Design and Implementation of a Machine Learning State Estimation Model for Unobservable Microgrids

BYRON ALEJANDRO ACUÑA ACURIO, (Graduate Student Member, IEEE), DIANA ESTEFANÍA CHÉRREZ BARRAGÁN, (Graduate Student Member, IEEE), JUAN CAMILO LÓPEZ AMEZQUITA, (Member, IEEE), MARCOS J. RIDER, (Senior Member, IEEE), AND LUIZ CARLOS PEREIRA DA SILVA

Department of Systems and Energy, School of Electrical and Computer Engineering, University of Campinas (UNICAMP), Campinas 13083-852, Brazil

Corresponding author: Byron Alejandro Acuña Acuño (byronacunio@gmail.com)

This work was supported in part by the Brazilian Institution São Paulo Research Foundation Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP in Portuguese) under Grant 2020/03069-5, Grant 2019/01906-0, Project 2021/11380-5, Project 2015/21972-6, and Project 2016/08645-9; and in part by the Electricity Sector Research and Development Program, Microgrids for Efficient, Reliable and Greener Energy (MERGE), regulated by the National Electricity Agency (Agência Nacional de Energia Elétrica (ANEEL in Portuguese)), in partnership with Companhia Paulista de Força e Luz (CPFL in Portuguese) Energia (Local Distribution System Operator), under Grant PD-00063-3058/2019-PA3058.

ABSTRACT

An observable microgrid may become unobservable when sensors are at fault, sensor data is missing, or data has been tampered by malicious agents. In those cases, state estimation cannot be performed using traditional approaches without pseudo-measurements. To address the lack of observability, this article presents the design and implementation of a novel three-phase state estimation method for unobservable and unbalanced AC microgrids, using machine learning techniques, without pseudo-measurements, and under heteroscedastic (i.e., non-constant variance) noise. The proposed machine learning state estimation (MLSE) makes full use of multiple candidate models trained with a small number of power flow simulations via OpenDSS, through random levels of demand and renewable generation in every simulation, enhanced through a proposed Tikhonov regularization operator. To deal with the heteroscedastic nature of measurements, a recursive average model is proposed to accurately estimate the state variables. Results are obtained using real data from a microgrid located on the main campus of the State University of Campinas (UNICAMP), in Brazil. The method can be easily adapted to microgrids with different configurations, distributed energy resources, and measurements. It is shown that the proposed MLSE outperforms the traditional weighted least square (WLS) state estimator.

INDEX TERMS

Machine learning, microgrids, Moore-Penrose left pseudoinverse, tikhonov regularization operator, recursive average model.

I. INTRODUCTION

State estimation calculates the most-likely values of the state variables in an electrical grid, from real-time available measurements. This information is helpful for the identification of erroneous measurements, on-line parameter estimation, cyber-security assessment, and autonomous energy management systems (EMS) [1]. State estimation has been mostly applied in transmission systems due to their high level of digitization [2]. The direct translation of state estimation techniques from transmission systems to the distribution level is not always possible [3], so that the approaches used in distribution systems and AC microgrids are still challenging [4], [5], [6]. On the other hand, incorporating uncertainty into...
state estimation models has also become a relevant research topic due to the proliferation of low-cost smart meters in distribution systems as a replacement of phasor measurement units (PMUs). In this case, noise measurements and uncertainty due to malfunction associated with smart meters are generally unknown and difficult to characterize in real-time applications [7].

The main challenges of applying a traditional state estimation in microgrids can be summarized as follows: (1) microgrids are typically unbalanced three-phase AC distribution systems. Hence, there is an increase in the required number of variables and equations, compared to classical approaches; (2) the X/R ratios in microgrids are significantly low. As a result, fast decoupled state estimators [8] cannot be used, and conventional DC state estimators become irrelevant [9]; (3) limited observability with only a few measurements available [10]; (4) topological changes (i.e., transition from main grid-connected to islanded mode, and vice-versa) can lead to different operating points in a short amount of time [11]; and (5) PMUs are not as ubiquitous as in transmission systems.

In this context, considerable efforts have been made to develop state estimators for distribution networks [12]. However, very few works have addressed the development of state estimators for AC microgrids [13], [14], [15], [16]. Most static estimators are based on the classical weighted least squares (WLS) method [17]. In [18], the WLS method was applied to multiple microgrids, resulting in high computational complexity. Other works have focused on scrutinizing the problem of dynamic state estimation for microgrids [19]. For instance, in [20] a modified extended Kalman filter (EKF) [21] was proposed to estimate the dynamic variables of generator units, as well as the static variables related to the network. In [22], a distributed dynamic state estimator was proposed to estimate the operation of the energy resources and the status (connected or islanded) of the microgrid. In [23], two techniques for state estimation in droop-controlled islanded microgrids were proposed, via an unscented Kalman filter (UKF) [24] and a non-linear particle filter. Microgrids usually comprise high-speed conversion systems associated with extremely fast dynamics, on the order of nano-seconds. Thus, dynamic state estimation would require expensive resources for very limited applications. On the other hand, estimation of steady-state electricity variables with scarcity of smart meters, communication networks, and computation technology is more reasonable and cost-benefit approach.

The main differences among the aforementioned works lie in the choice of the state variables and how the measurements are considered. By their choice of state variables, state estimation approaches can be broadly classified into two major groups: (1) node voltage estimators and (2) branch current estimators [12]. Both approaches can be formulated in polar or rectangular coordinates. Regarding data, some methods use raw measurements (i.e., originally measured or forecasted values), and others use equivalent voltage and current measurements. [12]. Forecasted data approaches are not adequate for microgrids able to seamlessly disconnect from the grid and operate in an islanded mode in the event of a disturbance. Nevertheless, most state estimation methods for microgrids are mainly based on WLS using analytical formulations to model the system and truncated iterative methods to solve the non-linear set of equations [25].

Traditional state estimation using WLS is widely used in transmission systems [26], and it has been found to be suitable for observable grid-connected microgrids [27]. The non-linear relationship between the state variables and measurements results in a state-dependent Jacobian matrix that has to be updated at each iteration of the state estimation process. Because of its state-dependent Jacobian (and thus gain) matrix, applying this method to microgrids is computationally expensive [28]. Moreover, WLS requires the meter error variance (usually noted as $\sigma$) as an input parameter to perform the estimation.

Classical WLS approaches consider $\sigma$ as a known constant. This work addresses the possibility of the measurements having different noise levels. The accurate knowledge of noise characteristics in state estimation methods is a prerequisite for designing a high-performance estimator [3]. However, characterizing noise using low-cost smart-meters can be challenging because the sampling rate ranges from minutes to hours [7]. Previous works considered measurements from (PMUs) which typically have a sampling rate of more than 50 samples per second where noise characterization is feasible [2]. Other works have used machine learning only to assist classical state estimators (as in [29], [30]). However, very few works [31], [32] have addressed the direct use of machine learning state estimation (MLSE) for AC microgrids. State estimation can be considered as a regression problem [33]. Therefore, if the noise ($\sigma$) is assumed known, constant, independent, and identical distributed (iid) from the parameters (i.e., homoscedastic), it is possible to develop a model for mapping the available measurements to states variables using supervised learning to train any regression model [33] with a suitable dataset. In this work, the noise is considered input-dependent (i.e., heteroscedastic), which is a more realistic assumption when low-cost smart meters with low sampling rates are used instead of PMUs [7]. Heteroscedastic noise encodes the possibility of the measurements to have different variances $\sigma$ for various real-time observations. On the other hand, the proposed approach estimates the state without specify (i) an initial guess estimation point (a.k.a., flat start), nor (ii) an explicit definition of the measurements’ variances ($\sigma$), as is the case of the classic WLS approach.

While many existing methods focus on solving state estimation problems under simplified assumptions, such as homoscedastic measurements, this manuscript argues the importance of exploring ways for merging simulation and machine learning to solve state estimation under varying conditions, such as heteroscedastic measurements. A primary motivation of this study is the potential future applications to handle the unexpected changes in the dynamic and stochastic nature of the microgrids that can be numerically intractable using classical approaches.
The proposed MLSE uses a synaptic matrix of weights $W$ that captures the functional relationship between state variables and the available measurements, in order to estimate the state variables accurately. The training model is based on Moore-Penrose left pseudoinverse, enhanced through a Tikhonov regularization. A robust MLSE against heteroscedastic uncertainty is obtained using an approach named in this paper as recursive average model. AC power flow simulations via OpenDSS [34] are used as the source of knowledge. Results are obtained using data from a real-world microgrid located at the main campus of the State University of Campinas (UNICAMP), in São Paulo, Brazil. An overview of the microgrid is shown in Fig. 1 [35]. The performance and accuracy of the proposed MLSE is analyzed during grid-connected and islanded operation modes with noisy measurements, each with a different signal-to-noise ratio (SNR). In summary, the contributions of this paper are as follows:

- A novel MLSE for unobservable three-phase unbalanced AC microgrids, that does not require previous knowledge of the meter error variance $\sigma$ to deal with the heteroscedastic nature of measurements;
- A proposed regularization operator that captures the functional relationship between state variables and available measurements;
- A novel learning approach named recursive average model for robustness against heteroscedastic uncertainty.

II. CLASSICAL STATE ESTIMATION

Given a set of $m$ available noisy measurements $z = (z_1, z_2, \ldots, z_m)^T$, state estimation determines the most-likely $n$ state variables $\hat{x} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)^T$ of the microgrid. In this case, the required state variables are the voltage magnitudes ($V_{i,p}$) and phase voltage angles ($\theta_{i,p}$), where $i$ is the bus number, given by $i \in \{1, 2, 3, \ldots, n_B\} \equiv B$, $n_B$ is the total number of buses, and $p$ is the phase defined by $p \in \{a, b, c\}$. Thus, the set of state variables consists of $3n_B$ elements of the voltage magnitudes, plus $3(n_B - 1)$ elements of the phase voltage angles. Note that an estimation of $x$ is sufficient to determine all remaining electrical variables [25].

Set $z$ contains different types of measurements, such as active power injection ($P_{i,p}$), reactive power injection ($Q_{i,p}$), current magnitude injection ($I_{i,p}$), active power flow ($P_{ij,p}$), reactive power flow ($Q_{ij,p}$), current magnitude ($I_{ij,p}$), among others, where $j \in B$ and $ij$ is a branch. Thus, given a set of measurements $z$, an analytical expression that maps the estimated state variables $\hat{x}$ with $z$ is given by (1).

$$
\begin{bmatrix}
  z_1 \\
  z_2 \\
  \vdots \\
  z_m
\end{bmatrix}_{m \times 1} =
\begin{bmatrix}
  h_1(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) \\
  h_2(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n) \\
  \vdots \\
  h_m(\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_n)
\end{bmatrix}_{m \times 1} + 
\begin{bmatrix}
  e_1 \\
  e_2 \\
  \vdots \\
  e_m
\end{bmatrix}_{m \times 1}
$$

(1)

In which $\hat{x} \in \mathbb{R}^n$ contains the estimated state variables, $z \in \mathbb{R}^m$ contains the available measurements, and $e \in \mathbb{R}^m$ is the added noise for each measurement. Note that if the number of real-time measurements $z$ is lower than the number of required state variables $\hat{x}$, observability might not be achievable and, as a consequence, it is impossible to calculate state variables without using pseudo-measurements [9]. The minimum amount of measurements $m_{\text{min}}$ needed for the method to work is $m_{\text{min}} = 2n - k$, where $k$ is the number of defined slack buses. In this case, the number of state variables is $2n - k$, assuming the microgrid contains $n$ buses, then the microgrid is described by $2n$ variables, namely $n$ voltage absolute values and $n$ voltage angles. However, to improve accuracy, the number of redundant measurements should be higher. In practice, a value of $m \approx 4n$ is often considered reasonable [9]. Thus, (1) can be written as:

$$
\begin{bmatrix}
  z_1 \\
  z_2 \\
  \vdots \\
  z_m
\end{bmatrix}_{m \times 1} =
\begin{bmatrix}
  h(\hat{x}) + e
\end{bmatrix}_{m \times 1}
$$

(2)

where, vector $h(\hat{x})$ are functions that maps the state variables to the measurements and the stochastic noise $e$ is modeled as multivariate Gaussian $e \sim \mathcal{N}(0, \Sigma_e)$. The solution of the state estimation problem using WLS approach can be expressed as follows:

$$
\min \left\{ \sum_{i=1}^m W_i r_i^2 \right\}
$$

(3)

where $i \in \{1, 2, 3, \ldots, m\}$ is the measurement used in the state estimation, $W_i$ is the weighting factor obtained from the stochastic noise model of each measurements. This is considered to be constant white Gaussian noise with zero mean. Thus, $W_i = \sigma_i^{-2}$ and $r_i$ are residuals computed as:

$$
r_i = h_i(\hat{x}) - z_i
$$

(4)

As indicated by equation (3), the presence of different noise parameters $\sigma$ in the objective function implies that dissimilar evaluations of the same set of measurements $z$ will lead to erroneous states values $\hat{x}$.

III. DATA GENERATION

Training a machine learning model with heteroscedastic noise measurements introduces uncertainty in the machine learning model. In order to deal with this challenge, the following data generation process is performed. From the statistical perspective, state variables generate a random vector $x$ obtained
from random variables that follow a multivariate normal distribution $\mathcal{N}(\mu, \Sigma)$. In this case, given $\mathbf{x}$, it is possible to obtain a $m$-tuple of noiseless measurements $\mathbf{h}(\hat{\mathbf{x}})$ using well-known analytical or computational microgrid models $h(\cdot)$, such as those in [36]. In practice, the information about those parameters that define the multivariate normal distribution $\mathcal{X}$, is not always available. However, using samples generated from power flow simulations, it is possible to train the machine learning model $\mathbf{W}_{\mathcal{D}_{\text{train}},\lambda}$ to capture the non-linear relationship between the state variables $\mathbf{x}$ and measurements $\mathbf{z}$. Subscript $\mathcal{D}_{\text{train}},\lambda$ refers to the training dataset and the regularization hyperparameter used to fit the model $\mathbf{W}$.

This will be explained extensively in the next sections. In this paper, the microgrid was first modeled in OpenDSS [34] to generate a noiseless dataset $\mathcal{D}_{\infty} = \{(\mathbf{z}^{(i)}) \in \mathcal{Z}, (\mathbf{x}^{(i)}) \in \mathcal{X}\}_{i=1}^{\infty}$, using traditional power flow simulations, with random levels of demand and renewable generation in every event $\delta$. As shown in Fig. 2 one thousand and five hundred events ($\delta = 1500$) were generated.

For numerical stability, a normalizing process to make all the state variables $\mathbf{x}$ lie between 0 and 1, in the generated data $\mathcal{D}_{\infty}$.

A. HETEROSCEDASTIC NOISE MEASUREMENTS

The noise used in the proposed MLSE is heteroscedastic. Thus, it can be defined as follows:

$$\mathbf{z} = \mathbf{h}(\hat{\mathbf{x}}) + \mathbf{e}(\mathbf{z}), \quad \mathbf{e}(\mathbf{z}) \sim \mathcal{N}(0, \sigma_{\mathbf{e}}^2)$$

(5)

To transform the noiseless dataset $\mathcal{D}_{\infty}$ into heteroscedastic noise measurements, a signal-to-noise (SNR) ranging from 20 to 40 dB was defined, with increments of 1 dB, based on previous publications [37], [38], [39], [40], [41], [42] that worked with state estimation over homoscedastic measurements. SNR quantifies the measurement error for each measurement using (6).

$$\text{SNR}_{[\text{dB}]} = 10 \log_{10} \frac{\mathbb{E}[\mathbf{e}_m^2]}{\sigma_{\mathbf{e}}^2}$$

(6)

The error $e_m \sim \mathcal{N}(0, \sigma_{\mathbf{e}}^2)$ is injected as in (5). Note that the state variables $\mathbf{x}$, computed in $\mathcal{D}_{\infty}$, correspond to a unique microgrid operating point despite the fact that the measurements $\mathbf{z}$ are randomly perturbed, causing an increment of uncertainty in each measurement $e_m$. The set $\mathcal{D}_{\text{set}} = \{\mathcal{D}_{\infty}, \mathcal{D}_{\text{SNR}20}, \mathcal{D}_{\text{SNR}21}, \ldots, \mathcal{D}_{\text{SNR}40}\}$, that contains several noisy datasets, eliminates the need to specify the nature of the measurement error $\mathbf{e}(\mathbf{z})$ during the training process. As shown in Fig. 2, from the available data $\mathcal{D}_{\text{set}}$ a stratified sampling strategy is employed to select $\mathcal{D}_{\text{test}} = 500$ records for test at random [43]. Then, a second stratified sampling method is used to divide the rest of the set into a training set $\mathcal{D}_{\text{train}} = 500$ and a validation set $\mathcal{D}_{\text{val}} = 500$. In this case, the available datasets are: (i) $\mathcal{D}_{\text{train-set}} = \{\mathcal{D}_{\text{train, }\infty}, \mathcal{D}_{\text{train, }20}, \ldots, \mathcal{D}_{\text{train, }40}\}$, (ii) $\mathcal{D}_{\text{val-set}} = \{\mathcal{D}_{\text{val, }\infty}, \mathcal{D}_{\text{val, }20}, \ldots, \mathcal{D}_{\text{val, }40}\}$, and (iii) $\mathcal{D}_{\text{test-set}} = \{\mathcal{D}_{\text{test, }\infty}, \mathcal{D}_{\text{test, }20}, \ldots, \mathcal{D}_{\text{test, }40}\}$. Due to the noisy heteroscedastic nature of the measurements, traditional multi-output linear regression methods without regularization are inadequate because different measurements could return the same state variables. Other well-known approaches that use classical regularization methods, such as the Ridge, Lasso & Elastic Net Regression, were tested using Scikit-learn [44] Python library without achieving good results. Therefore, the next section in this paper is concerned with the computation of a machine learning model $\mathbf{W}_{\mathcal{D}_{\text{train}},\lambda}$ able to perform an approximate solution to obtain estimate state variables from noisy heteroscedastic measurements in unobservable microgrids, since there is no analytical equation to solve this problem.

IV. PROPOSED MLSE AND REGULARIZATION PROCESS

The analysis explained for $\mathcal{D}_{\text{train}}$ and $\mathcal{D}_{\text{val}}$ will be the same for all the available datasets, for each SNR level at a time. Then, as shown in Fig. 2, a tuple of training and validation datasets $\{\mathcal{D}_{\text{train}}, \mathcal{D}_{\text{val}}\}$ is selected for a determined SNR level. In general, a machine learning problem can be solved using the following three steps [45]:

1. Define the machine learning model, for instance single layer networks, feed-forward neural network, recurrent neural networks, etc.
2. Define a metric to evaluate the learning process, like mean squared error (MSE), cross entropy, etc.
3. Define an optimization algorithm to improve the machine learning model such as least squares, gradient descent, etc.
A. MACHINE LEARNING MODEL

Different from classical WLS approaches that perform an optimization process iteratively, the proposed MLSE is based in a geometric point-of-view, in which machine learning is used to fit a parametric model matrix $W$ that projects the information of available measurements to state variables. It is possible to use the least square method to find a projection operator $W_{\text{train}, \lambda_{\text{best}}}$, which is a matrix whose dimensions depend on the number of $n$ state variables and $m + 1$ measurements, as follows:

$$
W_{\text{train}, \lambda_{\text{best}}} = \begin{bmatrix}
W_{1,0} & W_{1,1} & W_{1,2} & \ldots & W_{1,m} \\
W_{2,0} & W_{2,1} & W_{2,2} & \ldots & W_{2,m} \\
W_{3,0} & W_{3,1} & W_{3,2} & \ldots & W_{3,m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
w_{n,0} & w_{n,1} & w_{n,2} & \ldots & w_{n,m}
\end{bmatrix}
$$

where, the regression parameters of the model $w_{n,m} \in \mathbb{R}$, are not known, and must be estimated from the dataset $D_{\text{train}}$. In this case, $n$ by $m + 1$ regression parameters are used to estimate $n$ state variable, from $m$ available measurements, in which the regression parameters $(w_{i,0})_{i=1}^{n}$ correspond to the fixed offset for a basis function, a.k.a., intercept [45].

In this work, the machine learning model $W_{\text{train}, \lambda_{\text{best}}}$ was defined as a single layer network, since it has no memory, i.e., the estimated state variables depend only on current measurements. In practice, $W_{\text{train}, \lambda_{\text{best}}}$ performs an estimation of the state variables from available measurements, using equation (8):

$$
\begin{bmatrix}
\hat{x}_1 \\
\hat{x}_2 \\
\hat{x}_3 \\
\vdots \\
\hat{x}_n
\end{bmatrix} =
\begin{bmatrix}
w_{1,0} & w_{1,1} & w_{1,2} & \ldots & w_{1,m} \\
w_{2,0} & w_{2,1} & w_{2,2} & \ldots & w_{2,m} \\
w_{3,0} & w_{3,1} & w_{3,2} & \ldots & w_{3,m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
w_{n,0} & w_{n,1} & w_{n,2} & \ldots & w_{n,m}
\end{bmatrix}
\begin{bmatrix}
1 \\
z_1 \\
z_2 \\
\vdots \\
z_n
\end{bmatrix}
$$

B. LEAST-SQUARE OPTIMIZATION MODEL

In state estimation, it is required that estimated state variables $\hat{x}$ to be as close as possible to the real values of $x$, for all feasible scenarios of microgrid operation, i.e., $\forall \delta$, where $\delta$ is the number of available events in the training dataset $D_{\text{train}}$. To do so, the well-regarded mean squared error (MSE) [46] is used as the metric to train and validate the proposed model:

$$
\text{MSE}_{\text{\delta}}(\hat{x}^{\delta}, x^{\delta}) \triangleq \mathbb{E}[\|x^{\delta} - \hat{x}^{\delta}\|^2_2]
$$

where $\mathbb{E}[\cdot]$ denotes the expected value and $\|\cdot\|^2_2$ is the squared Frobenius norm. Since the MSE can be assessed as a measure of the aggregated contributions of all estimated variables, it is possible to rewrite (9) as:

$$
\text{MSE}_{\text{\delta}}(\hat{x}^{\delta}, x^{\delta}) = \frac{1}{n} \sum_{n=1}^{n} \left[ \text{MSE}_{\text{\delta}_n}(x_n^{\delta}, \hat{x}_n^{\delta}) \right]^2
$$

Given that the number of events using $\delta$ AC power flow simulations is a finite set, the expected value can be approximated by the sample mean computed as:

$$
\text{MSE}_{\text{\delta}_n}(x_n^{\delta}, \hat{x}_n^{\delta}) = \frac{1}{\delta} \sum_{j=1}^{\delta} (x_n^{(j)} - \hat{x}_n^{(j)})^2
$$

Which is the same metric used to evaluate linear least-squares models in estimation theory and in linear regression models [47]. From equation (11), a natural method to determine the best estimation of $\hat{x}$ is to obtain the matrix $W_{\text{train, } \lambda_{\text{best}}}$, which minimizes the MSE applied to the difference between the true state variables $x$ and the estimated counterparts $\hat{x}$:

$$
W_{\text{train, } \lambda_{\text{best}}} = \arg \min \text{MSE}_{\text{\delta}_n}(x_n^{\delta}, \hat{x}_n^{\delta})
$$

Using (8) and dataset $D_{\text{train}}$, equation (13) is obtained.

$$
\begin{bmatrix}
\hat{x}_1 \\
\hat{x}_2 \\
\hat{x}_3 \\
\vdots \\
\hat{x}_n
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & \ldots & 0 \\
0 & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{bmatrix}
\begin{bmatrix}
W_{1,0} & W_{1,1} & W_{1,2} & \ldots & W_{1,m} \\
W_{2,0} & W_{2,1} & W_{2,2} & \ldots & W_{2,m} \\
W_{3,0} & W_{3,1} & W_{3,2} & \ldots & W_{3,m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
w_{n,0} & w_{n,1} & w_{n,2} & \ldots & w_{n,m}
\end{bmatrix}
\begin{bmatrix}
z_1 \\
z_2 \\
z_3 \\
\vdots \\
z_n
\end{bmatrix}
$$

Throughout this paper, superscript $T$ denotes transposition, hence, equation (13) can be summarized as follows:

$$
\hat{X} = \Phi W_{\text{train, } \lambda_{\text{best}}}^T
$$

Replacing (14) in (12), one obtains:

$$
W_{\text{train, } \lambda_{\text{best}}} = \arg \min \left\{ \text{tr} \left[ (X - \Phi W_{\text{train, } \lambda_{\text{best}}}^T)^T (X - \Phi W_{\text{train, } \lambda_{\text{best}}}^T) \right] \right\}
$$

The above statement can be solved as:

$$
\frac{\partial}{\partial W_{\text{train, } \lambda_{\text{best}}}} \left\{ \text{tr} \left[ (X - \Phi W_{\text{train, } \lambda_{\text{best}}}^T)^T (X - \Phi W_{\text{train, } \lambda_{\text{best}}}^T) \right] \right\} = 0
$$

A possible closed-form solution known as a Moore-Penrose left pseudo-inverse [48] for (16) is shown in (17). However, this is not a feasible solution since the Gramian matrix $(\Phi^T \Phi)^{-1}$ is singular without a regularization process, as explained in the next section.

$$
W_{\text{train}}^T = (\Phi^T \Phi)^{-1} \Phi^T X
$$

1) APPLICATION OF THE TIKHONOV REGULARIZATION

In practice, the linear system of equations in (13) is an ill-posed problem because it does not meet the following three Hadamard criteria for well-posedness: (i) for all admissible data, a solution exists, (ii) for all admissible data, the solution is unique, and (iii) the solution depends continuously on the data [49]. Thus, small perturbations in measurements generate large errors in $\hat{x}$, when the Moore-Penrose left pseudoinverse is used (i.e., equation (17)) to fit the parametric model $W_{\text{train}}$ without regularization. This solution is useless due to the severe propagation of large errors, caused by large norm of $W_{\text{train}}$ matrix. A matrix with a large norm is called
an ill-conditioned matrix. An ill-conditioned matrix can take a unit-length vector and stretch it by a large amount. Thus small uncertainties in the domain vector get magnified and lead to large uncertainties in the range. Linear systems of equations, such as (13), with a matrix of the weights $W_{\text{train}}$, are often referred as linear discrete ill-posed problems [50]. The standard way to obtain stable solutions is to modify the problem by replacing (17) with a nearby problem, whose solution is less sensitive to large errors. This replacement is known as regularization [51], [52], [53]. While many methods focus on use of zeroth order Tikhonov without considering $w_{n,0}$ [54], such as Ridge regression model, this work proposes a new regularization operator $R$ that adds terms containing regularization parameters to (17) in order to control the norm of $W_{\text{train}}$ and obtain an stable estimation, as shown in Fig. 3.

The remainder of this section describes the regularization process shown in Fig. 3, which details the process to select the best regularization hyperparameter $\lambda$. This is of central importance for equation (18) instead of the not regularized least squares solution, in equation (17), aiming at properly characterizing $W_{\text{train,}\lambda}$, for a given set of noisy measurements.

To the best of the authors’ knowledge, the regularization operator $R$ presented in this work has yet to be proposed. Therefore, the solution is obtained by a closed-form solution, similar to the one based on the pseudo-inverse, given by:

$$W_{\text{train,}\lambda}^T = (\Phi^T \Phi + \sqrt{\lambda} R)^{-1} \Phi^T X$$

(18)

The design of the regularization is governed by available knowledge or model assumptions about the state estimation (i.e., prior information) and, to lesser extent, by the implementation or computational complexity. To solve linear discrete ill-posed problems $\lambda$ in equation (18) is known as the Tikhonov regularization hyperparameter, and $R$ is known as the regularization operator [55]. The idea is to shift the spectrum of $\Phi^T \Phi$, i.e., shift the singular values away from zero. A singular value is the positive square root of an eigenvalue of the symmetric matrix $\Phi^T \Phi$. Note that this expression $(\Phi^T \Phi)^{-1}$ in equation (17) is invertible if and only if all the singular values of $\Phi^T \Phi$ matrix are non-zero. It is worth mentioning that $\lambda > 0$ because it serves as the coefficient that shifts the diagonals of the $\Phi^T \Phi$ matrix, a.k.a, singular moment matrix [56]. Parameter $\lambda$ provides balance between the data fidelity (first term: $\Phi^T \Phi$) and prior information assumptions (the remaining terms $R$). Note that, with $\alpha \lambda = 0$, equation (18) is equal to (17), making the term $(\Phi^T \Phi)^{-1}$ singular again.

The proposed MLSE regularization process is shown in Fig. 3. A suitable value of $\lambda$ generally is not explicitly known and must be determined using the validation dataset $D_{\text{val}}$ during the training process. To do so, a $\lambda$ search space with dissimilar values must be defined as $\lambda \in \{\lambda_1, \lambda_2, \ldots, \lambda_{1000}\}$. In this work, equation (18) is used to train candidate MLSE models with different $\lambda$’s, creating the following model search space: $W_{\text{train,}\lambda} \in \{W_{\text{train,}\lambda_1}, W_{\text{train,}\lambda_2}, \ldots, W_{\text{train,}\lambda_{1000}}\}$. To determine the best MLSE model $W_{\text{train,}\lambda_{\text{best}}}$, obtained with a specific Tikhonov regularization hyperparameter $\lambda_{\text{best}}$, the $L$-curve search methodology is performed [57]. The $L$-curve is a log-log plot of the norm of a regularized solution $\|RW_{\text{train}}^T\|_2^2$ versus the norm of the corresponding residual norm $\|x^{(0)} - \hat{x}\|_2^2$. It is a convenient graphical tool for characterizing the trade-off between the size of regularized solutions and their fit to the training dataset, as the hyperparameter $\lambda$ changes. Graphically, the best $\lambda$ is the one located in the corner of the $L$-curve, as shown in Fig. 3. This work, it was empirically determined that the regularization operator $R$ in equation (19) shows a stable solution for the non-linear three-phase state estimation problem in unbalanced AC microgrids (19), as shown at the bottom of the next page, where, $R \in \mathbb{R}^{m \times m}$.

The regularization operator $R$ is inspired by the work in [58]. In this case, $R$ plays the role of a penalizing filter, it is done by considering the anti-reflective and high-order boundary conditions [59]. Both these conditions were introduced in [60] and [61], respectively. In this case, the boundary conditions of $R$ is determined by the first and last rows based on the first-order Tikhonov regularization [62], which penalizes deviations from a constant model, i.e., it favors “flat” (constant) solutions and penalizes gradients, working as a first-difference operator in the boundary conditions. On the other hand, the interior rows (rows 2 to $m$) of $R$ penalize model “roughness” or bumpiness (curvature) rather than model gradient, based on the second-order Tikhonov regularization [62]. This favors “smooth” (constant gradient) solutions, working as a second-difference operator.

C. RECURSIVE AVERAGE MODEL

As is shown in Fig. 2, instead of using multiple $W_{\text{SNR,}\lambda_{\text{best}}}$, one for each SNR level, a novel recursive average model is proposed to generate a single feasible model $\tilde{W}$ for the complete SNR interval of noisy measurements, as follows:

1) Train a model $W_{\text{SNR,}\lambda_{\text{best}}}$ using the MLSE regularization process of Fig. 3 with each pair of available training and validation datasets for different SNR levels, obtained as in Fig. 2.

2) Use equation (20) to average the models recursively, as explained in Fig. 4.

$$\tilde{W}_{k+1} = \frac{1}{2} \left( W_{\text{SNR,}\lambda_{\text{best}}} + \tilde{W}_k \right) \quad \forall k \in \mathbb{N}$$

(20)

V. MATERIALS AND METHODS

This section first describes the microgrid system used as a case study. Then, section V-A explains the mathematical background of hypothesis tests and the criteria used for validation purposes of the proposed MLSE model. Numerical results obtained using all different scenarios are presented and comprehensively discussed in section VI.

A real-world AC microgrid located at the main campus of the State University of Campinas (UNICAMP), in São Paulo, Brazil [35], with 320 state variables $(V_{i,p}$ and $\theta_{i,p})$
and 198 measurements \((P_{i,p}, Q_{i,p}, I_{i,p})\) was modeled using OpenDSS [34], and the proposed MLSE via the workflow in Fig. 5 programmed in Python 3.8 [63]. From the WLS perspective this case of study is unobservable, without using pseudo-measurements.

**A. STATISTICAL ANALYSIS**

To analyze the accuracy of the proposed MLSE, a box-plot diagram is computed using the available \(D_{test-set}\) for each SNR, as is shown in Fig. 2. The results of the estimation are compared against the actual power flow calculations using (9). In this case, two statistical tests are performed. First, a hypothesis test on the mean value is performed with the purpose of evaluating whether the estimated state variables on each SNR level are statistically equal to the reference given by the power flow simulations. In the second test, known as the homogeneity test, each state variable \(x\) is analyzed separately with the intention of evaluating whether

\[
R = \begin{bmatrix}
1 & -1 & 0 & 0 & \cdots & 0 & 0 & 0 \\
-1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 \\
0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\
0 & 0 & -1 & 2 & -1 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \cdots & -1 & 2 & -1 & 0 \\
0 & 0 & 0 & 0 & \cdots & 0 & -1 & 2 & -1 \\
0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1 & 1
\end{bmatrix}
\]  

(19)
the set of estimated values \( \hat{x} \) (using different noise levels) has the same distribution.

1) TEST OF THE MEAN
According to [64], a hypothesis is a statement about the parameters of one or more populations. In this case, one hypothesis should indicate that there is no difference between the estimated state variables and the references values of the test dataset, while the other hypothesis indicates that they are different, i.e.:

\[
H_0: \; \hat{x} = x, \\
H_1: \; \hat{x} \neq x, 
\]

where \( H_0 \) is the null hypothesis and \( H_1 \) is the alternative hypothesis.

To test whether the null hypothesis is true or not, the \( t \)-score was chosen as the test statistic since the standard deviation of the population is unknown. A critical region was computed, following the procedure described in [64], with a significance level of \( \alpha = 0.01 \). 

\[
\chi^2_{ref} = 37.566, \quad \text{and} \quad \chi^2 < \chi^2_{ref} \quad \text{for computed} \quad \chi^2 \quad \text{lower than} \quad \chi^2_{ref}. \quad \text{In this case, one could conclude that there are no errors in the estimated state variables.}
\]

The critical region and \( \chi^2 \) were computed using the hypothesis test functions of Python software [63].

VI. NUMERICAL RESULTS
In all experiments, the basic metric used to assess the performance of the MLSE is the mean-squared error (MSE) between the true value of the state variables \( x \) (i.e., ground truth) and the output offered by the MLSE model (\( \hat{x} \)). As shown in Fig. 6, the best \( \lambda \) was 0.0004, using a search space that ranges 1e-10 to 1e+10.

\[
\chi^2 = \sum_{i=1}^{n} \frac{(\hat{x}_i - x_i)^2}{x_i} \tag{21}
\]

The hypotheses were the following:

\( H_0 \): The populations of \( \hat{x} \) are homogeneous, i.e., the set of estimated state variables do not contain errors.

\( H_1 \): The populations of \( \hat{x} \) are not homogeneous, i.e., the set of estimated state variables contains errors.

The decision is made considering the \( \chi^2_{ref} \) statistic with \( c = (n-1)(k-1) \) degrees of freedom. In this case, the test was run for each level of noise. This is, with 20 populations defined by the estimated state variables with different values of SNR, each with two categories (voltage magnitude or voltage angle) and a significance level of \( \alpha = 0.01 \). 

\[
\text{In this case, the best result using noiseless measurements was approximately 1e-5, as shown in Figs. 8 to 11.}
\]

\( 2 \) All tests were performed using a workstation with an Intel(R) Core(TM) i7 870 processor and 8 GB RAM. In this case the average elapsed time for estimating the state variables was 0.000054 s.
worst result was 1e-4, approximately. These results suggest that the MLSE model is adequate for heteroscedastic noisy measurements. This is in contrast with traditional WLS state estimation approaches previously tested in the University of Campinas, that reported about 1e-3, using homoscedastic measurements and pseudo-measurements to deal with lack of observability(cf. [65] for a broader discussion on this).

Regarding testing of the mean, the $t$-values of the test are presented on Figs. 12 - 13 (tops). It can be seen that all $t$-values are close to zero. This means that there are no $t$-values within the rejection regions determined by the critical $t$-value = ±1.974. Instead, all the estimated state variables lay inside the critical region $-1.974 < t$-value $< 1.974$. This shows that the reference state variables and the estimated state variables are equivalent. Note that, since it is known that the $t$-distribution has zero mean, the $t$-values can be either positive or negative.

In addition, the $p$-values are shown in Figs. 12 - 13 (mid- dles). They quantify the probability of observing a more extreme test statistic, in the direction of the alternative hypothesis. In a nutshell, if all $p$-values are larger than the chosen threshold (5%), this would indicate that the observation cannot occur by mere chance, and the null hypothesis of equal population means would not be rejected. In this case, all $p$-values were larger than the threshold. Consequently, the test fails to reject the null hypothesis i.e. all the estimated state variables $\hat{x}$ should not be considered different than the ground truth $x$ on any $V_{i,p}$ or $\theta_{i,p}$.

Finally, the distribution of the $p$-values is presented in Figs. 12 - 13 (bottoms), to show how much the test correctly concludes that the estimated state variables are equivalent to the reference state variables. In this work, the deployed test of the mean concludes that estimated and reference state variables are equivalent, with a confidence level of 95%.
In order to test whether the estimated state variables and the references have the same distribution, the homogeneity test explained in Section V-A was deployed. In this case, all tests return p-values greater than the $\alpha = 0.01$, as shown in Fig. 14. Thus, with a level of significance of 1%, there is no evidence to conclude that the distribution of the reference variables is different than the distribution of the estimated state variables.

A. COMPUTATIONAL COMPLEXITY ANALYSIS

In order to have a performance value of the proposed MLSE algorithm on running time, the asymptotic execution time as a fundamental measure of the computational complexity efficiency is used. The asymptotic analysis makes it possible to determine the complexity of the asymptotic time. In this paper, the asymptotic notations $\Theta$, $O$, $\Omega$ are used to describe the worst-case time complexity of the MLSE. To be consistent with the notation in equation (8) and Fig. 7, the entry corresponding to the $i$-th row and $j$-th column is denoted as $W[i,j]$. Similarly, the $i$-th entry in the measurements vector $z$ is denoted as $z[i]$. Thus, the vector of estimated state variables $\hat{x}$ of length $n$ and its $i$-th entry for $0 \leq i < n$ is defined as follows:

$$x[i] = \sum_{j=0}^{m} W[i,j] \cdot z[j]$$

Assuming that each operation can be done in $O(1)$ time. The number of operations in the worst-case scenario is $O(nm)$:

$$O(nm) = \sum_{i=0}^{n} \sum_{j=0}^{m} 1 = nm$$

This implies that the worst-case computational complexity value of the proposed MLSE in running time is $\Theta(nm)$.

The advantages of the MLSE over the methods in the literature are: (i) Reduction of the implementation complexity, i.e., the formulation yields a unique constant matrix based on the trained weights to perform state estimation resulting in a non-iterative model. Hence, the number of iterations required to obtain a solution is one. This contrasts with the traditional WLS state estimation approach, which requires many iterations. Consequently, a significant computational time reduction is observed; (ii) versatility, the proposed approach can be used with any set of measurements. Furthermore, it is possible to run parallel implementations with models trained with different groups of measures in order to improve the reliability of the proposed approach; (iii) the MLSE can perform state estimation in unobservable microgrids without pseudo-measurements, while Newton-like approaches do not; (iv) MLSE reports a stable behavior under heteroscedastic noise without explicitly knowing measurements’ variances ($\sigma$), while classical approaches require computing variances before performing state estimation.

VII. CONCLUSION AND DISCUSSION

Machine learning state estimation reports high performance in heteroscedastic noisy measurements. The proposed machine learning state estimation is suitable for unobservable microgrids with low-cost smart meters. The MLSE model requires just a small number of samples to train a model. Only 500 power flow simulations were needed, in contrast with other approaches such as [31] and [32] models that required 10 000 and 12 000 events for train a suitable model respectively, on the other hand the proposed approach, can handle heteroscedastic uncertainty on measurements. The proposed MLSE approach estimates the state without specify (i) an initial guess estimation point (a.k.a., flat start), nor (ii) an
explicit definition of the measurement’s variances (\(\sigma\)), as is the case of the classic WLS approach. Herein, MLSE, is relevant for expanding existing knowledge in the state estimation area for three-phase, unbalanced and unobservable microgrid systems that are eligible for operating in grid-connected or islanding model. In this paper, gross errors are not considered at the available measurements for the state estimation. Future work will explore how to incorporate a pre-processing stage into the MLSE model to identify measures without gross errors, such as data latency and false data injection attacks, before performing state estimation.

The main difficulties of the proposed MLSE reside in the microgrid modeling used to perform precise simulations. A detailed model of each circuit component was necessary to accurately represent the microgrid’s response to different scenarios. To overcome this challenge, the University of Campinas (UNICAMP) is currently deploying a project [35] to study and develop specialized models in the field of (i) power systems, (ii) renewable energy sources, (iii) power electronics, (iv) control systems, (v) optimization, and (vi) communications and information networks. A microgrid system modeled in OpenDSS with those specialized components was employed to perform this study.

REFERENCES

[1] C. N. Lu, J. H. Teng, and W.-H. E. Liu, “Distribution system state estimation,” IEEE Trans. Power. Syst., vol. 10, no. 1, pp. 229–240, Feb. 1995.
[2] M. E. El-Hawary, Advances in Electric Power and Energy: Static State Estimation. Hoboken, NJ, USA: Wiley, 2010.
[3] F. Ahmad, A. Rasool, E. Oszosy, R. Sekar, A. Sabanovic, and M. Elitant, “Distribution system state estimation—A step towards smart grid,” Renew. Sustain. Energy Rev., vol. 81, pp. 2659–2671, Jan. 2018.
[4] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, “A survey on state estimation techniques and challenges in smart distribution systems,” IEEE Trans. Smart Grid, vol. 10, no. 2, pp. 2312–2322, Mar. 2019.
[5] Q. Yang, J. Yang, W. Yu, D. An, N. Zhang, and W. Zhao, “On false data-injection attacks against power system state estimation: Modeling and countermeasures,” IEEE Trans. Parallel Distrib. Syst., vol. 25, no. 3, pp. 717–729, Mar. 2014.
[6] S. Fajic and K. A. Clements, “Power system state estimation via globally convergent methods,” IEEE Trans. Power Syst., vol. 20, no. 4, pp. 1683–1689, Nov. 2005.
[7] A. Ailimandari, S. Zadhkhast, J. Jatskevich, and E. Vahedi, “Using smart meters in state estimation of distribution networks,” in Proc. IEEE PES Gen. Meeting | Conf. Expo., Jul. 2014, pp. 1–5.
[8] A. Monticelli, “Fast decoupled state estimator,” in State Estimation in Electric Power Systems. Cham, Switzerland: Springer, 1999, pp. 313–342.
[9] A. Abur and A. G. Exposito, Power System State Estimation: Theory and Implementation. Boca Raton, FL, USA: CRC Press, 2004.
[10] S. Bhela, V. Kekatos, and S. Veeramachaneni, “Enhancing observability in distribution grids using smart meter data,” IEEE Trans. Smart Grid, vol. 9, no. 6, pp. 5953–5961, Nov. 2018.
[11] Y. Weng, M. D. Ilić, Q. Li, and R. Negi, “Convexification of bad data and topology error detection and identification problems in AC electric power systems,” IET Gener., Transmiss. Distrib., vol. 9, no. 16, pp. 2760–2767, Dec. 2015.
[12] A. Primadianto and C.-N. Lu, “A review on distribution system state estimation,” IEEE Trans. Power Syst., vol. 32, no. 5, pp. 3875–3883, Sep. 2017.
[13] J. Huang, V. Gupta, and Y.-F. Huang, “Electric grid state estimators for distribution systems with microgrids,” in Proc. 40th Annu. Conf. Inf. Sci. Syst. (CISS), Mar. 2012, pp. 1–6.
[14] Y. Hu, A. Kuh, A. Kavcic, and D. Nakafulji, “Real-time state estimation on micro-grids,” in Proc. Int. Joint Conf. Neural Netw., Jul. 2011, pp. 1378–1385.
[15] M. Rana, L. Li, and S. Su, “Kalman filter based microgrid state estimation and control using the IoT with 5G networks,” in Proc. IEEE PES Asia-Pacific Power Energy Eng. Conf. (APPEEC), Nov. 2015, pp. 1–5.
[16] M. Rana, W. Xiang, and E. Wang, “Hi-infinity-based microgrid state estimations using the iot sensors,” in Smart Power Distribution Systems. Amsterdam, The Netherlands: Elsevier, 2019, pp. 285–295.
[17] J. Zhao, A. Gómez-Exposito, M. Netto, L. Mili, A. Abur, V. Terzića, I. Kamwa, B. Pal, A. K. Singh, J. Qi, and Z. Huang, “Power system dynamic state estimation: Motivations, definitions, methodologies, and future work,” IEEE Trans. Power Syst., vol. 34, no. 4, pp. 3188–3198, Jul. 2019.
[18] G. N. Korres, N. D. Hatzigiouyi, and P. J. Katsikas, “State estimation in multi-microgrids,” Eur. Trans. Elect. Power, vol. 21, no. 2, pp. 1178–1199, Mar. 2011.
[19] J. Zhao, M. Netto, Z. Huang, S. S. Yu, A. Gomez-Exposito, S. Wang, I. Kamwa, S. Akhlaghi, L. Mili, V. Terzića, A. P. S. Meliopoulos, B. Pal, A. K. Singh, A. Abur, T. Bi, and A. Rouhani, “Roles of dynamic state estimation in power system model, monitoring and operation,” IEEE Trans. Power Syst., vol. 36, no. 3, pp. 2462–2472, May 2021.
[20] N. Vieyra, P. Maya, and L. M. Castro, “Dynamic state estimation for microgrid structures,” Electr. Power Compon. Syst., vol. 48, no. 3, pp. 320–332, Feb. 2020.
[21] M. Brown Do Couto Filho and I. C. S. de Souza, “Forecasting-aided state estimation—Part I: Panorama,” IEEE Trans. Power Syst., vol. 24, no. 4, pp. 1667–1677, Nov. 2009.
[22] S. Choi and A. P. S. Meliopoulos, “Effective real-time operation and protection scheme of microgrids using distributed dynamic state estimation,” IEEE Trans. Power Del., vol. 32, no. 1, pp. 504–514, Feb. 2017.
[23] H. E. Z. Farag, S. Saxena, and A. Asif, “A robust dynamic state estimation for droop controlled islanded microgrids,” Electr. Power Syst. Res., vol. 140, pp. 445–455, Nov. 2016.
[24] E. A. Wan and R. Van Der Merwe, “The unscented Kalman filter for non-linear estimation,” in Proc. IEEE Adapt. Syst. Signal Process., Commun., Control Symp., Oct. 2000, pp. 153–158.
[25] U. Kuhar, G. Kosce, and A. Švigelj, Observability of Power-Distribution Systems: State-Estimation Techniques and Approaches. Cham, Switzerland: Springer, 2020.
[26] A. Monticelli, State Estimation in Electric Power Systems. Boston, MA, USA: Springer, 1999.
[27] R. Singh, B. C. Pal, and R. A. Jabr, “Choice of estimator for distribution system state estimation.” IET Gener., Transmiss. Distrib., vol. 3, no. 7, pp. 666–678, Jul. 2009.
[28] N. I. M. Gould, T. Rees, and J. A. Scott, “Convergence and evaluation-complexity analysis of a regularized tensor-Newton method for solving nonlinear least-squares problems,” Comput. Optim. Appl., vol. 73, no. 1, pp. 1–35, May 2019.
[29] Y. Weng, R. Negi, C. Faloutsos, and M. Ilić, “Robust data-driven state estimation for smart grids,” IEEE Trans. Smart Grid, vol. 8, no. 4, pp. 1956–1967, Jan. 2017.
[30] Y. Weng, R. Negi, and M. D. Ilić, “Historical data-driven state estimation for electric power systems,” in Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm), Oct. 2013, pp. 97–102.
[31] K. R. Mestav, J. Luengo-Rozas, and L. Tong, “Bayesian state estimation for unobservable distribution systems via deep learning,” IEEE Trans. Power Syst., vol. 34, no. 6, pp. 4910–4920, Nov. 2019.
[32] K. R. Mestav and L. Tong, “State estimation in smart distribution systems with deep generative adversary networks,” in Proc. IEEE Int. Conf. Commun., Control, Comput. Technol. Smart Grids (SmartGridComm), Oct. 2019, pp. 1–6.
[33] C. Molnar, G. Casalicchio, and B. Bischl, “Interpretable machine learning—A brief history, state-of-the-art and challenges,” in Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases. Cham, Switzerland: Springer, 2020, pp. 417–431.
[34] C. Roger Dugan and D. Montenegro, Reference Guide: The Open Distribution System Simulator (OpenDSS). Washington, DC, USA: EPRI, Jun. 2019.
[35] R. Quadros, J. L. Jucá, J. G. I. Cypriano, R. P. B. D. Silva, L. C. P. D. Silva, and R. G. Bento, “Implementation of microgrid on the university campus of UNICAMP - Brazil: Case study,” J. Electron. Adv. Electr. Eng., vol. 1, no. 2, pp. 21–25, 2021.
[36] W. H. Kersting, Distribution System Modeling and Analysis. Boca Raton, FL, USA: CRC Press, 2017.
[37] D. Shi, D. J. Tylavsky, and N. Logic, “An adaptive method for detection and correction of errors in PMU measurements,” IEEE Trans. Smart Grid, vol. 3, no. 4, pp. 1575–1583, Dec. 2012.

[38] E. Ghahremani and I. Kamwa, “Local and wide-area PMU-based decentralized dynamic state estimation in multi-machine power systems,” IEEE Trans. Power Syst., vol. 31, no. 1, pp. 547–562, Jan. 2016.

[39] Q. Huang, L. Shao, and N. Li, “Dynamic detection of transmission line outages using hidden Markov models,” IEEE Trans. Power Syst., vol. 31, no. 3, pp. 2026–2033, May 2016.

[40] N. Zhou, “A cross-coherence method for detecting oscillations,” IEEE Trans. Power Syst., vol. 31, no. 1, pp. 623–631, Jan. 2016.

[41] L. Xie, Y. Chen, and P. R. Kumar, “Dimensionality reduction of synchronphasor data for early event detection: Linearized analysis,” IEEE Trans. Power Syst., vol. 29, no. 6, pp. 2784–2794, Nov. 2014.

[42] M. Brown, M. Biswal, S. Brahma, S. J. Ranade, and H. Cao, “Characterizing and quantifying noise in PMU data,” in Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM), Jul. 2016, pp. 1–5.

[43] R. Arnab, Survey Sampling Theory and Applications. New York, NY, USA: Academic, 2017.

[44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, and J. Vanderplas, “Scikit-learn: Machine learning in Python,” J. Mach. Learn. Res., vol. 12, pp. 2825–2830, Oct. 2011.

[45] T. Mitchell, Machine Learning. New York, NY, USA: McGraw-Hill, 1997.

[46] K. Shinzui, Lognormal Distributions: Theory and Applications. Evanston, IL, USA: Routledge, 2018.

[47] J. O. Ogundare, Understanding Least Squares Estimation and Geometrics Data Analysis. Hoboken, NJ, USA: Wiley, 2018.

[48] R. Penrose, “On best approximate solutions of linear matrix equations,” J. Global Optim., vol. 46, no. 3, pp. 383–393, Mar. 2010.

[49] L. R. Dykes, “New methods for solution of discrete ill-posed problems,” SIAM J. Sci. Comput., vol. 20, no. 4, pp. 1307–1325, Jan. 2004.

[50] C. A. Balanis, Advanced Engineering Electromagnetics. Hoboken, NJ, USA: Wiley, 2005.

[51] T. Mitchell, Understanding Least Squares Estimation and Geometrics Data Analysis. Hoboken, NJ, USA: Wiley, 2018.

[52] L. R. Dykes, “New methods for solution of discrete ill-posed problems,” Ph.D. dissertation, Dept. Math. Sci., Kent State Univ., Kent, OH, USA, 2016.

[53] Å. Björck and L. Eldén, Methods in Numerical Algebra for Ill-Posed Problems. Linköping, Sweden: Linköping Univ., 1979.

[54] A. Göettel, “The theory of Tikhonov regularization for Fredholm equations,” Lect. Notes Comput. Sci., vol. 29, no. 6, pp. 2784–2794, Nov. 2014.

[55] A. N. Tikhonov and V. Y. Arsenin, Solutions of Ill-Posed Problems. Linköping, Sweden: Linköping Univ., 1979.

[56] G. Groetsch, “The theory of Tikhonov regularization for Fredholm equations,” Dept. Appl. Math., Boston Pitman Publication, Boston Way, NJ, USA, Tech. Rep., 1984.

[57] A. N. Tikhonov and V. Y. Arsenin, Solutions of Ill-Posed Problems, vol. 1, no. 30. New York, NY, USA: Faculty of Computational Mathematics and Cybernetics, 1977, p. 487.

[58] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2. Cham, Switzerland: Springer, 2009.

[59] K. Ito and B. Jin, Inverse Problems: Tikhonov Theory and Algorithms, vol. 22. Singapore: World Scientific, 2014.

[60] G. Khalaf and G. Shukur, Choosing Ridge Parameter for Regression Problems. Oxford, U.K.: Taylor & Francis, 2005.

[61] G. Van Rossum and F. L. Drake, Python 3 Reference Manual. Scotts Valley, CA, USA: CreateSpace, 2009.

[62] D. C. Montgomery and G. C. Runger, Applied Statistics & Probability for Engineers, 3rd ed. Hoboken, NJ, USA: Wiley, 2002.

[63] M. A. Acuña Acuña, “Design and Implementation of a Machine Learning State Estimation Model for Unobservable Microgrids,” M.Sc. thesis, Dept. of Systems and Energy, UNICAMP, Campinas, Brazil, in 2020, where she is currently pursuing the Ph.D. degree. Her research interests include artificial intelligence, machine learning, convex optimization, and state estimation.

[64] L. F. Ugarte, J. Tuo, and M. C. de Almeida, “Assessing approaches to modeling voltage magnitude measurements in state estimators devoted to active distribution networks,” Electr. Power Syst. Res., vol. 200, Nov. 2021, Art. no. 107440.

[65] L. F. Ugarte, J. Tuo, and M. C. de Almeida, “Assessing approaches to modeling voltage magnitude measurements in state estimators devoted to active distribution networks,” Electr. Power Syst. Res., vol. 200, Nov. 2021, Art. no. 107440.