Predicting Power Distribution Network Characteristics and Load Profiles with Limited Operational Data

Junyi Lu

A thesis submitted for the degree of Doctor of Philosophy to Department of Electronic and Electrical Engineering

University of Strathclyde

September 2025

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

Abstract

Electricity distribution systems were originally designed for unidirectional power flows and not for on-site generation. However, with the increasing integration of Distributed Energy Resources (DERs) and other Low-Carbon Technologies, the role of distribution networks is evolving. This shift introduces several challenges, including the need for more frequent network reconfiguration to identify potential voltage and thermal violations. Addressing these issues often relies on state estimation techniques, which depend on large and accurate datasets. Despite the importance of accurate monitoring, the vast number of secondary substations makes it economically impractical to install low-voltage monitoring equipment and store data over extended periods of monitoring data. Consequently, this thesis proposes three methods designed to maximise the utility of minimal data: (i) transfer learning, which utilises data from other locations; (ii) few-shot learning, which predicts outcomes with limited data from the same location. The thesis then introduces a comprehensive method for constructing a network model that leverages actual operational GIS data, ensuring the resulting representation of the network closely reflects real-world conditions. This model is subsequently utilised to perform forecasting analyses, enabling the accurate prediction of future operational scenarios. The approach is designed to align with the Distribution Future Energy Scenarios (DFES) framework used in Great Britain (GB), thereby ensuring that the forecasting results are both relevant and practical for contemporary network planning and operation. Grounded in real operational data, this thesis designs and implements a novel GIS-based model of a real-world distribution network and integrates forecasting methods based on transfer learning and few-shot learning to achieve forecast in data-sparse areas. The aim is to provide robust decision support for the GB electricity distribution sector. As a result, DNOs can achieve more cost-effective yet reliable performance across the network, reducing the need for comprehensive data sets or detailed forecast models at every substation. The significance of this research lies in its ability to extract and transfer salient information from both limited and data-rich sources. The knowledge can be applied in situ or moved to data-sparse substations, enabling cost-effective, feasible and accurate power-flow analysis and state estimation across the distribution network. When combined with a GIS-based model built from real operational data, the approach yields results that are directly actionable for DNOs.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my primary supervisor, Dr Bruce Stephen, for granting me the invaluable opportunity to undertake this research and for guiding me through such a formative and memorable journey. His constant support, insightful advice, and inspiration have been truly invaluable, not only for the completion of this PhD but also for shaping my future career.

I am equally indebted to my second supervisor, Dr Blair Brown, for his unique perspective on shaping this work and for encouraging me to approach research with clarity and accessibility. His emphasis on knowledge exchange and making complex ideas understandable to a wider audience has been especially influential in my development as a researcher.

My sincere thanks also go to my colleagues in the Advanced Electrical Systems Research Group. I am especially grateful to Amy and Jennifer for helping to lay the foundations of my PhD during the challenging early stages of the Covid pandemic. I also want to thank Alistair, Riu, Marina, Guo, and Thomas for their collaboration and encouragement throughout this journey.

Finally, and most importantly, I owe my deepest thanks to my girlfriend and my parents. Their unwavering encouragement, personal sacrifice, and financial support have made it possible for me to pursue my studies to this stage. Without their constant support and belief in me, none of this would have been possible. This thesis is dedicated to them.

Contents

List of Figures

List of Tables

List of Abbreviations

ANN	Artificial Neural Networks
ARIMA	Autoregressive Integrated Moving Average
BST	British Summer Time
CDF	Cumulative Distribution Function
CP	Cardinal Point
DER	Distributed Energy Resource
DFES	Distribution Future Energy Scenarios
DNO	Distribution Network Operator
DSSE	Distribution System State Estimation
DSE	Dynamic State Estimation
EPRI	Electric Power Research Institute
ESO	Electricity System Operator
EV	Electric Vehicle
FASE	Forecasting-Aided State Estimation
GAM	Generalised Additive Model
GB	Great Britain
GIS	Geographic Information System
GMT	Greenwich Mean Time
GSP	Grid Supply Point
HV	High Voltage
LCT	Low Carbon Technology
LV	Low Voltage
LSTM	Long Short-term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MBE	Mean Bias Error
NESO	National Energy System Operator
Opendss	Open Distribution System Simulator
PV	Photovoltaic

RMSE	Root Mean Square Error
RMU	Ring Main Unit
RNN	Recurrent Neural Network
SOP	System Operating Plan
SPEN	ScottishPower Energy Networks
SSEN	Scottish and Southern Electricity Networks
WLS	Weighted Least Squares
WLAV	Weighted Least Absolute Value
WPD	Western Power Distribution
UKGDS	UK Generic Distribution System

List of Symbols

α	Coefficient
β	Coefficient
C_θ	Copula function
d	Day
e	Residual
$E(\cdot)$	Expectation (Average)
h(x)	Measurement function vector
H	Jacobian matrix
i	Index
I	Current
L	Load
Λ	Precision matrix
μ	Mean
M	Month
N	Normal Distribution
P	Active Power
Q	Reactive Power
r	Coefficient
R	Set of real numbers
RA	Rolling Average
σ	Standard deviation
σ^2	Variance
ρ	Correlation coefficient
Σ	Covariance matrix
t	Time step
T	Temperature
θ	Phase angle
Φ	Standard normal CDF

Φ^{-1}	Inverse standard normal CDF
u	Uniform variable
V	Voltage magnitude
x	Independent variable
y	Dependent variable
$\hat{y}$	Forecasted dependent variable
Y	Dependent variable
z	Standard normal variable
Z	Measurement vector

1 Introduction

1.1 Overview

With the growing share of EVs and the increasing penetration of DERs on distribution networks, new operational challenges have emerged. One key issue is the occurrence of sudden voltage rises, particularly at the far end of distribution feeders with high PV generation, as distribution networks exhibit heightened sensitivity to active power injection [1]. Conversely, substantial voltage drops may arise due to the simultaneous charging of EVs during evening peak periods, potentially driving the transformer towards its operational constraints [2]. Furthermore, the reversal of power flow can lead to significant voltage increases, which may exceed permissible operational limits [3]. Collectively, these factors contribute to voltage fluctuations, which, in turn, induce rapid variations in renewable energy output from PV and wind generation, thereby compromising power quality [4].

All these factors contribute to the network becoming progressively unstable and more prone to faults or loss of supply. Meanwhile, DNOs are required to supply all LV customers within the voltage range defined in EN50160, which is +10% and −6% in GB [5]. To monitor stability, DNOs typically rely on periodic measurements, with voltage regulation carried out solely by tap-changing transformers at the primary substation, to let most LV network voltages remain within statutory limits [6]. However, distributed generation has already caused half of the secondary substation readings in the Western Power Distribution to exceed 6-8% the system nominal voltage [7].

With the growing penetration of LCTs, power quality risks are increasing as load profiles become more variable, which may further impact voltage stability. Therefore, predicting the overall performance of load profiles in the LV network and their impact on voltage stability is essential for future networks. However, while load forecasting is widely studied in the literature, most research focuses on the national level, primarily addressing generation-demand balancing. At the distribution level, relevant studies are scarce [8], and prediction performance is significantly impacted by the following constraints:

1. Lack of aggregation effect: In LV networks at the secondary substation level, a typical substation area serves approximately 50 to 300 customers. On a national scale, load profiles exhibit a strong correlation with weather conditions and calendar-based variations. However,

at the LV level, these influences (e.g., weather or calendar effects [9]) are significantly reduced. Instead, electricity usage in LV networks is more affected by local events, such as the "TV pickup" phenomenon in GB [10], it occurs when large numbers of people watch the same television programme simultaneously. For example, during a popular show or Sports fixtures, the electricity demand in sub-regional demand will rise at half-time. In contrast, other regions, such as Edinburgh, are typically less affected These events are often unmonitored, leading to significant fluctuations in load profiles and causing the statistical distribution of distribution network loads to deviate from commonly accepted probability distribution functions [11].

2. Lack of data and the influence of bad data: Loads in distribution networks have long been assumed to be balanced and stable. However, due to economic constraints, investment in distribution network monitoring has been neglected for an extended period. Even when monitoring data are available, the presence of bad data is a significant issue. Gross errors can affect bad data detection, while load models are influenced by these errors, leading to inaccurate predictions [12].

These challenges reveal three connected research gaps. First, LV load forecasting remains under-studied compared with national-scale forecasting [9], and the literature pays even less attention to the operational reality of LV networks where measurements are sparse, historical records are limited, and data quality is inconsistent. Consequently, many existing methods are difficult to deploy at scale across data-poor areas [13].

Second, while GIS-based LV network modelling has progressed [14], operational GIS datasets often contain connectivity omissions, missing nodes at genuine connection points, and other inconsistencies that limit direct use for electrical studies. Although several studies have proposed reconstruction or correction approaches, far less attention has been given to systematic validation of the resulting network models [15], and this gap is particularly evident for the GB context. This matters because distribution network designs, data practices, and operational characteristics can vary substantially across regions, meaning that methods demonstrated elsewhere may not transfer directly to GB LV networks without careful assessment.

Third, existing work rarely links reconstructed GIS-based models with forecast demand in a way that enables electrical interpretation of predicted loads. As a result, the literature lacks an integrated

pathway that maps data-driven LV load forecasts onto an electrically credible GIS-derived network model to generate actionable insights for DNO and DSO planning and operation.

This thesis presents the development of a method to address challenges associated with LV distribution network data and to model the network for monitoring overall performance. The proposed approach leverages the principles of transfer learning and few-shot learning. Each of these approaches being applied to forecasting tasks that are used to enhance the accuracy of load capacity or resolution predictions under various data limitations: lack of data from different areas (transfer learning), lack of historical data (few-shot learning). Additionally, a novel cardinal point forecast calibration model is introduced to mitigate the aggregation effect problem. Furthermore, a network model is constructed using actual operational GIS data, ensuring that the resulting representation of the network closely reflects real-world conditions. Based on the predicted load data, the model is employed to help DNOs identify extreme voltage and current values, enabling day-ahead operational planning, monitoring of peak fluctuations, and the implementation of active control measures.

This chapter introduces the contents of the thesis, outlines relevant analysis techniques, and details the methodology and publications associated with this research.

1.2 Contributions to Knowledge

Research question:

How can DNOs/DSOs infer large-scale LV network behaviour with minimal monitoring by transferring knowledge from a fully observed area to another area with limited or no observations to provide useful network information for planning and operation? (Chapter 3) How can be sparse within-area measurements be used to support data-driven point forecasts that are operationally meaningful and help assess overall network performance under limited observability? (Chapter 4) How can an operational network model be built by combining GIS-based topology data with engineering knowledge to map predicted loads onto network characteristics, and can this approach be automated and scaled beyond test feeders such as the IEEE 123-bus system to cover all DNO licence areas using only limited existing data? (Chapter 5)

The following list specifies the unique contributions to knowledge occurring from this thesis:

- An approach is proposed to investigate how active aggregation of smart meter data impacts the total secondary substation load profile. Additionally, the relationship between weather variables and calendar variables with the secondary substation load profile is examined, highlighting the challenges in accurate load prediction.
- A structured review of existing approaches to modelling distribution networks, with particular attention to their specific requirements. Building on this, an analytical review is conducted, applying power flow and state estimation methods to transfer load forecast results into meaningful insights about network characteristics, highlighting the challenges and the need for distribution system state estimation in the future distribution system.
- A novel approach is proposed for using transfer learning to enhance the Bayesian method in forecasting the residual error of a transfer model across different areas of the distribution network. This approach integrates dynamic state estimation techniques to improve observability in distribution network model. By leveraging knowledge from one area and applying it to another, this method enhances prediction reliability, even in regions with limited historical data.
- A novel approach for defining cardinal point forecasting in the distribution network is introduced. This method identifies three key tuning points in the distribution network, which are valuable for DNOs in day-ahead planning. Additionally, it offers a more feasible alternative to time-series forecasting
- A global model for the LV network is introduced, based on customer numbers. This model enables zero-shot learning for cardinal point time forecasting and few-shot learning for cardinal point scale forecasting in the LV network, even with limited or few observations. This approach is beneficial for DNOs in day-ahead planning and decision-making.
- A topology reconstruction method for reconstructing LV network models from operational GIS data is presented. In alignment with the DFES framework, this electrically coherent model supports accurate, low-data forecasting for GB LV networks and enables practical applications of the few-shot and transfer learning approaches developed in this thesis to assist DNO planning and operational decisions

Collectively, these contributions advance the field by integrating data-driven methods and engineering knowledge into a unified framework that enhances observability, forecast accuracy, and decision-making for distribution network operators, particularly in low-observability environments.

Answer to the research question:

Together, these methods provide a practical, low-cost workflow that lets DNOs quantify aggregation effects, transfer load forecast to network characteristics, and make planning and operational decisions using existing available data.

1.3 Thesis Structure

This chapter provides a brief introduction to the thesis, highlighting the contributions to knowledge and associated publications.

Chapter 2 provides background information on the research, discussing the current state and future challenges of distribution networks, particularly in terms of observability in the GB distribution network. It examines the equipment currently used to monitor the LV network and highlights the necessity of developing a day-ahead forecast model for the distribution network with limited data. This chapter also introduces the concepts of power flow and state estimation, which are crucial for monitoring overall network performance rather than focusing solely on individual substations. Furthermore, it presents dynamic state estimation, which could be essential for the future development of distribution networks.

Chapter 3 describes novel transfer learning methods used to transfer model fitting from data-rich areas to predict data in data-sparse areas. It introduces an updating Bayesian method for predicting residuals based on observed data in these sparse areas. Additionally, the chapter discusses the application of Forecasting-Aided State Estimation in the predicted data to gain insights into all voltage magnitudes and phase angles within the network.

Chapter 4 introduces a novel few-shot learning approach at the distribution level, defines the cardinal point in the distribution network, and presents the concept of a global model for distribution-level load forecasting based on customer numbers. This approach enables the forecasting of a substation with fewer observations by leveraging patterns from other substations.

Furthermore, a model calibration process is introduced after the global model, using a Gaussian Copula to further enhance forecast accuracy.

Chapter 5 examines how the thesis's methodologies, including few-shot learning and transfer learning, can be applied in real-world network operations. It begins with an overview of the Distribution Future Energy Scenarios (DFES) as a strategic framework. The chapter then demonstrates how data-driven approaches, supported by topology reconstruction GIS network reconstruction, address emerging challenges and enable the integration of low-carbon technologies through case studies. Overall, it highlights the practical value of the research presented in this thesis and provides a clear pathway for translating these academic advances into tangible improvements within the GB electricity distribution network.

Chapter 6 summarises the implementation and results of the thesis. It also discusses future work, emphasising the need for more flexible and integrated approaches to prediction and modelling in the distribution network.

1.4 Publications

Journal Papers:

- J. Lu, B. Stephen, and B. Brown, "Intraday residual transfer learning in minimally observed power distribution networks dynamic state estimation," Data-Centric Engineering, vol. 5, no. 1, p. e13, 2024. [Online]. Available: https://doi.org/10.1017/dce.2024.10

Journal Papers Under Review:

- J. Lu, B. Brown, and B. Stephen, "Cardinal Point Load Forecasting in LV Networks Using Few-Shot Residual Calibration," *IEEE Transactions on Industrial Informatics*, under review.
- A. Brash, J. Lu, B. Stephen, B. Brown, R. Atkinson, C. Michie, F. MacIntyre, and C. Tachtatzis, "Coherent Load Profile Synthesis with Conditional Diffusion for LV Distribution Network Scenario Generation," *Applied Energy*, under review.

- J. Lu, G. Shi, T. Liu, R. R. Sakamoto, B. Brown, A. Riccardi, G. Wilson, and B. Stephen, “Predictive Modelling of SF_6 Escape Rates for Optimal Maintenance Scheduling of HV Switchgear,” *IEEE Transactions on Power Delivery*, under review.

Related Conference Papers:

- J. Lu, J. Fallman, B. Brown, B. Stephen, S. D. J. McArthur and P. Papadopoulos, "Surrogate Models for Investigating Dynamic Security Regions of Renewables-Dominated Grids," *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, Dubrovnik, Croatia, 2024, pp. 1-5, doi: 10.1109/ISGTEUROPE62998.2024.10863663.

Conference Papers Under Review:

- J. Lu, T. Storey, M. Cassidy, C. McGarry, R. Rodriguez Sakamoto, Q. Hong, C. Booth, M. Elsheikh, C. Higgins, A. Meek, S. McArthur, B. Brown, A. Campbell, and B. Stephen, “Automated calibration of LV distribution network digital twins for future scenario generation,” IEEE Power & Energy Society General Meeting (PESGM), 2026, under review.

1.5 References

[1] H. Sun *et al.*, "Review of Challenges and Research Opportunities for Voltage Control in Smart Grids," *IEEE Transactions on Power Systems,* vol. 34, no. 4, pp. 2790-2801, 2019, doi: 10.1109/TPWRS.2019.2897948.

[2] K. Clement-Nyns, E. Haesen, and J. Driesen, "The Impact of Charging Plug-In Hybrid Electric Vehicles on a Residential Distribution Grid," *IEEE Transactions on Power Systems,* vol. 25, no. 1, pp. 371-380, 2010, doi: 10.1109/TPWRS.2009.2036481.

[3] M. Hasheminamin, V. G. Agelidis, V. Salehi, R. Teodorescu, and B. Hredzak, "Index-Based Assessment of Voltage Rise and Reverse Power Flow Phenomena in a Distribution Feeder Under High PV Penetration," *IEEE Journal of Photovoltaics,* vol. 5, no. 4, pp. 1158-1168, 2015, doi: 10.1109/JPHOTOV.2015.2417753.

[4] N. B. G. Brinkel *et al.*, "Impact of rapid PV fluctuations on power quality in the low-voltage grid and mitigation strategies using electric vehicles," *International Journal of*

Electrical Power & Energy Systems, vol. 118, p. 105741, 2020/06/01/ 2020, doi: https://doi.org/10.1016/j.ijepes.2019.105741.

[5] D. J. Start, "A review of the new CENELEC standard EN 50160," 1995: IET, pp. 4-1.

[6] C. Zhao, C. Gu, F. Li, and M. Dale, "Understanding LV network voltage distribution-UK smart grid demonstration experience," in *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 18-20 Feb. 2015 2015, pp. 1-5, doi: 10.1109/ISGT.2015.7131851.

[7] T. Frost, P. D. Mitcheson, and T. C. Green, "Power electronic voltage regulation in LV distribution networks," in *2015 IEEE 6th International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, 22-25 June 2015 2015, pp. 1-7, doi: 10.1109/PEDG.2015.7223039.

[8] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," *IEEE Transactions on Smart Grid,* vol. 10, no. 3, pp. 3125-3148, 2019, doi: 10.1109/TSG.2018.2818167.

[9] S. Haben, G. Giasemidis, F. Ziel, and S. Arora, "Short term load forecasting and the effect of temperature at the low voltage level," *International Journal of Forecasting,* vol. 35, no. 4, pp. 1469-1484, 2019/10/01/ 2019, doi: https://doi.org/10.1016/j.ijforecast.2018.10.007.

[10] "EURO 2020 and the TV 'pick-up' effect," in *National Energy System Operator (NESO)*, ed, 2021.

[11] R. Singh, B. C. Pal, and R. A. Jabr, "Statistical Representation of Distribution System Loads Using Gaussian Mixture Model," *IEEE Transactions on Power Systems,* vol. 25, no. 1, pp. 29-37, 2010, doi: 10.1109/TPWRS.2009.2030271.

[12] S. H. Braunstein, N. G. Bretas, A. Rossoni, and A. Bretas, "Bad data analysis in distribution state estimation considering load models," in *2015 IEEE Power & Energy Society General Meeting*, 26-30 July 2015 2015, pp. 1-5, doi: 10.1109/PESGM.2015.7286541.

[13] C. J. Bennett, R. A. Stewart, and J. W. Lu, "Forecasting low voltage distribution network demand profiles using a pattern recognition based expert system," *Energy,* vol. 67, pp. 200-212, 2014.

[14] C. McGarry, A. Anderson, I. Elders, and S. Galloway, "A Scalable Geospatial Data-Driven Localization Approach for Modeling of Low Voltage Distribution Networks and Low Carbon Technology Impact Assessment," *IEEE Access,* vol. 11, pp. 64567-64585, 2023, doi: 10.1109/ACCESS.2023.3288811.

[15] M. Deakin, M. Vanin, Z. Fan, and D. Van Hertem, "Smart energy network digital twins: Findings from a UK-based demonstrator project," *International Journal Of Electrical Power & Energy Systems,* vol. 162, p. 110302, 2024.

2 Distribution Network Characteristics and Challenges

This chapter establishes the foundations for the forecasting and network-level methods developed later in the thesis. Section 2.1 introduces the background and key challenges of GB distribution networks, with emphasis on limited LV visibility and the impacts of LCTs. Section 2.2 reviews monitoring devices and explains how measurement coverage shapes LV data availability. Section 2.3 analyses LV load characteristics and identifies informative forecasting parameters, including lagged load, temperature-related features and calendar variables, which inform the methods in Chapters 3 and 4. Sections 2.4 and 2.5 then introduce distribution network modelling, power flow analysis, and state estimation approaches to translate uncertain load information into operationally meaningful network insights.

2.1 Distribution Network Background and Challenges in GB

The majority of power distribution networks in Great Britain (GB) were planned, designed, and constructed in the 1950s and 1960s [1]. At the time, assets were installed with sufficient spare capacity to accommodate anticipated growth in electricity demand. However, DNOs currently have very limited visibility of, and knowledge about, networks, particularly regarding real-time utilisation. Instead, they typically rely on static network information, such as the number and type of customers and their general electricity usage patterns to estimate annual or daily peak demand at an LV substation by aggregating typical customer load profiles [2].

As a result, the distribution system has remained largely unmonitored, with customer interruptions caused by faults or network outages usually addressed reactively by DNOs. Figure 2.1 illustrates a typical GB distribution network.

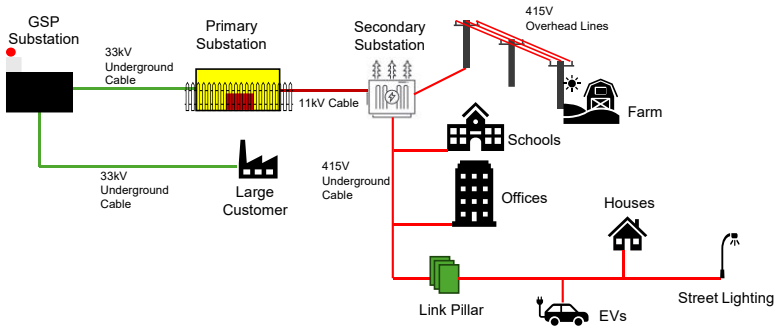

Figure 2.1 Hierarchical electrical distribution network in GB: From grid supply to end-users

The distribution network beginning at the GSP substation, which marks the boundary between the transmission and distribution networks [3]. The voltage is then stepped down from the Bulk Supply Point at 132 kV to 33 kV, from a primary substation at 33 kV to 11 kV, and finally from a secondary substation at 11 kV to 400 V for end users. In here, the LV network refers to the lowest level of the distribution network in GB, from 11 kV down to the end-customer level [4].

DNOs in GB currently have very limited real-time visibility of their LV networks. At present, they rely on static data such as customer numbers, load types, and historical electricity consumption to construct typical load profiles for each substation [2]. These profiles are then used to estimate annual or daily peak demands. However, since the forecasts are based on fixed assumptions, they have been shown to exhibit significant error margins when compared to actual measured peak values [5].

For a LV substation, the key safety metric is the utilisation factor *U*, which indicates whether the transformer is operating beyond its nameplate rating:

$$Utilisation\ Factor = \frac{Maximum\ kVA\ Demand}{Transformer\ kVA\ Rating} \tag{2.1}$$

Where the numerator represents the instantaneous peak apparent power drawn to the transformer, while the denominator denotes the transformer's rated capacity. Values of $U > 1$ indicate operation beyond the design limits, thereby increasing the risk of accelerated insulation ageing or potential failure.

DNOs typically obtain the non-coincident maximum demand of each premise either from advanced metering infrastructure or inferred from monthly energy consumption. To estimate the simultaneous peak demand at the feeder level, these individual maximum values are combined using the $diversity\ factor$:

$$Diversity\ Factor = \frac{Sum\ of\ individual\ peak\ demands}{Max\ simultaneous\ demand} \tag{2.2}$$

The diversity factor varies according to the number of connected customers. Traditionally, as the number of customers increases, the diversity factor also increases, but the rate of increase becomes slower. Power companies can use transformer load allocation management software to estimate the maximum aggregate energy consumption supplied by each transformer, thereby assessing how loads are distributed across the network. This is feasible because detailed monthly consumption

records are typically available for each customer. In theory, transformer-allocation software could apply equation (2.2) using these detailed consumption histories. In practice, however, few DNOs maintain comprehensive or up to date diversity factor tables, making this process both costly and prone to misjudgements.

Moreover, the increasing penetration of LCTs is undermining the empirical assumptions that underpin equation (2.2). Research in California [6] found that typical EV charging at residential locations occurs between 4–7 p.m., coinciding with peak demand. Uncontrolled EV charging could raise peak demand in residential feeders by an average of 64%, with 60% of feeders requiring reinforcement when each household has one EV. Similarly, studies in GB indicate that high EV uptake could cause voltage to drop below limits during winter peaks, affecting both LV feeders and substation transformers [7]. On the other hand, due to the low X/R ratio in distribution networks, power injection from distributed generation can cause significant voltage rise, making statutory limits easier to breach [8]. Traditional voltage regulation via upstream tap changers is often ineffective. While distribution network components (cables, lines, transformers) inherently support bidirectional flow within voltage and thermal limits, they were historically designed for unidirectional operation. Reverse power flow can also disrupt protective relays, which are configured to detect only one-way power flow [9]. Since peak PV generation occurs at noon while EVs are charged at night, this may lead to overvoltage issues at noon and undervoltage issues at night, potentially causing further damage to devices.

Therefore, the upgrades to substations and cables are inevitable. However, they will be very costly. The replacement of 11 kV underground cables will involve both long and short overlays. Long overlays span several hundred metres to a kilometre in a single installation process, while short overlays cover approximately 50 metres. Based on data from the Office of Gas and Electricity Markets [10] and Western Power Distribution [11], the cost of replacing 11 kV cables is estimated to be around £30k–80k per kilometre. Consequently, it is impractical to upgrade the cables all at once.

To mitigate risk in the meantime, it is essential for DNOs to carry out close monitoring of the LV network and conduct thorough analysis and modelling of distribution network for day ahead planning, even when only limited dataset is available. This enables the prevention of potential outages or asset damage caused by DERs or EVs because of overvoltage or undervoltage events.

2.2 Monitoring Device in the Distribution Network

Monitoring devices in power networks play a pivotal role in ensuring reliability and security. In transmission networks, these technologies are well-developed and extensively tested. Devices such as Phasor Measurement Units (PMUs) and Supervisory Control and Data Acquisition (SCADA) systems not only measure voltage and current waveforms at high sampling rates but also provide real-time data on phase angles and magnitudes [12]. Additionally, they enable remote control and adjustment of network elements, enhancing operational efficiency [13]. These advanced monitoring techniques are highly mature and support protection, fault recording, and event logging. High-frequency data acquisition allows these devices to deliver data multiple times per second, capturing rapid system dynamics and improving situational awareness in power networks [14].

However, the costs associated with installing PMUs and SCADA systems are considerable. Traditional PMUs typically range in price from £50,000 to £300,000 per unit [15], while large-scale SCADA systems can incur costs of several million [16]. Given that each DNO has typically numbers of significant 10000's secondary substations within license areas [17], it is impractical to deploy expensive, high-accuracy equipment in every substation. Furthermore, because the need to monitor the distribution network used to be relatively low, monitoring infrastructure is often inadequate. Most substations comprising transformers stepping down from 11 kV to 400 V, along with associated cabling and switchgear are equipped only with basic maximum demand indicators, which do not provide data on the timing, duration, or frequency of peak demand.

Recently, projects based on LV monitoring have gained increasing popularity, primarily driven by the growing penetration of DERs. DNOs such as SSEN, SPEN, and NGED have initiated trials of cost-effective LV monitoring devices in various regions of GB, including Oxford, St Andrews, and so on. For instance, SSEN is collaborating with two key equipment providers, Eneida and Lucy Electric GridKey [18]. These LV monitoring devices offer several advantages, particularly their compact size and intelligent design, owing to recent technological advancements. They can be installed and commissioned live on the network within minutes, and are remotely configurable, supporting firmware updates over the air. Additionally, they offer high measurement resolution and are capable of capturing key electrical parameters at the distribution substation level, broken down by feeder and phase. The devices also feature waveform capture capabilities, which support pre-fault detection and enable precise analysis of power quality across the grid [19]. The cost of

these monitoring devices is relatively low, with each unit priced at only a few hundred pounds. This represents a cost reduction of approximately 70% compared to the equipment used in earlier innovation projects.

Figure 2.2 Eneida LV monitor installed in LV feeder pillar [18]

Figure 2.2 illustrates the installation of the Eneida EdgeSense monitor at the top interior panel of a low-voltage distribution cabinet. The device is mounted above the rows of circuit breakers, with several black cables extending downward to connect with the monitored circuits. This positioning minimises interference with manual operations within the cabinet and allows efficient access to feeder measurements, such as voltage and current. And it shows that the device is intended for ease of installation within existing infrastructure.

Although LV monitors have been rapidly developed with the intention of reducing costs, there remains a significant gap in the reliability of the data they provide. One primary issue is the high frequency of data collection, which generates large volumes of raw data and creates challenges in storage, transmission, and processing. In some cases, load values remain constant for extended periods, sometimes lasting over an hour as many DNOs apply a deadband, ensuring that non-varying analogue generate fewer data points and therefore require less storage [20].

This may indicate communication delays or data buffering. If the device is unable to transmit new data, it might resend cached readings. Such behaviour can seriously affect the analysis of load

patterns, as it becomes difficult to distinguish between hardware errors and actual load characteristics.

The overall effectiveness of LV monitoring is still constrained by the relatively short asset life of LV monitors, as many of their embedded features, such as communication modules and batteries tend to degradation or obsolescence [21]. In addition, the monitoring data often lacks contextual information, such as network topology or customer usage patterns, which limits the scope of analysis unless supported by external data sources. Therefore, low observability remains a key challenge in the analysis of LV networks. Approaches that enable better use of existing observed data or learning from limited data sources continue to offer a promising pathway for understanding the overall characteristics of LV networks.

2.3 Challenges and Analysis of Load Characteristics in Distribution Networks

Most analyses of power system load data are based on two forecasting methods. The primary approach involves half-hourly time-series forecasting. A 30-minute interval keep a balance between providing sufficient resolution for capturing daily load patterns and maintaining manageable data volumes, also reducing the impact of noise [22]. The alternative method focuses on point forecasts, such as peak demand predictions, which are essential for ensuring that generation and transmission capacities are adequate to meet demand without risking blackouts or system failures. This section uses a one-year half-hourly load dataset from six substations in a rural region in the Northeast GB. The data therefore comprise six time series with 17,520 observations each substation (365 days × 48 half-hourly intervals). To validate the data for statistical analysis, an Augmented Dickey-Fuller (ADF) hypothesis test was performed on the load profiles. The test returned an ADF statistic of -4.257, which is lower than the 1% critical value of -3.431, and a p-value of 0.0005. These results indicate that the null hypothesis of a unit root can be rejected at a high level of significance, confirming that the series is stationary. This dataset is used for the analyses presented in the following subsection.

2.3.1 Challenges in Aggregating Smart Meter Data to LV Substation Load Data

Most load forecasts were traditionally conducted at the HV or national level, typically encompassing hundreds of thousands to millions of consumers. At this scale, demand tends to be more stable and regular, making it relatively easier to predict. In contrast, significantly less

research has been focused on forecasting at lower voltage levels, where demand is more volatile. The existing research in this area primarily concentrates on short-term load forecasting using smart meter data [23]. The prediction of LV network is usually based on the aggregation of multiple smart meters. Aggregations of smart meter data does not perfectly characterise LV load profiles, as it overlooks several important characteristics of real-world LV distribution networks. Firstly, it neglects cable losses within the network, which can significantly affect the actual power delivered to end-users. Secondly, it fails to account for customer consumption patterns. In LV networks, coincidental usage, such as multiple households operating appliances simultaneously during peak periods can result in sudden and sharp load fluctuations that aggregated data may not accurately capture.

This challenge is particularly notable in GB, where DNOs operate under strict regulatory frameworks regarding data privacy. DNOs are often constrained from utilising individual-level data, relying instead on aggregated smart meter feeds at the LV feeder level. This aggregation introduces temporal inconsistency; as the number of reporting meters fluctuates due to privacy thresholds or connectivity issues, it becomes difficult to verify consistency in the customer base.

For example, aggregated smart meter data from the National Grid [24] illustrates. The dataset comprises recordings from September and October, a period recognised as "shoulder months" [25]. illustrates this limitation.

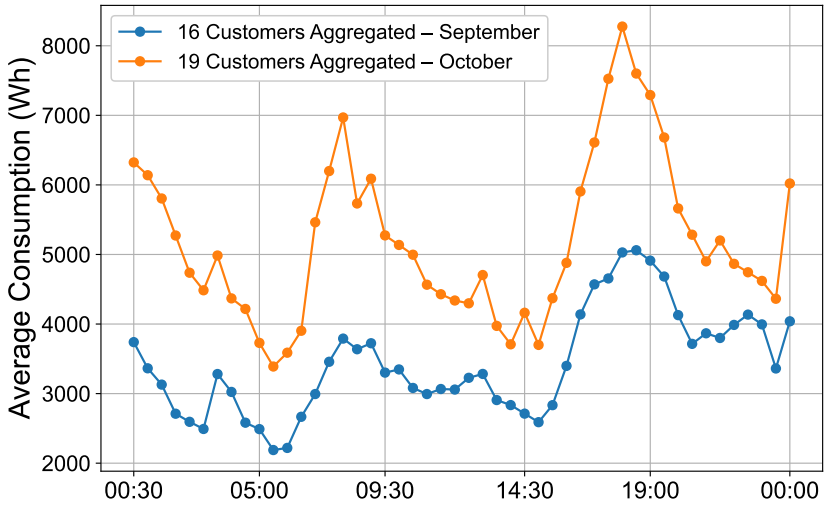

Figure 2.3 Half-hourly load profile based on aggregated smart meter data: Substation A

As shown in Figure 2.3, even a moderate rise in the number of aggregated customers from 16 to 19 (a 25% increase) results in disproportionate load shifts. Despite the small change in customer count, the peak load at approximately 18:00 surges from 5000 Wh to 8200 Wh with a 64% increase. Similarly, the trough load rises by approximately 30%.

This disproportionate increase indicates that the change in load is not merely a linear function of adding more meters. Instead, it reflects the combined impact of three compounding variables: 1) the increase in the number of aggregated units, 2) the temporal shift towards winter heating patterns in October, and 3) the likelihood that the specific composition of reporting customers has changed (i.e., the larger group may not be a direct superset of the smaller group). Consequently, this instability precludes the use of fluctuating aggregated data as a reliable control for long-term studies.

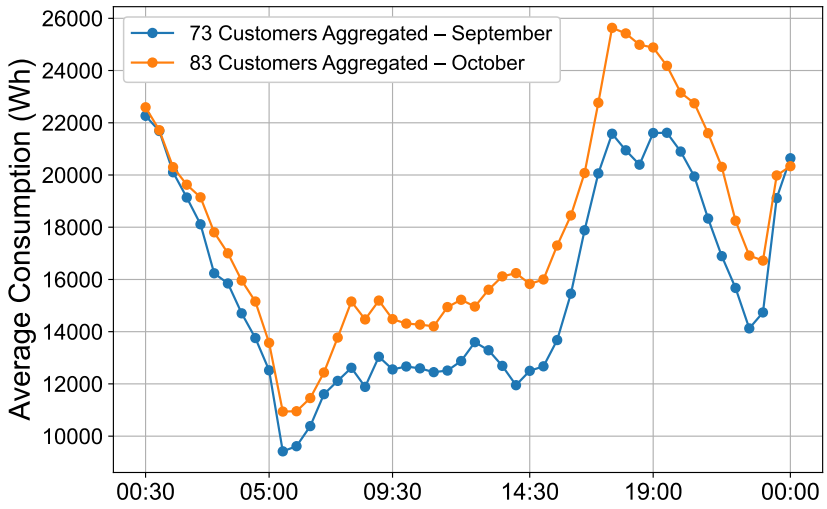

Figure 2.4 Half-hourly load profile based on aggregated smart meter data: Substation B

Figure 2.4 presents results from a second substation within the National Grid dataset, demonstrating similar volatility. Here, the load profile again shifts significantly more than the change in customer count would predict. This reinforces the observation that aggregated LV data is sensitive to coincidental usage and customer heterogeneity. Because the specific customers contributing to the aggregate can change alongside the total count and the time of year, it is impossible to isolate the root cause of the load variance. Consequently, relying on such aggregated feeds introduces significant uncertainty when attempting to characterise the true total load of the substation.

At the distribution level, identifying the factors that impact the load is particularly important, as it reveals trends in how the load changes and assists in selecting appropriate features, thereby improves model performance, particularly given the high diversity and noise in the data. Below are three main features that contribute to effective LV load forecasting and illustrate why achieving accurate forecasts at the LV level is particularly challenging. The primary method for evaluating whether a factor is suitable involves examining its correlation with the load data, as correlation directly indicates the degree of dependence between the feature and the load. Although machine learning and other advanced algorithms are capable of capturing non-linear relationships, correlation analysis remains a straightforward approach for identifying features that are directly related to the load data.

2.3.2 Characterisation and Analysis of LV Load Log Data

The primary factor considered in load forecasting is lagged data, reflecting the fact that individuals typically maintain consistent daily, weekly, and yearly routines. People tend to repeat similar behaviours at specific times each day, such as wake-up schedules and commuting hours, and follow weekly habits, including regular work habits and weekend activities [26]. These all contribute to stable electricity usage patterns. A common initial benchmark in load forecasting is that tomorrow's load will be the same as todays or the same as the equivalent day in the previous week. This simple approach is widely known as the Persistence method [27]. To illustrate the effect of day equivalency, autocorrelation analysis is conducted using lagged observations at one day and one week.

Autocorrelation is a fundamental statistical technique employed to quantify the linear relationship between the current observation and a lagged value of the same variable. This method facilitates the identification of repeating patterns and periodicities within the data. The normalised autocorrelation function for a given lag k is defined as:

$$r_k = \frac{\sum_{i=1}^{N-k}(Y_i - \bar{Y})(Y_{i+k} - \bar{Y})}{\sum_{i=1}^{N}(Y_i - \bar{Y})^2} \tag{2.3}$$

where k denotes the lag, N is the total number of the observations, $\bar{Y}$ is the mean of the data, Y_i is the value at i, r_k is the autocorrelation coefficient at lag k. The value of the normalised autocorrelation coefficient ranges between -1 and 1. A negative value indicates an inverse

relationship, whereas a positive value signifies a direct relationship. An absolute value closer to 1 implies a stronger linear relationship, while a value of 0 indicates no linear dependence.

Figure 2.5 and 2.6 present half-hourly time-series autocorrelation plots for two representative rural GB substations, each supplying approximately 100 customers. For each substation, the one-year half-hourly data were arranged so that each half-hour of the day forms a separate time series, and the normalised autocorrelation coefficient r_k, defined in Equation 2.3, was computed at daily ($k = 48$) and weekly ($k = 336$) lags. The resulting values of r_k for each half-hour constitute the curves shown in the figures.

The results reveal pronounced time variation in correlation, but both substations display similar patterns. Elevated autocorrelation is observed (i) between 23:00 and 03:00, when most households are asleep; (ii) from 07:00 to 09:00, coinciding with typical morning routines; and (iii) during the evening peak as residents return home from work. These intervals reflect the high degree of synchronisation in daily and weekly demand.

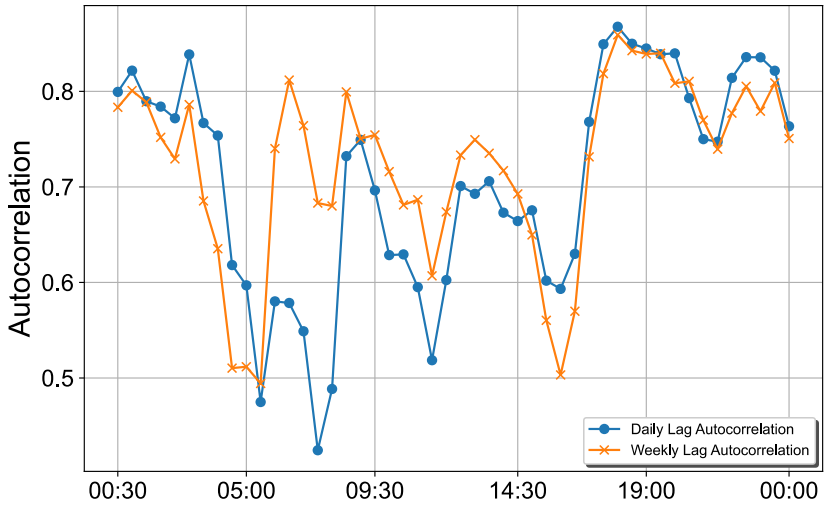

Figure 2.5 Time-series autocorrelation: daily and weekly comparison at Substation C

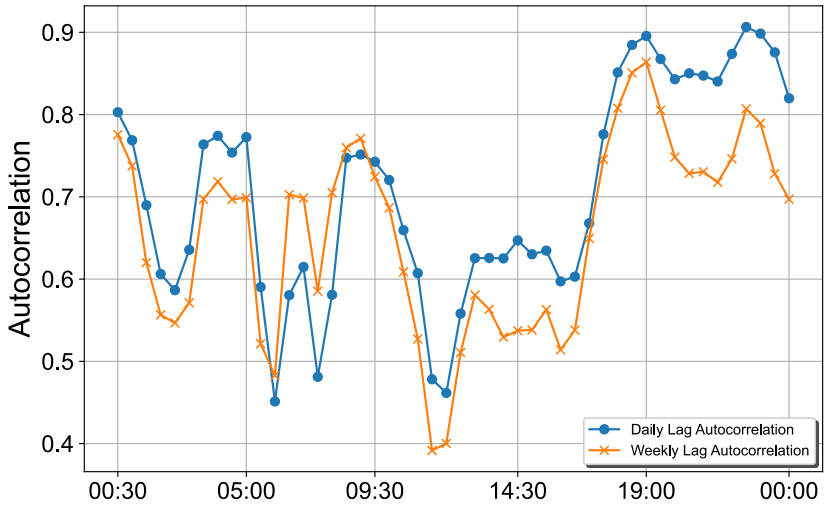

Figure 2.6 Time-series autocorrelation: daily and weekly comparison at Substation D

Conversely, autocorrelation falls markedly in the pre-dawn period like 04:00–06:00 and in the late afternoon. The diminished pre-dawn values are attributable to variability in wake-up times and in the use of high load appliances, whereas the late-afternoon decline is likely influenced by heterogeneous working patterns and weather-related factors.

Although time-series plots offer a useful overview of network performance, the autocorrelation results reveal that demand during the morning, evening and midnight periods shows high linear dependence, making it generally easier to identify underlying trends, whereas early-morning and mid-afternoon demand exhibits weaker linear dependence, resulting in more stochastic fluctuations. Because DNO system operators mainly focus on the magnitude and timing of demand peaks and troughs, the subsequent figures analyse the autocorrelation of demand at the evening and morning peaks, as well as at the daytime trough, across a range of customer-level substations.

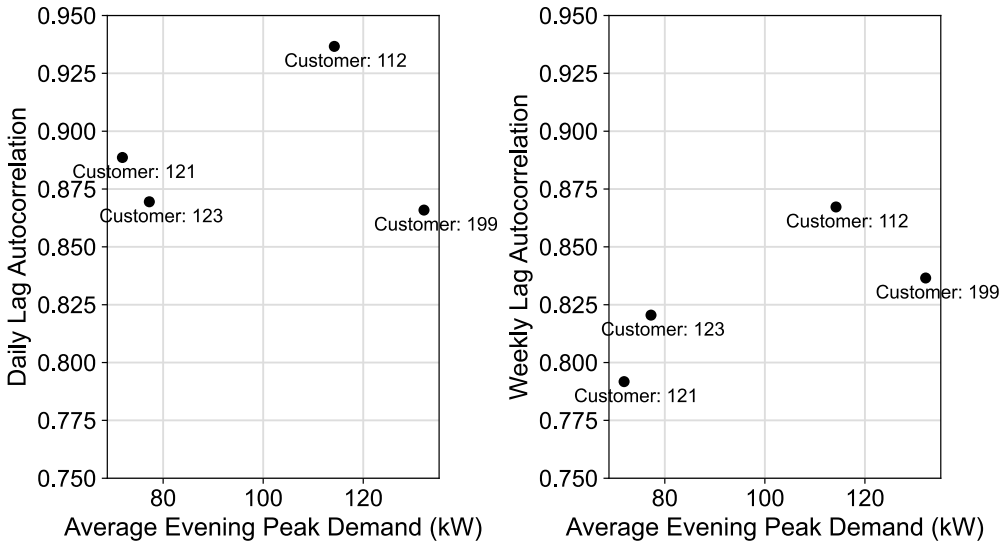

Figure 2.7 Autocorrelation vs. evening peak demand: daily and weekly comparison in GB for various customer aggregations

Figure 2.7 presents the autocorrelation of evening peak demand across substations, evaluated at both daily and weekly lags. The analysis demonstrates consistently high autocorrelation at the evening peak for all substations, as observed previously. Generally, substations with larger customer bases and higher peak demands display higher autocorrelation coefficients, suggesting more stable and repetitive consumption. Conversely, substations with lower aggregate demand often show reduced autocorrelation, reflecting the volatility inherent in smaller groups.

This relationship is not strictly proportional, as specific substation load characteristics can influence predictability regardless of customer number. For example, high-density housing can artificially boost predictability through synchronized usage, where similar demographics and building types lead to nearly identical consumption patterns. Consequently, the correlation between demand size and autocorrelation remains a broad trend rather than a fixed rule, modulated by the specific customer mix of each substation.

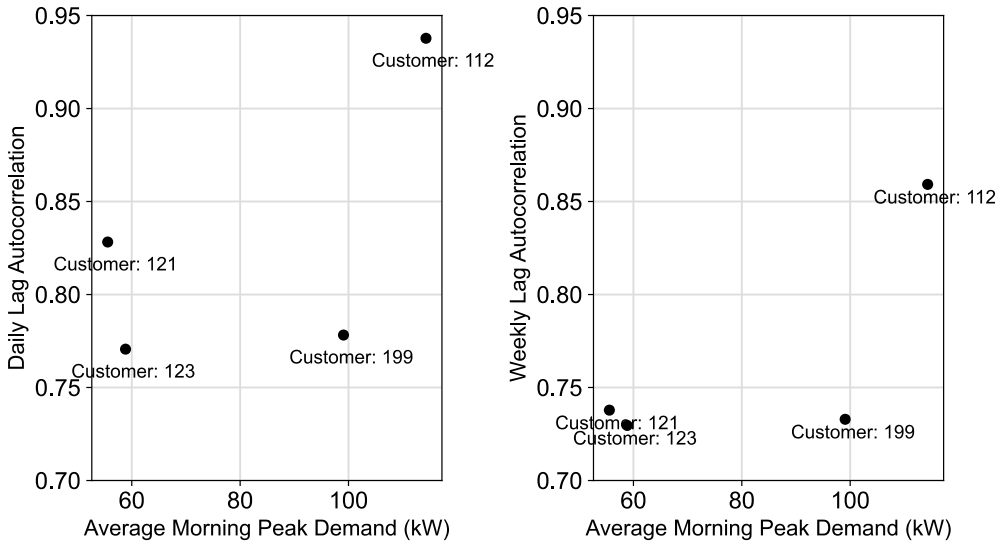

Figure 2.8 Autocorrelation vs. morning peak demand: daily and weekly comparison in GB for various customer aggregations

Figure 2.8 illustrates the autocorrelation of morning-peak demand across the substations. In general, for the same substation, autocorrelation coefficients for the morning peak are roughly 10 percentage points lower than those observed for evening peak above, implying that morning demand is more variable and influenced by external factors such as sunrise time. Unlike the evening-peak analysis, the number of customers and the total demand exert a weaker influence on autocorrelation during the morning peak. A few substations show greater predictability, likely because the consumers they serve use high-load appliances in a more regular pattern. However, most substations exhibit considerable variability, which makes morning demand harder to forecast.

Figure 2.9 shows the autocorrelation of the daytime trough across the surveyed substations. The coefficients are markedly lower than those for the morning peak, indicating that daytime troughs are subject to greater variability and are therefore more changeable and less predictable. Whereas peak loads typically reflect the aggregate behaviour of a major people behaviour, trough levels can be disproportionately influenced by the actions of a relatively minority of consumers; minor deviations in their demand can shift the trough appreciably, making it inherently harder to forecast. Moreover, the stronger autocorrelation at the one-day lag compared with the one-week lag suggests that the temporal dependence of the daytime trough is driven principally by short-term daily dynamics rather than by longer-term weekly cycles.

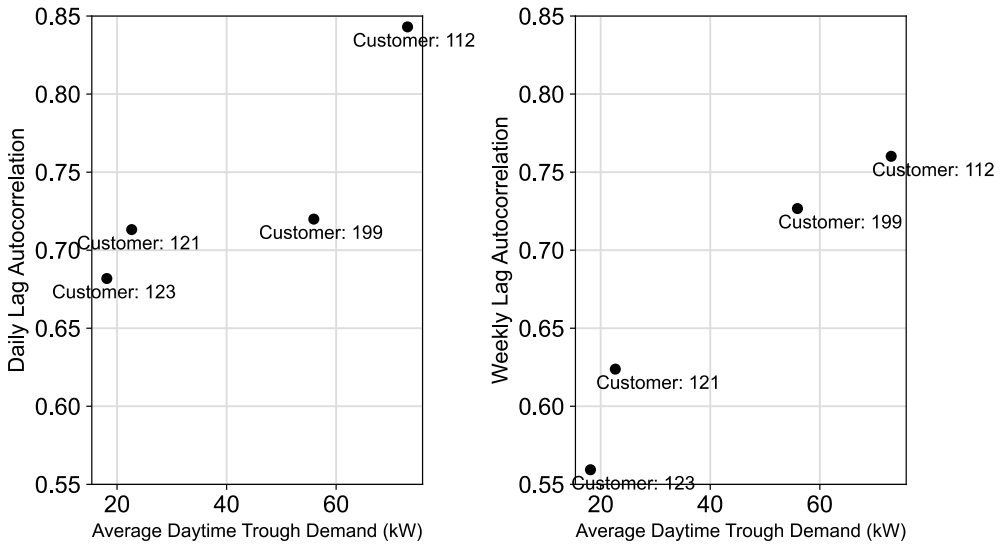

Figure 2.9 Autocorrelation vs. daytime Trough demand: daily and weekly comparison in GB for various customer aggregations

2.3.3 Characterising and Analysing LV Load Based on Temperature Data

The second major factor in load forecasting is weather data, because temperature heavily influences the operation of energy-intensive appliances such as heat pumps and electric water heaters. Although temperature is widely acknowledged as an effective predictor for aggregated HV demand, its utility at the LV level remains contested. Several studies indicate that incorporating weather variables can be neutral or negative to LV forecasts. [28] shows that the incremental explanatory power of calendar effects declines sharply once temperature covariates are added. Likewise, [23] finds no systematic improvement in forecast accuracy when either observed or modelled temperature data are included.

The relationship between temperature and LV load is quantified using the sample Pearson correlation coefficient, which provides a measure of the linear association between the two variables based on the observed sample data. The coefficient is calculated as follows:

$$r_{X,Y} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_{i+k} - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{N}(Y_i - \bar{Y})^2}} \tag{2.4}$$

where X represents temperature, Y denotes LV load, N is number of the paired observations, and $\bar{X}$ and $\bar{Y}$ are their respective sample means.

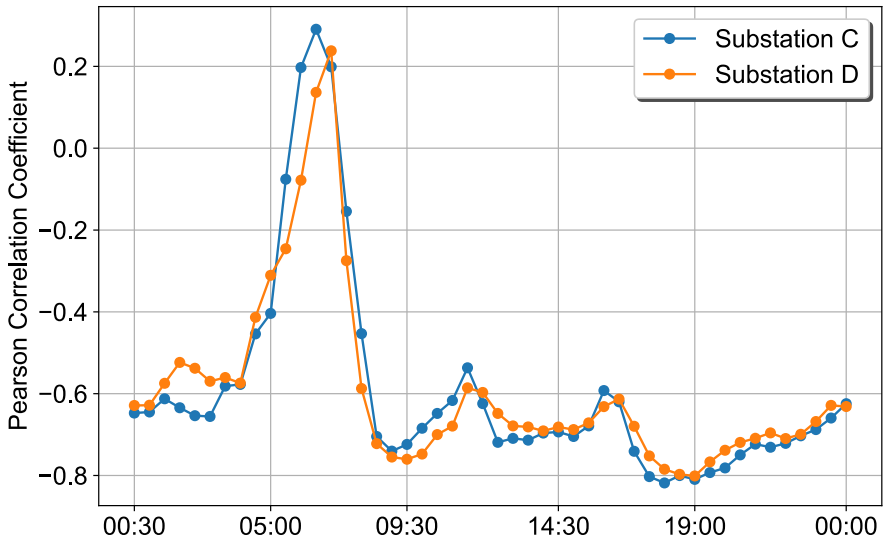

Figure 2.10 Correlation between temperature and load in the same time period at Substation C and Substation D

Figure 2.10 shows the correlation between temperature and load over the same time period at substation C and substation D at the same location. The results show that, for most of the time, there is a negative correlation between load and temperature, which is consistent with the findings in [23]. In the evening, there is a stronger negative correlation with temperature, meaning that as temperature increases, load decreases. This indicates that at night, higher temperatures reduce the need for heating, and therefore less load is required. In the context of GB this relationship is especially evident as air conditioning is quite rare in GB homes, so increased temperatures do not lead to a corresponding rise in electricity demand for cooling, unlike in some other countries. like the United States, the relationship between temperature and load is often not linear and can even look like a two-tailed curve (U-shaped) [29].

In the morning, however, there is a sharp rise in the correlation, even turning slightly positive for a short period. This is likely when people begin to wake up and start using kettles or heating. Although the temperature may also begin to rise, human activity has a greater influence on load than the outside temperature during these hours, resulting in very little linear relationship between temperature and LV load.

Additionally, substations C and D show very similar results, which is likely because they are in the same area. This suggests that substations in the same location may exhibit similar load behaviour, and it also indicates the potential for employing few-shot learning and transfer learning techniques, as these substations share the same pattern at the same location.

Although temperature has a strong relationship with LV load during the evening, the data used here are actual temperature values, which is not entirely representative of real-world forecasting scenarios. In practice, real-time temperature values are not available prior to making a load forecast. Therefore, temperature forecasts are typically used for practical load forecasting. However, data for temperature forecasts are rarely archived and can be difficult to obtain. In this study, the previous day's temperature is used as a proxy, as it represents the most recent available data.

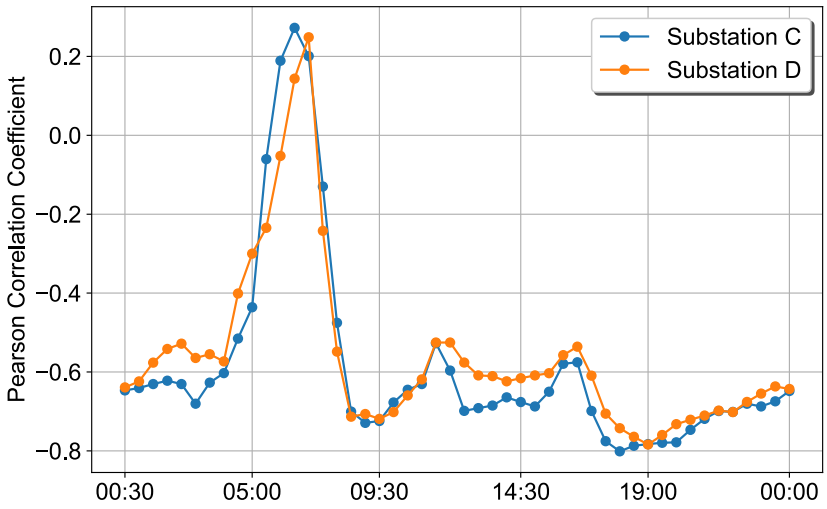

Figure 2.11 Correlation between day before temperature and load in the same period at Substation C and Substation D

Figure 2.11 shows that use of the day before temperature correlation with loads, it shows the similar results with the actual temperature in Figure 2.10, that is because daily temperatures often exhibit high autocorrelation, meaning the temperature at a given time on one day is usually similar to the temperature at the same time on adjacent days. As for GB, there are always for consecutive days in temperate climates, where weather patterns do not typically change drastically overnight.

On the other hand, human activities and the thermal inertia of buildings can cause electricity demand to respond not only to the immediate temperature but also to temperatures experienced over preceding hours or days [30]. For example, heating requirements are influenced by how warm buildings have been in the previous hours, rather than solely by the current temperature. Consequently, the previous day's temperature retains predictive value for electricity demand, particularly at certain times of day. The strong correlation observed between load and both current and previous day's temperatures validates this approach.

Additionally, the rolling average temperature can also serve as a significant explanatory variable in LV load forecasting. Rolling average features have demonstrated high importance in prior studies employing prediction models on time series load data [31]. The rolling average is calculated as follows:

$$RollingAverage_t = \frac{1}{N}\sum_{k=1}^{N} T_{t-k} \tag{2.5}$$

Where T_{t-k} represents the temperature at time step $t-k$, and the N denotes the window size over which the average is computed. The rolling average thus represents the mean temperature at time t based on the preceding N observations.

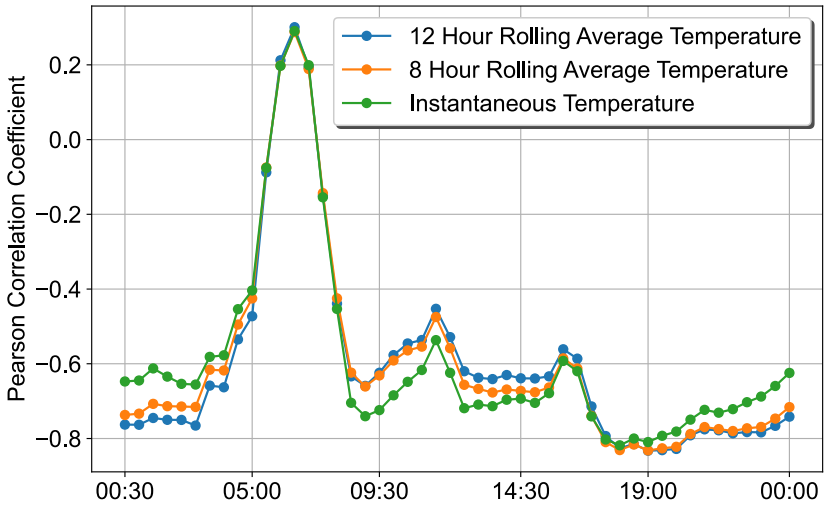

Figure 2.12 Comparison of rolling average temperature with instantaneous temperature

Figure 2.12 presents a comparison between the 8-hour and 12-hour rolling average temperatures and the instantaneous temperature, in terms of their correlation coefficient with electrical load.

Following [29], 8-hour and 12-hour rolling average temperatures are adopted as representative short-term and intra-day averaging windows that capture the delayed impact of outdoor temperature on building demand. The methodology is not restricted to these two windows and could accommodate alternative or additional averaging periods if required; here, 8 and 12 hours are used as typical examples from the literature.

The results demonstrate that, during the evening and at midnight, the Pearson correlation coefficients derived from the 8-hour and 12-hour rolling average temperatures are not only higher in magnitude, but also more stable and consistent in comparison to those obtained from instantaneous temperature. Specifically, the correlation coefficients for the rolling averages remain close to -0.8, indicating a strong and persistent relationship with load with less fluctuation, whereas the coefficients based on instantaneous temperature decrease to approximately -0.6 and exhibit greater variability. This indicates that electricity demand during these periods is more strongly influenced by the accumulated warmth of buildings over preceding hours, rather than by immediate fluctuations in outdoor temperature. Conversely, during the daytime, the correlation between instantaneous temperature and load becomes stronger, indicating that electricity demand is more directly influenced by short-term changes in temperature, potentially reflecting a range of temperature-dependent behaviours. These findings highlight the importance of considering both recent thermal history and real-time temperature when understanding and forecasting electricity demand across different periods of the day.

2.3.4 Characterising and Analysing LV Load Based on Calendar Data

Calendar performance is a key characteristic in load forecasting. In GB, seasonality exerts a significant influence, with recurring patterns each year reflected in calendar data [32]. Firstly, variations in daylight hours across different seasons substantially affect electricity consumption. These effects are observed through changes in lighting demand and the use of wet appliances [33]. As a high-latitude country, GB experiences extended daylight hours during the summer, encouraging people to spend more time outdoors which lower the lighting demand. Conversely, shorter daylight hours in winter tend to result in individuals returning home earlier, thus increasing electricity use in the evening. Secondly, difference between weekdays, weekends, and bank holidays cause marked changes in consumer behaviour, which, in turn, influence the usage patterns of electrical devices and alter the LV load profile. On weekdays, most individuals follow a regular

working schedule. In contrast, weekends are typically characterised by more flexible routines. This could lead to flatter load profiles during the day and shifts in peak demand periods. Additionally, bank holidays, such as Christmas or Easter, can lead to exceptional drops in demand, as many people are trend to go on holiday. Thirdly, seasonal weather patterns further impact electricity demand by modifying how consumers use certain appliances. For example, during the winter months, increased rainfall from November to January means people dry clothes with dryers rather than on clothes lines in winter. These weather-driven behavioural adjustments contribute to the observed seasonality in electricity demand profiles.

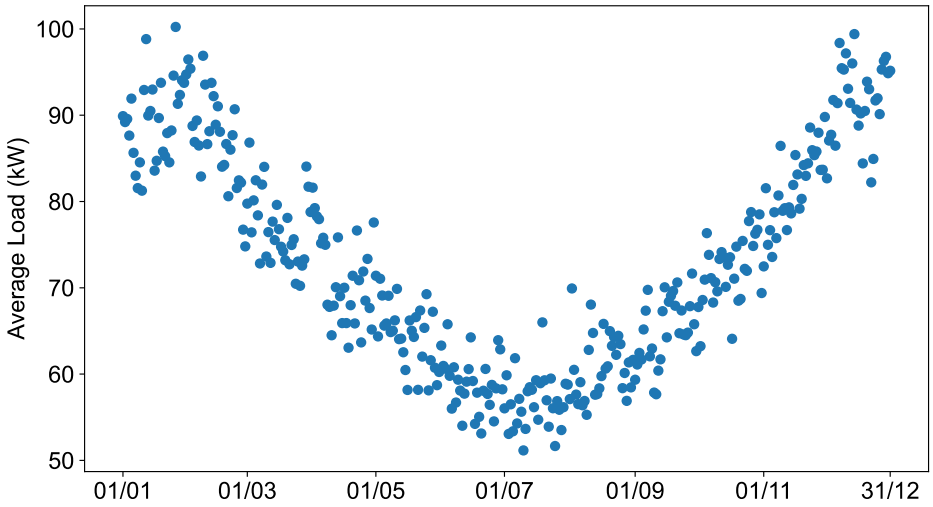

Figure 2.13 Annual load patterns of Substation C

Figure 2.13 presents the average annual load profile for Substation C. The pattern exhibits well-defined seasonal variation: demand climbs steadily as temperatures fall, producing pronounced peaks, in the winter months attributable to heating and lighting loads. In addition, the profile shows an abrupt decline over the Christmas and New Year holidays, reflecting that many people travel or visit family during this period.

To gain a clearer understanding of these patterns, correlation heatmaps are used to summarise how load varies in relation to common calendar variables, such as month and weekday. These heatmaps are effective visual tools that enable the rapid identification of relationships among multiple variables within a dataset, offering a comprehensive overview of all pairwise interrelationships briefly. In this context, the intensity of the colour serves as an intuitive indicator of correlation strength: lighter shades denote stronger correlations, while darker shades indicate weaker ones.

Notably, the diagonal of the heatmap, where each variable is compared to itself, will display the brightest colour, representing a perfect correlation of 1. By utilising correlation heatmaps, the interrelationships among all calendar related variables can be visually assessed. Examples of such variables include the month of the year, the day of the week, and the day of the month. The corresponding visualisations based on Substation C are presented below.

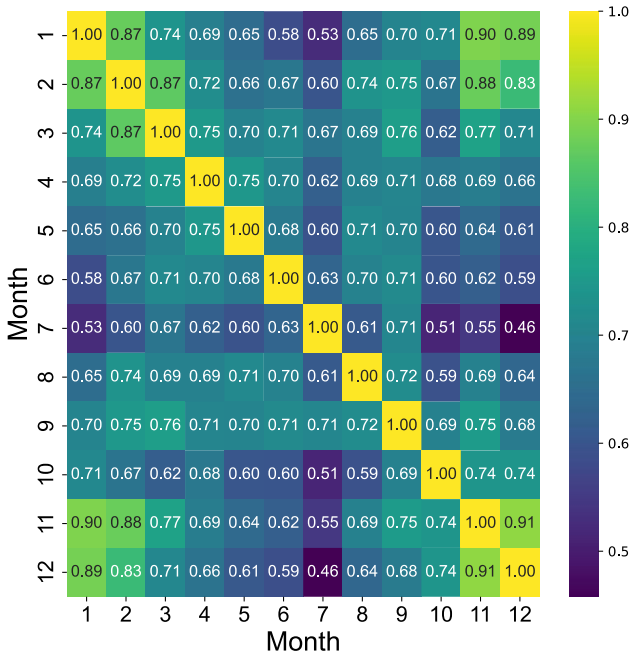

Figure 2.14 Heatmap of correlation between month profiles at Substation C

Figure 2.14 presents the correlation heatmap for each month of the year at Substation C. The results clearly demonstrate a pronounced seasonality effect in the load profiles. Adjacent months exhibit strong similarity, as reflected by high correlation coefficients. Furthermore, it reveals that the winter months like January, February, November, and December are highly correlated with each other, with coefficients around 0.9. In contrast, the summer months display lower inter-correlation, with values approximately 0.7. The correlation between winter and summer months is further reduced, with coefficients dropping to around 0.5. This pattern underscores the pronounced seasonality in GB's LV load profiles and highlights the importance of explicitly modelling the cyclical nature of the calendar year in load forecasting.

It is noteworthy that, despite being numerically the furthest apart on a linear calendar encoding, January and December exhibit a strong relationship in load behaviour. The month-to-month

similarity matrix indicates a high correlation between these two months, suggesting that seasonal demand transitions smoothly across the year boundary rather than changing abruptly. This similarity is influenced by shared winter conditions, including comparable weather-driven heating demand and daylight patterns. This implies that treating months as a simple linear variable can introduce an artificial discontinuity between December and January. This observation supports representing months as cyclical features, where January and December are adjacent points on a temporal circle, rather than as endpoints on a linear scale [28]. Incorporating this circularity can therefore improve the representation of annual seasonality in forecasting models.

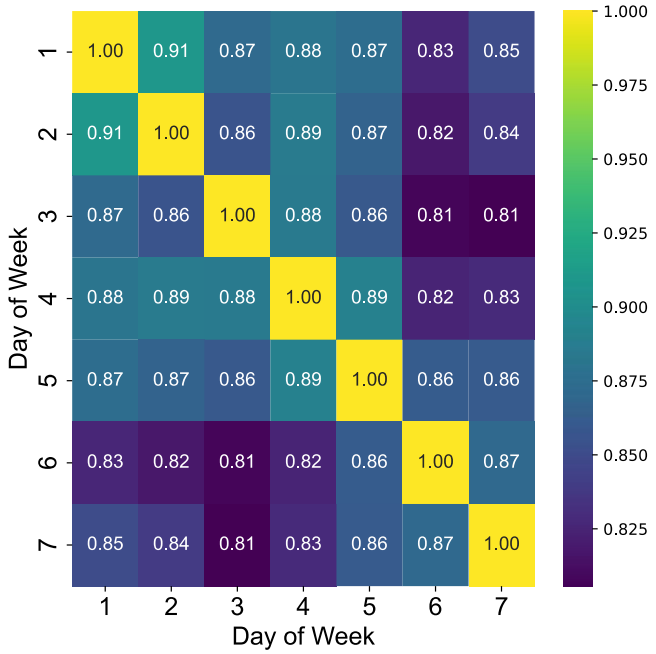

Figure 2.15 Heatmap of correlation between weekday profiles

Figure 2.15 presents the correlation heatmap for the day of the week at Substation C. Compared with the monthly data, the correlations within each week are notably higher, with values exceeding 0.8. As expected, the correlation between weekends and weekdays is lower, whereas the correlation among weekdays is relatively consistent, with a value of approximately 0.88. The correlation between weekdays and Friday is slightly lower, at around 0.86. Neglecting these distinctions can result in suboptimal load forecasting. However, because most datasets are limited to a year or less, developing individual models for weekends is often impractical due to insufficient data.

For this reason, incorporating day-of-week information through one-hot encoding is considered an effective strategy. A one-hot encoder is well suited to distinguishing between these categories. This method uses a set of binary variables, with only one variable active at a time while the others remain inactive. By representing each day as a separate binary feature, the model can learn specific patterns associated with each day type without the need for explicit segmentation or the construction of multiple models. This representation places the categorical variables into a low-dimensional space, which significantly improves convergence speed while maintaining accuracy [34].

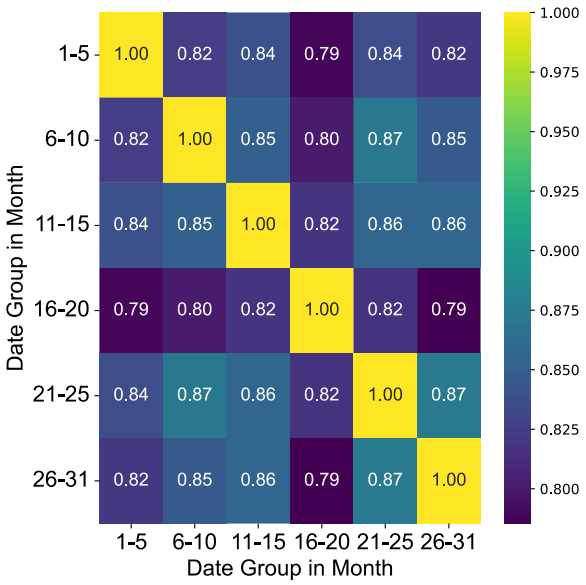

Figure 2.16 Heatmap of correlation between date group profiles

Figure 2.16 presents the correlation heatmap for date groups within a month at Substation C, a rural secondary substation on the east coast of Scotland serving 123 customers. The analysis uses a full year of half-hourly interval data. Each group corresponds to a distinct period: the beginning, middle, and end of the month. These periods capture notable variations in consumer behaviour. For example, towards the end of the month, many individuals receive their salaries or settle accounts, which can significantly influence both residential and commercial energy consumption. In some instances, individuals may defer certain activities such as laundry or household chores until after payday, resulting in a temporary surge in demand at the beginning of the subsequent month.

The correlations between the beginning (days 1-5) and end (days 26-31) of the month are relatively high, with values generally exceeding 0.82. By contrast, the mid-month period (e.g., days 16-20) exhibits lower correlations with both the start and end groups, with values around 0.79. This pattern indicates that consumer behaviour in the middle of the month differs from that observed at the beginning and end, potentially owing to the absence of salary payments, billing cycles, or other recurrent monthly events during this interval.

However, a simple bivariate analysis is insufficient for characterising energy demand because it fails to account for confounding factors such as seasonal trends. To address this limitation and identify which days of the week are most impactful, a multivariate Ordinary Least Squares (OLS) model is employed. This specification improves upon simple correlations by including one-hot encoded weekday indicators while controlling for the annual trend using a Day of Year variable. To avoid multicollinearity and ensure interpretability, Friday is used as the reference category. This allows the model to isolate the incremental effect of each weekday relative to the Friday baseline. The model is trained using observations from January to September and evaluated on October data to validate the structure out-of-sample.

Day of Week	Coefficient	p-value
Monday	-9.19	0.837
Tuesday	-50.31	0.260
Wednesday	-22.35	0.617
Thursday	-40.02	0.370
Saturday	71.62	0.109
Sunday	125.89	0.005

Table 2.1 Linear regression estimates of Day-of-Week impact

Table 2.1 indicates that Sunday is associated with a statistically significant increase in energy relative to the baseline ($p = 0.005$), suggesting that weekend behaviour or operational patterns may meaningfully influence demand. Saturday shows a positive but non-significant increase ($p = 0.109$), while the effects for Monday to Thursday are not statistically distinguishable from Friday in this specification. Overall, these findings address the limitation of bivariate analysis by demonstrating which weekday differences persist after accounting for broader seasonal structure.

Future work can strengthen this analysis by incorporating additional covariates known to confound simple correlations, such as temperature, public holidays, and special events. In addition, to substantiate the hypothesis of begin/end of month behavioural change, a dedicated regression specification could include an explicit end-of-month indicator, for example the final three to five days of each month, alongside weekday controls and seasonality terms, to test whether any end-of-month uplift remains statistically significant after adjustment for overlapping calendar effects. A richer seasonal representation, such as monthly indicators or Fourier terms, could also be compared against the quadratic day-of-year form to assess robustness. Given the likely serial dependence in daily energy data, further refinement could include autocorrelation-consistent (HAC) standard errors or a time-series regression framework with lagged demand. Collectively, these extensions would provide a more resilient quantification of calendar-related effects than a minimal model based on simple trend and weekday structure.

2.3.5 Challenges of Forecasting LV Load

Overall, the preceding analysis demonstrates that, although the aggregate effect of LV load is comparatively modest due to the smaller number of customers at this level, clear relationships exist between lagged load data, weather variables and calendar effects. However, at the LV and individual smart-meter level, customer load profiles are often highly volatile, stochastic and difficult to forecast directly because of the ‘noise’ associated with individual human behaviour.

To address this volatility, a widely adopted strategy in recent smart-meter load forecasting literature is the application of clustering techniques [35]. Clustering methods, such as k-means, are used to group customers with broadly homogeneous consumption patterns before training forecasting models. The idea is that by grouping similar customers together random noise is reduced and it becomes easier to see stable and predictable relationships with the explanatory variables [36]. This pre-processing step effectively transforms otherwise weakly predictable individual series into more predictable grouped series.

However, this approach has limitations for the specific objectives of this thesis. From a DNO perspective, operational decisions require forecasts for specific substation values rather than abstract customer clusters. Aggregating data into clusters can also hide critical local distinctions, such as seasonal variations. Therefore, to ensure predictions are operationally relevant, this thesis applies direct forecasting at the substation level instead of relying on customer clustering.

Furthermore, as LV load data are frequently at only the city or village level, the requirements for corresponding weather data become more stringent. Many datasets offer only a limited number of weather observations across GB, which can lead to the erroneous assumption that all locations experience identical weather conditions. Moreover, obtaining accurate weather forecast data remains a significant challenge. Lagged load data tend to exhibit the strongest correlations at the substation level and generally offer the highest predictive value. However, their use is constrained by the limited availability of monitoring infrastructure at LV substations, as most secondary substations do not routinely collect such data.

In summary, the correlation analyses in this section identify several parameters that are carried forward into the subsequent chapters of this thesis. Lagged LV load at daily and weekly horizons remains a core baseline feature, with stronger predictability observed for substations serving larger customer groups and during typical peak periods. Temperature is shown to be informative at specific times of day, and the use of previous-day temperature and short rolling-average temperatures provides a practical proxy for realistic forecasting conditions. Calendar variables, including month and day-of-week effects, capture clear seasonal and behavioural structure and are therefore retained as key explanatory features. Finally, the similarity observed between substations in the same region suggests potential value in leveraging cross-substation information to support forecasting under limited data. These findings directly inform the feature selection and forecast model design in Chapters 3 and 4.

This work proposes an approach that achieves acceptable forecasting errors for substations where data are unavailable or limited by utilising known variables such as customer count, weather, and calendar information, rather than relying solely on state-of-the-art LV load forecasting methods that require fully observed data. Given that LV load is predominantly influenced by customer behaviour, an additional consideration is whether lagged data from other substations can be systematically leveraged to enhance forecasting accuracy. Chapters 3 and 4 address these challenges by presenting two novel methods designed to improve LV load forecasting under such conditions.

Building on these substation-level forecasts, the subsequent section examines how load forecasts can be integrated across the network to generate operator-relevant information, such as voltage or current limits, supporting day-ahead decisions and month-ahead planning. It also considers

whether network-wide integration can improve outcomes even when individual forecasts contain errors. This motivates the next sections. Section 2.4 introduces distribution network modelling and power flow analysis to link load uncertainty to operating conditions, and Section 2.5 reviews state estimation and dynamic state estimation to enhance observability under limited measurements.

2.4 Modelling Distribution Networks and Power Flow Analysis

Historically, distribution networks attracted less analytical and modelling attention than generation and transmission infrastructure. In the absence of detailed studies, engineers adopted markedly conservative design philosophies, embedding generous capacity margins to accommodate every conceivable eventuality. However, the rapid spread of DERs, increased demand from technologies such as EVs, and increasingly stringent regulatory obligations now compel DNOs to exploit the full latent capability of their networks. To meet this imperative requires, first, an accurate quantification of the true maximum capacity of each distribution system and, second, the development of rigorous methodologies by which that capacity can be assessed. It further necessitates the systematic identification of all operational constraints: thermal, voltage, fault-level, and reliability, all of which depend on the availability of highly accurate models for every principal component of the network. Accordingly, detailed representations of lines, cables, transformers, protection devices, and embedded resources are indispensable [37].

Therefore, error metrics alone are insufficient for gaining deep insight into forecasting performance of individual distribution substation loads. Equally important is the evaluation of how forecast errors influence the behaviour of the wider distribution network. Simultaneous mis-forecasts at several substations can precipitate rapid voltage excursions outside statutory limits. Robust network modelling thus becomes essential: by tracing the spatial distribution of voltages and currents at every busbar under forecast conditions, such models enable operators to anticipate emerging operational challenges and devise proactive control strategies.

2.4.1 Modelling the Distribution Network

In this thesis the distribution network is modelled as a radial distribution feeder, the configuration most widely employed across the world. A radial feeder affords only a single path for power to travel from the distribution sub-station to each customer. Because it must serve a large number of unequal single-phase loads, its loading is inherently unbalanced.

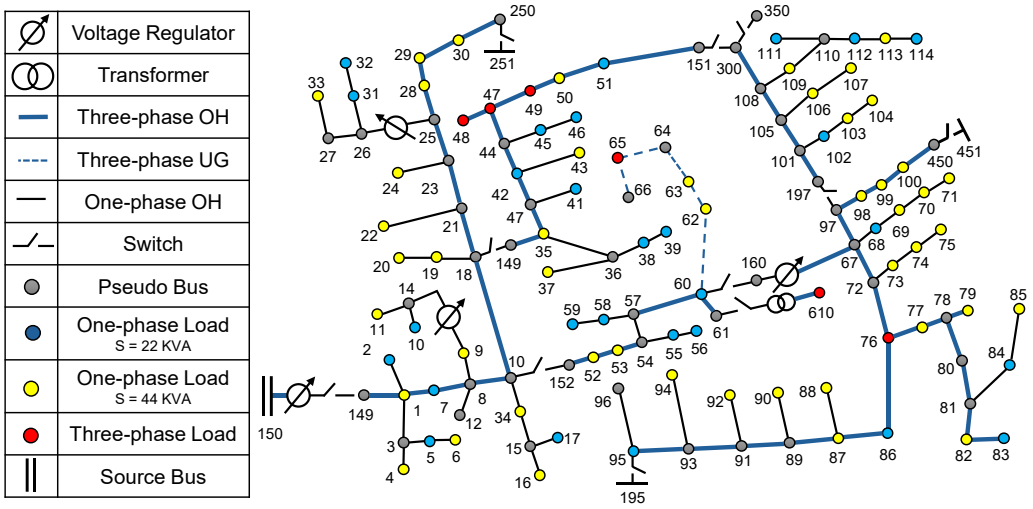

Figure 2.17 IEEE 123-bus test feeder map [38]

Figure 2.17 illustrates the IEEE 123-Bus Test Feeder, a benchmark system that contains a broad spectrum of components chosen to reflect the operational complexity of a real distribution network. In this North American based system the voltage is stepped down from 115 kV to 4.16 kV, and the primary circuits operate at a nominal 4.16 kV [38]. The HV substation transformer connects the 115 kV transmission source to Bus 150. Additionally, to enhance system reliability, the feeder architecture incorporates Normally Open switches, specifically the tie-link connecting Bus 250 and Bus 251, as well as interconnection points at buses 350 and 451. These nodes facilitate topology reconfiguration, allowing the network to transition from a radial structure to temporary loops for load transfer during contingency events.

A reliable network model enables the system operator to identify current operating conditions, formulate day-ahead plans, and issue actions on the basis of projected feeder loading and substation monitoring data. A typical feeder schematic therefore includes the following elements:

1. Voltage Regulators: These automatic devices raise or lower the voltage to maintain it within prescribed limits, usually close to the nominal value (for example, 33 kV, 11 kV, or 415 V, depending on the level of the GB distribution system). The parameters required include potential/current transformer ratios, compensator settings, and bandwidth.

2. Transformers: These electromagnetic devices adjust AC voltage levels, either stepping them up for efficient long-distance transmission or stepping them down to levels suitable for safe customer use. The key parameters are kVA rating, impedance, and no-load power loss.

3. Underground cables and overhead lines (single or three-phase): These conductors provide the physical pathways that convey electricity from substations to customers. Overhead lines, suspended from poles or towers, are comparatively inexpensive to install and repair but have a significant visual impact on the landscape. They are also vulnerable, as bird collisions, vandalism, and extreme weather can all damage them. Underground cables, laid in ducts, are protected from climatic effects and remove visual intrusion; however, they involve higher capital expenditure, restrict heat dissipation, and require specialised techniques for fault repair. Consequently, underground cables are generally preferred in urban and suburban areas, whereas overhead lines are more common in rural districts. Conductors may be single-core or three-core, allowing the feeder to operate in either single-phase or three-phase mode according to load requirements. The parameters needed include cable diameter (mm), resistance (Ω/km), reactance (Ω/km), and capacitance (μF/km) where Ω denotes ohms, μF denotes microfarads, mm denotes millimetres and km indicates kilometre. for positive-sequence and zero-sequence components.

4. Switches: These devices can open or close selected sections of the feeder. They permit fault isolation so that only a limited part of the network is de-energised after, for example, a short-circuit. Switches also facilitate network reconfiguration by redirecting power flows along alternative routes.

5. Pseudo buses: These are modelling constructs that represent points where phases split or merge. Although they do not correspond to physical plant, they clarify electrical connectivity and phase relationships among real buses.

6. One and three-phase loads: Most distribution-level demand in GB is one-phase, reflecting the modest power requirements of residential and other small customers. While economical, extensive one-phase load introduces phase imbalance, which can impair system performance. Three-phase loads, typical of larger commercial or industrial customers, receive a more stable and efficient supply but are comparatively uncommon at distribution system.

The network modelling presented in this thesis is relatively straightforward compared with the IEEE 123-bus network. As the primary aim of the experiments conducted in the thesis is to compare measured data with the equivalent outputs from transfer learning and few-shot learning in terms of overall network performance, as discussed in Chapter 3 and 4. To simulate steady-state conditions, the voltage regulators are omitted, and all switches are assumed to operate as usual with pseudo buses are ignored. The network is therefore presumed to operate under normal conditions, with no need for fault isolation. Transformer and cable parameters are taken from real specifications supplied by SSEN [39].

Nominal cross-section (mm²)	150
R_1: Resistance at 20 °C (Ω/km)	0.265
X_1: Reactance at 50 Hz (Ω/km)	0.091
C_1: Capacitance (μF/km)	0.33

Table 2.2 Network cable parameter

Rated Apparent Power (MVA)	26.4
Nominal Voltage on the High-Voltage Winding (kV)	33
Nominal Voltage on the Low-Voltage Winding (kV)	11
Positive and Negative Sequence Resistance (p.u.)	0.04707
Positive and Negative Sequence Reactance (p.u.)	0.6558
Zero Sequence Reactance (p.u.)	0.5559765
Starting Tap Ratio	0.97
Winding Connection	Dyn11

Table 2.3 Network transformer parameters [39]

As IEEE test networks are mainly based on North American voltage configurations, which differ markedly from those used in GB, two GB-based distribution networks are introduced here to reflect authentic GB voltage levels. The first is a relatively small but stable urban neighbourhood network, derived from substation geographical data and representative of a real GB urban distribution system. The second is a much larger GB test system, designed to assess the robustness and generalisability of the results.

The first urban neighbourhood network is derived from a subset of feeders within a GB distribution licence area that supplies both residential and light-commercial customers. It reflects typical GB distribution network secondary substation infrastructure: supply is delivered at 33 kV, feeders operate at 11 kV, and next lower voltage branches at 415 V serve the individual premises. In this network, the primary substation is balancing active and reactive power and providing the reference for voltage magnitude and phase angle across the network. A transformer within this substation steps the voltage down from 33 kV to 11 kV, and twenty loads are connected directly to the transformer's low voltage side. The corresponding network diagram is presented in Figure 2.18.

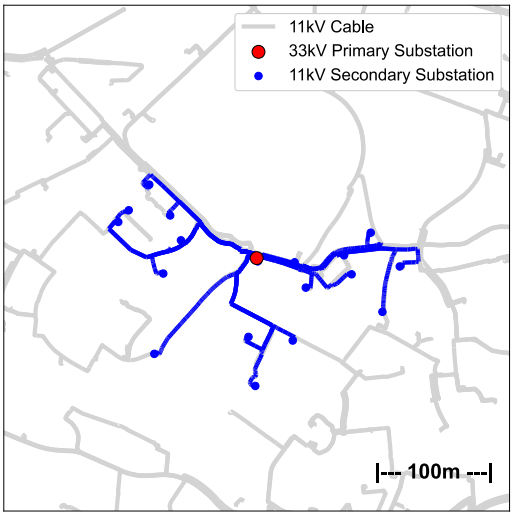

Figure 2.18 GB urban area 22-bus network

The second test system offers a broader representation of real GB distribution networks, particularly those in densely populated urban areas. It is based on the 77-bus variant of the UK Generic Distribution System (UKGDS). Developed by the Centre for Sustainable Electricity and Distributed Generation [40], the UKGDS provides stylised models designed to capture the extremes of UK distribution practice and to assess the scalability of new algorithms [40]. The full UKGDS comprises overhead lines and underground cables operating at 132 kV, 33 kV, and 11 kV, and contains 281 buses, 322 branches, and four grid supply points [41]. The 77-bus subset used in this study (illustrated in Figure 2.19) represents a section of the 11 kV network with 77 buses and 75 lines. In contrast to the 22-bus model discussed earlier, the 77-bus network includes a transformer that steps voltage down from 33 kV to 11 kV. Several hierarchical tiers radiate from the transformer's low-voltage side, covering an area of approximately 10 km². The system

incorporates one 33 kV substation that again serves as the slack bus and two transformers that step the voltage down to 11 kV. Seventy-five loads are connected to the remaining 75 substations distributed across the network, with an average separation of roughly 0.75 km between buses. The connections are organised in multiple levels so that distant substations operate at lower voltages. The network is designed to maintain voltage magnitudes within the range of 0.97–1.03 p.u., and the base apparent-power capacity is 10 MVA.

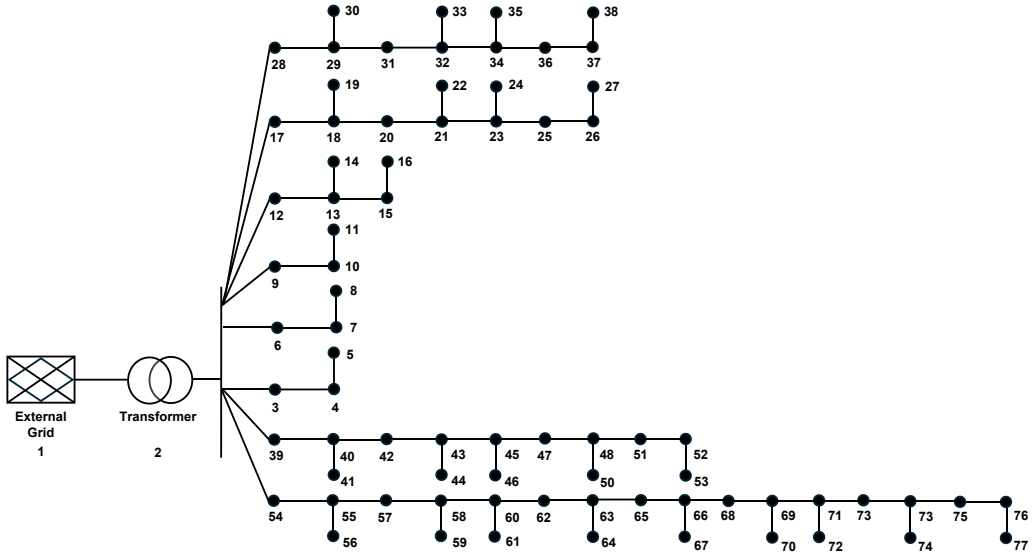

Figure 2.19 UKGDS 77-bus test network

After developing a detailed network model, distribution network analysis primarily involves examining feeder operation under steady-state conditions (power flow analysis) and fault conditions (short-circuit analysis). This thesis focuses exclusively on the steady-state case, where power flow analysis is used to determine the network's behaviours based on known data. It is one of the most widely employed tools in power system operation, enabling system operators to assess the real-time operating state of the network [42].

2.4.2 Power Flow in Distribution Networks

The power flow analysis of distribution systems is similar to that of transmission systems. By evaluating the power injected or drawn by each bus, methods such as the Newton-Raphson or Gauss-Seidel algorithms are applied to solve the non-linear equations governing power flow [43]. The results provide insights into bus voltage magnitudes, phase angles, branch power losses, and

voltage drops across feeders. These analyses allow operators to determine whether the system is operating within acceptable limits (e.g., ± 0.1 p.u. for voltage magnitude and ±10° for phase angles) [44] and to take appropriate actions such as opening or closing switches to maintain balance and ensure reliable operation .

In network analysis, a reference bus ('slack bus') is adopted as a point of reference, with a fixed voltage magnitude of 1 p.u. and a phase angle of 0 degrees. All other buses are modelled as PQ buses, where the injected complex power (comprising active (P) and reactive (Q) power) is specified and assumed to be independent of the bus voltage [45]. PV (voltage-controlled) buses are not required in this formulation, as the system is modelled without active voltage-regulating devices such as synchronous condensers or voltage-controlling distributed generation. This representation reflects the steady-state behaviour of most loads in practical distribution systems. In this thesis, the power flow solution is obtained using Newton's method.

Consider a network of N bus, where the relationship between the bus current I_n and bus voltage V_n is given by [46]:

$$I_n = \sum_{M=1}^{N} Y_{nm} V_m \tag{2.6}$$

And Y_{nm} denotes the entries of the admittance matrix. This matrix is square and encodes the network's electrical properties in terms of admittance: its diagonal elements contain the sum of all admittances connected to bus n , whereas its off-diagonal elements are the negative of the admittance of the branch connecting bus n to bus m. If there is no direct connection between bus n to bus m, then Y_{nm} is zero. The complex power injected at bus n, denoted by $P_n + jQ_n$, is given by:

$$P_n + jQ_n = V_n(\sum_{M=1}^{N} Y_{nm} V_m)^* \tag{2.7}$$

Where P_n and Q_n represent the real and reactive power injections at bus n, respectively, and $*$ denotes the complex conjugate.

To apply the Newton method, one constructs a Jacobian matrix capturing the partial derivatives of the power-flow equations with respect to the state variables (bus voltage magnitudes and angles).

The Jacobian relates small changes in real and reactive power to small changes in voltage magnitudes and phase angles. It may be represented as:

$$J = \begin{bmatrix} \frac{\partial P_n}{\partial \theta_n} & \frac{\partial P_n}{\partial V_n} \\ \frac{\partial Q_n}{\partial \theta_n} & \frac{\partial Q_n}{\partial V_n} \end{bmatrix} \tag{2.8}$$

Although the formulation is typically in polar coordinates, these partial derivatives are often evaluated using rectangular (real imaginary) arithmetic. The voltage at bus i rectangular form is:

$$V_i = |V_i|(cos\theta_i + jsin\theta_i) \tag{2.9}$$

Next, calculate the mismatches in real and reactive power at each bus n:

$$\Delta P_n = P_n^{scheduled} - p_n^{actual} \tag{2.10}$$

$$\Delta Q_n = Q_n^{scheduled} - Q_n^{actual} \tag{2.11}$$

Here, the "scheduled" power refers to the specified or desired real and reactive power injections at each bus. The "actual" power is the calculated real and reactive power injection at that bus, based on the current estimate of the system voltages during the iterative solution process.

$$J \begin{bmatrix} \Delta\theta_n \\ \Delta V_n \end{bmatrix} = - \begin{bmatrix} \Delta P_n \\ \Delta Q_n \end{bmatrix} \tag{2.12}$$

After solving for $\Delta\theta_n$ and ΔV_n, update the bus voltage. The power mismatch ΔP_n and ΔQ_n are then recomputed using these updated voltages, and the Jacobian is recalculated or updated (depending on the specific variant of Newton's method). This iterative process continues until all power mismatches fall below a specified tolerance, indicating convergence of the power-flow solution.

2.5 State Estimation and Dynamic State Estimation in Distribution Networks

In real-time operation, the measurements telemetered from remote-terminal units in power systems are often incomplete, redundant, inconsistent or noisy, so a reliable power-flow solution is rarely available straight from the raw data. To overcome this, Schweppe introduced state estimation [47], a statistical way that reconciles conflicting readings, suppresses noise, compensates for model inaccuracies and quantifies the remaining uncertainty. With the estimator in place, system operators regain a consistent view of system demand and bus-voltage magnitudes and can run on-line load-flow calculations that no longer diverge sharply from the off-line studies used for

planning. State estimation refers to the procedure of obtaining the voltage magnitude and phase at all the system buses at a given point in time.

2.5.1 State Estimation in Distribution Networks

In this thesis, the weighted least squares (WLS) state estimation technique is employed to analyse distribution networks. WLS minimises the weighted sum of squared measurement residuals, assigning each residual a weight equal to the inverse of the variance of its corresponding measurement, giving greater influence on more accurate data. The algorithm is the most widely adopted approach at the transmission level [48] and has been shown to perform reliably in distribution systems as well [49].

Let the state vector x comprise the voltage magnitude V_i and phase angle θ_i at every non-slack bus. Bus 0 is designated the slack bus and is fixed at $V_0 = 1\ p.u.$ and $\theta_0 = 0\ rad$.

$$x = [V_1, V_2, \dots, V_n\ \theta_1, \theta_2, \dots, \theta_n]^T \tag{2.13}$$

The measurement vector z contains active and reactive power injections, power flows, bus voltage magnitudes, phase angles and line current magnitudes. The nonlinear relationship between the state vector x, the measurement vector z, and the vector of known functions $h(x)$ is expressed as:

$$z = h(x) + e \tag{2.14}$$

where e is the measurement error vector, assumed to consist of independent, zero-mean Gaussian random variables with mean vector μ covariance matrix:

$$R = diag({\sigma_{11}}^2, {\sigma_{22}}^2 \dots {\sigma_{mm}}^2) \tag{2.15}$$

Consequently, the system comprises m measurements collected in the vector z and n state variables (voltage magnitudes and phase angles), $n < m$. The Jacobian matrix required for state estimation is therefore [50]:

$$H = \begin{bmatrix} \frac{\partial h_1(x)}{\partial \theta_1} & \dots & \frac{\partial h_1(x)}{\partial \theta_n} & \frac{\partial h_1(x)}{\partial V_1} & \dots & \frac{\partial h_1(x)}{\partial V_n} \\ \frac{\partial h_2(x)}{\partial \theta_1} & \dots & \frac{\partial h_2(x)}{\partial \theta_n} & \frac{\partial h_2(x)}{\partial V_1} & \dots & \frac{\partial h_2(x)}{\partial V_n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial h_m(x)}{\partial \theta_1} & \dots & \frac{\partial h_m(x)}{\partial \theta_n} & \frac{\partial h_m(x)}{\partial V_1} & \dots & \frac{\partial h_m(x)}{\partial V_n} \end{bmatrix} \tag{2.16}$$

In contrast with the power-flow Jacobian, this matrix is considerably larger and more complex, as it accommodates a wide variety of measurements, including active and reactive power injections, active and reactive power flows, voltage magnitudes, phase angles, and line current magnitudes.

The measurement vector z is modelled with independent, zero-mean, Gaussian-distributed errors with known variances. These variances are treated as known quantities obtained from meter accuracy classes and manufacturer specifications and where available from calibration data. This Gaussian error model is widely adopted in power system state estimation because metering errors arise from the combined effect of many small, uncorrelated error sources, so the resulting error distribution is well approximated by a normal law [48]. Under this model, the maximum-likelihood estimate $\hat{x}$ is the state vector that maximises the conditional probability density $f(z|x)$. Under the assumption of independent, normally distributed measurement errors with known variances, this maximum-likelihood formulation is mathematically identical to the weighted least-squares problem [51]. Therefore, maximising $f(z)$ is equivalent to minimising the objective:

$$\min J(z) = \sum_{i=1}^{m} (\frac{z-\mu}{\sigma})^2 \tag{2.17}$$

For the nonlinear state estimation measurement model $z = h(x) + e$, the WLS estimator solves:

$$\min J(x) = \sum_{i=1}^{m} \frac{(z-h(x))^2}{R} \tag{2.18}$$

which can be written compactly as:

$$\min J(x) = \frac{1}{2}[(z-h(x)]^T R^{-1}((z-h(x)) \tag{2.19}$$

A unique solution to this equation exists only when $m > n$, indicating that the system has more measurements than unknown parameters. If $m \leq n$, the system becomes underdetermined or exactly determined, resulting in infinitely many or non-unique solutions. This occurs because there are not enough independent measurements to uniquely determine all state variables, creating ambiguity in the state estimation.

The WLS estimator can be solved using Newton's method, a first-order Taylor expansion about the current estimate x^k gives:

$$h(x^k + \Delta x) \approx h(x^k) + H(x^k)\Delta x \tag{2.20}$$

By substituting this linearisation into the objective function and solving the resulting WLS problem, the Gauss–Newton normal equations are obtained as follows:

$$\Delta x = (H^T R^{-1} H)^{-1} H^T R^{-1} [z - h(x^k)] \tag{2.21}$$

The matrix $G = H^T R^{-1} H$ is known as the gain matrix, governs how the residual is mapped into the correction applied to the current estimate. Specifically, it determines the sensitivity of the solution to measurement errors and influences the convergence of the iterative process.

The state estimate is iteratively updated using:

$$x^{k+1} = x^k + \Delta x \tag{2.22}$$

Iterations continue until the magnitude of Δx falls below a prescribed tolerance (e.g. 10^{-6}). And x^{k+1} is taken as the final estimate.

State estimation in distribution networks is typically formulated using the full unbalanced AC power flow equations to iteratively adjust the state vector based on observed measurements. While conventional methods assume Gaussian errors, practical networks often exhibit non-Gaussian characteristics due to measurement and modelling anomalies. Consequently, non-Gaussian state estimation techniques are employed to explicitly model these error distributions, ensuring the estimated state is consistent with the specific probabilistic characteristics of the network.

Several robust alternatives to the conventional WLS distribution state estimator have been proposed to improve performance in the presence of gross measurement errors. Examples include the weighted least absolute value (WLAV) estimator and the Huber M estimator. WLAV is particularly robust to large outliers [52], while the Huber estimator offers a compromise between efficiency under Gaussian noise and robustness to bad data [53]. However, WLAV becomes less efficient when measurement errors are predominantly Gaussian, and the Huber estimator requires careful parameter tuning, incurs additional computational effort, and does not provide a straightforward way for bad data identification. As a result, WLS remains the most robust and widely adopted approach in power system applications is adopted in the remainder of this thesis.

Following several decades of research and refinement, state estimation for transmission networks has become a well-established engineering practice, supported by a mature and comprehensive methodological framework [54]. By contrast, state estimation for distribution networks is still evolving and is intrinsically more challenging owing to the networks' unique characteristics and data limitations [55].

A specific challenge concerns measurement density and network observability. Transmission networks are instrumented with a high density of measurements, which normally guarantees full observability. Distribution networks, however, are monitored far more sparsely. In GB, for example, most of the distribution system operates without real-time monitoring: although voltage, current and power are commonly measured at primary substations, secondary substations seldom provide comparable data.

Another challenge relates to methodological limitations in the treatment of incomplete data. At transmission level, insufficient measurements are commonly supplemented by pseudo-measurements, for instance, using zero-injection constraints or bounded inequalities, which can be assimilated within weighted least-squares. Transferring this stratagem to distribution systems is difficult: the pronounced heterogeneity of loads, together with the scarcity of historical data, undermines the statistical representativeness and reliability of such pseudo-measurements, thereby degrading estimator performance [49].

A further challenge pertains to load modelling and parameter variability. Conventional transmission state estimation studies approximate aggregated demand as constant impedance and power elements and treat network parameters as limited variant [56]. These assumptions are untenable in distribution contexts, where thermal temperatures fluctuate appreciably that change the network parameters [57], and consist of single-phase load connections resulting in inherently unbalanced loading [58]. Traditional state-estimation algorithms, predicated on a balanced three-phase model, consequently, yield biased or misleading results unless the underlying unbalance and parameter variability are explicitly accommodated.

Distribution System State Estimation (DSSE) has become indispensable for the practical operation of distribution networks. Research in this area began only in the late 1990s, when [59] introduced a three-phase current-based estimator that minimised the WLS objective while preserving a constant gain matrix. Shortly afterwards, [60] presented a WLS-based DSSE incorporating three-

phase modelling techniques; their formulation computes deviations in bus voltages and power flows, thereby allowing the uncertainty of the estimated states to be quantified. Subsequent work has largely refined the maximum-likelihood framework inherent in WLS [49], making use of pseudo-measurements for more realistic load representation [61] and adopting phase-wise estimation to address the unbalanced loading characteristic of distribution networks [62].

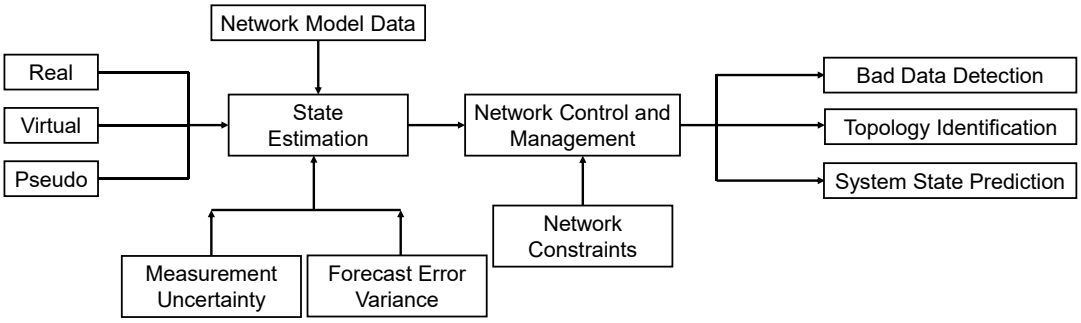

Figure 2.20 Flowchart of distribution system state estimation

Figure 2.20 shows schematic of DSSE architecture. The input data required for state estimation is more extensive than that for conventional power flow analysis and includes the following components:

1. Measurement Inputs: Three types of measurement inputs are used: real, virtual, and pseudo measurements.

    - Real measurements refer to telemetered values such as voltage magnitude, phase angle, or real and reactive power injections. Their accuracy depends on the precision of the measurement equipment used.

    - Virtual measurements represent zero-injection buses in the network and are typically considered highly accurate due to their deterministic nature.

    - Pseudo measurements are estimated or forecasted load values, derived from historical load profile data or predictive models based on contextual information such as weather conditions, consumer types, the number of consumers, and calendar-based demand patterns.

2. Network model: As discussed in Section 2.4, the network model is a fundamental tool for simulating the electrical behaviour of the system and forms the structural basis for state estimation.

3. Measurement and forecast variance: In distribution system state estimation, it is crucial to incorporate knowledge of the variances associated with each type of measurement. The variance of real measurements is determined by the accuracy of the corresponding sensors, while pseudo measurements typically exhibit higher uncertainty due to modelling and forecasting errors. Consequently, pseudo measurements are generally assigned higher variance values in the estimation process.

The objective of distribution system state estimation is to obtain the most accurate estimate of the network states based on the available information. These estimated states serve as a foundation for the control and scheduling block to execute a range of network functions that support operational decision-making. There are three key tasks involved:

- Bad data detection: This process identifies and filters out erroneous measurements using methods such as residual analysis. And diagnose the underlying causes of data anomalies.
- Topology identification: This ensures that the assumed network configuration reflects the actual system conditions by verifying breaker statuses and network connectivity.
- System state prediction: This task involves forecasting future operating conditions to facilitate proactive control strategies and long-term planning.

Traditionally, state estimators tackle these tasks in a static fashion: each time a new batch of measurements is processed, the previously estimated state is ignored. In reality, however, the power system is never truly in steady-state operation because of stochastic variations in demand and generation [63]. This situation is further aggravated by the large-scale integration of DERs on the generation side, and by increasingly new demand-response technologies like EVs.

2.5.2 Dynamic State Estimation in Distribution Networks

To address the limitations of static state estimation, a more effective approach is Dynamic State Estimation (DSE), also known as Forecasting-Aided State Estimation (FASE). This method establishes a temporal relationship among system states at successive time instances for both

present and future [64]. Consequently, it enables the use of mathematical models to generate predictions of system states and measurements, offering prior information that complements the incoming real-time data. By integrating this prior information with the newly acquired measurements, DSE significantly enhances the accuracy and robustness of state estimation [65].

In FASE, forecasts are usually produced either for bus voltages (voltage magnitudes and phase angles) or for bus power injections (active and reactive loads and generation). Selecting voltages is attractive because they already constitute the state vector in conventional static state estimation and, once voltages have been forecast, every other network variable can be derived directly [66]. However, time-series voltage data are highly cross-correlated and sensitive to unreported topology changes, so forecasting them are easily destabilised by configuration errors. Another approach uses power injection, however, the primary challenge associated with this method lies in mapping the uncertainty of predicted bus power injections to the resulting node voltages [67]. Since loads and generation are the fundamental drivers of system dynamics, bus loads typically follow regular patterns that can be forecast with a reasonable degree of accuracy. Furthermore, the statistical interdependencies among loads at different buses are generally weaker than those observed among bus voltages [65].

Accordingly, this thesis adopts a power injection forecast based FASE framework that utilises short-term load forecasts to characterise the network's dynamic behaviour. Building on the static state estimation formulation in Equation (2.13) to Equation (2.22), this thesis extends the state and measurement models into a forecast-aided form. Specifically, the state vector used in conventional WLS state estimation, x in Equation (2.13), is generalised as a time-indexed state to enable short-term prediction. To make this distinction explicit, the forecast-aided state is denoted by y_k, where:

- y_k contains the same physical variables as x, with bus voltage magnitudes and phase angles at all non-slack buses but indexed at time k.
- Thus, y_k can be viewed as the dynamic extension of the static state in Equation 2.13.

Under quasi-steady-state conditions, the forecast-aided state evolution is approximated by a linear transition model:

$$y_{k+1} = F_k y_k + g_k + w_k \tag{2.23}$$

Equation (2.23) is the FASE-specific state-transition model. Unlike full DSE, where the state transition is derived from the complete nonlinear differential-algebraic system dynamics, the FASE approximation neglects fast electromechanical transients over the short forecast horizon. This allows the state transition to be represented in linear form, where F_k is identified from historical operating data, g_k captures slowly varying trends such as gradual demand changes, and w_k represents modelling uncertainty.

$$z_k = h_k(y_k) + e_k \tag{2.24}$$

Equation (2.24) is the time-indexed counterpart of the conventional nonlinear measurement model in Equation (2.14). The key difference is that both the measurements and the measurement functions are now indexed by k. This notation emphasises that the measurement set and the network-dependent load-flow functions may vary over time due to operating-point changes, forecast updates, or configuration changes. Here, z_k is the measurement vector at time k, $h_k(\cdot)$ is the load-flow-based measurement function consistent with the network configuration at time k, and$e_k$is the corresponding measurement error vector.

In summary, Equation (2.23) and Equation (2.24) are introduced by extending the static state estimation framework in Equation (2.13) to Equation (2.16) and the measurement model in Equation (2.14) into a discrete-time forecast-aided form. Equation (2.23) provides the FASE state prediction step, while Equation (2.24) provides the time-aligned measurement model used for estimation and correction.

When load or generation forecasts are available, the state-transition model is augmented to incorporate a forecast input:

$$y_{k+1} = F_k y_k + B_k \hat{u}_{k+1} + w_k \tag{2.25}$$

where $\hat{u}_{k+1}$ is the forecasted power-injection vector, B_k is the sensitivity matrix that maps these injections to their influence on the voltage states. This enables the estimator to anticipate forthcoming operating conditions and thereby improves the accuracy and robustness of the state predictions while retaining the computational efficiency of a linear model.

In practical distribution system applications, DSE is often integrated with auxiliary algorithms to address real-world challenges. For instance, Kalman filters are commonly employed to mitigate

impulsive communication noise and counteract malicious cyber intrusions, which can introduce significant bias into incoming measurements [68].

2.6 Conclusion

Chapter 2 traces the historical evolution of GB distribution network and shows recent developments in its structure and operation. It highlights the growing need for enhanced observability and control of the distribution network, particularly in light of the increasing penetration of DERs and EVs, both of which can pose significant challenges to network stability.

The chapter also highlights the intrinsic challenges of forecasting demand on LV networks, whose distinctive characteristics render them markedly more complex than HV network. Because LV feeders serve only a small cohort of customers, aggregate demand exhibits greater diversity: the consumption patterns of individual users exert a proportionally larger influence on the overall load. In addition, the measurement infrastructure in LV networks is frequently sparse, outdated, incomplete, limiting both the availability and the quality of data. Collectively, these factors severely constrain the effectiveness of conventional time-series forecasting techniques in the LV context. To address this, the chapter 4 proposes the use of cardinal point forecasting, a method focused on predicting daily extreme values and key transitional points, which are comparatively more stable and hence more forecastable than full time series.

The chapter 3 contends that, similar to TSOs, distribution networks need also be more actively controlled. It demonstrates the fundamentals of how to model the modern distribution network and explores the attendant difficulties of conducting power flow analysis difficulties that arise principally from the network's limited observability. This discussion culminates in an examination of DSSE, a pivotal process that enables DNOs to comprehend both the instantaneous operating state and the anticipated behaviour of their systems.

To address the observability challenge, two broad strategies are available. The first approach is installing monitoring equipment at every node, which is both impracticable and prohibitively expensive. Consequently, the remainder of this thesis concentrates on the pragmatic implementation of DSSE under conditions of data sparsity, adopting data-driven approaches to enhanced network observability. It investigates the use of advanced machine learning techniques like few-shot learning and transfer learning to enhance state estimation performance in contexts

where measurement data are scarce. Both time-series and cardinal-point forecasting ways are evaluated within these frameworks.

The methodology proposed in subsequent chapters of this thesis aims to provide DNOs with a practical, scalable approach to resolving the observability gap in DSSE. By enabling accurate and robust state estimation under constrained data conditions, the approach supports the secure and resilient operation of distribution networks throughout GB, while preserving the required standards of estimation accuracy.

2.7 Reference

[1] C. J. Oatley, B. Ramsay, A. McPherson, R. Eastwood, and C. S. Ozveren, "A decision support system for electricity distribution network refurbishment projects," *Electric Power Systems Research,* vol. 40, no. 1, pp. 27-35, 1997.

[2] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of Low Voltage Network Templates—Part I: Substation Clustering and Classification," *IEEE Transactions on Power Systems,* vol. 30, no. 6, pp. 3036-3044, 2015, doi: 10.1109/TPWRS.2014.2371474.

[3] SPEN, "2024 Network Development Plan: SP Distribution Parts 1 & 2 – Main Report," SP Energy Networks, 2024/05 2024. [Online]. Available: https://www.spenergynetworks.co.uk/userfiles/file/SP_Energy_Networks_2024_Network_Development_Plan_-_SP_Distribution_Parts_1and2_-_Main_Report.pdf

[4] S. Haben, S. Arora, G. Giasemidis, M. Voss, and D. Vukadinović Greetham, "Review of low voltage load forecasting: Methods, applications, and recommendations," *Applied Energy,* vol. 304, p. 117798, 2021/12/15/ 2021, doi: https://doi.org/10.1016/j.apenergy.2021.117798.

[5] S. Haben *et al.*, "Mathematical solutions for electricity networks in a low carbon future," in *22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013)*, 10-13 June 213 2013, pp. 1-4, doi: 10.1049/cp.2013.0972.

[6] J. Coignard, P. MacDougall, F. Stadtmueller, and E. Vrettos, "Will Electric Vehicles Drive Distribution Grid Upgrades?: The Case of California," *IEEE Electrification Magazine,* vol. 7, no. 2, pp. 46-56, 2019, doi: 10.1109/MELE.2019.2908794.

[7] P. Papadopoulos, S. Skarvelis-Kazakos, I. Grau, L. M. Cipcigan, and N. Jenkins, "Electric vehicles' impact on British distribution networks," *IET Electrical Systems in Transportation,* vol. 2, no. 3, pp. 91-102, 2012.

[8] T. C. Green, T. Frost, A. Junyent‐Ferré, and P. D. Mitcheson, "Power Electronics for Flexible Distribution Networks," *Smart Grid Handbook,* pp. 1-19, 2016.

[9] H. Sadeghian and Z. Wang, "A novel impact-assessment framework for distributed PV installations in low-voltage secondary networks," *Renewable Energy,* vol. 147, pp. 2179-2194, 2020/03/01/ 2020, doi: https://doi.org/10.1016/j.renene.2019.09.117.

[10] Ofgem, "Electricity distribution price control review final proposals – allowed revenue – cost assessment appendix," Ofgem, December 2009.

[11] J. Stewart, J. Taylor, A. Smith, and J. Deacons, "Review of WPD Unit Costs," Parsons Brinckerhoff, United Kingdom, 2013/06/06.

[12] G. Cheng, Y. Lin, A. Abur, A. Gómez-Expósito, and W. Wu, "A Survey of Power System State Estimation Using Multiple Data Sources: PMUs, SCADA, AMI, and Beyond," *IEEE Transactions on Smart Grid,* vol. 15, no. 1, pp. 1129-1151, 2024, doi: 10.1109/TSG.2023.3286401.

[13] C. Shand, G. Taylor, E. Stewart, C. Roberts, A. McMorran, and P. Mohapatra, "Improving actionable observability of large distribution networks for transmission operators to support improved system control, fault detection and mitigation," *CIRED 24,* vol. 2017, no. 1, pp. 1215-1218, 2017.

[14] S. Almas, L. Vanfretti, S. Løvlund, and J. O. Gjerde, "Open source SCADA implementation and PMU integration for power system monitoring and control applications," in *2014 IEEE PES General Meeting | Conference & Exposition*, 27-31 July 2014 2014, pp. 1-5, doi: 10.1109/PESGM.2014.6938840.

[15] A. Pal, A. K. S. Vullikanti, and S. S. Ravi, "A PMU Placement Scheme Considering Realistic Costs and Modern Trends in Relaying," *IEEE Transactions on Power Systems,* vol. 32, no. 1, pp. 552-561, 2017, doi: 10.1109/TPWRS.2016.2551320.

[16] S. Phuyal, D. Bista, J. Izykowski, and R. Bista, "Design and implementation of cost efficient SCADA system for industrial automation," *International Journal of Engineering and Manufacturing,* vol. 10, no. 2, p. 15, 2020.

[17] SPEN, "Enhanced Substation Monitoring Deployment," SP Energy Networks, 2015/09 2015. [Online]. Available: https://www.spenergynetworks.co.uk/userfiles/file/EnhancedSubstationMonitoringDeployment07.pdf

[18] SSEN, "Electric vehicle ownership supported by SSEN through low-cost, data-gathering substation monitors," 2019/06/25 2019. [Online]. Available: https://www.ssen.co.uk/news-views/2019/2019-electric-vehicle-ownership-supported-by-ssen-through-low-cost-data-gathering-substation-monitors/.

[19] Eneida.io, "EdgeSense: Next Generation Smart Sensor Technology for the LV Grid," 2025/04/14 2025. [Online]. Available: https://eneida.io/edgesense/.

[20] ElectricityNorthWest, "Low Voltage Network Solutions (LVNS) Project: Appendix 4," Electricity North West Limited, 2015/12 2015. [Online]. Available: https://www.enwl.co.uk/globalassets/innovation/lvns/lvns-closedown/closedown-appendices/lvns-appendix-4.pdf

[21] SSEN, "SSEN Network Visibility Strategy," SSEN, 2024/02 2024. [Online]. Available: https://www.ssen.co.uk/globalassets/about-us/stakeholder-engagement/annual-engagement-plans-and-reports/ssen-network-visibility-strategy-feb-2024.pdf

[22] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *International Journal of Forecasting,* vol. 32, no. 3, pp. 914-938, 2016.

[23] S. Haben, G. Giasemidis, F. Ziel, and S. Arora, "Short term load forecasting and the effect of temperature at the low voltage level," *International Journal of Forecasting,* vol. 35, no. 4, pp. 1469-1484, 2019/10/01/ 2019, doi: https://doi.org/10.1016/j.ijforecast.2018.10.007.

[24] D. National Grid Electricity. *Aggregated Smart Meter Data – Secondary Substation*, National Grid Electricity Distribution, 2023/07/01. [Online]. Available: https://connecteddata.nationalgrid.co.uk/dataset/aggregated-smart-meter-data-secondary-substation

[25] F. Taylor, "Lights, appliances and smart technologies: Final report," Department for Business, Energy and Industrial Strategy, 2021. [Online]. Available: https://assets.publishing.service.gov.uk/media/61449c3cd3bf7f05b2ac20a4/efus-light-appliances-smart-tech.pdf

[26] K. Gram-Hanssen, "Understanding change and continuity in residential energy consumption," *Journal of Consumer Culture,* vol. 11, no. 1, pp. 61-78, 2011/03/01 2011, doi: 10.1177/1469540510391725.

[27] S. Dutta *et al.*, "Load and renewable energy forecasting for a microgrid using persistence technique," *Energy Procedia,* vol. 143, pp. 617-622, 2017.

[28] P. Lusis, K. R. Khalilpour, L. Andrew, and A. Liebman, "Short-term residential load forecasting: Impact of calendar effects and forecast granularity," *Applied energy,* vol. 205, pp. 654-669, 2017.

[29] L. Semmelmann, O. Resch, S. Henni, and C. Weinhardt, "Privacy‐preserving peak time forecasting with Learning to Rank XGBoost and extensive feature engineering," *IET Smart Grid,* vol. 7, no. 2, pp. 172-185, 2024.

[30] J. Yang, C.-Q. Ou, Y. Ding, Y.-X. Zhou, and P.-Y. Chen, "Daily temperature and mortality: a study of distributed lag non-linear effect and effect modification in Guangzhou," *Environmental Health,* vol. 11, pp. 1-9, 2012.

[31] M. Anwar, M. K. Alam, and S. E. Gleason, "Traction power inverter design for EV and HEV applications at general motors: A review," 2019: IEEE, pp. 6346-6351.

[32] B. Drysdale, J. Wu, and N. Jenkins, "Flexible demand in the GB domestic electricity sector in 2030," *Applied Energy,* vol. 139, pp. 281-290, 2015.

[33] J.-P. Zimmermann *et al.*, "Household Electricity Survey: A study of domestic electrical product usage," *Intertek Testing & Certification Ltd,* pp. 213-214, 2012.

[34] P. Rodríguez, M. A. Bautista, J. Gonzàlez, and S. Escalera, "Beyond one-hot encoding: Lower dimensional target embedding," *Image and Vision Computing,* vol. 75, pp. 21-31, 2018/07/01/ 2018, doi: https://doi.org/10.1016/j.imavis.2018.04.004.

[35] R. Li, F. Li, and N. D. Smith, "Multi-Resolution Load Profile Clustering for Smart Metering Data," *IEEE Transactions on Power Systems,* vol. 31, no. 6, pp. 4473-4482, 2016, doi: 10.1109/TPWRS.2016.2536781.

[36] F. McLoughlin, A. Duffy, and M. Conlon, "A clustering approach to domestic electricity load profile characterisation using smart metering data," *Applied Energy,* vol. 141, pp. 190-199, 2015/03/01/ 2015, doi: https://doi.org/10.1016/j.apenergy.2014.12.039.

[37] W. H. Kersting, "Distribution system modeling and analysis," in *Electric power generation, transmission, and distribution*: CRC press, 2018, pp. 26-1.

[38] W. H. Kersting, "Radial distribution test feeders," presented at the IEEE Winter Meeting Power Engineering Society, 2001.

[39] C. Eland, "Eland Cables Power Networks Catalogue," Eland Cables, 2023/06 2023. [Online]. Available: https://www.elandcables.com/media/tbmh23s1/eland-cables-power-networks-catalogue.pdf

[40] *UKGDS: United Kingdom Generic Distribution System*. (2025). Smart Energy Research Lab, Imperial College London. [Online]. Available: https://github.com/sedg/ukgds

[41] R. Singh, "State estimation in power distribution network operation," 2009.

[42] G. Artale *et al.*, "Real-Time Power Flow Monitoring and Control System for Microgrids Integration in Islanded Scenarios," *IEEE Transactions on Industry Applications,* vol. 55, no. 6, pp. 7186-7197, 2019, doi: 10.1109/TIA.2019.2932967.

[43] F. Milano, "Continuous Newton's Method for Power Flow Analysis," *IEEE Transactions on Power Systems,* vol. 24, no. 1, pp. 50-57, 2009, doi: 10.1109/TPWRS.2008.2004820.

[44] P. A. Lipka, R. P. O'Neill, and S. Oren, "Developing line current magnitude constraints for IEEE test problems," *rap. tech., Federal Energy Regulatory Commission,* 2013.

[45] S. Bolognani and S. Zampieri, "On the Existence and Linear Approximation of the Power Flow Solution in Power Distribution Networks," *IEEE Transactions on Power Systems,* vol. 31, no. 1, pp. 163-172, 2016, doi: 10.1109/TPWRS.2015.2395452.

[46] W. F. Tinney and C. E. Hart, "Power flow solution by Newton's method," *IEEE Transactions on Power Apparatus and systems,* no. 11, pp. 1449-1460, 1967.

[47] F. C. Schweppe and J. Wildes, "Power System Static-State Estimation, Part I: Exact Model," *IEEE Transactions on Power Apparatus and Systems,* vol. PAS-89, no. 1, pp. 120-125, 1970, doi: 10.1109/TPAS.1970.292678.

[48] M. Meriem, C. Bouchra, B. Abdelaziz, S. O. B. Jamal, E. M. Faissal, and C. Nazha, "Study of state estimation using weighted-least-squares method (WLS)," in *2016 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM)*, 26-28 Oct. 2016 2016, pp. 1-5, doi: 10.1109/CISTEM.2016.8066820.

[49] R. Singh, B. C. Pal, and R. A. Jabr, "Choice of estimator for distribution system state estimation," *IET generation, transmission & distribution,* vol. 3, no. 7, pp. 666-678, 2009.

[50] A. Mukhtar, *Power System State Estimation*. Artech, 2012, p. 1.

[51] R. de Levie, "When, why, and how to use weighted least squares," *Journal of chemical education,* vol. 63, no. 1, p. 10, 1986.

[52] R. A. Jabr and B. C. Pal, "Iteratively reweighted least-squares implementation of the WLAV state-estimation method," *IEE Proceedings - Generation, Transmission and Distribution,* vol. 151, no. 1, pp. 103-108, 2004/01/14 2004, doi: 10.1049/ip-gtd:20040030.

[53] R. A. Jabr, "Power system Huber M-estimation with equality and inequality constraints," *Electric power systems research,* vol. 74, no. 2, pp. 239-246, 2005.

[54] A. Abur and A. G. Exposito, *Power system state estimation: theory and implementation*. CRC press, 2004.

[55] G. Strbac, N. Jenkins, T. Green, and D. Pudjianto, "Review of innovative network concepts," *DG-GRID project report,* 2006.

[56] F. U. Nazir, B. C. Pal, and R. A. Jabr, "Approximate Load Models for Conic OPF Solvers," *IEEE Transactions on Power Systems,* vol. 36, no. 1, pp. 549-552, 2021, doi: 10.1109/TPWRS.2020.3029370.

[57] M. Z. Degefa, M. Koivisto, R. J. Millar, and M. Lehtonen, "Dynamic thermal state forecasting of distribution network components: For enhanced active distribution network capacity," in *2014 International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 7-10 July 2014 2014, pp. 1-6, doi: 10.1109/PMAPS.2014.6960607.

[58] K. Ma, L. Fang, and W. Kong, "Review of distribution network phase unbalance: Scale, causes, consequences, solutions, and future research directions," *CSEE Journal of Power and Energy Systems,* vol. 6, no. 3, pp. 479-488, 2020, doi: 10.17775/CSEEJPES.2019.03280.

[59] C. N. Lu, J. H. Teng, and W. H. E. Liu, "Distribution system state estimation," *IEEE Transactions on Power Systems,* vol. 10, no. 1, pp. 229-240, 1995, doi: 10.1109/59.373946.

[60] L. Ke, "State estimation for power distribution system and measurement impacts," *IEEE Transactions on Power Systems,* vol. 11, no. 2, pp. 911-916, 1996, doi: 10.1109/59.496174.

[61] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac, "Distribution System State Estimation Using an Artificial Neural Network Approach for Pseudo Measurement Modeling," *IEEE*

Transactions on Power Systems, vol. 27, no. 4, pp. 1888-1896, 2012, doi: 10.1109/TPWRS.2012.2187804.

[62] R. Singh, B. C. Pal, and R. A. Jabr, "Distribution system state estimation through Gaussian mixture model of the load as pseudo-measurement," *IET generation, transmission & distribution,* vol. 4, no. 1, pp. 50-59, 2010.

[63] J. Zhao *et al.*, "Power System Dynamic State Estimation: Motivations, Definitions, Methodologies, and Future Work," *IEEE Transactions on Power Systems,* vol. 34, no. 4, pp. 3188-3198, 2019, doi: 10.1109/TPWRS.2019.2894769.

[64] A. Primadianto and C. N. Lu, "A Review on Distribution System State Estimation," *IEEE Transactions on Power Systems,* vol. 32, no. 5, pp. 3875-3883, 2017, doi: 10.1109/TPWRS.2016.2632156.

[65] M. B. D. C. Filho and J. C. S. d. Souza, "Forecasting-Aided State Estimation—Part I: Panorama," *IEEE Transactions on Power Systems,* vol. 24, no. 4, pp. 1667-1677, 2009, doi: 10.1109/TPWRS.2009.2030295.

[66] R. Dobbe, W. v. Westering, S. Liu, D. Arnold, D. Callaway, and C. Tomlin, "Linear Single- and Three-Phase Voltage Forecasting and Bayesian State Estimation With Limited Sensing," *IEEE Transactions on Power Systems,* vol. 35, no. 3, pp. 1674-1683, 2020, doi: 10.1109/TPWRS.2019.2955893.

[67] A. K. Sinha and J. K. Mondal, "Dynamic state estimator using ANN based bus load prediction," *IEEE Transactions on Power Systems,* vol. 14, no. 4, pp. 1219-1225, 1999, doi: 10.1109/59.801876.

[68] E. Ghahremani and I. Kamwa, "Online State Estimation of a Synchronous Generator Using Unscented Kalman Filter From Phasor Measurements Units," *IEEE Transactions on Energy Conversion,* vol. 26, no. 4, pp. 1099-1108, 2011, doi: 10.1109/TEC.2011.2168225.

3 Residual Transfer Learning in Minimally Observed Power Distribution Networks

In recent years, DNOs have increasingly prioritised the deployment of advanced monitoring devices in substations to address significant operational challenges. These devices enable real-time data acquisition and storage, thereby enhancing network visibility and facilitating more effective system management. Nevertheless, the majority of GB distribution network remains unmonitored. In this thesis, substations equipped with advanced monitoring devices are characterised as 'data-rich', whereas those without such devices are 'data-sparse', i.e. lacking real-time data acquisition despite the possibility of limited, post-operational data collection [1].

Given the extensive scale of distribution network asset fleets, installing monitoring equipment in every substation across GB would require substantial time, financial investment, and ongoing maintenance. In many locations, newly monitored data may be insufficient for training a forecasting model, and there may be limited personnel available to maintain devices or process raw data to keep monitored load data continuously usable. Nevertheless, it is imperative for DNOs to understand the network's operating characteristics daily to inform their decisions. To address these challenges, we propose a novel transfer learning approach that leverages insights gained from data-rich areas to guide models deployed in data-sparse regions. Its key advantage lies in expediting the training process, thereby significantly reducing data requirements compared to existing methods and rendering it a more efficient and cost-effective solution. It is particularly beneficial for substations with minimal on-site computing resources. Furthermore, results obtained from dynamic state estimation, which reveal overall network voltage performance, indicate that there is no need to construct complete predictive models for every substation. Instead, these results serve as pseudo-measurements that inform the estimation of the network state at future time points.

The structure of this chapter is as follows: Section 3.1 introduces transfer learning and outlines its application in load forecasting, Section 3.2 describes the day-ahead forecast benchmark model and examines its performance at the distribution level, Section 3.3 presents the novel use of intraday residuals to adjust forecasts for various metering points, Section 3.4 reports the forecast errors observed under both fully monitored and minimally monitored network scenarios, Section 3.5

compares the resulting state estimates with pseudo-measurements derived from both levels of observation, as well as power flow calculations (which represent the ideal case), and Section 3.6 concludes with a discussion of the overall findings and potential avenues for future research.

3.1 Transfer Learning Background

An assumption underlying traditional machine learning methods is that the feature space and data distribution of the training set and the test set are identical. However, in real-world applications, labelled training data are often limited. For example, obtaining data such as LV load data can be difficult, which constrains the development of a machine learning model. In many cases, sufficient training data may be available only in a different domain of interest, where the data might reside in another feature space or follow a different distribution. In such circumstances, transfer learning can be employed to develop a more general model using readily available (source) data from a similar but distinct domain, and then adapt this model to a smaller dataset that specifically represents the target application [2].

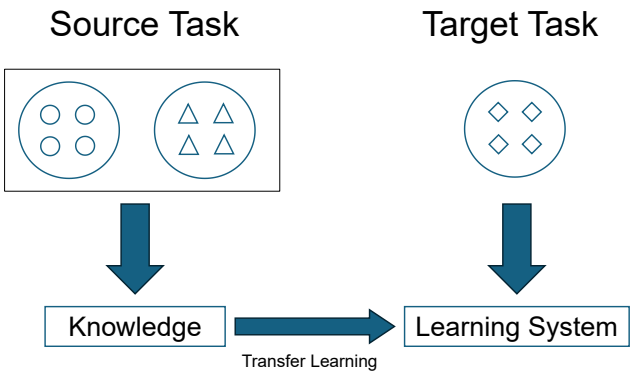

Figure 3.1 General framework of transfer learning

Figure 3.1 illustrates the simplified learning processes involved in transfer learning. The source task is associated with a plentiful dataset, denoted by D_S, which supports successful model training, represented by M_S. By contrast, the target task features a more limited dataset, denoted by D_T.situated in a different domain. Knowledge extracted from the source domain is transferred to the target domain, thereby enabling the learning system to adapt more efficiently to the target task despite fewer training samples. The resulting model for the target task is represented by M_T. Overall, the transfer learning process can be expressed as follows:

$$M_T = Transfer(M_S, D_T) \tag{3.1}$$

The field of transfer learning recognises four main methods: transductive, inductive, unsupervised, and negative transfer learning. In addition, it can be categorised into four types of learning: learning on instances, learning on features, learning on parameters, and learning on relations [3]. Transductive transfer learning applies when the tasks in the source and target domains are the same, but the domains themselves may differ. In this scenario, algorithms transfer knowledge from the source to the target domain by re-weighting or re-sampling source instances [4]. Conversely, inductive transfer learning handles scenarios where the source and target tasks differ. Because source domain data may be outdated and new data costly, it aims to build an accurate target model with minimal well-labelled data. A key step is identifying which portions of the source data remain suitable for adaptation [5]. In unsupervised transfer learning, there are no labelled data in either the source or the target domain. Such methods are particularly beneficial for niche or specialised tasks, for which sufficient labelled data are unavailable. A prominent application is clustering, Self-taught Clustering (STC) [6] aims to cluster a small set of unlabelled target data using a larger pool of unlabelled data from the source domain. Finally, negative transfer learning arises when an excessive amount of unrelated knowledge is transferred from the source domains. Despite its prevalence, negative transfer is often characterised in an informal manner, lacking a rigorous definition, careful analysis, and systematic treatment. [7].

In the LV network load forecasting tasks associated with this paper, transfer learning is mainly based on the Inductive Transfer Learning approach. The load data from data-rich areas are available, while information from data-sparse areas is difficult to obtain, limited in coverage, or instantaneous only. The objective is to begin modelling in areas where the data is known and then transfer and adapt that model to areas where there is less abundant data. In data-rich areas, there are multiple models available for day ahead forecasts as benchmarks. However, whether using mathematical or machine learning approaches, both require substantial data, which is impossible to collect in data-sparse areas.

3.2 Day Ahead Load Forecast Baseline Model

In the context of load forecasting, while forecasts are inherently subject to error, some can still provide significant value to utility companies. For forecasts to be considered useful, they must as

a minimum satisfy the operational and planning requirements of the utility. Furthermore, they should be accurate, explainable, and defensible but should not be too focused on accuracy, as minor improvement of forecasting accuracy beyond certain level may not result in significant economic benefits [8]. At the transmission level, day-ahead forecasts are typically expected to fall within an acceptable range of accuracy, often referred to as a "comfort zone," which generally corresponds to a MAPE between 2% and 6% for hourly load predictions for the following day [9], In contrast, distribution networks may tolerate larger errors, potentially up to 15% with the most LV forecast results reported in [10]. In particular, reliable load forecasts can support voltage regulation by helping operators maintain voltage profiles within statutory limits and enable more effective thermal control to avoid exceeding equipment temperature ratings and prolong asset life.

To undertake short-term load forecasting for LV networks, statistical approaches are frequently employed. These methodologies can generally be categorised into three principal types. The first is multiple linear regression [11], which posits that the relationship between explanatory variables and the dependent variable can be adequately represented within a linear framework. The second category comprises nonlinear modelling techniques, such as polynomial regression which is capable of modelling more complex patterns, including generalised additive models (GAMs) [12], which allow each explanatory variable to influence the dependent variable through smooth, nonlinear functions. The third category encompasses traditional time series methods, such as the autoregressive integrated moving average (ARIMA) model, which utilises historical values and their autocorrelations to forecast future trends [13].

Conversely, with the rapid advancement of machine learning, a variety of new techniques have also been applied to LV load forecasting. Artificial neural networks (ANNs) are among the most widely adopted due to their ability to capture intricate, nonlinear relationships between input features and the target variable [14]. Moreover, long short-term memory (LSTM) networks are becoming increasingly prevalent in time series forecasting tasks, such as wind or load forecasting [15]. These models employ several gating mechanisms that enable the retention of long-term dependencies, thereby expanding the number of parameters that may be learned from the training data and enhancing their forecasting capability.

Whilst such advanced methods can substantially improve predictive performance, it is important to recognise that increasing model complexity often brings a heightened risk of overfitting,

particularly when only limited data are available. Both statistical and machine learning approaches typically require substantial quantities of high-quality data to ensure effective model training and robust performance. In practice, state-of-the-art LV load forecasting generally relies on access to highly detailed information and comprehensive historical records. However, such data are frequently unavailable in real-world LV networks.

Given these practical constraints, the primary focus of the present study is not to maximise the absolute accuracy of the LV load forecasting model. Rather, the forecasting model is employed as a baseline to facilitate the evaluation of transfer learning techniques. To this end, two models are selected: the Hong Naïve model [9] and an LSTM Recurrent Neural Network [15]. An LSTM is selected as a widely used non-linear benchmark in short-term load forecasting, providing a common and well-established point of comparison against the Hong Naïve baseline. These models are used to establish benchmark performance in LV load forecasting. Furthermore, the Hong Naïve model also serves as a reference point for assessing the outcomes of the proposed transfer learning approach.

The first benchmark is Hong Naïve model, it is widely used as a standard model for the day ahead load forecast benchmark model is used here, which is Hong Naïve model [9]:

$$\begin{aligned} L(load) = {} & \beta_0 + \beta_1 \times Trend + \beta_2 \times Day \times Hour + \beta_3 \times Month + \beta_4 \times Month \times TMP \\ & + \beta_5 \times Month \times TMP^2 + \beta_6 \times Month \times TMP^3 + \beta_7 \times Month \times TMP \\ & + \beta_7 \times Hour \times TMP + \beta_8 \times Hour \times TMP^2 + \beta_9 \times Hour \times TMP^3 \end{aligned} \quad (3.2)$$

In the model, $\beta_0, \beta_1 \cdots \beta_9$ denote the regression coefficients, each quantifying the influence of their respective explanatory variables on the dependent variable, 'Trend' refers to the sequential hourly trend in the data, represented by ascending natural numbers to account for the temporal progression. The variables ‘Hour’, ‘Day’, and ‘Month’ respectively represent the hour of the day (ranging from 1 to 24), the day of the week (ranging from 1 to 7), and the month of the year (ranging from 1 to 12). ‘TMP’ indicates the local ambient temperature, measured in degrees Celsius (°C).

The model captures both linear and nonlinear interactions between the calendar factors (such as ‘Month’ and ‘Hour’) and temperature, by including cross-product terms as well as polynomial

terms in temperature like TMP^2 and TMP^3 . Such construction allows the model to flexibly accommodate complex seasonal and meteorological effects on load demand.

The rationale for employing this model is that it relies exclusively on temperature and calendar variables for load forecasting, intentionally omitting historical load data. This approach is particularly relevant in contexts where access to past load information may be limited or restricted due to regulatory requirements or privacy concerns. In addition, excluding historical loads helps to preserve the interpretability of the model by reducing the complexity that can arise from incorporating autoregressive elements. As a result, this modelling strategy is especially well-suited to transfer learning scenarios. It draws solely upon exogenous variables, which are generally available and consistent across both source and target domains. These characteristics enhance the generalisability and practical applicability of the model across a variety of settings.

Another test model employed for LV load forecasting is the Long Short-Term Memory (LSTM) network, which fundamentally differs from traditional feedforward neural networks. LSTM, as a type of recurrent neural network (RNN), can capture temporal correlations between past information and current conditions. In the context of time series problems, this means that the decision made by an RNN at time step $t-1$ can influence the outcome at a subsequent time step t. In this study, a basic LSTM model is utilised to assess whether the Hong native model is sufficiently robust to serve as a benchmark for the LV forecasting task. This approach is widely adopted in forecasting applications involving LCTs, such as PV systems, wind farms generate, and load forecast.

3.3 Bayesian Transfer Learning Method

A Bayesian transfer learning approach is adopted in this study, as Bayesian learning is capable of incorporating prior knowledge to reduce the decline in generalisation performance that arises from limited data [16]. Within this Bayesian framework, transfer learning is specifically applied to the residuals of the Hong Native model. Direct application of a model trained in one domain to another often results in poor generalisation. However, when the domains are related, appropriate transfer learning and domain adaptation techniques can exploit shared information across domains, thus enabling the development of more robust and generalisable models in the target domain.

3.3.1 Bayesian Transfer Learning Flowchart

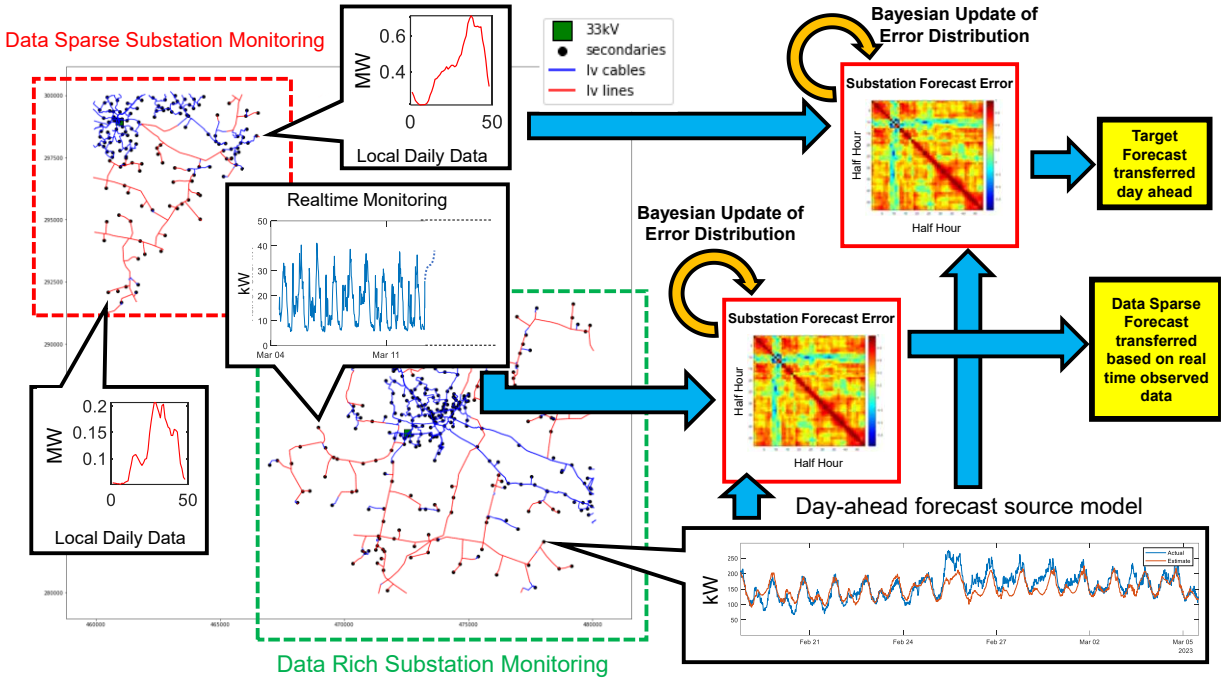

Figure 3.2 Bayesian transfer learning with online estimation of forecast errors for day-ahead load forecast models

Figure 3.2 illustrates the process of Bayesian transfer learning for day-ahead load forecasting in data-sparse substations, leveraging models trained in data-rich substations through an online Bayesian estimation of the forecast error distribution. The procedure begins with the initial collection and preprocessing of data from data-rich areas. A day-ahead forecast baseline model is then developed using this data. Next, errors are generated by applying the baseline model which is trained on data-rich areas to input data from data-sparse areas, using corresponding weather and calendar information. These errors serve as priors for the Bayesian model. Two scenarios are considered: in the absence of online monitoring in data-rich areas, an online Bayesian error updating model is employed to estimate the day-ahead posterior error distribution based on historical data. Conversely, when real-time monitoring is available in the data-rich area, a conditional multivariate Gaussian model is used to further refine the forecasts.

The final prediction for the data-sparse substation is obtained by integrating the posterior error estimate from the transfer learning model into the original baseline forecast, thereby enhancing the overall forecasting accuracy.

3.3.2 Online Bayesian Transfer Learning Method in Load Forecast

Although load profiles differ across locations, they may exhibit similar underlying seasonality and weather-driven patterns. Thus, residuals from the Hong baseline model in the source domain can be utilised to inform transfer learning strategies aimed at improving load profile accuracy in the target domain [17]. In this approach, the residuals are transferred from the source domain to the target domain through a joint prior multivariate normal probability density function. Based on the updating errors, the posterior distribution is assumed to follow a Normal-Wishart distribution, which serves as the conjugate prior of a multivariate normal distribution [18] and can be continuously updated as new observations become available.

A dynamic estimate of the forecast error at a target substation is thereby derived, enabling real-time adjustment of forecast outputs generated for the source substation. Let the observed load measurement for a substation be denoted as L and the predicted load based on the baseline model for the same substation as $\hat{L}$. The actual load for the source substation is L_S, whilst the actual load for the target substation is L_T. For a given substation, the day-ahead load forecast residual is defined as the difference between the actual and predicted loads. Thus, for the source substation, the forecast residual e_S is given by:

$$e_S = L_S - \hat{L}_S \tag{3.3}$$

Similarly, for the target substation, the forecast residual e_T is:

$$e_T = L_T - \hat{L}_T \tag{3.4}$$

Here, $\hat{L}_S$ denotes the forecast generated by the model trained using the Hong Naïve model on source data, and $\hat{L}_T$ represents the forecast for the target substation produced by applying the same trained model. In both cases, the identical baseline model, fitted on data from the source domain, is utilised.

However, each substation is subject to distinct local weather conditions and varying load magnitudes. Such differences can result in substantial forecast errors when applying a single fitted model across different substations. To address this, it is assumed that the day-ahead forecast residuals for both the source and target substations are drawn from a multivariate normal distribution with mean vector m and covariance matrix E [19]:

$$e_T \sim N(m, E) \tag{3.5}$$

To estimate substation forecast residual values e_T, a novel transfer learning framework is introduced, utilising a running Bayesian estimate of the error. Within this framework, it assumes the prior e_T follows multivariate normal distribution. Upon the arrival of new data, Bayesian inference is employed to update the posterior distributions of both the mean and the precision (the inverse of the covariance matrix) of the residuals. This is accomplished by specifying a conjugate Normal-Wishart prior for these parameters, thereby guaranteeing that the posterior distribution retains the Normal-Wishart form upon the incorporation of observed data [20].

Equation 3.6 expresses the conditional posterior distribution of the mean vector μ of the residuals, given the updated precision matrix Λ and observed data.

$$p(\mu|\Lambda, e_T) = N(\mu; \mu_N, \theta_N\Lambda^{-1}) \tag{3.6}$$

Here, μ_N denotes the posterior mean, and $\theta_N\Lambda$ represents the associated covariance matrix.

Equation (3.7) describes the posterior distribution of the precision matrix Λ, which follows a Normal Wishart distribution:

$$p(\Lambda|e_T) = W(\Lambda; \alpha_N, B_N) \tag{3.7}$$

In this expression, α_N and B_N are the updated parameters based on the observed data.

For a day-ahead forecast with half-hourly resolution, which produces a 48-dimensional forecast vector, the priors are specified under the assumption that the error values are initially unknown. The initial parameters are defined as follows:

1. Prior Mean: $\mu_0 = 0$ (This represents a 48-dimensional zero vector.)
2. Prior Precision of the Mean: $\theta_0 = 1$ (A scalar, applied uniformly across all 48 dimensions.)
3. Prior Precision: $\Lambda_0 = I_{48}$ (This represents 48 × 48 identity matrix.)

4. Wishart Distribution Parameters: $\alpha_0 = 49$; $B_0 = I_{48}$ (The identity matrix.)

Upon observing data, the posterior mean of the error vector is updated as follows:

$$\mu_N = \frac{\Lambda_0 \mu_0 + N\overline{e_T}}{\Lambda_N} \tag{3.8}$$

where N denotes the number of observed daily load forecast error vectors, and $\overline{e_T}$ is their empirical mean. The precision parameter is updated according to:

$$\theta_{N+1} = \theta_N + 1 \tag{3.9}$$

As indicated in Equation (3.9), each time a new observation is made, the expected value of the error distribution, μ_N, is updated to reflect the influence of both prior information and the newly observed data. This iterative updating process continually refines the estimate of the expected error, thereby improving the model's accuracy as more data become available.

The updated precision matrix is given by:

$$\Lambda_N = [\frac{1}{\theta_N(\alpha_N - 1)}]B_N \tag{3.10}$$

The updating process involves a sequence of steps, where each parameter is computed based on the most recent observations.

$$\alpha_{N+1} = \alpha_N + \frac{1}{2} \tag{3.11}$$

Each additional observation adjusts the shape parameter α_N by one half-unit, reflecting the cumulative weight of evidence.

$$B_N = B_0 + \frac{N}{2}[\bar{\theta} + \frac{\Lambda_0}{\Lambda_N}(\overline{e_T} - \mu_0)(\overline{e_T} - \mu_0)^T] \tag{3.12}$$

The matrix B_N is recalculated incorporates both the empirical covariance of the observed errors and the deviation of the empirical mean from the prior mean. Thus, assimilating prior knowledge and observed data in a statistically coherent manner.

And the empirical covariance $\bar{\theta}$ is defined as:

$$\bar{\theta} = \frac{1}{N}\sum_{n=1}^{N}(e_N - \overline{e_T})\,(e_N - \overline{e_T})^T \tag{3.13}$$

Equations (3.10) to (3.13) collectively constitute a conjugate Bayesian updating scheme. At each assimilation step, the posterior hyperparameters are recalibrated based on newly observed data. This recursive mechanism enables the transfer learning framework to adaptively respond to streaming data, thereby enhancing the accuracy and reliability of day-ahead load forecasts. By linking the empirical mean and covariance of the observations to the Bayesian prior, the model incrementally refines its posterior mean and precision matrix, allowing for dynamic adaptation as additional data become available.

3.3.3 Conditional Bayesian Transfer Learning Method

After generating the posterior mean of the error and adding it to the original forecast to obtain the final predictions, further model calibration can be performed using the true observations. If only a limited number of observations can be collected at the target substation each day, for example, by transmitting basic information through mobile communication networks using SIM (Subscriber Identity Module) cards, using minimal bandwidth and without the need for data storage, these sparse observations can be used to refine subsequent predictions. In this context, it is assumed that both the data and the errors follow a normal distribution, characterised by a Gaussian distribution. Building upon these observations, an improved approach is proposed that utilises a conditionally Gaussian-distributed residual. For the joint residual distribution across an entire day, it is assumed to follow a 48-dimensional multivariate Gaussian distribution, corresponding to the 48 half-hour intervals in a day. This distribution is represented by the matrix x, with the mean vector denoted by μ, the covariance matrix by Σ. The multivariate Gaussian distribution can be expressed as:

$$f(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{n}{2}}} exp[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)] \tag{3.14}$$

This distribution may change based on specific conditions or observed data. Let x_h denote the actual observed error data, which can be a single half-hourly value or a matrix containing several half-hourly values. Here, h represents time interval during which the data is observed, and μ_h is the corresponding vector of mean values of the history data. Let μ_t denote the mean vector, and Σ_t the covariance for a different time interval t of predicted residual on the same day. Let $\Sigma_{t,h}$

denote the covariance between intervals t and h. Adopting this approach offers a considered model that can refine residual forecast by estimating the discrepancies between observed and model-predicted values, as implied by the intra-day dependency structure.

Consider the joint residual vector as:

$$e = \begin{pmatrix} e_t \\ e_h \end{pmatrix} \tag{3.15}$$

Where e_h is the observed residual at interval *h,* and e_t is the unobserved residual to be predicted. All subsequent analyses proceed by conditioning on e_h and deriving the predictive mean $\mu_{t|h}$ and covariance $\Sigma_{t|h}$ for e_t.

Thus, the corresponding partition of the mean vector μ is given by:

$$\mu = \begin{pmatrix} \mu_t \\ \mu_h \end{pmatrix} \tag{3.16}$$

And for the covariance matrix Σ:

$$\Sigma = \begin{pmatrix} \Sigma_{t,t} & \Sigma_{t,h} \\ \Sigma_{h,t} & \Sigma_{h,h} \end{pmatrix} \tag{3.17}$$

Similarly, the precision matrix (the inverse of the covariance matrix) is:

$$\Lambda = \Sigma^{-1} = \begin{pmatrix} \Lambda_{t,t} & \Lambda_{t,h} \\ \Lambda_{h,t} & \Lambda_{h,h} \end{pmatrix} \tag{3.18}$$

For a multivariate Normal distribution, the log-density depends on the quadratic form

$$-\frac{1}{2}\Delta^2 = -\frac{1}{2}(e-\mu)^T\Sigma^{-1}(e-\mu) \tag{3.19}$$

The quantity Δ is called the Mahalanobis distance from μ to e and reduces to the Euclidean distance when Σ is the identity matrix. The quadratic form Δ^2 represents the scaled squared distance between e and μ taking into account the covariance structure of the data [18]. The conditional Gaussian is derived in this manner because the quadratic form naturally encapsulates the way in which the Gaussian distribution measures the similarity between an observation and the mean, scaled by the covariance structure of the data. This enables an accurate characterisation of dependencies and conditional relationships within the distribution.

Expanding the quadratic form, which exhibits the characteristic bell-shaped deviation from the mean:

$$-\frac{1}{2}(e-\mu)^T\Sigma^{-1}(e-\mu) = (e_t-\mu_t)^T\Lambda_{t,t}(e_t-\mu_t) + (e_t-\mu_t)^T\Lambda_{t,h}(e_h-\mu_h) \\ +(e_h-\mu_h)^T\Lambda_{h,t}(e_t-\mu_t) + (e_h-\mu_h)^T\Lambda_{h,h}(e_h-\mu_h) \tag{3.20}$$

To complete the square, consider the exponent in a general Gaussian distribution

$$-\frac{1}{2}(e-\mu)^T\Sigma^{-1}(e-\mu) = -\frac{1}{2}e^T\Sigma^{-1}e + e^T\Sigma^{-1}\mu + const \tag{3.21}$$

Now, focus on the functional dependence on x_t, treating x_h as constant. The quadratic terms in x_t are:

$$(e_t)^T{\Lambda_{t,t}}^{-1}e_t \tag{3.22}$$

The linear terms in x_t are:

$$(e_t)^T(-\Lambda_{t,t}\mu_t - \Lambda_{t,h}(e_h-\mu_h)) \tag{3.23}$$

To obtain the conditional distribution of e_t given e_h, rewrite the quadratic and linear terms as a shifted quadratic centred at the conditional mean, plus a constant:

$$(e_t)^T{\Lambda_{t,t}}^{-1}e_t + (e_t)^T(\Lambda_{t,t}\mu_t - \Lambda_{t,h}(e_h-\mu_h)) = \left(e_t-\mu_{t|h}\right)^T\Lambda_{t,t}(e_t-\mu_{t|h}) + con. \tag{3.24}$$

where $\mu_{t|h}$ chosen so that the linear coefficient is cancelled out by the shift:

$$\Lambda_{t,t}\mu_{t|h} = \Lambda_{t,t}\mu_t + \Lambda_{t,h}(e_h-\mu_h) \tag{3.25}$$

Therefore

$$\mu_{t|h} = \mu_t + {\Lambda_{t,t}}^{-1}\Lambda_{t,h}(e_h-\mu_h) \tag{3.26}$$

Equivalently, when expressed in terms of the covariance matrix, the conditional mean is given by:

$$\mu_{t|h} = \mu_t + \Sigma_{t,h}{\Sigma_{h,h}}^{-1}(e_h-\mu_h) \tag{3.27}$$

and the associated conditional covariance is:

$$\Sigma_{t|h} = \Sigma_t - \Sigma_{t,h}{\Sigma_{h,h}}^{-1}\Sigma_{h,t} \tag{3.28}$$

Therefore, the conditional multivariate Gaussian distribution is expressed as:

$$f(e_t|e_h = x_h) = N(\mu_{t|h}, \Sigma_{t|h}) \tag{3.29}$$

Here, $\mu_{t|h}$ denotes the conditional predicted mean, and $\Sigma_{t|h}$ represents the conditional predicted variance at time interval t. This result provides a critical reference point for the adjustment and refinement of the predictive model. Specifically, the conditional mean incorporates a linear adjustment based on the observed value at interval h, reflecting the influence of this observation on the estimate at t. In turn, the conditional covariance reflects the reduction in predictive uncertainty at t once the information at h is considered. Collectively, these expressions constitute the standard results for the conditional distribution of a multivariate normal, forming the theoretical foundation for statistical inference in this context.

3.4 Practical Case Study Illustrations

In this section, a range of results are presented, including the baseline load forecast and the application of transfer learning for pseudo-measurement generation, as well as analyses of voltage magnitude and phase angle for the state estimation and power flow across both networks. The overall performance is evaluated for all buses, as well as for a representative bus. In addition, the outcomes of various observation scenarios for state estimation are compared. The section begins with the presentation of performance metrics for the load forecasting models.

3.4.1 Evaluation Metrics for Transfer Learning Performance

A comprehensive evaluation framework is essential for assessing the effectiveness of the proposed transfer learning methodology. This section begins by describing the quantification of transferability for instance pairs with identical labels across source and target domains. Four widely recognised metrics are employed to benchmark the proposed approach against baseline prediction methods: mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). These metrics are selected to provide complementary perspectives on forecasting performance. MAE offers a direct and interpretable measure of expected magnitude error. RMSE is included because it penalises larger deviations more heavily, making it sensitive to performance during higher-impact periods such as peaks. MAPE provides a scale-normalised assessment of error, which is particularly important in this study because LV substations can operate at markedly different demand levels; using MAPE enables meaningful comparison of models across substations with different load magnitudes on a

consistent basis. R^2 is used to summarise overall goodness of fit and indicate how well variability in the observed load is captured by the model.

Together, these metrics facilitate analysis of predictive accuracy, sensitivity to significant errors, and overall model fit, providing a robust basis for evaluating the impact of transfer learning relative to conventional approaches. A positive transfer learning effect is indicated if the transfer learning methodology outperforms the baseline across all four metrics, demonstrated by lower MAE, RMSE and MAPE and a higher R^2 value relative to the baseline model. In addition, performance is compared against a locally trained model to assess whether transfer learning can achieve results comparable to, or approaching, the best attainable performance under local training. The metrics are defined as follows:

$$MAE = \frac{\sum_{k=1}^{N} |y_k - \widehat{y_k}|}{N} \tag{3.30}$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N} (y_k - \widehat{y_k})^2}{N}} \tag{3.31}$$

$$R^2 = 1 - \frac{\sum_k (y_k - \widehat{y_k})^2}{\sum_k (y_k - \bar{y})^2} \tag{3.32}$$

$$MAPE = 100 \times \frac{1}{n} \sum_{t=1}^{n} \left| \frac{y_k - \widehat{y_k}}{y_k} \right| \tag{3.33}$$

In comparing the results of transfer learning efficiency within power flow analysis, voltage magnitude estimates derived from load forecast models trained on sparse datasets are evaluated against benchmark 'ideal case' power flow solutions obtained from fully instrumented network data. This study examines whether such data-constrained models can still provide reliable voltage estimates across the network, even in the presence of significant load forecast inaccuracies, and whether the predicted voltage extrema (maximum and minimum values) remain within acceptable tolerance thresholds.

3.4.2 Substation Dataset Comparison and Baseline Forecasting Results

The substation data utilised in this study for the source and target substations are drawn from two distinct regions of GB. The source substation, located in the north, is regarded as a data-rich area. It is a typical rural 11 kV substation, with data available at a 30-minute resolution over the course

of one year. In contrast, the target substation is situated in the west of GB and represents a data-sparse area, comprising several urban 11 kV substations. For the target substations, data are also recorded at a 30-minute resolution, but only for several weeks, which are generally insufficient for adequately training predictive models, which highlights the necessity of employing transfer learning techniques in this context.

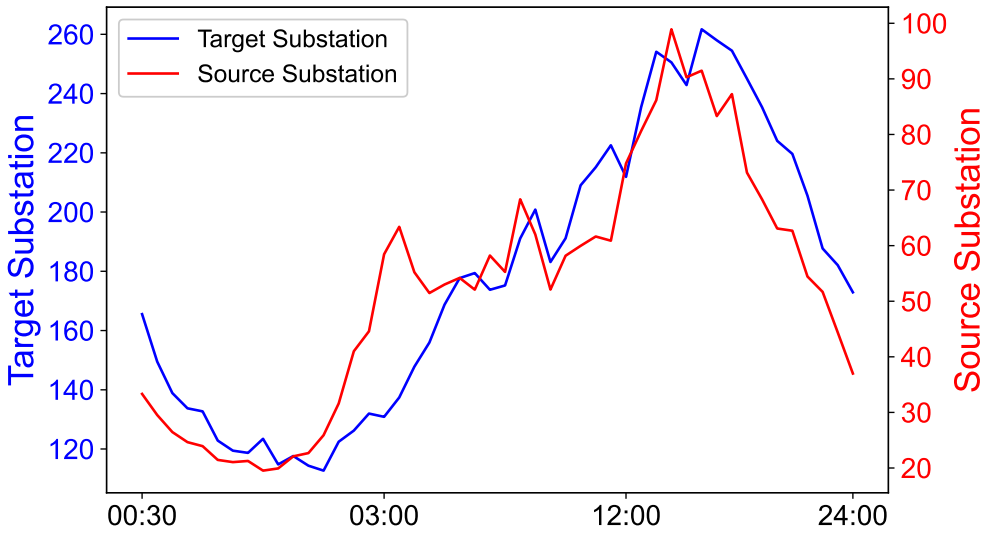

Figure 3.3 Daily load profiles for source and target substations with different scales

Figure 3.3 illustrates the daily load profiles for both the source and target substations. The blue line represents the target substation, which exhibits a higher average load of approximately 200 kW per day, consistent with its location in an urban area. In contrast, the red line corresponds to the source substation, where the average load is lower, at around 75 kW per half-hour interval. Compared to the target substation, the source substation demonstrates greater variability in its load profile, due to serving fewer customers. This increased variability presents additional challenges for accurate load forecasting.

The baseline load forecasting results were obtained using both the Hong Naïve Model and a basic LSTM network, allowing for a comparative assessment of baseline forecasting methods at the distribution level. The models were trained on data from the first three months for the source substation in the rural area, while the test data comprised load measurements from April and May at the same substation.

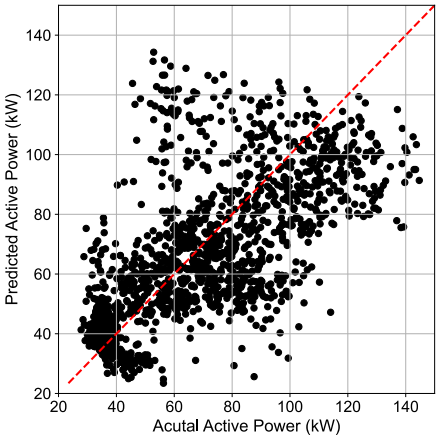

Figure 3.4 Comparative prediction plot for Hong Naïve model

Figure 3.4 presents the prediction results using the Hong Naïve Model. The model consistently underestimates peak values and overestimates troughs in active power, indicating a tendency to predict values towards the mean rather than capturing the full variability present in the data. As a result, the predictive performance is particularly limited when forecasting extreme values, highlighting the challenge of achieving accurate forecasts for LV substation loads. Nevertheless, the Hong Naïve Model still offers a basic level of predictive capability, making it a suitable benchmark for the application of transfer learning methods.

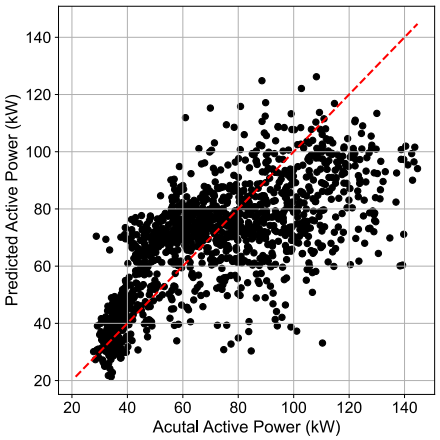

Figure 3.5 Comparative prediction plot for LSTM model

Figure 3.5 displays the prediction results for the same substation using the LSTM model. Here, a clear improvement over the Hong baseline model is evident, with notably better predictions for both the lower and mid-range values of active power. However, the model still exhibits some

dispersion in the predicted peak values, suggesting ongoing difficulties in accurately capturing extremes in demand. One key advantage of the LSTM model is its ability to utilise day-ahead data, which is often unavailable in data-sparse regions. Consequently, the Hong Naïve Model remains an appropriate baseline for evaluating the effectiveness of transfer learning methods in such contexts. In addition, the increasing spread of prediction errors at higher load levels suggests possible heteroscedasticity. This is a plausible characteristic of LV demand. Future work could explore variance-aware modelling approaches and assess whether variance-stabilising transformations or weighting improve performance during high-demand periods. In addition, the use of multi-year data would allow a more robust evaluation of level-dependent uncertainty across different weather patterns and annual operating conditions.

Load Forecasting Model	MAE	RMSE	MAPE	R^2
Hong-Baseline	17.308	22.768	25.56%	0.384
LSTM	14.865	19.622	20.97%	0.497

Table 3.1 Performance comparison of day ahead LV substation load forecast

Table 3.1 compares the performance of the two day-ahead load forecasting methods at the LV substation. The results show that the LSTM, as a machine learning method, outperforms the Hong Baseline Model, achieving approximately a 25% improvement in MAPE and a higher R^2 value. This demonstrates the advantage of machine learning approaches, which are better suited to modelling the non-linear relationships present in load data. However, it should be noted that forecasting performance at the LV substation does not reach the levels typically observed in transmission networks. Even so, the baseline model continues to provide reliable and meaningful results, serving as a valuable benchmark for the further development of transfer learning models.

The transfer learning methodology described below is employed to assess the accuracy of load profile prediction in the target domain. The metric defined in Equation (3.2) is used for the day-ahead forecasting of the source data. Subsequently, two transfer learning approaches are applied: Bayesian transfer learning, as defined in Equation (3.8), and Conditional Bayesian transfer learning, also defined in Equation (3.27). The performance of these methods is evaluated using the specified metrics to quantify the associated prediction errors. In addition, to examine the practical applicability of transfer learning, a benchmark model is implemented as a comparative baseline model within a local learning framework, thereby demonstrating the effectiveness of the proposed

transfer learning approach. LSTM models are not employed in transfer learning, as their inherent complexity makes it more difficult to provide a clear interpretation of both the transfer learning process and the state estimation outcomes compared to the Bayesian approaches considered here.

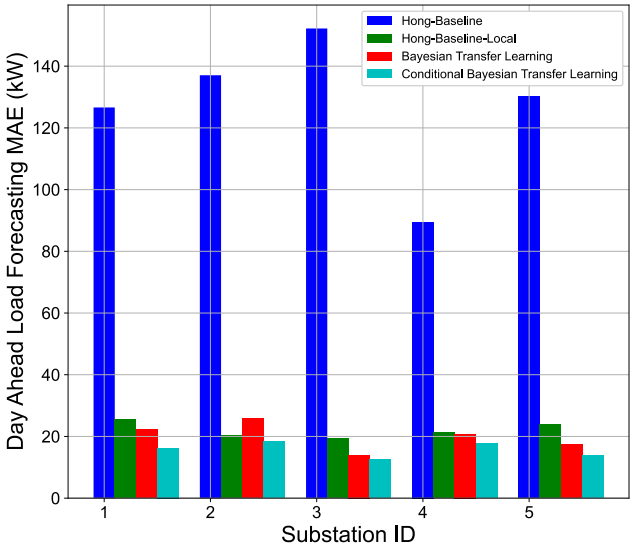

Figure 3.6 MAE comparison for five substations using four prediction approaches

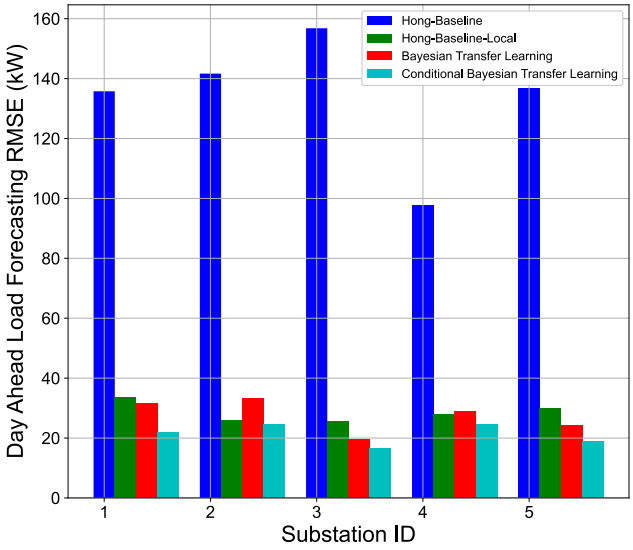

Figure 3.7 RMSE comparison for five substations using four prediction approaches

Figure 3.6 presents the MAE for day-ahead load forecasting across four methods. The advantage of the transfer learning approaches is clear. For all five substations, the application of transfer learning yields a substantial reduction in forecasting error compared to the baseline model. No

negative transfer effects are observed, with the forecasting error decreasing from approximately 120–150 kW in the baseline to around 20 kW when using transfer learning, closely matching the performance achieved by the local learning benchmark.

Based on Figures 3.6 and 3.7, when observed data at the start of the day are incorporated to calibrate the transfer learning forecasts, the conditional Bayesian transfer learning method demonstrates superior accuracy across all substations. It consistently achieves lower MAE and RMSE values compared to the other methods, even outperforming the local benchmark model. This improvement can be attributed to the model's ability to update daily, capturing short-term variations in substation load driven by calendar effects and weather conditions. In contrast, the benchmark model remains static, relying solely on historical training data as a linear regression model without adaptation. These results highlight the advantage of transfer learning, particularly when condition monitoring is integrated into the modelling process. Furthermore, the figures reveal variations in forecasting errors across substations, indicating that certain substations may exhibit load patterns that are inherently more complex and challenging to predict.

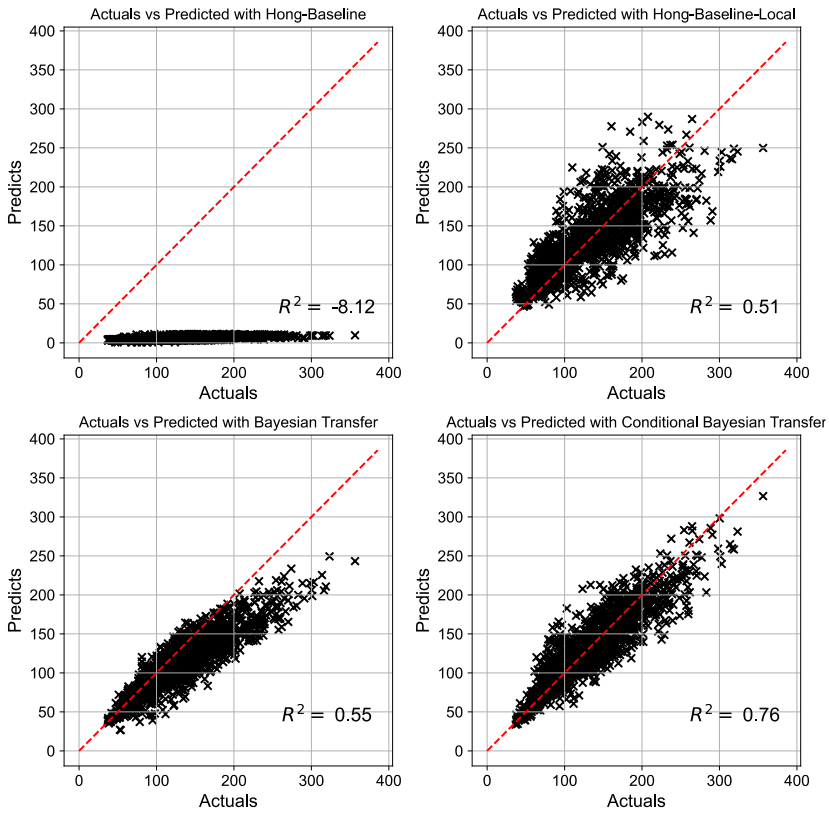

Figure 3.8 Comparative scatter plots of prediction (R^2) for four forecasting approaches

Figure 3.8 presents four scatter plots comparing the actual versus predicted values for each forecasting method. Each point represents a half-hourly load data point, where the true value is shown on the x-axis and the corresponding predicted value on the y-axis. The direct application of the baseline model is unsuitable due to differences in substation sizes, resulting in highly inaccurate predictions, as indicated by the poor alignment and negative R^2 value. Both the local Hong-Baseline and Bayesian Transfer Learning models demonstrate reasonable prediction accuracy for low and mid-range load values; however, their performance deteriorates at peak loads. In particular, the local Hong-Baseline model exhibits a wide dispersion of predictions, indicating significant errors, while the Bayesian Transfer Learning model shows a systematic underestimation of peak loads, introducing bias into the predictions. In contrast, the Conditional Bayesian Transfer Learning model achieves the best overall performance. By incorporating observation data, it reduces variability and produces predictions that are more tightly clustered around the diagonal, although some dispersion remains at the higher load levels. Overall, this method demonstrates superior predictive capability relative to the other approaches.

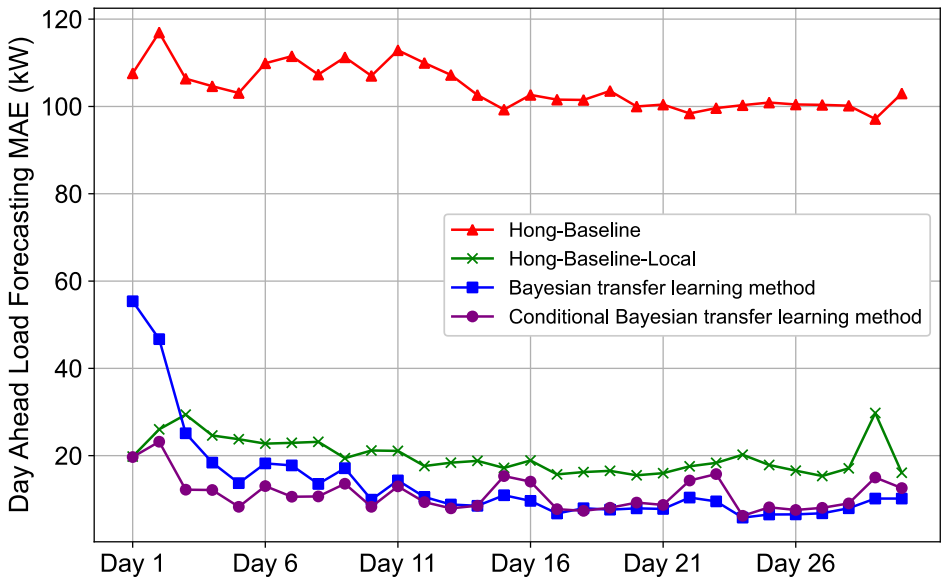

Figure 3.9 Daily MAE trends of four forecasting methods at a single substation over 30 days

Figure 3.9 illustrates the temporal evolution of forecasting performance for daily average MAE across the different transfer learning methods, highlighting their capacity to leverage previously accumulated error information. The results show that the standard Bayesian transfer learning method typically requires approximately 3 to 4 days to effectively assimilate this information and

progressively reduce the forecasting errors as the model adapts to the target domain. In contrast, the conditional Bayesian transfer learning method exhibits superior accuracy from the outset, benefiting immediately from the incorporation of real-time observations to adjust its forecasts in response to the most recent substation behaviour. This rapid adaptation allows the model to achieve a consistently lower error level throughout the forecasting period. As expected, the local training approach maintains a relatively stable error level across the entire period, since its model parameters remain fixed once trained on local historical data, and it does not benefit from continual updates or new information.

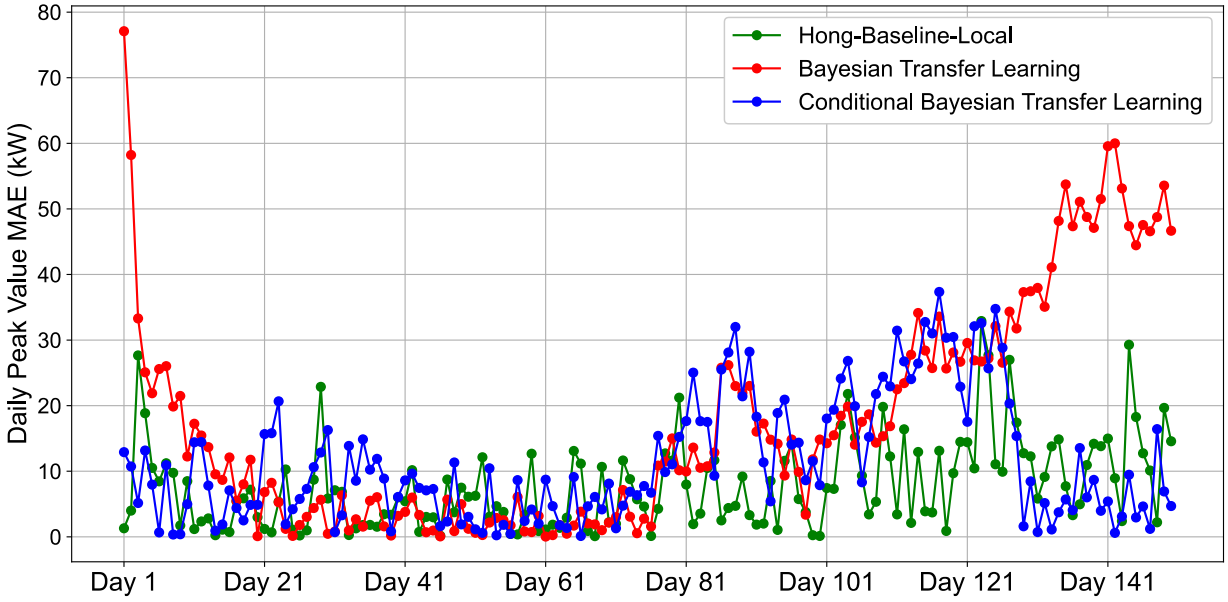

Figure 3.10 Daily MAE trends of peak load forecasts at a single substation over 150 days

Figure 3.10 illustrates the temporal evolution of forecasting performance for the different transfer learning methods, focusing specifically on the daily peak load forecasting error, in contrast to Figure 3.9, which examined the daily average MAE. The results indicate that peak load forecasting errors exhibit a periodic pattern. During the initial period, from Day 1 to approximately Day 80, all methods perform reasonably well, with errors generally remaining below 10 kW. However, after around Day 80, both the weather conditions and calendar effects begin to shift. As the test period starts in late July, the 90-day mark corresponds to late October, coinciding with the occurrence of St Jude's Storm (Cyclone Christian), which brought hurricane-force gusts of up to 99 mph to GB [21]. This extreme weather event likely influenced electricity consumption

behaviour, introducing unusual patterns into the load profiles. The Bayesian transfer learning method, relying on prior assumptions, struggles under these abnormal conditions, resulting in significantly higher errors and difficulties in maintaining accurate predictions. In contrast, the Conditional Bayesian Transfer Learning method, benefiting from real-time observations, is able to adjust more effectively after the storm and maintains lower error levels despite the disturbance. The local training approach, however, shows increasing error over time, potentially due to its inability to capture seasonal patterns as more months accumulate without cyclical adjustments.

Overall, these results highlight a key limitation of Bayesian transfer learning: extreme weather events can lead to prior mis-specifications, emphasising the importance of incorporating observational data for model recalibration. Furthermore, the findings suggest that including seasonal or monthly cycles within the modelling framework may improve the model's ability to track real-world variations over longer forecasting horizons.

3.5 Dynamic State Estimation Case Studies

After comparing the load forecast errors, it is necessary to further investigate their broader impact on network operation. Error metrics alone are insufficient to comprehensively assess forecasting performance at individual distribution substations. Equally important is understanding how forecast errors propagate across the wider distribution network. Simultaneous forecasting errors at multiple substations can result in rapid voltage excursions beyond statutory limits. Network modelling is therefore essential, as it captures the spatial distribution of voltages and currents under forecast conditions. While the distribution network may exhibit a degree of resilience when a significant forecasting error occurs at a single substation while others remain accurate, its vulnerability increases when multiple substations simultaneously experience moderate forecasting errors.

In this study, two network sizes are evaluated: the GB Urban 22-bus distribution network (Figure 2.18) and the UKGDS 77-bus test network (Figure 2.19). For the performance evaluation of the network state estimator, power flow results derived from the actual load data are designated as the true values, representing perfect information. The transfer learning forecasts, which include both active and reactive power at the low-voltage buses, serve as pseudo-measurements for state estimation. Simultaneously, power flow values for voltage magnitudes and angles at buses 0 and

1 are used as true measurements. These buses correspond to the high and low-voltage sides of the transformers, where such measurements are typically available in real-world systems. Specifically, the slack bus (bus 0) is assigned a voltage magnitude of 1 p.u. and an angle of 0 degrees, where "p.u." (per unit) denotes values normalised to a defined base, ensuring consistency across different voltage levels. The transformers are modelled as ideal, neglecting minor losses such as core and winding losses; consequently, the low-voltage side of the transformer is assumed to neither produce nor consume power [22]. For the state estimator parameters, true measurements are assigned a standard deviation error of 0.5%, while pseudo-measurements are assigned a standard deviation error of 5%.

3.5.1 Urban 22-bus Distribution Network Results

The first test is conducted using the GB Urban 22-bus distribution network, with Bus 10 selected as a representative example. The voltage magnitude and phase angle estimate obtained from the state estimation over a 10-day period, using the four forecasting methods, are presented below.

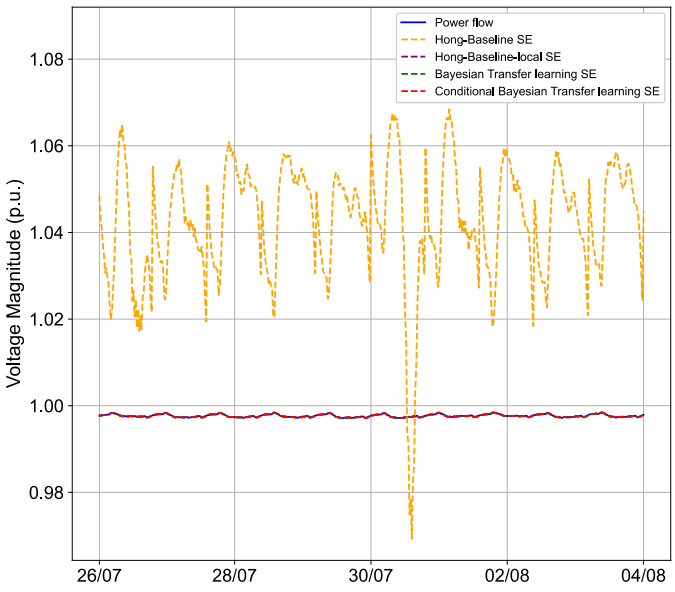

Figure 3.11 Comparison of voltage magnitude estimates from four state estimation methods at a specific bus in local network model

Figure 3.11 compares the four methods. It is evident that the Hong-Baseline method performs the poorest, as it fails to provide meaningful estimations for the localised substation cases. This demonstrates the ineffectiveness of direct using forecast model solution in this context, leading to

significant inaccuracies in the network results. A clearer comparison removing poor-quality data is provided in Figure 3.12. Furthermore, a statistical summary of the model performance is visualized in Figures 3.14 (scatter plots) and 3.16 (violin plots).

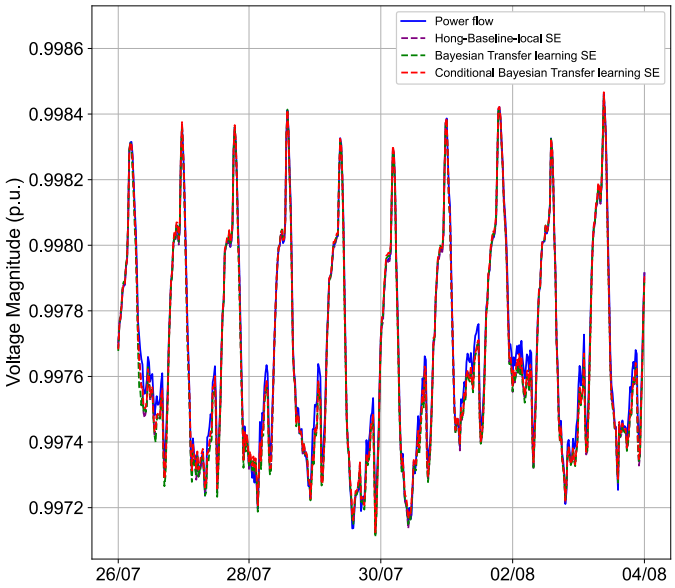

Figure 3.12 Comparison of voltage magnitude estimates from three state estimation methods at a specific bus in local network model

Figure 3.12 presents the results for the other three methods. At this specific bus, all three methods provide voltage magnitude estimates that closely follow the power flow results over the observed period. Although there are some errors in the load forecasts, the network itself has sufficient operational margin, which allows these state estimation methods to still achieve good accuracy. This highlights that, even when the forecasts are not perfect, the inherent operational margin within the network helps maintain reliable voltage magnitude estimation.

Figure 3.13 presents a comparison of the four methods when considering phase angle. Compared to the voltage magnitude results, the errors in phase angle estimation are relatively small. This is because distribution networks are dominated by residential loads, which are largely resistive and therefore have relatively small reactive power components. As a result, the estimation errors are reduced. Even the Hong-baseline model provides reasonable results, whilst the other three methods offer greater accuracy.

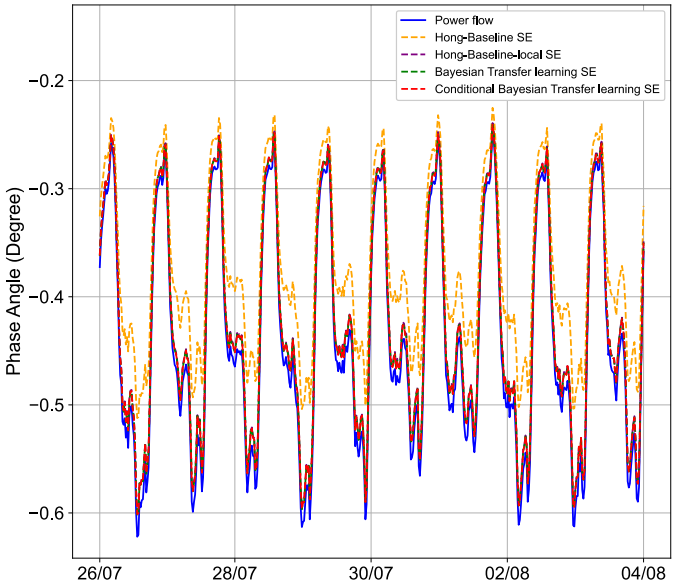

Figure 3.13 Comparison of phase angle estimates from four state estimation methods at a specific bus in the local network model

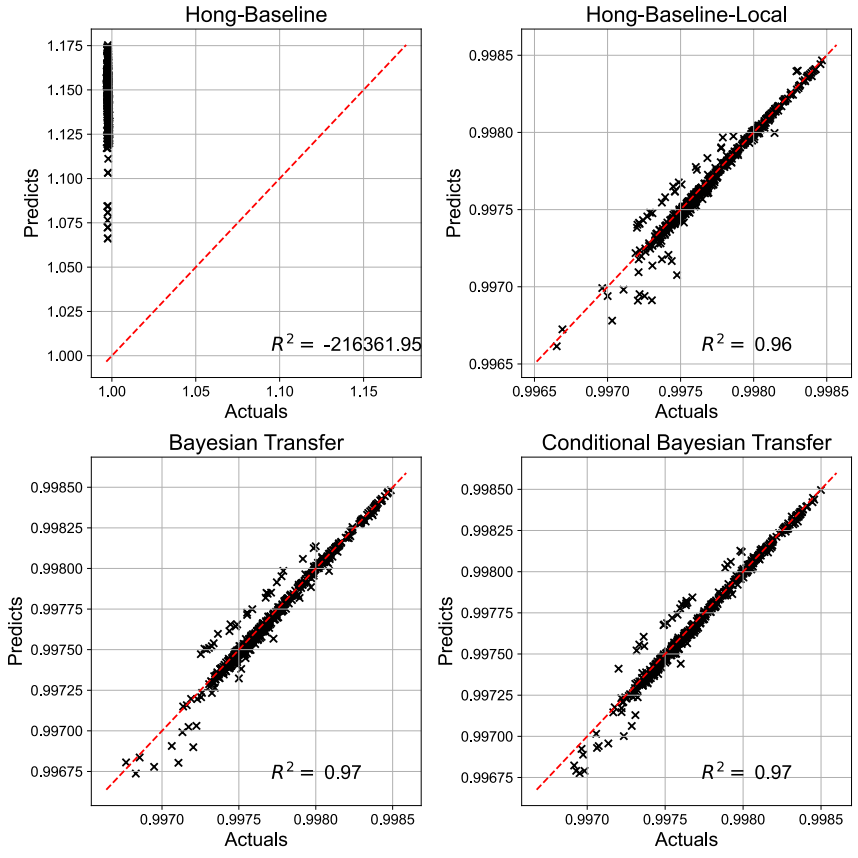

Figure 3.14 Comparison of state estimation voltage magnitude results across four methods

Figure 3.14 presents scatter plots comparing the state estimated voltage magnitudes at all buses (excluding the slack bus). The results demonstrate a consistent trend across individual buses. The Hong-Baseline model fails to produce meaningful results, as evidenced by the significant deviation from the ideal y=x line and the highly negative R^2 value. In contrast, the other three forecasting models closely match the power flow results, with R^2 values of approximately 0.96 or higher.

It is noteworthy that only one bus exhibits a significant bias error. This may be attributed to the presence of an industrial load at the target substation, whereas the source substation consists primarily of residual loads, potentially introducing this bias. Additionally, the main errors tend to occur when the voltage is low, corresponding to periods of peak load. This suggests that forecasting peak values remains challenging, as most models tend to underestimate peak magnitudes, drawing them closer to the average.

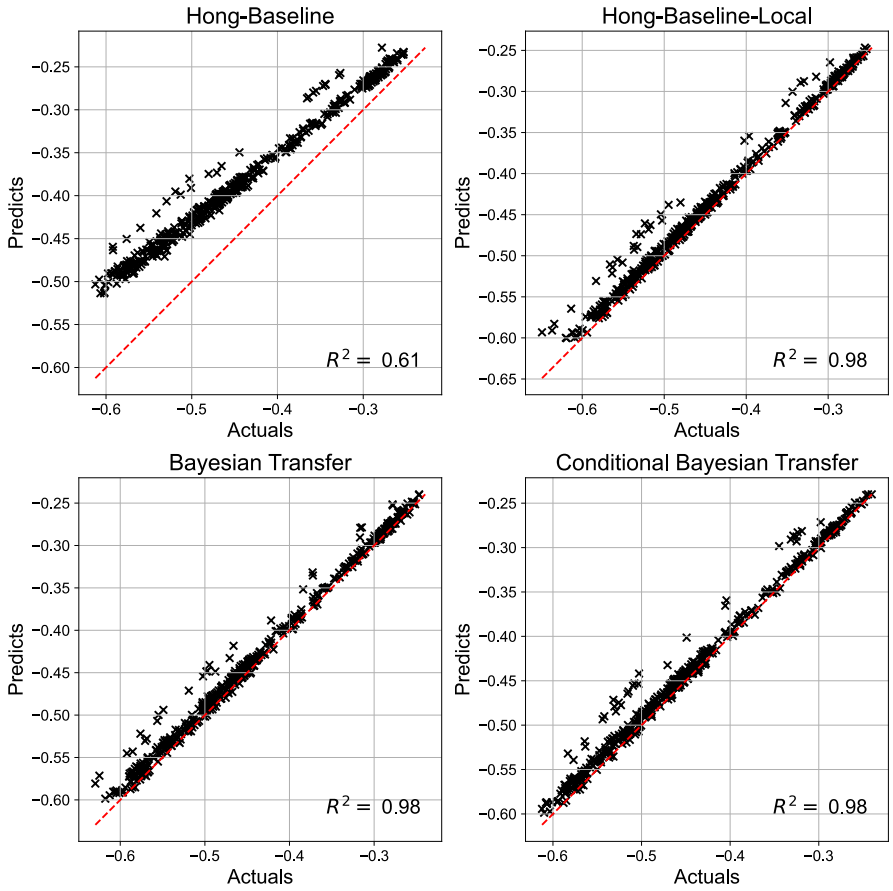

Figure 3.15 Comparison of state estimation phase angle results across four methods

Figure 3.15 presents scatter plots comparing the estimated phase angles at all buses (excluding the slack bus). The results reveal consistent trends across individual buses. The Hong-Baseline model yields reasonable results for phase angle estimation, as indicated by an R^2 value of 0.61, though its accuracy is notably lower than that of the other three models. The Hong-Baseline-Local, Bayesian Transfer, and Conditional Bayesian Transfer models achieve substantially higher accuracy, with R^2 values of 0.98.

It is also observed that a single bus exhibits significant error, which may be attributed to the presence of an industrial load. Industrial loads typically have much higher reactive power compared to residential loads, which could contribute to the increased estimation error at this location.

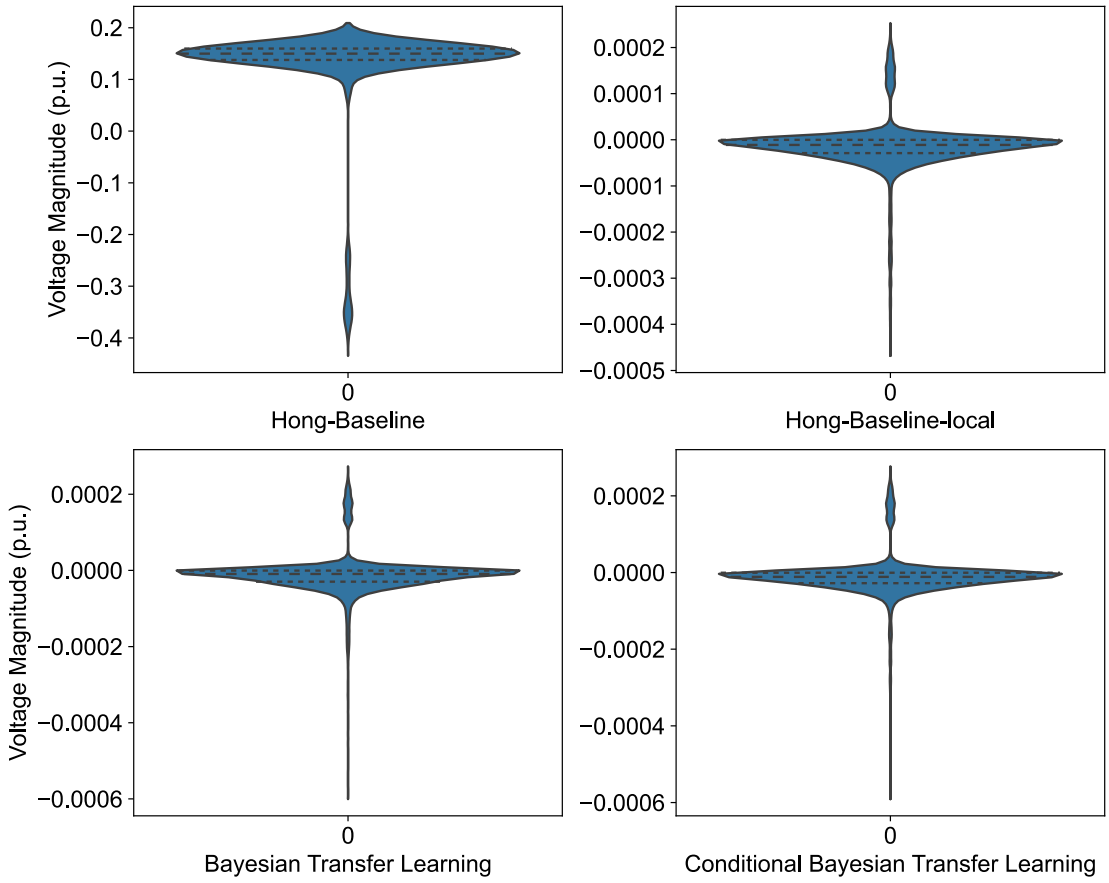

Figure 3.16 Comparison of voltage magnitude error distributions across four methods in a GB urban network

Figure 3.16 illustrates the error distributions for each method compared with the power flow results. These plots show the kernel density estimates of the data distributions, which provide insight into the spread and concentration of errors at different values. The inner dashed lines within each violin represent the quartiles (25%, 50%, and 75%), highlighting the spread of the data. The results indicate that the Hong-Baseline method exhibits the largest errors, with a range from approximately 0.1 to 0.2 p.u. Since the data are expressed in per unit (p.u.), this error is substantial and corresponds to a real voltage error multiplied by the base voltage 11000 V. In comparison, the other three methods show much smaller errors, with narrow distributions centred close to zero. These three methods result shows that in the relatively small network with Sufficient design capacity, such as larger substation capacity that the estimated voltages remain within the standard distribution network limits of 0.95 to 1.05 p.u., meeting practical operation requirements.

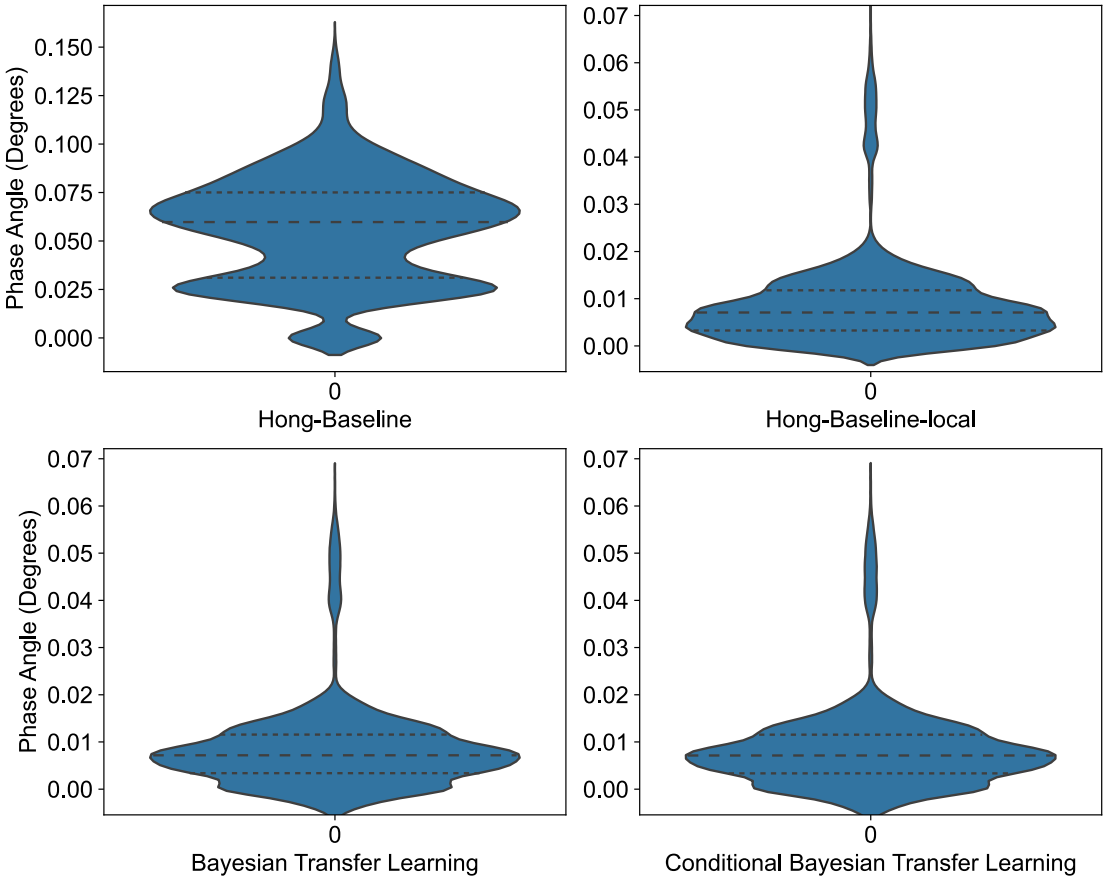

Figure 3.17 Comparison of phase angle error distributions across four methods in a GB urban network

Figure 3.17 further highlights the inferior performance of the Hong-Baseline method for phase angle estimation, as evidenced by its wider and more dispersed error distribution. In contrast, the Hong-Baseline-Local method displays a much narrower distribution, with most errors falling between 0.005 and 0.015 degrees, which demonstrates the effectiveness of incorporating local information through linear regression. Both the transfer learning methods yield similarly compact distributions, centred close to zero, indicating superior accuracy and robustness in phase angle estimation.

In practical distribution network operation, phase angle differences between adjacent buses are typically kept below 5 degrees, to maintain system stability and ensure reliable protection coordination. The results show that all three advanced methods maintain phase angle estimation errors comfortably within this acceptable range, meeting operational requirements for distribution networks.

3.5.2 UKGDS 77-Bus Test Network Results

The second experiment is conducted using the UKGDS 77-bus test network, with Bus 10 again selected as a representative example. The testing procedure follows the same approach as in the first experiment, allowing for a comparison of how different network characteristics influence performance based on the load forecast. In this case, the network is more complex and vulnerable. The state estimation results obtained over a ten-day period are presented below.

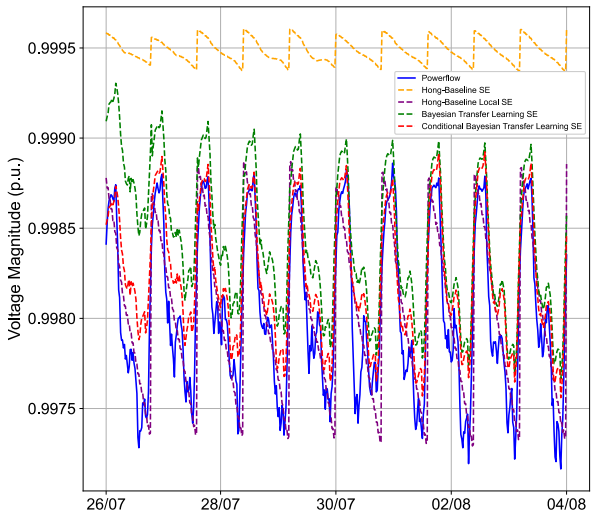

Figure 3.18 Comparison of voltage magnitude estimates from four state estimation methods at a specific bus in the UKGDS network model

Figure 3.18 presents the results for the four state estimation methods at this specific bus in the UKGDS network. The Hong-Baseline state estimation method cannot provide meaningful results, as its estimated values are consistently higher and do not represent the true voltage profile. The other three methods can capture the general trend of the actual voltage values but fail to reflect the sharp changes that occur over short periods. Nevertheless, the errors for these methods remain within acceptable limits, indicating that the transfer learning approaches are still reasonably effective, even in a larger network.

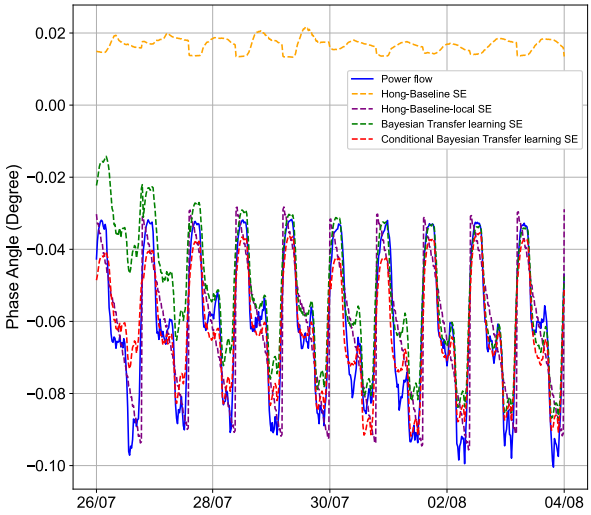

Figure 3.19 Comparison of phase angle estimates from four state estimation methods at a specific bus in the UKGDS network model

Similarly, Figure 3.19 presents the phase angle results at the same bus. The overall performance of the four methods mirrors what is observed for voltage magnitude. Once again, the Hong-Baseline state estimation method does not provide meaningful estimates. The other three methods manage to follow the overall trend of the actual phase angles, though they struggle to capture rapid fluctuations. Despite this, the estimation phase angle errors remain within acceptable bounds, further supporting the effectiveness of the transfer learning approaches in the larger network scenario. The difference in phase angle results arises from the hierarchical nature of the UKGDS compared to the simpler 22-bus system. The UKGDS, with multiple branching feeders and a diverse mix of loads, exhibits longer feeder paths, varying impedances. By contrast, the 22-bus test case has a simpler structure in which individual feeders are directly connected to a transformer,

resulting in shorter electrical paths, more consistent impedance, and therefore produce different phase angle behaviour.

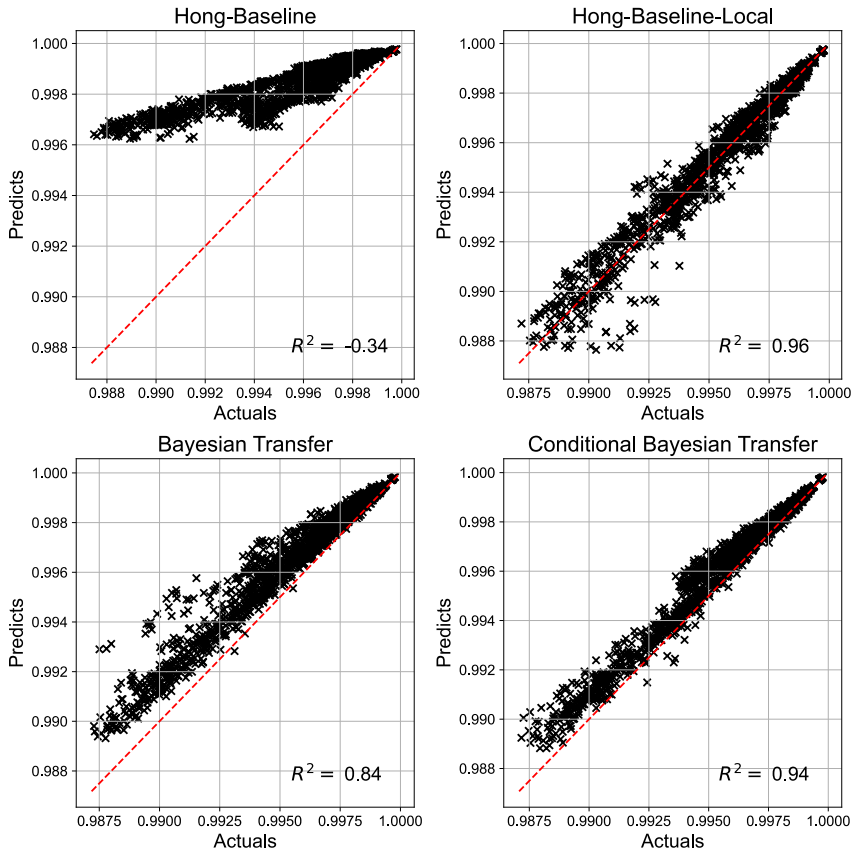

Figure 3.20 Comparison of state estimation voltage magnitude results across four methods in UKGDS 77-bus network

Figure 3.20 presents scatter plots comparing the state-estimated voltage magnitudes to the actual values at all buses (excluding the slack bus). Compared with the local network results, the estimations in the UKGDS network display greater errors. The Hong-Baseline model is still clearly incorrect, as its predictions are consistently biased and poorly correlated with the actual values. Both the Bayesian Transfer Learning and Conditional Bayesian Transfer Learning methods exhibit a noticeable bias error in the UKGDS network. However, with the assistance of observations, the bias in the Conditional Bayesian Transfer Learning method is much smaller than that in the standard Bayesian Transfer Learning method. On the other hand, the local model does not display a significant bias issue, but its results at the lower voltage end are more scattered. This suggests

that peak load forecasts tend to be inaccurate across most loads at the same time, since peak demand coincides with the low-voltage conditions in the network behaviour.

Overall, these results highlight the challenges of state estimation in a larger and more complex network, and demonstrate the value of incorporating transfer learning techniques, particularly when enhanced with observer information, for improving estimation accuracy.

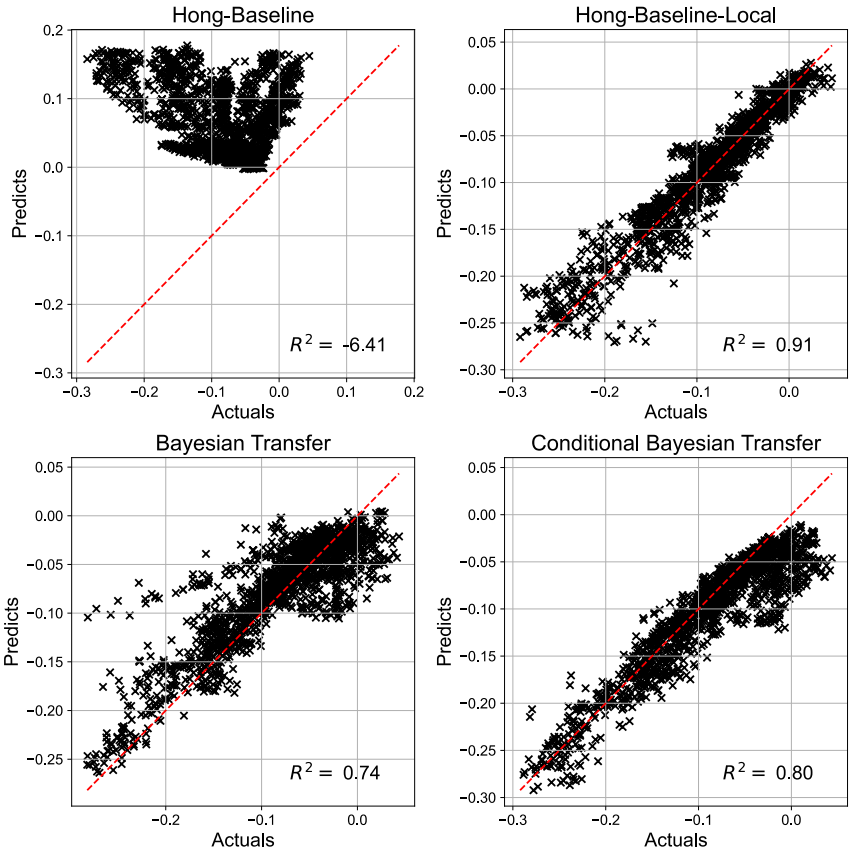

Figure 3.21 Comparison of state estimation phase angle results across four methods in UKGDS 77-bus network

Figure 3.21 presents scatter plots comparing the state-estimated phase angles with the actual values at all buses, except for the slack bus. The Hong-Baseline model again shows a poor fit, with significant errors and a negative R^2, highlighting its unsuitability for this task. While both Bayesian Transfer Learning methods achieve improved alignment with the actual phase angles compared to the Hong-Baseline model, some bias and dispersion remain, particularly in the standard Bayesian Transfer Learning approach. The Conditional Bayesian Transfer Learning method benefits from

observer information, resulting in a tighter clustering around the ideal line and a higher R^2. The local model continues to offer the best performance, with estimates that are closely matched to the actual phase angles.

Although Conditional Bayesian Transfer Learning method provides the most accurate load forecasts, it does not achieve the lowest errors in state estimation. The local model is able to yield smaller errors overall, indicating that improved load forecasting alone is insufficient to guarantee accurate state estimation. These results highlight the necessity of modelling the network and performing comprehensive state estimation in order to accurately assess the overall performance of the system after load forecasting. Given the presence of hundreds of loads in a typical distribution network, relying on simple error metrics from load forecasts is inadequate. Detailed state estimation is essential to understand how errors in load prediction propagate through network.

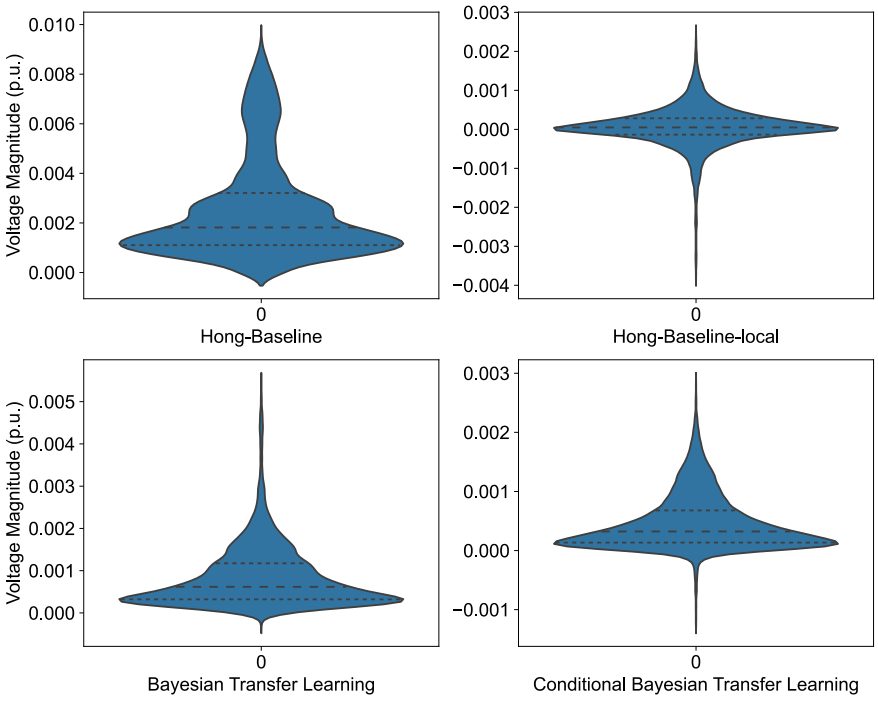

Figure 3.22 Comparison of voltage magnitude error distributions across four methods in UKGDS test network

Figure 3.22 presents the error distributions for voltage magnitude estimation, visualised using violin plots, with the layout and definitions consistent with those in Figure 3.16. The findings mirror previous results: the Hong-Baseline method displays the largest errors, with a broader

spread and longer tails, indicating a higher likelihood of large forecasting errors and outliers. In contrast, the errors for the other three methods are significantly smaller. The medians for other three methods are all close to zero, demonstrating their accuracy. The error distribution for the locally trained model closely approximates a Gaussian distribution, further validating its effectiveness as a benchmark. The transfer learning methods, however, show greater errors in the upper tail, suggesting a tendency to underestimate peak values in the load profile. With the inclusion of additional observations, the Conditional Bayesian Transfer Learning method substantially reduces the length of these tails, highlighting its improved identification of peak values.

It is also notable that the error associated with the UKGDS network is larger than that observed in the local network case. This increased error reflects the greater complexity and lower network reliability of the UKGDS network, indicating that it is less tolerant to modelling inaccuracies than the local network.

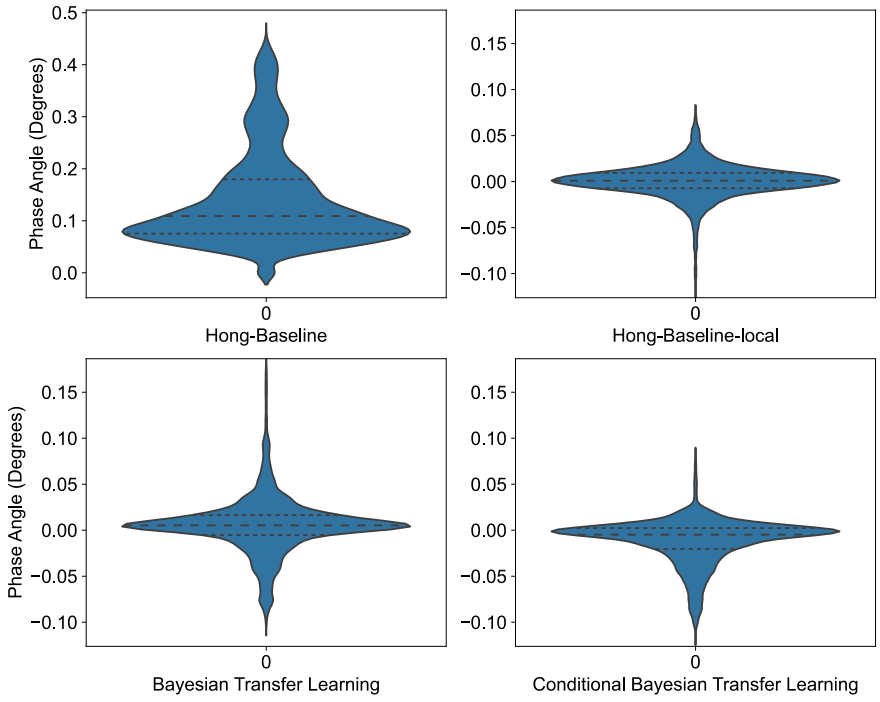

Figure 3.23 Comparison of phase angle error distributions across four methods in UKGDS test network

Figure 3.23 further illustrates the error distributions for phase angle estimation. The Hong-Baseline method again demonstrates the poorest performance, with substantially higher errors and a wider spread. Consistent with the results for voltage magnitude, the Bayesian Transfer Learning method can produce more extreme errors, while the inclusion of additional observations in the Conditional Bayesian Transfer Learning method effectively reduces these extremes. The Local model once more achieves the best overall accuracy, with a distribution tightly centred around zero, confirming the advantages of local training.

These findings suggest that, compared to active power, reactive power and corresponding phase angles display greater variability across different regions. This increased variability may be attributed to the more diverse mix of industrial and commercial consumers present in the network, underscoring the challenges of accurate state estimation in complex distribution systems.

Load Forecast Model	MAE	RMSE	MAPE
Hong–Baseline	108.4	122.6	99.3%
Hong–Baseline–local	19.2	26.2	21.1%
Bayesian transfer learning	18.5	26.3	16.7%
Conditional Bayesian transfer learning	**14.2**	**20.1**	**15.2%**

Table 3.2 Performance comparison of day ahead substation load forecast models

Table 3.2 presents the load forecasting results for the four methods, with bold values denoting the best performance in each metric. The Hong-Baseline method yields the largest errors by a significant margin across all metrics, indicating its unsuitability for practical applications. In contrast, the Hong-Baseline Local model shows substantial improvement, with much lower MAE, RMSE, and MAPE values. Both Bayesian Transfer Learning and Conditional Bayesian Transfer Learning methods further improve forecasting accuracy. Notably, the Conditional Bayesian Transfer Learning method achieves the best performance in all three metrics, demonstrating the benefit of incorporating even limited data monitoring. The Bayesian Transfer Learning method also achieves results very similar to the local training approach, underscoring the robustness and adaptability of transfer learning for deployment across different substations.

Overall, these results illustrate that transfer learning, particularly when supplemented with some local observations, can provide highly accurate load forecasts even when data is limited, making it a promising strategy for power network with limited monitoring.

State Estimation: Voltage Magnitude (Local Network)	MAE	RMSE	MAPE
Hong–Baseline	108.4	122.6	99.3%
Hong–Baseline–local	3.09×10^{-5}	5.94×10^{-5}	3.09×10^{-5} %
Bayesian Transfer Learning	**2.93×10^{-5}**	**5.54×10^{-5}**	**2.95×10^{-5} %**
Conditional Bayesian Transfer Learning	2.94×10^{-5}	5.69×10^{-5}	**2.95×10^{-5} %**

Table 3.3 Performance comparison of transfer learning methods in state estimation of voltage magnitude in local network

State Estimation: Phase angle (Local network)	MAE	RMSE	MAPE
Hong–Baseline	5.64×10^{-2}	6.29×10^{-2}	13.31%
Hong–Baseline–local	9.01×10^{-3}	1.34×10^{-2}	2.08%
Bayesian transfer learning	8.88×10^{-3}	1.29×10^{-2}	2.05%
Conditional Bayesian transfer learning	**8.87×10^{-3}**	**1.28×10^{-2}**	**2.04%**

Table 3.4 Performance comparison of transfer learning methods in state estimation of phase angle in local network

State Estimation: Voltage magnitude (UKGDS)	MAE	RMSE	MAPE
Hong–Baseline	2.54×10^{-3}	2.54×10^{-3}	25.6%
Hong–Baseline–local	**3.49×10^{-4}**	**5.40×10^{-4}**	**3.51%**
Bayesian transfer learning	8.47×10^{-4}	1.12×10^{-3}	8.52%
Conditional Bayesian transfer learning	4.83×10^{-4}	6.63×10^{-4}	4.86%

Table 3.5 Performance comparison of transfer learning methods in state estimation voltage magnitude in UKGDS network

State Estimation: Phase angle (UKGDS)	MAE	RMSE	MAPE
Hong–Baseline	1.42×10^{-1}	1.69×10^{-1}	319.83%
Hong–Baseline–local	**1.27×10^{-2}**	**1.89×10^{-2}**	**37.97%**
Bayesian transfer learning	2.11×10^{-2}	3.18×10^{-2}	102.98%

Conditional Bayesian transfer learning	1.77×10^{-2}	2.74×10^{-2}	117.11%

Table 3.6 Performance comparison of transfer learning methods in state estimation phase angle in UKGDS network

Tables 3.3, 3.4, 3.5, and 3.6 summarise the performance of transfer learning methods for dynamic state estimation of both voltage magnitude and phase angle, within the local network and the UKGDS network. Across all scenarios, the Hong-Baseline method produces the largest errors by a considerable margin, rendering it unsuitable for reliable state estimation.

When comparing the two networks, the errors associated with the UKGDS network are substantially higher than those observed in the local network. This can be primarily attributed to the reduced network reliability and increased complexity of the UKGDS network, which poses greater challenges for accurate state estimation.

For voltage magnitude estimation in the local network (Table 3.3), both Bayesian Transfer Learning and Conditional Bayesian Transfer Learning achieve extremely low errors, with MAE and MAPE values on the order of 10^{-5}, closely matching the performance of the local model. Notably, the increase in percentage error for transfer learning methods compared to local training is minimal with only about 0.05%, indicating that transfer learning can provide highly accurate results even when full local data is limited. This demonstrates the effectiveness of transfer learning when applied to state estimation in a less complex, more redundant, residential-dominated network.

For phase angle estimation in the local network (Table 3.4), a similar trend is observed: both transfer learning approaches achieve nearly the same low error rates as local training, with the Conditional Bayesian Transfer Learning model performing slightly better. This demonstrates the effectiveness of transfer learning when applied to state estimation within a less complex, more redundant residual dominant network.

However, when moving to the UKGDS network, the results in Tables 3.5 and 3.6 reveal a significant increase in error across all methods. For voltage magnitude estimation (Table 3.5), local training still outperforms all other methods, but the Conditional Bayesian Transfer Learning model continues to deliver reasonably accurate results, with a MAE of 4.83×10^{-4} and a MAPE of 4.86%. The Bayesian Transfer Learning method also maintains relatively low errors, though somewhat higher than the conditional approach. These results confirm that, despite the added complexity,

transfer learning methods remain viable and effective alternatives, especially when local measurements are sparse. In contrast, phase angle estimation in the UKGDS network (Table 3.6) exhibits much larger errors for all methods, particularly for the transfer learning approaches. While local training achieves a MAPE of 37.97%, both Transfer Learning methods see a considerable drop in accuracy, with MAPEs exceeding 100%. This outcome highlights the limitations of existing transfer learning frameworks when deployed in more complex networks characterised by a diverse mix of load types. It underscores the need for improved methodologies for reactive power forecasting and how its residuals behaviour in distribution networks, in order toto enhance the accuracy of phase angle estimation.

In summary, these results underscore that while transfer learning, especially with conditional or observer integration, can offer performance close to that of local training in less complex networks, additional challenges remain in larger and more intricate systems. Moreover, the findings reinforce the necessity of combining accurate network modelling and comprehensive state estimation to provide a true assessment of system performance, rather than relying solely on simple error metrics from load forecasting. This is particularly important given the hundreds of loads present in distribution networks, where small errors can aggregate and have wide-reaching impacts on operational security and reliability.

3.6 Conclusion

This chapter has introduced a methodology that applies transfer learning based on Bayesian inference, leveraging knowledge from data-rich area substations and transferring it to data-sparse substations through adaptive learning. Two distinct approaches have been evaluated: the first does not require continuous data support for the transfer process but relies on approximately 4–5 days of initial data to establish prior knowledge and enable meaningful forecasting; subsequent observations can then be used to update the prior and further enhance forecasting accuracy. The second approach employs a limited number of intraday observations to calibrate the transfer learning model, using a conditional Gaussian distribution to perform the model calibration.

The evaluation considers both load forecasting error metrics and voltage performance across two distribution network models to provide a comprehensive assessment of forecasting performance. The results demonstrate that transfer learning offers significant advantages for load forecasting

and state estimation at data-sparse substations, with forecast errors substantially reduced in comparison to benchmark models. Forecasting accuracy is further improved when intraday observations are incorporated for model calibration, in some cases exceeding the performance of locally trained baseline models. These improvements directly contribute to enhanced dynamic state estimation, where forecasting errors across multiple substations collectively influence the accuracy of the overall system state. In practical applications, this approach not only reduces monitoring costs by utilising external data sources but also delivers considerable commercial and operational benefits, particularly as distribution networks face increasing demands for balancing services driven by growing levels of distributed generation.

However, one limitation of the Bayesian approach lies in its vulnerability to extreme weather events, which can distort the prior knowledge and result in substantial forecast errors that require additional observational data for correction. Addressing this limitation, future work outside the scope of the thesis will focus on enhancing the load forecasting model's ability to capture extreme load profiles, improving its robustness under atypical demand conditions with a better forecast method to forecast the reactive power. In parallel, the transfer learning framework will be further developed to better account for the impact of weather conditions on the forecasting process. Additionally, the state estimation method will be refined to more effectively handle the variability and uncertainty inherent in distribution network load co-behaviours.

3.7 References

[1] J. Lu, B. Stephen, and B. Brown, "Intraday residual transfer learning in minimally observed power distribution networks dynamic state estimation," *Data-Centric Engineering,* vol. 5, p. e13, 2024.

[2] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering,* vol. 22, no. 10, pp. 1345-1359, 2010, doi: 10.1109/TKDE.2009.191.

[3] S. Niu, Y. Liu, J. Wang, and H. Song, "A Decade Survey of Transfer Learning (2010–2020)," *IEEE Transactions on Artificial Intelligence,* vol. 1, no. 2, pp. 151-166, 2020, doi: 10.1109/TAI.2021.3054609.

[4] Y. Yao and G. Doretto, "Boosting for transfer learning with multiple sources," 2010: IEEE, pp. 1855-1862.

[5] Y. Freund, R. Schapire, and N. Abe, "A short introduction to boosting," *Journal-Japanese Society For Artificial Intelligence,* vol. 14, no. 771-780, p. 1612, 1999.

[6] Dai, Q. Yang, G.-R. Xue, and Y. Yu, "Self-taught clustering," 2008, pp. 200-207.

Duan, D. Xu, and S.-F. Chang, "Exploiting web images for event recognition in consumer videos: A multiple source domain adaptation approach," 2012: IEEE, pp. 1338-1345.

[8] D. K. Ranaweera, G. G. Karady, and R. G. Farmer, "Economic impact analysis of load forecasting," *IEEE Transactions on Power Systems,* vol. 12, no. 3, pp. 1388-1392, 1997, doi: 10.1109/59.630486.

[9] Hong, P. Wang, and H. L. Willis, "A Naïve multiple linear regression benchmark for short term load forecasting," in *2011 IEEE Power and Energy Society General Meeting*, 24-28 July 2011 2011, pp. 1-6, doi: 10.1109/PES.2011.6038881.

[10] S. Haben, G. Giasemidis, F. Ziel, and S. Arora, "Short term load forecasting and the effect of temperature at the low voltage level," *International Journal of Forecasting,* vol. 35, no. 4, pp. 1469-1484, 2019/10/01/ 2019, doi: https://doi.org/10.1016/j.ijforecast.2018.10.007.

[11] Amral, C. S. Ozveren, and D. King, "Short term load forecasting using Multiple Linear Regression," in *2007 42nd International Universities Power Engineering Conference*, 4-6 Sept. 2007 2007, pp. 1192-1198, doi: 10.1109/UPEC.2007.4469121.

[12] D. Obst, J. d. Vilmarest, and Y. Goude, "Adaptive Methods for Short-Term Electricity Load Forecasting During COVID-19 Lockdown in France," *IEEE Transactions on Power Systems,* vol. 36, no. 5, pp. 4754-4763, 2021, doi: 10.1109/TPWRS.2021.3067551.

[13] G. Juberias, R. Yunta, J. G. Moreno, and C. Mendivil, "A new ARIMA model for hourly load forecasting," in *1999 IEEE Transmission and Distribution Conference (Cat. No. 99CH36333)*, 11-16 April 1999 1999, vol. 1, pp. 314-319 vol.1, doi: 10.1109/TDC.1999.755371.

[14] K. Y. Lee, Y. T. Cha, and J. H. Park, "Short-term load forecasting using an artificial neural network," *IEEE Transactions on Power Systems,* vol. 7, no. 1, pp. 124-132, 1992, doi: 10.1109/59.141695.

[15] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network," *IEEE Transactions on Smart Grid,* vol. 10, no. 1, pp. 841-851, 2019, doi: 10.1109/TSG.2017.2753802.

[16] A. Karbalayghareh, X. Qian, and E. R. Dougherty, "Optimal Bayesian Transfer Learning," *IEEE Transactions on Signal Processing,* vol. 66, no. 14, pp. 3724-3739, 2018, doi: 10.1109/TSP.2018.2839583.

[17] N. Zou, Y. Zhu, J. Zhu, M. Baydogan, W. Wang, and J. Li, "A transfer learning approach for predictive modeling of degenerate biological systems," *Technometrics,* vol. 57, no. 3, pp. 362-373, 2015.

[18] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*. Springer, 2006.

[19] B. Stephen, R. Telford, and S. Galloway, "Non-Gaussian Residual Based Short Term Load Forecast Adjustment for Distribution Feeders," *IEEE Access,* vol. 8, pp. 10731-10741, 2020, doi: 10.1109/ACCESS.2020.2965320.

[20] W. Penny, "Bayesian inference for the multivariate normal," *Wellcome Trust Center for Neuroimaging: University College, London,* 2014.

[21] B. B. C. News, "Metal theft: UK Energy Networks at risk, MPs warn," ed: BBC, 2013.

[22] H. Saadat, *Power system analysis*. McGraw-hill, 1999.

4 Few-shot learning for Cardinal Points Forecasts with Residual in LV Network

In this chapter, a point forecasting method is introduced and used to forecast cardinal points. As discussed in Chapter 3, achieving highly accurate time-series load forecasting may not be feasible for LV substations and could potentially mislead peak demand predictions. Therefore, instead of relying on overly smooth time-series forecasts, accurately identifying both the magnitude and timing of the turning points in the load curve enables system operators to make timely operational decisions, thereby supporting network balance and facilitating the rapid recognition of critical contingencies. This approach offers a more robust solution for forecasting and managing distribution networks, as it focuses on the aspects of demand that matter most to DNO operations and allow for timely intervention. By aligning the forecasts with points where intervention is most critical, the model becomes more targeted and delivers improved performance. Given the limited availability of monitoring data, a novel few-shot learning technique based on global model calibration is proposed. This method utilises minimal data to forecast both the magnitude and timing of the cardinal points. The chapter will begin by introducing the current definition of cardinal points in power systems at the transmission level. Subsequently, the background of few-shot learning will be presented, followed by a comprehensive explanation of the proposed methodology and a discussion of the results. The chapter concludes with a summary of the principal findings and directions for future research.

While transfer learning was explored in Chapter 3 as a means of addressing data scarcity, the specific framework developed there is not repeated in this chapter. Instead, this work advances the same underlying principle through few-shot learning. Few-shot learning can be regarded as an extension of transfer learning, as both approaches leverage knowledge acquired from a source domain to improve learning in a target domain with limited data availability [1]. In this chapter, the approach is further enhanced by replacing the linear regression baseline used in Chapter 3 with Generalised Additive Models (GAMs) and by incorporating fixed information available to DNOs. In this sense, few-shot learning represents an evolution of the methodology introduced earlier, making it well suited to the task of point forecasting.

4.1 Background and Definition of Cardinal Point Forecasting at the Distribution Level

Forecasting is a fundamental aspect of business decision-making, particularly within the energy sector, where accurate predictions of load, generation, and prices are essential for both power system operation and business management [2]. In GB, the National Energy System Operator (NESO) bears responsibility for the daily planning of electricity generation and transmission at the national level. Each day, NESO undertakes a comprehensive planning process within its control centre to develop System Operating Plans (SOPs) for key intervals of anticipated demand peaks and troughs, referred to as 'cardinal points' [3].

Based on the forecasts for each cardinal point, the corresponding SOP is subsequently provided to the real-time operations team. This procedure ensures that those responsible for operational decisions in real time are equipped with a clear, flexible, and economically efficient plan, grounded in the most reliable information available. By facilitating informed real-time operational decisions, the SOPs enable the Electricity System Operator (ESO) to maintain a secure and reliable balance of the system. In this manner, the forecasting and planning process ensures the most secure, economic, and adaptable operation of the national electricity system possible.

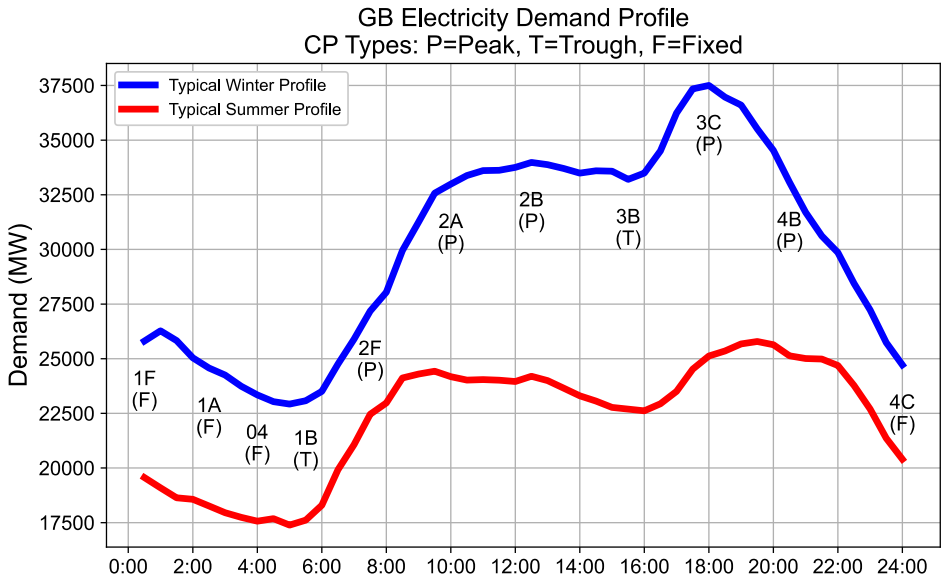

Figure 4.1 Identification of cardinal points in GB national electricity demand

In Figure 4.1, cardinal points are categorised by NESO as peaks (P), troughs (T), or fixed points (F), which together characterise the daily electricity demand profile. Peaks and troughs, in particular, are critical for system operation, reflecting periods such as the overnight demand minimum, daytime demand maximum, daytime demand minimum, and evening demand maximum. Electricity demand fluctuates throughout the day in response to the varying energy requirements of households, businesses, and industry. Fixed cardinal points, meanwhile, serve as important reference markers or baselines for the overnight load profile, include 1F at 00:30, 1A at 02:00, L04 at 04:00, and 4C at 24:00 [4].

The overnight minimum (1B) typically occurs in the early morning, usually between 04:30 and 07:30, as shown by 1B(T) in Figure 4.1. This is the period when most people are asleep, resulting in minimal electricity usage. The daytime minimum (3B), often referred to as the daytime trough, generally occurs between 13:30 and 16:30, a time when people are still at work and there is little additional electricity consumption beyond what is already in use for essential services such as computers and lighting. In recent years, daytime peaks and troughs have been increasingly influenced by embedded solar generation, which alters the traditional demand profile. Daytime peaks are both behaviourally and temporally dependent as indicated by the following cardinal points:

- 2F – Early morning demand peak, usually occurs between 08:00 and 09:00.
- 2A – The mid-morning peak, typically between 09:30 and 10:30, arises as the majority of people arrive at offices and schools.
- 2B – The midday peak, usually between 11:00 and 13:00, is often associated with a lunchtime increase in demand.

Evening peaks are also seasonally and operationally dependent:

- During British Summer Time (BST), 3C represents the afternoon peak as people return home, prepare dinner, and use additional appliances, whilst 4B captures the later evening peak as the sun sets and lighting is required.

- During Greenwich Mean Time (GMT), the "Darkness Peak" (DP) typically represents the highest demand of the day, coinciding with people returning home when it is already dark, leading to simultaneous increases in both appliance and lighting use.

For forecasting and consistency purposes, the relevant value is usually the maximum demand observed among these cardinal points. For example, in late June, the evening maximum for weekdays (Monday to Friday) is most frequently assigned to cardinal point 3C (17:00-20:00), which aligns with the earlier evening period consistent with standard commuting and household routines. Conversely, cardinal point 4B (20:30-22:00) is most often used for the evening maximum on weekends reflecting a shift in behaviour and a later peak demand associated with more flexible schedules on non-working days.

The demand at each cardinal point is broken down into separate forecast components according to the following formula [5]:

$$\begin{aligned} Forecast\ demand \\ &= Basic\ Demand + Weather\ Correction \\ &- Day\ of\ Week\ Correction - Embedded\ Wind\ and\ Solar \\ &+ Special\ Event \end{aligned} \tag{4.1}$$

The NESO utilises advanced statistical methodologies to generate cardinal point forecasts over a 2 to 14 days ahead horizon. For each cardinal point, NESO develops two distinct predictive models: a conventional model and a trend model. These models are constructed using a comprehensive array of historical data, encompassing historic electricity demands, day-of-week effects, historical meteorological data, and additional variables such as school holidays and the time of year [6]. The integration of these diverse datasets enables NESO to produce forecasts that are both robust and sensitive to the multifaceted drivers of electricity demand.

The outputs of these forecasting models determine a range of critical operational decisions. For example, the forecasts inform the requirement for additional balancing mechanism units to ensure that system demand is met, support assessments of system balance, and identify potential transmission constraints, such as network bottlenecks, that may impede system reliability. Furthermore, the forecasts are instrumental in determining contingency requirements; specifically, they quantify any anticipated shortfall or surplus in reserve margins necessary to guarantee that

generation assets can be synchronised in sufficient time to provide full output for the relevant cardinal point [3].

Within NESO's operational planning processes, cardinal points have been established as pivotal reference intervals for system monitoring and decision-making. Rather than relying on continuous time-series data, NESO utilises these key moments to inform the development of SOPs, thereby ensuring that planning efforts are both targeted and responsive to the most critical periods of system operation. This strategic approach is specifically intended to enhance operational focus and efficiency within the control room. The rationale for not employing full time-series data stems from the nature of system management: the energy system is actively overseen at these identified stress points, where rapid changes in system conditions are most likely to occur. In contrast, the periods between cardinal points are generally stable or predictable, and thus do not require constant, detailed intervention. Consequently, cardinal points serve as the primary "decision moments" during which system conditions can change abruptly, and prompt action may be necessary.

By concentrating operational attention on these critical intervals, NESO can formulate plans that are not only economic and secure, but also sufficiently flexible to accommodate the dynamic conditions characteristic of modern electricity systems. These plans are designed to be adaptable in real time as system dynamics evolve, thereby supporting the continuous and effective management of the system.

In contrast, constructing detailed SOPs for every possible time increment would impose a significant operational burden on both human and technical resources and, in most cases, would not deliver additional practical value. Such an approach would likely overwhelm operators with unnecessary detail and detract from their ability to concentrate on the most strategically important periods of system operation. By focusing on cardinal points, NESO can allocate resources and expertise precisely where they are most needed, ensuring that the operational response remains both targeted and robust. Ultimately, this methodology promotes a more resilient and responsive electricity system, capable of maintaining stability and security even under conditions of uncertainty or unexpected fluctuations in demand.

Furthermore, in the context of cardinal point forecasting, a number of studies have explored the adaptation of the profiling heuristic to operate using cardinal points, principally through the judgemental selection of a past load curve likely to resemble the forthcoming day's load profile

[7], [8]. The base curve may be chosen from the previous day, the corresponding day in the previous week, or a day in the previous year with similar weather conditions. This approach hinges on the premise that certain historic days provide a useful analogue for forecasting purposes.

To align the base curve with the cardinal points of the target day, scaling ratios are typically calculated to define the ratio of the forecast demand to the base curve demand at each cardinal point [9]. Between these points, linear interpolation is applied to estimate scaling ratios for intermediate periods, and the minute-by-minute forecast is then obtained by multiplying the base curve demand by the interpolated scaling ratio for each respective minute. A notable limitation of this method is that the cardinal points of the resultant forecast curve often do not coincide temporally with those of the original base curve. To address this misalignment, it is common practice to iterate the profiling process several times and using the most recent forecast curve as the new base curve for each subsequent iteration. Results indicate that this profiling heuristic outperforms more traditional interpolation techniques, such as cubic splines [10]. Several directions for further improvement have been identified, including the integration of multiple past daily profiles into the heuristic.

While substantial research exists on cardinal point forecasting at the transmission level, the distribution level remains largely unexplored. Yet this line of work could be highly beneficial and cost-effective for the active management of distribution networks. Producing comprehensive, granular time-series forecasts, for example at 30-minute intervals, for every feeder, substation, or district is resource-intensive in computation and oversight, and many distribution networks lack sufficient load data for reliable time-series models. Forecasting at cardinal points remains feasible under such constraints and is therefore particularly valuable. By concentrating on a small number of operationally salient points, operators can streamline forecasting and prioritise interventions. It also enables distribution system operators and DNOs in GB to plan more efficiently for peak or worst-case scenarios, such as transformer loading at the evening peak or voltage drops during minimum demand.

At the transmission level, a daily load shape is often summarised with about fourteen points, typically ten turning points and four fixed points. Distribution networks are more sensitive to local and unpredictable effects, for example clusters of electric-vehicle chargers or rapid swings in photovoltaic output. Because aggregation is weaker and volatility higher at the distribution level,

a faithful representation would usually require a richer set of cardinal points than the transmission-level fourteen. This chapter does not attempt that full catalogue. Instead, it provides a scoped example and a general framework that can be extended when data quality and computing resources allow.

Accordingly, in this chapter only focus on three most operational extremes: the daytime peak (06:00 to 13:00), the daytime trough (11:00 to 16:00), and the evening peak (16:00 to 23:30). These points are chosen because they dominate operational risk and can be estimated reliably with sparse data. Forecasts are produced at 30-minute resolution, reporting both expected timing and load magnitude.

At LV network, variability is amplified because timing and magnitude are shaped by human behaviour, especially on predominantly residential feeders. Historically, evening peaks align with cooking, heating, entertainment, and lighting [11]. However, the proliferation of remote working practices has begun to shift these dynamics, with energy consumption now potentially redistributed to other periods of the day [12], and negative electricity prices may incentivise electric-vehicle charging or heating shifts during low or negative price periods [13]. These effects can flatten traditional peaks or create cardinal points at atypical times, increasing the value of accurate timing and magnitude forecasts for day-ahead operations.

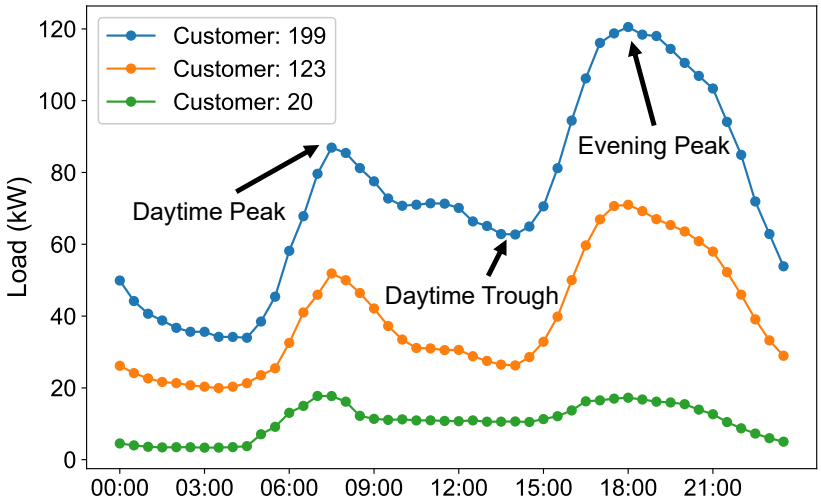

Figure 4.2 Average annual load profiles for three customer aggregations in a rural GB distribution network

Figure 4.2 presents the average load profiles, including cardinal point examples, for three distinct 11kV substations located within the same area of a GB distribution licence region over the course of one year. It is evident that all substations display a broadly similar pattern: a morning peak occurring between 07:00 and 09:00, corresponding to the start of most people's daily activities; a trough between 12:00 and 14:00, when the majority of individuals are typically at work; and a pronounced evening peak between 18:00 and 20:00, as people return home and increase their use of electrical appliances such as cooking and heating applications [14]. Furthermore, as the level of customer aggregation increases, aggregate load behaviour becomes more pronounced and regular, whereas at lower levels of aggregation load profiles are less consistent. For the three substations shown in Figure 4.2, the substation with the low customers tends to experience earlier daytime and evening peaks, and a later daytime trough, than the substations with higher customer numbers.

To address the challenge of substation data sparsity, the following section proposes employing a few-shot learning approach, utilising a global model based on fixed customer aggregation information, followed by model calibration using observed load data. This approach leverages fixed network information that is readily available to DNOs and does not change rapidly [15], helping to overcome the issue of limited observations at the distribution level.

4.2 Few-shot learning background

Machine learning has achieved remarkable progress in recent years, particularly towards solving complex tasks in area such as natural language processing and computer vision; however, most state-of-the-art methods require substantial amounts of data for training and tend to perform poorly when only limited data are available. In many domains, data is often scarce, not only due to the intrinsic nature of the problem or concerns regarding privacy, but also as a result of the considerable cost involved in data collection and preparation. Examples include medical imaging and diagnostics for rare diseases [16], fault detection in power systems, such as in deep-buried transmission cables [17], and building structural health monitoring, including post-disaster safety assessment and construction risk evaluation [18].

To address this challenge, few-shot learning has been proposed [19]. By leveraging prior knowledge, few-shot learning is able to rapidly generalise to new tasks using only a small number of labelled samples. Consequently, few-shot learning offers a promising and cost-effective solution,

significantly reducing the time required to develop machine learning applications that achieve acceptable performance.

Based on the manner in which prior knowledge is utilised to mitigate the core issue of data scarcity, few-shot learning methods can be categorised from three perspectives [1]. The first is data augmentation, in which prior knowledge is employed to enrich the supervised learning process. This includes augmenting the training data by generating multiple variations of existing samples [20], or by aggregating and adapting input and output from similar but larger datasets [21].

The second category is model improvement, where prior knowledge is used to constrain or reduce the size of the hypothesis space. Techniques such as multitask learning [22] or embedding learning [23] make the empirical risk minimiser more reliable, thereby reducing the risk of overfitting.

The third perspective is algorithmic enhancement, which leverages prior knowledge to guide the search for an optimal hypothesis within the hypothesis space. Methods such as refining existing parameters [24] can help reduce the search effort within the hypothesis set.

The few-shot learning approach employed here is based on data augmentation, which involves aggregating and adapting input and output data from similar but larger datasets. In this context, a global model is utilised. Global models exploit existing data by identifying shared patterns across equivalent components of the system, thereby enhancing generalisation and accuracy while reducing the risk of overfitting. This methodology has gained significant traction in the field of classification [25], [26]. Few-shot learning represents each class whether base or novel as a single point within an embedding space. Since the representation is learned jointly using training samples from both base and novel classes, it is referred to as a global representation. In regression, each task can be represented in the same way, and since the representation is learned jointly, it is referred to as a global representation. Furthermore, sample synthesis is employed to increase intra-class variation for novel classes. By randomly sampling data points from a subspace of samples belonging to the same class, the proposed synthesis strategy effectively augments intra-class variance. On the other hand, the straightforward use of global models is now becoming more common in time-series problems [27], global models treat a set of time series as a single dataset, allowing the model to be trained across multiple series when there is sufficient data. Consequently, global methods regarded as competitive alternatives to local modelling approaches, in which a separate model is trained for every specific time series. By learning shared representations across

related series, global models can achieve higher forecast accuracy and improved generalisation to unseen data.

Therefore, as most references on few-shot learning in global models are centred on classification tasks, while the present work addresses regression, a novel few-shot learning method in global models is proposed based on model calibration. Model calibration, which is widely adopted in weather forecasting [28], involves tuning model parameters using specific techniques to minimise the discrepancy between predictions and observations. In this context, model output variables act as predictors to explain persistent model biases, transforming a single-valued numerical weather prediction output into a corrected single-valued forecast [29]. This process effectively adjusts systematic errors in the average values of model outputs at observation sites [30].

Calibration is understood as a multi-faceted, joint property of both forecasts and observations: forecasts are considered calibrated if the observations can plausibly be regarded as samples drawn from the predictive distributions. Calibration refers to the statistical consistency between the distributional forecasts and the corresponding observations, and it is a joint property of both the predictions and the realised outcomes [31]. By narrowing the gap between modelled predictions and observed reality, calibration enhances both accuracy and reliability. In general, a forecast that exhibits greater concentration of probability mass around its central values is generally more useful, provided it remains well calibrated [32]. By efficiently capturing consistent relationships within the data, model calibration ensures both scalability and simplicity, as global model calibration extending scalability across regions and preserving simplicity through straightforward bias corrections.

Consequently, individual datasets can benefit from the knowledge encapsulated within a global model, thus enabling few-shot learning based on a limited number of observations.

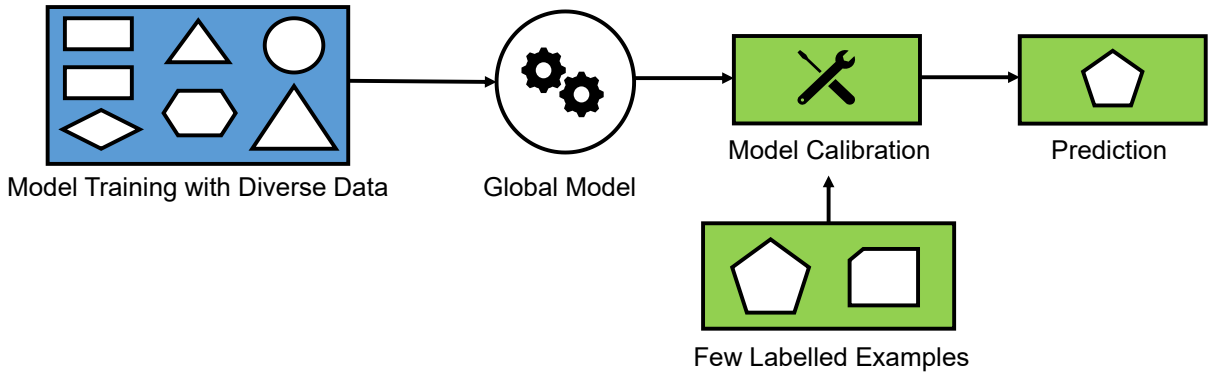

Figure 4.3 Few-Shot learning framework using a calibrated global model

The few-shot learning process is summarised in Figure 4.3, which illustrates the workflow of the proposed few-shot regression framework. The different shapes in the training set represent diverse datasets used to construct a robust global model, while the few labelled examples in the green box indicate the limited task-specific data employed for model calibration. For a given task with only a few labelled examples, the global model is calibrated through model calibration to account for task-specific characteristics. This calibrated model can then produce reliable predictions, effectively leveraging both global knowledge and limited local data.

In addition to few-shot learning, a simple zero-shot learning approach is also proposed here to serve as baseline models. Zero-shot learning seeks to enable a model to recognise novel data without the need for explicit training on those specific examples [33]. In this chapter, an attribute-based zero-shot learning framework is adopted, which leverages a predefined set of shared attributes to enable the model to make accurate predictions for previously unseen classes or tasks. Specifically, in the context of load cardinal point forecasting, zero-shot learning is employed to forecast unseen data where no labelled training dataset exists, relying solely on the relationships established by the global model [34]. By drawing upon the shared attribute space, attribute-based zero-shot learning facilitates the transfer of knowledge from known classes or tasks to entirely new ones, thereby enhancing the model's ability to generalise without direct supervision [35].

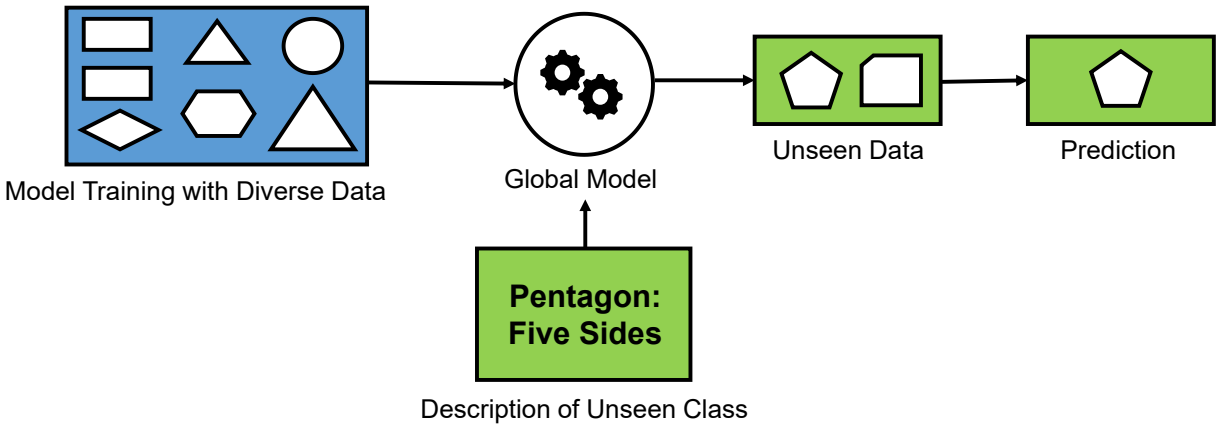

Figure 4.4 Zero-shot learning framework with a global trained model

The zero-shot learning process is depicted in Figure 4.4, which demonstrates the workflow of the proposed zero-shot learning framework. The model is initially trained on a diverse dataset encompassing multiple datasets to build a robust global model. For previously unseen datasets, only a high-level class description is provided, without any labelled examples. The global model leverages these semantic descriptions to identify and predict the unseen dataset, thus enabling generalisation to novel categories without requiring additional training data.

4.3 Global Model for Baseline Cardinal Point Forecasting

To predict day-ahead substation load data, local models are frequently employed, with each substation modelled individually using weather data, calendar effects, and lagged historical observations. However, these local models are subject to several limitations, including small sample sizes, a tendency towards overfitting, and poor scalability.

In response to these challenges, this study proposes a global modelling baseline approach. Substations within the same region often display similar patterns, attributable to shared environmental factors. As such, a global model can capitalise on pooled data, allowing the identification of subtle yet meaningful shared statistical structures [27]. Within a given region, the primary source of variation among substations is the number of customers, as substations with larger customer bases generally exhibit higher electricity demand. Therefore, it is both reasonable and practical to develop a global model for substations, explicitly incorporating the number of customers as a key explanatory variable. Based on this fixed information, a baseline approach is

presented here, which reflects the minimally achievable level of forecast performance.

The Generalised Additive Model (GAM) is employed as the global modelling approach in this study due to its capacity to flexibly capture non-linear associations between predictors and the response variable. GAMs are specifically designed to accommodate complex, non-linear relationships by using smoothing functions, making them particularly well-suited for situations where the connection between predictors and the response variable is not a straight-line relationship. Notably, the additive structure of GAMs preserves interpretability, as it models the overall effect as the sum of each predictor's contribution. This means the influence of individual variables can still be examined separately, rather than being entangled in complex interaction [36]. Within a GAM framework, the assumed distribution of the response variable determines the model fitting procedure. Specifically, a link function is utilised to relate the predictors to the expected value of the response variable. For the cardinal point forecast, where the load magnitude is predicted, the response variable is assumed to follow a Gaussian distribution, resulting in a linear GAM. Unlike smaller loads (e.g., smart meters) where a Gaussian assumption is often invalid due to high skewness, this assumption is reasonable for the secondary substation load data used here. The load magnitude acts as a continuous variable that, after accounting for known covariates, tends to exhibit approximately symmetric variation around its mean, consistent with the Normal distribution. To check this assumption, alternative distributional families were also tested, including Gamma and lognormal specifications. In this case, neither alternative led to an improvement in out-of-sample forecast accuracy, so the Normal assumption is used here.

By contrast, for the cardinal-point temporal forecast, the response variable is assumed to follow a Poisson distribution, resulting in a Poisson GAM. In this case, the response is expressed as event counts over discretised time intervals, which are non-negative integers, and a Poisson distribution provides a natural way of modelling such discrete outcomes. Other generalised additive model families, such as Gaussian or Gamma, assume a continuous response and are therefore less appropriate here, whereas the Poisson family is specifically designed for modelling the count of half hours between midnight and the peak, making it a practical choice for this application. These distributional assumptions are summarised in equations (4.2) and (4.3):

$$y_{scale} \sim N(\mu_{scale}, {\sigma_{scale}}^2) \tag{4.2}$$

$$y_{time} \sim \text{Poisson}(\mu_{time}) \tag{4.3}$$

Here, μ_{scale} and σ_{scale} denote the mean and standard deviation of the load magnitude data, respectively, whilst μ_{time} represents the mean of the time data, with the variance of the time data taken to be equal to its mean with μ_{time}, as implied by the Poisson distribution.

As GAMs already employ smooth functions to capture non-linear relationships, polynomial terms are not included. Instead, continuous numerical predictors are modelled using smooth functions, whilst categorical or binary variables (such as indicators for bank holidays or days of the week) are incorporated as parametric terms. This enables the model to account for abrupt shifts in the response variable that cannot be captured by smooth functions alone. The detailed GAM scale model for each substation is formulated as follows:

$$\begin{aligned}\mu_{scale} = f_1(\text{Day}) + f_2(\text{Month}) + f_3(T_{mean}) + f_4(T_{max}) + \\ f_5(RA_4) + f_6(RA_8) + + f_7(N_{cust}) + \sum_{j=8}^{15} \beta_j x_j + \beta_0\end{aligned} \tag{4.4}$$

Here, $f\_i(\cdot)$ denotes the i-th smooth function, and β_j are linear coefficient. Each smooth term is represented using a cubic B-spline basis, which provides an excellent balance between flexibility and smoothness, it can capture non-linear patterns whilst ensuring the fitted curve remains smooth and realistic [37].

For any such covariate z (for example, $f_1(\text{Day})$), the corresponding smooth function is expressed as:

$$f_i(z) = \sum_{k=1}^{K} \alpha_{ik}\, B_{ik}(z) \tag{4.5}$$

Where α_{ik} are basis coefficients and $B_{ik}(z)$ are cubic B-spline basis functions defined on a sequence of knots $t_{i,1} < t_{i,2} < \cdots < t_{i,k+4}$, Each basis function is a piecewise cubic polynomial that is non-zero only on the interval $[t_{i,k}, t_{i,k+4})$, outside this range, it is identically zero. This local support is what gives splines their computational efficiency and interpretability. In essence, splines represent a curve as a smooth combination of simple polynomial pieces joined at pre-specified knot points, allowing flexibility while avoiding overfitting.

T_{mean} and T_{max} refer to the average temperature and maximum temperature, respectively. In this study, both variables are derived from the previous day's weather data. These variables show a

strong relationship with the cardinal load points and temperature, as illustrated in Figure 2.10. This suggests that temperature dynamics are influenced by prevailing weather patterns throughout the day.

The variable rolling temperature averages also play an important role in peak load forecasting, primarily due to the thermal inertia of buildings [38]. Thermal inertia means that buildings respond slowly to changes in external temperature, so recent temperatures can have a lasting impact on energy demand. This relationship is clearly visible in Figure 2.12, which highlights the effect of rolling temperature averages on the load profile. In this study, the rolling temperature is defined as the previous day's rolling average, calculated from 6:00 p.m. onwards. This approach captures the ongoing influence of temperature on building energy use, particularly during evening peak periods. The variable N_{cust} represents the number of customers, reflecting how the size of the customer base influencing overall demand.

The variables Day and $Month$ in here are cosine transformations of the day and month, designed to capture the cyclical nature of seasonal trends and to ensure smooth transitions across time boundaries (e.g. from December to January). These transformations are defined as follows:

$$Day = cos(\frac{2\pi \cdot d}{365}) \tag{4.6}$$

$$Month = cos(\frac{2\pi \cdot m}{12}) \tag{4.7}$$

where d represents the day of the year and m denotes the month of the year. By using these cosine terms, the model can effectively account for repeating seasonal patterns without introducing discontinuities at the ends of the calendar.

For the categorical or binary variables represented as linear terms in the GAM model, $\sum_{j=8}^{17} \beta_j x_j$ includes the following predictors:

1. Binary indicator variables taking a value of 1 if the condition is met and 0 otherwise. Specifically:
    - Whether the day is a holiday.
    - Whether it is followed by a holiday.
    - Whether it is preceded by a holiday.

2. Dummy variables for each day of the week (Monday through Sunday), represented as a one-hot encoding. For example, Monday is coded as [1, 0, 0, 0, 0, 0, 0], Tuesday as [0, 1, 0, 0, 0, 0, 0], and so forth.

Finally, β_0 denotes the intercept, which is included as a fixed parameter in the model.

For the Poisson GAM, the mean and variance are calculated as:

$$\mu_{time} = \exp\left(\beta_0 + \sum_{j=1}^{8} f_j(x) + \sum_{j=8}^{15} \beta_j x_j\right) \tag{4.8}$$

where the f_i and β_j function are defined in the same way as in Equation (4.4) for the linear GAM.

In conclusion, the proposed GAM framework is intended to provide a flexible and interpretable methodology for substation-level load forecasting, modelling complex non-linear relationships through smooth functions and incorporating categorical influences via parametric terms. For substations located within the same or neighbouring areas, a GAM model structure and associated parameterisation are adopted, enabling the construction of a global model in which the only substation-specific variable is the number of customers. The suitability of this modelling choice is assessed in the empirical results presented in the following sections.

By substituting the relevant data for each substation into Equations (4.4) and (4.8), an attribute-based zero-shot learning baseline model can be established for forecasting cardinal points at similarly located distribution substations. In this context, the description of the unseen class in the zero-shot learning framework corresponds to a substation for which no historical load data are available, but which can still be characterised by fixed information with the number of connected customers, with the possibility of incorporating further fixed attributes if further data are available. For such a nearby substation, where only fixed information is known, the zero-shot learning model can be applied since all other explanatory variables are shared across substations, with only the customer-related attribute differing.

4.4 Few-Shot Learning Global Model Calibration

After utilising a global model to generate forecasts, calibration is a necessary subsequent step to refine these predictions for a target substation with limited observations. In this section, a few-shot

learning approach is proposed to enhance load cardinal point forecasting at substations with scarce data by calibrating the global model. This method leverages a Gaussian copula, incorporating observations from neighbouring substations to improve accuracy.

The section begins by outlining the main concept through a flowchart, which illustrates the entire methodological process. Subsequently, a detailed procedure for forecast calibration within the Gaussian copula framework is presented. The effectiveness of the approach is then examined using both bivariate and multivariate techniques.

4.4.1 Workflow of Few-shot Learning Calibration Structure

A novel approach to few-shot learning is presented here for the regression problem, achieved through calibration based on the global model. Since the number of customers alone cannot fully capture all the heterogeneity across different substations, forecast models inevitably fail to represent overall performance of the resulting load profile prediction. Although it is impractical to observe all substation data, a sparsely observed substation can be calibrated by leveraging data from a fully observed substation in the same area. This remains feasible even with limited information after applying the global model, as substations in the same area are likely to share similar residual changes relative to the global model output, given that customer behaviour trends tend to be similar.

Figure 4.5 provides an overview of the proposed few-shot learning approach using Gaussian Copula calibration of a global model. The flowchart begins by extracting cardinal points from available time-series substation data, which capture key load characteristics. Building on the global model, the approach then incorporates a few-shot learning procedure. In this phase, a limited number of local observations from a particular substation are used to calculate residuals. These residuals are subsequently calibrated against the fully observed substation data through a conditional Gaussian distribution in the Copula domain. By combining the corrected residuals with the output of the global model, the few-shot learning method is obtained. This approach capitalizes on both the broad representativeness of the global model and the fine-tuning afforded by local calibration.

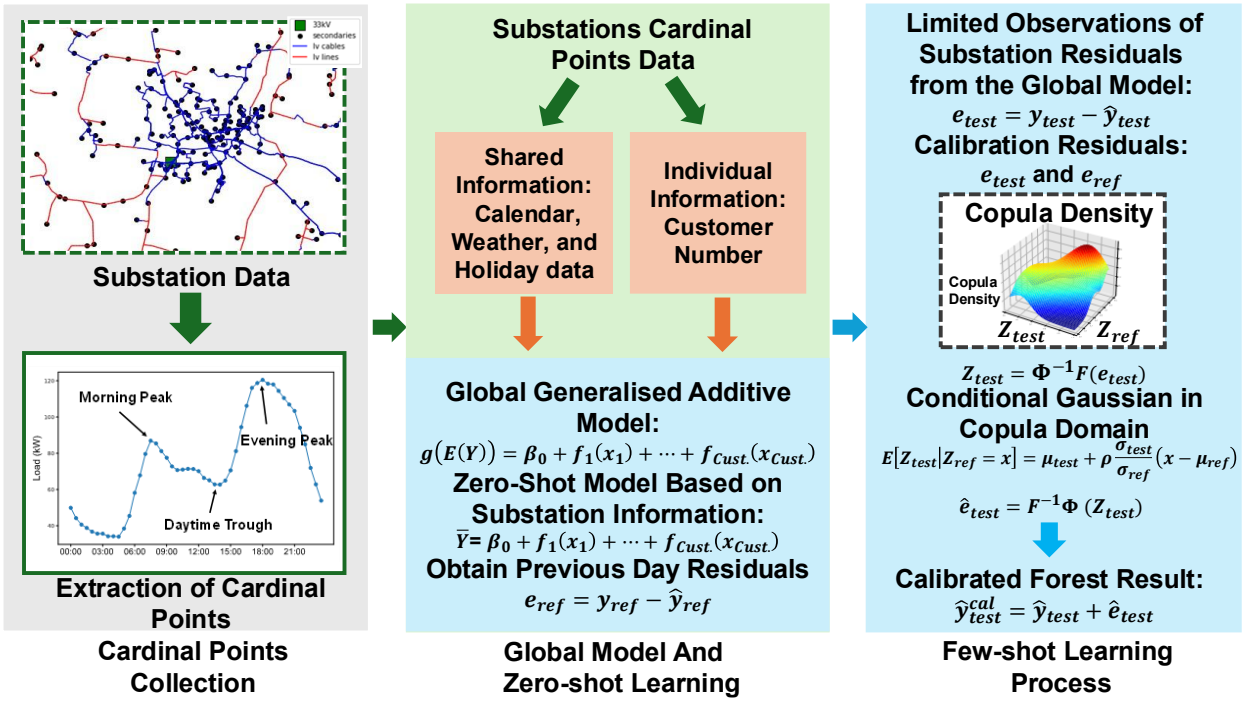

Figure 4.5 Few-shot learning LV load cardinal points forecast flowchart

4.4.2 Bivariate Gaussian Copula Cardinal Point Scale Calibration Method

In this section, a bivariate Gaussian Copula is introduced to calibrate the global model using only one week of observations (seven data points) for a particular substation. A key advantage of the copula framework is its independence from assumptions regarding the form of the marginal distributions. In empirical settings, these marginals often deviate often deviate from standard parametric families. Copulas address this by modelling the marginals and the dependence structure separately, thereby providing the flexibility to construct valid joint distributions even when the marginals do not conform to conventional forms. For example, in wind energy applications, wind speed is typically modelled with a Weibull distribution, whereas wind power output follows a highly non-linear and non-standard distribution; copulas provide a rigorous means of combining these marginals while separately representing their dependence [39].

This few-shot learning approach assumes that only seven samples are available, whereas another neighbouring substation is fully monitored and recorded on a daily basis. Given the complexity of non-linear dependence structures in the residuals of the global load forecasting model, copulas provide a robust methodological tool for modelling these relationships as a joint distribution [40].

Rather than modelling the load directly, it is preferable to model the residuals, as they more accurately capture the uncertainties inherent in the forecasting process [31].

Accordingly, a conditional copula derived from a joint distribution of residuals can be used to calibrate the forecast by incorporating information from previously observed errors. A copula is a mathematical function that captures the dependencies between several variables by linking their marginal distributions to construct a joint distribution, thereby enabling a more precise representation of the interrelationships within load forecast residuals. According to Sklar's Theorem, any joint cumulative distribution function (CDF) $F_Y(y_1, \ldots, y_n)$ with specified marginal distributions $F_{Y_i}(y_i)$ can be represented using a copula function C_θ as follows [41]:

$$F_Y(y_1, \ldots, y_n) = C_\theta(F_{Y_1}(y_1), \ldots, F_{Y_n}(y_n)) \tag{4.9}$$

In this study, the Gaussian copula is employed, which utilises a multivariate normal distribution to characterise the dependency structure among the variables. Extending this idea, the conditional Gaussian copula introduces conditioning variables by transforming the variables into standard normal space via their marginal cumulative distribution functions and subsequently capturing the underlying dependence structure.

The evening peak demand serves as an example of a cardinal point (e.g., one of three cardinal points), where $i \in \{1,2, \ldots,7\}$ to represent seven consecutive observations. Here, y denotes the observed demand and $\hat{y}$ denotes the corresponding global model forecast. The residuals are defined as:

$$e_{test,i} = y_{test,i} - \hat{y}_{test,i} \tag{4.10}$$

$$e_{ref,i-1} = y_{ref,i-1} - \hat{y}_{ref,i-1} \tag{4.11}$$

The subscript $test$ denotes a substation with limited observation data and ref denotes a fully observed substation. The index i corresponds to the current observation at the test substation and $i-1$ refers to the observation from the day before at the reference substation. The lagged residual $e_{ref,i-1}$ serves as a conditioning variable for modelling the statistical dependence structure of forecast residuals.

The residuals are first scaled to the uniform interval [0,1] using their respective marginal cumulative distribution functions. To prevent boundary issues, such as values being exactly 0 or 1, an empirical CDF is used in practice. For a dataset x_i, the empirical CDF is calculated as [42]:

$$F_{emp}(x_i) = \frac{rank(x_i)}{n+1} \tag{4.12}$$

This approach is applied to both residual sequences, i.e., the forecast errors from the test substation and the lagged errors from the reference substation.

$$u_{test,i} = F_{emp}(e_{test,i}) \tag{4.13}$$

$$u_{ref,i-1} = F_{emp}(e_{ref,i-1}) \tag{4.14}$$

The Percent-Point Function (PPF), which serves as the inverse of the cumulative distribution function, is then used to transform these uniform values into standard Gaussian space:

$$z_{test,i} = \Phi^{-1}(u_{test,i}) \tag{4.15}$$

$$z_{ref,i-1} = \Phi^{-1}(u_{ref,i-1}) \tag{4.16}$$

Here, Φ^{-1} denotes the inverse of standard normal CDF, and z represents the resulting values in Gaussian space.

It is assumed that the pair ($z_{test,i}$, $z_{ref,i-1}$) follows a bivariate normal distribution with means μ_1 and μ_2, standard deviations σ_1 and σ_2, and a correlation coefficient ρ. Under the Gaussian copula with this correlation structure, the conditional probability is expressed as follows [43]:

$$P(V \leq v|U = u) = \Phi\left(\frac{\Phi^{-1}(v) - p\Phi^{-1}(u)}{\sqrt{1-\rho^2}}\right) \tag{4.17}$$

For a set of days indexed by j, with the day before observation residual denoted by x, the conditional expectation for the substation with limited observations in Gaussian space can be expressed as [44]:

$$E[z_{test,j}|z_{ref,j-1} = x] = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x - \mu_2) \tag{4.18}$$

The corresponding conditional value in the Gaussian domain is then converted back to the uniform scale as follows:

$$u_{test,j} = \Phi(\hat{z}_{test,j}) \tag{4.19}$$

Finally, this uniform value is mapped back to the original scale by applying the inverse empirical CDF:

$$\hat{e}_{test,j} = F_{emp}^{-1}(u_{test,j}) \tag{4.20}$$

Here, $\hat{e}_{test,j}$ denotes the estimated residual for the unobserved substation, as determined from the global model. As a result, the predicted load for the substation with limited observations is given by:

$$\hat{y}_{test,j}^{cal} = \hat{y}_{test,j} + \hat{e}_{test,j} \tag{4.21}$$

where $\hat{y}_{test,j}^{cal}$ represents the calibrated prediction for the given substation with limited data in jth day, obtained by adjusting the output of the global model.

4.4.3 Bivariate Gaussian Copula Cardinal Point Time Calibration Method

For peak-time data recorded at half-hourly intervals, a discrete dataset comprising 48 thirty-minute points per day is obtained. Sklar's Theorem, as expressed in Equation (4.9), is conventionally applied to continuous variables. However, in the case of discrete data, the CDFs take the form of step functions, which can result in several copulas yielding the same joint distribution. To overcome this, the discrete data are mapped into a continuous Gaussian space via the probability integral transform [45], thereby ensuring a unique copula. This uniqueness is crucial for providing a clear representation of the dependence structure, enabling consistent modelling and inference.

As an illustrative case, suppose the evening peak residuals observed over seven days follow a Poisson distribution. The rate parameters for the Poisson distributions can be estimated as follows:

$$\lambda_{test} = E\left[e_{test,i}^{(t)}\right] \tag{4.22}$$

$$\lambda_{ref} = E\left[e_{ref,i-1}^{(t)}\right] \tag{4.23}$$

The Poisson CDF is then calculated by summing the probability mass function from 0 up to the relevant peak time value:

$$u_{test,i}^{(t)} = \sum_{i}^{k} \frac{\lambda_{test}^{r} e^{-\lambda_{test}}}{r!} \tag{4.24}$$

$$u_{ref,i-1}^{(t)} = \sum_{i}^{k} \frac{\lambda_{ref}^{r} e^{-\lambda_{ref}}}{r!} \tag{4.25}$$

In these equations, r is a dummy summation index used to compute the cumulative probability up to the observed value k; it is independent of the day index i used to refer to temporal observations.

The resulting values test represents the residuals transformed into the uniform domain via the Poisson CDF.

Finally, the calibrated forecast for the test substation at time t on day j follows the same formulation as in Equation (4.21), where $\hat{e}_{test,j}^{(t)}$ denotes the estimated residual for the unobserved substation, as determined from the global model. Accordingly, the predicted load $\hat{y}_{test,j}^{(t),cal}$ is given by the global model's output $\hat{y}_{test,j}^{(t)}$ adjusted with this residual. This correction leverages statistical dependencies learned from a fully observed reference substation, thereby enhancing accuracy during peak demand periods.

4.4.4 Multivariate Gaussian Copula Calibration Method

In the bivariate framework, the assumption is that a single factor, such as the magnitude or timing of a peak, influences only the prediction residuals from the previous day at another substation. However, a single predicted variable, for example the evening peak, may also have relevance at different temporal scales, such as the morning peak period. For instance, adverse weather conditions may prompt sudden changes in consumer behaviour. If poor weather is expected, individuals may return home earlier from social activities or weekend breaks, leading to increased energy consumption for heating and lighting. This shift in behaviour can cause a higher morning peak as people use electric heaters and lights earlier and can also result in an earlier evening peak as people arrive home sooner due to the weather. These examples demonstrate the complexity involved in accurately forecasting load patterns and highlight the limitations of a bivariate model, which might not capture the interactions among multiple influencing factors in limited datasets.

To address these multifaceted interactions, it is necessary to move from a bivariate to a multivariate modelling framework. In this expanded approach, all six previous-day variables are included as dependent factors within a Gaussian copula, with their correlations considered. This framework helps to overcome the limitations associated with small observational datasets, as it enables the model to capture a broader range of dependencies and to better reflect the complex behavioural patterns underlying load variability. The dependency factor is now represented as a vector consisting of six residual error variables: three corresponding to the cardinal point scale and three corresponding to the time-of-day residuals from the previous day at another substation. This is expressed as follows:

$$E_{ref,i-1} = \left[e_{ref}^{CP1}, e_{ref}^{CP2}, e_{ref}^{CP3}, e_{ref_t}^{CP1}, e_{ref_t}^{CP2}, e_{ref_t}^{CP3}\right]^T \in R^{6\times 1} \quad (4.26)$$

Here, $CP1$, $CP2$, $CP3$ refer to the morning, evening, and trough periods, respectively, while t refers to time. The empirical CDF is applied to the three scale residuals as described in Eq. (4.15), and the time-of-day residuals are calculated using the Poisson CDF as specified in Eq. (4.24). Therefore,

$$U_{ref,i-1} = \left[U_{ref}^{CP1}, U_{ref}^{CP2}, U_{ref}^{CP3}, U_{ref_t}^{CP1}, U_{ref_t}^{CP2}, U_{ref_t}^{CP3}\right]^T \in R^{6\times 1} \quad (4.27)$$

The transformed variable is then obtained by applying the inverse standard normal CDF:

$$Z_{ref,i-1} = \Phi^{-1}(U_{ref,i-1}) \in R^{6\times 1} \quad (4.28)$$

The mean and variance of $Z_{ref,i-1}$ are denoted by μ_2 and σ_2, both in $\in R^{6\times 1}$. The correlation matrix between $Z_{test,i}$ and $Z_{ref,i-1}$is given by $\mathrm{P} \in R^{7\times 7}$. With the previous day's observed residual denoted by X, the multivariate conditional expectation in Gaussian space for the substation with limited observations is given by:

$$E\left[Z_{test,j} \middle| Z_{\mathrm{ref},j-1} = X\right] = \mu_1 + \mathrm{P} * \frac{\sigma_1}{\sigma_2} * (X - \mu_2) \quad (4.29)$$

Applying Equation (4.19) to (4.21), the predicted load for the limited-observation substation is given by the calibrated formulation in (4.21), where $\hat{y}_{test,j}^{cal}$ represents the multivariate calibrated load prediction for the substation with limited data, based on the output from the global model.

4.5 Results

Two distinct datasets are utilised to assess the effectiveness of the proposed methodology. The first dataset pertains to a secondary substation network located in a rural area of northern Scotland. This network supplies electricity to between 50 and 200 customers, with a typical demand of approximately 100 kW, and involves a step-down in voltage from 11 kV to 400 V to facilitate final distribution to end users. Conversely, the second dataset encompasses transformer flow data from a primary substation situated in an urban region of Wales. Here, the substation caters to a much larger population, ranging from 5,000 to 15,000 customers, with an average load of around 10 MW and a voltage transformation from 132 kV down to 33 kV.

It is also important to note that certain substations, such as those supplying farms, parks, or wind turbines, have no officially registered customers, yet they still exhibit electricity consumption. This scenario underscores the necessity for a few-shot learning approach to accurately calibrate the underlying load profiles.

Both datasets cover an entire calendar year. The initial nine months are dedicated to training both global and local forecasting models, whilst the following two months are reserved for testing and validation. December is intentionally excluded from the analysis, as the Christmas period can have a significant and unique impact on electricity demand. Accurate forecasting for this period would require multi-year data. By omitting December, no unexpected errors from festive season anomalies are introduced into the results. By employing datasets of markedly different sizes and characteristics, the proposed few-shot learning method is rigorously evaluated for its robustness, reliability, and capacity to accommodate variability in substation cardinal point forecasting.

Figures 4.6 to 4.11 present six illustrative diagrams that showcase results at evening peak, morning peak and daytime trough cardinal points, offering a comparative analysis of four distinct forecasting methods. These comprise multivariate and bivariate few-shot learning, which use only a limited number of observations for calibration; local training, which makes use of fully observed local datasets in conjunction with the global model; and zero-shot learning, which depends only on the number of customers as input to the global model, without requiring any observational data.

Each figure depicts forecasting errors along two dimensions. For the scale-based analysis, Percentage Error (PE) is used as the performance metric. This expresses errors as a proportion of the actual value, which enables fair comparison between substations with different load magnitudes. The percentage error is calculated as follows:

$$PE = \frac{y_{test} - \hat{y}_{test,j}^{cal}}{y_{test}} \times 100\% \tag{4.30}$$

For the time-based forecasts, the Normal Error (E) is used, which remains constant across time and does not require normalisation. The normal error is defined as:

$$E = y_{test} - \hat{y}_{test,j}^{cal} \tag{4.31}$$

Since errors can be both positive and negative, these metrics indicate not only the magnitude of the forecasting errors but also any systematic tendency to overestimate or underestimate the actual values.

The results are visualised using two-axis box plots, which offer a clear representation of the error distributions. The x-axis corresponds to the scale-based percentage error, while the y-axis displays the time-based absolute error. Each box plot displays the 25th percentile (lower quartile), the 50th percentile (median), and the 75th percentile (upper quartile). Outliers are indicated separately in order to highlight significant deviations from the main body of the data.

Percentiles are calculated as follows: if the dataset is arranged in ascending order and contains N values, the pth percentile (where p is between 0 and 100) is located at the position P_p:

$$P_p = \frac{p}{100} \times (N + 1) \tag{4.32}$$

The interquartile range (IQR) is defined as the difference between the third quartile (Q_3) and the first quartile (Q_1), correspond to the 75th and 25th percentiles, respectively:

$$IQR = Q_3 - Q_1 \tag{4.33}$$

Data points are classified as outliers if they fall outside the following range: $[Q_1 - 1.5IQR, Q_3 + 1.5IQR]$. Observations outside these bounds are regarded as anomalous.

A skewed median within the box suggests a tendency towards overestimation or underestimation, providing insight into potential data bias. The diagrams also highlight the distribution of smaller errors, while any percentage errors above 30 percent are plotted using a symmetrical logarithmic scale on the x-axis, with a linear threshold set at 30. This technique compresses the scale for larger deviations and enhances the visual clarity for the range of more moderate errors. This overall approach enables a nuanced and transparent assessment of forecasting performance across the different methods.

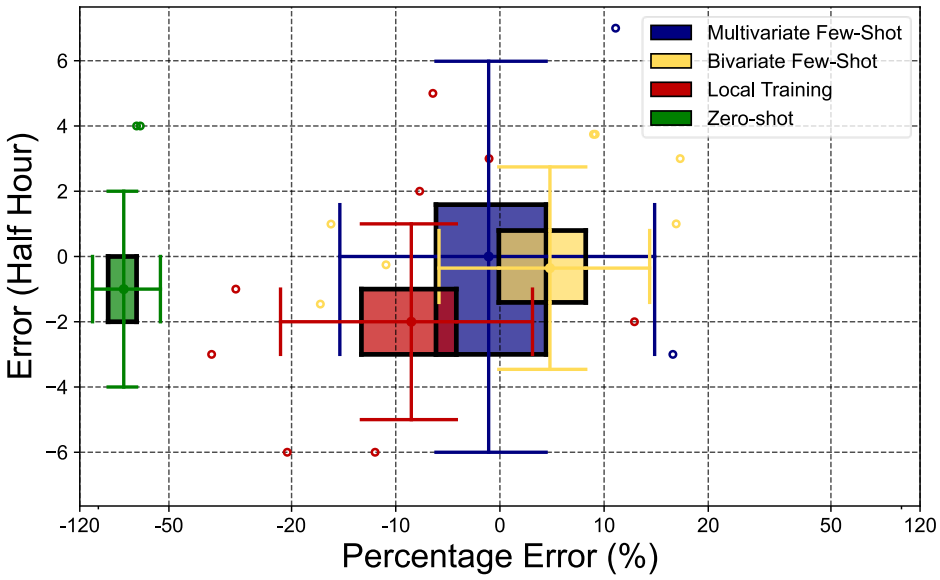

Figure 4.6 Comparison of scale (percentage error) and timing (error) for the evening peak cardinal point at secondary substation

Figure 4.6 shows the evening peak forecasting results for the four methods at the secondary substation level. When comparing the scale-based forecasts, it is clear that the zero-shot learning approach does not provide accurate predictions at the cardinal points. This issue likely arises from missing data in the substation customer dataset, such as information on PV systems, EV charging, or farming equipment. In contrast, the few-shot learning methods consistently outperform local training, while zero-shot learning remains the least effective. This gap in performance suggests that, with only one year of available data, local training tends to bias the model towards patterns seen in the training set, which then leads to underestimation on the test set. Few-shot learning methods, however, help to reduce this bias. The bivariate few-shot method shows much less bias, and the multivariate few-shot method achieves nearly zero bias.

Turning to the time forecasts, the zero-shot learning approach performs comparatively well. This may be because substations in the same area tend to have similar human activity patterns, making zero-shot learning a reasonable baseline model. The multivariate method achieves a mean error that is close to zero, but with greater variability, possibly because it responds more to scaling factors. The bivariate method, in comparison, has lower variability. Local training continues to show bias errors, probably because it fits too closely to the training data.

In summary, the multivariate few-shot method produces a wider range of errors, indicating higher variability, yet its median error remains close to zero, which suggests strong central accuracy. In contrast, both the local training and bivariate methods have smaller error ranges but are more biased, as shown by their median errors being shifted away from zero.

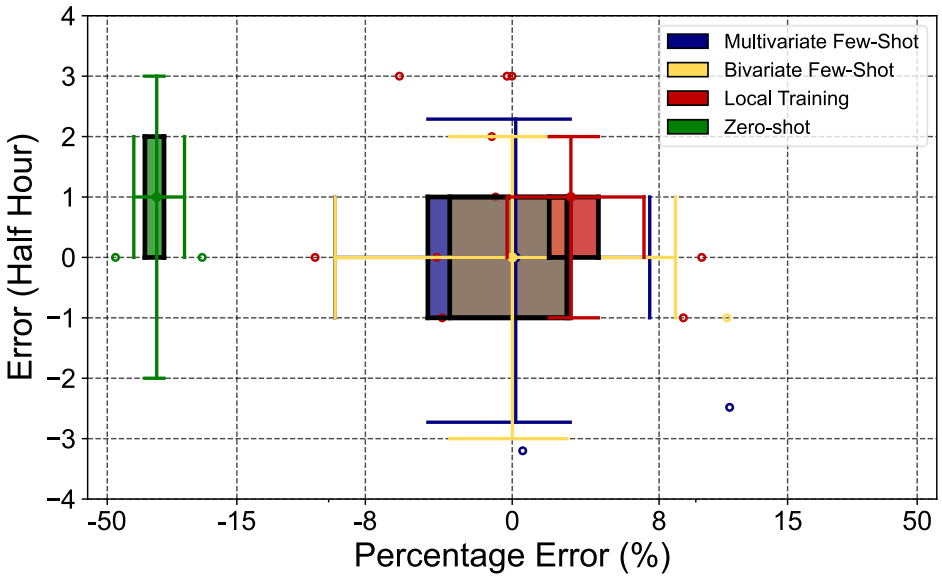

Figure 4.7 Comparison of scale (percentage error) and timing (error) for the evening peak cardinal point at primary substation

Figure 4.7 shows the evening peak forecasting results for the cardinal point at the urban primary substation. In comparison with Figure 4.6, it is evident that the urban primary substation consistently outperforms the rural secondary substation across all four approaches, with noticeably lower interquartile ranges, which indicates greater consistency and fewer extreme fluctuations. The zero-shot learning method still yields poor results. Its inability to produce accurate forecasts at key points for primary substations can be traced to missing data in the substation dataset, such as information on farms or schools.

At the primary substation level, the locally trained model exhibits much less forecast deviation, suggesting that aggregate effects make the forecasting task less complex. However, it retains a consistent bias by persistently overestimating the outcomes. On the other hand, the two few-shot learning methods produce nearly identical results. This suggests that the evening peak at primary substations depends far more on the previous day's evening peak than on the other two cardinal

points. While the few-shot approach greatly reduces forecast bias, it shows increased deviation, which may be a result of differences between the observed and target values.

In terms of time-based forecast error, the locally trained model achieves the smallest deviation, typically within half an hour. Even so, the few-shot learning methods also perform well, with most forecast errors remaining within one hour, demonstrating the effectiveness of the few-shot approach. Overall, at the primary substation level, few-shot learning remains capable of generating usable forecasts even with limited observations. This makes it a practical option in situations where data are scarce, although its overall accuracy is still not quite as high as that of the local training method.

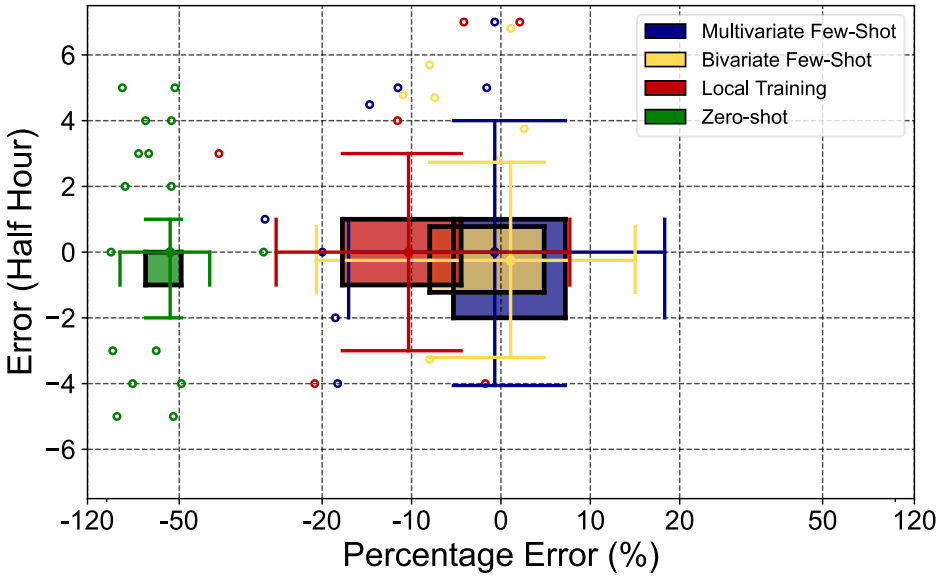

Figure 4.8 Comparison of scale (percentage error) and timing (error) for the morning peak cardinal point at secondary substation

Figure 4.8 presents the morning peak scale forecasts for secondary substations, as produced by the four evaluated methods. The results show that zero-shot learning still lacks the accuracy needed for precise scale forecasting. Additionally, the forecasting errors for the morning peak are much higher than those seen during the evening peak. This suggests that morning demand is more influenced by external factors, such as weather or sunlight duration. As a result, a more detailed division of cardinal points may be required to better identify and address this bias.

In contrast, the multivariate few-shot learning method does not display this type of bias,

highlighting the value of including all cardinal points from the previous day in the forecasting process. The results also indicate that local training continues to struggle with overfitting, repeatedly underestimating forecasts. However, the few-shot learning approach helps to overcome this issue by drawing on observations from other substations. This clearly demonstrates the advantage of using few-shot learning techniques for achieving more accurate forecasting.

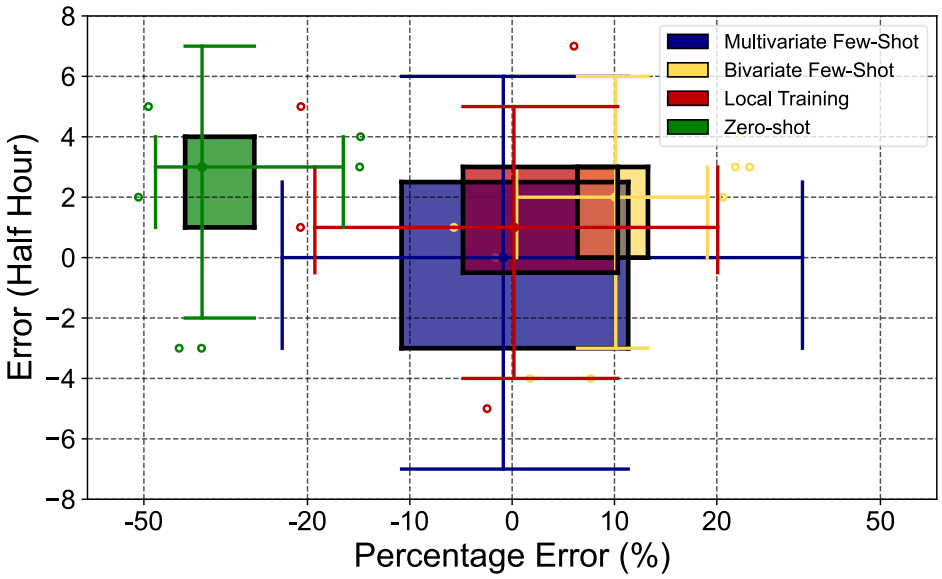

Figure 4.9 Comparison of scale (percentage error) and timing (error) for the morning peak cardinal point at primary substation

Figure 4.9 presents the forecasting results for the morning peak cardinal point at primary substations, utilising four different methodologies. Notably, during the morning peak, secondary substations exhibit superior performance compared to primary substations in both time-based and scale-based forecasts. This discrepancy may be attributed to the distinctive electricity consumption patterns associated with certain sites, such as schools or parks, which tend to be more active in the morning. However, these types of buildings are not included in the global model due to the unavailability of relevant data. The presence of such usage patterns can introduce additional complexity into the forecasting models for primary substations.

Furthermore, the zero-shot learning approach demonstrates robust performance in temporal forecasting, with the majority of errors confined to within one hour. This indicates a high degree

of reliability, a characteristic of particular significance in operational contexts, where accurate and timely forecasts are essential to support effective planning and management of electricity networks.

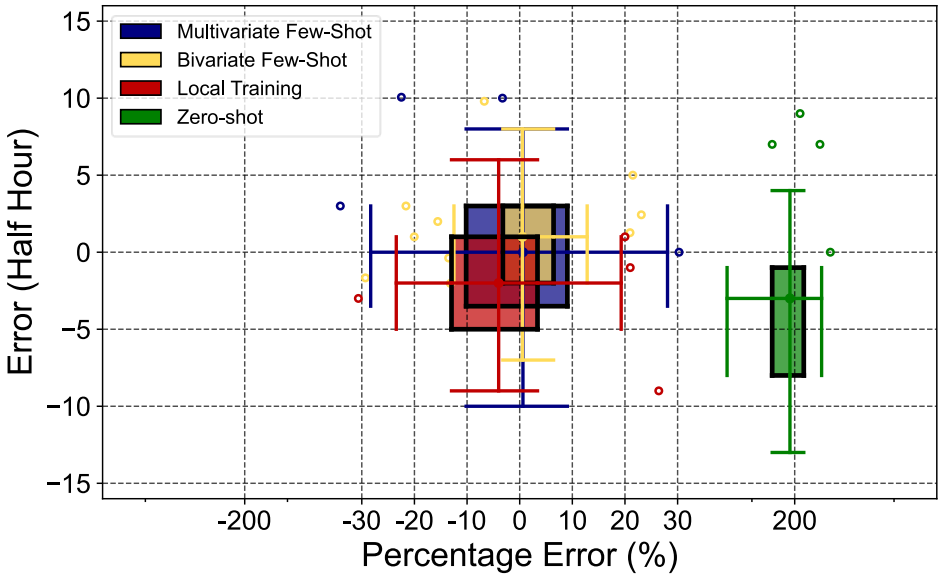

Figure 4.10 Comparison of scale (percentage error) and timing (error) for the daytime trough cardinal point at secondary substation

Figure 4.10 presents the forecasting results for the daytime trough cardinal point at secondary substations, evaluated using four different methodologies. It is evident that forecasting the daytime trough is considerably more challenging than predicting the evening or morning peaks. Specifically, the average scale error for both local training and few-shot learning exceeds 10%, with the bivariate few-shot learning approach achieving the lowest error. This suggests that the daytime trough is more strongly correlated with the previous day's daytime trough, whereas incorporating data from other cardinal points may introduce noise and mislead the forecasting model.

In terms of time-based forecasting, the results for the daytime trough are also the least accurate. In particular, the zero-shot learning method proves unreliable, with most forecasts underestimating the trough by more than two and a half hours. This highlights the inadequacy of the global model for forecasting the daytime trough. In practice, the occurrence of the trough in each individual area may be determined by the behaviours of a minority of people or specific commercial activities, which further complicates the forecasting task. Therefore, an individualised model may be

necessary to achieve more accurate results, or it may be beneficial to consider larger aggregation areas rather than focusing solely on the 11/0.415 kV level.

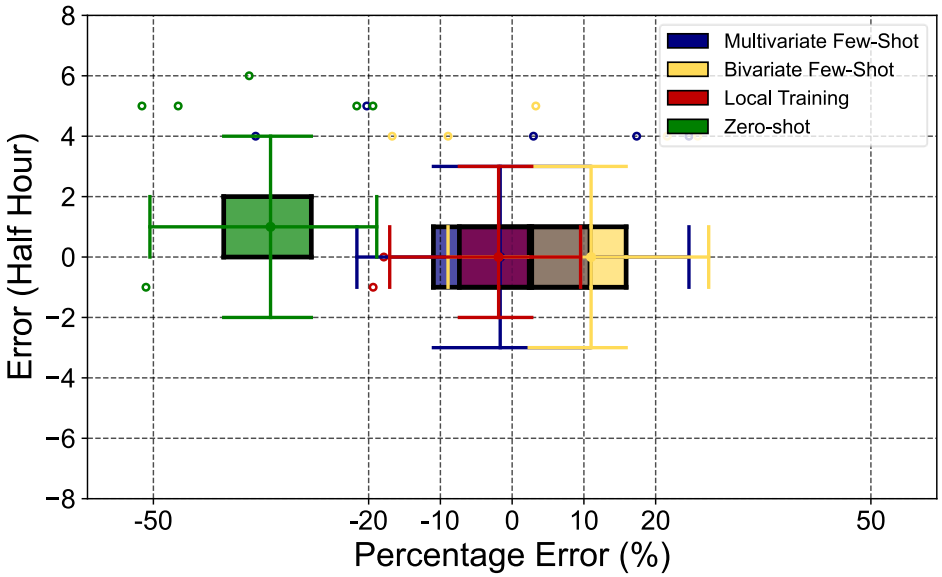

Figure 4.11 Comparison of scale (percentage error) and timing (error) for the daytime trough cardinal point at primary substation

Figure 4.11 presents the forecasting results for the daytime trough cardinal point at primary substations, evaluated using four different methodologies. It is evident that, with respect to the timing of the cardinal point, the forecasting performance at the primary substation level is significantly improved compared to that at the secondary substation level; most errors in time forecasting are now confined to within one hour, and the zero-shot learning approach demonstrates particularly strong results. This suggests that, at the primary substation level, the occurrence of the daytime trough is comparatively stable and can be reliably predicted within a certain area.

However, regarding the scale-based forecast, although the magnitude of errors has improved compared to those observed at secondary substations, particularly with respect to reducing large deviations, the performance remains less robust than that achieved for the peak forecasts. Notably, while the forecasting models are now better at predicting the evening and morning peaks, there are still significant challenges in accurately capturing the magnitude of the daytime trough, which typically occurs between 11:00 and 15:00. This suggests that, despite improvements at the primary

substation level, further refinement of the forecasting models is necessary, particularly to address the complex patterns of load reduction during different periods of the day.

Overall, the results depicted in the four figures suggest that zero-shot learning effectively forecasts timing but struggles with accurately predicting scale. In general, primary substations tend to provide more accurate forecasts compared to secondary substations; however, morning peaks display varying patterns influenced by user categories. With one year of data, local training methods risk overfitting, resulting in biased outcomes. In contrast, few-shot learning methods effectively calibrate the data, thus avoiding bias. However, few-shot learning methods also display greater deviation due to the limited number of observations, potentially resulting in an overly confident estimation of covariance. Nevertheless, the findings remain promising: cardinal point magnitudes and times can be reliably forecasted without historical observations, achieving low error results, while scale forecasts maintain an error margin of approximately 10%.

The following tables provide a comparison of these methods across secondary substation evening peak scenarios, assessed both by scale and time for the cardinal points. The evaluation of forecast performance utilises several key metrics. The Root Mean Square Error (RMSE) is used because it places greater emphasis on larger errors, thereby highlighting significant inaccuracies. RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{k=1}^{N}(y_k - \widehat{y_k})^2}{N}} \tag{4.34}$$

The Mean Absolute Percentage Error (MAPE) is employed, as it allows for direct comparison of forecasting performance across datasets of differing scales. MAPE is given by:

$$MAPE = 100 \times \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_k - \widehat{y_k}}{y_k}\right| \tag{4.35}$$

Meanwhile, the Mean Bias Error (MBE) measures any systematic tendency of the forecasting method, indicating whether the model consistently overestimates or underestimates actual values. MBE is defined as:

$$MBE = \frac{\sum_{k=1}^{N}(y_k - \widehat{y_k})}{N} \tag{4.36}$$

In the context of forecasting at cardinal point times, MAPE is replaced by the Mean Absolute Error (MAE), since time-based measurements do not lend themselves to percentage-based interpretation. MAE is calculated as:

$$MAE = \frac{\sum_{k=1}^{N} |y_k - \widehat{y_k}|}{N} \tag{4.37}$$

This substitution ensures that the error metric remains meaningful and appropriate for temporal forecasts.

	RMSE	MAPE	MBE
Multivariate Few-shot Learning	**10.46**	**5.76%**	**0.09**
Bivariate Few-shot Learning	12.45	6.46%	6.23
Local Training	17.40	10.25%	-13.50
Zero-shot Learning	116.31	79.87%	-115.78

Table 4.1 Comparison across secondary substation evening peak scale cardinal points

	RMSE	MAE	MBE
Multivariate Few-shot Learning	2.99	2.47	-0.34
Bivariate Few-shot Learning	**1.68**	**1.38**	**-0.30**
Local Training	2.76	2.26	-1.81
Zero-shot Learning	1.76	1.41	-0.62

Table 4.2 Comparison across secondary substation evening peak time cardinal points

	RMSE	MAPE	MBE
Multivariate Few-shot Learning	10.29	7.80%	**-0.53**
Bivariate Few-shot Learning	**8.50**	**7.09%**	-0.81
Local Training	13.29	11.56%	-10.47
Zero-shot Learning	72.29	70.56%	-71.71

Table 4.3 Comparison across secondary substation morning peak scale cardinal points

	RMSE	MAE	MBE
Multivariate Few-shot Learning	2.30	1.71	**0.12**
Bivariate Few-shot Learning	**2.07**	**1.45**	0.17
Local Training	2.47	1.75	0.28

Zero-shot Learning	2.12	1.54	-0.28

Table 4.4 Comparison across secondary substation morning peak time cardinal points

	RMSE	MAPE	MBE
Multivariate Few-shot Learning	8.98	11.46%	**0.54**
Bivariate Few-shot Learning	8.06	10.73%	3.84
Local Training	**7.98**	**10.67%**	-2.08
Zero-shot Learning	286.98	480.80%	286.89

Table 4.5 Comparison across secondary substation daytime trough scale cardinal points

	RMSE	MAE	MBE
Multivariate Few-shot Learning	4.80	3.78	**-0.14**
Bivariate Few-shot Learning	**3.75**	**3.13**	0.65
Local Training	4.19	3.35	-1.92
Zero-shot Learning	5.96	4.85	-3.60

Table 4.6 Comparison across secondary substation daytime trough time cardinal points

Tables 4.1 to 4.6 display the forecast performance at the three cardinal points for secondary substations. In terms of scale-based forecasts, the results indicate that few-shot learning delivers the most effective outcomes in local models for both evening and morning peak periods. These results also emphasises the intrinsic challenges involved in forecasting secondary substation loads with high precision, even when comprehensive datasets are available. The presence of persistent bias errors suggests that ongoing calibration against observed data is necessary to improve model accuracy.

Moreover, when working with data that exhibits consistent patterns, few-shot learning approaches benefit significantly from global models. This enables the models to mitigate overfitting and to refine their accuracy through cross-validation with results from other substations. On the other hand, for forecasts of load troughs, locally trained models show the best performance. This likely reflects the substantial variability in load profiles across individual substations during low-demand periods, possibly due to site-specific factors such as embedded photovoltaic (PV) generation. These observations highlight the importance of incorporating more granular cardinal points in the

analysis, to better account for the diverse operational characteristics present across different substations.

With respect to cardinal point time forecasts, zero-shot learning proves to be remarkably effective, achieving results within approximately 5 per cent of the most accurate methods, all without the need for any prior training data. This demonstrates the considerable potential of zero-shot learning as a robust option for temporal forecasting, especially in contexts where historical data is scarce or unavailable.

4.6 Conclusion

Forecasting electricity demand in distribution networks is a complex task, largely because of limited visibility and the economic impracticality of thoroughly monitoring the large number of secondary low-voltage substations. Despite these challenges, accurate forecasting is vital for DSOs, who must balance network demand against the financial constraints of cable upgrades as they support the transition to low-carbon technologies. It is equally important to predict both the timing and magnitude of substation loads at key cardinal points, as these forecasts are crucial for effective day-ahead planning and for managing both voltage and thermal constraints on the network.

In response to these needs, this chapter presents a new approach to cardinal point forecasting that requires only minimal or no local data to produce reliable predictions. The proposed methodology employs a global model that uses data collected from a range of substations. This allows the method to create an initial, attribute-based zero-shot learning forecast for a given substation. The forecast is then further refined using few-shot learning, a process that is calibrated on a small number of observations from other substations to improve both the accuracy of load magnitude and the timing of the prediction.

The results show that for forecasting the timing of cardinal points, zero-shot learning on its own can already deliver a robust baseline prediction, even if no previous observations are available. This baseline can then be improved through few-shot learning as more information becomes accessible. However, for scale forecasting, zero-shot learning on its own is not effective. In these situations, few-shot learning, combined with a global model, can significantly reduce forecasting

errors compared to models trained solely on local data at secondary substations. This approach also enables better estimation and correction of bias in the forecasts.

By relying on existing datasets, these cardinal point forecasting strategies not only help to reduce the costs of monitoring and data storage but also improve overall forecasting accuracy. This improvement is particularly important as the demand for distribution network balancing services increases in response to the widespread adoption of low-carbon technologies. By providing greater forecasting accuracy at the most critical moments, these methods enable DSOs to manage the network more effectively and achieve better outcomes during periods that have the greatest impact.

Although a range of alternative point forecasting approaches has been reported in this chapter, including weather-informed regression and ensemble methods [9], state-space models [36], and machine-learning approaches such as tree-based models [38], this chapter does not attempt an exhaustive comparison across all model families. Instead, the focus is on establishing a clear and interpretable point-forecasting baseline using the GAM framework, which aligns with the operational aim of forecasting substation cardinal points under limited observability. A broader evaluation against additional point forecasting models will be pursued in future work to further quantify the performance trade-offs under different data availability conditions.

Future work will concentrate on refining the use of additional cardinal points to develop a more detailed understanding of load profiles on distribution networks. By introducing more cardinal points during the morning and midday periods, it may become possible to gain a clearer insight into the influence of LCTs and the effects of dynamic tariff responses on demand throughout the day.

4.7 References

[1] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM computing surveys (csur),* vol. 53, no. 3, pp. 1-34, 2020.

[2] T. Hong, P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy Forecasting: A Review and Outlook," *IEEE Open Access Journal of Power and Energy,* vol. 7, pp. 376-388, 2020, doi: 10.1109/OAJPE.2020.3029979.

[3] NESO, "An Introduction to System Operator Plans," 2025/06/25 2025. [Online]. Available: https://www.neso.energy/data-portal/system-operating-plan-sop/an_introduction_to_system_operator_plans.

[4] National Grid, "Forward Plan Technical Annexe," 2025/06/25 2018. [Online]. Available: https://www.nationalgrid.com/sites/default/files/documents/P2209_forward_plan_technical_annexe_1202_2.pdf.

[5] C. Madden, A. R. Ahmadi, and J. West, "Open Networks Project – Operational Forecasting Report," 2025/06/26 2020. [Online]. Available: https://www.energynetworks.org/assets/images/ON20-WS1B-P3%20Operational%20Forecasting-PUBLISHED.23.12.20.pdf.

[6] J. Caplin, "Demand Forecasting," 2025/06/25 2017. [Online]. Available: https://bpb-eu-w2.wpmucdn.com/blogs.reading.ac.uk/dist/f/83/files/2017/04/1.1_J_Caplin_National_demand_forecast.pdf.

[7] J. W. Taylor, "An evaluation of methods for very short-term load forecasting using minute-by-minute British data," *International Journal of Forecasting,* vol. 24, no. 4, pp. 645-658, 2008/10/01/ 2008, doi: https://doi.org/10.1016/j.ijforecast.2008.07.007.

[8] A. B. Baker, D. Bunn, and E. Farmer, "Load forecasting for scheduling generation on a large interconnected system," in *Comparative models for electrical load forecasting*: Wiley, 1985, pp. 57-67.

[9] J. W. Taylor and R. Buizza, "Using weather ensemble predictions in electricity demand forecasting," *International Journal of Forecasting,* vol. 19, no. 1, pp. 57-70, 2003/01/01/ 2003, doi: https://doi.org/10.1016/S0169-2070(01)00123-6.

[10] J. W. Taylor and S. Majithia, "Using combined forecasts with changing weights for electricity demand profiling," *Journal of the Operational Research Society,* vol. 51, no. 1, pp. 72-82, 2000.

[11] K. Gram-Hanssen, "Understanding change and continuity in residential energy consumption," *Journal of Consumer Culture,* vol. 11, no. 1, pp. 61-78, 2011/03/01 2011, doi: 10.1177/1469540510391725.

[12] E. Shove and G. Walker, "What is energy for? Social practice and energy demand," *Theory, culture & society,* vol. 31, no. 5, pp. 41-58, 2014.

[13] E. Kyritsis, J. Andersson, and A. Serletis, "Electricity prices, large-scale renewable integration, and policy implications," *Energy Policy,* vol. 101, pp. 550-560, 2017.

[14] S. Haben, C. Singleton, and P. Grindrod, "Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data," *IEEE Transactions on Smart Grid,* vol. 7, no. 1, pp. 136-144, 2016, doi: 10.1109/TSG.2015.2409786.

[15] R. Li, C. Gu, F. Li, G. Shaddick, and M. Dale, "Development of Low Voltage Network Templates—Part I: Substation Clustering and Classification," *IEEE Transactions on Power Systems,* vol. 30, no. 6, pp. 3036-3044, 2015, doi: 10.1109/TPWRS.2014.2371474.

[16] J. Kotia, A. Kotwal, R. Bharti, and R. Mangrulkar, "Few shot learning for medical imaging," in *Machine learning algorithms for industrial applications*: Springer, 2020, pp. 107-132.

[17] Z. Zhang, D. Cai, C. C. Ukwuoma, and Q. Huang, "A few-shot learning model for renewable energy transmission fault diagnosis in tunnels," *Electric Power Systems Research,* vol. 247, p. 111798, 2025.

[18] X. U. Yang, F. A. N. Yunlei, B. A. O. Yuequan, and L. I. Hui, "Few-shot learning for structural health diagnosis of civil infrastructure," *Advanced Engineering Informatics,* vol. 62, p. 102650, 2024.

[19] A. Parnami and M. Lee, "Learning from few examples: A summary of approaches to few-shot learning," *arXiv preprint arXiv:2203.04291,* 2022.

[20] B. Hariharan and R. Girshick, "Low-shot visual recognition by shrinking and hallucinating features," 2017, pp. 3018-3027.

[21] H. Gao, Z. Shou, A. Zareian, H. Zhang, and S.-F. Chang, "Low-shot learning via covariance-preserving adversarial augmentation networks," *Advances in Neural Information Processing Systems,* vol. 31, 2018.

[22] W. Yan, J. Yap, and G. Mori, "Multi-task transfer methods to improve one-shot learning for multimedia event detection," 2015, pp. 37-1.

[23] M. Vartak, A. Thiagarajan, C. Miranda, J. Bratman, and H. Larochelle, "A meta-learning perspective on cold-start recommendations for items," *Advances in neural information processing systems,* vol. 30, 2017.

[24] S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, and L. Van Gool, "One-shot video object segmentation," 2017, pp. 221-230.

[25] S. Wang, X. Fu, K. Ding, C. Chen, H. Chen, and J. Li, "Federated few-shot learning," 2023, pp. 2374-2385.

[26] A. Li, T. Luo, T. Xiang, W. Huang, and L. Wang, "Few-shot learning with global class representations," 2019, pp. 9715-9724.

[27] P. Montero-Manso and R. J. Hyndman, "Principles and algorithms for forecasting groups of time series: Locality and globality," *International Journal of Forecasting,* vol. 37, no. 4, pp. 1632-1653, 2021/10/01/ 2021, doi: https://doi.org/10.1016/j.ijforecast.2021.03.004.

[28] A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski, "Using Bayesian model averaging to calibrate forecast ensembles," *Monthly weather review,* vol. 133, no. 5, pp. 1155-1174, 2005.

[29] L. J. Wilson and M. Vallée, "The Canadian updateable model output statistics (UMOS) system: Design and development tests," *Weather and forecasting,* vol. 17, no. 2, pp. 206-222, 2002.

[30] Q. Duan *et al.*, "Automatic model calibration: A new way to improve numerical weather forecasting," *Bulletin of the American Meteorological Society,* vol. 98, no. 5, pp. 959-970, 2017.

[31] T. Gneiting, F. Balabdaoui, and A. E. Raftery, "Probabilistic forecasts, calibration and sharpness," *Journal of the Royal Statistical Society Series B: Statistical Methodology,* vol. 69, no. 2, pp. 243-268, 2007.

[32] T. Gneiting, *Calibration of medium-range weather forecasts*. European Centre for Medium-Range Weather Forecasts Reading, UK, 2014.

[33] F. Pourpanah *et al.*, "A Review of Generalized Zero-Shot Learning Methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 45, no. 4, pp. 4051-4070, 2023, doi: 10.1109/TPAMI.2022.3191696.

[34] C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-Based Classification for Zero-Shot Visual Object Categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 36, no. 3, pp. 453-465, 2014, doi: 10.1109/TPAMI.2013.140.

[35] D. Huynh and E. Elhamifar, "Fine-grained generalized zero-shot learning via dense attribute-based attention," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 4483-4493.

[36] J. D. Vilmarest and Y. Goude, "State-Space Models for Online Post-Covid Electricity Load Forecasting Competition," *IEEE Open Access Journal of Power and Energy,* vol. 9, pp. 192-201, 2022, doi: 10.1109/OAJPE.2022.3141883.

[37] H. Inoue, "A least-squares smooth fitting for irregularly spaced data: Finite-element approach using the cubic B-spline basis," *Geophysics,* vol. 51, no. 11, pp. 2051-2066, 1986.

[38] L. Semmelmann, O. Resch, S. Henni, and C. Weinhardt, "Privacy‐preserving peak time forecasting with Learning to Rank XGBoost and extensive feature engineering," *IET Smart Grid,* vol. 7, no. 2, pp. 172-185, 2024.

[39] S. Gill, B. Stephen, and S. Galloway, "Wind Turbine Condition Assessment Through Power Curve Copula Modeling," *IEEE Transactions on Sustainable Energy,* vol. 3, no. 1, pp. 94-101, 2012, doi: 10.1109/TSTE.2011.2167164.

[40] M. Sun, I. Konstantelos, and G. Strbac, "C-Vine Copula Mixture Model for Clustering of Residential Electrical Load Pattern Data," *IEEE Transactions on Power Systems,* vol. 32, no. 3, pp. 2382-2393, 2017, doi: 10.1109/TPWRS.2016.2614366.

[41] F. Durante and C. Sempi, *Principles of copula theory*. CRC press Boca Raton, FL, 2016.

[42] S. Coles, J. Bawa, L. Trenner, and P. Dorazio, *An introduction to statistical modeling of extreme values*. Springer, 2001.

[43] G. J. Crane and J. v. d. Hoek, "Conditional expectation formulae for copulas," *Australian & New Zealand Journal of Statistics,* vol. 50, no. 1, pp. 53-67, 2008.

[44] B. Stephen, R. Telford, and S. Galloway, "Non-Gaussian Residual Based Short Term Load Forecast Adjustment for Distribution Feeders," *IEEE Access,* vol. 8, pp. 10731-10741, 2020, doi: 10.1109/ACCESS.2020.2965320.

[45] P. Berkes, F. Wood, and J. Pillow, "Characterizing neural dependencies with copula models," *Advances in neural information processing systems,* vol. 21, 2008.

5 Future Operational Consideration in GB Distribution Networks

This chapter builds on the theoretical and methodological foundations of transfer learning and few-shot learning methods established in earlier chapters, this chapter shifts the focus to the practical application of advanced forecasting and modelling techniques within the operational environment of GB's distribution networks. As the industry faces growing pressures from the integration of low-carbon technologies, decentralised energy resources, and changing patterns of demand, DSOs must adopt innovative approaches to ensure continued reliability and efficiency. This chapter explores how the methodologies developed throughout this thesis, such as few-shot learning for cardinal point forecasting and transfer learning, can be implemented in real-world network operations for voltage control in the peak-load scenario, as well as for providing an overall assessment of network voltage and current performance with minimal observation. It discusses their potential role in industry projects and in the day-to-day decision-making of DSOs. The discussion begins with an overview of the Distribution Future Energy Scenarios (DFES) process, which provides a strategic framework for considering how distribution networks may evolve over the coming decades. Subsequent sections illustrate the use of data-driven approaches and modelling techniques to address emerging challenges and emphasise the importance of modelling the distribution network using actual network data and available GIS data, in line with the principles set out by the DFES. Both radial and mesh network topologies are analysed to provide a comprehensive understanding of different network configurations. The chapter then assesses the impact of future low-carbon technology (LCT) integration through case studies and industry examples. Finally, opportunities for deploying transfer learning and few-shot learning within network operation are highlighted, demonstrating how these innovations can support more adaptive, resilient, and forward-looking distribution systems.

Overall, this chapter demonstrates the practical significance and industry relevance of previous academic research by addressing a key prerequisite for real-world application: building a distribution network model that reflects the true electrical connectivity rather than only the mapped geometry. The contribution of this chapter is therefore not simply the construction of a GIS-based model, but the development of a topology reconstruction workflow that identifies and corrects

common connectivity issues so that the resulting network more accurately represents how the system is electrically organised. This provides a clearer and more defensible pathway for translating academic advances into improvements within the GB distribution network, particularly in contexts where asset records may be incomplete or inconsistently represented in GIS. The reconstructed network model can be directly applied to practical use cases, including studies of LCT integration such as PV panels and the low-data load forecasting methods presented in Chapters 3 and 4. Formal validation and LV state estimation using the reconstructed model are next steps and are identified as future work to further quantify performance and strengthen operational confidence.

5.1 Distribution Future Energy Scenarios (DFES) in GB Distribution Network

The DFES is a strategic initiative undertaken by each distribution network license holder in GB. Its primary purpose is to examine how electricity demand, generation, and the adoption of low-carbon technologies may develop across regional networks over the coming decades, typically with projections extending to 2050. Within the DFES, DSOs present a variety of plausible future scenarios. These encompass the increased uptake of electric vehicles and associated charging infrastructure, the expansion of heat pumps and electrified heating, the growth of local renewable generation and battery storage, as well as changes in traditional patterns of electricity demand.

Each scenario functions as an analytical tool to evaluate how distribution networks might be affected, and to determine what forms of investment, operational change, or technological innovation may be necessary to ensure a secure, reliable, and cost-efficient electricity supply as the GB pursues decarbonisation. The national energy regulator, Ofgem, requires each DNO to publish a DFES [1]. This document supports business planning and provides evidence to justify investment in the distribution network. Every DSO, including SPEN [2], [3], SSEN [4], [5], Northern PowerGrid [6] and others, operates within a specific region. Each region is characterised by a distinct mix of customers, industrial activities, climate conditions, and local policy priorities.

In addition, some local authorities have established net zero targets that are earlier than the national objectives or have ambitions for the electrification of transport or heating. The characteristics of distribution networks, including urban or rural settings, the level of renewable integration, and the strength of the grid, also differ significantly between licence areas.

A detailed analysis of the Distribution Future Energy Scenarios for SP Distribution is presented here [2]. SP Energy Networks' SP Distribution license covers Central and Southern Scotland, supplying electricity to approximately two million homes and businesses through around 390 primary substations. The forecasting methodology adopts a bottom-up approach, with forecasts developed for each of the 390 primary substations and 88 Grid Supply Points. These forecasts are rigorously stress-tested annually, incorporating extensive stakeholder engagement.

Key findings from the report in DFES in SP Distribution 2024 include the following:

- Doubling of Demand: Peak demand on the distribution network is projected to double by 2050, even after accounting for demand-side flexibility.
- Significant Growth in Low Carbon Technologies: The report anticipates a substantial increase in low carbon technology (LCT) devices over the long term. Specifically, the number of connected distributed generation and storage devices is expected to increase fivefold over the same period.
- Eight Million Low Carbon Devices: By 2050, up to eight million electric vehicles and heat pumps may be connected to the network.
- Expansion of distribution network capacity: Scenario studies indicate that flexibility requirements and connected generation capacity could reach around seven times their current levels by 2030 and up to twenty times by 2050. This massive influx, driven by DER, will necessitate substantial upgrades to the firm capacity of cables, feeders, and transformers to prevent overloads [2].

To ensure a secure and reliable electricity supply whilst facilitating decarbonisation, SP Distribution identifies two main priorities:

- Timely Capacity Upgrades: Implementing upgrades where demand and generation hotspots emerge. To ensure a secure and reliable electricity supply whilst facilitating decarbonisation, SP Distribution identifies two main priorities:
- Digitalisation and Real-Time Control: Enhancing digital infrastructure and real-time control to enable the dispatch of thousands of flexibility contracts.

To minimise the cost of capacity upgrades whilst ensuring the network remains stable and reliable, digitalisation is being considered as a practical solution. In particular, SP Distribution has implemented the RIIO-ED2 Digitalisation Strategy [7], a transformative plan designed to support the transition to a modern, data-driven, and customer-focused energy network.

These developments are driven by the rapidly changing landscape facing electricity distribution networks. DFES predict significant transformation of the network in the coming years. This transformation is expected to lead to a substantial increase in both the volume of data generated and the complexity of network operations. In response, there is an increasing need to introduce advanced analytics, making network planning, connections, maintenance, and control more dynamic and responsive. It is crucial that network data can be flexibly analysed to assess both long-term scenarios and near-real-time operational states.

The first major step in the digitisation of the network is to create a connected network view using one GB DNO GIS data. This foundational process leverages GIS data and systematically maps it to electric network models, thereby enabling a comprehensive digital representation of the physical infrastructure. Achieving this has required the identification and collation of all relevant GIS asset data, followed by an extensive data cleansing exercise to enhance the quality and accuracy of the dataset, with rigorous practical testing. Such thorough data preparation ensures that the resulting network model is both reliable and robust, providing a solid basis for subsequent analysis and operational decision-making.

Section 2 will present the methodology for constructing a real-world distribution network model from GIS data, considering both radial and mesh network configurations. It will also address the unique challenges encountered in this process, particularly in contrast to the more standardised IEEE test networks, which are often used in academic studies but do not fully capture the complexities of actual operational environments.

Section 3 will then explore the prospective integration of LCTs within the distribution network. This section will detail the necessary steps and considerations for future-proofing the network in response to the increasing penetration of LCTs, such as electric vehicles and distributed generation.

Finally, section 4 will demonstrate the practical application of the methodologies outlined in Chapters 2 and 3 to the distribution network model, showcasing how these approaches can be

implemented within a real-world GIS network environment to support advanced network management and planning in situations with little or no monitoring.

5.2 Modelling and data-driven approach in Distribution Network and its challenges

In the context of distribution network modelling and testing, the most common approach is to utilise the IEEE test networks, as they encompass a comprehensive range of voltage levels and can be selected according to specific modelling requirements. Consequently, to evaluate and validate modification in distribution system design, a number of IEEE standard test feeders such as the IEEE 123-bus, IEEE 34-bus, and IEEE 13-bus networks have been developed to represent different types of networks [8]. These test feeders are widely used within the research community. However, these feeders are essentially static, single-snapshot models that generally lack real time-series measurements. This limitation significantly constrains their application for quasi-static time-series analysis of distribution systems, where it is necessary to capture the time-varying nature of load behaviours [9].

Moreover, there are inherent structural and operational differences between distribution systems in the United States and Europe. For example, in North America, a secondary transformer rated between 10 kVA and 50 kVA typically serves between one and ten customers via radial triplex cables [10]. In contrast, in Europe, a secondary transformer is usually rated between 100 kVA and 1,000 kVA and serves between fifty and over two hundred and fifty customers through a networked secondary system [10]. As a result, the IEEE test networks, which are primarily based on NA distribution practices, do not accurately represent the structure and operational characteristics of distribution networks in GB, where the distribution level differs significantly.

While these test networks are suitable for hypothetical and design-based studies, such as investigating whether designed mid-feeder compensation provides adequate dynamic voltage control after disturbances [11] or assessing the operational viability of converters within distribution networks [12], they are not sufficient for the day-to-day analytical requirements of DSOs. Each real-world network possesses unique characteristics and operational constraints, making standard test feeders overly simplistic and unrepresentative. Therefore, the development of GIS-based models derived from actual distribution network data is essential to realistically

represent the implications of substantial increases in demand and to support more accurate operational analysis.

In the GB distribution network, there are two types of distribution systems: radial network and mesh network. A radial network is a straightforward configuration where each customer is supplied via a single path from the source transformer. This arrangement is cost-effective and easy to manage, but it does have its drawbacks: any fault along the line can result in a loss of supply for all customers downstream of the fault. A mesh network also known as a ring, on the other hand, incorporates multiple interconnections between different points in the system as well as connections to the MV network. This set-up enables electricity to flow along various routes, significantly enhancing the reliability and flexibility of supply. If there is a fault on one path, electricity can usually be rerouted through alternative routes, minimising disruption and improving the resilience of the network.

Here, the network model is created using OpenDSS (Open Distribution System Simulator), an open-source, scriptable power system simulator developed by the Electric Power Research Institute (EPRI) for simulating electrical distribution systems [13]. OpenDSS focuses on steady-state and time-series analysis of distribution networks, and excels because it natively handles unbalanced, multi-phase systems, large numbers of distributed energy resources, and detailed load and voltage control modelling. Unlike the standard IEEE test network, which each parameter is well designed and correct, the GIS data from the DSOs often has data quality issue and not ready for direct network creation, with one of the major issues being connection problem between cables and loads [14]. Below, a detailed process for a 33 to 11 kV network will be presented, along with a test of the given 11 kV to 415 V network.

Figure 5.1 illustrates a segment of the 33 kV to 11 kV network, encompassing one 33 kV primary substation and 56 11 kV secondary substations. Approximately 1,400 straight-line segments, extracted as linestrings (a type of geometric object in GIS that stores a sequence of coordinate pairs to represent a path on a map) from the raw geographical data, are used to interconnect these substations. While the network appears generally satisfactory when viewed at a macroscopic level, the majority of loads and transformers are not correctly connected to their respective lines, and many lines themselves are not accurately linked. This issue reflects the results observed in the Electricity North West dataset, as reported by the University of Manchester [14] , and highlights

the necessity for network reconnection. For the benefit of the DSO, all coordinates have been processed and shifted so that no actual locational data is exposed; this ensures that no practical, sensitive information is shown in this figure.

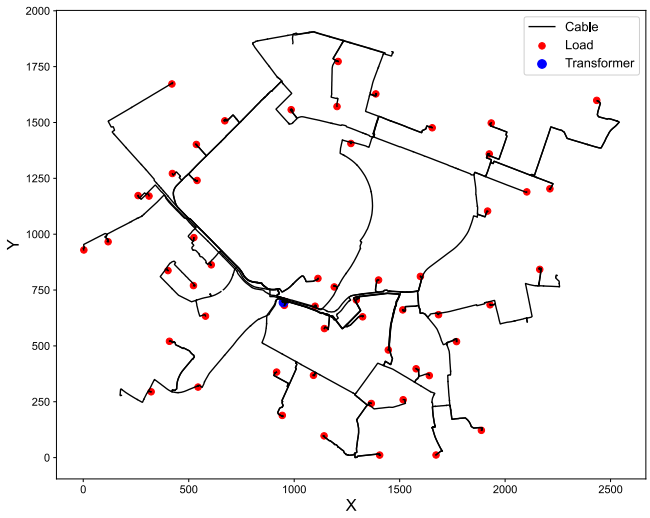

Figure 5.1 Spatial layout of a 33/11 kV distribution network in GB urban area based on raw geographical data

A straightforward approach to establishing network connectivity is to link each node to its nearest neighbour or, alternatively, to define a small circular region around each node and connect it to all points in proximate. Many existing methods employ algorithms such as breadth-first search (BFS) to solve the connection problem [15]; however, a key limitation of these approaches is that they focus primarily on achieving connectivity, rather than ensuring that the resulting connections accurately reflect the true electrical relationships within the network.

In practice, the allocation of loads within power distribution networks is governed by factors that are considerably more complex than simple geographic proximity. While connecting each load to its closest node may offer mathematical convenience, real-world network connectivity is shaped by the actual topology of the system, the operational requirements, and the historical development of the network infrastructure. As a result, straight-line distance alone is insufficient to capture the intricacies of practical distribution network design.

With regard to physical connections, distribution lines are seldom constructed to minimise physical distance; instead, their routing is influenced by the layout of roads, the need to circumvent physical

obstacles, and various operational constraints. Similarly, loads are typically connected to designated substations, which are themselves interconnected according to a rigorously defined network structure, with assignments based on network configuration rather than simple spatial proximity. Furthermore, each load may be allocated to a particular feeder for reasons of load balancing, network reliability, or operational efficiency, rather than solely for geographical convenience.

Therefore, in this context, an approach based on distinct feeder IDs and substation design rule is proposed. In a 33/11 kV distribution network, feeders are typically high-capacity cables that distribute power from the primary substation at 33 kV to the secondary substations at 11 kV, serving as the main pathways for electrical energy flow. Beyond their physical form, feeders play a strategic role in defining the network topology, as their configuration determines how substations and loads are interconnected. In this network, there are 14 different feeder IDs.

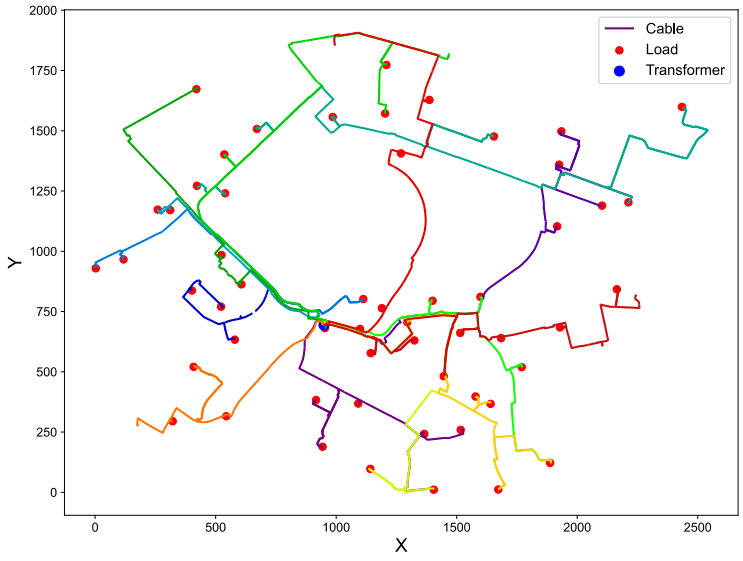

Figure 5.2 Feeder-level visualisation of a GB 33/11 kV distribution network: cables coloured by Feeder ID

Figure 5.2 presents the distribution network topology with a spatial mapping of feeders, each distinguished by a unique feeder ID. It is evident that multiple feeders can traverse the same geographical corridor and may even overlap visually in the figure. These represent parallel cables. When the GIS data and network model indicate that two feeders with different IDs occupy the

same route, this typically signifies the presence of two physically separate cables running in parallel along that path. Figure 5.3 also distinguishes individual feeders and transformers within the network.

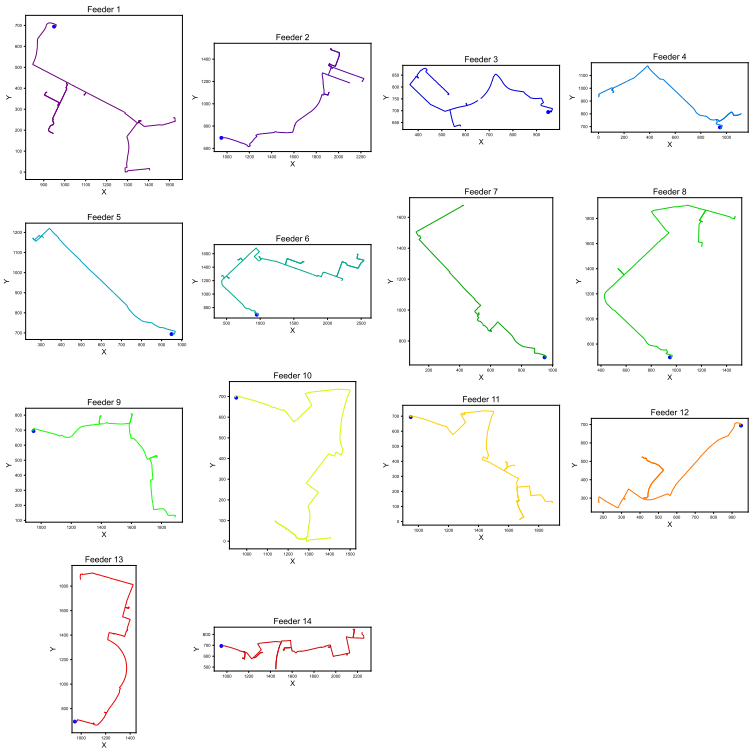

Figure 5.3 Layout of individual feeders and transformer connections in GB 33/11 kV distribution network

Figure 5.3 demonstrates that, within this network, each feeder is connected individually to the transformer. Consequently, a method has been developed that determines feeder connections based on the reconnection of each end node. This approach contrasts with the existing method proposed by the University of Manchester [14] , which utilises the Breadth-First Search (BFS) algorithm to explore all neighbouring nodes and establish complete connectivity, assuming uniform network design throughout.

The objective of the 33/11 kV network model developed in this chapter is to produce a realistic, yet computationally tractable, representation of the SPEN primary network that can be used for time-series power flow simulations and to provide consistent boundary conditions for the LV substation models in later chapters. Importantly, the aim is not only to reconnect missing cables,

but also to reconstruct a realistic set of cable connections that is consistent with known DNO design rules, using the available asset records together with engineering knowledge of typical feeder and substation configurations, rather than creating arbitrary connections simply to achieve network connectivity.

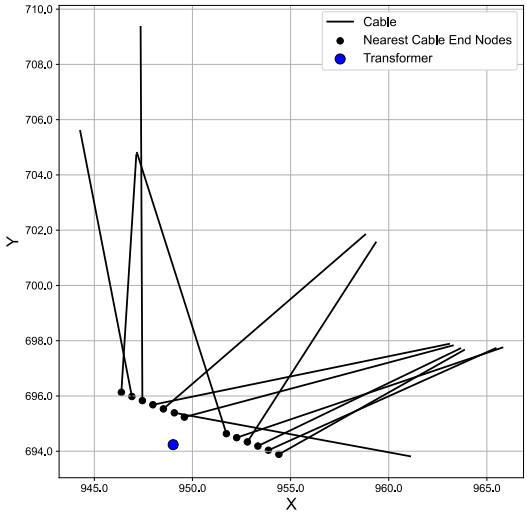

Figure 5.4 Detailed layout of fourteen nearest cable end nodes in relation to transformer

Figure 5.4 shows a detailed map of the nearest cable end nodes and the 33 kV transformer. The figure indicates that the cable ends are not connected to the transformer. The number of nearest end nodes matches the total number of feeders operating with this transformer. It can be concluded that, for transformer connection, an additional cable is required for each feeder to connect it to the transformer. At the same time, the cable ends should not be connected to each other, as this would change the network topology to a mesh network, potentially causing issues with power flow.

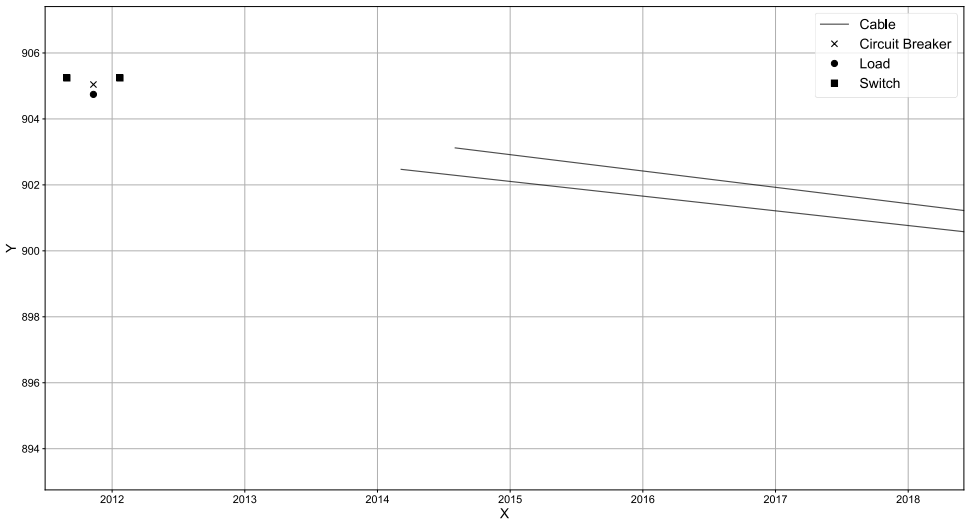

Figure 5.5 Layout 1 – schematic representation of cable, load, switch, and circuit breaker

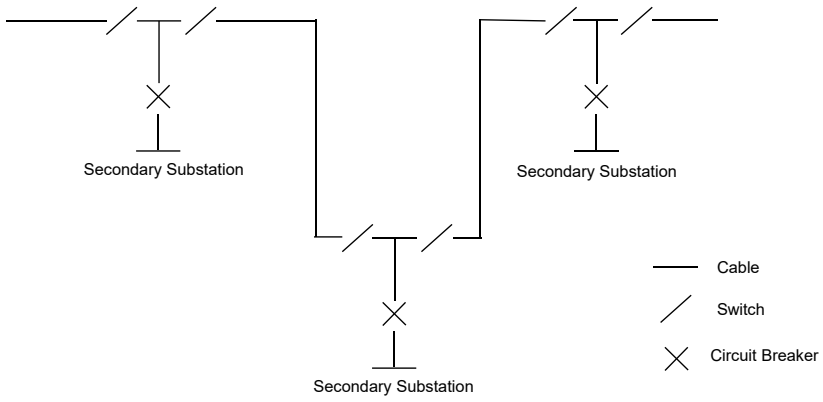

Figure 5.6 Representative RMU-based network configuration as depicted in SPEN documentation [16]

Figures 5.5 and 5.6 illustrate a typical design for an RMU-based distribution network. A RMU (Ring Main Unit) is a compact, metal-enclosed switchgear unit intended for medium-voltage distribution networks. It typically comprises several switch disconnectors and either a circuit breaker or fuse for transformer protection. The RMU facilitates flexible operation of the distribution network by enabling sectionalisation, isolation, and rapid restoration of supply following a fault. Its configuration supports both incoming and outgoing feeders, thereby allowing the implementation of a ring network as depicted in the diagrams. Although the actual GIS data

do not represent the final connections between the cables and the load, switch, and circuit breaker, the fundamental design remains consistent.

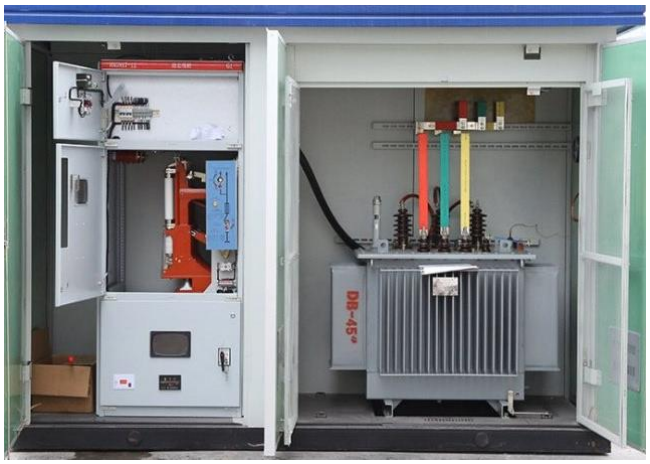

Figure 5.7 Physical arrangement of a typical 11/0.415kV secondary substation [17]

To illustrate the secondary substation configuration discussed in Figure 5.6, Figure 5.7 shows a real-world example. The transformer is located in the right-hand compartment, identifiable by its large grey tank and three coloured busbars representing the three-phase output (corresponding to the load shown in Figure 5.6). The circuit breaker and switch mechanism are housed in the left-hand compartment, visible as the orange and red mechanical assembly inside the open cabinet.

In practice, the connection logic is such that each switch is connected by two cables, two switches are connected to the circuit breaker, and the circuit breaker is subsequently connected to the load. This arrangement underpins the reliability and operational flexibility of RMU-based distribution systems.

Figure 5.8 presents the second distribution network layout, which demonstrates a typical configuration in which a branch, or sub-tree, of cables extends from the main feeder. In this arrangement, the re-connection logic specifies that the cable end node is connected to a circuit breaker, which subsequently supplies the load. This structure represents the standard radial or branched feeder arrangement that is commonly observed in distribution networks.

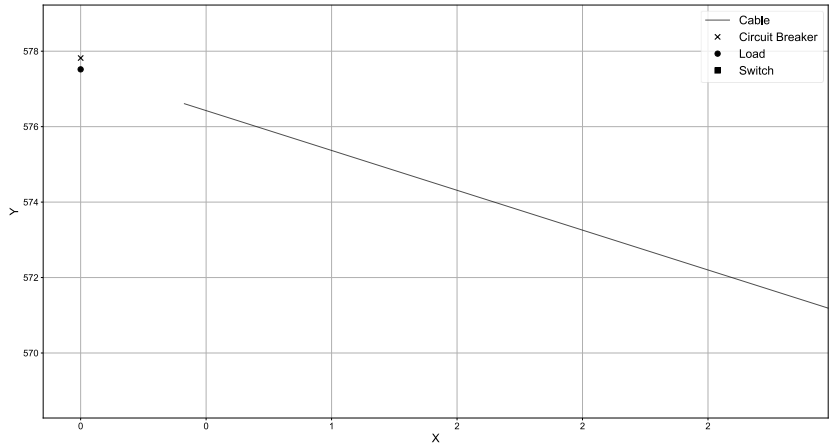

Figure 5.8 Layout 2 – schematic representation of cable, load, switch, and circuit breaker

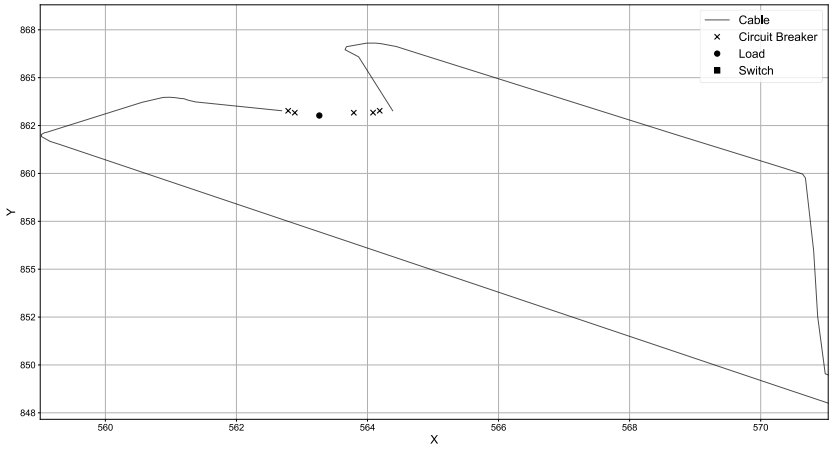

Figure 5.9 Layout 3 – schematic representation of cable, load, switch, and circuit breaker

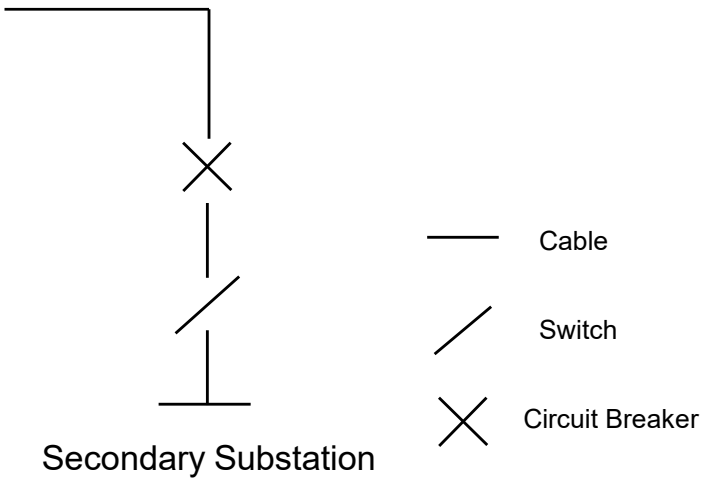

Figure 5.10 Single-line diagram corresponding to the network topology shown in Figure 5.8

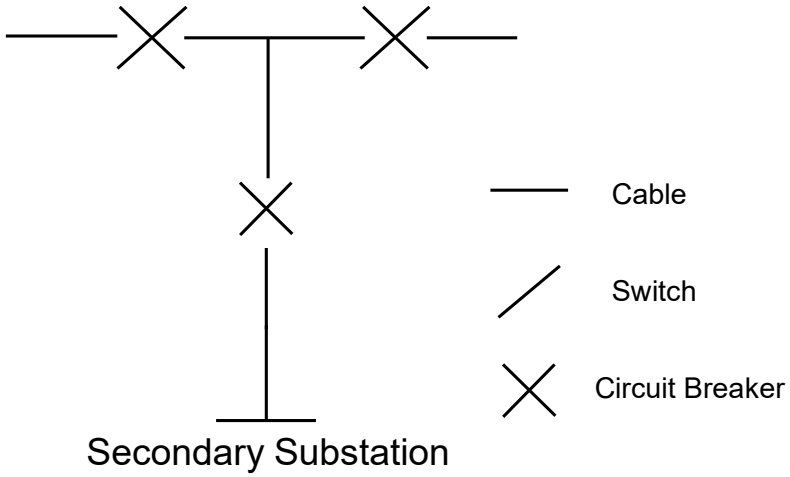

Figure 5.11 Single-line diagram corresponding to the network topology shown in Figure 5.9

Figure 5.9 illustrates the third distribution network layout, comprising an RMU-based (Ring Main Unit) configuration. In this scenario, the physical network contains multiple circuit breakers, specifically five, which are installed to provide additional protection.

Figure 5.10 and Figure 5.11 introduce schematic single-line diagrams for Figure 5.8 and Figure 5.9 to clarify the underlying electrical topology. These schematics are necessary to define the specific configuration of switches, cables, and connections represented by the spatial data, ensuring the operational logic is accurately captured for analysis.

For the purposes of power flow modelling and to reduce complexity, these circuit breakers are consolidated into a single equivalent device within the network model. As a result, only the circuit breaker in closest proximity is retained in the model, while the others are omitted, and the load is connected directly. This approach enables more efficient analysis of the network's operational characteristics without compromising the validity of the power flow study [18], noting that detailed protection coordination is not the primary focus of the analysis.

For the same reason, switches are assumed to be in their normal operating state. Normally closed switches are therefore treated as continuous lines and are not explicitly modelled, while only a small number of normally open switches are retained in the distribution model.

Accordingly, the following design rules are applied:

1. If cable end nodes are already connected, then no additional connections introduced.

2. All closed switches are disregarded in order to simplify the connection logic.

3. Circuit breakers located within five metres of each other are consolidated and represented as a single device to simplify the connection logic.

4. For an individual feeder, load, or transformer does not have a nearby circuit breaker, a pseudo circuit breaker is created one metre in the appropriate direction from the load or transformer to maintain consistency in the connection logic.

5. End nodes are connected to a circuit breaker if the distance between them is less than three metres and they are not already connected; subsequently, the circuit breaker is connected to the load or transformer.

6. Any instances that do not conform to these design rules are subject to manual verification and adjustment as necessary.

Based on the design rules, there are no unintended mergers or alterations to the network configuration resulting in a mesh distribution network due to misconnected points or points being placed too closely together. This approach ensures that the minimum necessary cable length is added to connect any disconnected segments of the network, while preserving the dataset's format. Furthermore, this method reserves space for the protection component, enabling the incorporation of the network model via circuit breakers in future versions.

Taking Feeder 1 as an example, it contains a total of 116 end node entries, of which 51 are duplicates, indicating that these nodes are already connected. This leaves 14 nodes unconnected and therefore requiring reconnection. After establishing the necessary connections between the circuit breaker, load, and transformer, the corresponding results are presented in Figures 5.10, 5.11, and 5.12.

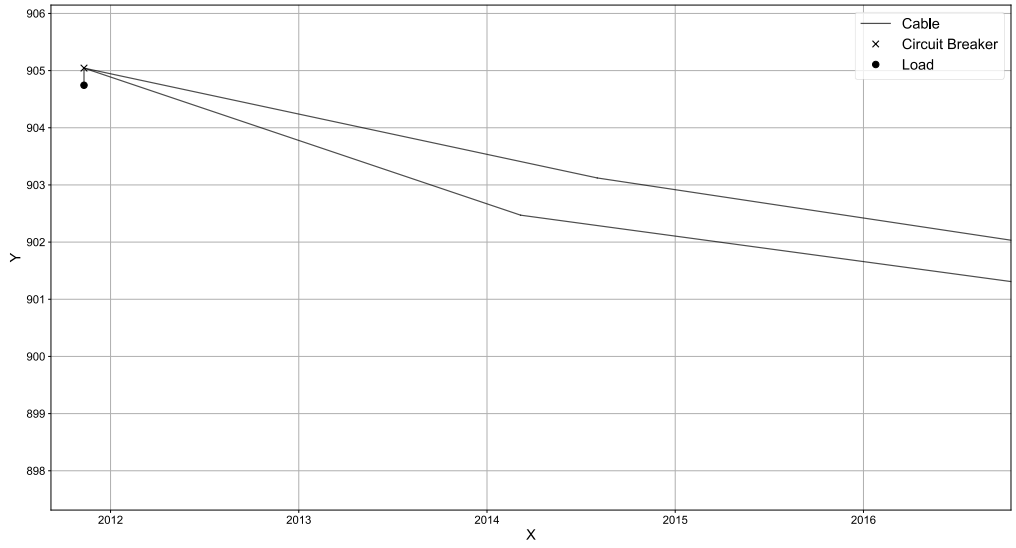

Figure 5.12 Schematic illustration of layout 1 – reconnection outcomes

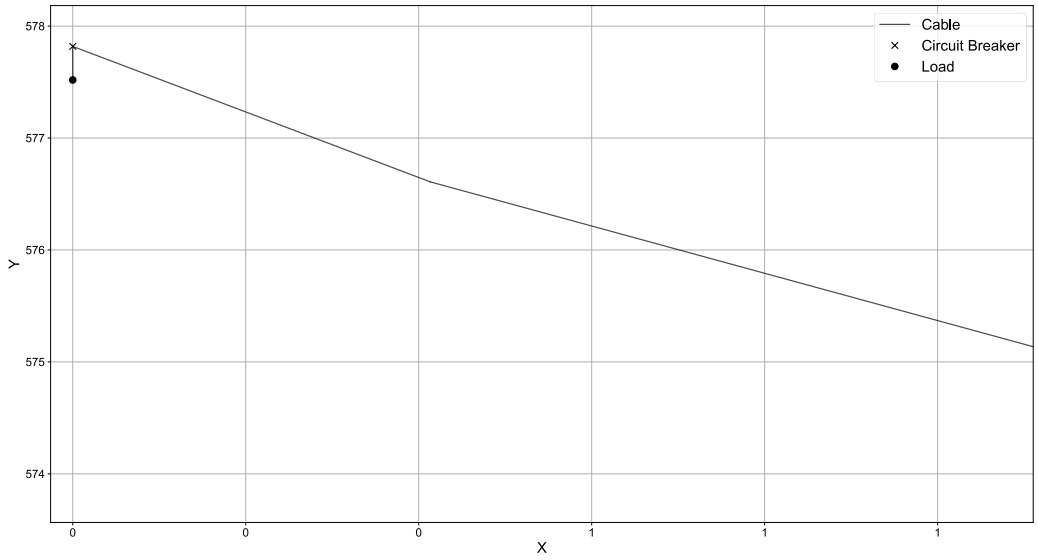

Figure 5.13 Schematic illustration of layout 2 – reconnection outcomes

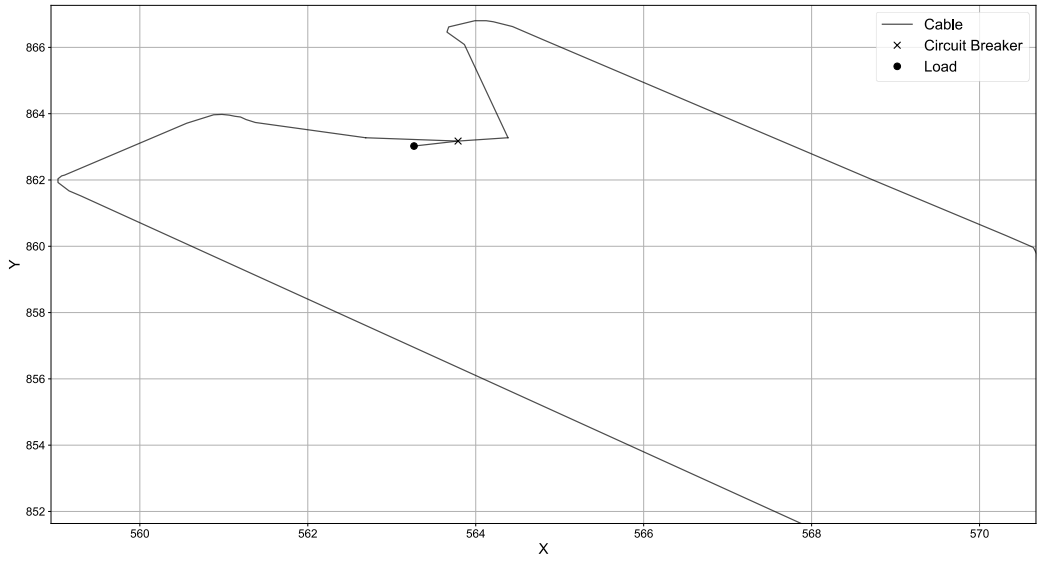

Figure 5.14 Schematic illustration of layout 3 – reconnection outcomes

The three figures above confirm that each load can now be connected to the network via the line equipped with a circuit breaker. This ensures that the majority of the network is now properly connected between the loads and the circuit breaker through the cable infrastructure. Any remaining disconnected sections may require manual adjustment, ensure that no additional loops are inadvertently created in the process.

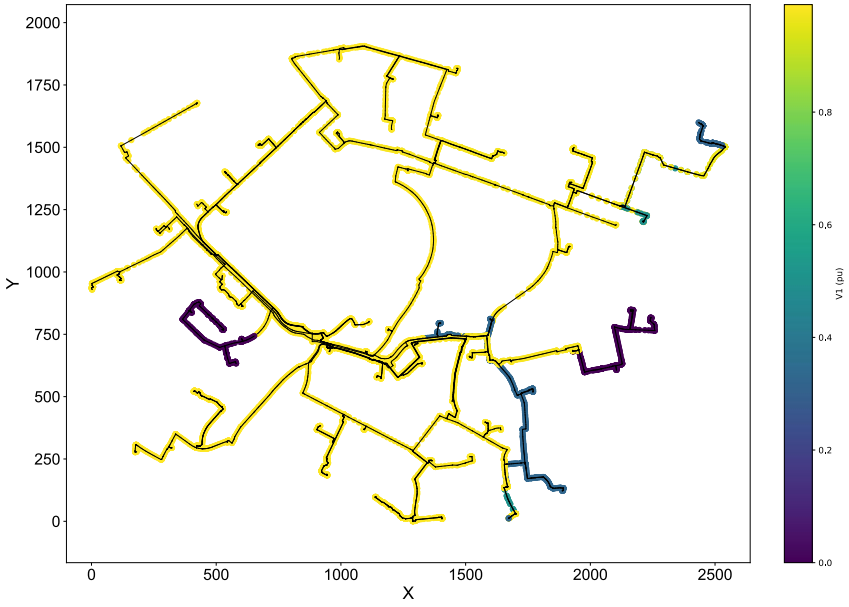

Figure 5.15 Voltage magnitude profile: OpenDSS power flow analysis

Figure 5.15 presents the results obtained directly from the design rules. The figure demonstrates that the majority of the network is connected to the transformers, with most voltage magnitudes close to 1.0 p.u. However, it is also evident that approximately five feeders remain incompletely connected, as indicated by multiple end nodes shown in purple. These nodes have a voltage magnitude of 0.0 p.u., signifying no connection to the transformer. This issue may have arisen due to inaccuracies or omissions in the dataset provided by the network operator, which contains nearly a thousand records for the linestrings and is therefore prone to some errors. Consequently, the remaining work involves manually connecting the incomplete sections of the network by examining each feeder individually to identify problems and then adding or extending lines as necessary to resolve these issues.

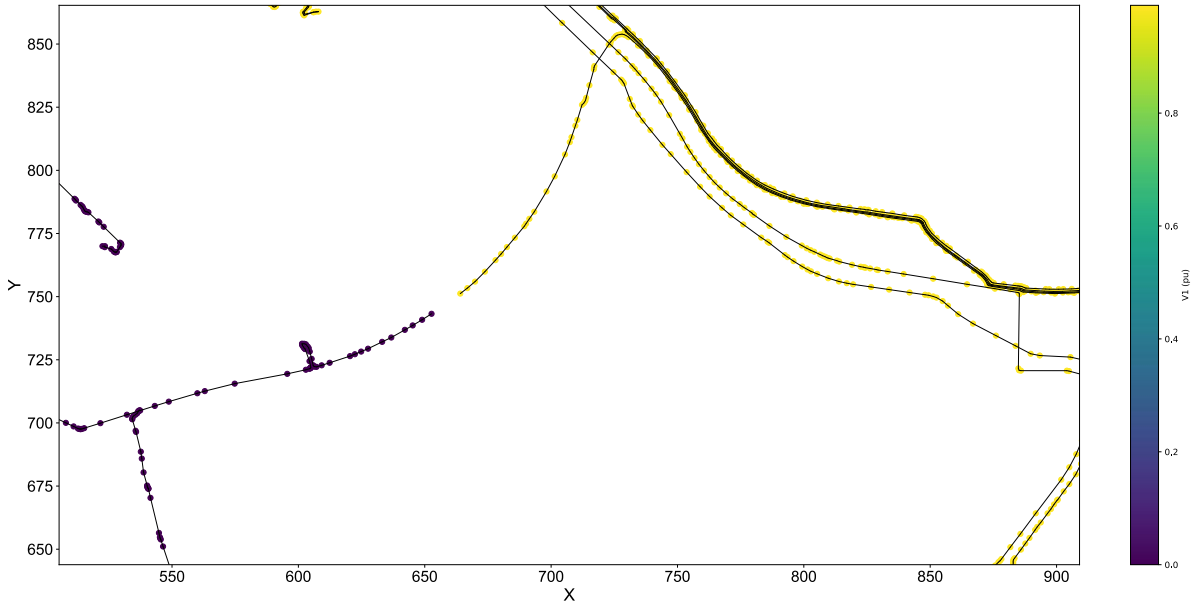

Figure 5.16 Power flow results highlighting network disconnection points

Figure 5.16 illustrates an example of the data omission issue, wherein two lines are disconnected in the middle of the map, resulting in network discontinuity. Another common issue arises when one line is connected to another, but there is no explicit node present in the linestring to represent the connection. In such cases, it is necessary to split the original linestring to accurately depict the connection point within the network model.

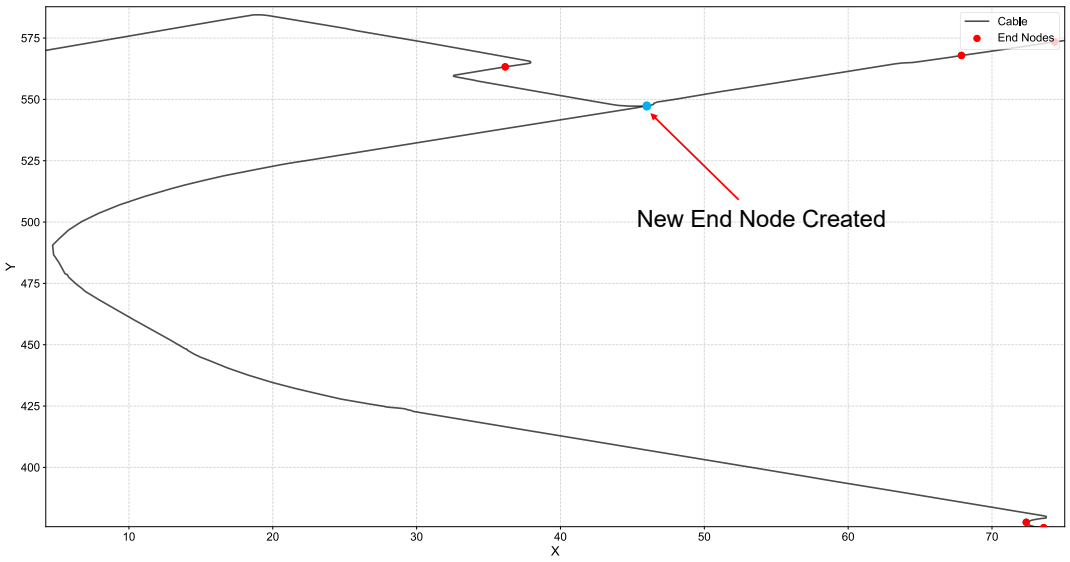

Figure 5.17 Example of new end node created for the cable

Figure 5.17 shows an example of this splitting process. The blue point represents an end node of the left cable. However, it is far from the existing end node of the right cable. The gap between these end nodes is too large to justify adding a new connecting line, as this would be unrealistic. Instead, the nearest point on the right cable is split to create a new end node that connects to the blue end node. In this way, the network is reconstructed with minimal change to the original geometry.

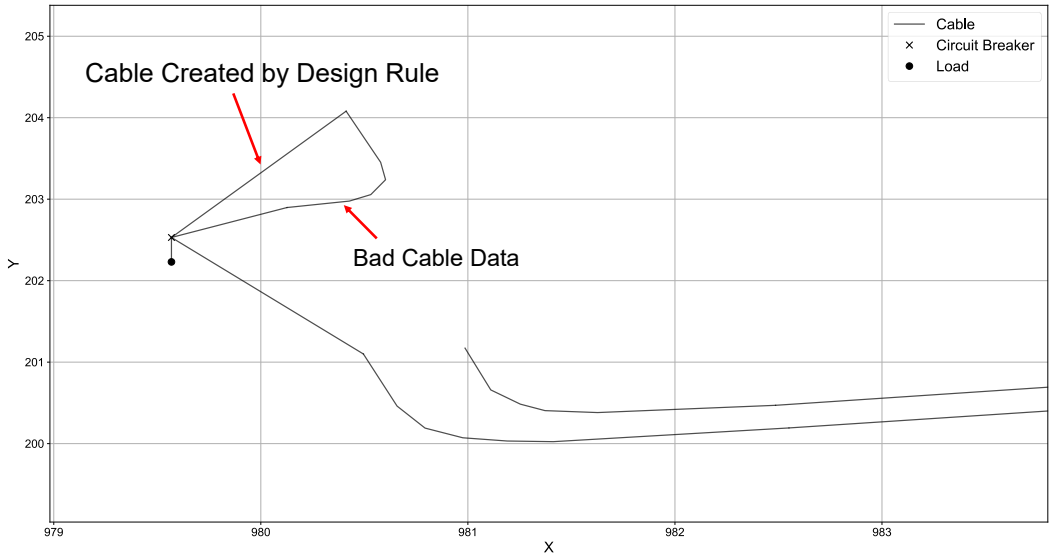

Figure 5.18 Example of erroneous cable data in a network model

Figure 5.18 presents another example of bad data within the network model. In this case, the erroneous cable data is isolated and occupies a position that should be used to connect to the circuit breaker. As a result, the downstream section of the radial network is left unconnected.

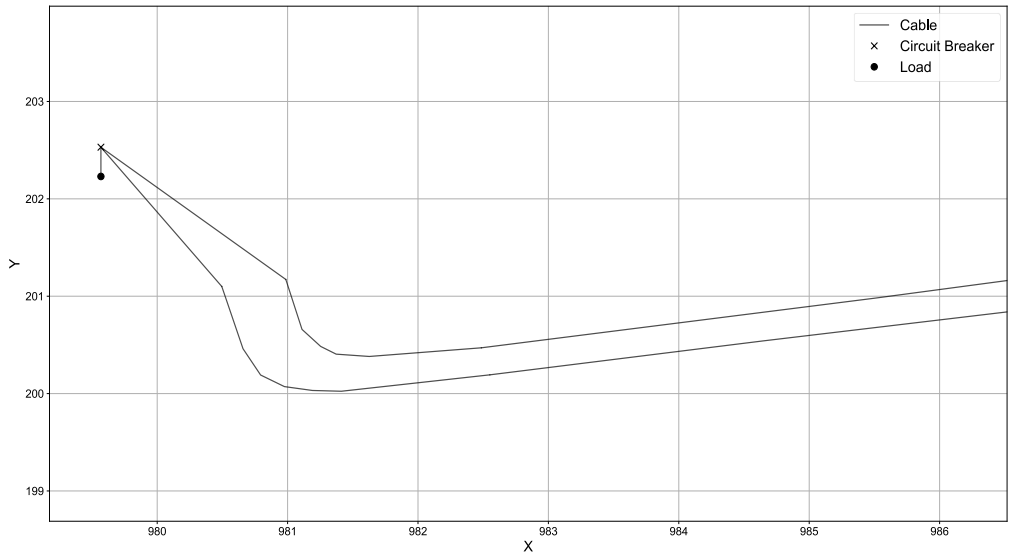

Figure 5.19 Corrected cable connection following data rectification

Figure 5.19 illustrates the correction process applied to resolve the issue highlighted in Figure 5.16. The previously isolated and erroneous cable, which prevented the downstream section of the radial network from being connected, has been removed. In its place, a new cable line has been added, following the design rules set out in the DNO network diagram (see Figure 5.6). This correction ensures that the circuit breaker and load are now properly connected within the network. The revised configuration adheres strictly to the principles of a radial network, restoring connectivity and maintaining the logical structure of the feeder.

This example demonstrates the importance of accurate cable data and network topology, as well as the effectiveness of manual intervention in rectifying modelling errors.

After correcting all the errors that were inadvertently introduced into the dataset, the fully connected network is illustrated in the figure below. This visual representation reflects the final structure of the network once all necessary amendments have been made, ensuring the data is now accurate and reliable for further analysis.

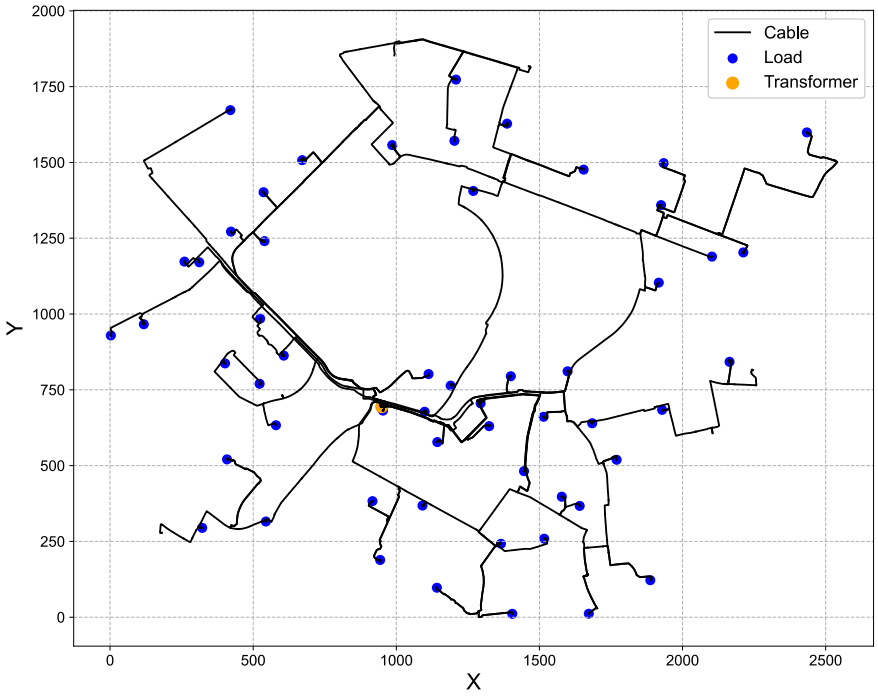

Figure 5.20 Structure of the fully connected network following data correction

The Figure 5.20 presents the geographical layout of the fully connected network. The network cables are clearly delineated, providing a comprehensive overview of the system's connectivity. All nodes and connections shown reflect the amended dataset, free from previous errors. This ensures that the visualised topology accurately corresponds to the actual physical layout, thereby supporting more reliable analysis and interpretation. Such a geographical representation is essential for network planning, operational studies, and future expansion, as it allows for the identification of potential vulnerabilities or areas for improvement within the system.

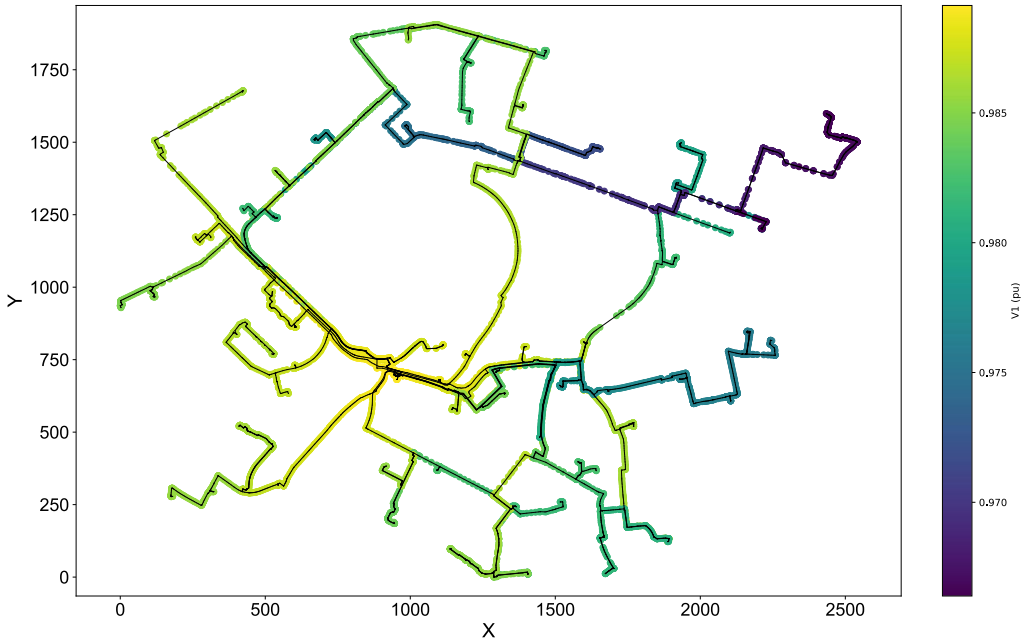

Figure 5.21 Voltage magnitude profile of the fully connected network – results from OpenDSS power flow analysis

Figure 5.21 presents the power flow results for the fully connected network. Compared to Figure 5.14, the effect of manual reconnection is significant; it is now evident that the entire network is interconnected. The minimum voltage magnitude across the system is approximately 0.97 p.u. Furthermore, the voltage magnitude tends to decrease as the distance from the substation increases along each feeder. Importantly, the voltage profiles of the individual feeders do not affect one another, which demonstrates the effectiveness of the network construction based on authentic, raw GIS data. This highlights a very successful implementation of the fully connected network configuration. With minimal changes to the network, only 200 additional lines have been created, with a total length of approximately 200m.

Nevertheless, the network shown in Figure 5.21 is a reconstructed model and therefore requires validation. In practice, the inferred connectivity should be validated by comparing simulated voltage profiles against measurements obtained from smart meters and substations. This comparison serves to confirm whether the model accurately reflects the behaviour of the actual LV network.

After the successful completion of the 33/11 kV network, the next step is to examine the network closer to the customer at the 11/0.415 kV level. The major challenge here is privacy, since it

already concerns household-level data, and the exact location of each household should not be published. Therefore, no dataset exists for the location of loads in the 11/0.415 kV network, and such data can only be created through assumptions and manual adjustments.

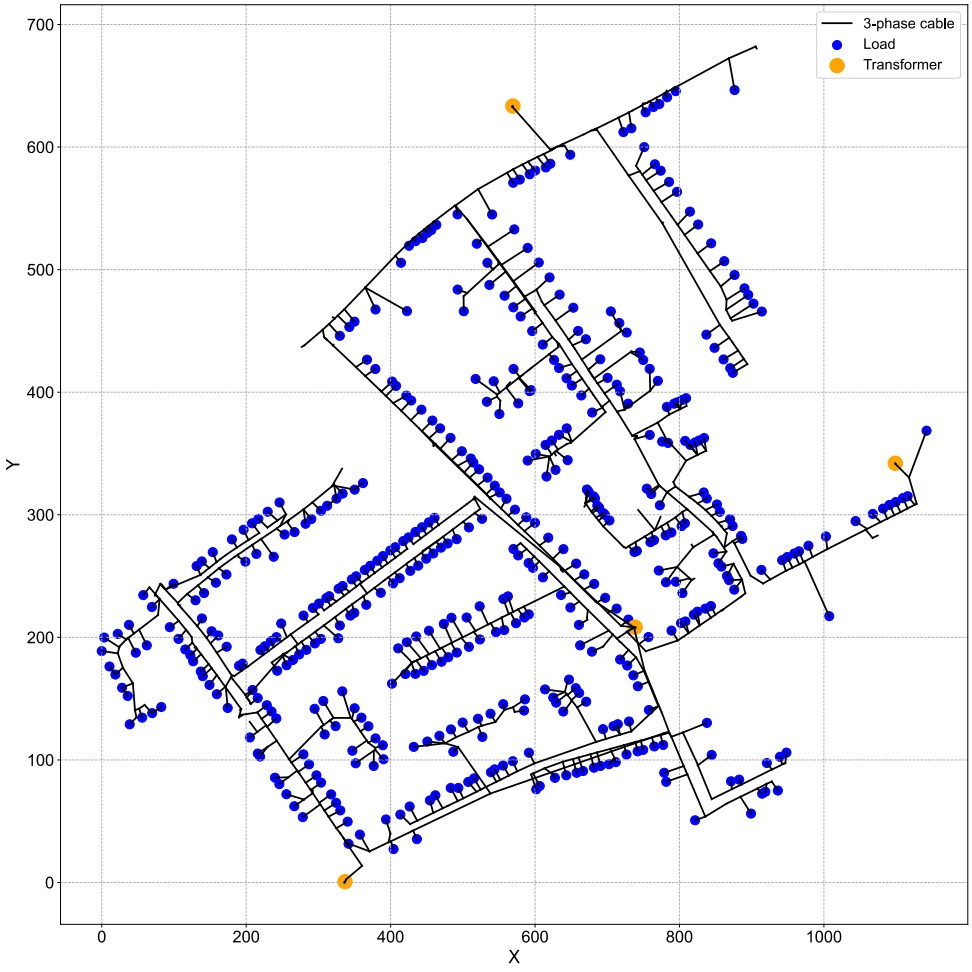

Figure 5.22 Layout of an 11/0.415 kV urban distribution network in GB

Figure 5.22 illustrates a representative 11/0.415 kV urban distribution network taken from the GIS dataset. The network comprises approximately 460 loads and 1,250 cables, all interconnected through four transformers to form a meshed distribution system. In such a meshed configuration, loads can be supplied by any or all of the four transformers, depending on the network's operational state and switching arrangement. This enhances both the reliability and flexibility of supply, as electricity can reach the loads via multiple alternative routes in the event of faults or maintenance. However, the increased interconnection and complexity of meshed networks can

make protection coordination and fault location more challenging, potentially leading to the risk of unintentional supply interruptions.

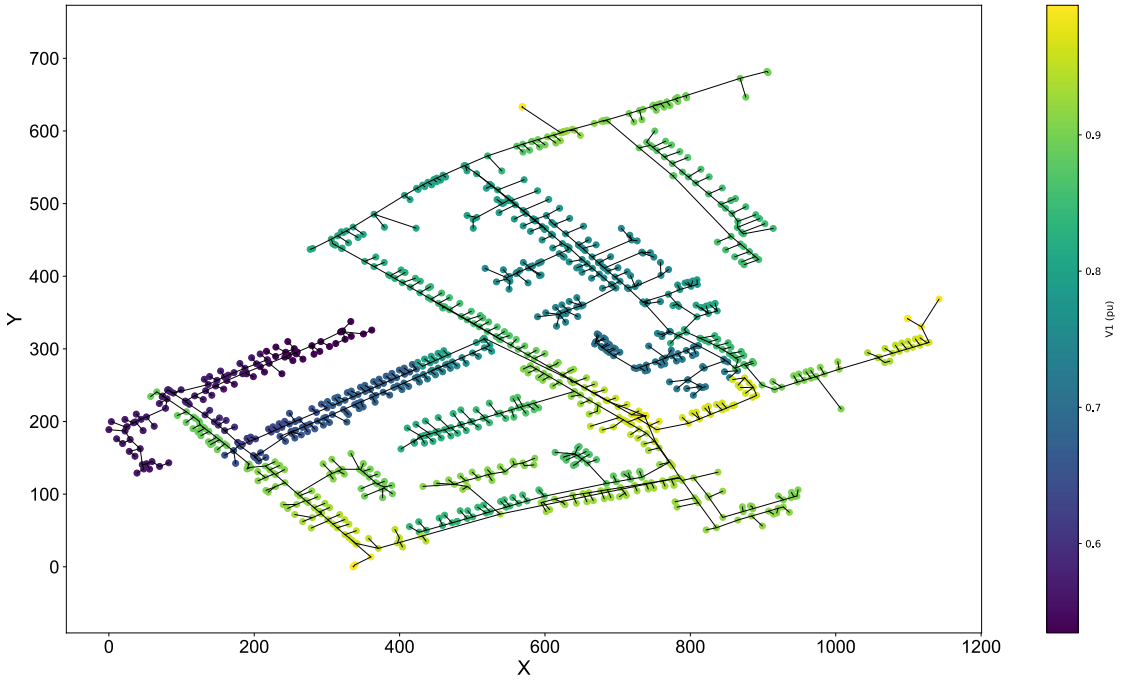

Figure 5.23 Voltage magnitude profile of the 415V test network – results from OpenDSS power flow analysis

Figure 5.23 shows the results for the 415V test network. Unlike Figure 5.19, which presents a radial network structure, Figure 5.19 displays a mesh network. The presence of four distinct yellow areas, each with a voltage magnitude close to 1.0 p.u., indicates that four transformers are supplying power simultaneously and are interconnected in a mesh arrangement. This configuration allows multiple pathways for power flow, in contrast to the single-path structure of a radial network.

However, despite this mesh structure, the results also reveal a significant issue within the network. There are purple regions where the voltage magnitude drops below 0.6 p.u., which is well below the minimum allowable limit of 0.95 p.u. These low-voltage areas are not being supplied by the bottom left transformer, but rather by another transformer, suggesting that one or more feeder connections may be missing or incorrectly modelled during the construction of the network.

This example highlights a major challenge in creating real-world distribution network models from GIS data. The datasets are vast and complex, and although the plotted network may initially appear correct, closer inspection often reveals errors that only become apparent at a detailed level. Such issues are frequently the result of data quality problems or recording errors, and addressing them requires a thorough, metre-level review, as well as a strong familiarity with the actual network configuration. This process can be extremely time-consuming.

5.3 Impact of future LCT integration in the Distribution Network

After constructing the network based on GIS data, a significant advantage is the ability to investigate the impact of LCTs such as PV panels, heat pumps, and EVs on the network in much greater detail. In contrast, the IEEE test network typically provides only a general overview of the effects of LCTs, treating them as generic loads placed somewhere in the network without regard to specific locations. The GIS-based model, however, enables detailed, high-resolution analysis of how and where these technologies can be integrated.

For example, with a GIS-based network, it is possible to identify particular areas where the installation of PV systems or EV chargers would create operational issues. PV panels installations should be avoided in the yellow regions of the network, as these areas already experience high voltage and could be prone to overvoltage problems if further generation is added. Similarly, installing EV chargers in the purple regions, where voltage magnitudes are already significantly low, could lead to even more severe undervoltage issues.

This level of precision allows network planners to make informed decisions about the optimal placement of new technologies, helping to maintain both the stability and efficiency of the system. By digitising the network and its impact through GIS-based modelling, it becomes possible to assess local conditions and operational challenges that would otherwise remain hidden. Such insights are difficult to obtain using standard test networks, which demonstrates the clear value of digital, GIS-based approaches in modern power system analysis.

Below is a simple test designed to assess the integration of PV systems in the 33/11 kV test network. In this experiment, the impact of PV installation is evaluated by placing PV systems both near to the transformer and further away from the transformer within the network. By comparing the network's response to PV integration at different locations, it is possible to gain insight into how

the position of distributed generation affects voltage profiles. This approach helps to identify the most suitable locations and installation sizes for PV deployment, thereby ensuring reliable network operation.

As real-time monitoring of the secondary substations is not available, an idealised scenario is adopted for this simple analysis. Specifically, a snapshot model is used for the power flow study.

In this scenario, each of the 56 secondary substations is assumed to be operating at a fixed load of 300 kW. First assumed that a PV system with a capacity of 1200 kW is installed at the end of Feeder 6.

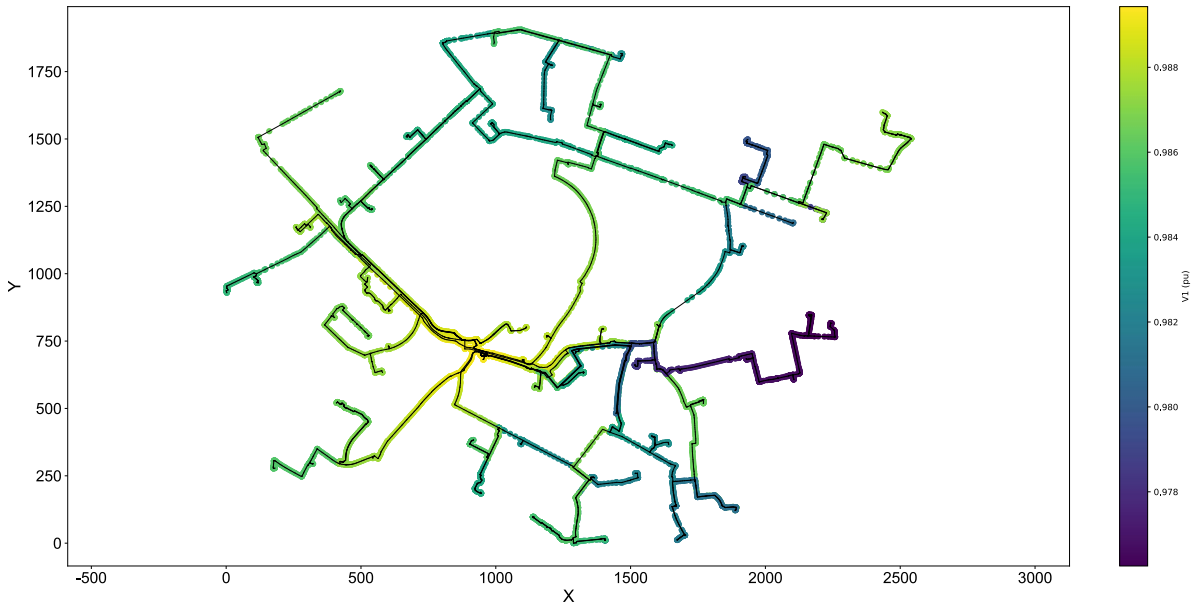

Figure 5.24 Voltage magnitude profile of the 11 kV test network with a 1200 kW PV system installed at the feeder end

Figure 5.24 illustrates the power flow results when a 1200 kW PV system is installed at the end of Feeder 6. The results demonstrate a positive impact on voltage magnitude, as the PV installation helps to raise voltage levels at the feeder's extremity, which is typically where voltage drops are most pronounced. Compared to the case with no PV installed, shown in Figure 5.17, the minimum voltage at the end of the feeder increases from 0.965 p.u. to 0.975 p.u. This improvement in voltage profile highlights the benefit of distributed generation in supporting network voltages.

In addition to enhanced voltage regulation, another significant advantage is the reduction in feeder losses. Since the PV supplies a portion of the local demand, less current needs to be delivered from the substation through the entire length of the feeder. This decrease in current flow results in lower resistive losses, thereby improving the overall efficiency of the distribution network.

Overall, the installation of a 1200 kW PV system at the end of Feeder 6 offers clear benefits in terms of both voltage support and loss reduction.

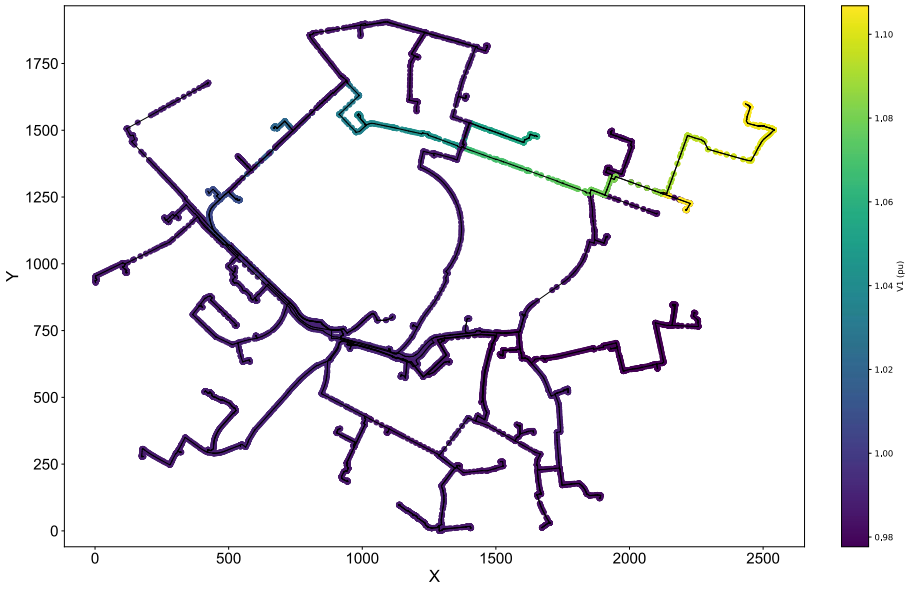

Figure 5.25 Voltage magnitude profile of the 11 kV test network with a 4500 kW PV system installed at the feeder end

Figure 5.25 illustrates the power flow results when a 4500 kW PV system is installed at the end of Feeder 6. In contrast to the scenario depicted in Figure 5.22, these results highlight significant adverse impacts on the network due to excessive PV generation. In this case, the PV output greatly exceeds the local load, resulting in reverse power flow back towards the substation. Consequently, the voltage at the end of the feeder rises well above acceptable limits, exceeding 1.10 p.u., which constitutes a clear breach of GB statutory voltage limits. Such overvoltage conditions pose serious risks to both customer equipment and the integrity of network infrastructure.

Traditionally, distribution networks are designed for power to flow unidirectionally—from the substation outwards towards the loads. The introduction of large-scale PV at the feeder's extremity reverses this flow, which can disrupt conventional protection schemes and complicate network

operation. Protection devices, typically coordinated for single-directional fault currents, may fail to operate correctly or may even trip unnecessarily in the presence of reverse power flow, thereby undermining the reliability and safety of the network.

This example clearly demonstrates the importance of thorough network analysis and prudent integration of distributed energy resources. By simply running power flow studies, DSOs can readily identify locations where problems such as voltage violations or reverse power flows are likely to occur. This proactive approach enables DSOs to implement appropriate measures, such as restricting the connection of LCTs, reinforcing network cables, or upgrading protection equipment to accommodate new operating conditions.

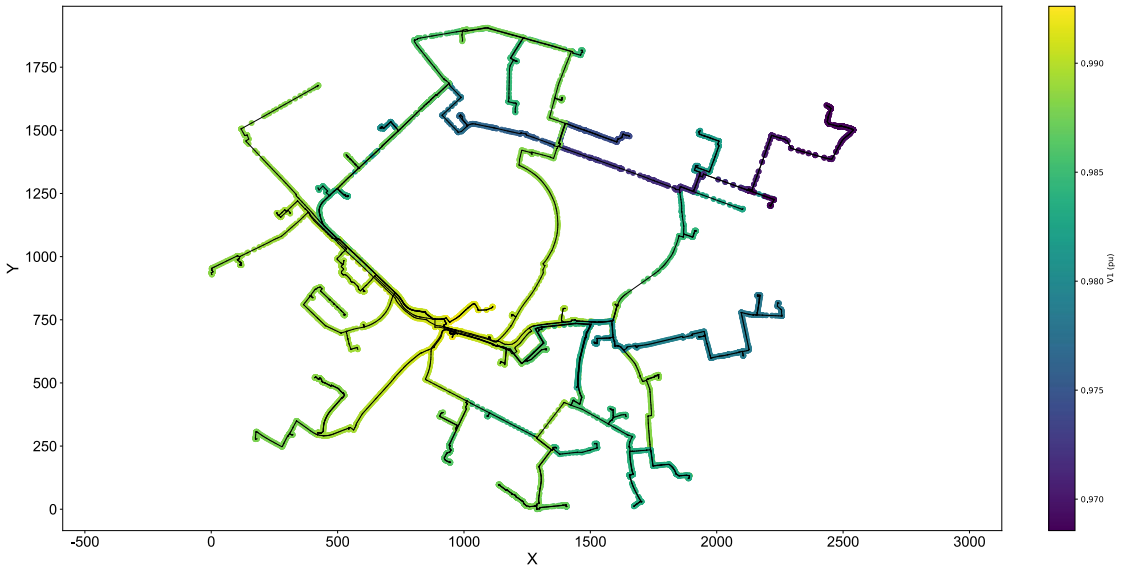

Figure 5.26 Voltage magnitude profile of the 11 kV Test Network with a 4500 kW PV system installed near the feeder origin

Figure 5.26 presents the results for a 4500 kW PV system installed near the beginning of Feeder 6. The power flow analysis indicates that the resulting voltage increase remains within statutory limits when compared to the case where the PV is installed at the end of the feeder. This outcome is primarily attributable to the relatively low line impedance between the PV system and the substation when the installation is close to the feeder's origin. Consequently, the power injected by the PV system causes only a minor voltage rise at the point of connection, thereby mitigating the risk of significant overvoltage issues. Furthermore, the substation's voltage control equipment can regulate voltage more effectively for connections situated near the feeder head, making it

easier to maintain voltage within statutory requirements. These findings demonstrate that installing PV systems near the beginning of the feeder is generally a safer and more straightforward approach for the distribution network than placing them further away. However, it should be noted that PV systems located near the substation provide limited benefits in terms of alleviating voltage drops or losses at the far end of the feeder; areas furthest from the substation still experience the largest voltage drops. In future work, the utilisation of network models with real operational data could enable DSOs to determine the optimal siting and capacity of PV systems, thereby ensuring that the network remains balanced and free from overvoltage.

5.4 Application of Transfer Learning and Few-Shot Learning to Network Operation

A real operational GIS-based network model can be extremely valuable, as it enables the acquisition of high-quality real-world data from power flow analyses that accurately reflect actual network behaviour. However, achieving this level of insight traditionally requires extensive network monitoring. For example, as demonstrated in Section 5.2, monitoring the network would necessitate the installation of 56 monitors at secondary substations to obtain a comprehensive overview of network performance and to conduct time-series power flow analyses. Such analyses are essential for assessing the impact of variable PV generation under different solar irradiance and weather conditions or understanding weekday and weekend variations in EV charging demand.

This highlights the potential role of the advanced machine learning techniques developed earlier in this thesis in the context of GIS based network analysis. In this section, we apply the transfer learning model introduced in Chapter 3 together with the few-shot learning model developed in Chapter 4 to a distribution network case study.

The transfer learning approach is used to exploit information from well monitored parts of the LV network and apply it to secondary substations with limited or no monitoring. This is particularly valuable in GB rural areas, where substations are often far from operational teams and the installation, maintenance and repair of monitoring equipment are challenging and costly. By reusing knowledge learned from substations that already have good measurements, transfer learning can reduce the need for extensive on-site instrumentation and the associated staffing effort.

The few-shot learning approach is then used for dense urban networks. Large cities such as London can contain hundreds or even tens of thousands of secondary substations, making it both impractical and unnecessary to install monitors at every substation and to stream high resolution data, for example every half hour, to the cloud. Few-shot learning enables accurate estimation of substation loads by monitoring only a small subset of substations within a given network. This allows DSOs to gain valuable insights into network conditions during the most critical periods without comprehensive monitoring across the entire LV network. The following section presents the results of applying transfer learning and few-shot learning to a GIS-based network, demonstrating their potential to reduce the need for extensive monitoring and instrumentation in distribution networks.

The first experiment applies the evening peak cardinal-point scenario within the GIS-based network model. Power-flow simulations are performed, and voltage magnitudes are measured at two monitoring locations: the primary transformer and the end of feeder bus, i.e., the load bus furthest from the substation. This configuration improves robustness and spatial coverage by sampling voltage magnitude at the feeder head and at its point of greatest cumulative drop. Measurement at the primary transformer bus provides the upstream reference level for the feeder. Measurement at the end-of-feeder bus captures the maximum voltage reduction along the line and the highest variability under changing demand. These two locations bound the feeder's voltage profile, permit inference of the voltage gradient, and ensure potential limit violations are observable where they are most likely to occur. Four methods are evaluated, consistent with Chapter 4: a local training model, multivariate few-shot learning, bivariate few-shot learning, and zero-shot learning, assessed over a 53-day period.

Figure 5.25 illustrates the transformer voltage magnitude over a 53-day period. The results compare the actual measured values with four transfer learning approaches: multivariate few-shot, bivariate few-shot, local and zero-shot. The local model provides the closest agreement with the actual profile, with minimal error across the observation period, as it is trained directly on data from this feeder. In contrast, the multivariate and bivariate few-shot models tend to overestimate the voltage magnitude, showing a slight positive bias relative to the measured values. The zero-shot method consistently underestimates the voltage magnitude and shows limited responsiveness to short-term variations, indicating weaker adaptability to the specific feeder characteristics.

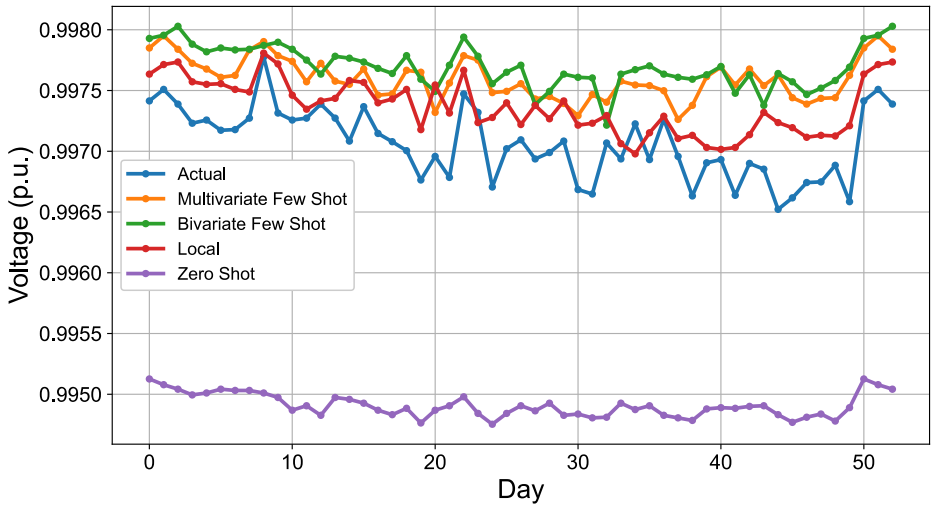

Figure 5.27 Comparison of few-shot learning methods for 33kV transformer voltage magnitude over 53 days

Figure 5.27 presents the end-of-feeder voltage magnitude over a 53-day period. Compared with the beginning of the feeder, the end-of-feeder voltage exhibits substantially larger deviations. In the case shown, the error at the end of the feeder is approximately ten times greater than that observed at the beginning, highlighting the greater difficulty in maintaining voltage control at this location. Both the multivariate few-shot and bivariate few-shot models follow the measured series closely, with the bivariate model performing marginally better during periods of sharper fluctuations, such as around days 20 to 25 and 48 to 53. The local model displays a consistent downward bias and larger errors during low-voltage periods, particularly near days 33 to 41. The zero-shot model remains significantly below the measured values and exhibits muted variability, indicating poor adaptation to the target feeder. Overall, the few-shot approaches deliver the most accurate and stable performance at this critical compliance point.

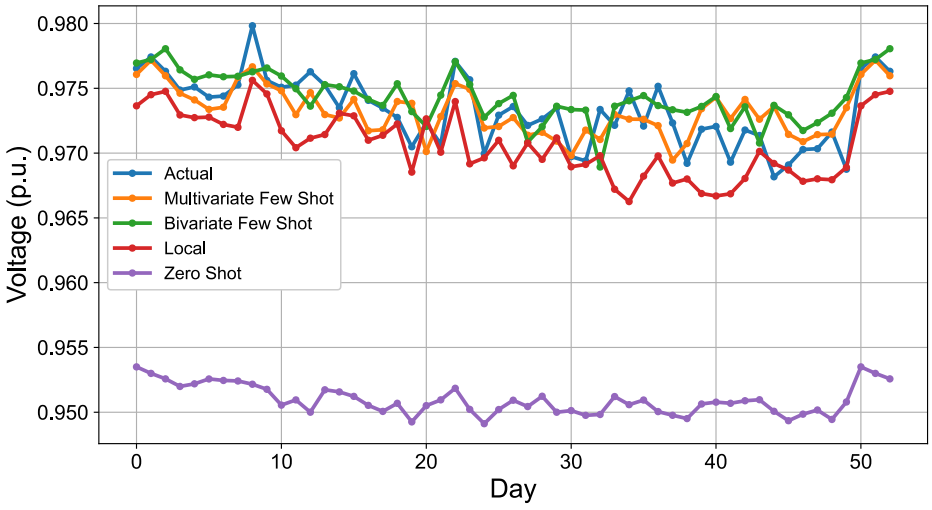

Figure 5.28 Comparison of few-shot learning methods for end-of-feeder voltage magnitude over 53 days

The second experiment evaluates a transfer-learning method that generates multi-substation time-series forecasts and assesses their impact on network behaviour in the GIS-based model. Day-ahead load profiles produced by the Chapter 3 framework are applied as nodal demands over the evaluation horizon, and time-series power-flow simulations are run. Voltage magnitudes are recorded at the primary transformer and the end-of-feeder bus, using the same monitoring pair as in Experiment 1 to maintain comparability.

Figure 5.29 presents a seven-day trace comparing power-flow results. The measured voltage exhibits a clear diurnal cycle, with pronounced troughs and modest intra-day rebounds. The Hong baseline is nearly flat and systematically biased high, offering limited operational value. Both transfer-learning approaches replicate the daily shape and timing; however, the Unconditional Bayesian Transfer shows an upward bias, whereas the Conditional Bayesian Transfer more faithfully tracks the amplitude and phase of the measured series, particularly around the evening minimum, indicating reduced systematic bias and improved trend capture. Relative to the evening cardinal-point forecast, errors here are higher, underscoring the benefit of focusing on cardinal points: extreme operating conditions are represented more accurately, as the model concentrates on peak values, and the number of power-flow runs can be reduced substantially. Nevertheless, the transfer-learning approaches provide time-series outputs that support an overall assessment of network performance for DSO operators.

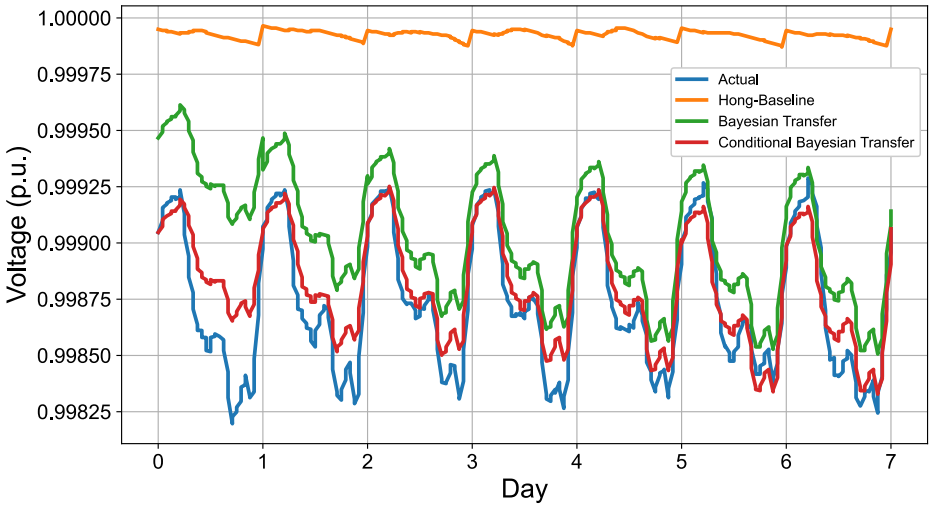

Figure 5.29 Comparison of transfer learning methods for transformer voltage magnitude over 7 days

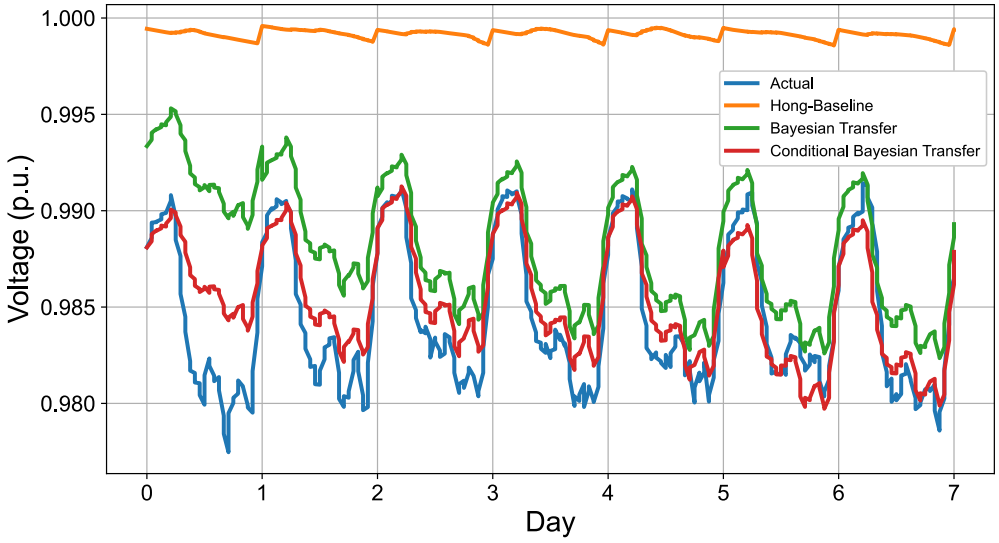

Figure 5.30 Comparison of transfer learning methods for end-of-feeder voltage magnitude over 7 days

Figure 5.30 presents the end-of-feeder voltage magnitude results obtained using the transfer learning approach. The patterns are broadly consistent with those at the cardinal points. However, the errors at the end of the feeder are much larger, approximately twenty times those at the primary transformer. Both peaks and troughs are overestimated at certain days. Even so, the forecasts remain within operational limits and perform markedly better than the benchmark.

Overall, the few-shot and transfer learning results show that meaningful forecasts can be produced in data-sparse areas with limited history or monitoring by transferring knowledge from other locations or by applying a global model. When combined with the GIS-based network model, these forecasts enable network operators to run time-series power-flow studies using fixed network parameters, which supports more informed day-ahead operational planning and management.

5.5 Conclusion

This chapter has highlighted the practical value and industry relevance of applying advanced forecasting and modelling techniques within the operational environment of GB's distribution networks and how it can be impacted in the real world. As DSOs face increasing complexity from the growth of low-carbon technologies and changing demand patterns, the adoption of innovative, data-driven approaches is becoming essential to maintain both reliability and efficiency.

By presenting the DFES process and demonstrating the construction of real-world GIS-based network models, this work highlights the clear advantages of using authentic network data over standard test feeders. Rigorous data cleansing and reconstruction ensure that these models are both robust and representative of actual network conditions. The process of correcting and validating network connectivity, often achieved through a combination of automated and manual methods, underscores the necessity for meticulous data preparation and regular quality assurance.

Moreover, this approach not only ensures accurate network connectivity but also provides flexibility for future developments, such as incorporating switches for fault simulation. Each newly added cable is carefully recorded, further evidence of the robustness of the network design. Compared to established platforms such as in Australian LV network model [19], the proposed method demonstrates advantages in terms of network data integrity and adaptability.

Modern machine learning approaches, including transfer learning and few-shot learning, have also been shown to help DSOs overcome monitoring challenges and improve network management using limited resources. By connecting fundamental algorithmic research with real-world applications, the methods and insights discussed here provide a practical pathway for developing smarter and more resilient distribution networks.

In summary, the methodologies established in this chapter form a solid foundation for addressing the ongoing challenges of informing network planning and operation, helping to secure the reliability, efficiency, and sustainability of electricity distribution in GB.

5.6 Reference

[1] Ofgem, "Future Energy Pathways: Guidance Document," Office of Gas and Electricity Markets (Ofgem), 2025/02 2025.

[2] SPEN, "Distribution Future Energy Scenarios (DFES) 2024," SP Energy Networks, 2024/03 2024. [Online]. Available: https://www.spenergynetworks.co.uk/userfiles/file/DFES_SP_Distribution_2024.pdf

[3] SPEN, "Distribution Future Energy Scenarios (DFES) 2024 – SP Manweb," SP Energy Networks, 2024/03 2024. [Online]. Available: https://www.spenergynetworks.co.uk/userfiles/file/DFES_24_SP_Manweb_2024.pdf

[4] SSEN, "SSEN DFES 2024: North of Scotland Technical Report," SSEN, 2024/03 2024. [Online]. Available: https://www.ssen.co.uk/globalassets/about-us/dso/dfes/ssen-dfes-2024-north-of-scotland-technical-report.pdf

[5] SSEN, "SSEN DFES 2024: Southern England Technical Report," SSEN, 2024/03 2024. [Online]. Available: https://www.ssen.co.uk/globalassets/about-us/dso/dfes/ssen-dfes-2024-southern-england-technical-report.pdf

[6] NorthernPowergrid. *Distribution Future Energy Scenarios (DFES)*, Northern Powergrid Open Data Portal, 2025/08/11. [Online]. Available: https://northernpowergrid.opendatasoft.com/pages/dfes/

[7] SPEN, "ED2 Digitalisation Strategy," SP Energy Networks, 2025/03 2025. [Online]. Available: https://www.spenergynetworks.co.uk/userfiles/file/SPEN_ED2_Digitalisation_Strategy_March_2025.pdf

[8] H. Kersting, "Radial distribution test feeders," in *2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194)*, 28 Jan.-1 Feb. 2001 2001, vol. 2, pp. 908-912 vol.2, doi: 10.1109/PESW.2001.916993.

[9] Bu, Y. Yuan, Z. Wang, K. Dehghanpour, and A. Kimber, "A Time-Series Distribution Test System Based on Real Utility Data," in *2019 North American Power Symposium (NAPS)*, 13-15 Oct. 2019 2019, pp. 1-6, doi: 10.1109/NAPS46351.2019.8999982.

[10] K. P. Schneider *et al.*, "Analytic Considerations and Design Basis for the IEEE Distribution Test Feeders," *IEEE Transactions on Power Systems,* vol. 33, no. 3, pp. 3181-3188, 2018, doi: 10.1109/TPWRS.2017.2760011.

[11] Frost, P. D. Mitcheson, and T. C. Green, "Power electronic voltage regulation in LV distribution networks," in *2015 IEEE 6th International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, 22-25 June 2015 2015, pp. 1-7, doi: 10.1109/PEDG.2015.7223039.

[12] S. Harrison *et al.*, "Review of multiport power converters for distribution network applications," *Renewable and Sustainable Energy Reviews,* vol. 203, p. 114742, 2024/10/01/ 2024, doi: https://doi.org/10.1016/j.rser.2024.114742.

[13] *OpenDSS: The Open Distribution System Simulator*. (2025). EPRI. [Online]. Available: https://www.epri.com/pages/sa/opendss

[14] A. Navarro and L. F. Ochoa, "Deliverable 1.2" Tool for translating network data from ENWL to OpenDSS"," 2012.

[15] C. McGarry, A. Anderson, I. Elders, and S. Galloway, "A Scalable Geospatial Data-Driven Localization Approach for Modeling of Low Voltage Distribution Networks and Low Carbon Technology Impact Assessment," *IEEE Access,* vol. 11, pp. 64567-64585, 2023, doi: 10.1109/ACCESS.2023.3288811.

[16] SPEN, "SPEN Connection Methodology," SP Energy Networks, 2024/11 2024. [Online]. Available: https://www.spenergynetworks.co.uk/userfiles/file/SPEN_connection_methodology.pdf

https://kangdaele.en.made-in-china.com/product/yanrAMLDaBhI/China-Manufacturer-Box-Type-Outdoor-Mobile-Prefabricated-Compact-Power-Transformer-Substation-Electrical-Substation-with-Good-Price.html

[18] Schumilin, H. K. Çakmak, K.-U. Stucky, C. Duepmeier, and V. Hagenmeyer, "Semantic CIM2Matpower: Ontology-based network topology processor and converter for CIM node-breaker models," 2019: IEEE, pp. 31-36.

[19] E. Karunarathne and L. N. Ochoa, "Data-Driven Construction of Detailed Three-Phase LV Network Models," 2025.

6 Conclusions and Future Work

6.1 Conclusions

This thesis has presented and validated a data-efficient framework for forecasting and monitoring LV distribution networks under realistic data constraints. By integrating techniques from transfer learning, few-shot learning, and GIS-derived network modelling, the work significantly improves observability and forecast accuracy in networks where historical data are limited, sparse, or unevenly distributed.

The original contributions outlined in Section 1.2 are substantiated across the thesis chapters as follows. In Chapter 2, a methodological approach is proposed to quantify why smart meter data cannot accurately regenerate the secondary substation load profile. The chapter also investigates the influence of exogenous factors, such as weather and calendar variables, on the load profile. This analysis provides a basis for understanding the limitations of naïve time-series forecasting approaches, particularly in the context of data sparsity and temporal bias. Also in Chapter 2, a structured review of existing approaches to modelling distribution networks is presented, with particular attention to their operational characteristics, data requirements, and configuration-specific constraints. Building on this foundation, an analytical review is conducted, in which power flow and state estimation methods are applied to translate load forecast outputs into meaningful insights about voltage profiles. This integrated approach enables a deeper understanding of the network's operating conditions, while also highlighting key challenges such as measurement sparsity, model uncertainty, and phase imbalance. These findings underscore the critical importance of Distribution System State Estimation for ensuring reliable operation of future distribution networks.

Chapter 3 introduces a novel transfer learning framework that augments Bayesian forecasting techniques for modelling residual errors across distinct geographical areas of the distribution network. By transferring knowledge from data-rich regions to data-poor

ones, the approach enables more reliable predictions in low-observability contexts. This is complemented by the application of Forecasting-Aided State Estimation, which extends network observability by estimating voltage magnitudes and phase angles based on forecasted load profiles.

In Chapter 4, a new forecasting paradigm, known as cardinal point forecasting, is defined and implemented. This method focuses on three operationally salient tuning points relevant for day-ahead network planning. It provides a more tractable and data-efficient alternative to time-series forecasting, particularly when model complexity or data availability poses a constraint. Chapter 4 also introduces a global forecasting model, based on customer numbers, that supports zero-shot learning for forecasting the timing of cardinal points and few-shot learning for estimating their magnitude. To improve the calibration of these forecasts and mitigate aggregation errors, a Gaussian copula-based correction is applied, enhancing local accuracy despite limited observational data.

Finally, in Chapter 5, a practical methodology for constructing an operational network model from actual geospatial topology data is proposed. This model is aligned with DFES practices and serves as a physical foundation for applying the earlier forecasting methods. The resulting tool supports operational decision-making by enabling DNOs to anticipate extreme voltages and currents with the integration of LCTs, identify the risk of constraint violations, and deploy day-ahead interventions with greater precision. Taken together, these contributions comprise a coherent framework that addresses the pressing challenge of operating and forecasting within low-observability LV networks. The thesis demonstrates that meaningful improvements in forecast reliability and network visibility are achievable without requiring widespread deployment of monitoring infrastructure. This is made possible through the intelligent transfer of load-profile knowledge, the prioritisation of operationally meaningful cardinal points, and the integration of forecasting into physically grounded, GIS-based network models.

Nevertheless, certain limitations remain. The accuracy of the proposed methods is sensitive to input data quality, particularly the accuracy of customer counts, and to the suitability of

the donor areas selected for transfer learning. Furthermore, the global model assumes stability in the statistical relationships across time and geography, which may be challenged as low-carbon technologies become more prevalent and consumer behaviour evolves.

Despite these caveats, the proposed framework provides DSOs with a practical and scalable solution for day-ahead forecasting and operational planning. It consistently outperforms benchmark models in the majority of tested scenarios, all without the need for additional monitoring infrastructure. This represents a significant advancement in the field of distribution network operation and short-term forecasting.

6.2 Future Work

Looking ahead, several promising directions can further enhance the practical impact of this research. A priority is to develop more comprehensive and granular network models, with particular emphasis on the 11/0.415 kV level, to deepen insight into LV operation and enable more precise assessment of impacts at customer level. These models should be systematically validated against actual network monitoring data so that simulated behaviours are checked against reality and confidence in results is strengthened.

This project is closely aligned with recent and ongoing initiatives in several countries that aim to advance LV network understanding and capability. In GB, the University of Manchester has delivered the Low Voltage Network Solutions (LVNS) project [1], which has provided foundational contributions to LV modelling and monitoring through the development of detailed GIS-based OpenDSS network models validated using empirical field data. In Australia, the University of Melbourne is undertaking the project Integrated MV–LV Network Studies to Assess Electrification Impacts [2], which seeks to develop detailed three-phase integrated MV–LV models to capture the effects of widespread electrification, particularly from technologies such as heat pumps, electric vehicles and distributed solar generation. In the United States, Iowa State University is leading the project Robust Real-Time Modelling of Distribution Systems with Data-Driven Grid-Wise

observability [3], which integrates high-resolution sensor data with advanced machine learning algorithms to improve real-time modelling and observability of LV networks.

These initiatives share a common vision of equipping DSOs with enhanced visibility and control at the LV level, thereby enabling more effective utilisation of existing infrastructure and improved integration of low-carbon technologies. Collectively, they reflect a widely held belief that high-fidelity and data-validated LV network models, incorporating GIS information, customer demand profiles and real-time measurements, are essential to the modernisation and decarbonisation of electricity distribution systems.

Applying forecast load profiles to these models will facilitate advanced voltage control strategies such as Conservation Voltage Reduction, using forecast information to support energy efficiency while maintaining compliance with statutory voltage limits [4]. High-fidelity GIS-based models also make it possible to develop three-phase low-voltage network representations with the help of smart meter data [5], and can make N-1 contingency analysis feasible in distribution networks [6], allowing operators to identify vulnerabilities and improve resilience. As the availability and quality of distribution data continue to improve, analytical and operational techniques long established at transmission level, including state estimation, dynamic security assessment and real-time contingency analysis, can be adapted to the distribution context to deliver greater visibility, flexibility and control. With richer integration of actual GIS data, new possibilities open up for real-time decision support and more detailed planning for the integration of future low-carbon technologies. By extending the modelling, validating outcomes with real-world measurements and adopting these advanced operational strategies, the work can continue to support a smarter, more flexible and more sustainable electricity distribution system in GB.

6.3 Reference

[1] ElectricityNorthWest, "Low Voltage Network Solutions (LVNS) — Closedown Report," 2014. [Online]. Available: https://www.enwl.co.uk/future-energy/innovation/smaller-

projects/low-carbon-networks-fund/low-voltage-network-solutions/low-voltage-network-solutions-closedown-report/

[2] UniversityofMelbourne, "Integrated MV-LV Network Studies to Assess Electrification Impacts," University of Melbourne, Melbourne, Australia, 2025 2025. [Online]. Available: https://electrical.eng.unimelb.edu.au/power-energy/projects/integrated-mv-lv-network-studies-to-assess-electrification-impacts

[3] Z. Wang, "Smart Meter Data: A Gateway for Reducing Solar Soft Costs with Model-Free Hosting Capacity (MoHCa) Maps," Iowa State University, Ames, IA, USA, 2025 2025. [Online]. Available: https://wzy.ece.iastate.edu/project/mohca_1.pdf

[4] Gutierrez-Lagos and L. F. Ochoa, "CVR assessment in UK residential LV networks considering customer types," in *2016 IEEE Innovative Smart Grid Technologies - Asia (ISGT-Asia)*, 28 Nov.-1 Dec. 2016 2016, pp. 741-746, doi: 10.1109/ISGT-Asia.2016.7796477.

[5] E. Karunarathne, M. Z. Liu, L. F. Ochoa, and T. Alpcan, "Using Real Smart Meter Data to Construct Three-Phase Low Voltage Network Models," *IEEE Transactions on Power Systems,* vol. 40, no. 3, pp. 2465-2477, 2025, doi: 10.1109/TPWRS.2024.3474672.

[6] F. Luo, C. Wang, J. Xiao, and S. Ge, "Rapid evaluation method for power supply capability of urban distribution system based on N−1 contingency analysis of main-transformers," *International Journal of Electrical Power & Energy Systems,* vol. 32, no. 10, pp. 1063-1068, 2010/12/01/ 2010, doi: https://doi.org/10.1016/j.ijepes.2010.01.021.