2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)

State Estimation in Smart Distribution Systems with Deep Generative Adversary Networks

Kursat Rasim Mestav† and Lang Tong†

Abstract— **The problem of distribution system state estimation using smart meters and limited SCADA (Supervisory Control and Data Acquisition) measurement units is considered. To overcome the lack of measurements, a Bayesian state estimator using deep learning is proposed. The proposed method consists of two steps. First, a deep generative adversary network is trained to learn the distribution of net power injections at the loads. Then, a deep regression network is trained using the samples generated from the generative network to obtain minimum mean-squared error (MMSE) estimate of the system state. Our simulation results show the accuracy and the online computation cost of the proposed method are superior to the conventional methods.**

Index Terms— **Distribution system state estimation, deep learning, generative adversary networks, deep regression network, SCADA, smart meter, and Bayesian inference.**

I. Introduction

The problem of state estimation is considered for distribution systems. A major obstacle to state estimation in distribution systems is that such systems are nominally *unobservable* [1], [2]. By unobservable it means that there is a manifold of uncountably many states that correspond to the same measurement. System unobservability arises when the number of sensors is not sufficiently large—typical in distribution systems—or sensors are not well placed in the network. An observable system may become unobservable when sensors are at fault, sensor data missing, or data tampered by malicious agents [3].

The popular weighted least-squares (WLS) estimator and its variants can no longer be used when the system is unobservable because a small WLS error in model fitting does not imply a small error in estimation; a large estimation error may persist even in the absence of noise. A standard remedy of unobservability is to use the so-called *pseudo measurements* based on interpolated observations or forecasts from historical data. Indeed, the use of pseudo measurements has been a dominant theme for distribution system state estimation. These techniques, however, are ad hoc and do not assure the quality of estimates.

The advent of smart meters and advanced metering infrastructure provide new sources of measurements. Attempts have been made to incorporate smart meter data for state estimation [4]–[6]. Not intended for state estimation, smart meters measure accumulative consumptions. They often arrive at a much slower timescale, *e.g.,* in 15-minute to hourly intervals, that is incompatible with the more rapid changes of DER. Unfortunately, existing techniques rarely address the mismatch of measurement resolution among the slow timescale smart meter data, the fast timescale real-time measurements (*e.g.,* current magnitudes at feeders and substations), and the need for fast timescale state estimation.

State estimation for unobservable systems must incorporate additional properties beyond the measurement model defined by the power flow equations. To this end, we pursue a *Bayesian inference* approach where the system states (voltage phasors) and measurements are modeled as random variables endowed with (unknown) joint probability distributions. Given the highly stochastic nature of the renewable injections, such a Bayesian model is both natural and appropriate.

The most important benefit of Bayesian inference is that observability is no longer required. A Bayesian estimator exploits probabilistic dependencies of the measurement variables on the system states; it improves the prior distribution of the states using available measurements, even if there are only a few such measurements. Unlike the least squares techniques that minimize *modeling error*, a Bayesian estimator minimizes directly the *estimation error*.

The advantage of Bayesian inference, however, comes with significant implementation issues. First, the underlying joint distribution of the system states and measurements is unknown, and some type of learning is necessary. Second, even if the relevant probability distribution is known or can be estimated, computing the actual state estimate is often intractable analytically and prohibitive computationally.

A. Summary of Results and Contributions

The main contribution of this work is an application of deep learning technology for distribution system state estimation when the system is unobservable by the deployed SCADAs. To this end, we develop a data-driven generative model coupled with a deep neural network that provides SCADA timescale state estimates. As a major departure of the predominantly pseudo-measurement approaches to state estimation when the power system is unobservable, the proposed solution to state estimation for the unobservable systems takes a Bayesian inference perspective, assuming the system states as random quantities. Consequently, the proposed approach is not bound by the observability assumption as required by the traditional weighted least-squares solutions.

The Bayesian inference that minimizes the mean squared error (MSE) of the state estimate is given by the conditional

Kursat Rasim Mestav†, and Lang Tong† are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY, 14850 USA e-mail: {krm264,lt35}@cornell.edu.

This work was supported in part by the National Science Foundation under Awards 1809830 and 1932501.

978-1-5386-8099-5/19/$31.00 ©2019 IEEE

mean of the system state. Simple as it may appear, the conditional mean can be difficult to compute. Fundamentally, the underlying joint probability distribution of the measurement and the system state is required. Even when such a probability distribution is given in closed-form, the computation of the conditional mean is intractable in general. Furthermore, the lack of measurement-state samples makes it impossible to learn the joint distribution directly.

The proposed deep learning approach consists of unsupervised learning of the generative model of the network injection and supervised learning of the conditional mean of the system states. Specifically, given direct or indirect measurements of the net-injection, we consider several machine learning techniques for the underlying generative models of network injections. For distribution systems with smart meter measurements, such models can be learned from smart meter measurements or estimates of network injections using parametric or nonparametric techniques, including deep learning techniques such as the generative adversary network (GAN) method.

The unsupervised learning of injection generative model is followed by supervised learning of the conditional mean of the network state. To this end, we exploit the physical model of the power system by embedding the power flow equation in generating training samples. A deep neural network with prewhitening first layer is proposed. We show that the proposed state estimator achieves several orders of magnitude improvement in accuracy and online computation costs over the classical weighted least-squares (WLS) estimates.

B. Related Work

State estimation based on deterministic state models has been extensively studied. See [1], [2] and references therein. We henceforth highlight only a subset of the literature with techniques suitable for distribution systems.

In some of the earliest contributions [7]–[10], it was well recognized that a critical challenge for distribution system state estimation is the lack of observability. Different from the Bayesian solution considered in this paper, most existing approaches are two-step solutions that produce *pseudo measurements* to make the system observable followed by applying WLS and other well-established techniques.

From an estimation theoretic perspective, generating pseudo measurements can be viewed as one of forecasting the real-time measurements based on historical data. Thus the pseudo-measurement techniques are part of the so-called forecasting-aided state estimation [11], [12]. To this end, machine learning techniques that have played significant roles in load forecasting can be tailored to produce pseudo measurements. See, *e.g.,* [13]–[18].

Bayesian approaches to state estimation are far less explored even though the idea was already proposed in the seminal work of Schweppe [19]. Bayesian state estimation generally requires the computation of the conditional statistics of the state variables. An early contribution that modeled explicitly states as random was made in [20] where load distributions were used to compute moments of states, although real-time measurements were used as optimization constraints rather than as *conditioning variables* in Bayesian inference. One approach to calculating conditional statistics is based on a graphical model of the distribution system from which belief propagation techniques are used to generate state estimates [21]. These techniques require a dependency graph of the system states and explicit forms of probability distributions. Another approach is based on a linear approximation of the AC power flow [22].

The approach presented in this paper belongs to the class of Monte Carlo techniques in which samples are generated and empirical conditional means computed. In our approach, instead of using Monte Carlo sampling to calculating the conditional mean directly as in [23], [24], Monte Carlo sampling is used to train a neural network that, in real-time, computes the MMSE estimate directly from the measurements.

The proposed technique builds on to our work on distribution system state estimation [25] with several notable differences. First, the techniques of learning of generative models are different. In the referred research it was assumed that the power injections follow a Gaussian mixture distribution. Here we propose a more comprehensive technique using generative adversarial networks (GANs). The GANs are not only a more generic method to learn distributions without having strong assumptions, it also allows us to learn a distribution when the samples are not directly observable, but observable under an operation. We proposed to change the objective function of the GANs to train it to learn the distribution of power injections using the aggregated smart meter measurements. The state estimation algorithm using regression learning is improved with a prewhitening technique and more regularization techniques.

II. System Model and Bayesian State Estimation

The system state vector of the power grid at time t is defined by $x_t^i = V_t^i \angle \theta_t^i$ where V_i^t is the voltage magnitude and $\angle \theta_t^i$ is the phase angle for the state variable of bus i. The overall system state $x_t = [x_t^1, \cdots, x_t^N]^\intercal$ is the column vector consisting of voltage phasors at all buses. A SCADA measures active/reactive power injections, power flows measurements and voltage magnitude. The SCADA measurement vector y_t and system state x_t are related by

$$y_t = h(x_t) + w_t, t = 1, 2, \dots \tag{1}$$

where t is the time index at the SCADA timescale (millisecond), $h(\cdot)$ is the measurement function, w_t the measurement noise.

A. Weighted Least Squares Solutions and Observability:

The WLS estimator is optimal for observable systems in the absence of measurement:

$$\hat{x}_{\text{WLS}}(y_t) = \arg\min_{x_t} ||y_t - h(x_t)||^2, \tag{2}$$

The goal of least squares is to minimize the modeling error. When there are not enough SCADAs installed, or some SCADAs are faulty, the system becomes *unobservable*. WLS methods need pseudo measurements or extra constraints to estimate states when the system is unobservable.

B. Smart Meters in State Estimation:

Smart meter measurements that, say, T times slower than the SCADA measurements. Then the smart meter measurement vector $z[n]$ is aggregated every T time:

$$z[n] = \sum_{i=nT-T}^{nT-1} g(x_i) + v[n] \tag{3}$$

where $g(\cdot)$ is the power injection function and $v[n]$ the measurement noise.

SCADAs are not widely deployed such that most distribution systems are unobservable. The common practice is to calculate pseudo-measurements using smart meters and solve (2).

C. SCADA and Smart Meter Based Bayesian State Estimation

We present an alternative way to exploit smart meter measurements. As a significant departure from the WLS-based techniques, we take a Bayesian viewpoint in which system state x_t is *random* and jointly distributed with the measurement y_t. Given that the power system is driven by increasingly stochastic renewable generations and demands, this assumption seems both natural and appropriate. By modeling x_t as a random variable, we are able to make use of its prior distribution $f(x_t)$ and the *discriminative model* through the conditional distribution $f(y_t|x_t)$.

Let y_t be the SCADA measurement vector and x_t the system state as defined in (1). Unlike the WLS techniques that minimize the difference between the actual measurement y_t and predicted measurement Hx_t, a Bayesian state estimation method places the optimization objective directly on the actual state x_t and its estimate $\hat{x}_t$. Using the Euclidean norm * as the measure of error, the Bayesian estimator is referred to as the minimum mean-squared error (MMSE) state estimator given by a deceptively simple solution in the form of the *conditional mean*:

$$\min_{\hat{x}_t(\cdot)} \mathbb{E}(||x_t - \hat{x}(y_t)||_2^2) \rightarrow \hat{x}_t^{MMSE} = \mathbb{E}(x_t|y_t) \tag{4}$$

A major advantage of the Bayesian formulation is that system observability is no longer required, and *the Bayesian estimator gives the lowest possible mean squared error among all estimators including all existing WLS based estimators regardless whether the system is observable*.

III. Bayesian State Estimation via Deep Learning

The Bayesian estimator comes with several nontrivial implementation challenges. First, the Bayesian formulation requires knowledge on the *generative model* - the joint distribution $f(x_t, y_t)$ - of x_t and y_t. Without access to samples of (x_t, y_t), it is difficult even to estimate the generative model. Second, even if we have $f(x_t, y_t)$, computing the conditional mean are often intractable. Several early attempts [21], [26] employed belief propagations on a graphical model to compute the solution efficiently. These methods still require the underlying graphical model; they deserve a new look through the lenses of modern machine learning.

The advent of powerful deep learning tools and computation resources such as GPU and cloud computing make it possible to overcomes the above challenges of Bayesian state estimation. The key idea of our approach is to embed the underlying physical law in the neural network learning process.

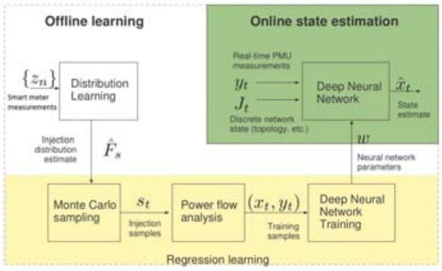

Fig. 1: Flowchart of the algorithm.

Shown in Fig. 1 is a schematic that highlights the building blocks of the proposed deep learning approach to Bayesian state estimation. The proposed scheme includes the online state estimation in the upper right and the offline learning in the rest of the figure. The online state estimation is through a deep neural network that approximates the MMSE estimator defined in with SCADA measurements y_t and discrete network state J_t (such as breaker state) as its input, and the state estimate $\hat{x}_t$ the output. The computation cost of producing state estimate in real-time is orders of magnitude lower than the WLS-based solutions and can be further reduced with special hardware implementations [27].

The offline learning includes distribution and regression learning modules. Taking samples from smart meters, weather, and other external data, distribution learning produces an estimate of the probability distribution of the (fast timescale) power injection. For each net injection sample s_t drawn from $\hat{F}_s$, the solution of the power flow equation gives the system state x_t that gives a measurement sample y_t. We now have a training sample (x_t, y_t). A collection of these training samples are used to set the neural network parameter w via an empirical risk minimization.

A. State Estimation via Deep Neural Networks:

Using the samples of states and measurements, we trained a multilayer feedforward neural network to approximate the optimal MMSE estimator. The neural network, Fig 2 consists of multiple layers of neurons. Neurons at each layer produce a vector output for the next layer using a (parameterized) nonlinear function of the output from the previous layer.

The universal approximation theorem (see, *e.g.,* [28]) has established that a neural network with a single hidden layer

*Other norms such that l_1- norm or the probability of error measure are also of interest

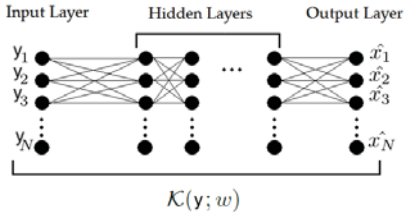

Fig. 2: Multilayer Feedforward Neural Network.

is sufficient to approximate an arbitrary continuous function. This means that with a sufficiently large neural network and appropriately chosen parameters, a neural network can well approximate the MMSE state estimator.

Given the set of training samples

$$\mathcal{S} = \{(x_t, y_t) \mid k = 1, \cdots, |\mathcal{S}|\}$$

generated, the weight matrix w is chosen to minimize the *empirical risk* defined by

$$\begin{aligned} L(w; \mathcal{S}) &= \frac{1}{|\mathcal{S}|} \sum_{k:(x_t, y_t) \in \mathcal{S}} ||x_t - \mathcal{K}(y_t; w)||^2, \\ w^* &= \arg\min_w L(w; \mathcal{S}). \end{aligned}$$

The empirical risk minimization problem above is well studied for deep learning problems, and an extensive literature exists. See, *e.g.,* [29].

For the state estimation problem at hand, the class of stochastic gradient descent algorithms is considered. The Adam algorithm [30] designed for non-stationary objectives and noisy measurements is particularly suitable.

A characteristic of deep learning is over-fitting, which means that the number of neural network parameters tends to be large relative to the training data set. To overcome over-fitting regularization techniques should be used that constraint in some way the search process in neural network training. In training, we used dropout and batch normalization. We used a constant prewhitening layer as the first layer to preprocess the inputs.

B. *Learning from Smart Meters via Generative Deep Learning :*

A barrier to train a machine learning model for state estimation is the lack of training samples. Distribution learning can be parametric or non-parametric [31], [32]. Although parametrical methods are practical and easy to use, they have biasses and their learning capacities are often limited. We propose to use aggregated historical power injection data from Smart Meters to learn the distributions using a generative machine learning model called Generative Adversarial Networks (GAN) described in [33].

As shown in Fig. 3 GAN consists of a Generative and a Discriminative networks which are simultaneously trained. While the generative network learns to generate samples similar to the data, the discrimination network learns how to detect it. After training, the generative network implicitly learns the probability distribution of the data. The objective function of GANs as described in [33] is a two-player minimax game,

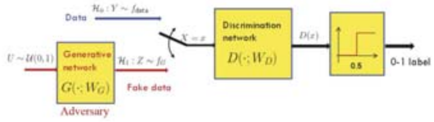

Fig. 3: Generative Adversarial Network.

$$\begin{aligned} &\min_{W_g} \max_{W_D} \mathbb{E}_{y \sim f_{data}}[log(D(y, W_D))] + \\ &\mathbb{E}_{u \sim U}[log(1 - D(G(u, W_G), W_D))], \end{aligned} \tag{5}$$

where D is the Discriminative network, G the Generative network, W_g and W_d are the parameters to learn, U the uniform distribution and f_{data} the distribution of data.

To be able to learn the fast-timescale power injection distribution using the smart meter measurements, we changed the algorithm and objective function. Instead of generating a sample from the generator, we generated T samples and aggregated them to imitate the smart meter measurements in each step. The objective function is replaced by,

$$\begin{aligned} &\min_{W_g} \max_{W_D} \mathbb{E}_{y \sim f_{data}}[log(D(y, W_D))] + \\ &\mathbb{E}_{u \sim U}[log(1 - D(\sum_{i=1}^{T} G(u_i, W_G), W_D))]. \end{aligned} \tag{6}$$

IV. Simulations Results and Discussions

A. *Simulation Settings*

a) Systems simulated: IEEE 118-Bus Test System is used in the simulations [34]. It is assumed the 99 of the busses have both load and generation. To solve the power flow equations MATPOWER toolbox [35] is used.

It is assumed there are two kind of meters in the system. (i) Smart meters measures aggregated power injection at loads every 15 minutes. They are placed at all nodes. (ii) SCADA meters are assumed to measure the active and reactive power injection and power flow. It is assumed that the number of SCADA measurements are not enough to make the system observable. It is assumed there is an independent and identically distributed additive Gaussian measurement noise with zero mean and variance set at 1% of the average net consumption value.

b) Performance measure: The average squared error(ASE) per-node defined by,

$$ASE = \frac{1}{MN} \sum_k ||\hat{x}[k] - x[k]||^2, \tag{7}$$

is used as a performance measure, where M is the number of Monte Carlo runs, k the index of the Monte Carlo run, N the number of nodes, $\hat{x}[k]$ and $x[k]$ the estimated and the state vectors, respectively. The objective function in the deep learning model is chosen as minimize mean-squared error to approximate the optimal estimator.

c) Distribution Learning via generative deep learning: We used aggregated power injection data from the Pecan Street collection† for distribution learning. It is beneficial to select the training samples of deep learning model using the historical data with similar features such as the season, hour of the day, weather, etc. We collected the net power injection data for each bus fixing the hour of the day at 5 pm from the 1st of May to the 31st of August in 2018.

We assumed that the power injection distribution at each bus is a linear transformation of one distribution, they differ only by mean and variance. We normalized all measurements to obtain the samples of that distribution. Then, we used these samples to train the GAN. As the data samples are from smart meters, we modified the objective function of training as described in (6).

We trained the generative network with 2 hidden layers and 100 neurons at each layer. Batch normalization and dropout with rate 0.2 are used in layers. Adam optimization [30] algorithm with mini-batches is used as the optimizer. Leaky-ReLU at hidden layers and a linear activation function at the final layer are used as the activation functions. For the discriminative network, we used two hidden layers with 30 neurons. Leaky-ReLU at hidden layers and a sigmoid activation function at the final layer are used as the activation functions.

After training the GAN and we generated many power injection vectors using the generator network. To verify the results, we aggregated the generated samples to imitate the smart meter measurements and plotted the cumulative distribution function (CDF) of it and empirical CDF of the raw data in Fig. 4. The figure shows empirical CDF obtained by the samples and CDF obtained from the learned distribution are similar. It verifies the training was successful. We also observed the discriminative function's output is converged to a constant value of 0.5 as expected, described in [33].

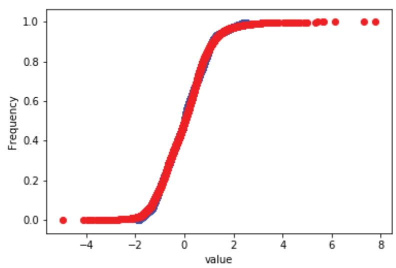

Fig. 4: The comparison of the CDFs. The red curve is the empirical CDF of the data, the blue curve is CDF of the imitated measurements from the learned distribution.

d) Neural network specification and training: We separated the power injection vectors generated by the generator into training (12000 samples), validation(4000 samples) and test(4000 samples) sets. For each one, we solved the power flow equations by MATPOWER to obtain the power flow values and states. For this network, 32 optimally placed SCADA meters make the system observable. We repeated the experiment with 8, 14, 20 and 26 SCADAs which are guaranteed to be unobservable. For each case, measurements are imitated by adding measurements noise and states. The noisy measurements of the SCADA meters are chosen as the input of the network. Prewhitening is used on the inputs. The states are chosen as the output. Then we trained a deep neural network with 5 to 10 hidden layers to estimate states on the test set.

The ReLU (Rectified Linear Units) activation function was used for neurons in the hidden layers and linear activation functions in the output layer. The Adam algorithm was used to train the neural network with mini-batches of 60 samples. Early stopping was applied by monitoring validation errors. To select an initial point for the optimization, He's normal method [36] was used. To have a better regularization, batch normalization and dropout with 0.3 dropping rate are used at hidden layers.

e) Comparison of Performances: We implemented the proposed deep learning approach to Bayesian state estimation on the IEEE 118-bus. We compared the proposed *Bayesian state estimation with deep neural network* (herein abbreviated as Bayesian NN) with two WLS-based methods in the literature:

1) WLS with pseudo measurements: referred to as *Regularized WLS* generates injection pseudo measurements by normalizing the smart meter measurement over the interval.
2) Augmented WLS: uses only SCADA measurements to estimate states. An extra constraint is added on the (2).

Fig. 5 presents the performance of three algorithms on five scenarios. It is demonstrated that for highly unobservable systems, Bayesian methods are advantageous over conventional WLS methods. Note that the MSE floor of the augmented WLS method due to the SCADA - Smart Meter mismatch.

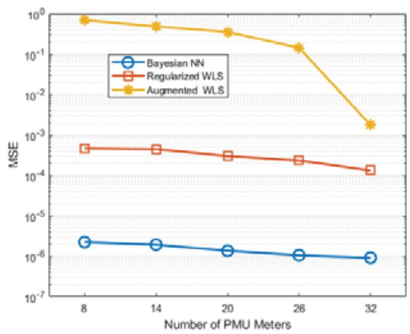

Fig. 5: The comparison of algorithms on 5 test cases.

†http://www.pecanstreet.org/

V. Conclusion

This paper presents a comprehensive Bayesian state estimation algorithm for unobservable distribution systems using SCADA and smart meters under a stochastic demand model with generation. The proposed approach employs two machine learning techniques: distribution learning of power injection using generative adversary networks and regression learning of MMSE estimator using deep learning. The performance of traditional methods are not fulfilling when the system is unobservable, where Bayesian alternatives are still viable options. The objective function of GANs is modified to learn the distribution of power injections from the historical smart meter measurements. The use of deep neural network also plays a crucial role in overcoming computation complexity in Bayesian estimation, making the online computation significantly lower than the traditional WLS solutions. Simulation results demonstrate the potential of the Bayesian state estimation for cases that are otherwise intractable for conventional WLS-based techniques.

References

[1] A. Abur and A. G. Expósito, *Power System State Estimation: Theory and Implementation*. CRC Press, 2004.

[2] A. Monticelli, *State Estimation in Electric Power Systems: A Generalized Approach*. Power Electronics and Power Systems, Springer, 1999.

[3] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, “Malicious data attacks on the smart grid,” *IEEE Transactions on Smart Grid*, vol. 2, pp. 645–658, Dec 2011.

[4] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, and E. Vaahedi, “Distribution system state estimation based on nonsynchronized smart meters,” *IEEE Transactions on Smart Grid*, vol. 6, pp. 2919–2928, Nov 2015.

[5] A. Al-Wakeel, J. Wu, and N. Jenkins, “State estimation of medium voltage distribution networks using smart meter measurements,” *Applied Energy*, vol. 184, pp. 207 – 218, 2016.

[6] Y. Gao and N. Yu, “State estimation for unbalanced electric power distribution systems using ami data,” in *2017 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, April 2017.

[7] I. Roytelman and S. M. Shahidehpour, “State estimation for electric power distribution systems in quasi real-time conditions,” *IEEE Transactions on Power Delivery*, vol. 8, pp. 2009–2015, Oct 1993.

[8] M. E. Baran and A. W. Kelley, “State estimation for real-time monitoring of distribution systems,” *IEEE Transactions on Power Systems*, vol. 9, pp. 1601–1609, Aug 1994.

[9] C. N. Lu, J. H. Teng, and W. . E. Liu, “Distribution system state estimation,” *IEEE Transactions on Power Systems*, vol. 10, pp. 229–240, Feb 1995.

[10] A. K. Ghosh, D. L. Lubkeman, M. J. Downey, and R. H. Jones, “Distribution circuit state estimation using a probabilistic approach,” *IEEE Transactions on Power Systems*, vol. 12, pp. 45–51, Feb 1997.

[11] M. B. D. C. Filho and J. C. S. de Souza, “Forecasting-aided state estimation—part i: Panorama,” *IEEE Transactions on Power Systems*, vol. 24, pp. 1667–1677, Nov 2009.

[12] M. B. D. C. Filho, J. C. S. de Souza, and R. S. Freund, “Forecasting-aided state estimation—part ii: Implementation,” *IEEE Transactions on Power Systems*, vol. 24, pp. 1678–1685, Nov 2009.

[13] A. Bernieri, C. Liguori, and A. Losi, “Neural networks and pseudo-measurements for real-time monitoring of distribution systems,” in *Proceedings of 1995 IEEE Instrumentation and Measurement Technology Conference - IMTC ’95*, pp. 112–119, April 1995.

[14] R. Singh, B. C. Pal, and R. A. Jabr, “Distribution system state estimation through gaussian mixture model of the load as pseudo-measurement,” *IET Generation, Transmission Distribution*, vol. 4, pp. 50–59, January 2010.

[15] 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, pp. 1888–1896, Nov 2012.

[16] J. Wu, Y. He, and N. Jenkins, “A robust state estimator for medium voltage distribution networks,” *IEEE Transactions on Power Systems*, 2013.

[17] A. Onwuachumba, Y. Wu, and M. Musavi, “Reduced model for power system state estimation using artificial neural networks,” *IEEE Green Technologies Conference*, 2013.

[18] F. Adinolfi, F. D’Agostino, A. Morini, M. Saviozzi, and F. Silvestro, “Pseudo-measurements modeling using neural network and fourier decomposition for distribution state estimation,” in *IEEE PES Innovative Smart Grid Technologies, Europe*, pp. 1–6, Oct 2014.

[19] F. C. Schweppe, J. Wildes, and D. P. Rom, “Power system static state estimation, Parts I, II, III,” *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, pp. 120–135, 1970.

[20] A. K. Ghosh, D. L. Lubkeman, and R. H. Jones, “Load modeling for distribution circuit state estimation,” *IEEE Transactions on Power Delivery*, vol. 12, pp. 999–1005, April 1997.

[21] P. Chavali and A. Nehorai, “Distributed power system state estimation using factor graphs,” *IEEE Transactions on Signal Processing*, vol. 63, pp. 2864–2876, June 2015.

[22] L. Schenato, G. Barchi, D. Macii, R. Arghandeh, K. Poolla, and A. V. Meier, “Bayesian linear state estimation using smart meters and pmus measurements in distribution grids,” in *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 572–577, Nov 2014.

[23] K. Emami, T. Fernando, H. H. Iu, H. Trinh, and K. P. Wong, “Particle filter approach to dynamic state estimation of generators in power systems,” *IEEE Transactions on Power Systems*, vol. 30, pp. 2665–2675, Sep 2015.

[24] A. Angioni, T. Schlösser, F. Ponci, and A. Monti, “Impact of pseudo-measurements from new power profiles on state estimation in low-voltage grids,” *IEEE Transactions on Instrumentation and Measurement*, vol. 65, pp. 70–77, Jan 2016.

[25] K. R. Mestav, J. Luengo-Rozas, and L. Tong, “Bayesian state estimation for unobservable distribution systems via deep learning,” *IEEE Transactions on Power Systems*, pp. 1–1, 2019.

[26] Y. Hu, A. Kuh, T. Yang, and A. Kavcic, “A belief propagation based power distribution system state estimator,” *IEEE Computational Intelligence Magazine*, vol. 6, pp. 36–46, Aug 2011.

[27] V. Sze, Y. Chen, T. Yang, and J. S. Emer, “Efficient processing of deep neural networks: A tutorial and survey,” *Proceedings of the IEEE*, vol. 105, pp. 2295–2329, Dec 2017.

[28] S. Sonoda and N. Murata, “Neural network with unbounded activation functions is universal approximator,” *Applied and Computational Harmonic Analysis*, vol. 43, no. 2, pp. 233 – 268, 2017.

[29] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

[30] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” *International Conference for Learning Representations*, 2015.

[31] L. Wasserman, *All of Statistics*. Springer-Verlag New York, 2004.

[32] L. Wasserman, *All of Nonparametric Statistics*. Springer-Verlag New York, 2006.

[33] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Networks,” *arXiv e-prints*, p. arXiv:1406.2661, Jun 2014.

[34] R. Christie, “Power systems test case archive,” *Univ. Washington, Seattle, WA, USA.*, 1993. `https://www.ee.washington.edu/research/pstca/pf118/pg_tca118bus.htm`.

[35] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, “Matpower: Steady-state operations, planning, and analysis tools for power systems research and education,” *IEEE Transactions on Power Systems*, vol. 26, pp. 12–19, Feb 2011.

[36] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” *CoRR*, 2015.