Forecast-Aided State Estimation in Unbalanced Distribution Networks Using Smart Meter Data under Limited Communication Bandwidth

Aditya Rangarajan*, Rahul K. Gupta†, Daniel K. Molzahn‡, and Line Roald*

***Abstract*—With the growing integration of stochastic renewable generation and adaptable resources in electrical distribution systems, distribution utilities are increasingly eager to improve the visibility of their networks using distribution system state estimation (DSSE). However, scarcity of measurements and limited communication bandwidth challenges the ability of the distribution utilities to estimate distribution system states. This paper presents a forecast-aided real-time state estimation method for distribution networks, using forecasts for nodes lacking direct measurements. While other recent studies have also used forecast-aided state estimation methods, existing approaches require large amounts of historical data to train the forecasting model or depend on phasor measurements, both of which are not easily accessible to distribution utilities. In contrast, we introduce a joint forecasting and state estimation methodology. Our forecasts are generated by a Vector-Autoregressive (VAR) model, which is recursively trained as new measurements are acquired, and thus does not rely on a full set of historical data or phasor measurements. These forecasts, together with the available measurements, are subsequently used for DSSE. We validate the effectiveness of our approach on the IEEE 123-bus benchmark network, taking into account various correlation assumptions and differing quantities of accessible measurements.**

***Index Terms*—State estimation, distribution networks, vector autoregressive model (VAR), LinDistflow, low observability.**

ABBREVIATIONS

DER	Distributed Energy Resources
EV	Electric vehicle
PV	Photovoltaic
DSO	Distribution System Operators
DSSE	Distribution System State Estimation
VAR	Vector Autoregression
YW	Yule-Walker
RLS	Recursive least squares
RMSE	Root mean square error
MAE	Mean absolute error

I. INTRODUCTION

Distribution networks are constantly changing with the rapid integration of DERs, such as EV chargers and privately owned solar PV installations. This evolution has, on one hand, boosted flexible resources within distribution grids. On the other hand, it has also led to challenges with maintaining adequate voltages and avoiding overloads due to the significant magnitude and stochastic nature of solar PV generation and EV demand. These factors have increased the volatility of power injections within power distribution systems, causing line flow congestion, transformer overloads, voltage imbalance, and overvoltages. As DSOs are required to ensure quality of supply while adhering to the physical limits of the network [1]–[3], they are increasingly focused on enhancing the visibility of their networks through DSSE. Additionally, the growing number of flexible resources in the distribution system brings increased risks, such as cyber attacks [4]. DSSE provides DSOs with the real-time situational awareness needed to react appropriately when threats arise.

Normally, state estimation involves an overdetermined system of equations where the number of measurements is higher than the number of states to be estimated. However, implementing state estimation in power distribution systems is challenging as the number of available measurements is usually limited. In addition, distribution grids also have many branches and nodes, each with associated states, resulting in an under-determined system of equations. The extensive installation of the smart meters in recent years could fill the gap of measurements as illustrated in [5]–[7], therefore enhancing the observability of DSSE. However, the key challenge is the communication of the smart meter data to a central server where DSSE is hosted. This issue arises from the limited investment in the communication infrastructure in the distribution systems. Currently, smart meters are typically pinged either monthly for billing purposes or daily [8], [9], with a maximum time resolution of 15 minutes. This limitation is primarily due to the high cost of communication, as noted in [10], [11], with measurements recorded at intervals ranging from 15 minutes to an hour. Therefore, even if there are enough measurements, limitations in the communication bandwidth make it challenging to transmit these measurements at a high sampling rate to the DSO. Thus, the lack of measurements and limited communication bandwidth motivates the development of DSSE methods relying on sparse measurements.

A. Related Work

Several methods in the literature consider solving the state estimation problem in contexts with limited measurements. Such methods can broadly be categorized into three different

The authors acknowledge funding from PSERC project T-67: Smart Meter-Driven Distribution Grid Visibility and Control.

*Dept. of Electrical and Computer Engineering, UW Madison, WI, USA. {arangarajan4, roald}@wisc.edu

†School of Electrical Engg. and Computer Science, Washington State University, Pullman, WA, USA. rahul.k.gupta@wsu.edu

‡School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, USA. molzahn@gatech.edu

approaches. The first approach relies on algorithms for *compressed sensing* and *matrix completion*, where the problem of low observability is addressed by leveraging sparsity properties of the DSSE problem while it remains underdetermined. For example, [12] proposed compressing the measurements based on correlation information among the nodal voltages. The work in [13] assumed that the substation injected current is much higher than the load current, resulting in a sparse problem. References [14] and [15] propose approaches based on matrix completion, which is generally used to estimate missing values in low-rank matrices. The key drawback of these approaches is that they are highly sensitive to measurement noise and require significant computational effort.

The second approach relies on *pseudo-measurements* to address the problem of sparse measurements [16]. Pseudo-measurements specify estimated power consumption or generation based on historical observations. For example, [16] proposes using standard profiles of different consumers based on clustering. Reference [17] uses artificial neural networks to create pseudo-measurements. The key drawback of these methods is that they depend on fixed heuristics that are determined based on past observations and are not updated frequently. Heuristics based on historical data may no longer be valid due to the stochastic nature of injections in distribution systems with high levels of solar PV generation and EV demand and may be outdated particularly fast in areas where the number of solar PV installations and EVs is rapidly changing. Furthermore, the construction of heuristics requires access to historical data, which is not necessarily easy to obtain for technical and privacy reasons.

The third approach, which is the focus of our work, relies on a two-stage framework referred to as *forecast-aided* DSSE. In the first stage, a forecasting model is used for predicting measurement values for the non-observed nodes. In the second stage, a DSSE is solved considering the measurements from the observed nodes and the predicted values of the non-observed nodes, making the DSSE problem overdetermined. References [18]–[22] propose different approaches to achieve forecast-aided DSSE. Reference [18] uses load forecasts in the state estimation process; however, their motivation was to improve the estimation performance, rather than to tackle the challenge of low observability. References [20] and [21] consider using the forecasting model in a low-observable setting and showed that forecasting schemes can make the DSSE observable. However, the proposed scheme assumes measurements from phasor measurement units (PMUs), which are not usually available at the distribution level. Further, this scheme has a long offline training phase that assumes access to a large set of historical data from all nodes, which is not necessarily realistic. The method in [21] also utilizes a separate forecasting model for each node and ignores the spatial correlation with injections from the other nodes. Other drawbacks of the method in [21] is that they use a Gaussian process offline to train the forecast model and that the forecast model is not updated as new measurements arrive. Such a scheme may fail when distribution injections have substantial real-time variations. The works in [19] and [22] also propose forecast-aided DSSE, but rely on a large number of measurements from PMUs and remote terminal units, which are unavailable in typical distribution systems.

In summary, existing forecast-aided DSSE methods assume the availability of phasor measurements, a large number of historical measurements to train the forecast model, and/or do not account for the spatial correlation between different nodal measurements. We tackle these limitations of existing work as described in the next section.

B. Contributions

We address the above-mentioned drawbacks of the forecast-aided DSSE by (i) developing a new recursive forecasting model that relies only on existing smart meter measurement and communication infrastructure, and (ii) integrating this forecasting model with the state estimation step. We demonstrate the efficacy of the proposed method on a case study. We describe each contribution in more detail below.

- **Forecasting model:** We propose a forecasting model that relies only on smart meters measurements that are already present in the grid, thus allowing DSSE without the installation of PMUs, as required by the methods proposed in [20]–[22]. To comply with existing bandwidth limitations in the communication with smart meters, our models only rely on accessing a few measurements at a time and are updated recursively. We achieve this using a recursive least-square approach, which is computationally inexpensive to train and can be updated in each timestep based on measurements from a small number of locations in the network. As a result, we do not rely on a large amount of historical data to train the forecasting model, as assumed by the previous work in [20]–[22].
- **State Estimation:** We utilize the above-mentioned forecasts along with smart meter measurements from a limited number of nodes to solve a linear DSSE. The DSSE is formulated as a linear least squares estimator which is based on the linearized Lin3DistFlow approximation of the power flow equations [23]–[25]. Unlike previous research where the smart meter locations are fixed, the proposed DSSE is capable of reliably estimating the network state with a limited set of smart meter measurements whose locations can be changed dynamically.
- **Validation:** We validate the proposed scheme on the unbalanced IEEE 123-bus benchmark network and show that the proposed forecast-aided approach successfully achieves good estimation performance while directly measuring only 10% of the total states at any time.

The paper is organized as follows. Section II gives a schematic overview of the proposed approach. Section III introduces the linearized power flow model, which is later used for state estimation. Section IV presents the forecasting schemes. Section V describes the state estimation scheme. Section VI presents the numerical setup and results. Finally, Section VII concludes the work.

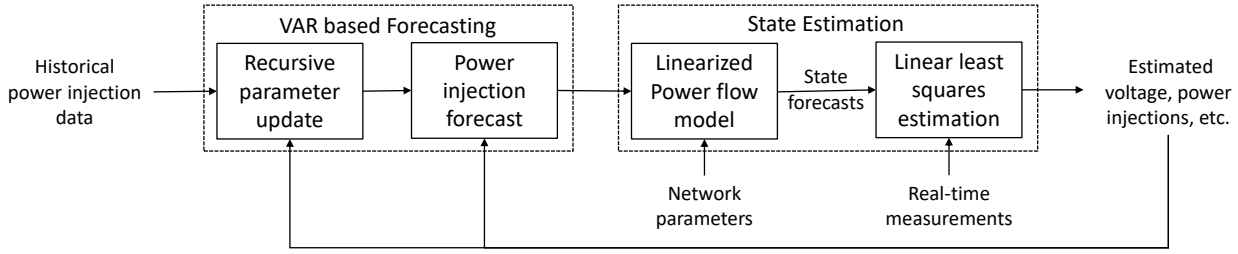


Fig. 1: Overview of the proposed state estimation framework, partial real time measurements

II. Proposed Approach

In this section, we describe the notation used in the paper and provide an overview of the proposed approach.

A. Notation

We consider a generic unbalanced power distribution network represented by the graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{E}$ is the set of all distribution lines. $\mathbf{A}$ is the incidence matrix of the graph $\mathcal{G}$ with the column corresponding to the slack bus removed. The nodes in the network are indexed from $1, 2, \ldots, N$ where node 1 represents the substation. The three phases are denoted by the superscripts a, b, and c. The lines are indexed from $1, \ldots, E$, where $E = N - 1$, as distribution networks are typically operated in a radial manner.

Throughout the paper, we use **boldface** to represent vectors and matrices, while normal fonts are used to represent scalars and elements of vectors/matrices. The vectors $\mathbf{v}$, $\boldsymbol{\theta}$, $\mathbf{P}$, and $\mathbf{Q} \in \mathbb{R}^{3N}$ represent the squared voltage magnitudes, phase angles, and active and reactive power injections, respectively. The vectors $\mathbf{p}$ and $\mathbf{q} \in \mathbb{R}^{3E}$ denote active and reactive power flows along lines. We use single subscripts i and j to represent quantities associated with the corresponding node in the network, e.g., $\mathbf{v}_i$ and $\mathbf{P}_i$ for node i. Double subscripts ij are use to represent quantities associated with the distribution line connecting nodes i and j, e.g., $\mathbf{p}_{ij}, \mathbf{q}_{ij}$. Let $\mathbf{I}_n$, $\mathbb{0}_n$ and $\mathbb{1}_n$ be the $n \times n$ identity matrix, the length-n zero vector, and the length-n vector of ones, respectively. The operators $\mathtt{Tr}(\cdot)$ and $\mathtt{vec}(\cdot)$ represent the trace and vectorization of matrices, respectively. The operators $\mathtt{diag}(\cdot)$ and $\mathtt{blkdiag}(\cdot)$ are used to represent diagonal and block-diagonal matrices, respectively. Let $^\top$ and $^{-\top}$ denote the transpose and inverse transpose of a matrix, and $\|\cdot\|$ represents the ℓ_2-norm of a vector. Finally, $\mathbb{E}(\cdot)$ and $\mathfrak{N}(\cdot)$ represent the expectation operator and normally distributed random variables, respectively.

B. Overview

Each node in the network is assumed to be equipped with a smart meter capable of measuring voltage magnitudes $\mathbf{v}$ as well as active and reactive power injections $\mathbf{P}$ and $\mathbf{Q}$. However, due to limited communication bandwidth, the DSO can only access measurements from a subset of smart meters, denoted by $\mathcal{M} \subset \mathcal{N}$, at any given time. Given this limitation, the DSO's objective is to estimate the network state, i.e., voltage magnitudes $\mathbf{v}$, phase angles $\boldsymbol{\delta}$, and power injections $\mathbf{P}$ and $\mathbf{Q}$ for all nodes, using the available measurements. This communication bottleneck challenges the DSO's ability to achieve an accurate distribution system state estimate. As only a few measurements are available at any given time, the state estimation problem is underdetermined. Therefore, the key difficulty lies in developing a DSSE approach that can accurately estimate the network state when only a limited number of smart meters are sampled due to bandwidth constraints.

In this paper, we tackle the challenge of limited bandwidth constraints by developing a forecast-aided DSSE that is capable of reliably estimating the network state using a small number of dynamically sampled smart meter measurements, i.e., the set of sampled smart meters changes at each time instant. The proposed scheme consists of two stages, as shown in Fig. 1. The first stage implements a recursive forecasting scheme based on vector auto-regressive (VAR) model, which is trained using power injection measurements from a time-varying subset of smart meters. The trained VAR model is used to predict power injections at all nodes in the network for the next time step using data from the previous time step. Subsequently, the second stage performs linear least-squares estimation to provide both the voltage magnitudes and phase angles and refine the power injection forecasts. This second step uses both voltage magnitude and power injection measurements from the observed nodes. To account for load variability during the day, the measurements are used to recursively refine the forecast model to keep it updated with the new load information. We model the grid using a linearized approximation of the power flow equations known as the Lin3DistFlow model [23]–[25], resulting in a linear DSSE. While we use the Lin3DistFlow approximation for state estimation, we evaluate our proposed method by comparing the estimated values to the true states obtained by solving a non-linear AC power flow. The prediction error of the forecasting model and the noise in smart meter measurements are modeled as additive white Gaussian noise with zero mean.

In the following sections, we present the key components of the proposed framework which are the power flow model, the forecasting scheme, and the state estimation formulation.

III. Three-phase LinDistFlow

In this section, we describe the linearized three-phase DistFlow approximation (Lin3DistFlow) [23]–[25] used in our model. To simplify the non-linear DistFlow equations, the

Lin3DistFlow approximation assumes that the system is lossless, the voltage angle differences across lines are small, and the voltage magnitudes are close to unity. As a consequence of these assumptions, the voltage phasors are assumed to be almost balanced. Under these assumptions, the voltages at adjacent nodes i and j are related as:

$$\Delta \mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j = \mathbf{H}^p_{ij}\mathbf{p}_{ij} - \mathbf{H}^q_{ij}\mathbf{q}_{ij}, \tag{1a}$$

$$\Delta \boldsymbol{\delta}_{ij} = \boldsymbol{\delta}_i - \boldsymbol{\delta}_j = -\mathbf{H}^q_{ij}\mathbf{q}_{ij} - \mathbf{H}^p_{ij}\mathbf{p}_{ij}, \tag{1b}$$

where $\Delta \mathbf{v}_{ij} \in \mathbb{R}^3$ represents the difference of squared voltage magnitudes $\mathbf{v}_i, \mathbf{v}_j \in \mathbb{R}^3$ at nodes i and j. Similarly, $\Delta \boldsymbol{\delta}_{\boldsymbol{ij}} \in \mathbb{R}^{\mathbf{3}}$ represents the difference in phase angles $\boldsymbol{\delta}_i, \boldsymbol{\delta}_{\boldsymbol{j}} \in \mathbb{R}^3$ at nodes i and j. The quantities $\mathbf{p}_{ij}$ and $\mathbf{q}_{ij}$ are the active and the reactive power flows on the distribution line connecting nodes i and j. Finally, the matrices $\mathbf{H}^p_{ij}, \mathbf{H}^q_{ij} \in \mathbb{R}^{3\times 3}$ are

$$\mathbf{H}^P_{ij} = \begin{bmatrix} -2r^{aa}_{ij} & r^{ab}_{ij} - \sqrt{3}x^{ab}_{ij} & r^{ac}_{ij} + \sqrt{3}x^{ac}_{ij} \\ r^{ba}_{ij} + \sqrt{3}x^{ba}_{ij} & -2r^{bb}_{ij} & r^{bc}_{ij} - \sqrt{3}x^{bc}_{ij} \\ r^{ca}_{ij} - \sqrt{3}x^{ca}_{ij} & r^{cb}_{ij} - \sqrt{3}x^{cb}_{ij} & -2r^{cc}_{ij} \end{bmatrix}, \tag{2a}$$

$$\mathbf{H}^Q_{ij} = \begin{bmatrix} -2x^{aa}_{ij} & x^{ab}_{ij} + \sqrt{3}r^{ab}_{ij} & x^{ac}_{ij} - \sqrt{3}r^{ac}_{ij} \\ x^{ba}_{ij} - \sqrt{3}r^{ba}_{ij} & -2x^{bb}_{ij} & x^{bc}_{ij} + \sqrt{3}r^{bc}_{ij} \\ x^{ca}_{ij} + \sqrt{3}r^{ca}_{ij} & x^{cb}_{ij} + \sqrt{3}r^{cb}_{ij} & -2x^{cc}_{ij} \end{bmatrix}. \tag{2b}$$

Using (1a), we can express the voltage difference $\Delta \mathbf{v} \in \mathbb{R}^{3E}$ across all lines as

$$\begin{aligned} \Delta \mathbf{v} &= \mathtt{vec}(\Delta \mathbf{v}_{ij}), \\ &= \begin{bmatrix} \mathtt{blkdiag}(\mathbf{H}^p_{ij}) & \mathtt{blkdiag}(\mathbf{H}^q_{ij}) \end{bmatrix} \begin{bmatrix} \mathtt{vec}(\mathbf{p}_{ij}) \\ \mathtt{vec}(\mathbf{q}_{ij}) \end{bmatrix}, \\ &= \begin{bmatrix} \mathtt{blkdiag}(\mathbf{H}^p_{ij}) & \mathtt{blkdiag}(\mathbf{H}^q_{ij}) \end{bmatrix} \mathbf{S}_{\mathbf{f}}, \end{aligned} \tag{3}$$

where $\mathbf{S}_{\mathbf{f}} = \begin{bmatrix} \mathtt{vec}(\mathbf{p}_{ij})^\top & \mathtt{vec}(\mathbf{q}_{ij})^\top \end{bmatrix}^\top \in \mathbb{R}^{6E}$ is the vector of active and reactive branch flows.

For radial networks, we can express the squared voltage magnitudes $\mathbf{v}$ and the nodal power injections $\mathbf{S}$ as

$$\mathbf{v} = \mathbf{v}_1 \mathbb{1} + \mathbf{A}^{-1}\Delta \mathbf{v}, \tag{4}$$

$$\mathbf{S} = \mathbf{A}^\top \mathbf{S}_{\mathbf{f}}, \tag{5}$$

where $\mathbf{v} = [\mathbf{v}_2^\top, \ldots, \mathbf{v}_N^\top]^\top \in \mathbb{R}^{3N}$, $\mathbf{S} = \begin{bmatrix} \mathbf{P}_2^\top, \ldots, \mathbf{P}_N^\top, \mathbf{Q}_2^\top, \ldots, \mathbf{Q}_N^\top \end{bmatrix}^\top \in \mathbb{R}^{6N}$, $\mathbf{v}_1$ is the voltage at the substation, and $\mathbf{P}_i$ and $\mathbf{Q}_i$ are the active and reactive power injections at node i, respectively. Combining (3), (4), and (5) results in

$$\mathbf{v} = \mathbf{v}_1 \mathbb{1} + \begin{bmatrix} 2\mathbf{R} & 2\mathbf{X} \end{bmatrix} \mathbf{S} = \mathbf{v}_1 \mathbb{1} + \mathbf{M}_v \mathbf{S}, \tag{6}$$

where $\mathbf{R} = \mathbf{A}^{-1}\,\mathtt{blkdiag}(\mathbf{H}^p_{ij})\mathbf{A}^{-\top}$, $\mathbf{X} = -\mathbf{A}^{-1}\,\mathtt{blkdiag}(\mathbf{H}^q_{ij})\mathbf{A}^{-T}$, and $\mathbf{M}_v = \begin{bmatrix} 2\mathbf{R} & 2\mathbf{X} \end{bmatrix}$. Similarly, we can derive the following expression for the voltage phase angles $\boldsymbol{\delta}$:

$$\boldsymbol{\delta} = \boldsymbol{\delta}_1 \mathbb{1} + \begin{bmatrix} \mathbf{X} & -\mathbf{R} \end{bmatrix} \mathbf{S} = \boldsymbol{\delta}_1 \mathbb{1} + \mathbf{M}_\delta \mathbf{S}, \tag{7}$$

where $\boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_N \end{bmatrix}^\top \in \mathbb{R}^{3N}$, $\boldsymbol{\delta}_1$ is the voltage angle at the substation, and $\mathbf{M}_\delta = \begin{bmatrix} \mathbf{X} & -\mathbf{R} \end{bmatrix}$.

IV. Forecasting

We next describe the forecasting model used to predict the power injections at every node. The power injection forecasts, in conjunction with the Lin3DistFlow model, are used to forecast the state variables, which include the squared voltage magnitude ($\mathbf{v}$) and the voltage angles ($\boldsymbol{\delta}$), and to perform state estimation, as shown in Fig. 1.

In the literature, there are two broad classes of methods used for forecasting load and renewable generation: (a) Machine learning (ML) based methods and (b) statistical time series models like autoregressive processes (AR), autoregressive moving average (ARMA) processes, etc. Recent studies highlight the application of ML-based models to forecast loads and renewable generation [26]–[29]. While ML-based methods can accurately capture complicated non-linear relationships between loads, these methods often require large datasets for effective training, which makes them computationally expensive. Consequently, ML-based methods are less effective at adapting to short-term variations in power injections.

On the other hand, statistical methods such as ARMA processes are simpler, require less data, and are computationally more efficient. Several studies have highlighted the utility of different statistical methods for short term load and renewable generation forecasts [19], [30]–[33]. Due to the low computational effort involved in training these statistical models, they can adapt quickly to changes in the network, making them suitable for short-term forecasting. Hence, AR-type models are a better choice in the context of forecast-aided DSSE. In addition to the short-term temporal correlations, power injections also exhibit spatial correlations influenced by geographical and behavioral factors [34]. To capture both spatial and temporal correlations effectively, we selected a vector autoregressive (VAR) model.

A. Vector Autoregressive Model (VAR)

We next provide a detailed description of the vector autoregressive (VAR) model for predicting nodal power injections and the associated training process. Consider the predicted nodal power injections at time t, $\hat{\mathbf{S}}_t$, described by an order-p VAR(p) model:

$$\hat{\mathbf{S}}_t = \sum_{i=1}^{p} \mathbf{B}_i \mathbf{S}_{t-i} + \boldsymbol{\varepsilon}_t, \tag{8}$$

where $\mathbf{S} = [\mathbf{P}, \mathbf{Q}]^\top \in \mathbb{R}^{6N}$ is the vector of active and reactive power injections, $\mathbf{B}_p, \ldots, \mathbf{B}_p \in \mathbb{R}^{6N\times 6N}$ are the VAR model's coefficient matrices, and $\boldsymbol{\varepsilon}_t \sim \mathfrak{N}(0, \boldsymbol{\Sigma}_s)$ is a zero-mean Gaussian noise with a covariance of $\boldsymbol{\Sigma}_s \in \mathbb{R}^{6N\times 6N}$. The temporal correlation in the power injections $\mathbf{S}$ is captured by the diagonal terms of $\mathbf{B}_i$ and the covariance matrix $\boldsymbol{\Sigma}_s$, while the spatial correlations are captured by the off-diagonal terms of $\mathbf{B}_i$ and $\boldsymbol{\Sigma}_s$. While potentially providing improved forecasting accuracy, there are certain drawbacks to using higher-order models. *First*, higher-order models have many more parameters, which increases the amount of data and

computational effort required for training. *Second*, higher-order models can result in overfitting, which can reduce the accuracy improvements outside of the training setting. Thus, we select a VAR(1) model, which has been shown to be adequately accurate for many short-term forecasting applications [32], [35]:

$$\hat{\mathbf{S}}_t = \mathbf{B}\mathbf{S}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{9}$$

The VAR(1) model, described by (9), is used to predict values for the power injections at time step t ($\hat{\mathbf{S}}_t$) based on the power injections at the previous time step ($\mathbf{S}_{t-1}$). While the forecasting model is simple, we face two hurdles: (i) we have to determine appropriate values for the matrix $\mathbf{B}$ and (ii) we do not have access to all the measurements $\mathbf{S}_{t-1}$ from the previous time step. In the following subsection, we will first describe two methods to determine $\mathbf{B}$ while assuming that we could access all measurements simultaneously. Then we will describe our proposed approach for relaxing this requirement.

B. Estimating the Coefficient and Covariance Matrices

This subsection describes three methods for estimating the coefficient matrix $\mathbf{B}$ and the covariance matrix $\boldsymbol{\Sigma}_s$ for the prediction error $\boldsymbol{\varepsilon}_t$ in (9).

1) Yule-Walker: The Yule-Walker (YW) method is a standard method for estimating the parameters of autoregressive models [36]. Multiplying both sides of (9) on the right with $\mathbf{S}_{t-1}^\top$ and taking the expected value yields

$$\boldsymbol{\Pi}(1) = \mathbf{B}\boldsymbol{\Pi}(0), \tag{10}$$

where $\boldsymbol{\Pi}(0)$ and $\boldsymbol{\Pi}(1)$ represent the auto-covariances of $\mathbf{S}$ at lags of 0 and 1 time periods, respectively. These can be estimated using past observations of $\mathbf{S}$ as follows:

$$\hat{\boldsymbol{\Pi}}(j) = \frac{1}{M}\sum_{i=1}^{M} \mathbf{S}_i \cdot \mathbf{S}_{i+j}^\top, \quad j \in \{0, 1\}, \tag{11}$$

where M is the number of past observations available for training the VAR(1) model. The coefficient matrix $\mathbf{B}$ is estimated by solving the system of equations (10). Finally, the forecasting error covariance is computed as

$$\hat{\boldsymbol{\Sigma}}_s = \hat{\boldsymbol{\Pi}}(0) - \hat{\mathbf{B}}\hat{\boldsymbol{\Pi}}(1), \tag{12}$$

where $\hat{\boldsymbol{\Sigma}}_s$ is the estimated forecasting error covariance and $\hat{\mathbf{B}}$ is the coefficient matrix estimated by solving (10) [36].

We highlight that the accuracy of the YW method improves with the number of available observations, M. Additionally, as the network size grows, the number of parameters in the VAR model increases, necessitating a larger dataset to accurately train the model [34]. Furthermore, the YW method assumes that power injection measurements from all nodes are available at every time instant. However, in reality, this assumption is often unrealistic due to the limited availability of sensors and constraints on communication bandwidth. To address this, we use a recursive least-squares algorithm to estimate the parameters of the VAR(1) model.

2) Recursive Least Squares (RLS): Consider the VAR(1) model described in (9). Rather than utilizing all historical data simultaneously, we can iteratively update our estimate of the matrix $\mathbf{B}$ at each time step, as power injection measurements are collected, using the following iterative update rule:

$$\hat{\mathbf{B}}_t = \hat{\mathbf{B}}_{t-1} + \big(\mathbf{S}_t - \hat{\mathbf{B}}_{t-1}\mathbf{S}_{t-1}\big)\mathbf{K}_t^\top, \tag{13}$$

where $\hat{\mathbf{B}}_t$ is our estimate of the matrix $\mathbf{B}$ at time t. The optimal gain matrix $\mathbf{K}_t \in \mathbb{R}^{6N}$ is found by solving the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{K}_t} \sum_{i=1}^{k} \lambda^{t-i} \|\mathbf{S}_t - \hat{\mathbf{B}}_{t-1}\mathbf{S}_{t-1}\|^2 \iff \\ &\min_{\mathbf{K}_t} \ \mathtt{Tr}\big((\bar{\mathbf{S}}_t - \hat{\mathbf{B}}_{t-1}\bar{\mathbf{S}}_{t-1})\boldsymbol{\Lambda}(\bar{\mathbf{S}}_t - \hat{\mathbf{B}}_{t-1}\bar{\mathbf{S}}_{t-1})^\top\big), \end{aligned} \tag{14}$$

where the scalar $\lambda \in (0, 1]$ is a forgetting factor, $\boldsymbol{\Lambda} = \mathtt{diag}(\lambda^{t-1}, \ldots, \lambda^{t-i+1}, 1)$, $\bar{\mathbf{S}}_t = [\mathbf{S}_2, \ldots, \mathbf{S}_t]$, and $\bar{\mathbf{S}}_{t-1} = [\mathbf{S}_1, \ldots, \mathbf{S}_{t-1}]$. The optimization problem (14) has a closed form solution that is expressed as

$$\mathbf{K}_t = (\bar{\mathbf{S}}_t\boldsymbol{\Lambda}\bar{\mathbf{S}}_{t-1}^\top)(\bar{\mathbf{S}}_{t-1}\boldsymbol{\Lambda}\bar{\mathbf{S}}_{t-1}^\top)^{-1}. \tag{15}$$

Using the Sherman-Morrison matrix inversion formula, we can write (15) using the following recursion:

$$\mathbf{K}_t = \mathbf{C}_{t-1}\mathbf{S}_k(\lambda + \mathbf{S}_t^\top\mathbf{C}_{t-1}\mathbf{S}_t)^{-1}, \tag{16a}$$

$$\mathbf{C}_t = \frac{1}{\lambda}(\mathbf{I}_{6N} - \mathbf{K}_t\mathbf{S}_t^\top)\mathbf{C}_{t-1}, \tag{16b}$$

where $\mathbf{C} = (\bar{\mathbf{S}}_{t-1}\boldsymbol{\Lambda}\bar{\mathbf{S}}_{t-1}^\top)^{-1}$ is the covariance of the estimated matrix $\hat{\mathbf{B}}$. The covariance $\boldsymbol{\Sigma}_s$ of the prediction error $\boldsymbol{\varepsilon}$ is estimated as

$$\hat{\boldsymbol{\Sigma}}_s = \frac{1}{t-1}\sum_{i=1}^{t}(\tilde{\mathbf{S}}_i - \hat{\mathbf{B}}_t\tilde{\mathbf{S}}_{i-1})(\tilde{\mathbf{S}}_i - \hat{\mathbf{B}}_t\tilde{\mathbf{S}}_{i-1})^\top, \tag{17}$$

where $\hat{\boldsymbol{\Sigma}}_s$ is the estimated covariance of the prediction error $\boldsymbol{\varepsilon}_t$. Therefore, $\mathbf{S}_t \sim \mathfrak{N}(\boldsymbol{\mu}_s, \hat{\boldsymbol{\Sigma}}_s)$, where $\boldsymbol{\mu}_s = \mathbf{S}_{t-1}$. Selecting an appropriate forgetting factor λ can yield convergence of the RLS algorithm with significantly fewer training data samples. Moreover, RLS provides computational advantages over the YW method. The YW method requires solving the system of equations in (10), which can be computationally prohibitive for large systems, and requires storing extensive historical data for accurate training. In contrast, RLS updates parameters iteratively, eliminating the need for large-scale data storage and reducing computational costs.

3) RLS with Partial Measurements: As formulated above, the RLS method assumes that power injection measurements from all the nodes are available at each time step to perform the iterative update. To address this, we next describe an approach that only accesses measurements from subset of smart meters at each time step. Let $\tilde{\mathbf{S}}_t$ represent the training data sample used to perform the iterative update of the estimated matrix $\hat{\mathbf{B}}$, described by (13) and (16). To perform this iterative update, we use the power injection measurements at time t for

the observed nodes and the previously measured values for the unobserved nodes, i.e.,

$$\tilde{S}_t^m = S_t^m, \qquad \forall m \in \mathcal{M}_t, \tag{18}$$

$$\tilde{S}_t^m = \tilde{S}_{t-1}^m, \qquad \forall m \in \mathcal{N} \backslash \mathcal{M}_t, \tag{19}$$

where $\mathcal{M}_t$ is the set of observed nodes at time t and $\tilde{S}_t^m$ and S_t^m represent the entries of the vectors $\tilde{\mathbf{S}}_t$ and $\mathbf{S}_t$ corresponding to node m. Here, we assume that the nodal power injections are in a quasi-steady-state, meaning they change slowly over time. While this assumption may not strictly hold in the presence of DERs, we will empirically demonstrate that the method remains effective under realistic conditions. The subset of observed nodes changes at each time step, ensuring full network coverage in a finite number of steps. The proposed method does not depend on how the set of observed nodes is chosen at every time step and optimally selecting the nodes to be observed for obtaining the measurements is part of our future work. However, it is important to ensure that the measurements are obtained from all the nodes in the network in the shortest time possible.

C. Forecasting the System States

Using (6) and (7), the system states $\mathbf{y} = [\mathbf{v}^\top, \boldsymbol{\theta}^\top, \mathbf{P}^\top, \mathbf{Q}^\top]^\top \in \mathbb{R}^{12N}$, consisting of the squared voltage magnitudes, voltage angles, and power injections, are expressed as

$$\mathbf{y} = \mathbf{y}_1 + \mathbf{M}\mathbf{S}, \tag{20}$$

where the matrix $\mathbf{M} = \begin{bmatrix}\mathbf{M}_v^\top & \mathbf{M}_\delta^\top & \mathbf{I}_{6N}\end{bmatrix}^\top \in \mathbb{R}^{12N \times 6N}$ and $\mathbf{y}_1 = \begin{bmatrix}\mathbf{v}_1^\top & \boldsymbol{\delta}_1^\top & \mathbb{0}_{6N}^\top\end{bmatrix}^\top \in \mathbb{R}^{12N}$. Using (20), we express the observed states $\mathbf{y}_o$ as

$$\mathbf{y}_o = \mathbf{Z}_o\mathbf{y} = \mathbf{y}_{1o} + \mathbf{M}_o\mathbf{S}, \tag{21}$$

where $\mathbf{Z}_o$ is a permutation matrix composed of the rows of the identity matrix $\mathbf{I}_{12N}$ corresponding to the observed states such that $\mathbf{y}_{1o} = \mathbf{Z}_o\mathbf{y}_1$ and $\mathbf{M}_o = \mathbf{Z}_o\mathbf{M}$. Similarly, we express the unobserved states $\mathbf{y}_u$ as

$$\mathbf{y}_u = \mathbf{Z}_u\mathbf{y} = \mathbf{y}_{1u} + \mathbf{M}_u\mathbf{S}, \tag{22}$$

where $\mathbf{Z}_u$ is a permutation matrix composed of the rows of the identity matrix $\mathbf{I}_{12N}$ corresponding to the unobserved states such that $\mathbf{y}_{1u} = \mathbf{Z}_u\mathbf{y}_u$ and $\mathbf{M}_u = \mathbf{Z}_u\mathbf{M}$. Since we do not actually measure the unobserved states, we do not model any measurement noise in the unobserved states.

Combining (21), (22), and the forecasted power injections $\hat{\mathbf{S}}$ from (9), we obtain forecasts of the state variables:

$$\hat{\mathbf{y}}_o = \mathbf{y}_{1o} + \mathbf{M}_o\hat{\mathbf{S}}, \tag{23a}$$

$$\hat{\mathbf{y}}_u = \mathbf{y}_{1u} + \mathbf{M}_u\hat{\mathbf{S}}, \tag{23b}$$

where $\hat{\mathbf{S}} \sim \mathfrak{N}(\boldsymbol{\mu}_s, \hat{\boldsymbol{\Sigma}}_s)$ are the forecasts of the nodal power injections and $\hat{\mathbf{y}}_o$ and $\hat{\mathbf{y}}_u$ are the forecasts of the state at the observed and the unobserved nodes, respectively. Since we assume that the power injection forecasts $\hat{\mathbf{S}}$ are normally distributed and, as described in (23), the states and the power injections are linearly related, the forecasted states are also normally distributed, i.e., $\hat{\mathbf{y}}_o \sim \mathfrak{N}(\boldsymbol{\mu}_o, \hat{\boldsymbol{\Sigma}}_o)$ and $\hat{\mathbf{y}}_u \sim \mathfrak{N}(\boldsymbol{\mu}_u, \hat{\boldsymbol{\Sigma}}_u)$. Their respective means are

$$\boldsymbol{\mu}_o = \mathbb{E}(\hat{\mathbf{y}}_o) = \mathbf{M}_o\boldsymbol{\mu}_s, \tag{24a}$$

$$\boldsymbol{\mu}_u = \mathbb{E}(\hat{\mathbf{y}}_u) = \mathbf{M}_u\boldsymbol{\mu}_s. \tag{24b}$$

The covariances of $\hat{\mathbf{y}}_o$ and $\hat{\mathbf{y}}_u$, $\hat{\boldsymbol{\Sigma}}_o$ and $\hat{\boldsymbol{\Sigma}}_u$, and the cross-covariance, $\hat{\boldsymbol{\Sigma}}_{uo}$, between $\hat{\mathbf{y}}_o$ and $\hat{\mathbf{y}}_u$ are computed as

$$\hat{\boldsymbol{\Sigma}}_o = \mathbb{E}\big((\hat{\mathbf{y}}_o - \boldsymbol{\mu}_o)(\hat{\mathbf{y}}_o - \boldsymbol{\mu}_o)^\top\big) = \mathbf{M}_o\hat{\boldsymbol{\Sigma}}_s\mathbf{M}_o^\top, \tag{25a}$$

$$\hat{\boldsymbol{\Sigma}}_u = \mathbb{E}\big((\hat{\mathbf{y}}_u - \boldsymbol{\mu}_u)(\hat{\mathbf{y}}_u - \boldsymbol{\mu}_u)^\top\big) = \mathbf{M}_u\hat{\boldsymbol{\Sigma}}_s\mathbf{M}_u^\top, \tag{25b}$$

$$\hat{\boldsymbol{\Sigma}}_{uo} = \mathbb{E}\big((\hat{\mathbf{y}}_u - \boldsymbol{\mu}_u)(\hat{\mathbf{y}}_o - \boldsymbol{\mu}_o)^\top\big) = \mathbf{M}_u\hat{\boldsymbol{\Sigma}}_s\mathbf{M}_o^\top. \tag{25c}$$

The estimated means and the covariances of the forecasts for the observed and unobserved states are used in conjunction with actual measurements to estimate the network states. We model the measurement noise in the observed states as additive Gaussian noise, i.e.,

$$\mathbf{y}_o^m = \mathbf{y}_o + \boldsymbol{\eta}, \tag{26}$$

where $\boldsymbol{\eta} \sim \mathfrak{N}(0, \boldsymbol{\Omega}_o)$ is the measurement noise with a covariance of $\boldsymbol{\Omega}_o \in \mathbb{R}^{M \times M}$ and $\mathbf{y}_o^m \sim \mathfrak{N}(\mathbf{y}_o, \boldsymbol{\Omega}_o)$ is the measured value of the observed states $\mathbf{y}_o$. Using (26) and forecasts of the observed states, $\hat{\mathbf{y}}_o$, we forecast the measurements

$$\hat{\mathbf{y}}_o^m = \hat{\mathbf{y}}_o + \boldsymbol{\eta}, \tag{27}$$

where $\hat{\mathbf{y}}_o^m \sim \mathfrak{N}(\boldsymbol{\mu}_o, \hat{\boldsymbol{\Sigma}}_o + \boldsymbol{\Omega}_o)$ is the forecast of the measurements of the observed states. We assume that the measurement noise $\boldsymbol{\eta}$ and the forecast error $\boldsymbol{\varepsilon}$ are uncorrelated, i.e., $\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\eta}^\top) = \mathbb{E}(\boldsymbol{\eta}\boldsymbol{\varepsilon}^\top) = 0$. However, in practice, this assumption may not hold as the VAR model used for forecasting is trained using power injection measurements, meaning that the accuracy of these measurements directly influences the accuracy of the forecasts. Consequently, any errors in the measurements could propagate into the forecasted states, leading to potential correlation between $\boldsymbol{\eta}$ and $\boldsymbol{\varepsilon}$. However, our numerical assessment shows that, despite the potential correlation in prediction errors and measurements noise, the proposed method performs well over the time periods considered. Furthermore, this motivates the periodic retraining of the VAR model to avoid the accumulation of errors caused due to this assumption. Since the measurement noise and the prediction errors are assumed to be uncorrelated, the covariance between the forecasts of the unobserved states $\hat{\mathbf{y}}_u$ and the measurements $\hat{\mathbf{y}}_o^m$ is also given by (25c).

V. State Estimation

Building upon the models described in the prior two sections, we next describe the state estimator based on a linear least-squares approach. We use the forecasts of nodal power injections to forecast the system states using the Lin3DistFlow approximation. Then, treating these state forecasts as priors, we incorporate limited smart meter measurements to refine our state estimate by solving a least-squares optimization problem. With the means $\boldsymbol{\mu}$ and covariances $\hat{\boldsymbol{\Sigma}}$ of the state forecasts

$\hat{\mathbf{y}}$ estimated in Section IV-C, we construct a linear estimator for the unobserved states $\mathbf{y}_u$ of the form $\tilde{\mathbf{y}}_u = \mathbf{L}\mathbf{y}_o^m + \boldsymbol{\rho}$, where $\boldsymbol{\rho}$ is a normally distributed random variable representing the estimation error, by minimizing the squared norm of the estimation residual:

$$\min \mathbb{E}\big(\|\tilde{\mathbf{y}}_u - \mathbf{y}_u\|^2\big), \tag{28}$$

where $\tilde{\mathbf{y}}_u$ is the vector of estimated states. The optimization problem described in (28) has a closed-form solution [37]:

$$\begin{aligned}\tilde{\mathbf{y}}_u &= \mathbb{E}(\mathbf{y}_u|\mathbf{y}_o^m),\\ &= \boldsymbol{\mu}_u + \hat{\boldsymbol{\Sigma}}_{uo}\big(\hat{\boldsymbol{\Sigma}}_o + \boldsymbol{\Omega}_o\big)^{-1}(\mathbf{y}_o^m - \boldsymbol{\mu}_o).\end{aligned} \tag{29}$$

The proposed estimator uses the statistical information of states computed in Section IV-C, which only depend on the forecasted power injections, the network topology, and the line parameters. The estimator given by (29) is linear with respect to the measurements and, thus, is computationally efficient and easy to implement. Moreover, this estimator does not require full network observability as the lack of measurements are compensated for by the spatial correlation in power injections estimated by the VAR(1) model. For accurate state estimation, it is important to choose the observed nodes such that $(\hat{\boldsymbol{\Sigma}}_o + \boldsymbol{\Omega}_o)$ is invertible, which will happen as long as the observed nodes are uniformly spread across the system.

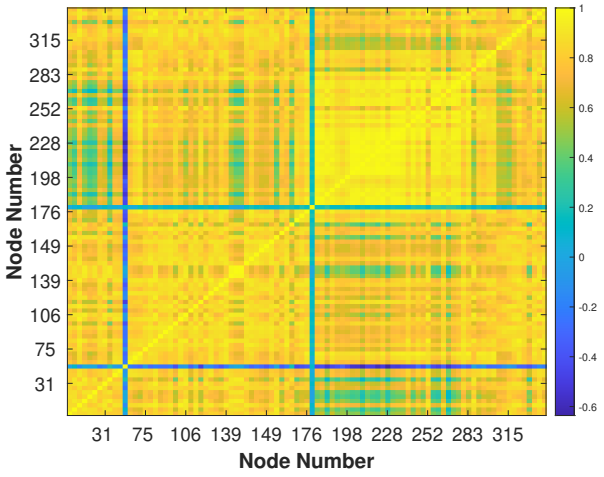
While our simulations will show that out proposed method performs well under many different scenarios, a possible drawback of this method as currently implemented is that it might not be robust to outliers, as we have not accounted for other sources of measurement error. Existing machinery for detecting and mitigating outliers (see, e.g., [38]) could be applied in this context, which is a direction for future work.

VI. Numerical Simulations

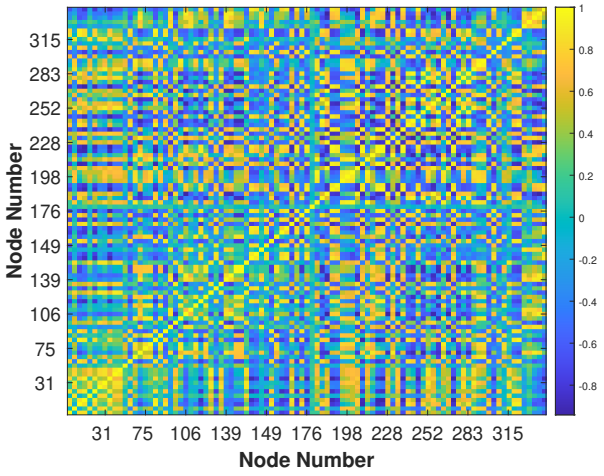
In this section, we describe the numerical simulations performed under various scenarios to validate the proposed method. First, we describe the test feeder utilized and the modifications applied to it. Next, we provide details on the load and renewable generation data used, along with the different simulation scenarios. Finally, we present a detailed analysis of the the results.

A. Setup

1) Test System: The performance of the proposed DSSE method is tested on the IEEE 123-bus distribution test feeder [39], using the network data available at [40]. The feeder includes capacitors banks, voltage regulators, and switches for topology changes. To simplify the modeling requirements, we neglect the capacitor banks, replace all the voltage regulators by short transmission lines of equivalent impedance, and remove all the switches in the network. The nodes connected to the ends of normally closed switches are merged to maintain connectivity, while normally open switches are left open. Extensions of the proposed approach to consider more detailed network models are possible without substantive changes to our proposed methodology.



(a) Base case.



(b) Low-correlation scenario.

Fig. 2: Correlation in the active power injections in (a) the base case and (b) in the low-correlation scenario.

2) Load and Renewable Data: We use load profiles based on actual measurements from a substation in Oak Park, Portland, OR [41]. These measurements include a full day of load data sampled at 2 kHz, which is subsequently down-sampled to a resolution of 1 minute. For data on renewable generation, we used actual active power measurements from a rooftop PV system from the EPFL Distributed Electrical Systems Laboratory [42], with data available at [43]. In our simulations, we assume that PV plants with a capacity of 50% of the peak load are installed at every node with non-zero power injection. Using the load and renewable generation data, we obtain the true values of the network states by solving the non-linear AC power flow (ACPF), which are used to evaluate the accuracy of our proposed scheme and to obtain noisy measurements as described in the following sections.

3) Measurements: We model the measurement noise as unbiased additive Gaussian noise according to the standard

defined by the Instrument Transformer (IT) class [44], [45]. For simulations, we consider IT 1.0 which translates to the maximum measurement noise of 1% relative to the true values. Additionally, we assume that only a small set of measurements, as low as 10%, are accessible in real-time at each time step. The measurement set can be updated at each time step to capture measurements from all the nodes in the network. To select the set of observed nodes, we begin by constructing a random permutation $\mathcal{N}_p$ by shuffling the set of nodes $\mathcal{N}$. The set $\mathcal{N}_p$ is then partitioned into subsets of equal size. At each time step, one subset is selected as the measurement set.

4) Scenarios: To validate the performance of the proposed method under different operating conditions, we consider the following scenarios:

a) *Base case (BC):* Each node in the feeder is randomly assigned one of the six available load profiles, which are described in Section VI-A2. The load profiles are first normalized to have a peak value of one. The load data for each node are then generated by multiplying the node's assigned load profile by the nominal load value specified in the network data. We note that this results in the load data having a high degree of spatial correlation, as shown in Fig. 2a, as may occur if the distribution feeder is spread over a small geographic area.

b) *Higher renewable penetration (PV):* To verify the effectiveness of the proposed method in networks with high renewable penetration, we simulate a scenario where each node is equipped with a PV plant with a capacity of up to 50% of its nominal load. In this scenario, we use the same load data as the base case scenario.

c) *Low correlation (LC):* Finally, we simulate a scenario with reduced correlation in the load data, as illustrated in Fig. 2b, to assess whether the results from the base case are overly influenced by the high correlation in the base case scenario. Since we do not have access to actual load data from many different customers, we generate a synthetic dataset with low correlation by shuffling each load profile by a random amount in time. This captures the underlying patterns in the typical customer behaviors inherent to the measured load data from [41] while modeling customers with offset daily schedules.

For each of the scenarios, we use the root mean square error (RMSE) and the mean absolute error (MAE) to evaluate the performance of the proposed method:

$$RMSE = \sqrt{\frac{1}{N_s}\sum_{i=1}^{N_s} \|\tilde{\mathbf{v}}_t - \mathbf{v}_t\|_2^2}, \tag{30}$$

$$MAE = \frac{1}{N_s}\sum_{t=1}^{N_s} \|\tilde{\mathbf{v}}_t - \mathbf{v}_t\|_1, \tag{31}$$

where N_s refers to the number of samples and $\tilde{\mathbf{v}}_t$ and $\mathbf{v}_t$ are vectors of the estimated and true voltage magnitudes.

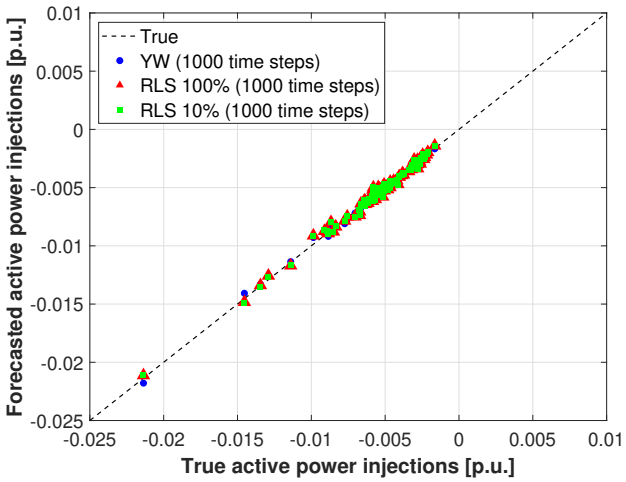
TABLE I: Forecasting accuracy with varying training dataset sizes and measurement sets

# Training samples	Method	MAE [p.u.]	RMSE [p.u.]
1000	YW	8.9120×10^{-5}	1.3291×10^{-4}
	RLS 100%	8.2374×10^{-5}	1.2946×10^{-4}
	RLS 10%	9.6501×10^{-5}	1.4401×10^{-4}
200	YW	1.3965×10^{-3}	2.1458×10^{-3}
	RLS 100%	6.7451×10^{-4}	4.4499×10^{-4}
	RLS 10%	7.0130×10^{-4}	4.4706×10^{-4}

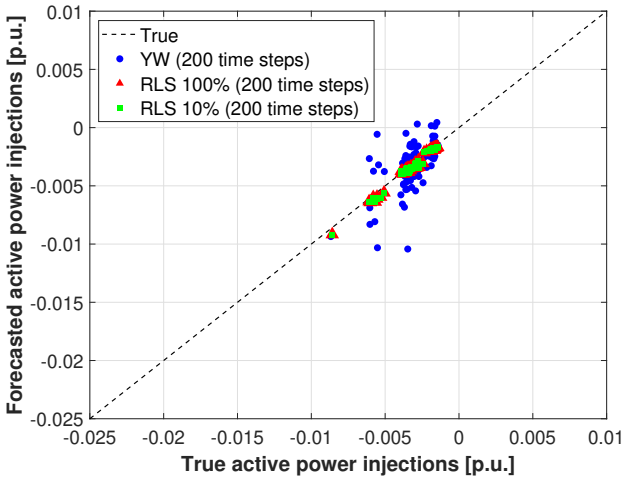
B. Forecasting Performance

In this section, we evaluate the performance of the VAR(1) model, which, as described by (9), predicts the power injections $\mathbf{S}_t$ at the current time step using power injection measurements $\mathbf{S}_{t-1}$ from the previous time step. The VAR(1) model is trained using the Yule-Walker (YW) and the Recursive Least Squares (RLS) algorithms explained in Section IV-B, with the YW algorithm providing a baseline. To compare the effectiveness of the YW and RLS algorithms under varying amounts of available data, we consider two training datasets that include either 1000 or 200 time steps. For the YW algorithm, the training dataset is gathered over time and then used simultaneously to estimate the parameters of the VAR(1) model. Conversely, for the RLS algorithm, the dataset size represents the number of time steps required for iteratively training the VAR(1) model. At each time step, the RLS algorithm updates the parameters of the VAR(1) model using measurements only from the current time step. In addition to varying the size of the training dataset in terms of the number of time steps, we also assess how varying the percentage of sampled nodes impacts the RLS algorithm's performance. To obtain a baseline, we first train the VAR(1) model using the YW algorithm with power injection measurements from all nodes. We then train the VAR(1) model with the RLS algorithm assuming that differing subsets of nodes can be observed at every time step. Specifically, we consider two cases where the operator can access measurements from 100% and 10% of nodes in the network. At every time step, the set of observed nodes is varied according to the procedure described in Section VI-A3. We evaluate the performance of both the methods via predicting the power injections at the current time step $\mathbf{S}_t$ by assuming that we have perfect knowledge of the power injections at the previous time step $\mathbf{S}_{t-1}$.

Fig. 3 and Table I summarize the forecasting performance of each method with different dataset sizes and percentages of sampled measurements. We first analyze the scenario where data from 1000 time steps are used to train the VAR(1) model. Here, we observe that all the considered methods have comparable prediction accuracy, as shown by Fig. 3a. Importantly, the RLS method accurately trains the VAR model even when only 10% of the nodal power injections are accessible at each time step due to bandwidth constraints. The forecasting accuracy of both methods deteriorates when we have access to less training data. We can observe this in Figs. 3b when compared to the plots in Figs. 3a. In the case of



(a) 1000 time steps



(b) 200 time steps

Fig. 3: Forecast accuracy of the VAR(1) model when trained using (a) 1000 time steps and (b) 200 time steps. In both plots, the dashed black line represents the true nodal power injections. Both methods perform comparably when the VAR model is trained using 1000 time steps of data. However, with 200 samples, the YW method fails to accurately predict the power injections, while the RLS method is more robust.

the RLS method with access to 200 samples of data, the RMSE increases by almost a factor three and the MAE increases by almost an order of magnitude, as shown in Table I. Despite the degraded performance, the RLS method is, in general, still capable of capturing the trends in power injections with an acceptable accuracy. Additionally, the performance of the RLS method is insensitive to the percentage of measurements accessed in each training iteration. The performance with only 10% of the measurements is only slightly worse than when we have access to all the measurements. On the other hand, with a training dataset of 200 samples, the YW method fails to accurately estimate the parameters of the VAR model, resulting in poor forecasting performance, as shown by Fig. 3b. The

TABLE II: Voltage estimation error in the base case

% Measurements	RMSE [p.u.]	MAE [p.u.]
Fixed Measurement set		
10	5.4991×10^{-4}	4.2027×10^{-4}
20	4.9729×10^{-4}	3.8687×10^{-4}
Dynamic Measurement set		
10	5.6587×10^{-4}	4.3026×10^{-4}
20	5.2314×10^{-4}	4.0294×10^{-4}

TABLE III: Voltage estimation error with PV integration

% Measurements	RMSE [p.u.]	MAE [p.u.]
Fixed Measurement set		
10	5.1374×10^{-4}	3.8790×10^{-4}
20	4.6347×10^{-4}	3.5473×10^{-4}
Dynamic Measurement set		
10	5.2448×10^{-4}	3.9446×10^{-4}
20	4.8866×10^{-4}	3.7037×10^{-4}

TABLE IV: Voltage estimation error under the low correlation

% Measurements	RMSE [p.u.]	MAE [p.u.]
Fixed Measurement set		
10	5.7959×10^{-4}	4.4772×10^{-4}
20	5.3669×10^{-4}	4.3154×10^{-4}
Dynamic Measurement set		
10	6.3749×10^{-4}	4.8921×10^{-4}
20	5.4284×10^{-4}	4.2703×10^{-4}

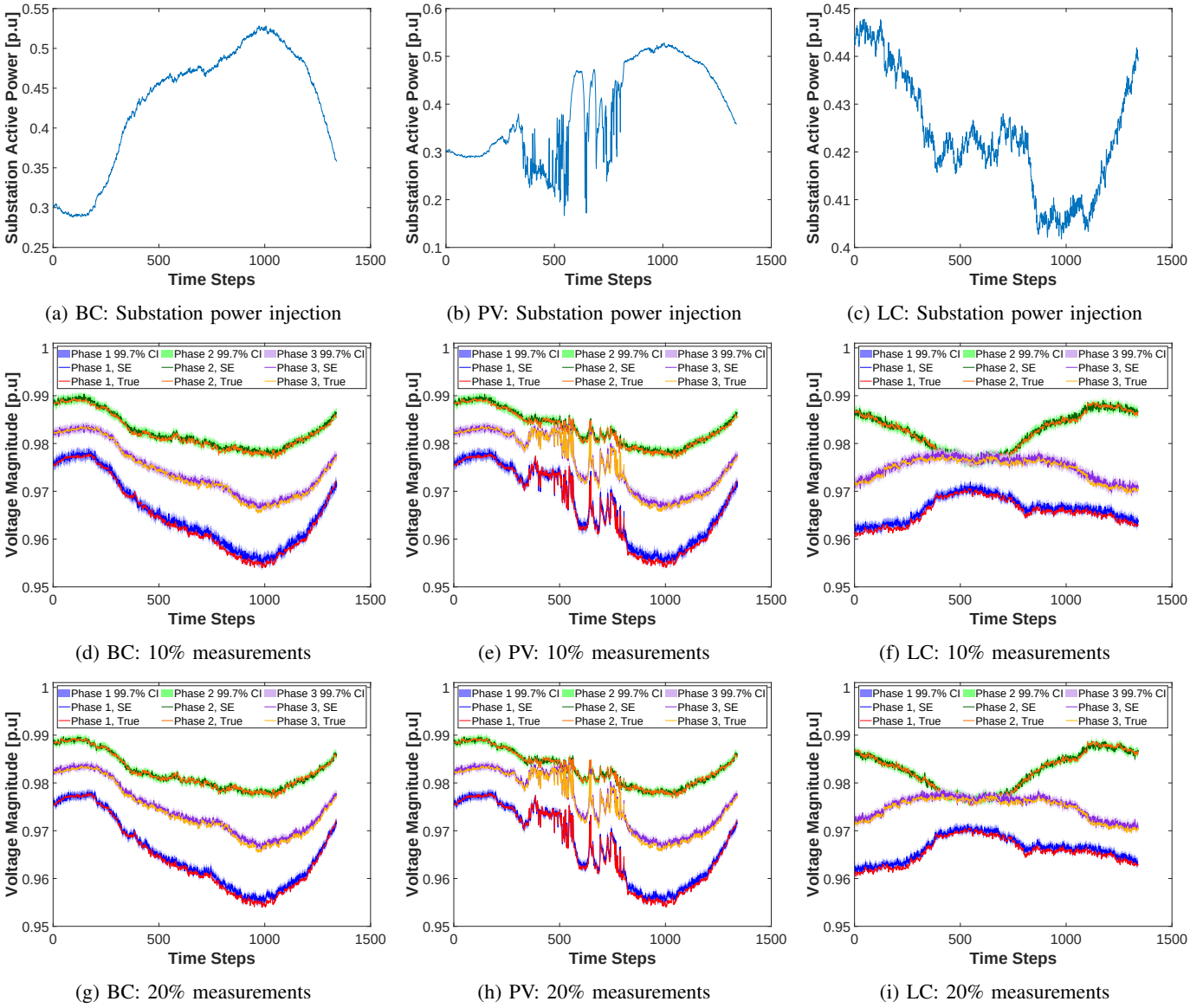
YW method involves solving (10), where $\mathbf{\Pi}_0$ is computed as a sum of M rank-1 matrices, with M denoting the number of training samples. Therefore, the YW method is only feasible with a sufficiently large M so that $\mathbf{\Pi}_0$ is not ill-conditioned.

From Fig. 3, it is evident that the forecasting accuracy of the RLS method is more sensitive to the size of the training dataset than to the number of sampled measurements. Thus, during the offline training phase, it is more important to train the VAR model for a longer duration, resulting in a more accurate estimate of its parameters $\mathbf{B}$ and $\mathbf{\Sigma}_s$.

C. State Estimation Performance

Next, we compare the performance of the proposed state estimation method under the scenarios described in Section VI-A4, with the results summarized in Fig. 4 and Tables II–III. In all three scenarios, we train the VAR(1) model using the RLS method with 200 data samples. As shown in Section VI-B, despite the lower accuracy, the VAR(1) model trained with 200 data samples still captures the general trends in the power injections. As we will show in this section, the state estimation step is able to account for this increased error to recover accurate state estimates. While the VAR(1) model can be updated regularly to adapt to changes in load and renewable generation, for simplicity, we train the VAR model at the beginning of the simulation and keep it fixed thereafter.

To evaluate the performance of the proposed method, voltage magnitudes estimated using the proposed method are compared against the true nodal voltages obtained by solving an AC power flow using the backward-forward sweep algorithm. To assess the performance of the proposed method under



(a) BC: Substation power injection (b) PV: Substation power injection (c) LC: Substation power injection

(d) BC: 10% measurements (e) PV: 10% measurements (f) LC: 10% measurements

(g) BC: 20% measurements (h) PV: 20% measurements (i) LC: 20% measurements

Fig. 4: Performance of the proposed state estimation method in different scenarios. Figs. 4a–4c show the substation active power injections under different scenarios. Figs. 4d–4i show the performance of the proposed method with 10% (top) and 20% (bottom) measurements, in the base case (left, BC), with PV integration (middle, PV), and in the low-correlation scenario (right, LC). As an illustrative example, we plot the voltage at one particular three-phase node but emphasize that the performance is consistent across all the nodes in the network as shown by Tables II–IV.

different communication infrastructures, we simulate both fixed and dynamically varying measurement sets, with sizes corresponding to 10% and 20% of all nodes in the network. The fixed measurement set represents a case where only a few nodes in the network are equipped with smart meters, while the dynamically varying measurement set represents a case where every node in the network is equipped with a smart meter but measurements can only be accessed from a few nodes each time step due to limited bandwidth. In the case of a fixed measurement set, we randomly choose one of the subsets constructed using the procedure described in Section VI-A3 at the beginning of the simulation. The measurement set is kept fixed during the entire simulation. In the case of a dynamically varying measurement set, a different subset of measurements is selected at each time step such that all nodes in the network are sampled within a finite number of time steps. State estimation is initialized using the final sample of the training data. The estimated power injections from the previous time step are used as inputs to the VAR model for predicting the power injections at the current time step.

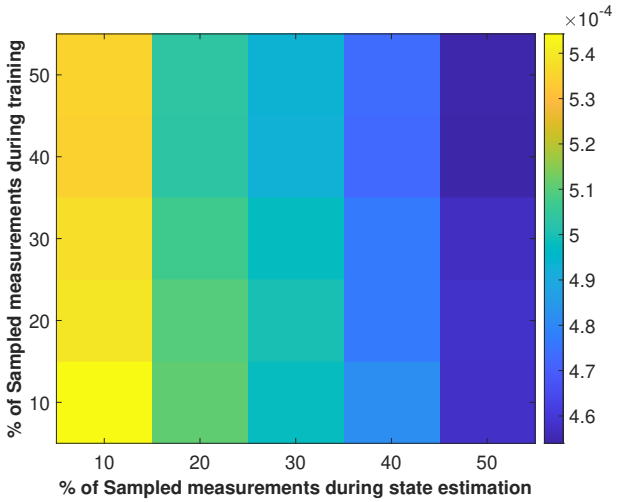
We first analyze the performance of the proposed method in the base case scenario. In this scenario, we assume there is no renewable generation in the network, ensuring that there are no sudden changes in the power injections, as shown in Fig. 4a.

We also observe that the loads are highly correlated. Under these conditions, the proposed method is expected to work well with few measured nodes. Thus, the base case scenario serves as a baseline, allowing us to compare the performance of our proposed method under other scenarios. Figs. 4d and 4g show the true and the estimated voltages of all phases of a particular node, computed using a fixed measurement set with 20% and 10% of all nodes, respectively, while the estimation errors with different measurement configurations are tabulated in Table II. Here, we can observe that our proposed method performs consistently across different measurement configurations. Notably, even with only 10% of the nodes sampled, the state estimation error remains low.

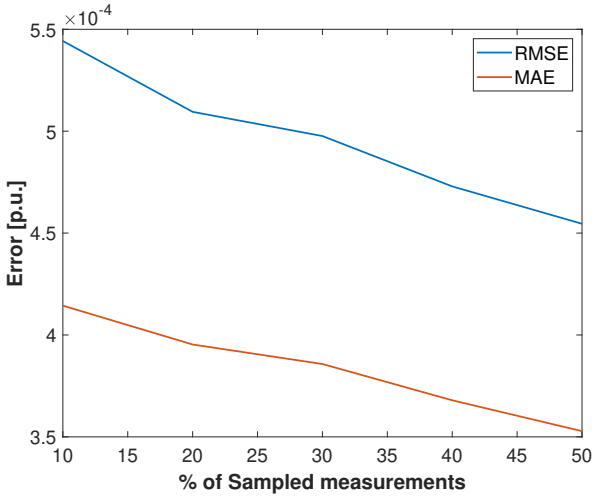
To assess whether the observed performance in the base case is related to the lack of renewable generation, we next analyze our proposed method in a scenario with high renewable penetration, the results of which are summarized in Figs. 4e and 4h and in Table III. The selected PV profiles exhibit significant fluctuations in active power injections, as shown by the substation active power injection in Fig. 4b. These sudden variations represent the uncertainty in PV generation due to factors such as cloud cover and partial shading. Despite the rapid changes in voltage magnitudes induced by the fluctuations in power injections, the proposed method accurately tracks the voltage, as shown by Figs. 4e and 4h. This indicates that method performs well even when our assumption of quasi-steady-state power injections (see Section IV-B3) does not hold in the presence of significant quantities of renewable generation. Additionally, from Tables II and III, we observe that our proposed method performs similarly in both the high renewable generation and the base case scenarios despite the sudden changes in active power injections.

Finally, to assess the impacts of load demand correlations, we simulate the case with low correlation and compare to the base case. Figs. 4f–4i and Table IV illustrate that the method remains effective for low-correlation scenarios. The proposed method performs comparably to the base case across the different measurement configurations, with some slight deterioration since the lack of measurements is partially compensated by the estimated correlation in nodal power injections. In the extreme case when the power injections are uncorrelated, the forecasted power injections do not provide additional information, necessitating a larger measurement set.

The results presented above illustrate that the proposed method performs consistently across all scenarios. However, we can observe from Tables II–IV that, for all three scenarios, the performance of the proposed method with a dynamic measurement set is slightly worse than when compared to fixed measurement set due to the random sampling of observed nodes at each time step. Since the network consists of both three- and single-phase nodes, we note that random sampling does not ensure a consistent number of measurements at each time step. Thus, choosing a measurement set more strategically is likely to improve our method's performance. This is a topic for our future work.



(a) Variation in the RMSE of estimated voltage magnitude as we vary the percentage of nodes sampled during VAR training (vertical axis) and during state estimation (horizontal axis).



(b) RMSE and MAE of the estimated voltage magnitude with an increasing percentage of sampled measurements for both VAR model training and state estimation.

Fig. 5: Impact of the number of measured nodes on state estimation performance.

D. Impact of the Number of Measurements

Finally, we provide a more detailed assessment on how the number of sampled measurements impacts the accuracy of the state estimate. To accomplish this, we vary the number of measurements sampled during VAR parameter estimation and state estimation. For each combination of sampled measurements across these two stages, we simulate the base case and the low-correlation scenarios to assess the sensitivity of the proposed method to the number of measurements under different conditions. To ensure that the comparison is fair across all cases, we fix the forgetting factor λ at 1. We use 200 data samples to train the VAR model and use a fixed measurement set for the state estimation.

As expected, the performance of the proposed state esti-

mation framework improves when we increase the number of measurements sampled at different stages, as shown in Figs. 5a–5b. From Fig. 5a, we observe that the performance of the state estimation is more sensitive to the number of measurements sampled at the state estimation step than to the number of measurements sampled during VAR training. While only the results for the base case scenario are shown, we observe similar trends for the low-correlation scenario. The results presented above suggest that careful selection of measurements during the state estimation stage is more critical.

VII. Conclusions

This paper has presented a forecast-aided state estimation scheme for distribution systems that addresses the problem of scarcity of measurements caused by limited communication bandwidth. This scheme consists of two stages. The first stage trains a forecasting model using power injection measurements from a subset of smart meters that are randomly sampled over a period of time. The second stage then uses the trained forecasting model along with the measurements from a few observable nodes for the state estimation. The proposed method improves on the recent literature by (i) introducing a recursive approach for the forecasting which does not require a large amount of historical measurements, (ii) developing a state estimation approach without any phasor measurements, and (iii) accurately estimating the system state with a limited set of dynamically sampled smart meter measurements.

We validated the proposed scheme on the IEEE 123-bus benchmark network. The results show that the proposed scheme can accurately estimate the system state with measurements from only 10% of nodes, with estimation error reducing as the number of observed nodes increases. We also analyzed the performance of the proposed scheme using three scenarios with different correlation properties in power injections. The proposed method performs consistently across scenarios with varying measurement configurations and thus remains robust under a range of scenarios.

Future work will improve upon the key assumptions used in this work as follows.

a) Our model assumes that every node has a smart meter (with few sampled simultaneously), which may not hold in distribution grids with limited smart meter deployment. However, we also demonstrated that our methods achieve accurate estimates with a constant set of measurements, representing scenarios with limited measurement units. We plan to further investigate the impacts that reduced smart meter coverage has on estimation accuracy.
b) We used a linearized power flow approximation, Lin3DistFlow, and compared the state estimation results against the true AC power flow solutions. In the future, we will extend the state estimation model to consider the AC three-phase non-linear power flow model.
c) We assumed that all the prediction errors and measurement noise are additive white Gaussian noise with zero mean, which may not reflect the real noise characteristics. Future work could relax this assumption by considering additional noise models.
d) We also assumed that the prediction errors and measurement noise are uncorrelated, which may not always be true. Our results show that, for the range of time periods we considered, the proposed method remains accurate despite possible correlations. To mitigate the potential issues that may arise if this assumption does not hold, the VAR model could be retrained periodically.

Furthermore, our future research will focus on developing an optimization framework for dynamically sampling smart meter data, with the aim of reducing the state estimation error while satisfying communication constraints. We will also develop algorithms for considering adjustable precision from smart meters and determining different bit rates for various smart meters to minimize the overall state estimation error, considering the impact of quantization error and communication bandwidth limits.

References

[1] Power Quality Application Guide, "Voltage disturbances," *Standard EN*, vol. 50160, 2004.

[2] CIGRE' Task Force C6.04.02, "Benchmark systems for network integration of renewable and distributed energy resources," CIGRE International Council on Large Electric Systems, Tech. Rep., July 2009.

[3] IEEE Std 1159-2009, "IEEE recommended practice for monitoring electric power quality," Tech. Rep., June 2009.

[4] J. Zhao, G. Zhang, M. La Scala, Z. Y. Dong, C. Chen, and J. Wang, "Short-term state forecasting-aided method for detection of smart grid general false data injection attacks," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1580–1590, 2015.

[5] FERC, "Assessment of Demand Response and Advanced Metering," December 2024. [Online]. Available: https://www.ferc.gov/power-sales-and-markets/demand-response/reports-demand-response-and-advanced-metering

[6] U.S. Dept. Energy, Washington, DC, USA, "Voice of experience: Leveraging AMI networks and data," March 2019. [Online]. Available: https://www.energy.gov/sites/default/files/2024-02/01-03-2019_doe-voe-leveraging-ami-networks-and-data-report_508_0.pdf

[7] ——, "Advanced metering infrastructure and customer systems," September 2016. [Online]. Available: https://www.energy.gov/sites/prod/files/2016/12/f34/AMI%20Summary%20Report_09-26-16.pdf

[8] N. Duan, C. Huang, C.-C. Sun, and L. Min, "Smart meters enabling voltage monitoring and control: The last-mile voltage stability issue," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 677–687, 2022.

[9] C. Huang, C.-C. Sun, N. Duan, Y. Jiang, C. Applegate, P. D. Barnes, and E. Stewart, "Smart meter pinging and reading through ami two-way communication networks to monitor grid edge devices and DERs," *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 4144–4153, 2022.

[10] 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.

[11] S. S. S. R. Depuru, L. Wang, V. Devabhaktuni, and N. Gudi, "Smart meters for power grid—Challenges, issues, advantages and status," in *IEEE/PES Power Systems Conference & Exposition*, 2011.

[12] S. S. Alam, B. Natarajan, and A. Pahwa, "Distribution grid state estimation from compressed measurements," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1631–1642, 2014.

[13] M. Majidi, M. Etezadi-Amoli, and H. Livani, "Distribution system state estimation using compressive sensing," *International Journal of Electrical Power & Energy Systems*, vol. 88, pp. 175–186, 2017.

[14] P. L. Donti, Y. Liu, A. J. Schmitt, A. Bernstein, R. Yang, and Y. Zhang, "Matrix completion for low-observability voltage estimation," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2520–2530, 2019.

[15] R. Madbhavi, B. Natarajan, and B. Srinivasan, "Enhanced tensor completion based approaches for state estimation in distribution systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 5938–5947, 2020.

[16] A. Mutanen, M. Ruska, S. Repo, and P. Jarventausta, "Customer classification and load profiling method for distribution systems," *IEEE Transactions on Power Delivery*, vol. 26, no. 3, pp. 1755–1763, 2011.

[17] 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.

[18] E. A. Blood, B. H. Krogh, and M. D. Ilic, "Electric power system static state estimation through Kalman filtering and load forecasting," in *IEEE Power and Energy Society General Meeting*, 2008.

[19] J. Zhao, G. Zhang, Z. Y. Dong, and M. La Scala, "Robust forecasting aided power system state estimation considering state correlations," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2658–2666, 2016.

[20] R. Dobbe, D. Arnold, S. Liu, D. Callaway, and C. Tomlin, "Real-time distribution grid state estimation with limited sensors and load forecasting," in *7th ACM/IEEE International Conference on Cyber-Physical Systems (ICCPS)*, 2016.

[21] R. Dobbe, W. van 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, 2019.

[22] R. Dutta, S. J. Geetha, S. Chakrabarti, and A. Sharma, "An l-1 regularized forecasting-aided state estimator for active distribution networks," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 191–201, 2021.

[23] L. Gan and S. H. Low, "Convex relaxations and linear approximation for optimal power flow in multiphase radial networks," in *18th Power Systems Computation Conference (PSCC)*, 2014.

[24] D. B. Arnold, M. Sankur, R. Dobbe, K. Brady, D. S. Callaway, and A. Von Meier, "Optimal dispatch of reactive power for voltage regulation and balancing in unbalanced distribution systems," in *IEEE Power and Energy Society General Meeting (PESGM)*, 2016.

[25] M. D. Sankur, R. Dobbe, A. v. Meier, E. M. Stewart, and D. B. Arnold, "Optimal voltage phasor regulation for switching actions in unbalanced distribution systems," in *IEEE Power & Energy Society General Meeting (PESGM)*, 2020.

[26] A. A. Mamun, M. Sohel, N. Mohammad, M. S. Haque Sunny, D. R. Dipta, and E. Hossain, "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," *IEEE Access*, vol. 8, pp. 134 911–134 939, 2020.

[27] J. Gorka and L. Roald, "Classification models for forecasting and real-time identification of solar curtailment in the California grid," in *15th ACM International Conference on Future and Sustainable Energy Systems (e-Energy)*, 2024, pp. 158–169.

[28] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting–A novel pooling deep RNN," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, 2018.

[29] D. Markovics and M. J. Mayer, "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction," *Renewable and Sustainable Energy Reviews*, vol. 161, p. 112364, 2022.

[30] M. He, V. Vittal, and J. Zhang, "A sparsified vector autoregressive model for short-term wind farm power forecasting," in *IEEE Power & Energy Society General Meeting*, 2015.

[31] D. C. Hill, D. McMillan, K. R. W. Bell, and D. Infield, "Application of auto-regressive models to U.K. wind speed data for power system impact studies," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 134–141, 2012.

[32] Y. Chakhchoukh, V. Vittal, and G. T. Heydt, "PMU based state estimation by integrating correlation," *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 617–626, 2014.

[33] S. Córdova, H. Rudnick, Á. Lorca, and V. Martnez, "An efficient forecasting-optimization scheme for the intraday unit commitment process under significant wind and solar power," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 4, pp. 1899–1909, 2018.

[34] J. Shi, Y. Liu, and N. Yu, "Spatio-temporal modeling of electric loads," in *North American Power Symposium (NAPS)*, 2017.

[35] M. Hassanzadeh, C. Y. Evrenosolu, and L. Mili, "A short-term nodal voltage phasor forecasting method using temporal and spatial correlation," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3881–3890, 2016.

[36] P. J. Brockwell and R. A. Davis, *Modeling and Forecasting with ARMA Processes*. Cham: Springer International Publishing, 2016.

[37] J. Walrand, *Tracking—A*. Cham: Springer International Publishing, 2021, pp. 163–192.

[38] A. Abur and A. Gómez-Expósito, *Power System State Estimation: Theory and Implementation*. Marcel Dekker, 2004.

[39] W. Kersting, "Radial distribution test feeders," in *IEEE Power Engineering Society Winter Meeting*, vol. 2, 2001, pp. 908–912.

[40] IEEE PES AMPS DSAS Test Feeder Working Group. [Online]. Available: https://cmte.ieee.org/pes-testfeeders/resources/

[41] Y. Gu, "Meteorological and customer composition factors in power load forecasting," Masters Thesis, University of Wisconsin–Madison, 2023.

[42] R. Gupta, F. Sossan, and M. Paolone, "Grid-aware distributed model predictive control of heterogeneous resources in a distribution network: Theory and experimental validation," *IEEE Transactions on Energy Conversion*, vol. 36, no. 2, pp. 1392–1402, 2021.

[43] EPFL's Distributed Electrical Systems Laboratory, "DESL-Photovoltaic-timeseries." [Online]. Available: https://github.com/DESL-EPFL/DESL-Photovoltaic-timeseries

[44] IEC, *Instrument transformers - Part 3: Additional requirements for inductive voltage transformers*, IEC Std 61869-3:2011, 2017.

[45] ——, *Instrument transformers - Part 2: Additional requirements for current transformers*, IEC Std 61869-2-:2012, 2012.