The sea of data: Orion’s approach to leveraging large datasets for enhancing network management practices

Yuyin Kueh

EEA Conference & Exhibition 2024, 10-12 September, Christchurch

Abstract
Like many Electricity Distribution Businesses (EDBs), Orion has been on a journey to increase its LV network visibility to prepare for increased adoption of electric vehicles (EVs), solar generation (PV) and energy storage solutions. From a long-running LV monitoring programme to becoming the first EDB in New Zealand to reach an agreement on the provision of 5-min smart meter data from Bluecurrent, we are now collecting data from network assets, meter and behind-the-meter sources.

However, visibility is only one piece of the puzzle. Leveraging these datasets to provide meaningful network insights requires the development of a robust analytics strategy and resilient data infrastructure to support it. In this paper, we will provide an overview of Orion’s Network Transformation Roadmap workstreams, our approach to developing power system insights from diverse datasets, the process of prototype to production and the challenges encountered along the way.

By harnessing the power of advanced analytics, EDBs of the future will be able to proactively identify and mitigate potential grid congestion, optimise asset performance, and anticipate future demand trends with a new level of precision. Advanced analytics can also enable the development of predictive maintenance strategies, minimising downtime and maximising operational efficiency for the benefit of the communities we serve.

Theme(s): Artificial intelligence, automation, digitisation, data and communication

1. Background

Orion’s low voltage (LV) network makes up approximately a third of its total network length across all voltage levels and supplies over 99% of its customers. However, existing methods of LV network management rely on lagging indicators which do not provide a comprehensive forward view of LV network performance.

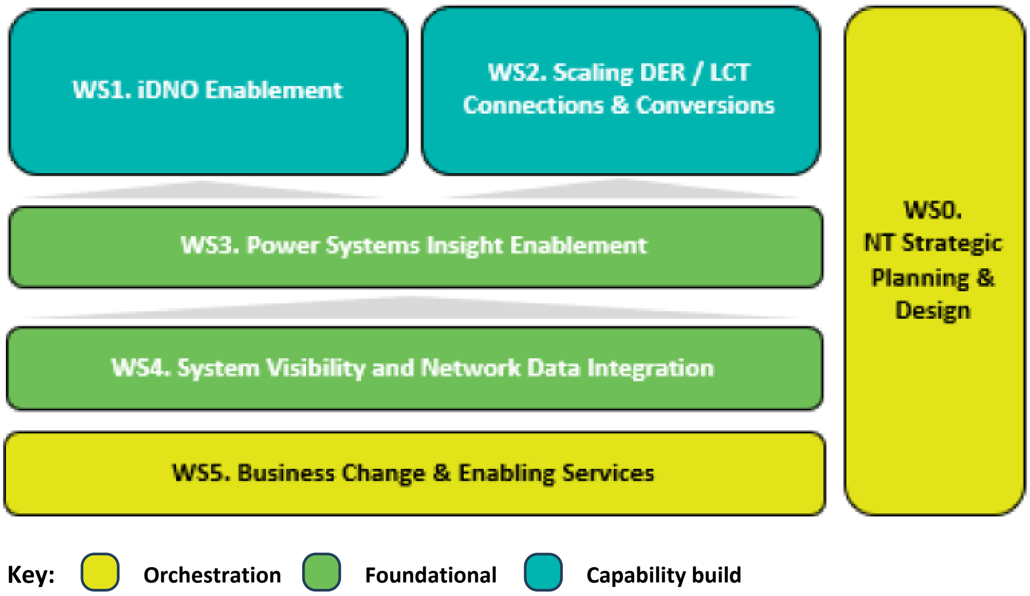
The Central Canterbury region is experiencing increasing levels of housing infill and adoption distributed energy resources (DERs) such as electric vehicles (EVs) and photovoltaic (PV) generation. Where these connections are added into existing network, capacity can be gradually eroded and result in customers experiencing degraded service in the form of power quality issues, solar curtailment or unplanned outages due to asset faults. These issues are often not evident through the traditional approach of monitoring transformer peaks and can lead to reactive and unpredictable expenditure.

Therefore, there is a need to adapt our network management approach to address issues more proactively which requires a significant uplift in network visibility and data-driven insights.

2. Introduction to Network Transformation Roadmap

Orion's Network Transformation Roadmap (NTR) has been developed to drive the business change required to for Orion to transform into an intelligent Distribution Network Operator (iDNO) and be an enabler to the energy transition by 2030.

The roadmap is structured into six workstreams (WS):



WS0. NT Strategic Planning & Design – Focuses on providing direction for the foundation and capability build workstreams by ensuring alignment to the Orion Group strategic direction and changes in the regulatory environment.

WS1. iDNO Enablement – Focuses on enabling the intelligent Distribution Network Operator (iDNO) capability required to seamlessly maximise the throughput of energy across our network with high penetrations of DER / LCT under utility-led, market-led or price-led operation / enablement modes.

WS2. Scaling DER / LCT Connections & Conversions – Focuses on ensuring our network enables / does not block the path to customers decarbonising at scale and pace, by efficiently accommodating on our network the growing numbers of renewable energy sources, electric vehicles, energy storage, and other Low Carbon Technologies (LCT).

WS3. Power Systems Insight Enablement – Focuses on enhancing capability in the Engineering Technology Domain (ETD) through integrated systems, spatial/temporal data analytics, and network modelling.

WS4. System Visibility and Network Data Integration – Focuses on ensuring comprehensive situational awareness of network utilisation and performance through an uplift of asset information and network visibility from behind-the-meter to GXP level, including bidirectional energy flows from DER / LCT.

WS5. Business Change & Enabling Services – Focuses on ensuring that there is effective planning, communication and orchestration of the NT business change.

To achieve the goals of scaling DER / LCT connections and enabling Orion to function as an iDNO, enhancing our capability for data-driven decision-making is crucial. Reliable data is fundamental to enabling this cultural shift. This paper provides an overview of work to date on key initiatives in Workstreams 3 & 4 to support this goal.

3. Key initiatives

3.1. Network topology & visibility improvements

Proactive management of the LV network requires significant uplift in LV data quality (both topology and visibility) and robust processes to ensure that quality is upheld.

Over the past few years, Orion has been focusing on increasing visibility through LV monitoring and smart meter data acquisition. We have now installed LV monitoring on approximately 12% of ground-mount distribution transformers and, in October 2023, became the first EDB in New Zealand to reach an agreement with Bluecurrent for the purchase of 5-min Network Operational Data (NOD).

Bluecurrent meters provide coverage to over 90% of Orion's ICPs, offering unprecedented visibility of LV network performance. However, coverage across the network is not uniform, with considerable blind spots around Banks Peninsula and other pockets throughout Christchurch City (Figure 1). Therefore, handling of missing data is an important consideration when carrying out analysis with this new data source.

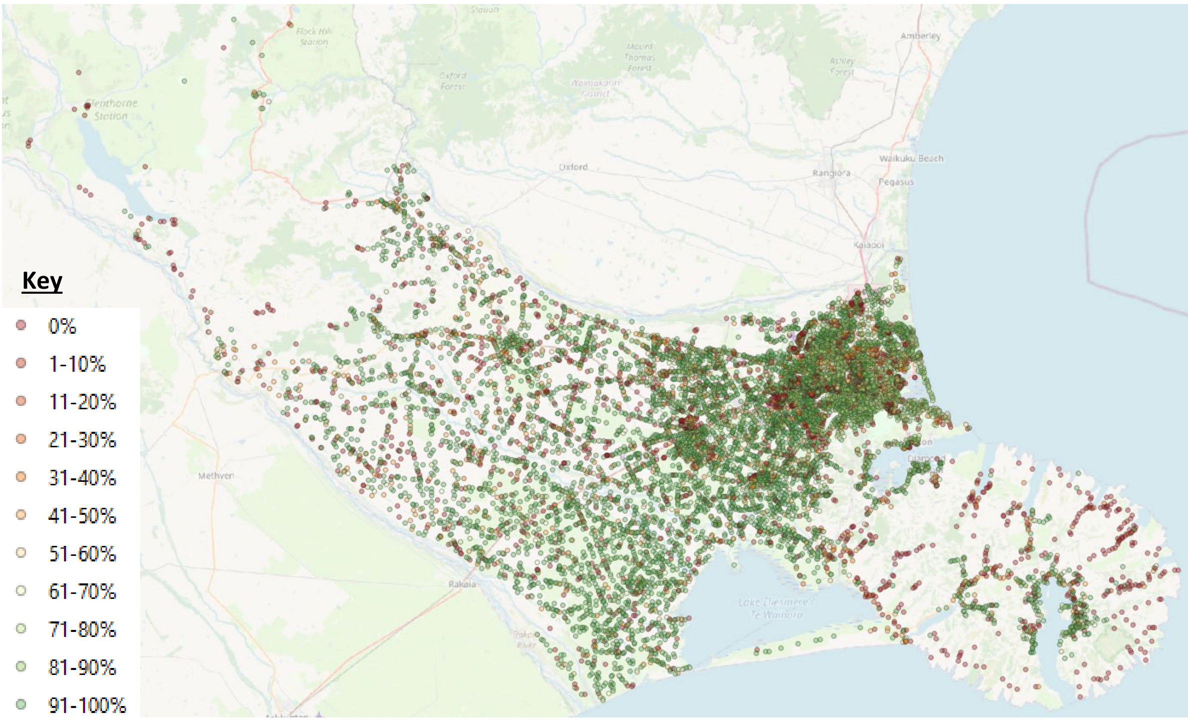


Figure 1: Bluecurrent smart meter penetration percentage across the Orion Network (by distribution transformer)

As visibility of the LV network increases, the gaps in topology data such as customer phasing, cable/line types and network configuration become the limiting factor in accurate modelling of the LV network and future scenarios. On Orion's network, approximately 40% of LV overhead line types are listed as 'unknown' and only 20% of customers have field verified phasing information. Service main sizes are also not well documented and we experience significant lag in updates to LV connectivity records.

We are currently undertaking field work to improve the quality of our LV network topology records. These datasets can then be used to validate the accuracy of topology correction algorithms to impute missing or incorrect records across the wider network. Generation of these synthetic datasets can save a significant amount of time for field teams. However, they require a distinct handling and tagging process to ensure that field verified records are given preference, where available, and never overwritten by an estimated value.

3.1.1. Field phase identification

In 2023/24, field phase identification was carried out in the Milton zone substation supply area using a non-contact proximity meter along the street. The survey covered ~14,000 ICPs with a success rate of 64%. Some common issues encountered with being able to identify the phase of specific customers were cabled service mains direct off overhead lines, missing or incorrect ICP labelling and multi-unit installations. Further overboundary work is now underway to complete the remaining ICPs.

3.1.2. Overhead conductor survey

To benchmark the accuracy of LV line data, an LV conductor survey was also carried out in the Milton zone substation supply area. The survey covered 2,147 spans of network lines as well as 1,716 overhead service mains to assess the quality of our existing data and collect up-to-date records. Survey results for network lines showed an accuracy of 81.8% while service mains were only 8.5% accurate (the remainder being either unrecorded or incorrect). Further survey work is pending upgrade of our GIS platform to ensure results can be incorporated into our master GIS dataset.

3.2. Network data analytics

Telemetry data such as HV SCADA points and LV monitoring provide ‘visibility’ of the electrical performance at different parts of the network. However, these datasets consist of large volumes of time-series observations which cannot be quickly analysed through manual means. Smart meter data compounds this issue by adding around 300 observations/ph/ICP/day (equating to approximately 80 million additional observations/day for Orion’s network).

This quantity of data requires the development of business use cases and automated analytics to efficiently interpret and form insights to inform decision-making. For example, smart meter current, voltage and phase angle data being used to identify asset overloads, outages, incorrect connectivity, conductor sizes, customer phasing and DER.

Over the last two years, Orion have been building capability to handle smart meter data and determining how to adapt our business-as-usual processes to make use of the new dataset. This includes establishment of a dedicated Power Systems team, DAVE (Data and Visualisation Engine) platform and data governance framework. These support our in-house use case development and prototyping in parallel to trials of the Gridsight and Future-Grid analytics platforms. By starting this capability development prior to the signing of our agreement with Bluecurrent in October 2023 we were able to rapidly make use of the new data as soon as it started flowing.

3.2.1. Analytic use case development

The shift towards more data-driven decision-making begins with understanding the problem to be solved and where current lack of data is resulting in non-optimal decisions. This is the beginning of use case development.

Once the objective of a use case is determined, stakeholders should be consulted to provide input on the derived data and accuracy levels required for decision-making. From here, data inputs, cleaning strategies and analytic methodologies can be developed and tested until the requirements are met.

Since there are a multitude of use cases which smart meter data unlocks, prioritisation is required to ensure that the focus is given to the use cases which either provide benefit to largest number of stakeholders or provide a foundation for other use cases. Orion has applied descriptors to provide an indication of complexity to implement. These are:

- **Basic** – Requires telemetry and basic asset data (e.g. asset ratings). Little to no reliance on network topology / configuration
- **Integrated** – Requires telemetry data, electrical asset data and network topology / configuration. May require deterministic power flow modelling
- **Advanced** – Requires telemetry data, electrical asset data, network topology / configuration and third-party data (e.g. weather, DER). May require probabilistic power flow modelling, machine learning and predictive analytics

Grouping use cases by complexity provides better visibility of those which have common dependencies and aids in identification of opportunities for modular development.

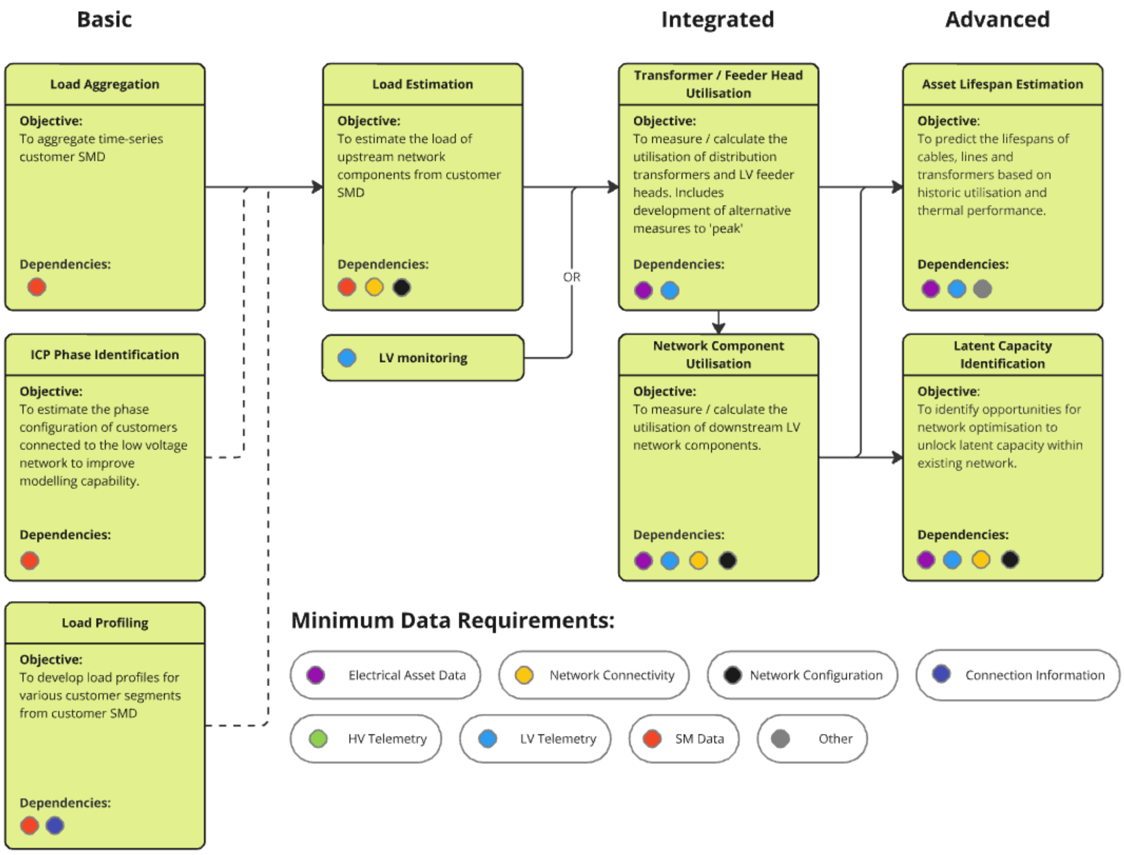


Figure 2: Modular analytic use cases, dependencies and data requirements

‘Off-the-shelf’ solutions such as Gridsight and Future-Grid Compass offer pre-built and tested applications which can be rapidly deployed to begin providing smart meter data insights without the need for in-house development. However, asset owners should not underestimate the internal capability uplift required to interpret the outputs, adjust existing workflows around new information and understand causes of anomalies.

3.2.2. Validation methods

When first using or developing new analytics, an understanding the required accuracy levels for different applications within the business are key. Results from the analytics should be assessed against field data or other high-confidence datasets to ensure that accuracy thresholds can be met for intended use cases. The validation process provides a benchmark of expected level of accuracy and an understanding of the influencing factors so that these can be communicated before being rolled out for wider business use.

Example 1 – Gridsight phase identification

One of the applications provided by the Gridsight platform is automatic detection of relative customer phasing per distribution transformer. Relative phasing can significantly improve the accuracy of transformer and circuit load imbalance estimations. The application performs clustering of customer smart meter voltages to infer which customers are connected to the same phase and assign a grouping of 1, 2 or 3.

To validate this application, the following process was carried out:

1. Non-contact phase identification was carried out in the Milton zone substation supply area to record actual customer phases.
2. Each customer phase was compared to the phase groupings assigned by Gridsight, with the group to phase colour association for each transformer being determined by majority (e.g. if group 2 correlated with 90% of phase survey results relating to Bph, it was assigned Bph, with the remaining 10% being error).
3. Discrepancies between inferred and surveyed phase were reviewed
4. Phase configuration accuracy was then weighted based on the number of ICPs in group/total number of ICPs
5. Weighted results were summed to determine the overall accuracy of the Phase ID application per transformer.

Table 1: Example phase ID accuracy calculation for a distribution transformer supplying 81 ICPs

	Phase configuration							
	R	**Y**	**B**	**RB**	**RY**	**YB**	**RYB**	**Overall accuracy**
ICPs (field survey results)	23	25	30	0	1	1	1	**96.3%**
% total ICPs	28.4%	30.9%	37.0%	0.0%	1.2%	1.2%	1.2%	
ICPs (Gridsight correct)	21	25	29	0	1	1	1	
Phase config accuracy	91.3%	100.0%	96.7%	0.0%	100.0%	100.0%	100.0%	
Weighted result	25.9%	30.9%	35.8%	0.0%	1.2%	1.2%	1.2%	

Validation across 67 transformers (3,139 ICPs) yielded an average overall accuracy of 89.5% (Figure 3) which was marginally below the target accuracy of 90% for this use case. Further investigation is required to determine the causes of lower accuracy as there was no clear correlation to number of ICPs or network type.

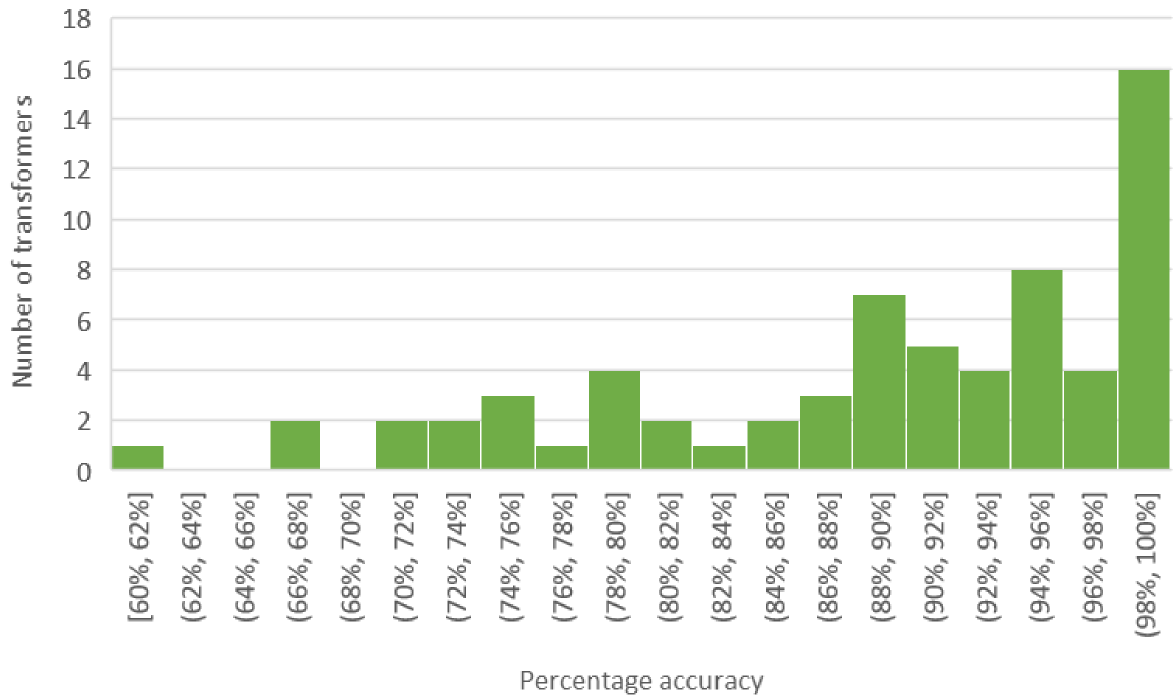


Figure 3: Weighted accuracy of Gridsight Phase ID application (per transformer) compared to Milton field survey results

Example 2 – Load estimation (aggregation w/ linear scaling)
In the absence of direct monitoring, smart meter aggregation can be used to estimate the loading of upstream assets such as distribution transformers, cables and lines. However, the accuracy of this is highly dependent on the available smart meter data downstream of the point of interest. Data gaps can be backfilled using a variety of methods, the simplest of which is a linear scaling based on the percentage of missing meters at each timestep.

$$Estimated\ P_{tot} = \frac{\sum P_{SM}}{\%\ SM\ penetration}$$

One method to validate the accuracy of this estimation method is simply to compare the scaled smart meter power aggregate to total transformer load measured by LV monitoring (Figure 4). In this example, the smart meter aggregation with scaling applied meets the target accuracy for this use case of within ± 10% of the actual measured values.

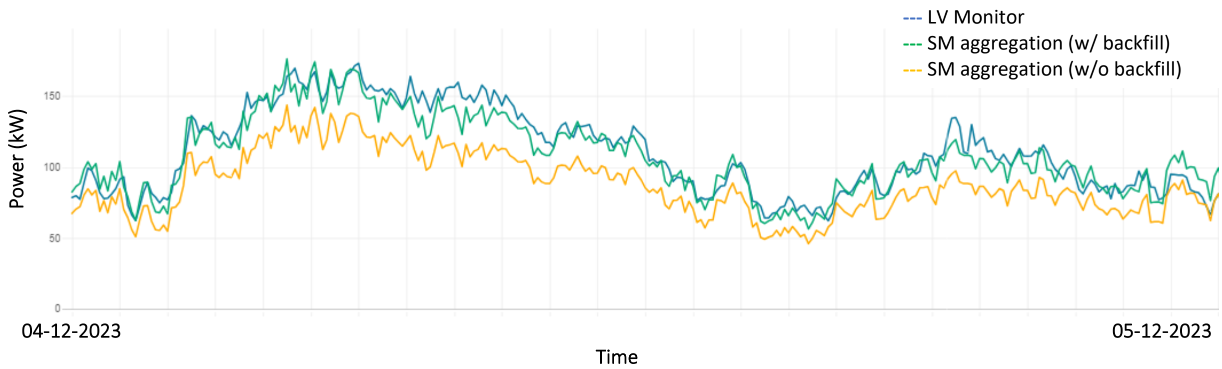


Figure 4: Aggregation of distribution transformer load (130 ICPs with 81% SM penetration) compared to actual load

However, since the purpose of this use case is to estimate loading of non-monitored assets, the accuracy cannot always be directly measured. Instead, the possible variance should be quantified based on an understanding of the factors influencing the backfilled result.

Variance in results for this method is driven by uncertainty in which smart meters are present in the dataset and whether they are representative of those which are missing. The magnitude of this can be assessed by selecting sites with 100% smart meter penetration and taking randomised sub-samples of ICPs to simulate a lower penetration level. From each starting point, the linear scaling method is applied and results compiled to determine the probable range (green) against the actual total (red). Figure 5 and Figure 6 provide an example of the variance in the linear scaling method at smart meter penetrations of 50% and 90%.

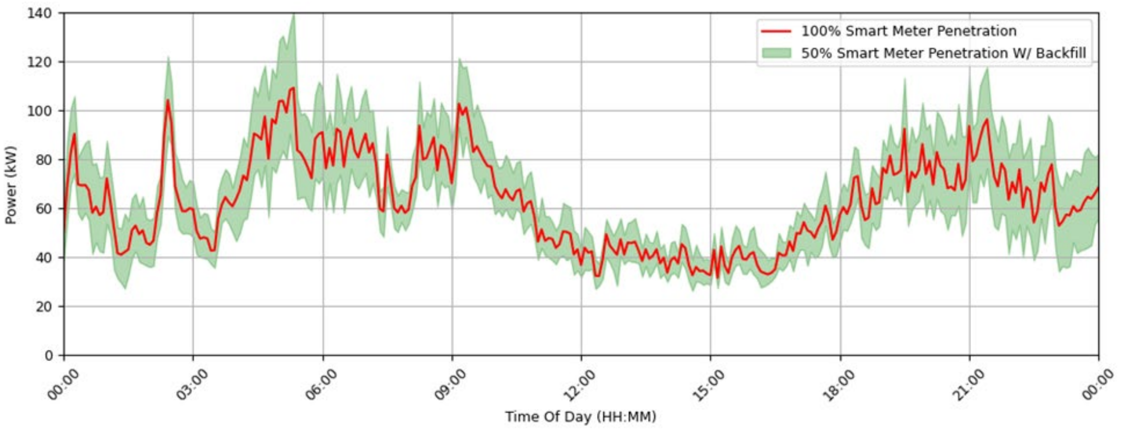


Figure 5: Aggregation of distribution transformer load (80 ICPs with 50% SM penetration) compared to actual load

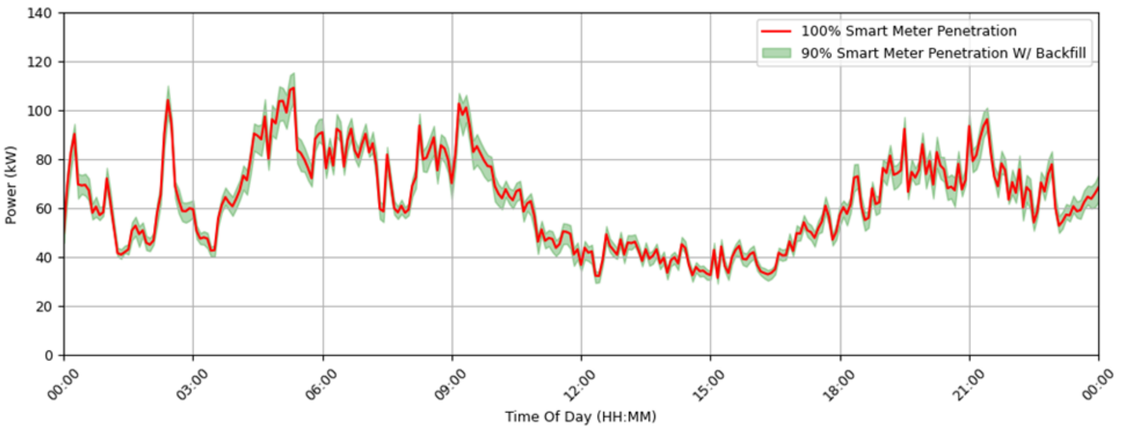


Figure 6: Aggregation of distribution transformer load (80 ICPs with 90% SM penetration) compared to actual load

There are several key influencing factors which impact the variance of load estimation using this method. These are:

- Smart meter penetration
- Customer connection size
- Mix of customer types
- Total number of customers
- Historical network configuration

When input conditions are insufficient to meet the maximum tolerance for variance, aggregated results should be flagged as lower confidence or discarded to prevent corruption of higher-level analytic measures (e.g. transformer utilisation).

3.2.3. Production

Once a use case has been built, iterated and validated against requirements, it is ready for production deployment. At this point, consideration should be made towards how often the data will need to be accessed and the complexity of the computation to determine whether insights should be pre-computed over the entire dataset or generated on user request. These options will have a bearing on the data storage requirements, compute time, user accessibility and, for solutions deployed to a cloud platform, cost.

Documentation and training around new use cases should be included when rolling out to users within the wider business to ensure that there is an understanding of the insights produced, accuracy and any limitations for their use.

3.3. Network model automation

Network analytics can provide an excellent view of current and historic network performance. However, to fully understand the interactions on a network level, particularly in future scenarios, detailed power flow modelling is required.

Historically, power flow models have been constructed manually, often based on a large number of conservative assumptions to account for gaps in data and low update frequencies. The impacts of these compounding assumptions are difficult to quantify and lack of model calibration can lead to high uncertainty in the results. Many networks have now shifted to models derived from Geographic Information Systems (GIS) or Advanced Distribution Management Systems (ADMS) but these often require manual interventions to correct data inconsistencies, particularly at the LV level.

To run distribution networks more efficiently and reduce unnecessary network build, there is a need for more accurate power flow models in which scenarios can be run more regularly to reduce the level of uncertainty. Due to an increasing volume of studies required, scalability and consistency around modelling practices are important considerations which can be addressed through automation.

There are three main aspects of power systems modelling which can be automated:

- **Model creation and calibration** – building an electrical model representing the physical network connectivity and loading
- **Scenario and model management** – creation of options for network alteration / augmentation and management of multiple study cases and scenarios
- **Power system studies** – running of studies based on chosen scenarios and interpretation of results

While Orion has some existing automations around HV and LV model creation and calibration which reduce the overhead of fully manual model maintenance, poor source data quality can still create bugs in the final model which require manual intervention to correct. In addition, power system study processes, methodologies, uncertainty and constraint thresholds are not well documented.

To lay the groundwork for increased end-to-end model automation, Orion is currently working to document and standardise modelling practices within the company. This will include current and future input data requirements from source systems to inform the data model design of system replacements, where applicable.

We are also investigating use of the IEC Common Information Model as a mechanism to improve transfer of network model information between multiple platforms internally and piloting an AI-powered connectivity correction tool to improve GIS data quality and consistency ahead of our GIS platform migration.

4. Challenges & lessons learned

The shift from reactive to proactive network management has posed a number of challenges and has required a cultural shift in our data handling and decision-making practices.

4.1.1. Data Quality, Governance, and Integration

A key lesson learned was the critical importance of data quality, governance, and integration capability. Initially, our LV monitoring programme did not account for the integration of data back into Orion systems, instead relying on vendor-side systems for data management and visualisation. While this approach may suit diagnostic and reactive management, it quickly became impractical for handling a mix of devices and enabling efficient analysis and reporting.

Establishing the DAVE data platform and standardising monitor data across multiple providers enabled the development of automations which help to identify network and data quality issues for further investigation. This platform also facilitated easier integration of telemetry data with other sources, such as customer data and asset ratings, to support the establishment of limit-based analytics.

As networks become more complex, there will be a need to transfer data between more systems and parties than ever before. Therefore, standardisation of network data and transfer mechanisms will be vital to reduce the overheads of custom integrations.

4.1.2. Tools and Training

To effectively handle large datasets, availability of big data processing / visualisation tools and prioritisation of data literacy among potential users is crucial. Training should cover fundamental concepts of data analysis, visualisation techniques, and interpretation skills to ensure that employees can engage with data in an informed manner. Developing the ability to identify key insights amidst large volumes of data is also vital for creating automated analytics tailored to specific use cases.

Where relevant, training on cloud computing is also recommended so that data and engineering teams understand and manage associated costs, ensuring that data processing and storage are both efficient and cost-effective.

4.1.3. Culture and Processes

Integrating new data and analytics into daily operations requires time and effort. Identifying early adopters within teams who are eager to embrace these changes can demonstrate the value of new insights, test workflows and help address initial challenges to foster buy-in from the wider team. This collaborative approach of engaging with stakeholders and subject matter experts to understand their data needs is crucial for reworking processes and incorporating new analytics effectively into business decision-making.

5. Conclusion

In conclusion, the foundational work in the "System Visibility and Network Data Integration" (WS4) and "Power Systems Insight Enablement" (WS3) workstreams of Orion's Network Transformation Roadmap represent an important step towards enabling proactive management of our LV networks. By building capability around how to leverage telemetry, data analytics and network modelling to serve new and existing business needs, we can significantly improve network and operational efficiency which will be essential for managing the complexities of operating our network with increased DERs and electrification.

The cultural shift to enable more data-driven decision making involves placing high importance on data quality, data governance, change management and training. While there are still significant challenges around network data quality, establishment of use cases, methodologies, and accuracy levels can assist decision-makers by providing transparency around confidence levels and limitations. Engaging early adopters and supporting business-wide data literacy will facilitate smoother integration of new insights into business-as-usual to build a foundation for network transformation.