Intraday residual transfer learning in minimally observed power distribution networks dynamic state estimation

Junyi Lu ⬦, Bruce Stephen ⬦ and Blair Brown

Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK
Corresponding author: Junyi Lu; Email: junyi.lu@strath.ac.uk

Abstract

Traditionally, electricity distribution networks were designed for unidirectional power flow without the need to accommodate generation installed at the point of use. However, with the increase in Distributed Energy Resources and other Low Carbon Technologies, the role of distribution networks is changing. This shift brings challenges, including the need for intensive metering and more frequent reconfiguration to identify threats from voltage and thermal violations. Mitigating action through reconfiguration is informed by State Estimation, which is especially challenging for low voltage distribution networks where the constraints of low observability, non-linear load relationships, and highly unbalanced systems all contribute to the difficulty of producing accurate state estimates. To counter low observability, this paper proposes the application of a novel transfer learning methodology, based upon the concept of conditional online Bayesian transfer, to make forward predictions of bus pseudo-measurements. Day ahead load forecasts at a fully observed point on the network are adjusted using the intraday residuals at other points in the network to provide them with load forecasts without the need for a complete set of forecast models at all substations. These form pseudo-measurements that then inform the state estimates at future time points. This methodology is demonstrated on both a representative IEEE Test network and on an actual GB 11 kV feeder network.

Impact Statement

As global efforts shift toward sustainable energy, electricity distribution networks face a transformative phase. Originally designed for one-way power flow, these networks now incorporate distributed energy resources (DERs) and other eco-friendly technologies, presenting substantial operational challenges, including necessary advanced metering and frequent reconfiguration. A pivotal issue is accurate state estimation in low voltage (LV) networks, where low observability and non-linearities hinder reliable estimates. Our research introduces a transfer learning methodology, employing an online Bayesian method to predict bus pseudo-measurements, thus elevating state estimation accuracy. For industry insiders, the benefits are numerous. Enhanced reliability translates to fewer downtime from less grid disruption, and better asset management, through fewer damaged assets from extremes of network operation. Improved state estimation allows greater DER integration without jeopardizing grid stability. Furthermore, this method can deter expensive infrastructure upgrades by optimizing current assets, leading to long-term cost savings.

1. Introduction

Renewable energy sources are often variable and intermittent, meaning their output can change rapidly and unpredictably. High levels of renewable energy penetration, particularly when they constitute more than 20–30% of the total energy supply, can pose challenges for distribution networks. For instance, with the rising penetration of low-carbon technologies in the low-voltage network, systems like photovoltaic installations and electric vehicle charging could lead to voltage excursions for a significant number of customers (Navarro-Espinosa and Ochoa, 2016). Today, the distribution network, which delivers power over the last few miles to customers, also faces monitoring challenges. While the growing popularity of smart meters enhances situational awareness, their measurements often are not transmitted in real-time. Even though more homes and businesses are using smart meters that could help a network operator gauge activity, there is a delay in relaying that information back to a control center. Additionally, there is an insufficient level of monitoring (e.g., SCADA) on low voltage (LV) feeders, and it will be extremely expensive as GB has more than 900,000 LV substations (Li et al., 2015). The majority of the power distribution networks in Great Britain (GB) were planned, designed, and constructed during the 1950s and 60s (Oatley et al., 1997). At that time, the cables were built with sufficient capacity to accommodate projected demand growth from end-use. As a result, the distribution system has remained largely unmonitored. However, in recent years, the changing usage of distribution networks has resulted in bi-directional power flows (from embedded generation such as PV panels) and extreme loads (resulting from energy-efficient appliances driving baseload down and electric heating and transport driving peaks up), more distribution network operators (DNOs) have begun planning and installing monitoring devices in substations to enable real-time data monitoring and storage (Rowe et al., 2014), making data available remotely. These would represent a data-rich area or measurement point. In contrast, the majority of the distribution network remains a data-sparse area, where real-time data may not be available, although some data might be collected separately and subsequently based on DNO operations. This complicates the process of creating accurate models of network behavior. Additionally, the scale of distribution network asset fleets means that installing monitors in every area would require significant time and investment. For instance, in 2008, GB decided to introduce smart meters to all households, but by 2023, only 31.3 million had been installed, covering just 55% of households (Kerai, 2023). Another issue is the complexity of load behaviors of residential and light commercial premises – a significant proportion of connections at LV. Low reactance to resistance ratios in the distribution system make the system more resistive, and resistive losses may become more significant. This could mean that the simplifications and assumptions made in current State Estimation methods do not provide an accurate representation of the true system state, leading to erroneous estimates (Ahmad et al., 2018). Furthermore, the relationships between loads on low-voltage distribution network buses are not linear or Gaussian, meaning that conventional least-squares state estimation results in a sub-optimal model (Vanin et al., 2023). Lastly, there is the issue of imbalance to consider. In practice, distribution systems often exhibit significant unbalance across their three phases (Ma et al., 2017), violating conventional state estimation assumptions of a three-phase balanced network, which can lead to inaccurate state estimations (Ahmad et al., 2018).

Distribution network reconfiguration is necessary given that excursions in thermal and voltage constraints could occur in the distribution network at short notice, resulting from weather, social, or behavioral routine factors. The reconfiguration process optimizes the system state by altering the status of line switches and power injection, thus reducing network losses. It ensures the balance of power supply and demand while meeting current and voltage constraints (Tang et al., 2022). Furthermore, this reconfiguration helps to anticipate and prevent potential overload and voltage constraint excursions.

To gain insights into the distribution network behavior, state estimation is required (Schweppe and Wildes, 1970). It is an approach that transforms available network information into an estimate of a vector representing the magnitudes and angles of voltage on all network buses. This vector is also referred to as the static-state vector. State estimation generally uses a mathematical procedure to process

real-time measurements to best estimate the current state of the entire system (Dehghanpour et al., 2019). The results of state estimation provide real-time data for other estimates, such as power injection requirements, power flow, and voltage angles, among others. While state estimation is extensively utilized in transmission and higher voltage levels (Táczi et al., 2021, adoption on LV distribution networks, such as the 11 kV network in GB, faces a number of challenges. Some of these challenges stem from insufficient observation data as well as the lack of tools tailored to incomplete measurements and non-Gaussian load behavior (Dehghanpour et al., 2019), which itself forms a major assumption in conventional state estimation methodology. Thus, there is a need to improve existing distribution network state estimation methods to ensure both system observability and the resulting quality of the state estimates.

During the state estimation process, there is a requirement for continuous load data; however, this is not always available, especially at the distribution level. Therefore, pseudo-measurements are used in anticipation of the real measurements becoming available. Pseudo-measurements essentially fill in for the missing load data that is anticipated during state estimation. At the high voltage level, these pseudo-measurements offer practical estimates of system states, enhancing precision in modeling efforts. Typically, load values derived from standard load profiles serve as the foundation for these pseudo-measurements (Manitsas et al., 2012). Yet, within the distribution system, these values are often completely unknown due to their high variability.

Although there have been many algorithms for estimating pseudo-measurements of power data in distribution networks, most of them use large amounts of data in practical applications, potentially resulting in considerable computational expenditure and impracticality for operational deployment at distribution. To address this lack of data, a method predicated on transfer learning is proposed in this paper to achieve dynamic state estimation on LV distribution networks. The novel contribution is a transfer learning methodology for load pseudo-measurement forecast at individual pseudo-measurement points on a day ahead basis. The new approach leverages transfer learning based on updating Bayesian estimates of intraday forecast residuals for load pseudo-measurement prediction, which allows the model to utilize knowledge gained while solving one problem (a source substation day ahead forecast relation) and apply it to a different but related problem (one or more target substation forecasts). This not only accelerates the training process but also requires significantly less data compared to existing state-of-the-art methods, making it a more efficient and cost-effective solution, especially when running it on substation computing devices with minimal resources.

The structure of this paper is as follows: Section 2 provides an overview of the foundational aspects of dynamic state estimation for power networks. Section 3 details the application of transfer learning for load forecasting, including the novel use of intraday residuals to adjust forecasts to different metering points. Section 4 details the day-ahead forecast benchmark model, elaborates on the functionality of transfer learning, and compares errors between fully observed and minimally observed network cases. The resulting state estimates are compared against pseudo-measurements obtained from both levels of observation as well as the power flow calculations, which provide the ideal case. Section 5 presents a discussion on the results and potential avenues for future work.

2. Background on estimating the state of power systems

2.1. Basics of state estimation

Fred Schweppe first introduced state estimation to the power system in 1970 (Schweppe and Wildes, 1970). This data processing algorithm converts available information like meter readings into an estimate of the static-state vector, which represents the magnitudes and angles of voltage at all network buses.

The mathematical relationship of the related static model is often represented as

$$z = h(x) + v, \tag{1}$$

The fundamental mathematical model of state estimation is built upon the relationship between measured variables and state variables, where z stands for the set of network measurements, **x** signifies

the vector of the state variables, h indicates the connection between the measured values z and the state variables $\mathbf{x}$, and $\mathbf{v}$ is the measurement error vector, accounting for all the discrepancies or errors in the observed values. Measured variables can encompass real and reactive power flows on overhead lines/underground cables, real and reactive power injections at buses, and voltage magnitudes at specific buses. State variables, which typically include bus voltage magnitudes and phase angles, define the state of the system, however, state variables are not directly measured. Instead, they are estimates based on the measured variables and the system model. The error vector v, which can be represented as a 1d vector $\{v_1, v_2, v_3, \ldots, v_m\}$, is assumed to consist of zero mean Gaussian noise, an assumption which may not hold in practice. The variance of the error, R, provides an indication of certainty about the measurement. Occurrences of high variances on the diagonal of the covariance matrix are indicative that the measurement is not accurate.

In the measurement-state relationship (1), x represents the unknown true state, a deterministic quantity. Since the errors v are random variables, this makes the measurements z random variables as well. As such, z is assumed to follow a Gaussian distribution with mean $h(x)$ and covariance R. The goal of state estimation is minimizing the error, which is equal to the maximum likelihood estimate of minimizing the squared error weighted by the measurement accuracy. The solution for the weighted least squares performance index, J, is given by

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

which is equivalent to

$$\min J(x) = \frac{1}{2} \sum_{i=1}^{m} \frac{\{z_i - h_i(x)\}^2}{\sigma_i^2}, \tag{3}$$

where m is the total number of measurements and σ_i^2 is the measurement error variance at ith measurement.

In the state estimation model, the main assumption is that there is not a full bus and line representation of the network of interest. Figure 1 illustrates a typical 14-bus test network, featuring 5 generators and 11 loads. To demonstrate the effectiveness of a state estimator, load data can be set within a network model such as this, and a full observation can be obtained from power flow calculations using the topology and line characteristics. Censoring this ground truth data and performing state estimation to understand the network operating parameters, yields accuracy measures for the estimator in an experimental environment.

Traditional state estimation methods assume that the system state remains constant during the estimation period. However, in modern distribution power systems, these assumptions often come into question, for example, integration of renewable energy sources, such as solar and wind, brings about variability and unpredictability from changing weather conditions and lack of diversity on distribution level loads. Additionally, system disturbances, which can range from unexpected equipment failures to sudden demand surges, add to the challenges. Given these complexities, there is a pressing need for a means to allow the network to be reconfigured accordingly.

2.2. Dynamic state estimation

Dynamic state estimation (DSE) emerged in response to the inadequacies of traditional state estimation methods. Unlike its predecessors, DSE acknowledges the dynamic nature of power systems, furnishing more precise estimates of the system state (Zhao et al., 2019). It provides essential near-real-time data for proficient system operation and control. Utilizing time-series data, including voltage, current, and power flow measurements, DSE forecasts the power system's future state based on both historical and current data. These dynamic measurements, such as phase angle difference measurements, highlight the system's future generation-load balance and provide real-time insights into its behavior (Liu et al., 2021). This prediction hinges on a dynamic system model and the presently estimated state. DSE incorporates system dynamics into the estimation by employing mathematical models of components like generators and
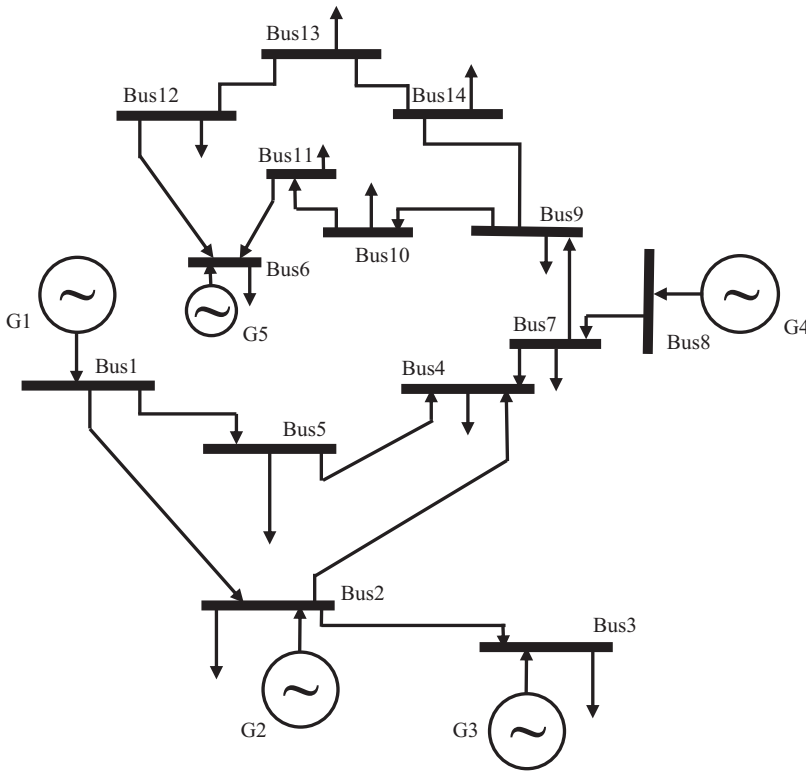
Figure 1. *IEEE 14-bus test network with 5 generators connected.*

transformers. These models account for the physical laws that dictate the operation of these components and their operational constraints.

One notable DSE method is the forecasting-aided state estimation (FASE) (Filho et al., 2009). While it typically yields satisfactory results for smooth-evolving input vectors, it outperforms simpler estimation techniques (Zhao et al., 2019). However, its assumption of Gaussian-distributed load data assume might not always be accurate at the distribution level. Despite this, FASE remains a valuable tool for security analysis and preventive control functions at higher voltage levels.

Most state estimation tools play a crucial role in real-time power system monitoring and control. However, its application in distribution networks is not widespread, primarily because customer loads change dynamically and are non-linear. This variability makes it challenging to obtain accurate pseudo-measurement estimates using Gaussian distributions, as commonly done at the transmission level. Typically, load values derived from standard load profiles serve as the foundation for these estimates (Manitsas et al., 2012). Yet, within the distribution system, these values are often completely unknown due to their high variability. This has led to a thorough examination of the characteristics of pseudo-measurements in the distribution network. Consequently, a number of statistical characteristics of pseudo-measurements in distribution networks have been modeled in various ways. A Gaussian mixture model, employed to estimate the accuracy of state estimation, is used in Singh et al. (2010a, 2010b), and a time-varying variance and mean model in Angioni et al., 2016). Furthermore, several machine learning methodologies have been deployed: an artificial neural network (ANN) model in tandem with a load profile approach is put forward in Manitsas et al. (2012), and a probabilistic neural network (PNN) is outlined in Gerbec et al. (2005).

To address this challenge, this paper proposes an innovative approach: employing transfer learning to predict these pseudo-measurements. This method expedites the training process and requires less data,

making it both efficient and cost-effective. The development of this approach, which is elaborated in Section 3, shows great promise for the safe and cost-efficient operation of power systems at the distribution level.

3. Using transfer learning to predict power distribution network pseudo-measurements

The assumption of traditional machine learning methods is that the feature space and data distribution characteristics of training data and test data are the same. When labeled training data is limited for the purpose of creating a machine learning model, transfer learning can be used to learn a more general model based upon easily available (source) data from a similar but different source, with subsequent adaption of the general model for application to a smaller data set that specifically represents the "targeted" application. Transfer learning is used to improve learners in one domain by transferring information from related domains (Pan and Yang, 2010).

Going further into the formalisms of transfer learning, two primary tasks emerge: the source task and the target task. The source task, denoted as D_S, is characterized by an abundance of data, which facilitates successful model training, resulting in a model represented as M_S. In contrast, the target task, denoted as D_T, serves as the principal objective but often struggles with limited data availability. The associated model for this task is represented as M_T. Central to transfer learning is its ability to leverage expertise—in the form of features, representations, or other insights—acquired from the source task to enhance performance on the target task (Pan and Yang, 2010). This process can be represented by the equation- $M_T = \text{Transfer}(M_S, D_T)$.

There are three main categories of transfer learning methods: inductive transfer learning, transductive transfer learning, and unsupervised transfer learning. Inductive transfer learning is used when the learning environment differs between the same task and source task. Transductive transfer learning seeks to learn within the same task but in a different environment and domain. Unsupervised transfer learning endeavors to discover the underlying structure of unlabeled data in both the target and source domains (Agarwal et al., 2021).

With the increasing complexity of power distribution networks and the mounting challenges in obtaining accurate measurements, there is an urgent need for innovative solutions to bridge this gap. To tackle this issue, this section develops an approach that harnesses the potential of transfer learning to predict pseudo-measurements at various substation buses on an LV distribution network. The scarcity of actual LV network measurements calls for an efficient method for pseudo-measurements prediction. Transfer learning, which draws upon knowledge from related domains, meets this challenge by speeding up the development time of machine learning models (i.e., training) and diminishing the associated data requirements.

In the LV network load forecasting tasks associated with this contribution, transfer learning is based on the Inductive Transfer Learning approach. In this scenario, 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 modeling 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.

The advantages of using this method are significant. First, training learning models for each specific application task starting from nothing requires substantial historical data and computational resources as well as data backhaul from the substation where the data was collected. Moreover, for many tasks, there might not be enough data available to train a deep-learning model without resulting in fitting issues. Inductive transfer learning can leverage pre-trained models that were trained on large datasets and adapt them to a specific task, even if the dataset is relatively small (Luo et al., 2019).

In Antoniadis et al. (2023), the authors use transfer learning to transfer the knowledge learned from the source electricity load data at a finer scale to improve predictions on target electricity load data at a network scale. The process begins by fitting a generalized additive model (GAM) to the source data.

The features estimated by the GAM are then used to create new features for the target dataset. Following this, the method computes an estimate of forecasting residuals on the target dataset. Finally, a random forest model is fitted on the augmented target dataset to predict the GAM residuals. The final forecasts are obtained by combining the GAM forecasts and the corrections provided by random forest.

In the field of energy forecasting, there are several research studies that have explored the concept of adjusting a forecast based on its anticipated errors or residuals, particularly in wind power prediction (Chen, 2022). They employ a multilayer deep neural network to predict errors, which are then used in a simpler, smaller turbine model. They leverage transfer learning to overcome the challenges associated with individually modeling each turbine. In the production forecasting of (Alolayan et al. (2022), they trained a deep neural network (DNN) model on abundant data from one county (the source model), then transferred the learned features (knowledge transfer layers) from this model to a new model (the target model) for a different county with limited data. To forecast the daily electric demand for specific customers (Hooshmand and Sharma, 2019), a CNN model is first trained on publicly available energy datasets to learn general features of the energy time series. The pre-trained CNN model is then fine-tuned using the limited data available for the target energy asset. Then, the model is evaluated on a held-out test set from the target asset's data to gauge its accuracy.

The transfer learning approaches used in energy forecasting often depend on machine learning models to articulate the relation between the predictor variables and the output variable. These models typically require learning from large volumes of data for optimal performance, which is nearly impossible to achieve on a distribution network where monitoring has only been deployed recently and in relatively small numbers. Additionally, maintaining and training numerous forecast models for each area of the network is very costly. In contrast, this paper introduces a novel approach utilizing online updated Bayesian methods. This mathematical and probabilistic framework offers greater transparency in the model's decision-making process and can effectively incorporate prior knowledge, thus reducing the reliance on extensive data.

4. A novel Bayesian transfer learning method in LV forecasting

In this section, a novel transfer learning prediction methodology will be introduced and developed, demonstrating how to satisfy the requirements of the transfer learning problem and the associated power network application conditions and constraints.

4.1. Day ahead forecast benchmark model

Before describing the transfer learning methodology, it is essential to define the "local" learning method, which serves as a base case representing no transfer. The benchmark for the day ahead forecast in load forecasting is derived from the model by Hong et al. (2011). This model, grounded in multiple linear regression, serves as a foundational approach for load forecasting. The mathematical representation of this model is provided below:

$$L(load) = \beta_0 + \beta_1 \times Trend + \beta_2 \times Day \times Hour + \beta_3 \times Month + \beta_4 \times Month \times TMP + \beta_5 \quad (4)$$

$$\times Month \times TMP^2 + \beta_6 \times Month \times TMP^3 + \beta_7 \times Month \times TMP + \beta_7 \times Month$$

$$\times TMP + \beta_8 \times Month \times TMP^2 + \beta_9 \times Hour \times TMP^3.$$

In the model, $\beta_0, \beta_1, \cdots, \beta_9$ are the regression coefficients, "Trend" represents the hour-by-hour trend, designated by natural numbers in ascending order. The variables "Hour", "Day", and "Month" correspond to the 24 hours in a day, the 7 days in a week, and the 12 months in a year. "TMP" is the local temperature (in °C). L represents data after the day following the day that all other parameters pertain to. The rationale for using this model is that it solely relies on temperature and calendar variables in load forecasting models, excluding past loads, which could be accessible due to regulatory restrictions or privacy concerns in many areas. The exclusion of past loads also serves to maintain the interpretability of

the model (Wang et al., 2016). Therefore, it is well-suited for transfer learning since it relates to weather and time (factors that do not hinder the transfer process with additional local data requirements), as these are common features in both the source and target domains. Moreover, it contains only the instantaneous load measurement (Singh et al., 2010a).

4.2. Updating Bayesian transfer learning method

Here, a running estimate of the forecast error at a target substation is derived to adjust the forecast model output at a source substation. Defining the observed load measurement for a substation as L and the predicted load for the same substation as $\hat{L}$. The real data for the source substation is then L_S. The real data for the source substation is then L_S and for the target substation is L_T. For a single substation, a day-ahead load forecast residual is defined as the difference between the real and predicted load. Hence, the residual e of a load L, for the transfer learning source will be

$$e_S = L_S - \hat{L}_S. \tag{5}$$

Similarly, for the target substation, the residual is

$$e_T = L_T - \hat{L}_T. \tag{6}$$

The relationship between $\hat{L}_S$ and $\hat{L}_T$ is that both utilize the same fitting model based on the source data, as defined in equation (4). However, each substation is influenced by different local weather conditions. This distinction can lead to significant errors when applying the same fitted model across different substations.

The day ahead residual of the load forecast is assumed to be drawn from a multivariate normal distribution with mean m and covariance E (Stephen et al., 2020):

$$e_T \sim N(m, E). \tag{7}$$

To infer substation level values of e_T, a novel transfer learning model using a running Bayesian estimate of error is proposed. It assumes the prior sample is e_T follows a multivariate normal distribution with mean and precision (which is the inverse of the covariance) m and E and with the posteriors over given e_T error distribution follows a Normal–Wishart distribution with mean μ and prior precision Λ.

$$p(\mu | \Lambda, e_T) = N(\mu; \mu_N, \theta_N \Lambda), \tag{8}$$

$$p(\Lambda | e_T) = W(\Lambda; \alpha_N, B_N). \tag{9}$$

For a half-hour resolution, day ahead forecast, which outputs a 48-dimensional forecast vector, the priors are set with the assumption that the value of the error is initially unknown. The initial parameters are defined as follows:

1. Mean prior: $\mu_0 = 0$ (This represents a 48-dimensional vector of zeros.)
2. Prior precision of the mean: $\theta_0 = 1$ (This scalar value is applied across all 48 dimensions.)
3. Prior precision: $\Lambda_0 = I_{48}$ (This represents a 48x48 identity matrix.)
4. Wishart distribution parameters: α_0 and B_0, α_0 is 49 and B_0 is a 48 x 48 identity matrix.

Then, the posterior mean of the observation of error will become

$$\mu_N = \frac{\Lambda_0 \mu_0 + N\overline{e_T}}{\Lambda_N}, \tag{10}$$

where N is the number of observed single-day load forecast error vectors.

$$\theta_N = \theta_0 + N. \tag{11}$$

From equation (8), it is evident that with each instance of newly observed data, its mean denoted as $\overline{e_T}$, the expected error data distribution μ_N, is updated. This update is based on the discrepancy between the most

recent prior data and the observed data. Consequently, the value of the expected error data is refined each time new true data is received, thereby enhancing its accuracy.

The updating precision matrix will become

$$\Lambda_N = \left[\frac{1}{\theta_N(\alpha_N - 1)}\right] B_N, \tag{12}$$

where

$$\alpha_N = \alpha_0 + \frac{N}{2} \tag{13}$$

and

$$B_N = B_0 + \frac{N}{2}\left[\overline{\theta} + \frac{\Lambda_0}{\Lambda_N}(\overline{e_T} - \mu_0)(\overline{e_T} - \mu_0)^T\right], \tag{14}$$

$$\overline{\theta} = \frac{1}{N}\sum_{n=1}^{N}(e_N - \overline{e_T})(e_N - \overline{e_T})^T. \tag{15}$$

The posterior mean of the transfer learning result based on the observed data is in equation (10) and its updating precision matrix is equation (12).

4.3. Conditional Bayesian transfer learning method

If a limited number of observations are gathered each day, subsequent predictions can be enhanced under Gaussian conditions, where both the data and errors are assumed to follow a normal distribution, characterized by a Gaussian distribution. Utilizing these observations, an improved method is proposed that employs a conditionally Gaussian distributed residual. For the joint residual distribution for a whole day, assume it follows a 48-dimensional multivariate Gaussian distribution. This distribution is represented by the matrix X, where the mean vector of the distribution is μ and its covariance matrix is Σ. The multivariate Gaussian distribution can be expressed as

$$f(x; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{n}{2}}} exp\left[-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right]. \tag{16}$$

This distribution may change based on specific conditions or observed data. Let ω denote the actual observed error data, represented as e_h to maintain clarity in the time format with other parameters. Here h represents the time interval during which the data is observed, and μ_h is its corresponding predicted value vector. Consider μ_t as the mean vector and Σ_t as the variance for a different time interval t of the predicted residual on the same day. Also, consider $\Sigma_{t,h}$ as the covariance between the intervals t and h. Adopting this approach offers a nuanced model that can refine residual predictions by estimating the discrepancies between observed and model-predicted values implied by the intra-day dependency structure. Formulating the conditional multivariate Gaussian with data and parameters can be expressed as (Stephen et al., 2020)

$$f(e_t | e_h = \omega) = N(\overline{\mu}, \overline{\Sigma}), \tag{17}$$

$$\overline{\mu} = \mu_t + \Sigma_{t,h}\Sigma_{h,h}^{-1}(\omega - \mu_h), \tag{18}$$

$$\overline{\Sigma} = \Sigma_t + \Sigma_{t,h}\Sigma_{h,h}^{-1}\Sigma_{h,t}, \tag{19}$$

where $\overline{\mu}$ represents the conditional predict mean and $\overline{\Sigma}$ represents the conditional predict variance at time interval t. This observation serves as a critical reference point in adjusting and refining the predictive model.
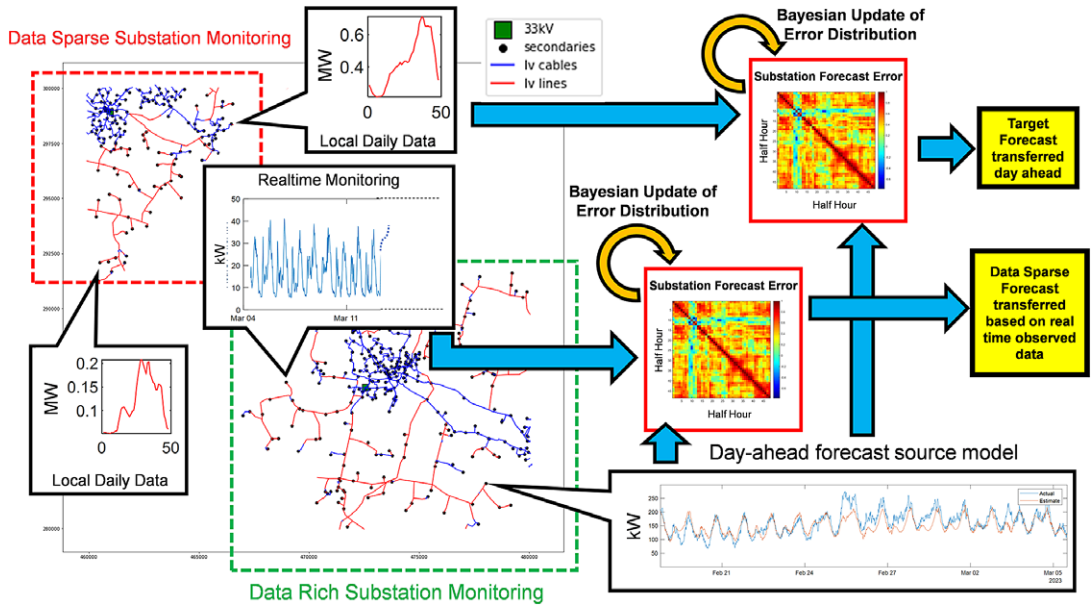
Figure 2. *Transfer learning of day ahead forecast for data-sparse substations from data-rich substation models using online Bayesian estimate of forecast error distribution.*

4.4. Process in Bayesian transfer learning

The flowchart in Figure 2 illustrates the process of the novel Bayesian transfer learning method. It begins with initial data collection and preprocessing in data-rich areas. Subsequently, a day ahead forecast benchmark model is applied in the data-rich area data, proceeds to the generation of errors generated from inputting data from data sparse into the model fit in data-rich area, serving as priors. The online Bayesian update error model is then employed to determine the day ahead posterior error. If the data-rich is supported by real time monitoring via a sim based modern, a conditional multivariate Gaussian model is utilized to further refine the model. The final prediction is obtained by the selection and application of the transfer model, integrating this error prediction into the original prediction.

5. Practical illustration case studies

Although transfer learning is carried out for forecasting pseudo-measurements of load, the operational value comes not from the forecast accuracy but from the increased accuracy in state estimation that it yields. Therefore, some additional metrics need to be investigated to understand the extent of operational value unlocked.

5.1. Forecast performance metrics

It is necessary to quantify the transferability of instance pairs with the same label from the source to the target domains. After testing the model, four different metrics are used to evaluate the effectiveness of the proposed transfer learning methodology against the baseline prediction methods. The four metrics (equations A, B, C, D) are used to gauge an improvement in performance (Weiss et al., 2016). At model testing, if all four error metrics show better performance when associated with the proposed transfer learning methodology, as compared to the baseline, this is taken as evidence of a positive transfer learning effect.

Mean absolute error (MAE) quantifies the average size of the error, regardless of its direction, to make sure there are no biases in the prediction model. Additionally, to assess whether the model tends to have

larger errors, the root mean square error (RMSE) is used, as it gives larger errors more weight, indicating the model's sensitivity to large discrepancies. Furthermore, since predictions occur at different scales, the mean absolute percentage error (MAPE) is also utilized. It expresses the error as a percentage of the actual value, offering an intuitive measure of prediction accuracy across various scales. Finally, R-squared (R^2) statistically represents the extent to which the predicted and actual values are correlated, offering insights into the model's goodness of fit across its entire range of outputs beyond mere error size. If the four error metrics showed better performance compared to the baseline, which involves less error measured and a higher R^2 value measured than the baseline model, which assumes the model is the same in the two areas, this is further evidence of a positive transfer learning effect. Additionally, a comparison is made between the transfer learning model and the locally trained model to determine whether transfer learning can achieve results that are nearly identical to the baseline model's best performance.

$$\text{MAE} = \frac{\sum_{k=1}^{N} |y_k - \widehat{y_k}|}{N}, \tag{20}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{k=1}^{N} (y_k - \widehat{y_k})^2}{N}}, \tag{21}$$

$$R^2 = 1 - \frac{\sum_k (y_k - \widehat{y_k})^2}{\sum_k (y_k - \overline{y})^2}, \tag{22}$$

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|. \tag{23}$$

5.2. Substation data and transfer learning result

The source substation data used in this study originates from an 11 kV substation in a typical rural area in GB – at 30 min resolution, over 12 month period, this can be considered data rich. As illustrated in Figure 3, the main load, represented by the red line, averages around 15 kW every half-hour. Conversely, the target dataset, represented by the blue line, is recorded at an urban substation in GB, with a load averaging around 200 kW.
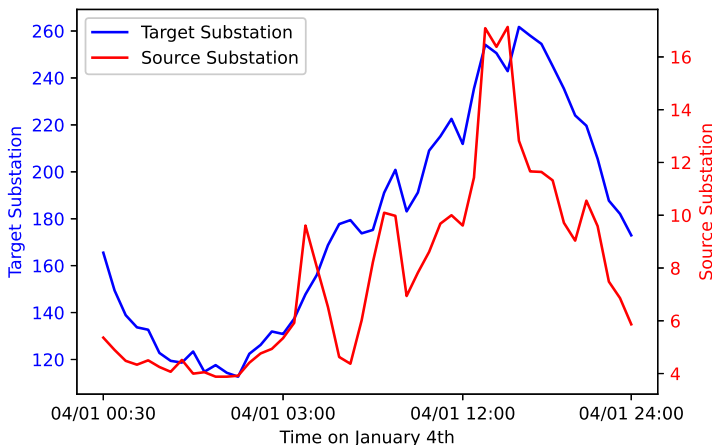

Figure 3. *Example source and target substation daily load profiles with varying scales.*

The transfer learning methodology outlined in Section 4 is now applied to determine the accuracy of predicting/forecasting load profiles in the target domain and assessed against the metric from Section 5.1. Equation (4) is used to forecast the source data 1 day ahead. In subsequent steps, equations (10) and (15) are employed to predict errors associated with transfer learning. Additionally, to validate the practicality of transfer learning, the benchmark model is also employed as a comparative measure in local learning to highlight the effectiveness of the transfer learning methodology, and localized forecasting is undertaken to demonstrate what the best-case scenario could be.

In Figure 4a, the benefits of the transfer learning approach are clearly demonstrated. Across all 10 substations, there is a significant reduction in prediction MAE error compared to the baseline, no negative transfer learning result was exhibited, decreasing the original forecast error from 200 to approximately 20 kW – almost in line with the ideal case of localized forecasting.

Based on Figure 4b and c, with data observed transmitted via a SIM-based modem at the start of the day, the conditional transfer learning method proves to be the most accurate for all substations. It exhibits a lower MAE and RMSE compared to other methods, even surpassing the benchmark in the same area. The reason for this is that with continuous observation of the target substation, the model can update daily to reflect short-term substation performance trends relative to short-term calendar and weather condition drivers. Meanwhile, the benchmark model remains static, retaining the training data characteristics as a linear regression model. This demonstrates the advantage of transfer learning, underscoring the efficacy of integrating condition monitoring into transfer learning models. Additionally, the plots highlight variations in prediction errors across substations, suggesting that some substations may have load patterns that are inherently more challenging to forecast.

Figure 5 provides a visualization of the true versus predicted values for each prediction method. Each point in the scatter plot represents a half-hourly load data point, with its true value given by the *x*-coordinate and its predicted value by the y-coordinate. In the upper tail of the transfer learning comparative scatter plot, there is a noticeable deviation between the *y* and *x* values. At peak values, the predicted values are significantly lower than the actual values. However, with the assistance of a few observations, conditional transfer learning addresses this issue effectively, highlighting the advantages of this method.

Figure 6 presents the temporal evolution of transfer learning performance, underscoring its ability to capitalize on previously acquired error knowledge. The trend indicates that standard transfer learning methods generally need a period of 3 to 4 days to integrate this knowledge efficiently. On the other hand, the conditional transfer learning method demonstrates notable efficacy from the beginning. Contrary to expectations, the local training approach fails to show the predicted incremental daily enhancement.

6. Dynamic state estimation case study

While error metrics highlight the predictive capability of a model, this does not translate into a model utility, that is, how the prediction affects the quality of a decision based on it. Accordingly, this section introduces two case studies for DSE utilizing the four different forecasting methods for pseudo-measurement prediction. The first case study focuses on an actual 22-bus GB local neighborhood network, while the second uses the more challenging UK General Distribution System (UKGDS) 77-bus network. Power flow results are taken as the ground truth, with state estimation serving as a metric to measure the effectiveness of transfer learning. State estimation plays a pivotal role in the proposed transfer learning methodology for predicting load data. It serves as a benchmark for evaluating the accuracy and efficacy of predictive models. By comparing estimated states with actual power flow data, a comprehensive view of the network's performance is obtained. This comparison not only reveals the network's tolerance for forecast errors but also facilitates a deeper understanding of its dynamics, aligning to enhance network analysis.

6.1. Representative GB test case

An urban neighborhood network is constructed based on a subset of feeders from a GB distribution license area - it serves a residential and light commercial customer base. The network is typical of GB Distribution
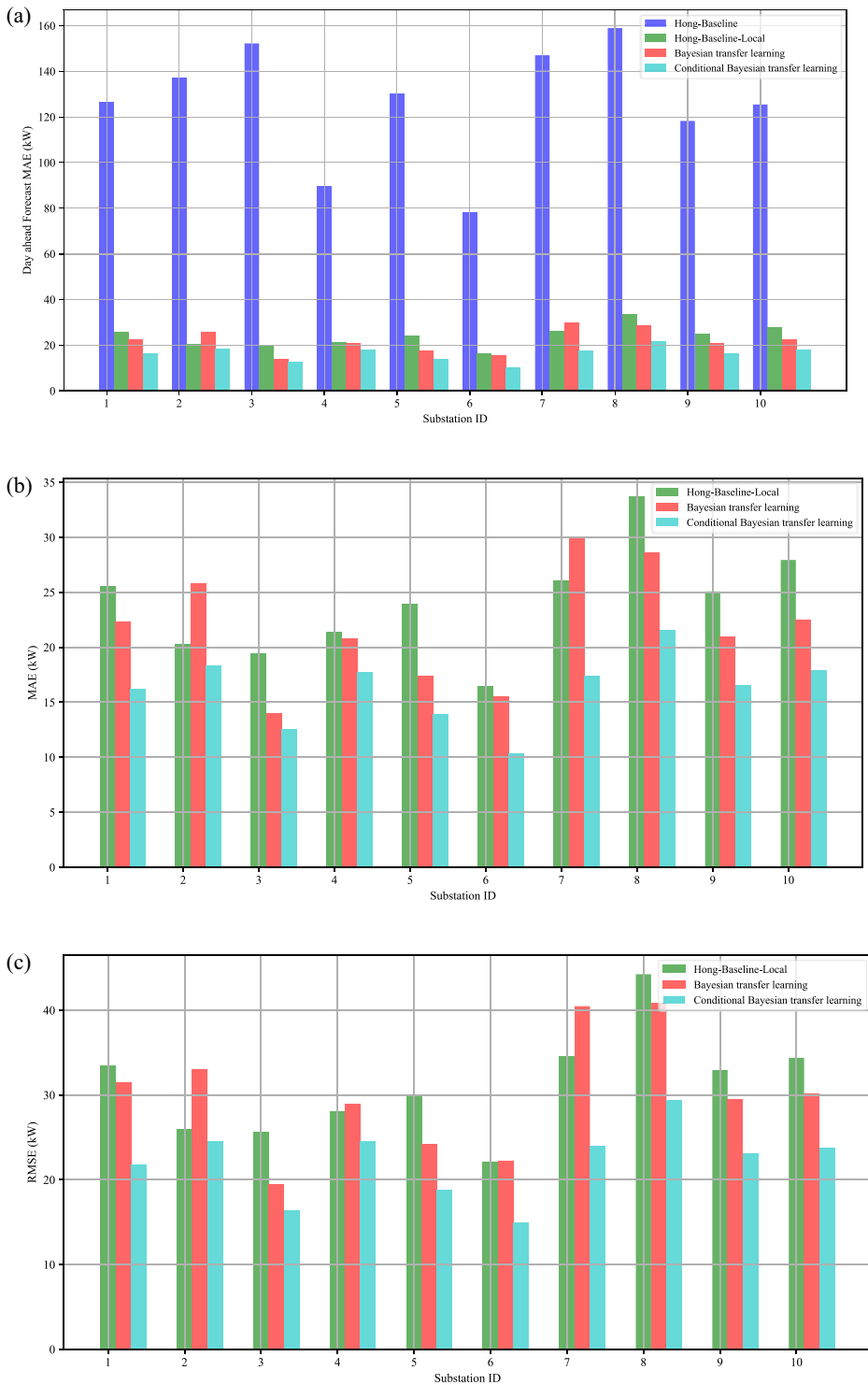
Figure 4. *a) 10 substations MAE comparison over four prediction methods; b) MAE comparison across three prediction methods at the substation. c) RMSE comparison across three prediction methods at the substation.*
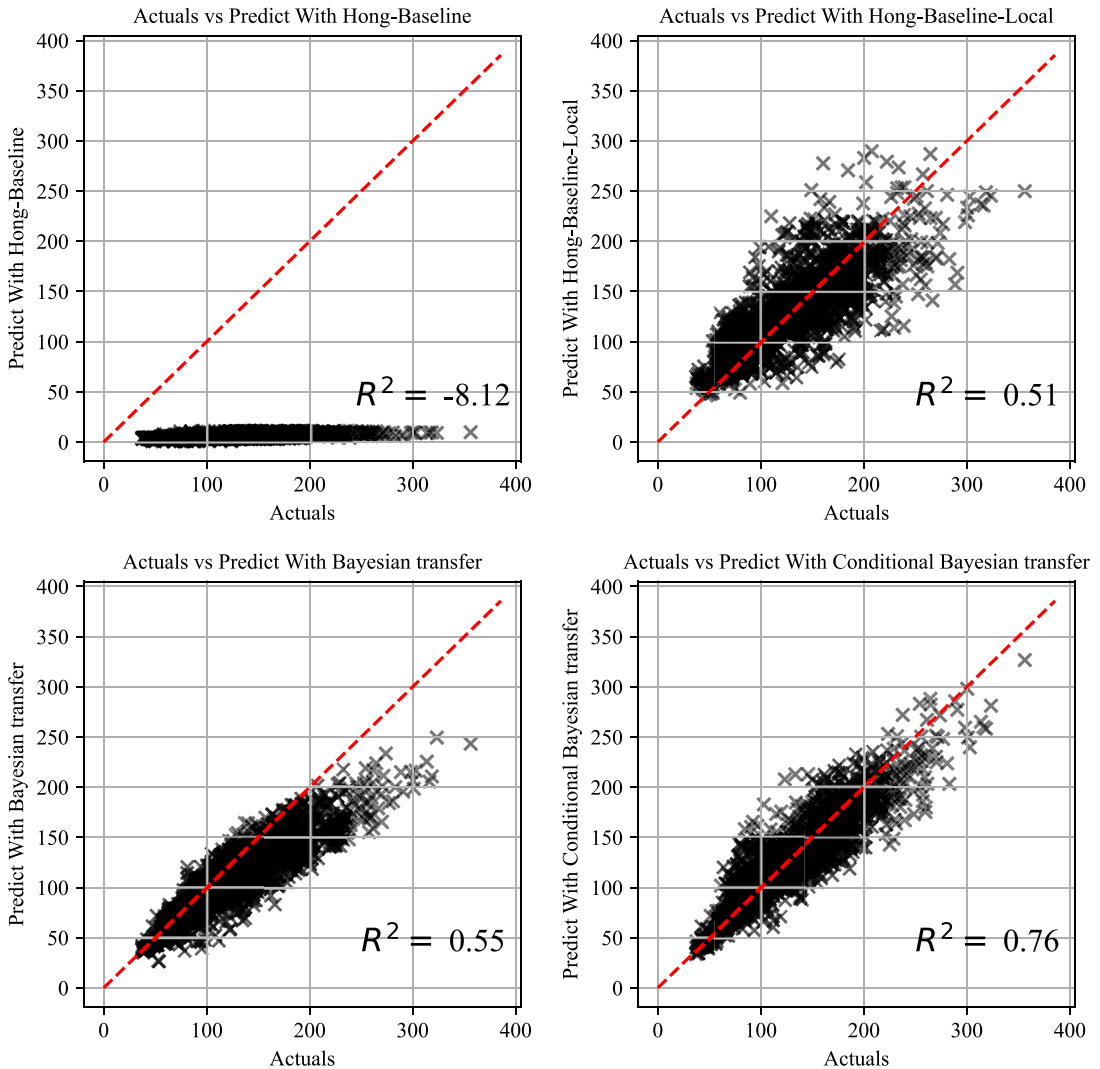
Figure 5. *Comparative prediction analysis for the four forecasting approaches.*

networks: it is fed at 33 kV, with a circuit voltage of 11 kV and the LV feeders branch off at 415 V before connecting to premises. The primary substation is designated as the Slack bus, which is used to balance the active and reactive power of the overall system. It is set as a reference point to measure angle and voltage throughout the network. This substation features a transformer line that steps down the voltage from 33 kV to 11 kV. Twenty loads are directly connected to the low-voltage bus of the transformer. Figure 7 presents the network map.

For the performance evaluation of the state estimators, the power flow result is designated as the true value or perfect information. The transfer learning value, derived from the active and reactive power of 20 low-voltage buses, is used as the pseudo-measurement for state estimation. Concurrently, the power flow values of the voltage magnitude and angles at bus 0 and 1 serve as the true measurements. These correspond to both the high and low-voltage transformer sides, which can be measured in the real world. Specifically, the slack bus 0 is set to have a voltage magnitude of 1 p.u. and an angle of 0, "p.u." means "per unit", representing normalized values relative to a base value. This normalization ensures consistency in representing quantities across different voltage scales. The transformers are set to be ideal,
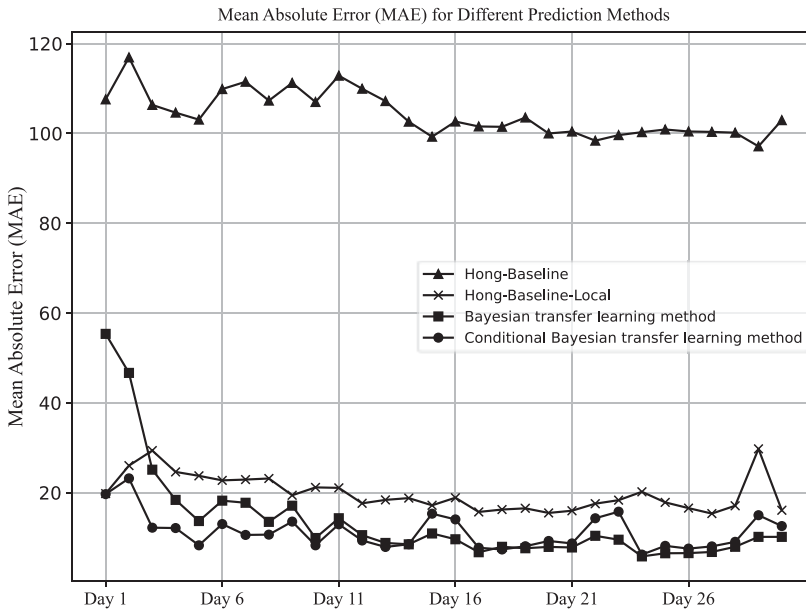
Figure 6. *MAE across days for the four distinct prediction methodologies for one substation.*
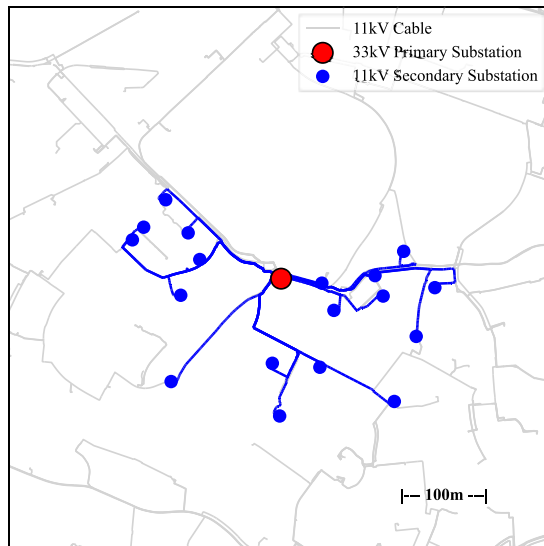

Figure 7. *GB urban neighborhood area 22-bus network.*

ignoring minor losses like core and winding losses. This means the bus on the low-voltage side of the transformer neither produces nor consumes power (Grainger and Stevenson, 1994). For the parameters in the state estimator, the true measurements are determined with a 0.5% standard deviation error, while a 5% standard deviation error is set for the pseudo-measurement. The state estimation voltage magnitude and angle comparisons for one low-voltage substation (bus #10), estimated over a period of 10 days, are displayed in Figures 8a, b and 9, respectively.

Figures 8a, b and 9 display the results of 10-day-long state estimations compared to the near ideal results obtained from the power flow calculations. Figure 9a presents the voltage magnitude results from
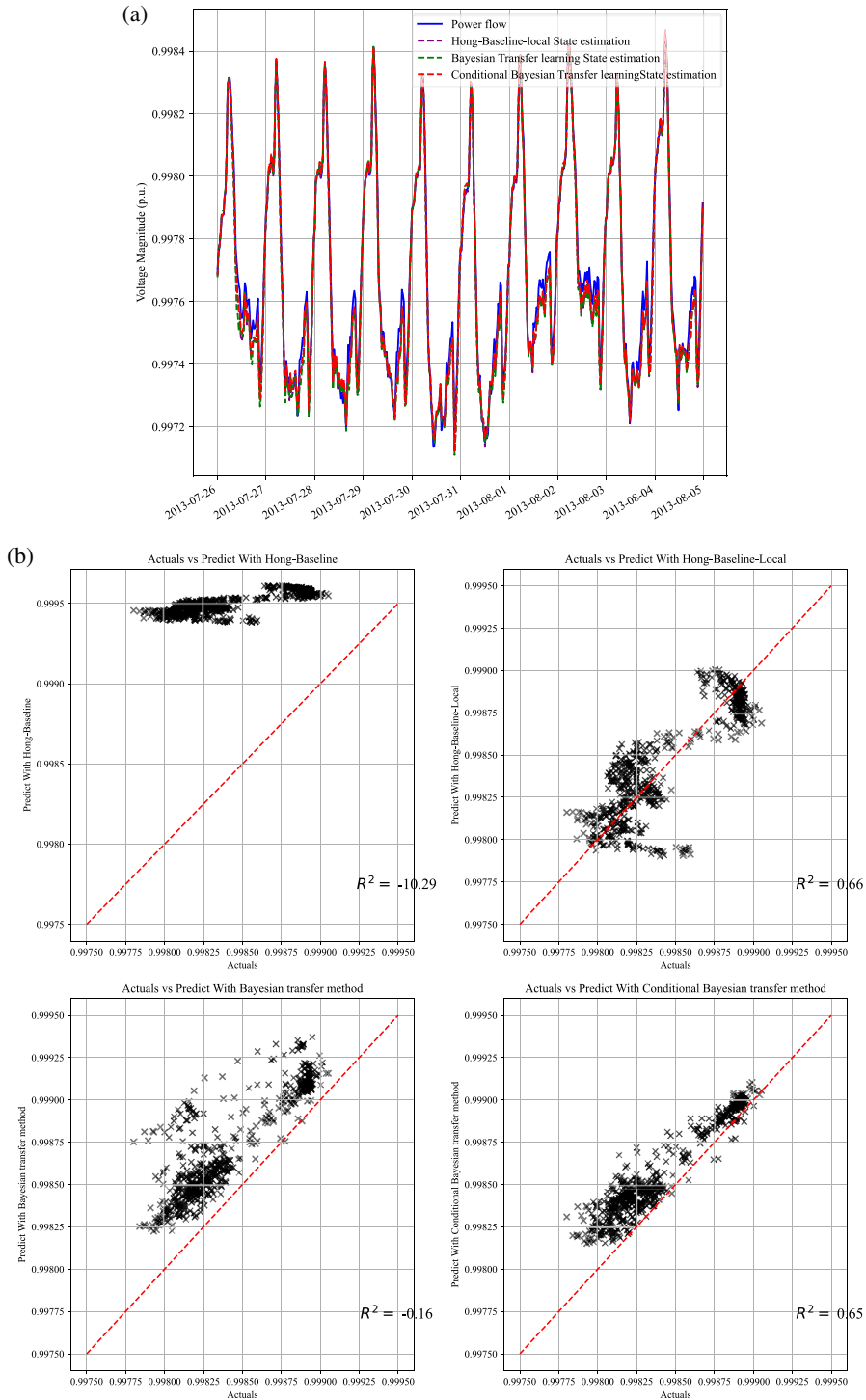
Figure 8. *a) Comparison of state estimation voltage magnitude results with those obtained through power flow, that is, the ideal case. b) Comparison of state estimation magnitude results with those obtained through power flow across the three transfer learning methods.*
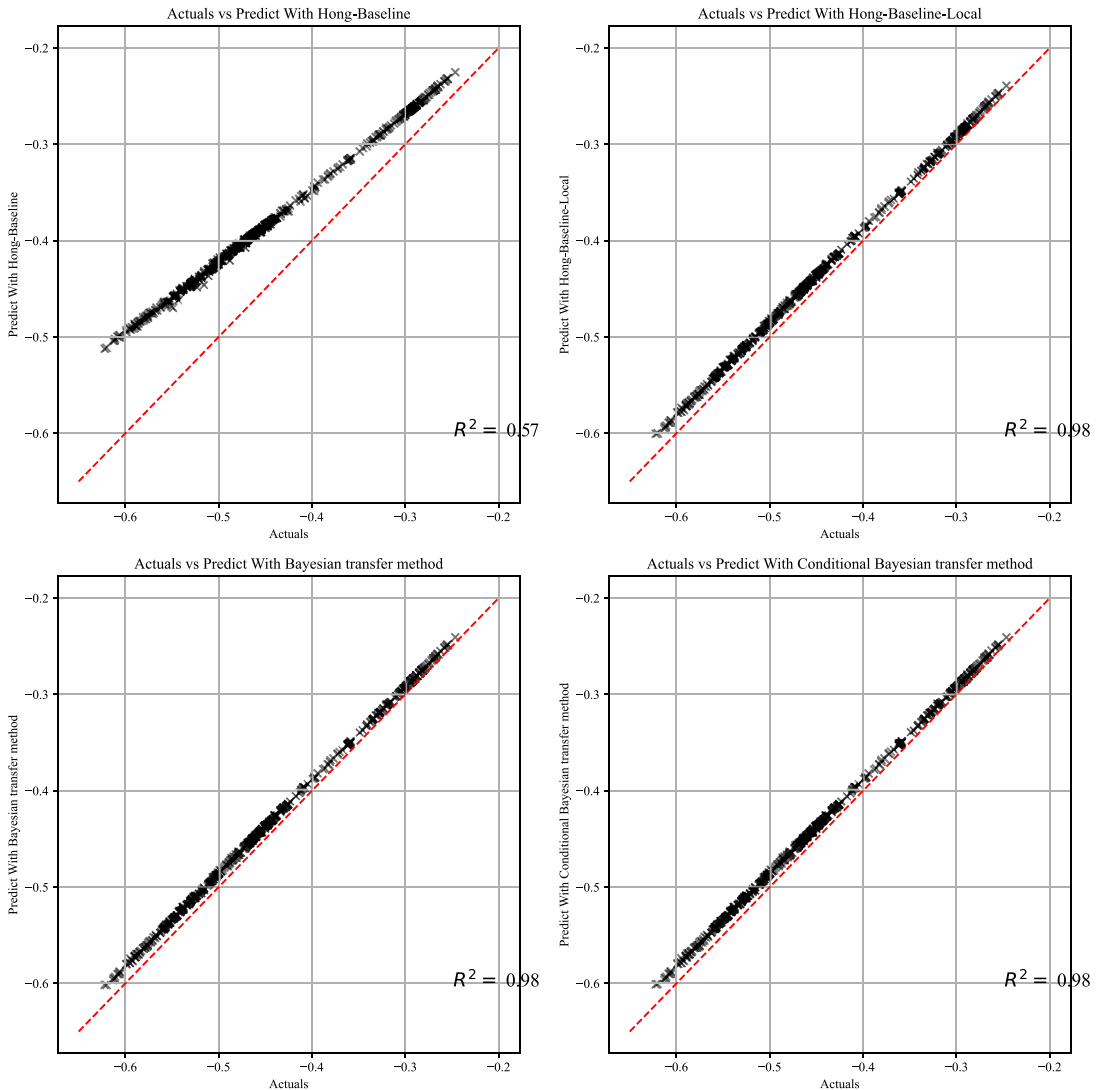
Figure 9. *Comparison of State Estimation Angle Results with those obtained through power flow.*

four methods. It is evident that the Hong-Baseline performs the poorest and is unable to provide meaningful estimation for the localized substation cases – demonstrating the ineffectiveness of a global forecast model solution. Figure 9b demonstrates that the other three methods closely follow the power flow results, indicating high accuracy. In Figure 10, the Hong-Baseline again performs the worst in terms of voltage angle, while the other three methods excel. These findings affirm the utility and effectiveness of transfer learning in forecasting network performance resulting from an anticipated load.

Figure 10 displays the error distribution results between each method and the power flow, depicted as a violin plot. This plot represents the kernel density estimate of the data distribution, showcasing a symmetrical spread and indicating data density at different values. The inner lines within the violin plot denote the data's quartiles, representing the 75%, 50%, and 25% quartiles. The results reveal that the Hong-Baseline has the highest error, ranging from 0.1 to 0.2. Given that the data is in per unit (p.u.), this error is substantial, equating to a real voltage error multiplied by 11,000. The remaining three methods exhibit considerably smaller errors.
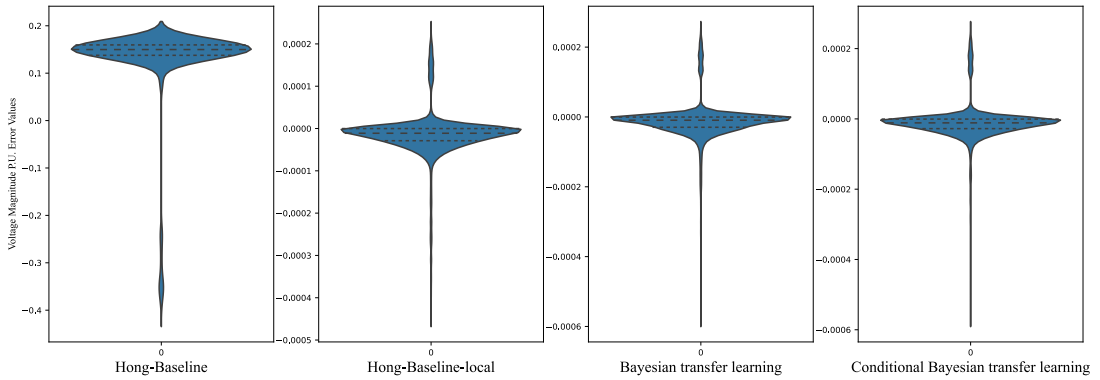
Figure 10. *Voltage magnitude error distributions on GB urban local network.*

Figure 11 further underscores the inferior performance of Hong-Baseline in voltage angle. However, Hong-Baseline-local primarily distributes between 0.005 to 0.015, highlighting the potential of linear regression. Both the transfer and conditional transfer methods yield similar results.

6.1.1 UKGDS network result

To achieve a more general representation of actual GB distribution networks, especially in urban areas or larger networks within the UK, and to test the scalability of the algorithm for deriving more robust and generalizable conclusions, it is also crucial to test the extremes, as UKGDS is intended to be a stylized representation of extremes. The UKGDS 77-bus network was also employed for testing. The UKGDS is a compilation of power system network models representative of UK distribution networks(UKGDS, 2015). Developed by the Centre for Sustainable Electricity and Distributed Generation (SEDG), the comprehensive UKGDS comprises overhead lines and cables at 132, 33, and 11 kV voltage levels with 281 bus bars and 322 branches, all supplied by four grid supply points. This research configures a segment of the UKGDS network, specifically utilizing 77 buses and 75 lines at the 11 kV level. The network map is shown in Figure 12.

Compared to the 22-bus model used in Section 6.1, this network features a transformer that steps down from 33 kV to 11 kV. Additionally, the 75-bus model is divided into several levels to connect to the transformer's low-voltage side. The transformer's high-voltage side is designated as the slack bus, with a voltage of 1 p.u. The intraday predicted load is input as pseudo-measurements into the 75-bus model. In this case study, the voltage magnitude and angle on both the high and low sides of the transformer are set as
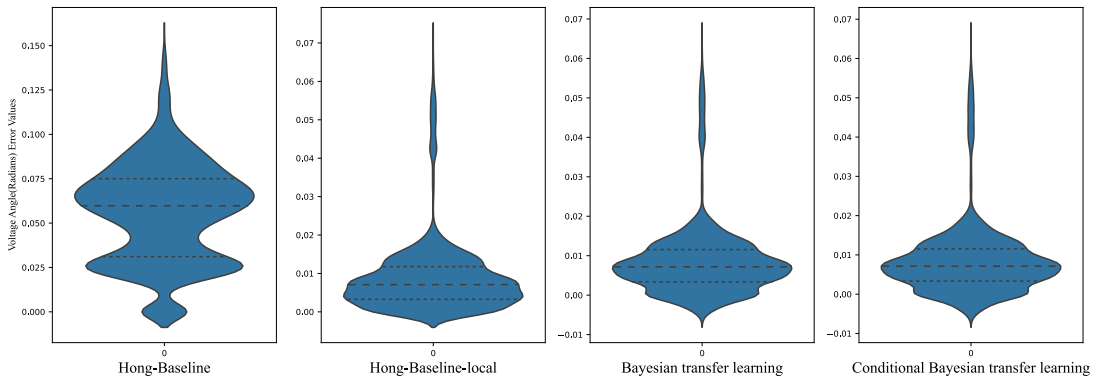

Figure 11. *Voltage Angle Error distributions on GB urban local network.*
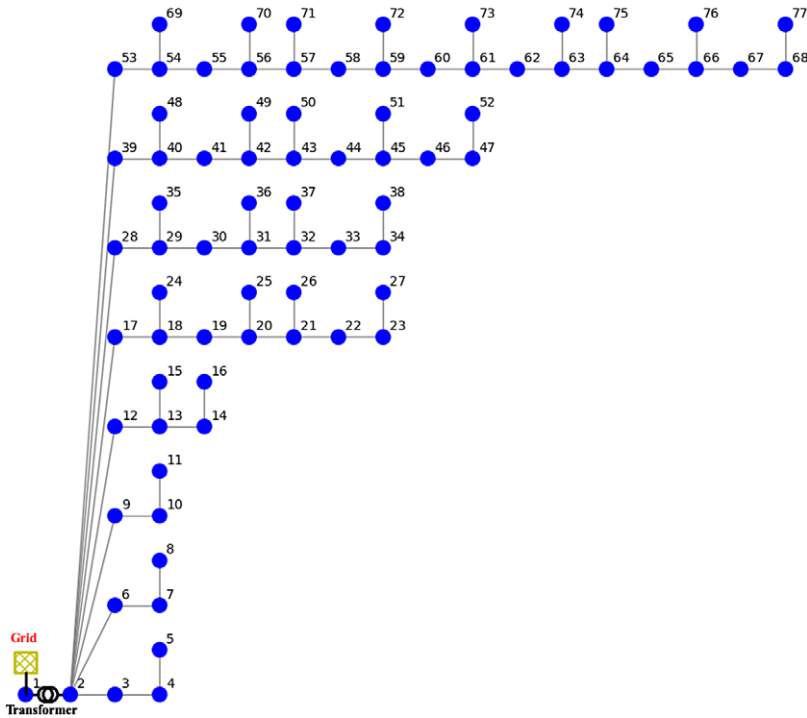
Figure 12. *UKGDS 77-bus low voltage test network.*

the true measurements. Similar to the 22-bus model, the true measurements are determined with a standard deviation error of 0.5%, while a 5% standard deviation error is applied to the pseudo-measurements.

Figures 13 and 14 depict the outcomes of 10-day-long state estimations in comparison to the power flow within the UKGDS network. Figure 8a illustrates the voltage magnitude results derived from four distinct methods. Even within a larger network, state estimation remains effective, closely mirroring the fluctuations of the power flow. This attests to the success of the purposed transfer learning methodology. Additionally, when compared with local training, it is evident that transfer learning can achieve comparable results.

Figure 15 presents the error distribution using a violin plot, with the layout and definition being the same as those in Figure 10. The results are also similar; the Hong-Baseline method exhibits the highest error, while the errors in other methods are significantly lower. The Baseline model exhibits the longest tails and a larger error magnitude, implying a significant possibility of incorrect forecasts and outliers. The medians of the remaining three methods are close to zero, indicating the accuracy of the predictions. The distribution from local training resembles a Gaussian distribution, validating the effectiveness of the benchmark. Transfer learning demonstrates more substantial errors in the upper tail, suggesting the model consistently underestimates the load profile's peak behavior. However, with the integration of additional observations, the conditional transfer learning method notably reduces the tail lengths, indicating its superior identification of peak values. Additionally, the error associated with the 75-bus model is greater compared to the 22-bus model. This indicates that the 75-bus model is less tolerant than the 22-bus model due to its increased complexity and reduced network redundancy.

Figure 16 further demonstrates the results in terms of voltage angle, highlighting the inferior performance of the Hong-Baseline method. Similar to the magnitude results, the transfer learning method may yield more extreme errors, whereas the conditional transfer learning method can reduce them by observation. Consistent with the outcomes from the 22-bus model, local training demonstrates superior
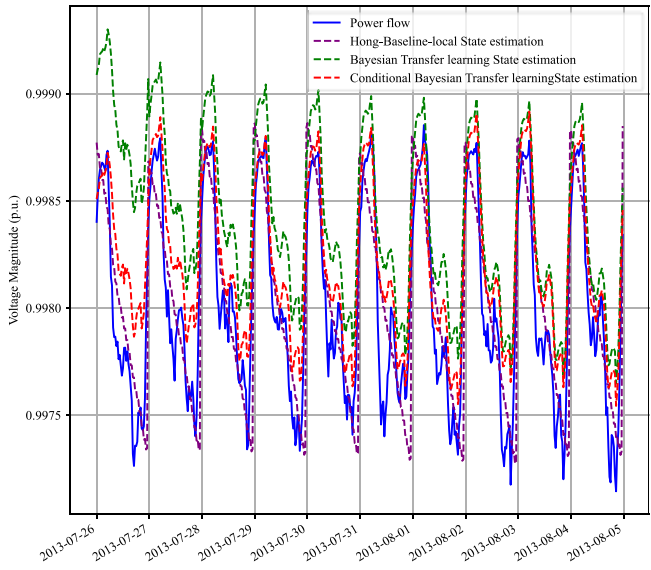
Figure 13. *Comparison of state estimation voltage magnitude results with power flow in UKGDS network.*
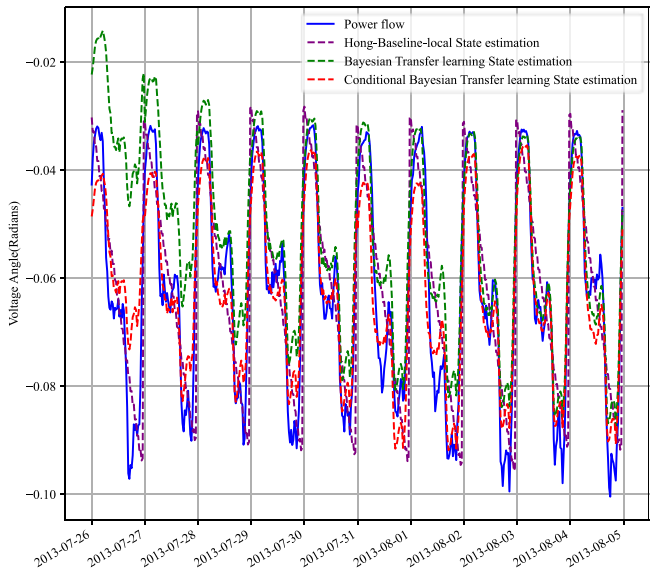

Figure 14. *Comparison of state estimation voltage angle results with power flow in UKGDS network.*

performance. This suggests that, in comparison to active power, reactive power exhibits greater variability across different regions, potentially owing to the diverse mix of industrial and commercial premises.

Table 1 presents the transfer learning result for the load forecasting in four methods, illustrating that with the help of a few data monitoring assist, the conditional Bayesian transfer learning has the highest accuracy in all three metrics, while the Bayesian transfer learning also has similar performance to the result compared to the local training, highlighting the effectiveness of the transfer learning across substations.
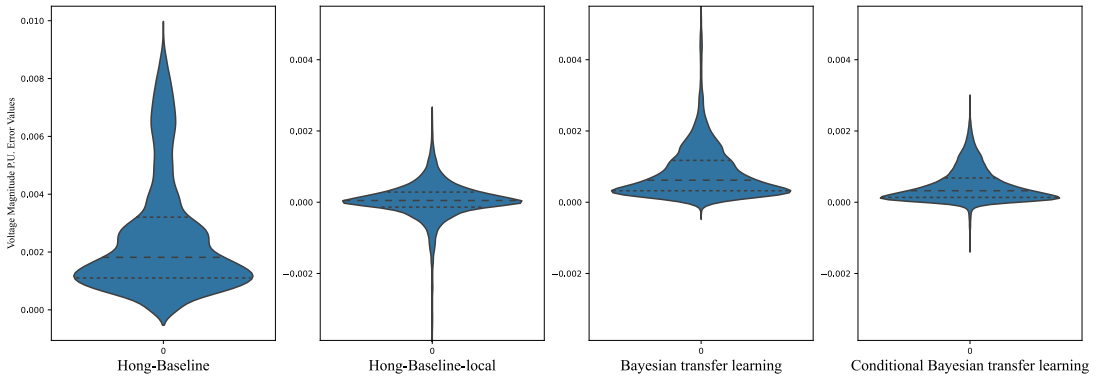
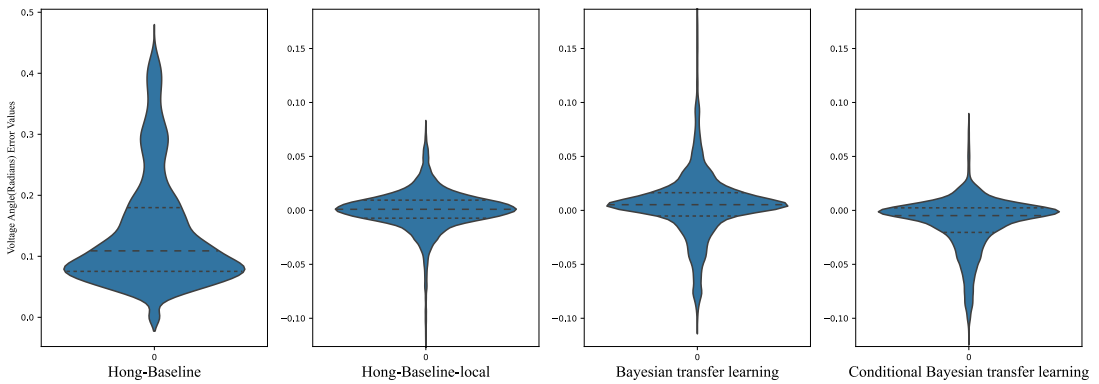***Figure 15.*** *Voltage magnitude error distributions in UKGDS network.*

Figure 16. *Voltage angle error distributions in UKGDS network.*

Table 1. *Performance comparison of substation day ahead active power forecast models (bold value represent best performance model)*

Forecast model	MAE	RMSE	MAPE
Hong–Baseline	108.371	122.643	99.304%
Hong–Baseline–local	19.202	26.157	21.109%
Bayesian transfer learning	18.509	26.330	16.740%
Conditional Bayesian transfer learning	**14.217**	**20.101**	**15.155%**

Table 2. *Performance comparison of transfer learning methods in state estimation of voltage magnitude in local network (bold value represent best performance model)*

State estimation: voltage magnitude	MAE	RMSE	MAPE
Hong–Baseline	0.157	0.162	15.717%
Hong–Baseline–local	3.086×10^{-5}	5.938×10^{-5}	0.00309%
Bayesian transfer learning	$\mathbf{2.940 \times 10^{-5}}$	$\mathbf{5.543 \times 10^{-5}}$	**0.00295%**
Conditional Bayesian transfer learning	2.943×10^{-5}	5.685×10^{-5}	**0.00295%**

Table 3. *Performance comparison of transfer learning methods in state estimation of voltage angle in local network (bold value represent best performance model)*

State estimation angle (Local network)	MAE	RMSE	MAPE
Hong–Baseline	0.0564	0.0629	13.31%
Hong–Baseline–Local	0.0091	0.0134	2.08%
Bayesian transfer learning	**0.0088**	0.0129	2.05%
Conditional Bayesian transfer learning	**0.0088**	**0.0128**	**2.04%**

Table 2, 3, 4 and 5 show the results of transfer learning in dynamic state estimation for voltage magnitude and voltage angle within the local network and UKGDS 77-bus network, respectively. When compared to the local network, the errors in the 77-bus network are significantly higher, which can be attributed to reduced network redundancy and increased complexity. The results indicate that local training yields the best outcomes in both categories. However, the results of transfer learning for voltage magnitude are also very good, with only a 0.05% increase in percentage error. This level of accuracy can lead network operators to make the right decisions. An incorrect voltage magnitude prediction can be a source of misinformation, potentially reducing the operators' situational awareness. Similar to the findings in Table 3, the performance in voltage angle estimation is notably poorer than expected. Further study is required to mitigate inefficiencies in power transfer in cables and to prevent system oscillations, which could potentially lead to network instabilities (Meegahapola and Littler, 2015).

7. Conclusion

Distribution System Operation will be essential to support low-carbon technology adoption in end-use and embedded generation applications. Flexibility to alleviate network congestion through either voltage or thermal limit management can only be realized with accurate and location-specific forecasting,

Table 4. *Performance comparison of transfer learning methods in state estimation voltage magnitude in UKGDS network (bold value represent best performance model)*

State estimation magnitude (UKGDS)	MAE	RMSE	MAPE
Hong–Baseline	0.00254	0.00326	0.255%
Hong–Baseline–local	**0.0003**	**0.000539**	**0.0350%**
Bayesian transfer learning	0.0008	0.00112	0.0851%
Conditional Bayesian transfer learning	0.0004	0.00066	0.0486%

Table 5. *Performance comparison of transfer learning methods in state estimation voltage angle in UKGDS network (bold value represent best performance model)*

State estimation angle (UKGDS)	MAE	RMSE	MAPE
Hong–Baseline	0.142	0.16881	319.83%
Hong–Baseline–local	**0.0127**	**0.01887**	**37.974%**
Bayesian transfer learning	0.0211	0.0318	102.983%
Conditional Bayesian transfer learning	0.0177	0.02740	117.114%

however, predicting load at the distribution network is challenging due to the limited observability of the network, high monitoring costs, and heterogeneous load behavior. This paper has introduced a methodology that utilizes transfer learning derived from Bayesian inference, leveraging data from data-rich operational environments and applying their adapted learnings to less well-observed network locations. Two distinct approaches have been compared: one in which a limited number of observations are reported each day and another where no continuous data is observed except at the measurement point. The comparison is based on the practicality and expense of monitoring power networks at the distribution level. The findings highlight that transfer learning significantly reduces forecast errors compared to benchmarks. The accuracy further improves when prior prediction data is available, even outperforming direct local predictions. This accurate estimate feeds through into practical use in dynamic state estimation, where multiple substations need to be forecast, and the ultimate consequence of error manifests in the quality of the state estimate. In operation, this method would not only reduce monitoring expenses by using external data sources but also offer considerable commercial and operational benefits, especially as the demand for distribution network balancing services grows as more connections are needed. Future work will focus on refining the model to better represent the extremes of load profiles and identifying the algorithmic improvements required to optimize state estimation for distribution network level load co-behaviors.

References

Agarwal N, **Sondhi A**, **Chopra K and Singh G** (2021) Transfer learning: Survey and classification. In *Smart Innovations in Communication and Computational Sciences: Proceedings of ICSICCS 2020*, UK government, pp. 145–155.

Ahmad F, **Rasool A**, **Ozsoy E**, **Sekar R**, **Sabanovic A and Elitaş M** (2018) Distribution system state estimation-A step towards smart grid. *Renewable and Sustainable Energy Reviews 81*, 2659–2671. https://doi.org/10.1016/j.rser.2017.06.071.

Alolayan OS, **Raymond SJ**, **Montgomery JB and Williams JR** (2022) Towards better shale gas production forecasting using transfer learning. *Upstream Oil and Gas Technology 9*, 100072. https://doi.org/10.1016/j.upstre.2022.100072.

Angioni A, **Schlösser T**, **Ponci F and Monti A** (2016) Impact of Pseudo-Measurements From New Power Profiles on State Estimation in Low-Voltage Grids. *IEEE Transactions on Instrumentation and Measurement 65*(1), 70–77. https://doi.org/10.1109/TIM.2015.2454673.

Antoniadis A, **Gaucher S and Goude Y** (2023) Hierarchical transfer learning with applications to electricity load forecasting. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2023.04.006.

Chen H (2022) Knowledge distillation with error-correcting transfer learning for wind power prediction. *arXiv e-prints, arXiv: 2204.00649*. 10.48550/arXiv.2204.00649.

Dehghanpour K, **Wang Z**, **Wang J**, **Yuan Y and Bu F** (2019) A survey on state estimation techniques and challenges in smart distribution systems. *IEEE Transactions on Smart Grid 10*(2), 2312–2322. https://doi.org/10.1109/TSG.2018.2870600.

Manitsas E, **Singh R**, **Pal BC and Strbac G** (2012) Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling. *IEEE Transactions on Power Systems 27*, 1888–1896.

Filho MBDC, **Souza JCS d and Freund RS** (2009) Forecasting-aided state estimation—Part II: Implementation. *IEEE Transactions on Power Systems 24*(4), 1678–1685. https://doi.org/10.1109/TPWRS.2009.2030297.

Gerbec D, **Gasperic S**, **Smon I and Gubina F** (2005) Allocation of the load profiles to consumers using probabilistic neural networks. *IEEE Transactions on Power Systems 20*(2), 548–555. https://doi.org/10.1109/TPWRS.2005.846236.

Grainger JJ and Stevenson WD (1994) *Power system analysis*. McGraw-Hill series in electrical and computer engineering.

Hong T, **Wang P and Willis HL** (2011). A Naïve multiple linear regression benchmark for short term load forecasting. Paper presented at *the 2011 IEEE Power and Energy Society General Meeting*.

Hooshmand Aand Sharma R (2019). Energy predictive models with limited data using transfer learning. Paper presented at the *Proceedings of the Tenth ACM International Conference on Future Energy Systems, Phoenix, AZ, USA*. https://doi.org/10.1145/3307772.3328284

Kerai M (2023) *Smart Meter Statistics in Great Britain.* Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1143890/Q4_2022_Smart_Meters_Statistics_Report.pdf (accessed 21 March 2024).

Li R, Gu C, Li F, Shaddick G and Dale M (2015) Development of low voltage network templates—Part I: Substation clustering and classification. *IEEE Transactions on Power Systems* 30(6), 3036–3044. https://doi.org/10.1109/TPWRS.2014.2371474.

Liu Y, Singh AK, Zhao J, Meliopoulos APS, Pal B, Ariff MAbM, Van Cutsem T, Glavic M, Huang Z, Kamwa I, Mili L, Mir AS, Taha A, Terzija V and Yu S (2021) Dynamic state estimation for power system control and protection. *IEEE Transactions on Power Systems* 36(6), 5909–5921. https://doi.org/10.1109/TPWRS.2021.3079395.

Luo G, Yang Y, Yuan Y, Chen Z and Ainiwaer A (2019) Hierarchical transfer learning architecture for low-resource neural machine translation. *IEEE Access* 7, 154157–154166.

Ma K, Li R and Li F (2017) Utility-scale estimation of additional reinforcement cost from three-phase imbalance considering thermal constraints. *IEEE Transactions on Power Systems* 32(5), 3912–3923. https://doi.org/10.1109/TPWRS.2016.2639101.

Manitsas E, Singh R, Pal BC and Strbac G (2012) Distribution system state estimation using an artificial neural network approach for pseudo measurement modeling. *IEEE Transactions on Power Systems* 27(4), 1888–1896. https://doi.org/10.1109/TPWRS.2012.2187804.

Meegahapola L and Littler T (2015) Characterisation of large disturbance rotor angle and voltage stability in interconnected power networks with distributed wind generation. *IET Renewable Power Generation* 9(3), 272–283.

Navarro-Espinosa A and Ochoa LF (2016) Probabilistic impact assessment of low carbon technologies in LV distribution systems. *IEEE Transactions on Power Systems* 31(3), 2192–2203. https://doi.org/10.1109/TPWRS.2015.2448663.

Oatley CJ, Ramsay B, McPherson A, Eastwood R and Ozveren CS (1997) A decision support system for electricity distribution network refurbishment projects. *Electric Power Systems Research* 40(1), 27–35.

Pan SJ and Yang Q (2010) A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* 22(10), 1345–1359. https://doi.org/10.1109/TKDE.2009.191.

Rowe M, Yunusov T, Haben S, Singleton C, Holderbaum W and Potter B (2014) A peak reduction scheduling algorithm for storage devices on the low voltage network. *IEEE Transactions on Smart Grid* 5(4), 2115–2124. https://doi.org/10.1109/TSG.2014.2323115.

Schweppe FC and Wildes J (1970) Power system static-state estimation, Part I: Exact model. *IEEE Transactions on Power Apparatus and Systems* PAS-89(1), 120–125. https://doi.org/10.1109/TPAS.1970.292678.

Singh R, Pal BC and Jabr RA (2010a) Distribution system state estimation through Gaussian mixture model of the load as pseudo-measurement. *IET Generation, Transmission & Distribution* 4(1), 50–59.

Singh R, Pal BC and Jabr RA (2010b) Statistical Representation of Distribution System Loads Using Gaussian Mixture Model. *IEEE Transactions on Power Systems* 25(1), 29–37. https://doi.org/10.1109/TPWRS.2009.2030271.

Stephen B, Telford R and Galloway S (2020) Non-Gaussian residual based short term load forecast adjustment for distribution feeders. *IEEE Access* 8, 10731–10741. https://doi.org/10.1109/ACCESS.2020.2965320.

Táczi I, Sinkovics B, Vokony I and Hartmann B (2021) The challenges of low voltage distribution system state estimation—An application oriented review. *Energies* 14(17). https://doi.org/10.3390/en14175363.

Tang W, Sun B, Feng R, Huang C and Zhao L (2022) A distribution network reconfiguration continuous method based on efficient solution space coding. Paper presented at *the CECNet*.

UKGDS (2015) UKGDS: The United Kingdom Generic Distribution System. Retrieved from https://github.com/sedg/ukgds#readme

Vanin M, Acker TV, D'hulst R and Hertem DV (2023) Exact modeling of non-Gaussian measurement uncertainty in distribution system state estimation. *IEEE Transactions on Instrumentation and Measurement* 72, 1–11. https://doi.org/10.1109/TIM.2023.3287253.

Wang P, Liu B and Hong T (2016) Electric load forecasting with recency effect: A big data approach. *International Journal of Forecasting* 32(3), 585–597. https://doi.org/10.1016/j.ijforecast.2015.09.006.

Weiss K, Khoshgoftaar TM and Wang D (2016) A survey of transfer learning. *Journal of Big Data* 3(1), 1–40.

Zhao J, Gómez-Expósito A, Netto M, Mili L, Abur A, Terzija V, … Meliopoulos APS (2019) Power System Dynamic State Estimation: Motivations, Definitions, Methodologies, and Future Work. *IEEE Transactions on Power Systems* 34(4), 3188–3198. https://doi.org/10.1109/TPWRS.2019.2894769.