that is NOT centrally controlled

ENA – Energy Networks Australia

ESB – Energy Security Board

HEMS – Home Energy Management Systems

Hosting Capacity – The amount of DERs that can be connected to a distribution network while still ensuring that the network remains within its technical limits. The real and reactive power contributions from DER that can be imported or exported into the electricity grid without breaching the physical or operational limits of the electricity distribution network

IoT – Internet of Things

MDP – Metering Data Provider

NEM – National Electricity Market

NMI – National Meter Identifier – unique 10- or 11-character reference that is associated with the electricity connection point at your home or business

OEM – Original Equipment Manufacturer

PV – photovoltaic

Executive Summary

'Alea iacta est' – the dice has been rolled and demonstrates there is no turning back from the transition to a decentralised and decarbonised energy future.

The current challenge for policy makers, industry and stakeholders is to manage this transition to ensure that it is just, affordable and safe. In the Australian context, special attention needs to be paid to households and small rooftop PV systems. Cumulatively, the 3 million+ small-scale rooftop PV systems are already Australia's biggest electricity generator, with households across the country installing these systems at a rate of between 250,000 to 300,000 per year.

While these rooftop PV systems are a blessing for households in terms of electricity bill savings, they are a growing problem for distributed network service providers (DNSPs), especially in the low-voltage part of the system.

Increasing clean energy generation from PV is essential to replace Australia's current reliance on fossil fuel generation. At the same time, the decentralised nature of rooftop PV often strains the existing network infrastructure (which is designed for one way flow of electricity) to its limits.

In order to avoid conflicts between the service and safety goals of the DNSPs and consumers' desire to install solar PV, new ideas and solutions are required.

In this context, Project SHIELD aims to use state-of-the-art technology (which has been developed in Australia) to enable DNSPs to analyse how much PV can be connected to a specific feeder in the network without having to invest in infrastructure updates. The project team uses the energy data platform solution designed by Luceo Energy, in conjunction with the state estimation algorithm developed by GridQube, to calculate the specific PV-hosting capacity for a number of selected feeders.

While simple calculation of such limits is reasonably well understood, Project SHIELD has initiated a reality check as part of the project. Calculation of PV-hosting capacity for a specific network feeder requires data from this feeder.

Many in the industry look to smart meter data to fill the data gap; however, with the exception of Victoria, smart meter penetration is low. Project SHIELD adopted a novel, open and inclusive approach to data in order to find answers to the following questions:

- How many smart meters are installed at the project feeders?
- Can other data sources be utilised to compensate for a lack of smart meters in many locations?
- What kind of other data sources are installed in these feeders?
- Can we get access to such data sources?
- What does it cost and how is privacy protected?
- Do these other data sources provide data that is useful?

- Is all the data available at feeder level adequate in terms of quality and quantity to calculate limits?
- Do we believe that the results of the calculation are sufficiently representative and reliable to permit network operators to adjust the feeder PV-hosting capacity from static to calculated?

The Project SHIELD team worked very collaboratively and constructively to find answers to these questions and overcame difficulties posed by working under COVID-19 restrictions. At this stage, the project's findings confirm the widely held assumption that data and modern analytics are the solution to transitioning to a dynamic network. However, in practice, access to, cost of and availability of data is still a huge challenge.

Some of the current key learnings and outcomes of the project are:

1. It is possible to use data from diverse data sources and of differing quality to run network-focused analytics.
2. Such data and analytics can deliver a reasonably good result to PV-hosting capacity at feeder level.
3. DNSP operators will need to test these results and determine if they are adequate to be adopted as general best practice.

These results are encouraging and support the concept of a dynamic and digital grid. Nonetheless, it is worth taking note of the related learnings and observations listed below to obtain an in-depth understanding of the current landscape.

1. Overall number of sensors connected to a communication device in network feeders are relatively low.
2. No currently installed sensors and solutions (except for sensors on network-owned assets) are designed to deliver data for network purposes.
3. Sensors that are installed behind customers' meters are designed to inform the users and, in some cases, the hardware OEM.
4. OEM and service-provider data platforms are capable of (but not necessarily equipped to) sharing data or parts of the data with external third parties via API.
5. Data quality and availability differs greatly between the various solutions that were proposed.
6. Some OEMs or service providers have no interest in sharing data with third parties. IP, security, and potentially conflicting business models were cited as reasons.
7. Consumers are generally sceptical to share their data with networks. Close engagement and incentivisation could alter this position on a case-by-case basis.
8. Cost of accessing data varies according to the supplier and their underlying business model.

So far, Project SHIELD has demonstrated that a data and analytics-based approach is plausible to align consumers' and networks' interests with regards to rooftop PV and safe and reliable network operations without major capital expenditure.

The project also highlights the fact that while access to various data sources and data useful for analytics is technically and legally possible, it is costly and delivers unreliable quality.

Smart meter penetration is too low, the data is not optimised for network purposes and access is too costly for DNSPs to solely rely on current smart meter infrastructure.

Given consumer scepticism about data sharing, networks need direct, low cost, high-fidelity data to permit dynamic, robust and reliable analytics and decision-making that will result in improved customer service and a safe network with high degree of rooftop PV penetration. Such digital data infrastructure will deliver benefits to networks and consumers beyond what has been analysed in this project.

Project Overview and Objectives

During the past five years, high and steady growth in the installation of household DER has seen the total numbers of systems almost double. At the same time, technological advances have meant that the capacity of the fleet has increased by 170% within this timeframe. The emerging disruption to DNSPs is evident.

Much work is underway to understand, manage and accommodate the growing DER (Distributed Energy Resource) fleet. Within the DER space, there are currently 16 ongoing ARENA projects, of which 8 are associated with managing or increasing hosting capacity of the network, as well as its stability. This demonstrates the collaboration between government and industry to support the renewable-energy transition in Australia. Project SHIELD is similar to other projects in this space; however, its aims are unique in terms of assessing how to achieve the best visibility (at the lowest cost) of the low-voltage network from multiple non-traditional power data sources in order to effectively estimate the network's available hosting capacity. It is hoped that this study will supplement current research work and provide networks with an alternate pathway to gain network DER visibility and estimate feeder capacity until a better alternative (such as ubiquitous AMI) becomes available.

Our activity seeks to achieve the following key objectives.

- To demonstrate a proof of concept for aggregating different types and sources of data to estimate DER-hosting capacity.
- To develop an initial business case for network stakeholders seeking to establish minimum level data requirements for assessing hosting capacity at a reasonable level of accuracy. This process will help DNSPs to avoid unnecessary or cost inefficient investment in network-data sourcing and monitoring.
- To develop new technologies which will allow networks to effectively use existing and new data sources to estimate the operating state of the LV feeder
- To increase the value of renewable deliverables by providing an estimation of hosting capacity that can be used to prevent premature curtailment of renewable-energy exports on partner networks.
- To share our findings with industry stakeholders and the wider research community.

Project SHIELD will be implemented in a 2-phased delivery approach [see Figure 1 below]. At the end of Stage 1, the project team will review the results of the project to date. They will also consider the benefit of proceeding to the next stage given that it will involve investment in additional data device installation. Finally, the project team will discuss their findings with ARENA and seek endorsement for Phase 2.

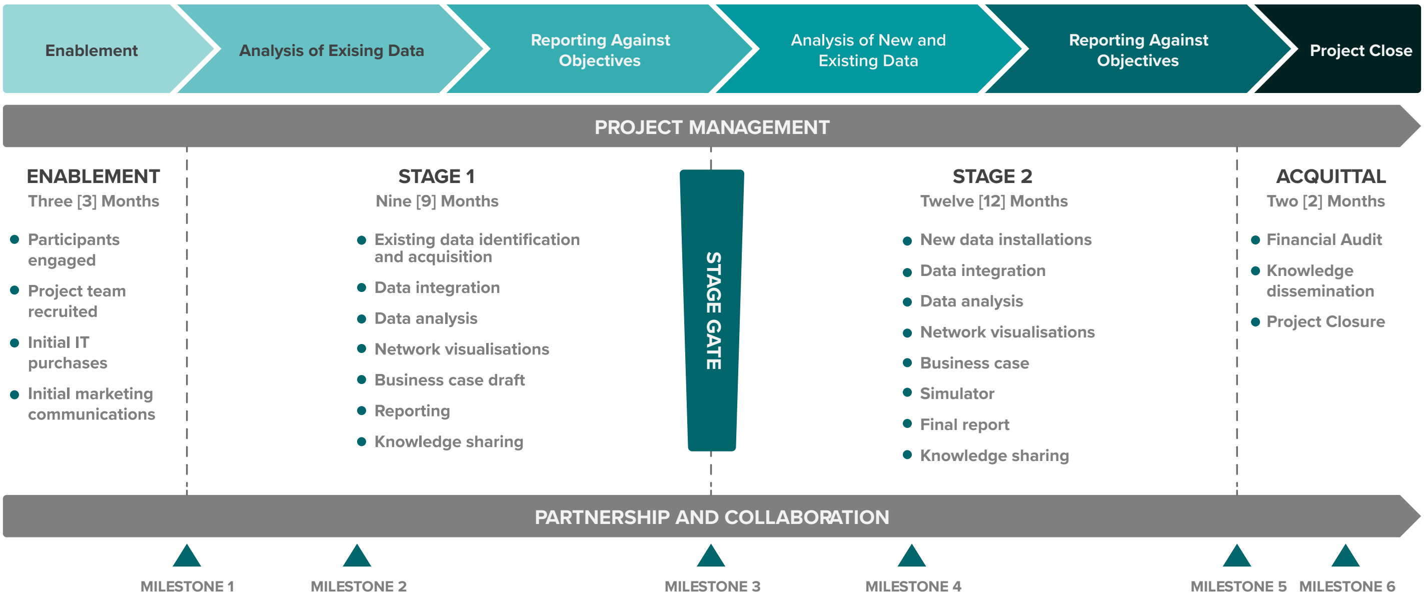


FIGURE 1 – Phased approach for delivery

This Outcomes Report includes the findings and results of the project at the end of Phase 1.

Project Approach

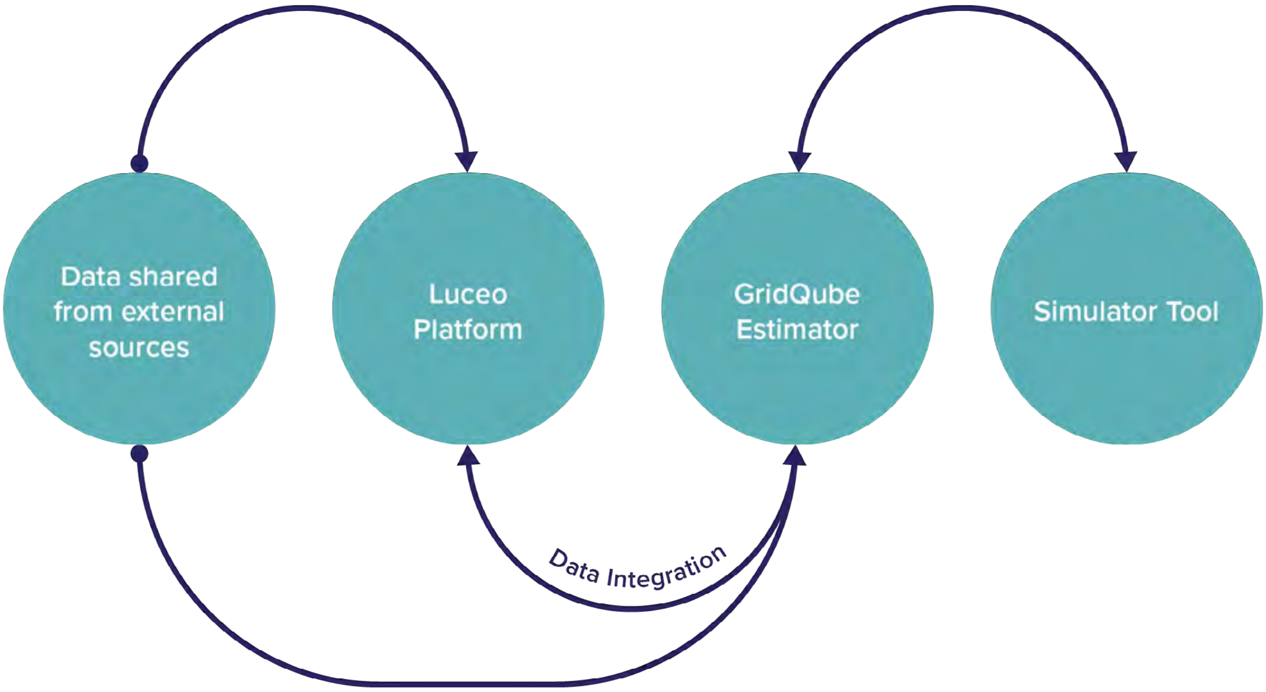


DIAGRAM 1 – Project SHIELD approach

Below is a high-level summary of our approach to Phase 1 of the project. Each area will be described in greater detail in the following sections of this report.

Data Acquisition

Project SHIELD aims to provide visibility into the low-voltage network by utilising the existing data sources that are currently available. These include non-traditional sources of power data that are typically not utilised by networks.

To achieve this objective, our strategy was to use data sources that are owned or controlled by the project participants and their partners, as well as to acquire data from third-party external vendors.

Data Storage (Luceo Platform)

This data will be received and stored in a newly created cloud-based data platform that has been developed and provided by Luceo Energy. In addition to using the platform as a data repository, we will be able to visualise and display the project's data, operating parameters and feeder architecture of our trial areas.

Data Analysis (GridQube State Estimation)

The combined data from the multiple data sources is connected to the GridQube State Estimation Algorithm. The GridQube software supplements the data provided with additional network-monitoring data. Using this data and an electrical network model, the algorithm calculates the most probable electrical state of the network and, in turn, generates dynamic operating envelopes for both export and import within that portion of the low-voltage network.

Scenario Analysis (GridQube Simulator)

As part of the project, different scenarios will be run to evaluate the impact of changes in the number, type and location of data sources have on the network model, as well as the accuracy of the estimated network state.

Data Acquisition

Based on the detailed information that was required for the project, it was determined that we should initially limit our proof of concept to a feeder from each participating DNSP (i.e. Essential Energy, Energex and Ergon Energy). Each feeder was expected to provide between 900–2000 customers.

A key objective of the project was to better understand the value of different data. Consequently, feeder selection was based on locations where the network partners believed they have access to higher concentrations of data.

Essential Energy determined that the Port Macquarie feeder would be the trial feeder for NSW as it has a high penetration of solar equipment that is known to cause voltage issues. The Port Macquarie–Hastings Council area is located on the mid north coast of New South Wales, about 420 kilometres north of Sydney CBD and 510 kilometres south of Brisbane CBD.

Similarly, it was felt that reviewing one feeder from Energex and one feeder from Ergon would demonstrate a variety of networks. The two feeders selected in Upper Coomera and Bohle Plains were identified as having low-voltage networks with a good penetration of data that Energex and Ergon could both access. Upper Coomera is located on the Gold Coast of Queensland and has a population density of about 1,421 persons per square km. In comparison, the Bohle Plains area near Townsville has a population density of roughly 450 persons per square km but it includes an extremely high penetration of rooftop solar power.

Figures 2, 3, and 4 present typical voltage and power profiles for a single distribution transformer on the Upper Coomera network. However, it should be noted that network performance was not a primary consideration in selecting this feeder.

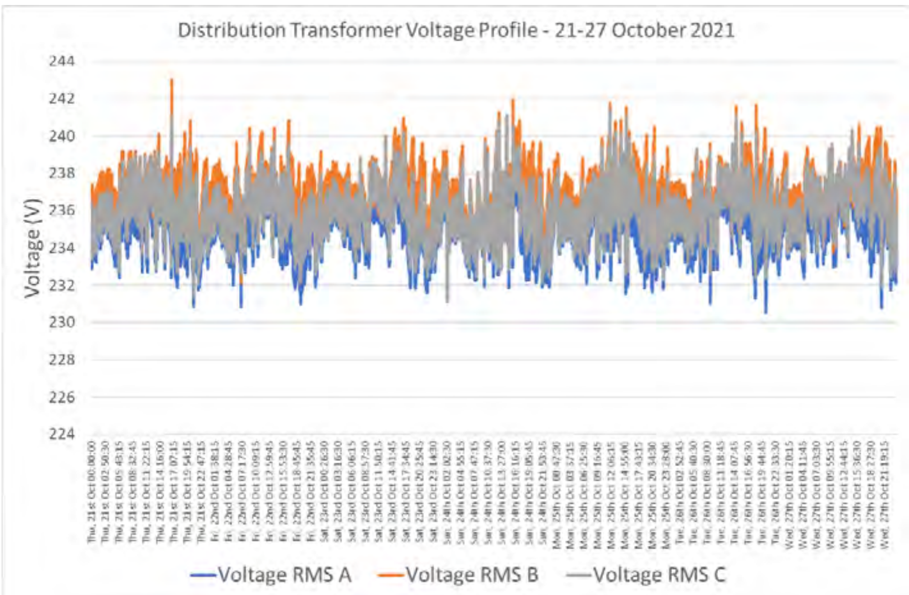


FIGURE 2 – Distribution transformer voltage profile

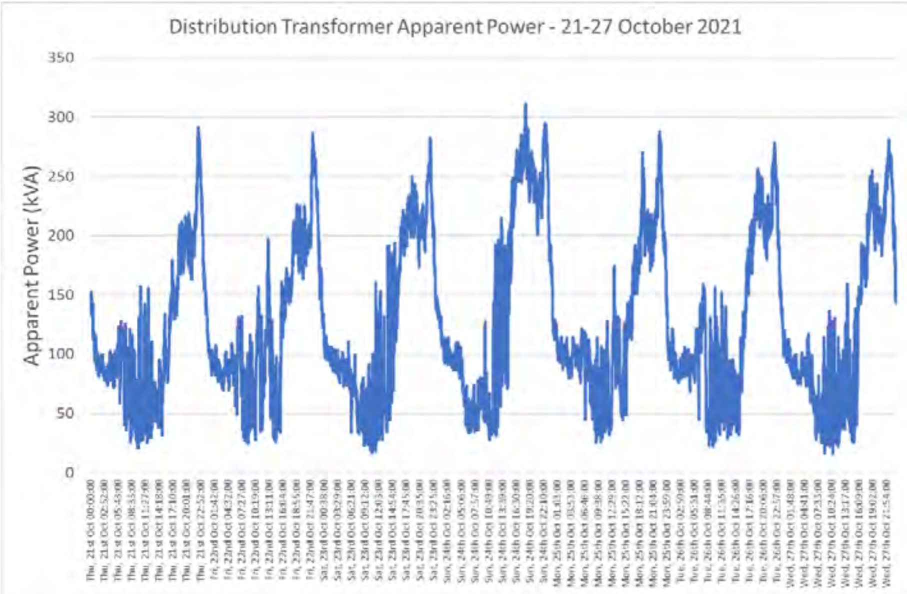


FIGURE 3 – Distribution transformer total apparent power

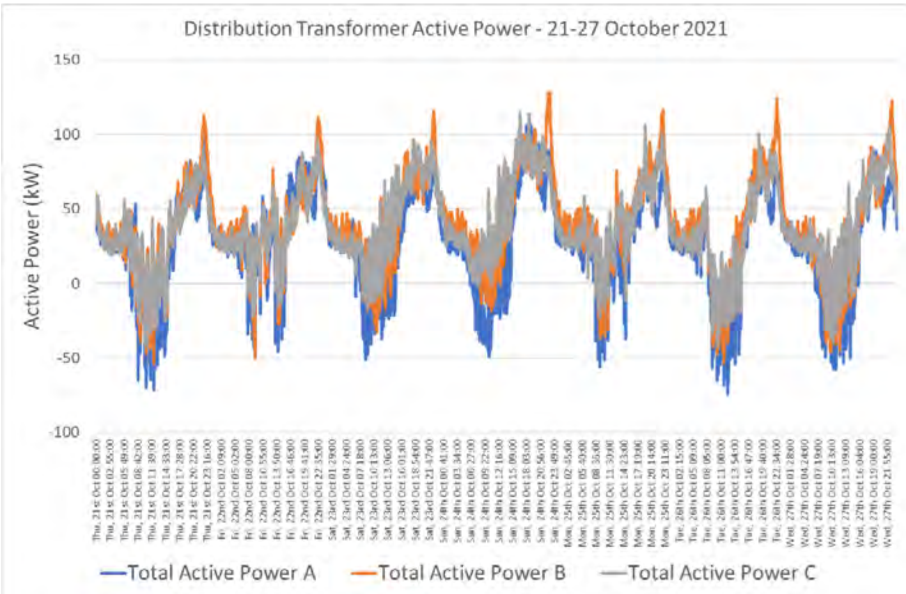


FIGURE 4 – Distribution transformer active power per phase; negative values indicate reverse power flow

Identifying and Approaching External Data Sources

The project team began the process of data acquisition for each of the trial feeders by identifying vendors of each data source (see Appendix 1). Following an initial desktop review, 68 companies were identified as potential data sources, including AMI suppliers, meter data providers, network monitoring equipment providers, network IoT device providers, home energy management system providers, inverter manufacturers, EV-charging station providers, plus other sources such as NBN equipment.

Further analysis identified that 30 of these companies could be removed as they were considered not viable various reasons including:

- they were already involved with the project
- they were subsidiaries of another company already on the list
- the project team members deemed they did not have any devices in our trial areas.

Of the remaining companies, we were not able to identify the appropriate contact person or way of contacting a further 12 of these companies. Where possible, we sent messages via the company's website form.

We emailed letters of invitation (see Appendix 2) to the remaining 26 companies to notify them about Project SHIELD and invite their participation. Where possible, we also utilised industry networks by reaching out to personal contacts of our team members to reach these remaining companies; an approach which proved to be the most successful. In the end, we communicated with a total of 14 companies (i.e. 54% of potential providers identified).

Discussions with the companies are summarised below:

- The majority of companies did not have any devices on our trial feeders. Many vendors expressed interest in the outcome of the project but had little or no data sources, either in the trial areas or, more specifically, on our trial feeders. In fact, there were so few matches that the project decided to expand the trial areas in Queensland to include more LV feeders in order to locate more devices on the network.
- Several of the companies we spoke to expressed concern over privacy and competition. Given that Redback Technologies is a direct competitor of many of these companies, there was some reluctance about sharing data with Luceo as its partner company. Other companies cited concerns about sharing customer-level data.
- Many of the companies do not hold or are willing to share data at the customer level. Due to the unique nature of this project, we needed to receive engineering data from each household (including its location on the electricity network). Many companies do not hold NMIs or it is their policy only to share data in aggregate to government, utilities or regulatory bodies.
- For those companies interested in sharing their data, several were willing to share data at no cost while others were willing to sell their data. However, we found this data was only available at a high fixed cost for a small amount of data. This was predominantly due to the high cost of bespoke development, coding and testing of new queries for the established datasets.
- Customer permissions were required by some companies while less of a concern to other companies.
- While every attempt was made to acquire data from every source by the milestone date, we ended up with only 4 external data sources.

The project is currently in discussions with 1 more source, but the dialogue was initiated too late to meet the deadlines for Stage 1. From the original identified 7 data source categories, we received data from 4 of them, including DNSP (transformer monitors), Consumer (home analytics), Smart Meter and Other. However, not all data sources were available in every trial feeder area. For example, Smart Meter data was available for Bohle Plains and Port Macquarie but not for Upper Coomera, while transformer monitor data was available for Port Macquarie and Upper Coomera but not for Bohle Plains.

Table 1 below shows the complete list of data sources by feeder.

Network Model	Bohle Plains				Coomera 9A				Port MacQuarie			
	Imported	Start Date	End Date	NMIs	Imported	Start Date	End Date	NMIs	Imported	Start Date	End Date	NMIs
	Y	1/10/2021	31/10/2021	831	Y	1/08/2021	1/09/2021	1391	Y	1/10/2021	31/10/2021	716

Telemetry	Bohle Plains				Coomera 9A				Port MacQuarie			
	Imported	Start Date	End Date	NMIs	Imported	Start Date	End Date	NMIs	Imported	Start Date	End Date	NMIs
OL1	Y	1/01/2021	31/12/2021	4	N				N			
OL2					Y	1/08/2021	31/08/2021	49	N			
Transformer					Y	1/08/2021	31/08/2021	0	Y	1/10/2021	31/10/2021	0
Smart Meters	Y	3/01/2022	9/01/2022	304	Y				Y	1/01/2021	12/10/2021	169
NBN	Y	29/10/2021	29/10/2021	0	Y	29/10/2020	29/10/2021	10	Y	29/10/2020	29/10/2021	2
Delta	Y	5/10/2021	31/10/2021	0	Y	1/04/2021	31/05/2021	0	Y	6/10/2021	31/10/2021	0
Solar Analytics	Y	1/01/2021	31/12/2021	2	Y	1/01/2021	31/12/2021	4	Y	1/01/2021	31/12/2020	0
Totals				113				63				171
Density of Sources				13.60%				5%				24%

TABLE 1 - Project SHIELD data sources

Data Collection

Historical vs Real Time

Originally, the objective of the project was to gather real-time data. However, we quickly realised that many of the data suppliers did not collect their data in real time or were willing to provide their real-time data. Additionally, those companies that were willing to provide an API to access data in real time were not able to do this within our project timelines.

Given the DNSPs' cyber security, network security and privacy considerations, it was not feasible (or necessary) to obtain access to real-time data.

The project team agreed that dynamic operating envelopes (DOEs) that were calculated on historical data over a set timeframe would be adequate for our requirements (without the data needing to have been passed from all the various data sources (monitoring devices) in real time). This also meant that we just needed to collect one snapshot of data and would not require additional API data.

Date Range

The project team requested providers to supply full data sets for 2021; however, it quickly became evident that not all providers had access to this information. Some suppliers had collection devices that had been on site for several years and so they were able to provide full datasets for 2021. In contrast, other suppliers had newly installed devices and so they were only able to provide partial amounts of data for this timeframe.

Once we had decided to use historical data for the project, we had to determine the most ideal test windows. This process involved identifying the time periods when the maximum number of devices were collecting data in a trial feeder area. The resulting date ranges are shown in Table 1 – Project SHIELD data sources above (with the following additional notes):

- The Bohle Plains feeder had the optimal date range for October 2021 for all telemetry.
- The Upper Coomera feeder optimal month was August 2021 (albeit for one data source).
- The Port Macquarie feeder had an optimal month of October 2021 with 29th October 2021 yielding telemetry from all data sources.

Frequency of Data

The team found that each device measures the data at different intervals (i.e. 1, 5 and 15 minutes). One benefit of the Luceo Platform was the ability to store the data regardless of the intervals at which it is provided. Therefore, while we had requested 5-minute intervals for the purposes of this project, Luceo accepted and stored data at all intervals.

Data Specifications

Project SHIELD requested voltage, active or reactive power from all vendors but was willing to accept any combination of these. In some cases, we received only voltage measurements while, in other cases, we received a majority or all the requested engineering data as shown in Table 2 – Project SHIELD telemetry sources. In the case of the Luceo Platform, the time intervals were accepted as is with no data modifications undertaken to convert the data back into a standard interval value.

Summary of Data Sources

The project received data from Evergen, Delta Electronics, NBN and Solar Analytics. Other data sources were provided by internal partnerships. At the outset of the project, Energex and Ergon did not have any formal arrangements in place for acquiring customer smart meter data. Formalising the agreements and data transfer mechanisms to acquire this data for the purposes of this project took time but have now been established.

Table 2 – Project SHIELD telemetry sources below provides a summary of the data received from

each data source, including engineering data collected and time intervals. Any gaps in the data were managed within the overall data analysis work without any additional modifications. Both instantaneous and average measurements were used.

Telemetry Data Source	Description	Data	Interval
OL1	Engineering data from Luceo OuijaLite 1.0 devices (Point of Connection)	Voltage Current Power factor	1 min Instantaneous
OL2	Engineering data from Luceo OuijaLite 2.0 devices (Point of Connection)	Voltage Current Power factor	1 min Instantaneous
Transformer	Engineering data from transformer monitor (Transformer)	Voltage Current Angle ImportVrah ImportWh ExportVrah ExportWh	10 min Average (30-min for energy data)
Smart Meters EQL	Customer smart meter data (Point of Connection) of EQL	Voltage Current Angle	10 min Average
Smart Meters Essential Energy	Customer smart meter data (Point of Connection) of Essential Energy	Voltage Current Angle	10 min Instantaneous
NBN	Engineering from NBN nodes (Point of Connection)	Voltage	45 min Instantaneous
Delta	Engineering from Delta inverters (Point of Connection)	Voltage Active power	15 min Instantaneous
Solar Analytics	Engineering from Solar Analytics devices (Point of Connection)	Voltage – Minimum Voltage – Maximum	5 min Min/Max

TABLE 2 – Project SHIELD telemetry sources

Data Quality

We found that the quality of the data received varied widely. In some cases, NMIs were missing and latitude/longitude values provided were not accurate enough for our purposes. Without the NMI data, it was not possible to determine exactly how the connection point joined the network. In the case of some vendors, the project team found their location was determined at time of commissioning via GPS signal from the phone used to set up the device which, in turn, explained why the device seemed to be in the middle of a road. Transferring data over from customer smart meter companies to the DNSPs' systems also required significant work and interpretation skills.

Data Definitions

Project SHIELD found that there was often some confusion regarding the characteristics of the data the project required. We needed to be very clear about the difference between historical data vs real-time data, as well as the frequency of data. In some cases, we received data for devices but were unsure whether it was aggregated data or device-level data. There were also some terms (such as 'energy data') which appeared to mean a variety of things to different vendors. Consequently, in this situation, the term 'engineering data' was used instead.

NO COMMON TERMINOLOGY USED ACROSS THE INDUSTRY MEANT MULTIPLE CONVERSATIONS NEEDED TO BE HELD TO CLARIFY REQUIREMENTS.

Data Security and Privacy Laws

The uncertainty and ongoing changes in relation to customer data privacy caused many vendors to feel uncomfortable about sharing data and this continues to be an ongoing area of discussion. The project decided to seek external legal advice about this issue. We were advised that there was some moderate risk associated with the usage of NMIs and customer data. However, it was recognised that the data is "not reasonably accessible to the public and will be protected and used within very strict bounds (no individual premise data will be identified publicly)".

PERCEPTIONS AROUND DATA PRIVACY WERE AN INHIBITING FACTOR AFFECTING MANY VENDORS. THIS WAS DISCUSSED WITH THE ESB AND IS BEING ADDRESSED AS PART OF THEIR DATA STRATEGY INITIATIVE.

Recommendations were provided to ensure that any identifiable personal information, such as NMIs and address information, was removed from any report or screenshot, etc. Any risks were identified and controls were put in place.

Unfortunately, this assurance was still insufficient to convince some companies to share their data.

Access and Sharing

Data from External Vendors

To address the issues that had been identified, data was received from external vendors, mostly by email or else uploaded securely to SharePoint. During the project, a secured SharePoint site was established as central store for all telemetry. Participants were provided with exclusive links to access and copy their files. After this site had been established, email was actively discouraged – primarily for security reasons of security and because it is not fit for purpose as a central storage repository.

Data between Partners

Data was shared between project team members via email and then moved to a Microsoft Teams folder. All project team members were given Guest access to a Microsoft Teams project folder that was hosted by Redback Technologies.

Data between GridQube and UQ was shared via RDM – the University of Queensland's Research Data Manager Cloud system. Each participant was able to access the project subfolders (both for uploading and downloading) which were password protected with expiration dates. Passwords were sent out via a different channel to the folder link.

Challenges and Findings

Some challenges the project faced were the amount of time required and the difficulty associated with actually acquiring the data. On average, it took about 3 months to progress from our first introductions to detailed discussions about data-sharing agreements and determining whether the data would be suitable for the project to use.

At times, it was difficult to determine whether a device would be included on our trial feeders without giving and receiving confidential customer information. For instance, in early discussions, vendors often could not share customer addresses as it was considered private information, so we were provided street names and postcodes instead. However, as discussions later progressed and formal arrangements were put in place to share data, we sometimes found that just because a device was on the same street, it did not necessarily mean it was connected to the trial feeder.

As expected, many vendors advised that they were not permitted to share their customer's data without a signed data agreement. Where payment to vendors for extracting datasets was required further complications arose since the project team did not want to enter into an agreement unless we could verify that the data would be relevant and this could not be determined until the agreement was signed and the data provided (and the financial commitment made). For example, in the case of one vendor, it was initially identified that they had a number of devices in the postcode provided. However, when narrowing down to street names, only 10% of the devices were identified to be on the right streets and, of those, fewer than 50% turned out to be on our trial feeders.

Further complexity arose from the amount of time it took to progress through the negotiation phase for collecting data. Each conversation raised issues that had to be explored offline with management or legal resources and this added more time. While vendors agreed with the objectives of Project SHIELD, they clearly prioritised their commercial activities over participation in the project which led to delays in responding to our phone calls and emails (often weeks or months overdue). In one case, a company received our initial communication in September 2021, but it took them until December 2021 to commence meaningful dialogue with us. Their data is currently being investigated and will be integrated into Stage 2 but, unfortunately, it was unable to be included in Stage 1 of the project due to timeline restrictions.

Another challenge we faced was the accuracy and detail of the network model for the low-voltage trial feeders. The state estimation algorithms required an understanding of the electrical connectivity layout of the trial feeders (e.g. What is connected? What are the operating parameters of items connected? What is the order/context of connection?). The project team quickly discovered that some of the required network data was incomplete and was restricted to associations between customers and the distribution transformer they were connected to. Historically, capturing and maintaining this level of detail for low-voltage networks has- not been prioritised as it is not a critical aspect of the distribution-network operation. In the case of one of our feeders, the decision was therefore made that it would provide a good test of quality of data sourced from low-voltage devices in medium-voltage state estimation.

Data Storage and Integration

Project SHIELD Data Architecture

As shown below in Figure 5 – Project SHIELD data architecture, network and telemetry data was received from external data sources and uploaded to the Luceo Platform in order to be available for use by GridQube via API and file extracts.

The design of the Luceo data model was based on a graph model consisting of nodes and edge tables which enables flexible network-centric storage of network and telemetry data.

GridQube processes the data and delivers the state estimation calculations and operating envelope information to Luceo through APIs.

DATA IS STORED DIFFERENTLY BY EVERY COMPANY AND IT WAS DIFFICULT TO ACQUIRE DATA FROM EXTERNAL SOURCES IN LARGE QUANTITIES FOR THE TRIAL FEEDERS THAT WAS APPROPRIATE FOR STATE ESTIMATION PURPOSES.

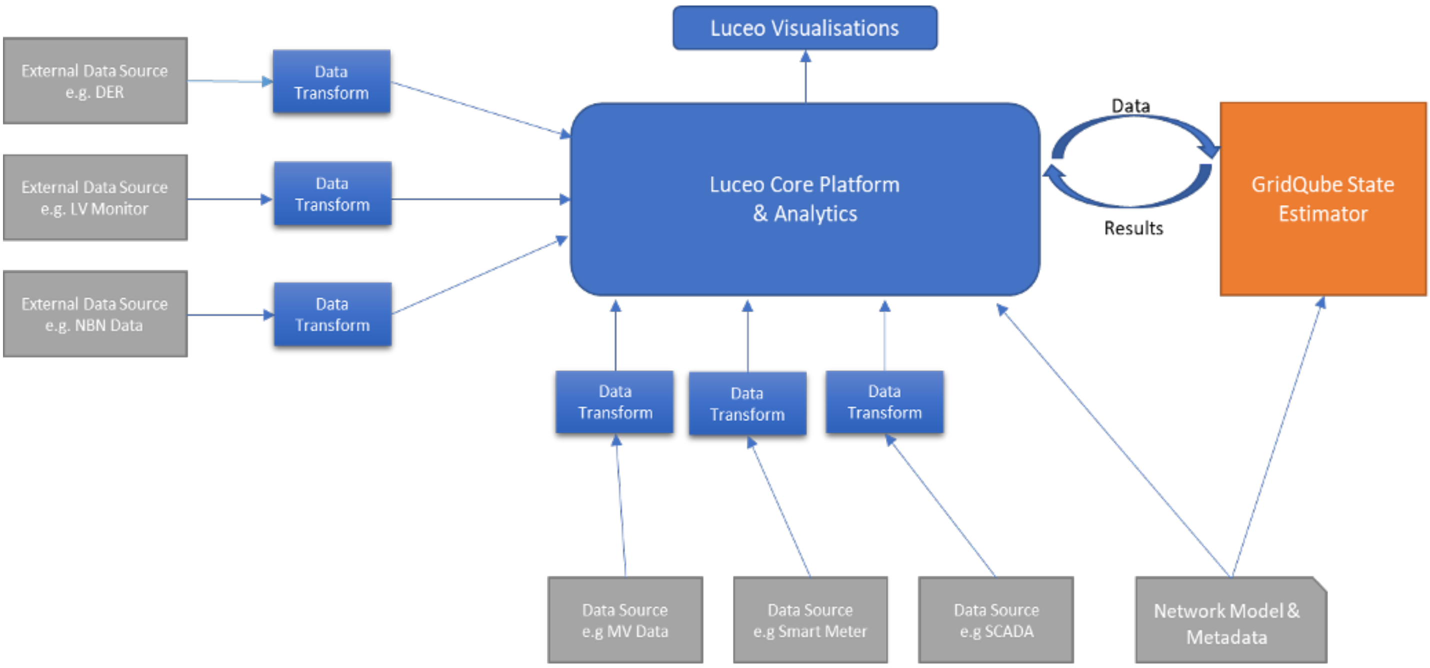


FIGURE 5 - Project SHIELD data architecture

Description of the Luceo Platform

Luceo's data platform was available for the project's activities. The platform is a cloud-based big-data platform that has been customised for electricity network operators' use. It is currently operational across a number of electricity networks, with over 20,000 devices providing 1-minute intervals data in real time from the low-voltage network. Prior to commencing the project, this platform was capable of accepting a range of data formats.

Luceo's unique suite of solutions (including an effective, real-time IoT platform and visualisation tools) have enabled users to manage their existing and emerging challenges. Distribution-network operators using the Luceo Platform obtained real-time insights and the data necessary to manage modern decentralised energy systems. The platform permits users to view critical network parameters from every feeder in the network on one dashboard.

THE LUCEO PLATFORM WAS MODIFIED TO STORE NETWORK AND TELEMETRY DATA OF HETEROGENOUS DEVICES

Modifications

As part of Project SHIELD, Luceo created a new network-centric data model to support an electricity network model. This graph-based model enabled a flexible and dynamic system of nodes and devices, including the ability to connect multiple devices to each node while also maintaining separate data for each device's respective measuring points and telemetry data.

The portal was adjusted to reflect the network-centric view to enable drilling down from transformers to devices. Specific information relating to a node type was available in common portal widgets. Device types were established as 'first-class citizens' (i.e. given secure credentials for authentication and security) to display the data for searching purposes. New APIs supported data integration to facilitate remote call integrations and ETL components for uploading network and telemetry data of different device types.

Description of the State Estimation Solution

- GridQube's technology is a fully automated network analysis and management system. The core technology stack offered by GridQube consists of the following 3 interrelated functional components.
- **GridQube DSSE:** A Distribution System State Estimation engine, purpose-built for use in distribution systems, including LV networks down to the individual customer connection point. The DSSE engine is applicable on any voltage level of a multi-phased AC-power network and captures all electrical and operational imbalances.
- **GridQube CA:** The Constraint Analysis engine is a general-purpose engine that relates controllable parameters such as single-phase and 3-phase active power that is

connected to a modelled and monitored network with a flexible list of applicable technical and operational limits throughout.

- **GridQube CCO:** The Capacity Constrained Optimisation engine is the decision-support/decision-making engine. Most decisions are influenced by technical and operational limits and dependent on answers to the following questions: “What is the maximum that can be done?” or “What is the minimum that needs to be done?”

The system accepts as any of the following sources of input and calculates dynamic operating envelopes for a range of different circumstances and allocation strategies as an output:

- metering data
- SCADA data
- GIS data / network models / electrical connectivity, etc
- load / generation forecasts
- weather data (including potentially pre-processed data-load/generation projections)

Importing Data into the Luceo Platform

When importing data into the Luceo Platform the following assumptions were made.

- Telemetry data from vendors would have a NMI reference as a first preference; alternatively, a Latitude Longitude would be utilised.
- Data provided by DNSPs would be similar and existing data imports could be used.
- Given the flexibility of the Luceo Platform, we could import data in any format with any amount of detail.

All vendors were told that in the interests of efficiency, it was acceptable for them to provide data in any format and schema. As a result, the data we received arrived by different methods and in varying formats. When importing the data into the Luceo Platform, we found that:

1. In most cases, telemetry was high-quality data provided at the required resolution and included a high percentage of time intervals data provided for a particular day.
2. Assuming there was a commonality of schemas between feeders and telemetry was optimistic rather than realistic which resulted in more effort required to design and complete data loaders. The project team accepted that an industry schema was not mandated – nor could have been enforced.
3. Upper Coomera and Bohle Plains network models had significantly different schemas as they are located in separate distribution networks with different and independent systems that contain the required data.
4. Data formats provided were not always in the expected CSV format and required custom import development. The project accepted that it was not possible to mandate the final

file format.

5. For some data sources, location data was not sufficiently accurate to pinpoint the location of a device in the network model which resulted in lower number of devices discovered and lower device density.

Data files were analysed to determine the type of schema and if NMIs were available for mapping to the network. Where there was insufficient NMI information, a latitude and longitude approach was implemented. This often involved reverting back to the original provider of the data source.

In the case of the Port Macquarie feeder, a concept of Premises ID was used for mapping the telemetry data. This telemetry data approach worked well as it was able to be translated through other means if required, such as Cabinet IDs. Any available engineering data was considered valuable and so we were able to upload all telemetry data. Time intervals presented no significant issues as they were consistent.

Data Validation

The data quality was assessed using basic thresholding, sanity checking and outlier detection. For example, it was determined that there should not be any solar output or export at night in any location. Any residential house data with solar data showing output only at night should be regarded as an error; perhaps due to the devices to record solar and house output having been swapped around. Solar and household power data should have basic thresholds. For example, in terms of instantaneous power, residential solar is rarely more than 10 kW single-phase and residential consumption is rarely more than 15 kW. Outlier detection is another way to check for spurious output rather than using a rules-based approach. This is accomplished by clustering energy time-series or examining energy models for each house.

For data sources where energy values could be derived on an hourly basis (for solar or house consumption), we developed a model to assess the relationship of the observed data with observed weather (spatial) and temporal data. The weather data was obtained on an hourly basis from ECMWF (European Centre for Medium-range Weather Forecasts). A house which exhibits no correlation with either weather or temporal data in terms of either solar or household power generation is considered suspicious. Over a particular timeframe, we would expect the residents to have a consistent circadian rhythm and for this to be detectable. The absence of any pattern raises questions about the data quality.

Other data on each feeder was tested by manually examining the correlation matrix for outliers. To date, we have not observed any significant problems with this data.

THERE WAS NO STANDARD DATA FORMAT CONSISTENT ACROSS ALL COMPANIES, INCLUDING DNSPS.

Passing Data to GridQube

Creating the Interface

The integration of data from the Luceo Platform to GridQube was via an API. This API operated by GridQube submitting a request using NMIs, start and end times and frequency of interval. The request is submitted to the Luceo environment, and a request token is generated. When data is available, it is downloaded, and that file is then accessed by GridQube. Next, the data is transformed and then uploaded into GridQube's systems. The data operating envelope for that time range is calculated and the solution is sent back and stored within the Luceo system.

An issue that became apparent was the fact that since data was flowing from Luceo to Grid- Qube, GridQube could not actually specify NMIs to request the data from Luceo as they didn't know what they were. We also found that the API contract did not provide filtering by the Feeder ID as this was not part of original design; however, it was subsequently found to inhibit a fully automated interface as it required a data extraction of NMIs to be generated and sent to GridQube. Although this manual process only took several minutes for the first two feeders, it still required knowledge about which feeder the data belonged to. In the case of the last feeder, the ability to request a feeder was added to the API which enabled this process to be fully automated.

Network Model Data

The network model is the essential data that was required for this project to be conducted successfully. Network model data for each feeder was obtained from the associated DNSP partners of this project. Redback Technologies received the connectivity models for the feeders while GridQube received the detailed electrical models for all feeders.

Analysing Data for State Estimation

Data Analysis

Once the measurement data had been passed to GridQube from the Luceo Platform, it was translated from Luceo's format into GridQube's format and imported into the GridQube system.

As shown in the diagram below, Project SHIELD data inputs into GridQube's State Estimator were categorised into two types: (i) transformer measurements and (ii) NMI measurements. The transformer measurements were provided directly from the DNSPs while our NMI measurements were provided by the Luceo Platform via API. GridQube then combined the measurements from both sources to prepare a single measurement data file that would be processed by the estimator.

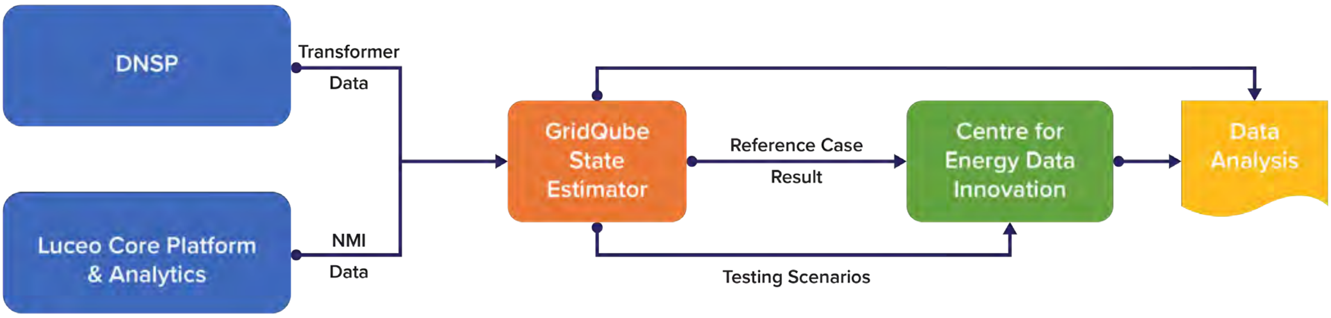
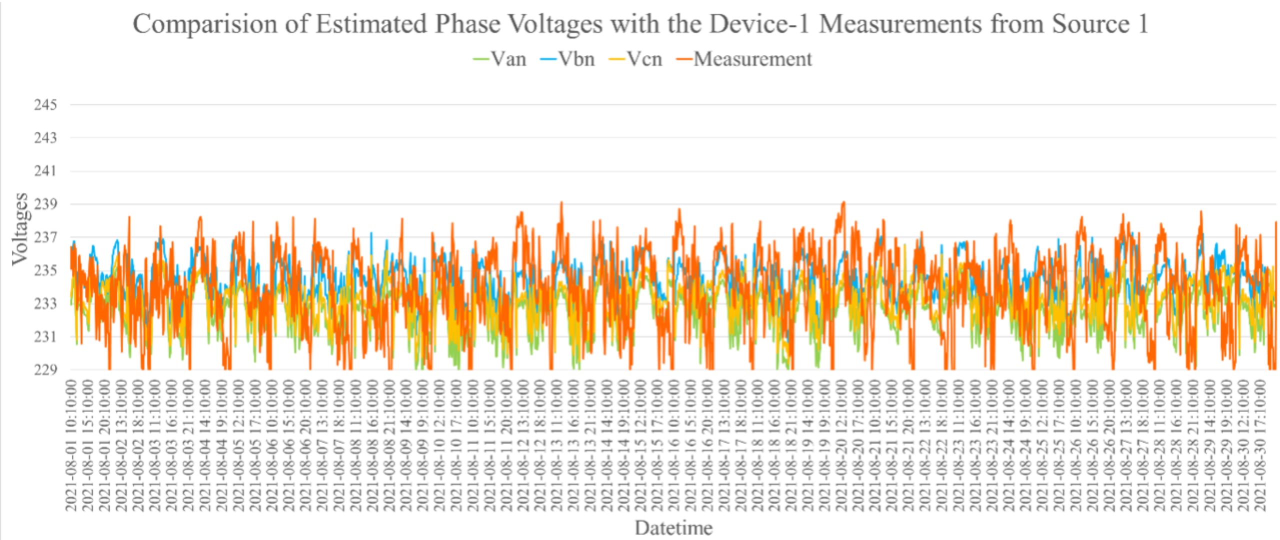
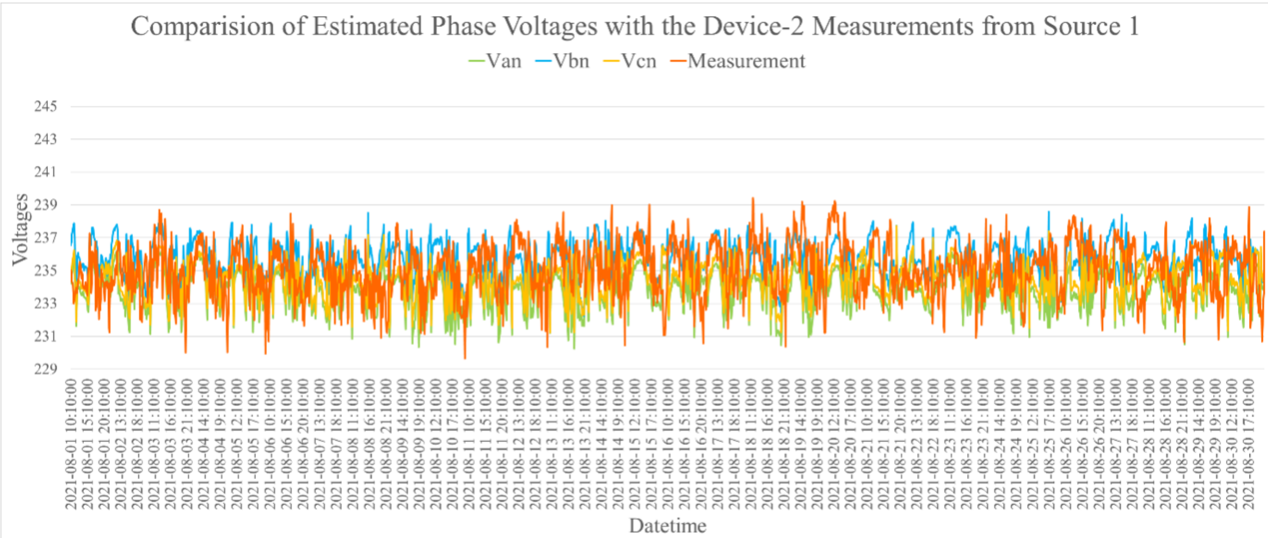


DIAGRAM 2

GridQube analysed the quality of the NMI measurements data by estimating voltages of different sources using the transformer measurements and the network models received from the DNSPs. The observed values of device measurements were then compared with the estimated values. Comparison of different sources of NMI measurements allowed us to determine whether these measurements were accurate.

For example, Figure 6 compares observed voltage values against the estimated values for two NMIs. In the figures, Van, Vbn, and Vcn are the estimated phase voltages obtained from GridQube's estimator while the measurement values are the device-observed voltages obtained from Source 1 and received from the Luceo Platform. In Figure 7, the device-recorded voltages are identical to the GridQube's estimated voltages. Since the estimated voltages are the same as device measurements, the devices' accuracy is considered to be acceptable.

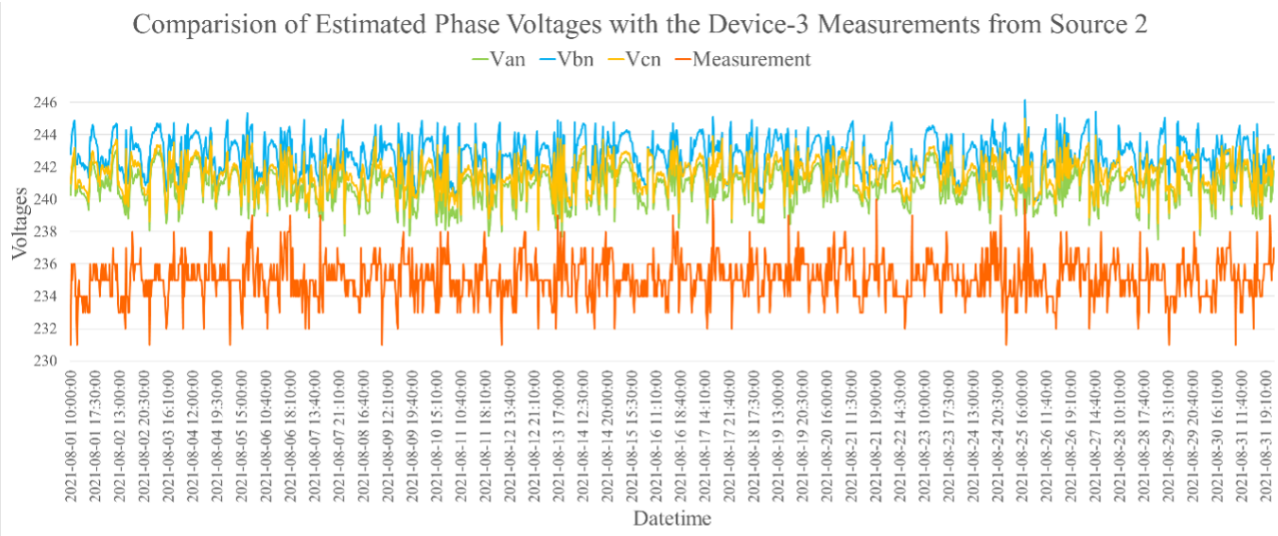


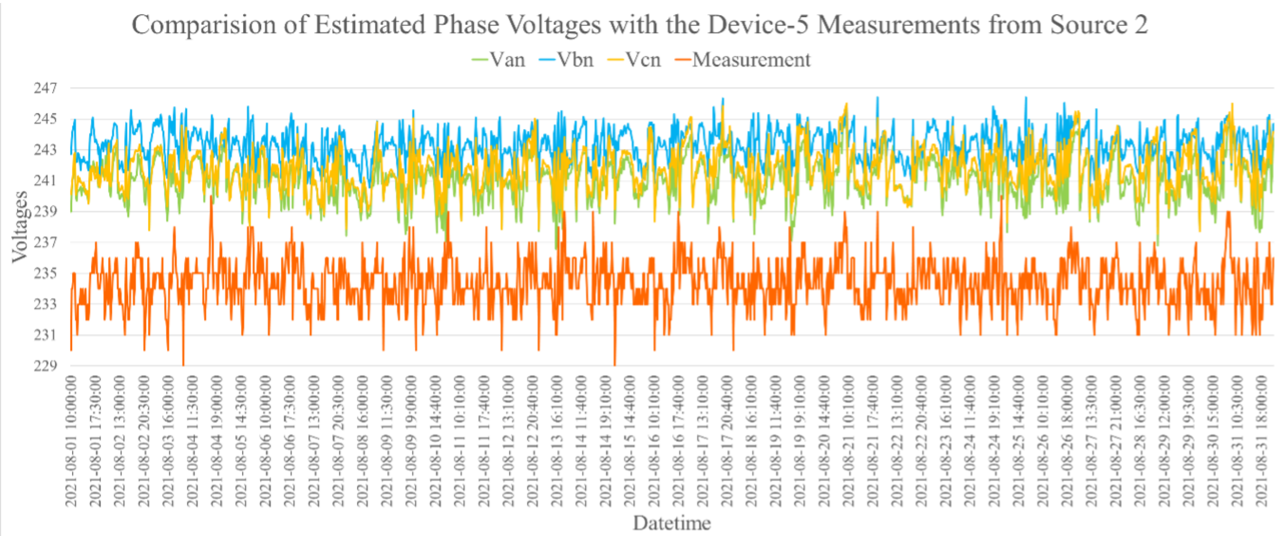


FIGURES 6 - 7 - Comparison of estimated phase voltages with NMI device measurements from Source 1 in Coomera Feeder

Investigations conducted on NMI measurements from Source 2 resulted in different conclusions. Figures 8–9 shown comparisons of some voltage measurements received from Source 2 with estimated voltages from GridQube. From the figures, it is obvious that the reported device measurements from Source 2 are far below the estimated voltages. The offset of voltage measurements was notably high for the majority of Source 2 devices received for the Coomera feeder. Under-measurement of the voltages can significantly upset the estimator, resulting in poor precision output. Therefore, we concluded that measurements obtained from Source 2 were not suitable for the state estimation application. It should be noted that the analysis results were based on a small number of Source 2 device measurements and so were not a generalised opinion in relation to Source 2 device measurements. Definitive conclusions would require further investigation which, in turn, involves more time.

THE DATA ACQUIRED FOR THE PROJECT PROVIDED VARYING LEVELS OF VALUE FOR STATE ESTIMATION PURPOSES.





FIGURES 8 - 9 - Comparison of estimated phase voltages with NMI device measurements from Source 2 on Coomera feeder

Additionally, GridQube's analysis detected an issue with a transformer monitor on one of the trial feeders. Further analysis by the DNSP identified that the transformer monitor had been incorrectly configured. As a result of resources being deployed for flood recovery, this problem could not be fixed in time for our project purposes. However, the project team decided to continue analysing data from the other transformer monitors which could lead to some valuable findings with regards to accuracy of the state estimation providing calculations using limited data. This scenario will be particularly relevant once the third transformer has been reconfigured and its data added to the blended data file in Phase 2. At this stage, we can evaluate the impact of the additional data.

Having been assimilated into the GridQube system, further analysis of data associated with the phase identification of NMIs and transformers was undertaken. This included comparing the frequency of data collected by devices to the transformer data frequency received from the DNSP, as well as identifying proper tap settings for the transformers. These two key analyses were performed using GridQube's advanced algorithms to ensure the estimator produced high-quality output based on the given input. In some cases, data was found to be unsuitable and so was discarded from the dataset as an outlier. The resulting file then was sent to the Centre of Energy Data Innovation (CEDI) to create a series of test scenarios.

Assessing the Impact of Data on State Estimation Accuracy

It was important to get a better understanding of how different factors of input dataset affect state estimation accuracy. As part of the investigations, we first focussed on how inaccurate or missing transformers' tap information might influence state estimation accuracy.

The project team chose to use the Upper Coomera feeder from Energex based on the fact that the feeder has different types of customers based on their loading profiles, and the feeder has more transformer measurements than the other trial feeders used in the project.

Scenario analyses are required to be designed for multiple timestamps for solid and valid statistical analysis-based findings. However, the measurement devices of the selected feeder did not have consistent measurement data across consecutive timestamps. Therefore, it was difficult to find a set of measurement devices with data available throughout the timestamps upon which scenarios are created. To overcome this limitation within the given measurement sets, we figured out the timestamps that ensured the maximum number of devices without any missing measurements. This was important to make a fair comparison of the estimation accuracy differences over multiple scenarios.

Estimation Accuracy Assessment for Transformer Tap Adjustment

To investigate how changing the transformer tap settings impacts estimation accuracy, we focused on developing scenarios considering only available transformers' measurements and then obtaining the estimation solutions for the entire network. We then performed a detailed comparison of the accuracy of estimation solutions against real measurements from customers' premises. Again, no other measurements other than transformer measurements were used in this assessment process.

Scenario Creation

For the scenario analysis, one transformer was chosen. In the selected feeder, 52 customer premises' devices were selected and the state estimation calculation for those premises was evaluated against the actual measurements. Forty-eight devices were located downstream of Distribution Transformer 3, while devices 3, 11, 24, and 27 were located beneath other transformers. Table 3 shows the Tap positions of the different service transformers for the various devices. Note that there were no measurements from Transformer 3. All other transformer taps were optimized while Transformer 3 tap settings were changed to create different scenarios. A total of three scenarios were created by adjusting the tap position of the Transformer 3:

- Scenario 1: Transformer 3 set at tap position 3
- Scenario 2: Transformer 3 set at tap position 6
- Scenario 3: Transformer 3 set at tap position 4

Measurement Device	Service Transformer	Tap Position of Service Transformer
Device 3	Other Transformers	Optimised
Device 11		
Device 24		
Device 27		
Other Devices	Transformer 3	Varied

TABLE 3 – Scenarios are created from measurements obtained from the transformers

Results

The assessment process created scenarios by adjusting the tap position of Transformer 3 at three distinct positions and getting the estimation results for the entire network. A total of 999 measurements files were created on different timestamps for each scenario, and then those were fed into the State Estimator. The Estimator then produced the most approximated network states named as estimated solutions. Next, the estimation of nodal voltages was compared with the actual measurements obtained from the measuring devices at each timestamp. The estimated node voltages were then subtracted from the actual measured voltages to obtain the deviation of the estimation precision in Volts. Figures 10 – 12 show the deviation of the estimation accuracy presented in box and whisker plots for Scenarios 1 to 3.

In Figure 10, the variation of estimation accuracy was +4 to +9 volts on average for most devices except Devices 3, 11, 24, and 27. The estimation deviation was relatively high, which indicates that the output was not very accurate showing that, with a selection of tap position 3 for Transformer 3, the network states were underestimated compared to the actual voltage measurements. However, the estimation deviation for Devices 3, 11, 24, and 27, where the associated transformers' tap positions was optimized, was on average 2 volts. Therefore, optimizing other transformers' tap positions significantly reduces the estimation error.

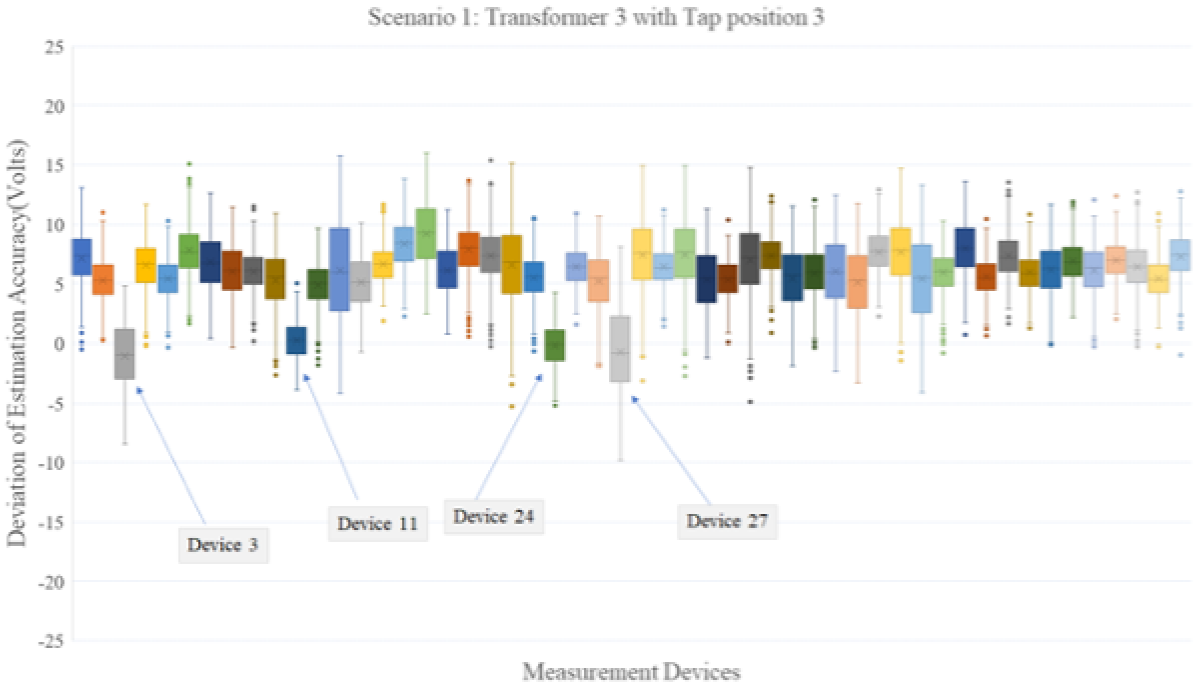


FIGURE 10 – Under-estimation of network voltages due to lower tap position 3 of Transformer 3

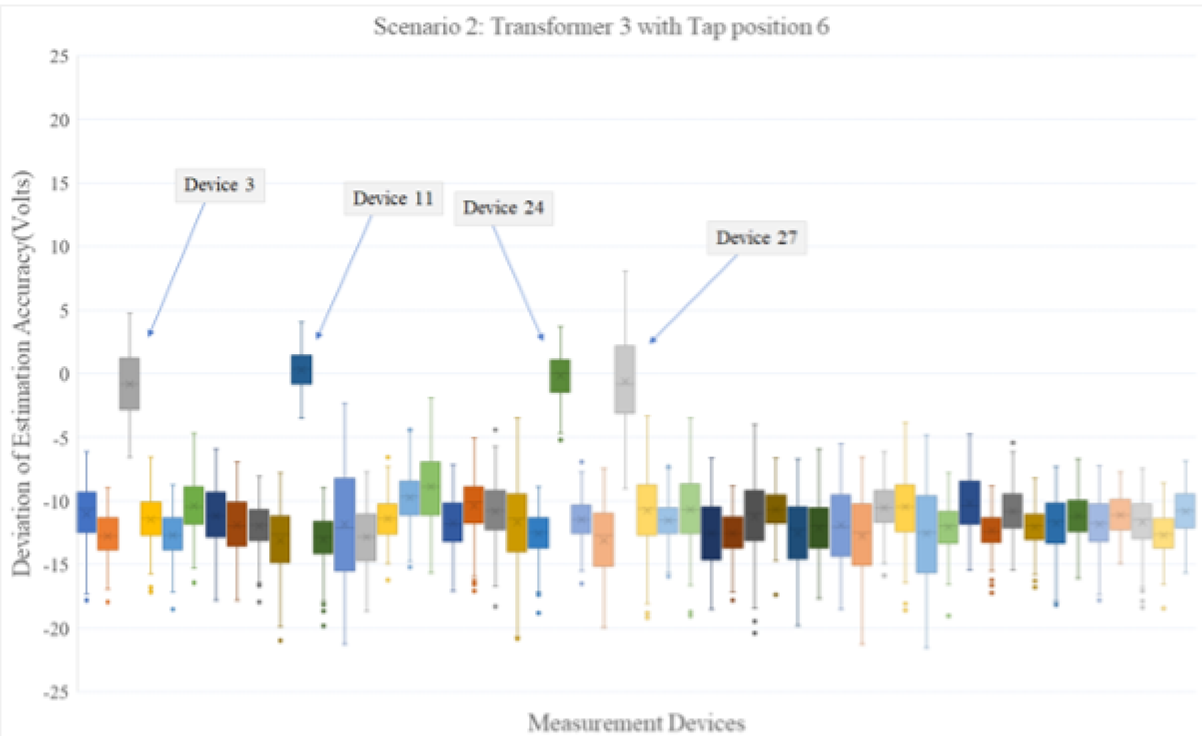


FIGURE 11 – Over-estimation of network voltages due to higher tap position 6 of Transformer 3

In Figure 12, the tap position of Transformer 3 was set at its most likely setting of 4. In this case, the estimation accuracy for all the devices was within +4 volts to -3 volts on average. This indicates a significant improvement in estimation accuracy from the results in Scenarios 1 and 2. The estimation quality for Devices 3, 11, 24, and 27 again did not change.

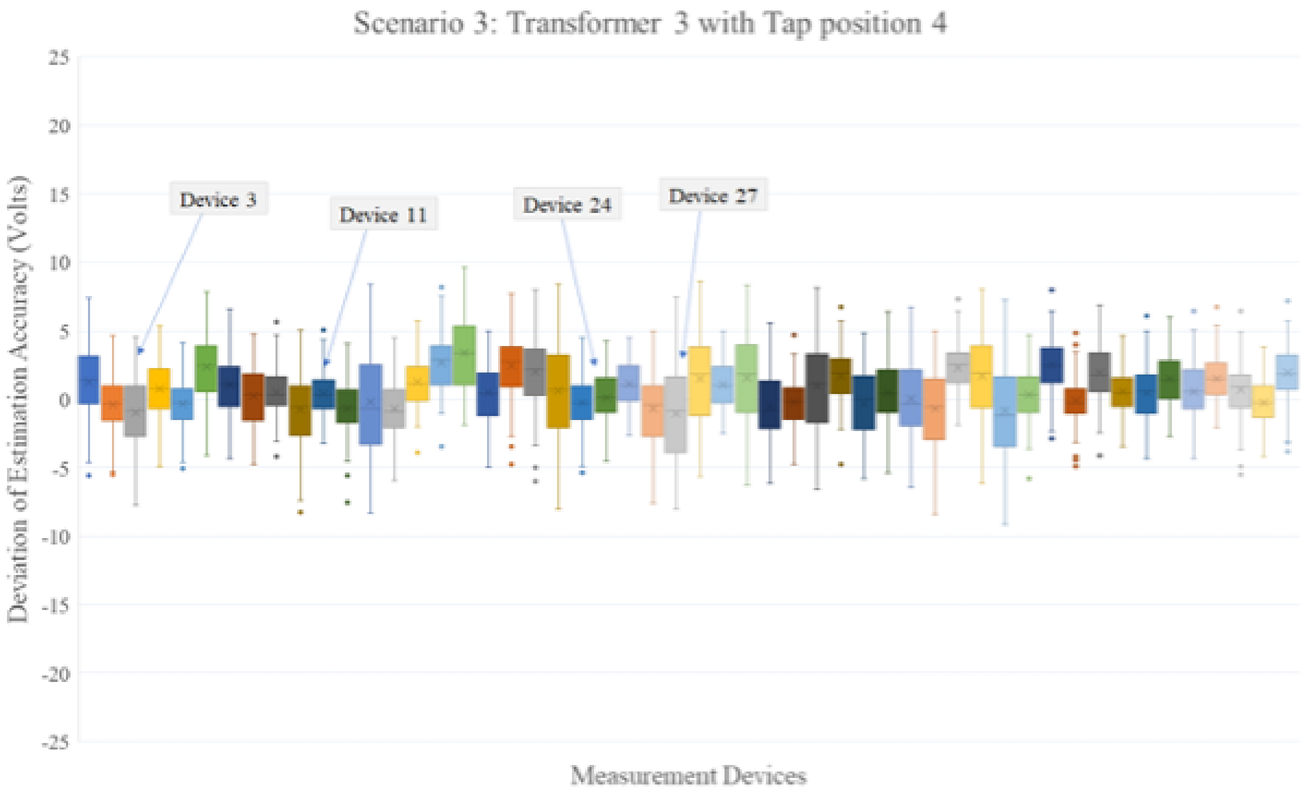


FIGURE 12 – Improving estimated network voltages by optimized tap position of Transformer 3

These results show that a high estimation accuracy of nodal voltages at a customer's premise can be obtained by using only transformers' measurements and adjusting tap position. These analyses from Scenarios 1 to 3 give us a clear indication of how important it is to obtain the transformer tap settings to improve the overall accuracy of the state estimation. If the tap settings are not obtained, it is necessary to find the best possible tap position for each transformer. As demonstrated here, the most probable transformer tap position can be obtained through analysis although it is time-consuming and an LV voltage reference is required to verify the prediction. Therefore, tap settings for each transformer are valuable but they are not absolutely required to improve estimation accuracy. If there is no information given about the transformer's tap position along with the measurements data, an optimal transformer tap position is provided by GridQube's transformer tap adjustment algorithm.

In summary, the following are the findings of the investigations:

- Measurements from the transformers are significantly crucial for estimating premise level estimation with relatively high accuracy.
- Information on transformer tap settings is essential. If the tap information is not present, optimized tap settings must be obtained to improve estimation accuracy.
- Ideally, DNSPs should collect and maintain accurate tap position data if they are not already doing so.

Results of State Estimation

DOE Calculation

Calculating the dynamic operating envelopes (DOE) is one of the most common applications of GridQube's technology stack. DOE are defined as an allocation strategy that maximises additional export (generation) and import (load) that the network can accept at any one time without breaching any specified technical and operational limits. An important difference between the GridQube CCO-based DOE calculation and other approaches is its ability to calculate DOEs as an additional allocation over and above uncontrolled essential services, or as a total allocation to the entire premise. If operated in real time, it greatly reduces the need for worst-case scenarios to be used in DOE calculation due to extensive forecast uncertainties. In this situation, there is no absolute need for forecasting beyond the next DOE-time interval. This, in turn, permits significantly larger export or import allocations while also ensuring the DOE calculations respond to network events as they occur.

In this way, DNSPs have greater flexibility in enabling customers to participate in a DOE-based network management scheme and it also encourages the maximum number of technically feasible allocations.

For further information about the verification of the DOE calculations, refer to Appendix 3.

Figure 7 presents an example of the DOE allocation result for a distribution transformer, highlighting the variable nature of the DOE and the seasonal differences which could be expected. For example, if we consider the orange line which represents the import allocation for January, it is apparent that there are days where the already heavily loaded network could not support as much additional import as days in April when there was a lighter load on the network.

For Project SHIELD, dynamic operating envelopes were calculated for 10-minute intervals from 1–8 August 2021 and then stored on the Luceo Platform.

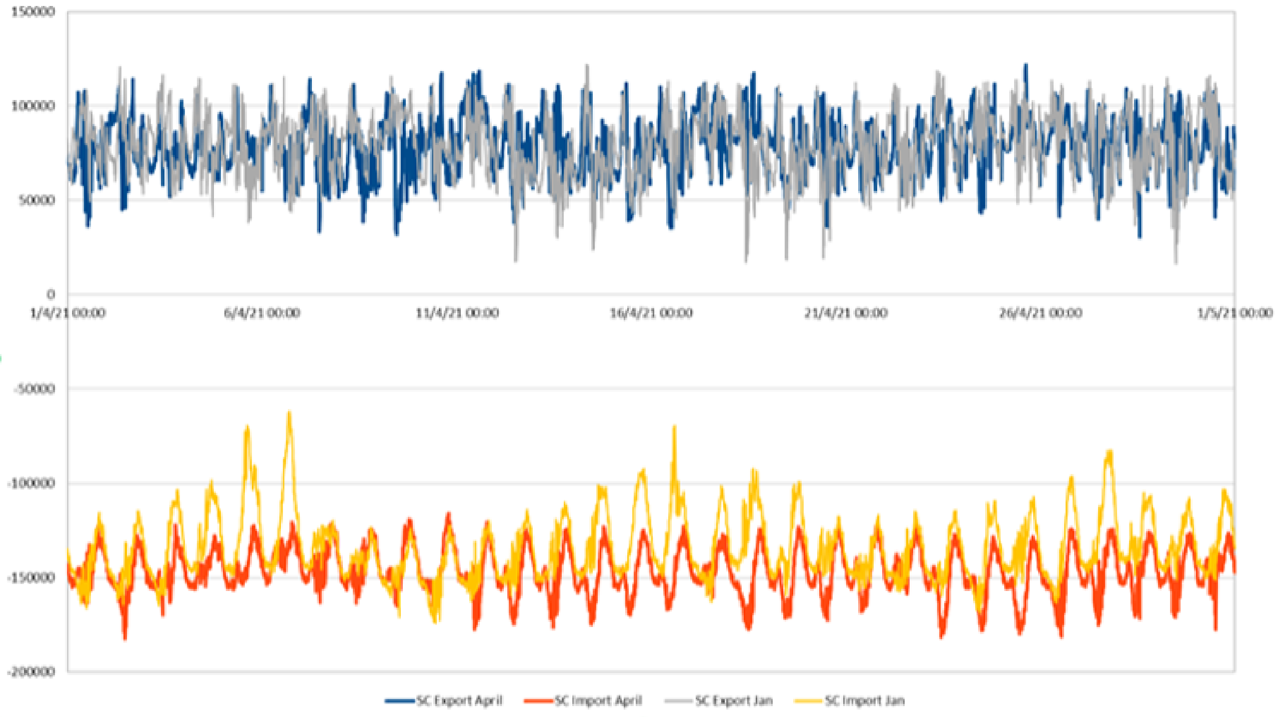


FIGURE 10 – Example of a DOE allocation with seasonal difference (blue, red: April, grey, orange: January) – Positive: export/generation, Negative: import/load

Viewing the Results

(Network Visibility Platform)

Once the state estimation and DOE metrics were calculated, they were returned via the API to be imported into the Luceo Platform.

Luceo provides a portal for visualisation of the data and its reporting requirements. A design workshop with members from the DNSP community to elicit ideas and requirements for the Network Visibility Platform was held. The design team then determined what parameters could be provided in the Luceo Portal.

Specifically, the portal includes the ability to view different devices and data sources, along with operating envelope metrics, using the existing suite of reporting capabilities. In addition, there are existing Table extraction capabilities which transparently incorporate reporting on the network model and telemetry data.

The Luceo Portal provides a network-centric view of the data using dashboards that may be custom built with interconnected 'widgets' that show such information as geographical, charts, alerts, statistics and metadata. As seen below in Figure 17, the top level displays the feeders for a DNSP.

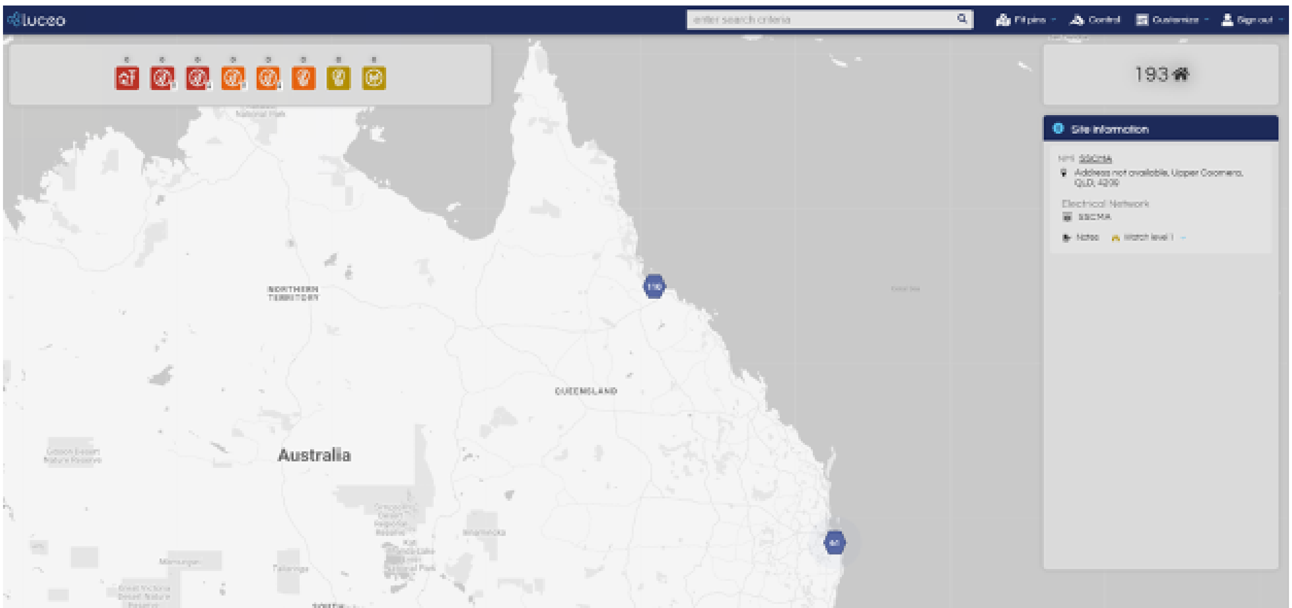


FIGURE 17 – Luceo Portal showing two feeders of a single DNSP

The Luceo Portal allows users to drill down through a network model and view different device types in a consistent manner as demonstrated.

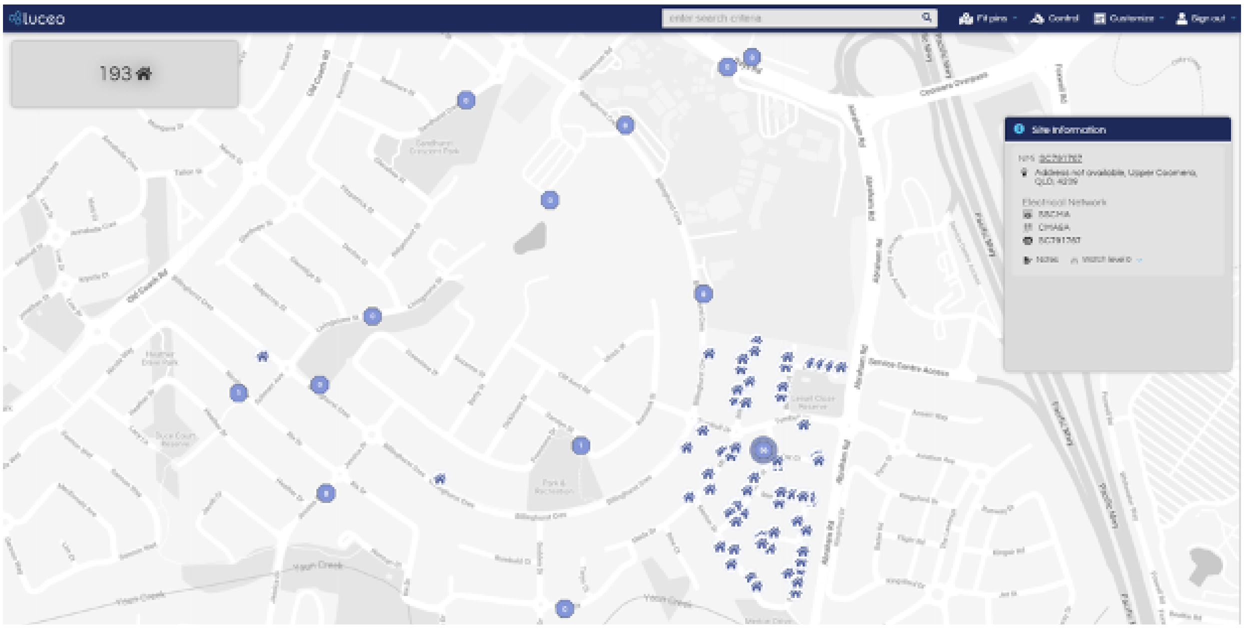


FIGURE 18 – Luceo Portal showing Upper Coomera feeder at different network levels

Site information can be displayed, including metadata and chart information, as well as incorporating operating envelope information as illustrated in Figure 19 – Luceo Portal showing transformer with DOE below. This information is available for any device type and its corresponding telemetry data.

FIGURE 19 – Luceo Portal showing transformer with DOE

Scenario Analysis and Simulation

The aim of the project was to investigate the impact of different measurement datasets on the accuracy of the state estimation results. In Phase 1, we focused on analysing the influence of data density (or the number of devices included) on the state estimation result. Given the small amount of data that we received, we are unable to conclude much about the optimal location and placement of these devices at this stage.

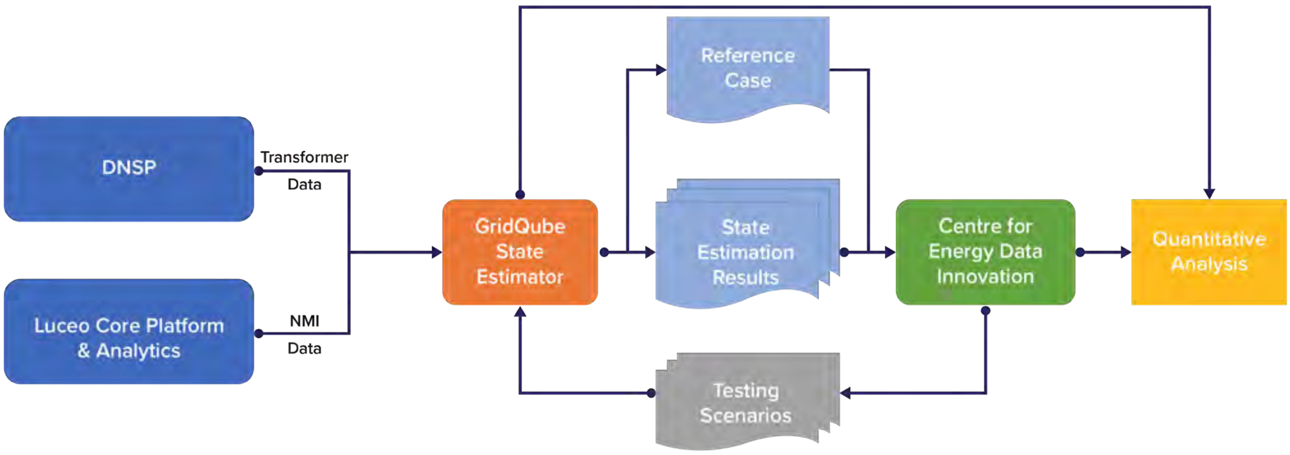


DIAGRAM 3

The influence of completeness of the measurements on network visibility was reviewed via scenario analysis by the University of Queensland's Centre for Energy Data Innovation (CEDI). As discussed above, GridQube created a reference measurements file by combining measurement data from the Luceo Platform and transformer data from the DNSPs and then passing on this information to CEDI. For the sake of simplicity, the dataset only contained the blended measurements at one single timestamp. Following the experiment design parameters, different scenarios files were then created from the reference file. The generated scenarios were passed back to GridQube for processing various scenarios in its estimation system in order to identify the networks' estimated operational parameters. Lastly, the state estimation results were sent back to CEDI where researchers conducted numerical analysis on the state estimation results.

Base Case

Initially, we created a base case as a proof of concept for the simulation. Data from 10 Energex transformers were collected for the base-case experiment (with no data from the Luceo Platform being included). A selected timestamp of 2020-09-30, 14:00 AM was used. There were 3 types of measurement associated with each transformer: phase-neutral phase angle, current and voltage. All 3-phase data was recorded. To investigate the impact of different levels of dataset completeness on the accuracy of the state estimation, we created different scenarios with varying levels of completeness. In particular, we gradually removed some transformer data from the whole dataset until

50% of the data visibility was achieved. In total, we generated 637 different scenarios. The original complete dataset was then established as the reference dataset.

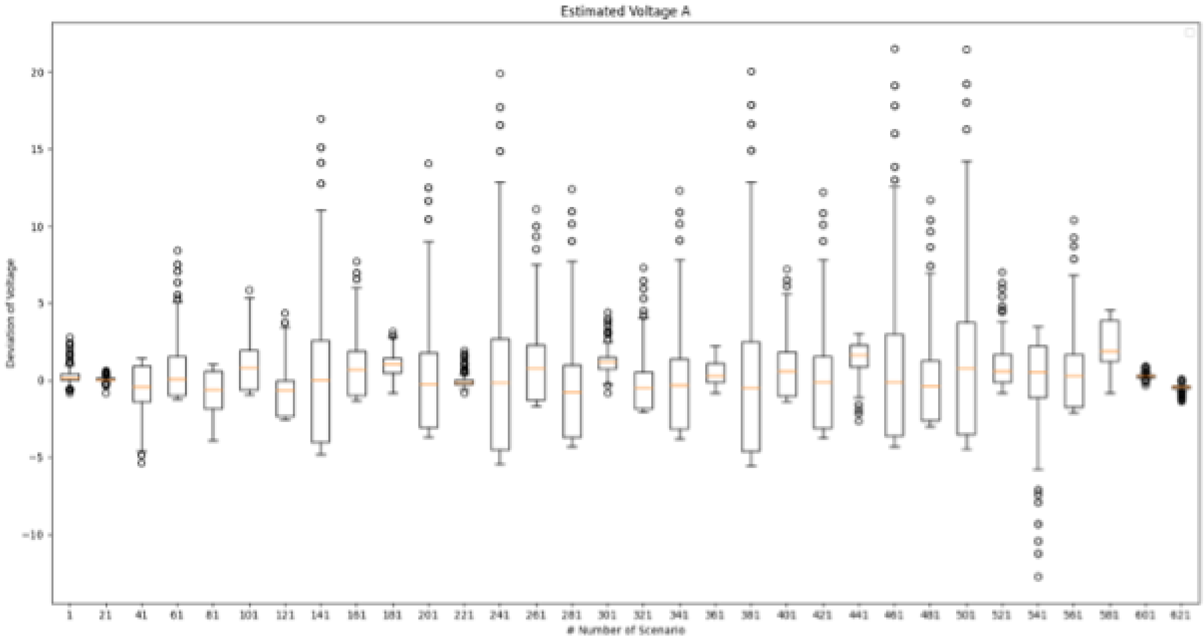


FIGURE 11 – Error deviation of voltage on Phase A

From the boxplot shown in Figure 11 the error deviation of the state estimation results between each scenario and the reference case was analysed. The scenario with smaller median deviation and smaller interquartile range (IQR) is superior to the other scenario. IQR denotes the distance between the upper and lower quartiles. In this example, the lower quartile is 25th percentile and the upper quartile is 75th percentile. We can also see that certain scenarios perform better than others which, in turn, shows that the accuracy of the state estimation is not proportional to measurement strictly in a certain data-density level. Some measurements carry more weight (importance) than the other measurements.

Based on this case study, we were able to design a scoring system that ranked the performance associated with each scenario. This scoring system contains 2 components: firstly, the deviation of mean of all measurements in each scenario is ranked against the reference case in ascending order. Secondly, the IQR is also ranked in ascending order. In both rankings, the smaller the score the better the accuracy.

The scoring system can be represented as:

$$\textit{Total Score} = \textit{rank (deviation of mean)} + \textit{rank (IQR)}$$

Scenario Cases for Upper Coomera Feeder

Data from 14 transformers and 53 devices were collected from the Upper Coomera feeder. The timestamp of 2021-08-08, 01:40 AM was selected to maximise the number of data points. To investigate the impact of different levels of NMI dataset completeness on the accuracy of the state estimation, we created 17 test scenarios, each with a different number of NMIs included. Initially, we removed three NMIs from the original 53 NMIs which left a total of 51 NMIs. Next, we slowly reduced the total number of NMIs by 3 for each new scenario. In this way, 6, 9, 12…,51 NMIs were gradually removed from the entire dataset. Following the random sampling strategy, the decision about which NMIs should be removed was made. The original dataset was then established as the reference dataset.

By observing the error deviation between the reference scenario and generated scenarios, we can investigate the correlation between the level of completeness of the dataset and NMIs. We can also study the importance of specific NMIs in relation to the state estimation result. In each result file, estimation of 2913 nodes were included. In this experiment, we focused on investigating the impact of data density on voltage magnitude.

Due to restrictions in mathematical computation, only 383 out of 1020 result files were received by CEDI for data analysis.

From the Figure 9, we can clearly discern that increasing the quantity of removed NMI data also leads to an increase in the average level of voltage magnitude deviation. The yellow lines show the envelope of this trend. We can conclude that in order to achieve satisfactory accuracy levels in state estimation, sufficient data density level is required.

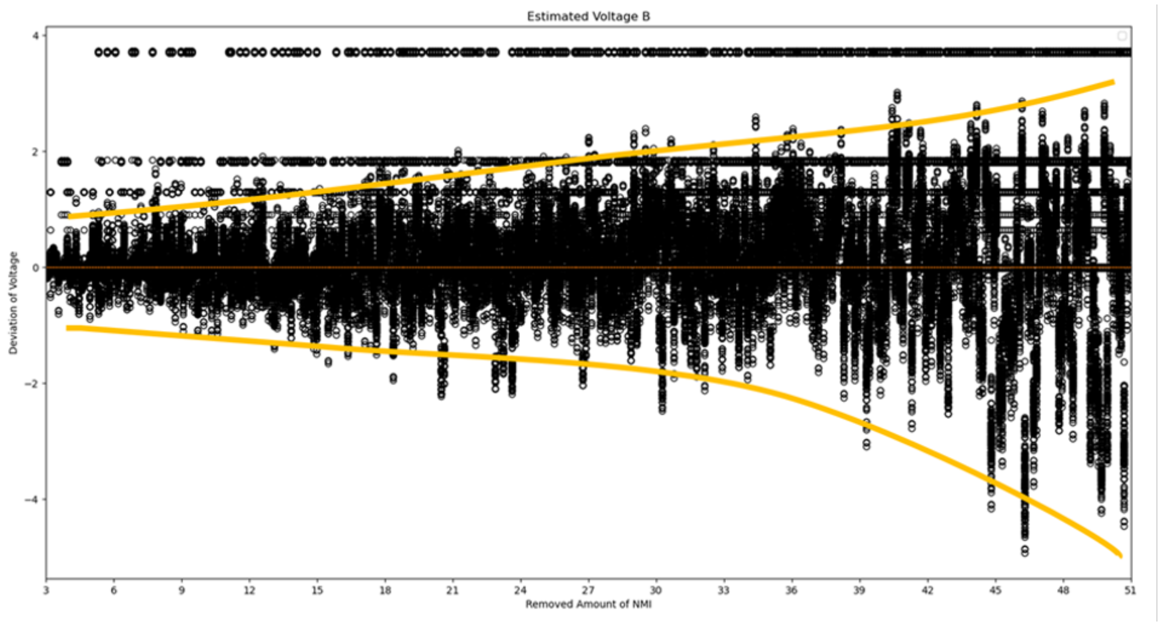


FIGURE 12 – Deviation of voltage magnitude on Phase B with different level of NMI data density

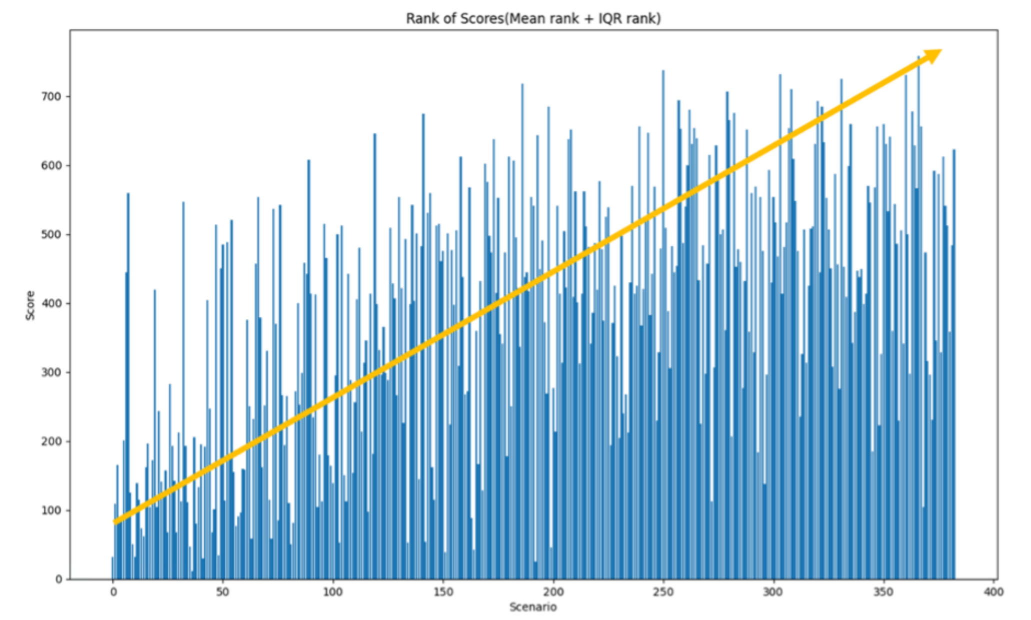


FIGURE 13 – Scoring of results for different scenarios with different level of data density

In Figure 10, we can observe the trend for ranking score to increase with a decrease in data density, with the smaller the rank score, the better. However, this linear relationship is not very strict within a small range of data density level. In other words, appropriate selection of measurements can help to achieve a satisfactory accuracy level when the data density is not high enough.

We analysed the top 20 scenarios (20 smallest) using the above scoring system. There were some measurements that were frequently removed in these 20 scenarios. The top 3 NMI that were removed are: **NMI 1**: 18 times, **NMI 2**: 17 times, **NMI 3:** 16 times. This result indicates that the measurements of these NMIs did not significantly affect the state estimate result.

FINDINGS SUPPORTED THE THEORY THAT STATE ESTIMATION WAS MORE ACCURATE WITH MORE DATA BUT ALSO THAT LOCATION OF THE DEVICE HAD AN IMPACT ON THE FINAL CALCULATION.

Commercial vs. Residential Devices

The project also investigated how additional measurements from different types of customer premises might improve the accuracy of the state estimation solution. This was done by developing scenarios mixing transformers' measurements with additional premise-level measurements.

In distribution networks, the pattern of residential customers (RC) loading is almost predictable, while commercial customers (CC) loading is much less predictable. Additionally, there are many more residential customers compared to commercial customers. To improve estimation accuracy, the project needs to consider installations of new measurement devices but needs to determine which types of measurements are more valuable than others.

Scenario Creation

Table 4 shows what measurement data was available for each type of customer. To narrow down the results, seven residential customers from the Upper Coomera feeder were initially considered, and two commercial customers as well.

Commercial (CC)/ Residential (RC)	Voltage Measurement Availability	Active Power Measurement Availability	Reactive Power Measurement Availability
CC1	YES	NO	NO
CC2	YES	NO	NO
RC1	YES	YES	YES
RC2	YES	YES	YES
RC3	YES	YES	YES
RC4	YES	YES	YES
RC5	YES	YES	YES
RC6	YES	YES	YES
RC7	YES	YES	YES

TABLE 4 – Deviation of voltage magnitude on Phase A with different level of NMI data density for Port Macquarie dataset

Additional measurements from residential premises were used with measurements from eleven transformers to estimate nodal voltages of commercial premises. The comparison indicated how additional residential premises measurements improved the voltage estimation for commercial premises. Similarly, a combination of commercial premises and transformer measurements aimed to compare the estimated nodal voltages with actual residential device measurements. To demonstrate the importance of the measurements of one type of premises in relation to another, four scenarios were developed with 120 timestamps, as follows:

- Scenario 4: Only available transformers measurements were used
- Scenario 5: Measurements from two residential premises were blended with transformers measurements.
- Scenario 6: Measurements from two commercial premises were blended with transformer measurements.
- Scenario 7: Measurements from all 50 residential premises were blended with transformer measurements.

Results

The state estimation results obtained from each scenario were subtracted from the actual device measurements. The difference was the deviation of estimation accuracy for the scenario. Scenario 4 was obtained using only the transformers' measurements, and therefore it was considered the reference case. Figure 14 shows the box and whisker plots of estimation accuracy deviation in volts for all four scenarios (4-7). For Scenario 5, the span of estimation deviations was identical to Scenario 4 for both commercial premises (CC1 and CC2). The solutions obtained from Scenario 5 were also slightly overestimated compared with the actual commercial measurements. Therefore, the additional two residential premise measurements did not improve the estimation quality of commercial premises even though it improved the accuracy for some residential premises.

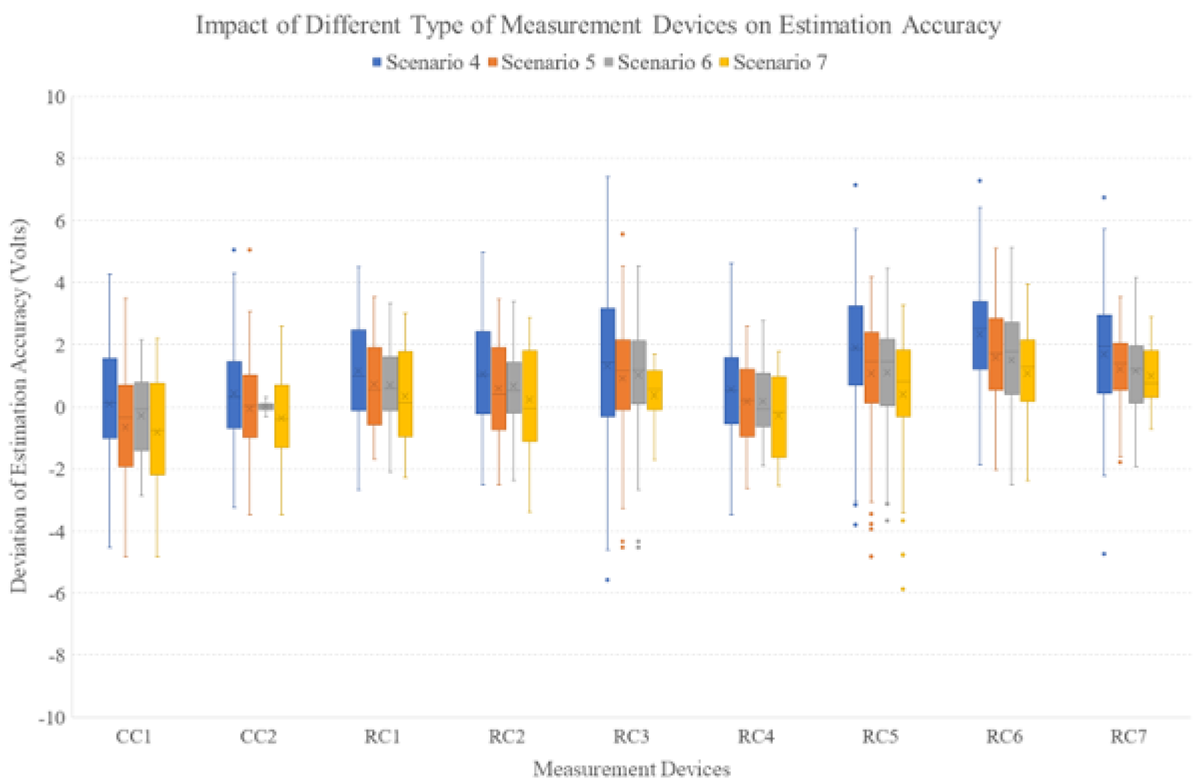


Figure 14: Assessment of Estimation accuracy using various types of measurement devices in Upper Coomera feeder through Scenario Creations.

In Scenario 7, a larger volume of residential premise measurements (increased from 2 device measurements to 50) was combined with the transformer measurements. Surprisingly, the estimation accuracy was not improved for either CC1 or CC2. This implies that residential premise measurements may not improve commercial premise estimations no matter how many devices are included. The residential measurements were mostly clustered around a single geographic area and the study was conducted on a very small subset of commercial measurement data so this finding warrants further exploration to determine if it is valid under other conditions.

Conversely, in Scenario 6, where two additional measurements were from commercial premises, the estimation accuracy was improved for most of the residential premises (RC1, RC2, RC4, RC5, RC6, RC7). Additionally, the increase in accuracy for CC2 was significantly noticeable, whereas the change in accuracy for CC1 was very slight. The lack of active and reactive power measurements from the commercial premises may have prevented an improvement in the accuracy for CC1.

Thus, it is clear that for this case study measurements from commercial premises improved the state estimation accuracy for both residential premises and commercial premises and were of greater value than residential premise measurements.

In summary, the following are the findings from this investigation:

- Commercial premise loading profiles are non-predictable, and therefore, measurements from commercial premises are of greater benefit in improving the accuracy of the estimated network state in distribution networks.
- Residential loading profiles are more predictable, and therefore, fewer residential premise measurements are required and adding more does not markedly improve the accuracy of estimation. Measurements from a minimum number of distributed residential premises would be sufficient.

Scenario Cases for Port MacQuarie Feeder

The Port MacQuarie dataset contains the data from three transformers and 142 NMI devices. The selected timestamp is 2021-04-15 08:00:00. As with the Upper Coomera feeder case studies, we generated scenarios based on different levels of NMI data density. In this case study, we established 10 different levels of data density and sampled 100 scenarios for each

level. In total, 1000 scenarios were generated to send to the state estimator. Initially, we re- moved two NMIs from the total of 142 NMIs which left 140 NMIs. Next, we gradually reduced the number of NMIs by 10 for each new scenario. The strategy for removal of NMIs involved a random sampling strategy. The original dataset was then established as the reference data- set. From this, 486 result files came back due to the restriction of the algorithms.

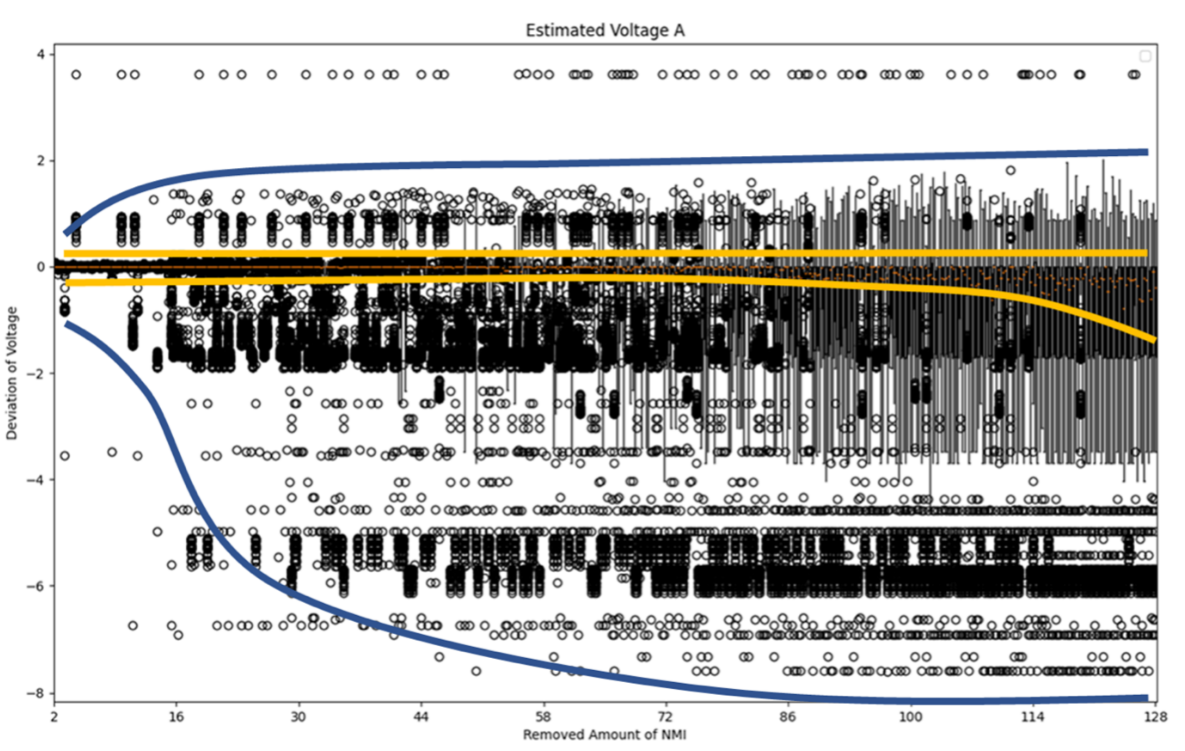


FIGURE 15 – Deviation of voltage magnitude on Phase A with different level of NMI data density for Port Macquarie dataset

From Figure 15, it is apparent that, along with the increasing number of removed NMIs, the deviation of both median and the outliers in Phase A is also increasing. This phenomenon matches what we observed in the Upper Coomera feeder case study. It demonstrates that it is important to maintain a certain level of data density in order to achieve satisfactory state estimation results. We also found that the divergence speed of median (shown by yellow envelope lines) is slower than the speed of outliers (shown by blue envelope lines). This indicates that if we place more significance on median in comparison to other attributes while analysing the network, we may expect to receive a higher level of data incompleteness.

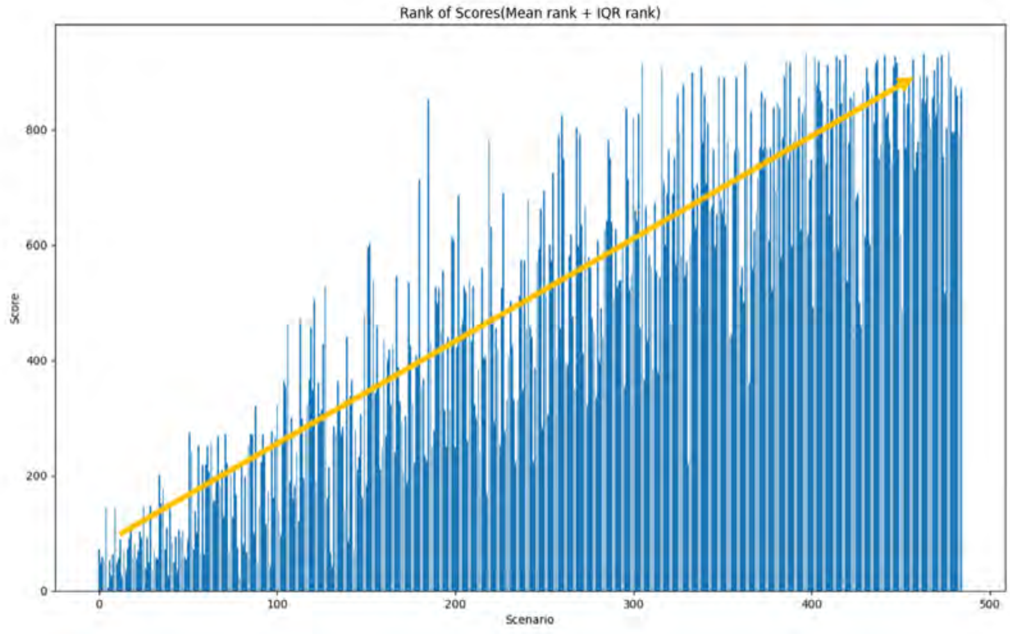


FIGURE 16 – Deviation of voltage magnitude on Phase B with different level of NMI data density for Port Macquarie dataset

In Figure 16, we can observe the same trend as in Upper Coomera feeder case study. Specifically, the ranking score increases with a decrease in the data density, with the smaller the rank score, the better. Basically, the less measurement that is removed, the more accurate state estimation result received. However, there are many scenarios that outperform other scenarios of similar data density which also indicates the importance of appropriate NMI measurement selection.

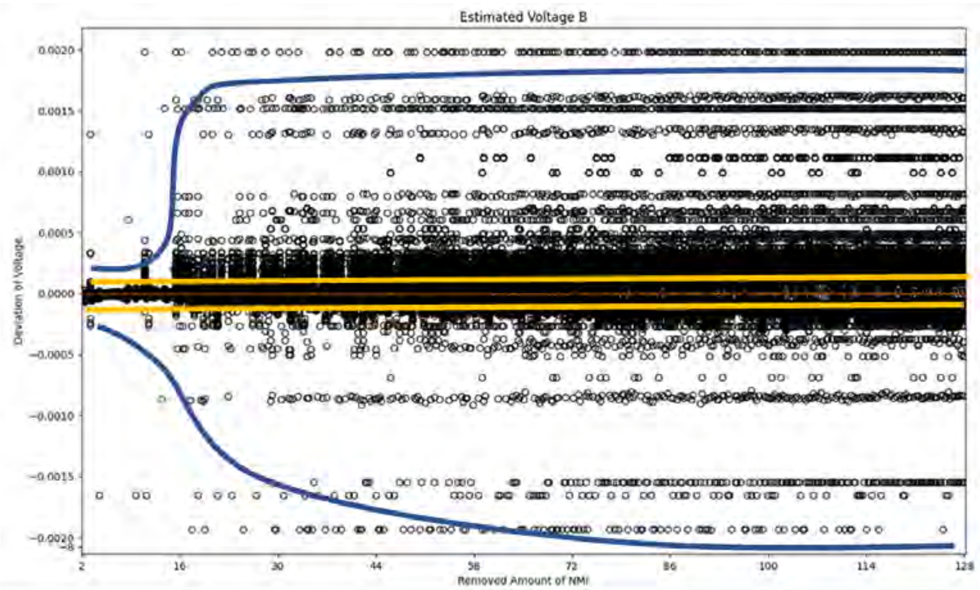


FIGURE 17 – Deviation of voltage magnitude on Phase B with different level of NMI data density for Port Macquarie dataset

The experiment in Phase B has shown some similarities with Phase A. For example, the deviation of outliers increases in tandem with the increasing levels of removed NMI data, and such divergent speed is very rapid. However, Figure 12 also illustrates a different median value pattern. The median has remained within a very narrow range with increasing data incompleteness. Moreover, in comparison with the experiment in Phase A, both median and outliers have a lower deviation level in Phase B.

There are some reasons that have resulted in this difference. Firstly, in the blended measurement file, all NMIs measurement are in Phase A so the state estimation results also show a clearer pattern in this phase. Secondly, in the Port Macquarie case study, the total number of NMIs is 142 which is significantly higher (i.e. almost 3 times the total number of NMIs present in the Upper Coomera feeder (CMA9A)). Consequently, the search space has also increased dramatically (as the dimension increases) and the number of scenarios created (1000) is considered to be a very tiny part of the entire space. As a result, it is not possible to obtain an incisive indication of voltage deviation.

Moreover, we are still faced with the lack of measurement in the Port Macquarie dataset. In the measurement dataset, there is no NMI with power measurements which could result in no significant difference in the estimated solution for different measurement scenarios. This could possibly be due to a lack of measurement data which means that the results diverge at such a slow pace. Thirdly, there is no knowledge about the phase and tap information. For this feeder, there is no information on phase identification and tap settings for NMIs and transformers. Thus, the data received is considered insufficient for a high-quality estimation solution.

Lastly, in the Port Macquarie case study, the total number of transformers with measurements is 3. One transformer data was not able to be used due to unreliable data patterns. Consequently, only two transformers' measurements could be used as input for the estimator. Transformer measurements are considered to be very significant in state estimation. Therefore, this could be another important reason for the Coomera feeder not to have achieved outcomes.

From the Upper Coomera feeder case study and Port Macquarie case study, we have seen shared results and consistent patterns in relation to state estimation accuracy and data density.

Conclusions

Network visibility is the prerequisite for calculating a DNSP's hosting capacity. To optimize the use of their assets while improving their network visibility with a high degree of accuracy, we focused on providing a strategic guideline for DNSPs.

The results found are based on extensive analysis of the limited measurement dataset and information gap in the selected project feeder. To validate this analysis, further work must be conducted on acquiring additional points of measurement. This will allow the initial outcomes to be refined and confirmed to an agreeable level for the DNSPs to be able to confidently make some decisions.

From above experiments and numerical analysis, we have reached the following conclusions:

- To achieve required levels of state estimation accuracy, a certain level of data density should be maintained.
- At the same or similar data density level, different scenarios performed very different- ly. Appropriate measurement selection can make a difference in terms of reaching a cost-effective solution.
- To obtain high-quality analysis results, the quality and completeness of the measure- ment dataset is important. Comprehensive attributes are very helpful in order to obtain valuable insights from the dataset.
- Transformer measurements are critical, and it is recommended to obtain them from every distribution transformer of a feeder if possible.
- A transformer's tap position is highly valued information and contributes to improving estimation accuracy.
- Measurements from the non-predictable customers, or commercial premises, are second-most important with high density.
- The measurements from predictable customers, or residential premises, are the third most important with less density.
- We have proposed an effective quantitative tool for ranking the scenarios. Using this tool, we can identify the optimal combination of measurement set at a certain level of data density.

Next stage of scenario analysis:

- In the next stage, we will use our knowledge of distribution-network topology to investigate the impact of NMI device locations on the state estimation result.
- We will validate our finding with real-world datasets to provide guidance for new NMI installations.

FINDINGS SUPPORTED THE THEORY THAT STATE ESTIMATION WAS MORE ACCURATE WITH MORE DATA BUT ALSO THAT LOCATION OF THE DEVICE HAD AN IMPACT ON THE CALCULATION. SPECIFICALLY, MEASUREMENTS FROM COMMERCIAL PREMISES PROVED MORE VALUABLE THAN ADDITIONAL DATA MEASUREMENTS FROM RESIDENTIAL PREMISES.

Conclusion and Recommendations

The first year of Project SHIELD has been very challenging. Covid has added a layer of complexity to our collaborations, particularly between our technical staff who benefit greatly whenever they have been able to be in the same room to share and explore the progress of their work. In addition, access to data and the speed at which data was available has greatly tested the technical team. As we approached our project milestone, floods, associated widespread power problems and a Covid resurgence threatened the project yet again. Despite this, a great deal of progress has been achieved.

Stage 1 of Project SHIELD's outcomes are summarised below:

- Many data source providers were initially identified and were generally interested to assist us with the project. As the project team worked with these suppliers, acquiring customer-level data proved to be increasingly challenging due to lack of data in the trial areas, concerns over privacy risks and risks related to intellectual property and interpretation of legislation. There were also issues related to resources to extract the data required to work with the project. The ESB independently identified such problems in their data strategy and are developing unified approaches to resolve any future issues.
- While the Luceo Platform was able to import engineering data from a range of heterogenous sources, the project found that this data was provided in a wide variety of formats and so it often required a different importing process for each source and pre-processing of datasets.
- A large amount of the data we acquired actually ended up providing inaccurate location or voltage measurements and, hence, did not offer equal levels of value to state estimation calculations.
- Initial analysis supports our expectation that more data sources will result in more accurate hosting estimates and that location of the device on the network, as well as the type of data source, has an impact on calculation accuracy.

In conclusion, we believe that further analysis is justified; however, the project has demonstrated that more accurate and larger amounts of data is needed to provide DNSPs with the answers they require.

APPENDIX 1

Potential Data Sources

Potential data sources identified by the project team.

Source	Provider
AMI	DMI Atlas MK10D – AGL EDMI EM1200 ETSA EDMI(EQL) Landis and Gyr E350 Citipower Powercor SP Ausnet Landis and Gyr (Ausgrid) PRI (United) Secure (Jemena) ISKRA
HEMS	Smappee Efergy Carbon Track AlphaESS EverGen GELI Gswitch Control4 Wonder Systems Solarhart Gateway
Inverters	Fronius SMA ABB Delta Enphase Sonnen LG
Network devices	SCADA Distribution Transformers Noja Power synchro phasers
EV chargers	Tritium CChargePoint Chargefox

Accredited meter data providers	ActewAGL Distribution Active Stream Pty Ltd (subsidiary of Plus ES) AMRS (Aust) Pty Ltd Acumen Metering Ausnet Services CitiPower Pty Ltd Essential Energy Energex Ltd Ergon Energy Corporation Ltd Endeavour Energy ActewAGL Distribution (trading as Evoenergy) IntelliHUB Pty Ltd Jemena Asset Management Pty Ltd Jemena Electricity Networks (Vic) Limited Metering Dynamics Pty Ltd Metropolis Metering Services Pty Ltd Plus ES Powercor Australia Ltd Powermetric Metering Pty Ltd SA Power Networks Secure Meters (Australia) Pty Ltd Select Solutions Group Pty Ltd Skilltech Consulting Services Pty Ltd Tasmanian Networks Pty Ltd (trading as TasNetworks) United Energy Distribution Pty Ltd Vector Advanced Metering Services (Australia) Pty Ltd
IoT	Solar Analytics Watt Watchers Luceo Edge Electrons Switchdin Phi Saver Reposit Solar Edge BillCap
ARENA projects	My Energy Marketplace
Other	NBN UPS nearby electric rail monitoring

APPENDIX 2

Invitation to to join Project SHIELD

The following communication was sent out in September 2021 via email.

Good morning,

<< Company Name >> is a well-known and trusted supplier to many Australian households and your technology is at the cutting edge of the digitisation of our energy industry. We suspect your products may collect data that could provide insights into the functioning of our neighbourhood electricity grids as they transition to a renewable future.

This letter is inviting <<company name>> to participate in a collaborative industry project designed to maximise the amount of renewable energy integrated into our neighbourhood networks. Your participation will also demonstrate and establish the value of your data to new markets (networks and network operators).

Project SHIELD is an ARENA-funded investigation aiming to unlock the potential of existing data sources like yours by creating visibility of the size and impact of local renewable generation in our suburbs. These insights will assist grid operators to better manage this renewable energy in real time as more installations occur.

The project may provide a pathway towards monetising existing data, which could in turn fund more services for your customers or new revenue streams to share with your customers.

Your data could be a vital and valuable part of this interim solution that will assist until the network is fully digitised. Every additional data source will increase the accuracy of our calculations.

Specifically, if your datasets include measurements of household voltage, active power or apparent power then we want to talk to you.

We estimate that within the Project's three trial suburbs (Townsville, Gold Coast, Port Macquarie) you may only have a small number of customers with valid data sources, as a result we don't expect the collation of this data to be a time consuming or challenging exercise. When eventually combined with other data sources, the Project will establish valuable insights while also demonstrating the underlying value associated with your data.

We understand the privacy and legal considerations associated with sharing data such as this and have a number of options to enable the supply of data. We are open to rewarding your customers and helping to cover the costs of your participation.

For our part, we will protect all data provided, ensure your organisation is kept updated on the project progress and recognise <<Company Name>> as a participant in this project.

The project will be of great interest to electricity networks globally and our partnerships with ARENA (Australian Renewable Energy Agency) and ENA (Energy Networks Australia) will ensure the results and associated value are widely discussed across the industry.

To express your interest or for further information, email us at contact@projectshield.com.au no later than 10 September 2021. Members of our team may also be reaching out to you to follow up.

APPENDIX 3

Dynamic Operating Envelope Calculations & Verification

The dynamic operating envelope (DOE) represents the amount of additional export or import the network can accommodate at a particular point in time without exceeding technical limits. In this way, export DOE represents the additional electricity generation the network could accept at a particular time whereas import DOE reflects the additional load the network could support in the form of battery or electric vehicle charging.

The project recommended that an independent check of GridQube's methods should be undertaken to verify that collectively adding exports to the limit of the determined envelopes would not exceed the technical network limits. PowerFactory (a power flow analysis software application widely used by distribution-network businesses) was utilised for this purpose.

The process is outlined below:

- Run Distribution System State Estimation (DSSE) to generate estimated network states over a 24-hour period. DSSE was supplied 10-min average voltage, current and phase angle data from available distribution transformer monitors. This resulted in a sequence of network states in 10-min timesteps for a single 11kV feeder for a single day.
- Import the network states into PowerFactory. Active and reactive power measurements generated by DSSE were applied to the 11kV terminals which corresponded to each distribution transformer's connection on the feeder. The voltage at the feeder start was also imported.
- Import the export-DOE allocations (assuming maximum export). The distribution transformer loads were reduced in line with the export DOE at each timestep to reflect additional generation exported from customers to the network.
- Run the PowerFactory power flow analysis as a quasi-dynamic simulation over each 10-min timestep in the 24-hour study period.
- Identify if network technical constraints were breached in any of the timesteps.

The generation of export-DOE allocation relies on specific technical and operational limits in relation to the CA and CCO processes. In this instance, the limits imposed were:

- 11kV +/- 5% for 11kV node phase-to-phase voltages
- 216V to 253V voltages at secondary side of distribution transformer terminals
- 130% kVA rating for distribution transformer
- 263A current limit on all 11kV branches (specific to network analysed)
- 216V to 253V voltages at the end of simulated LV lines

The PowerFactory results were inspected to determine if the export-DOE allocation led to any of the above limits being breached.

- 130% kW rating for distribution transformer in load direction
- 80% kW rating for distribution transformer in reverse/back-feeding direction

RESULTS – Export Scenario

11kV Feeder Limits

Figure X presents the PowerFactory quasi-dynamic simulation voltage results for the 11kV voltages measured at all distribution transformer primary terminals. Red lines reflect the ± 5% 11kV phase-to-phase voltage thresholds. The network remains within the desirable voltage range for the period analysed.

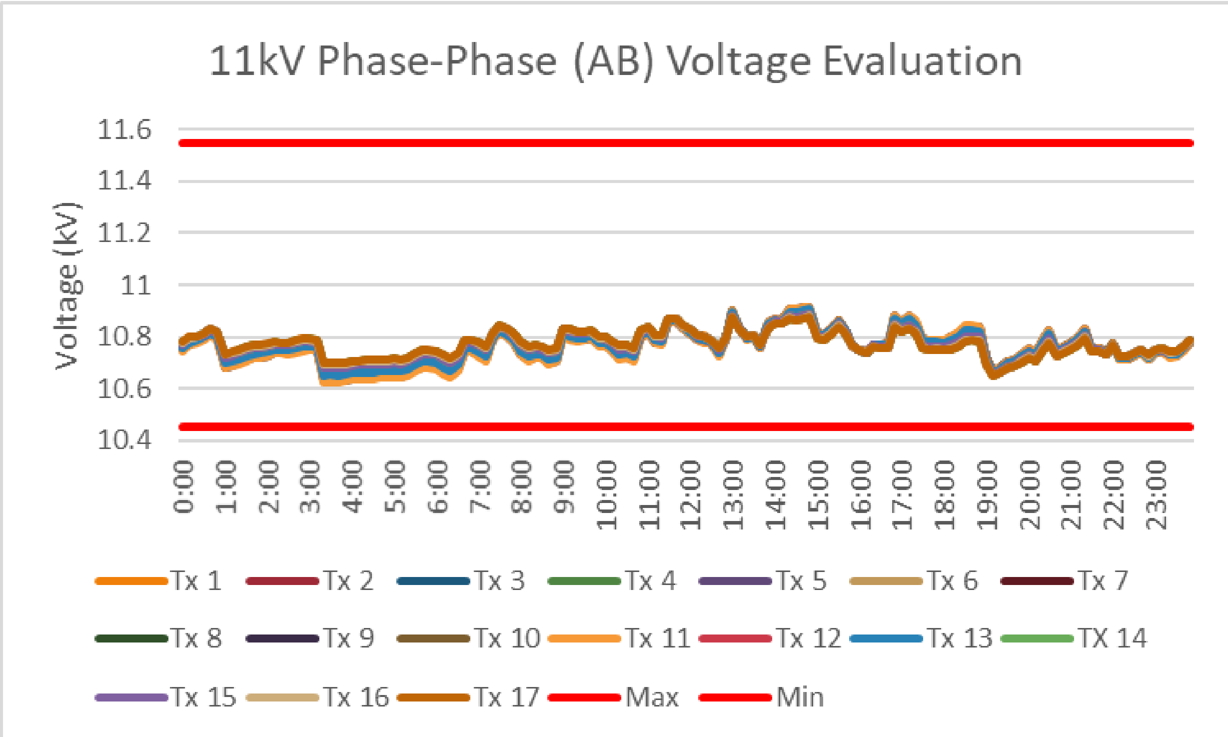


FIGURE X – (a) Voltage measurements Phase A – Phase B

Figure X (a-c) – PowerFactory phase-phase voltages measured at the primary terminals of distribution transformers for the DOE-export evaluation.

The current within all 11kV branches remained below the limit of 263A. The limit of 263A is specific to the feeder studied and this value will vary for other networks.

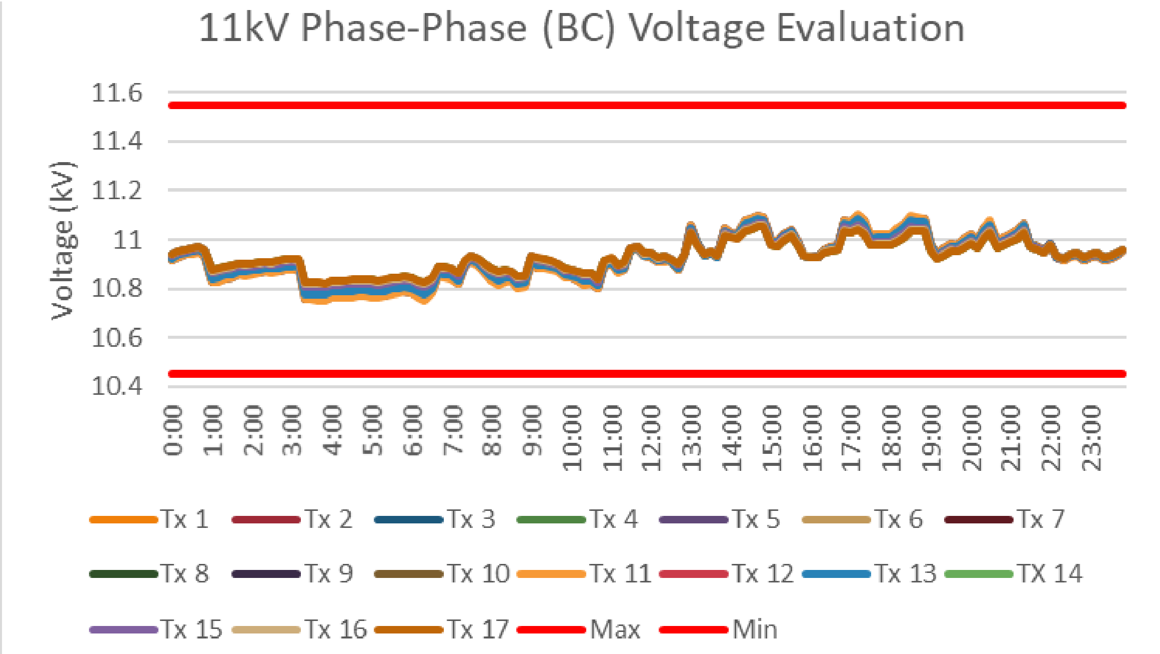


FIGURE X – (b) Voltage measurements Phase B – Phase C

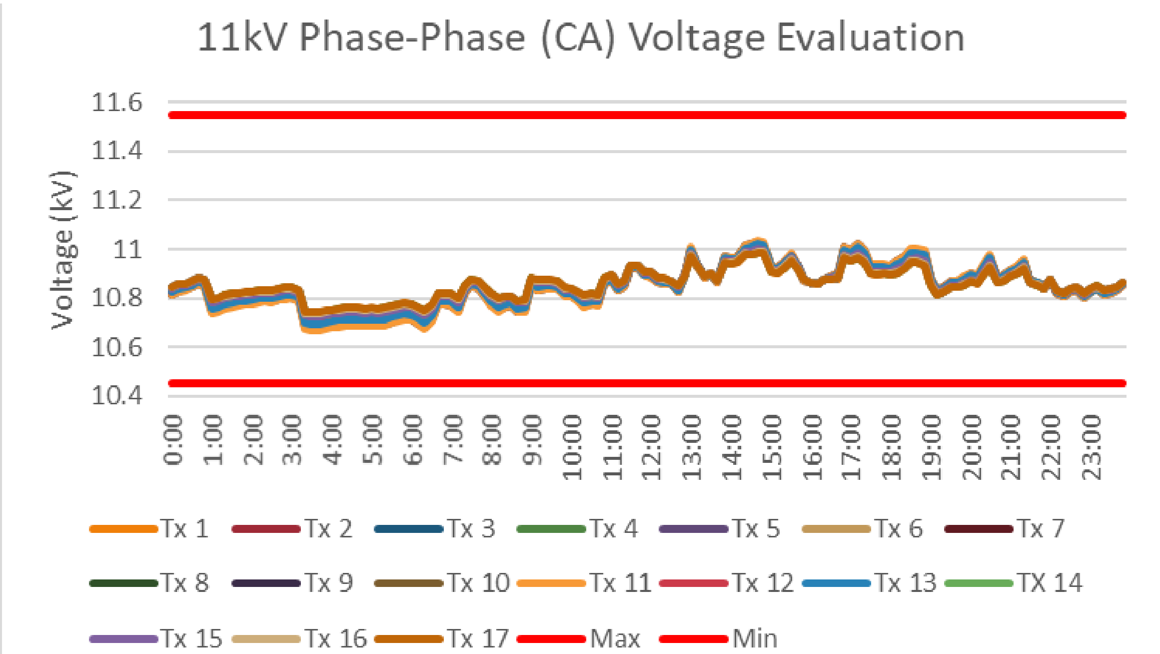


FIGURE X – (a) Voltage measurements Phase C – Phase A

Distribution Transformer Limits

Distribution transformer total apparent power remained below 130% of the transformer ratings. This was verified for all 17 transformers ranging in size from 200kVA – 500kVA (see Figure Y).

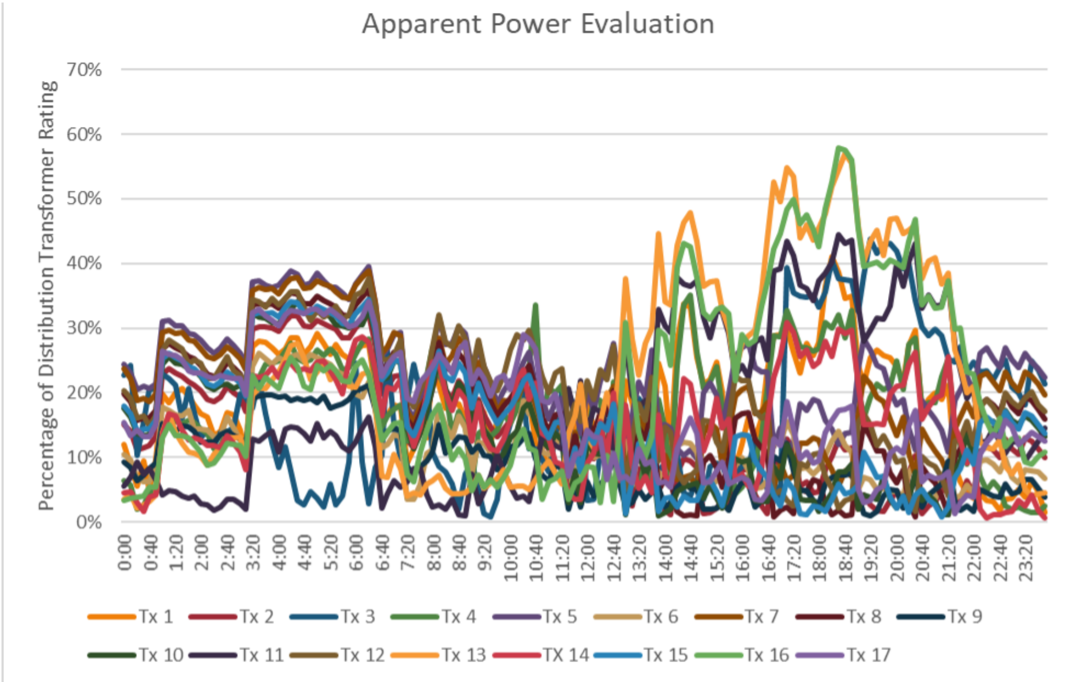


FIGURE Y – Apparent Power (kVA) as a percentage of distribution transformer ratings

Forward and reverse limits on active power were assessed. Results indicate 130% and 80% limits were maintained under forward and reverse active power flow, respectively (see Figure Z).

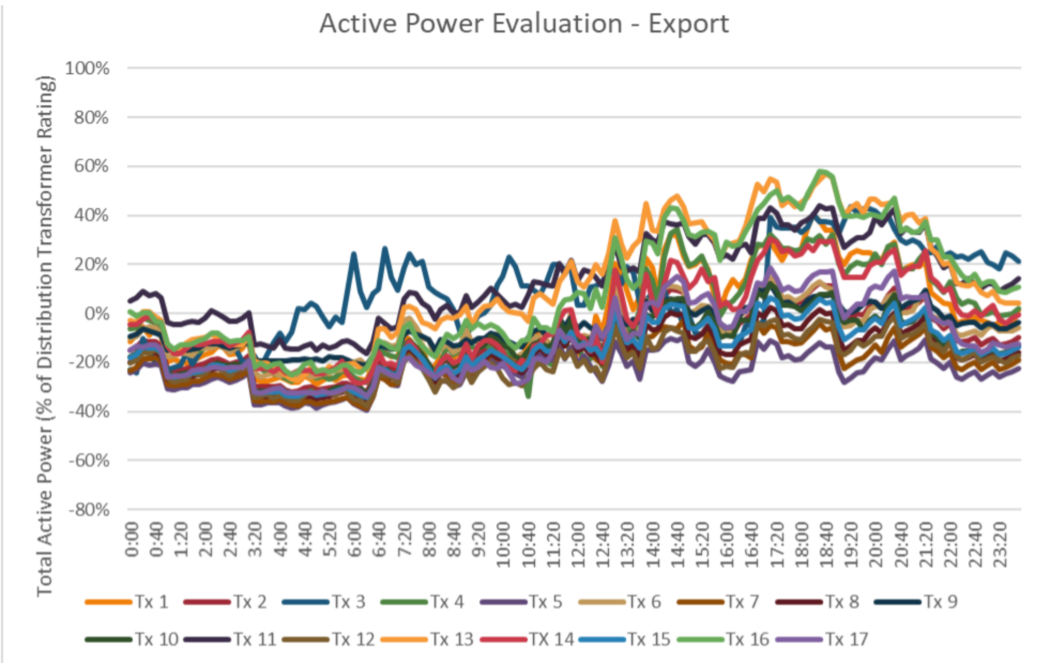


FIGURE Z – Distribution transformer active power as a percentage of transformer rating, DOE-export scenario

LV Voltage Limits

The voltages at the secondary terminals of all distribution transformers were evaluated to determine if they remained within the 216V–253V range under all the DOE-export scenarios for the 24-hour period analysed.

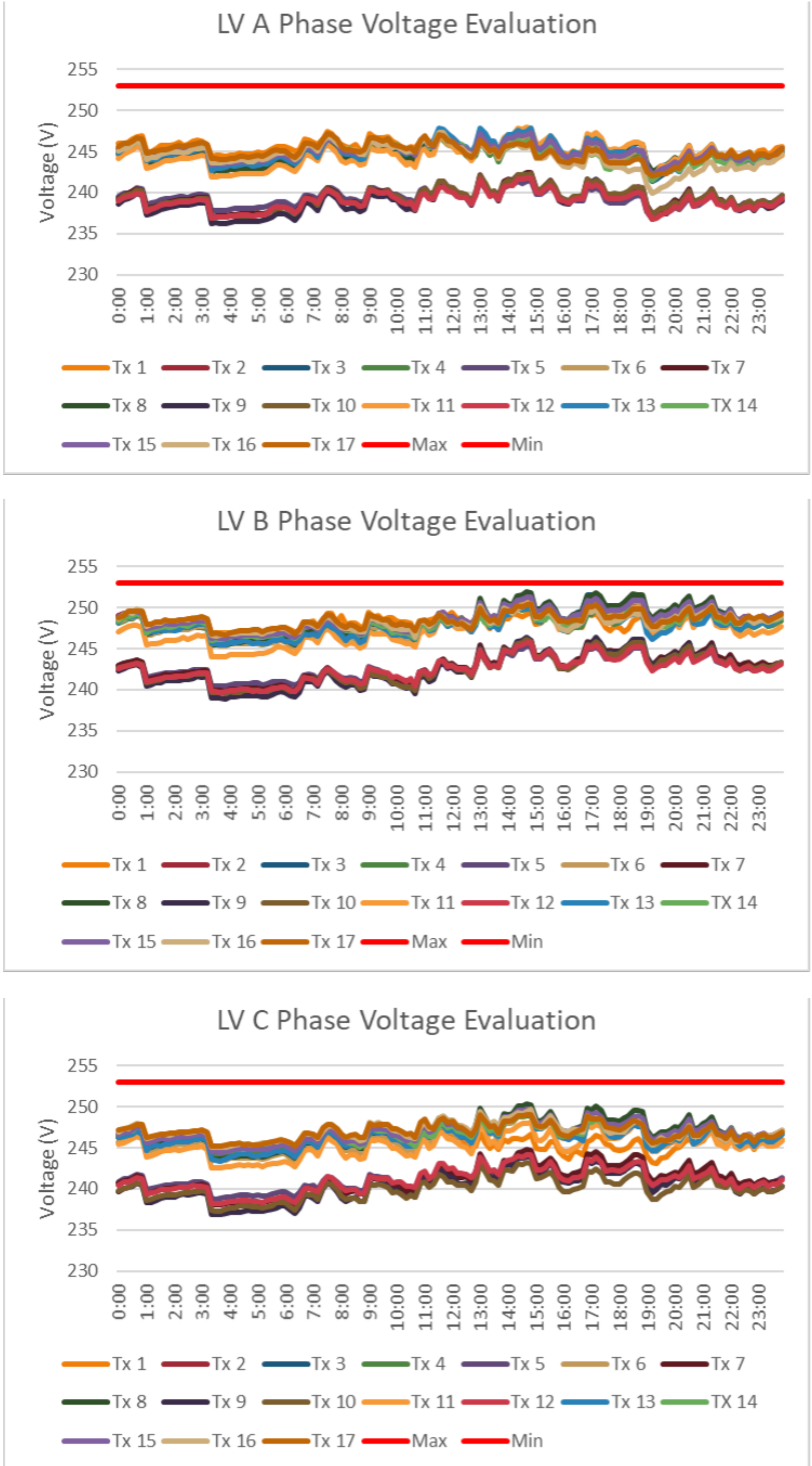


FIGURE AA – Voltage measurements at distribution transformer secondary terminals – DOE-export scenario.

The PowerFactory model available did not explicitly model the low-voltage network. Simulated end of LV line voltages could not be verified.

Results from this evaluation of GridQube’s calculated dynamic operating envelopes indicate that for the 24-hour period analysed (which equates to 144 individual network states and corresponding envelopes), the technical and operating limits of the network were maintained.