State Estimation for Unobservable Distribution Systems via Deep Neural Networks

Kursat Rasim Mestav, Jaime Luengo and Lang Tong

***Abstract*—The problem of state estimation for unobservable distribution systems is considered. A Bayesian approach is proposed that combines Bayesian inference with deep learning neural network to achieve the minimum mean squared error estimation of network states for real-time applications. The proposed technique learns probability distributions of net injection from smart meter data and generate samples for training a deep neural network. Results show that the proposed technique offers significant improvement in estimation accuracy and computation cost over weighted least squares methods with pseudo-measurements. Simulations are also used to evaluate robustness of the proposed Bayesian method against estimation errors in distribution learning and bad data.**

***Index Terms*—Distribution system state estimation, Bayesian inference, deep learning neural networks, smart distribution systems.**

I. Introduction

We consider the problem of state estimation for distribution systems that have limited measurements such that they are *unobservable* [1]. The states of such a system cannot be determined uniquely from the available measurements even when there is no measurement noise, and techniques such as the weighted least squared (WLS) method [1] fail in general. A standard remedy is to use the so-called *pseudo-measurements* based on interpolated measurements or historical data. Such techniques are ad hoc and in general suboptimal.

The present distribution systems are not well metered and in general unobservable. However, there have been compelling cases made for distribution system state estimation due to the rising presence of distributed energy resources (DER) in distribution systems [2]. To unlock the full potential of DER, a modernization of the distribution system is necessary to provide tighter control of power flow in real-time operations, which requires effective state estimation.

An essential barrier to state estimation for real-time control is unobservability. Although smart meters at the edge of the network have been deployed progressively, these type of measurements are typically at a much slower time scale incompatible with the more rapid changes of DER such as solar generations. Realizing state estimation for real-time operation in distribution systems, therefore, requires a fundamentally different approach from that used in transmission systems — one that overcomes the difficulty of lack of measurements.

‡ The authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853, USA.

This work is supported in part by the National Science Foundation under Award 1549989 and Iberdrola Foundation.

A. Summary of results and contributions

The main contribution of this work is a novel application of Bayesian estimation, probability distribution learning, and deep neural network learning techniques. We demonstrate the potential of such techniques for large distribution systems and provide insights into the architectural characteristics of deep neural networks.

Bayesian inference is based on the probabilistic modeling of system states. Given the highly stochastic nature of DER, modeling voltage phasors as random variables is natural. The key challenges of developing Bayesian state estimation, however, are (i) the need to learning the underlying probability distributions that define the system states and measurements; (ii) the complexity of computing conditional mean of the system states.

The main idea of the proposed technique, as illustrated in a schematic diagram in Fig. 1, is that historial measurements are used to learn the probability distributions of the net-injection. The learned distributions are used to generate samples to train a deep neural network to approximate the minimum mean squared error (MMSE) estimator of the system states. A stochastic gradient descent algorithm with early stopping is used in the training process. In real-time applications, the neural network directly computes the MMSE estimates with linear complexity of $O(N)$ where N is the size of the network. In contrast, WLS types of state estimator has the complexity roughly of the order of $O(N^3)$ per iteration.

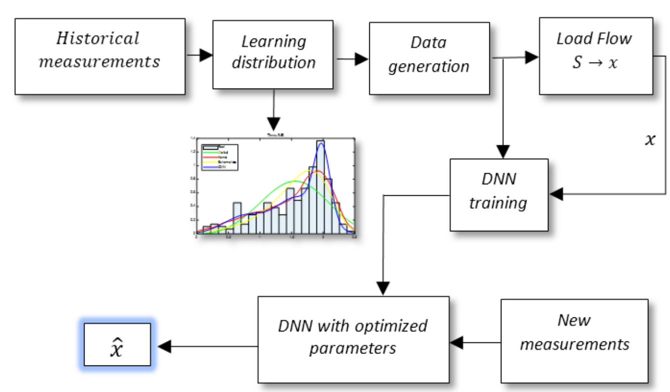

Fig. 1: Block diagram of the method.

Numerical results demonstrate several interesting features of the proposed approach. First, deep learning seems to be essential. Whereas existing neural network state estimation techniques typically use a flat network involving one or two layers, our results show that, for the tested 85 and 141 bus networks, a neural network of 10 layers or more can provide accurate estimates, achieving mean squared error (MSE) per bus at the level of 10^{-5} to 10^{-6} p.u. on test data sets. Such estimation errors are well within the typically required accuracy for voltage estimates. In contrast, the WLS methods with pseudo-measurements techniques using pseudo-measurements have errors several orders of magnitude higher.

Second, among various training techniques, the stochastic gradient decent algorithm with early stopping are effective for the cases tested. Finally, simulation results show that the proposed approach exhibits a promising level of robustness against estimation errors in probability distributions and sampling errors in training.

B. Related Work

State estimation based on deterministic models of states has been extensively studied. See [1] and references therein. Existing Bayesian techniques that model states as random are less common even though the idea was already proposed in the seminal work of Schweppe [3] where (extended) Kalman filtering techniques were proposed.

In the context of state estimation for unobservable distribution systems, Bayesian methods for state estimation can be classified into two categories: Bayesian pseudo-measurements [4]–[7] and Bayesian state estimation [8]–[17]; the former uses probability distributions to generate pseudo-measurements so that conventional (point) estimation techniques such as WLS can be applied. Such a hybrid techniques are suboptimal but can be easily incorporated in the conventional state estimation methods. The latter type uses distribution information explicitly and aims to minimize the mean squared error. These techniques vary in how the conditional mean of the system states are computed. The method proposed in this paper falls into this latter category.

Direct Bayesian state estimation requires the computation of conditional mean of the state variables. One approach is based on a graphical model of the distribution system from which belief propagation techniques are used to generate state estimates [8], [9]. 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 [10]. The proposed approach belongs to the class of Monte Carlo techniques where samples are generated and empirical conditional means are computed. The main difference between our approach and existing techniques [11]–[13] is the way conditional means are computed in real-time. Instead of using Monte carlo sampling to compute the conditional mean directly as in [11]–[13], Monte carlo sampling is used to train a neural network that, in real-time, computes the MMSE estimate directly from the measurements.

Neural networks have been proposed for state estimation as early as [14]. Different architectures of neural networks have been considered: parallel distributed processing in [15], auto-encoder in [16]. Although not casted as a Bayesian state estimation, the approach in [17] appears to be quite close to ours. In [17], a multilayer neural network is used to estimate states directly using load bus measurements as the input of the neural network. The distributions of power injections are assumed. In our approach, the distribution of net injections are learned from smart meter measurements and more sophisticated deep learning techniques and a deep neural network architecture are used.

II. Network and measurement models

We assume an unbalanced three phase distribution system. The three-phase voltage phasors at bus i is a complex column vector $x_i = [x_i^1, x_i^2, x_i^3]^\intercal$ where the superscripts are phase indices and $x_i^k = V_i^k \angle \theta_i^k$ where V_i^k is the voltage magnitude and θ_i^k is the phase angle for the state variable at phase k of bus i. The overall system state $x = [x_1, \cdots, x_N]^\intercal$ is the column vector consisting of voltage phasors at all buses.

The vector of measurements z is a function of the state x, and measurement error e modeled by

$$z = h(x) + e, \quad S = g(x) \tag{1}$$

where $h(x)$ is the measurement equation, S the vector of power injections, and $g(x)$ the power flow equation. When the system is unobservable, each measurement (even when $e = 0$) is associated with a manifold of states.

Different configurations of measurements can be assumed. The measurement vectors may include some but insufficient number of the following variables, for each phase $k \in \{1, 2, 3\}$,

(P_i^k, Q_i^k) : Active/reactive power injection at node i;
(P_{ij}^k, Q_{ij}^k) : Active/reactive power flow from node i to j;

Other measurements such as current magnitudes and accumulative power from smart meters can also be included.

Power flow equations are used to relate the state variables with the measurements:

$$P_i^k = V_i^k \sum_{l=1}^{3} \sum_{j=1}^{n} V_i^l [G_{ij}^{kl} \cos(\theta_i^k - \theta_j^l) + B_{ij}^{kl} \sin(\theta_i^k - \theta_j^l)] \tag{2}$$

$$Q_i^k = V_i^k \sum_{l=1}^{3} \sum_{j=1}^{n} V_i^l [G_{ij}^{kl} \sin(\theta_i^k - \theta_j^l) + B_{ij}^{kl} \cos(\theta_i^k - \theta_j^l)] \tag{3}$$

where G_{ij}^{kl} and B_{ij}^{kl} are the conductance and susceptance between node i and j from phase k to l. The LHS defines the three-phase net complex power injection S_i at node i whose elements are $S_i^k = P_i^k + jQ_i^k$. In absence of measurement noise, given the set of active and reactive power injections, the above equation can be solved to obtain the system states, which in turn give the branch power and current flows.

III. Bayesian solution via Deep Neural Network

A. Bayesian State Estimation

Bayesian state estimation starts with defining the probability space that specifies the joint distribution of the measurement z and state x. For a distribution system with stochastic injections, the probability space is defined by the independent random vector S of net-injection and measurement error e.

From the power flow equations (2)-(3), S determines the system state x (in the forms of (V, θ)) , which in turn determines measurement z. Thus the joint distribution of $F_{S,e}$ specifies the joint distribution $F_{x,z}$.

A Bayesian estimator $\hat{x}(z)$ of x is a measurable function of z. The MMSE estimator is given by

$$\hat{x}^*(z) = \arg\min_{\hat{x}} \mathbb{E}(||x - \hat{x}(z)||^2) = \mathbb{E}(x|z).$$

Unfortunately, the functional form of the MMSE estimator is highly complex. One approach is to use a Monte Carlo technique that, for given measurement z, compute samples from generated conditional distribution $F_{x|z}$. For real-time applications, such an approach is difficult to implement. We propose next an alternative that approximates $\hat{x}^*$ using a deep neural network.

B. Neural Network Approximation

Fig 2 shows a K layer network, where the (first) input layer receives the input z, and the output layer produces estimates of x.

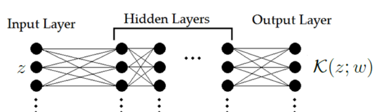

Fig. 2: Multi-layer Perceptron Model.

Each middle layer has a set of neurons connecting the outputs of the neurons from the previous layer and producing outputs for the next layer using a nonlinear function, in our case hyperbolic tangent function. For the ith neuron in the interior layer j, its output z_i^{j+1} is given by

$$z_{i,j} = \frac{\exp(u_{i,j}) - \exp(-u_{i,j})}{\exp(u_{i,j}) + \exp(-u_{i,j})}, \ \ u_{i,j} = w_{i,j}^{(0)} + \sum_k w_{i,j}^{(k)} z_{k,j-1},$$

where $\{w_{ij}^k\}$ is the set of weights associated with neuron (i, j). For the Kth (output) layer, the output of neuron j is an affine function of the output of the previous layer. Specifically, an estimate of a state variable given by

$$z_{i,K} = w_{i,K}^{(0)} + \sum_k w_{i,K}^{(k)} z_{k,K-1}.$$

The outputs collectively produce an estimate of the state x.

In approximating the MMSE estimator, the neural network weight parameter w is set to minimize the MSE of its estimate

$$w^* = \arg\min_w \mathbb{E}(||x - \mathcal{K}(z; w))||^2). \tag{4}$$

The above optimization is only conceptual, however, because the expectation operator requires explicit joint distribution of z and x. We show next how w^* can be obtained though a neural network training process.

C. Bad Data Detection and Mitigation

Bad data are anomalies in data collection that are common in transmission systems and potentially more significant in distribution systems. Bad data detection can be implemented using a generalized likelihood ratio test that has α as the probability of type I (false positive) error and β the probability of type II (false negative) error. When the bad data test is positive, the measurement used in the input of the neural network should be replaced by the mean of the prior distribution. This implies that α percentage good measurements are replaced by the mean of the prior distribution whereas β percentage of the bad data are missed, which introduces statistical deviation of the nominal measurements. Our numerical results indicate that Bayesian state estimates via neural networks is quite robust to bad data.

IV. Training Neural Network

A. Learning Net Injection Distributions

In our proposed method, the first step is to estimate net injection distribution using the historical load data. That distributions strongly depend on the days of the week, seasons and geographical locations. We cluster the historical data with common attributes.

We consider parametric models for net injection distributions. Among commonly used models such as Gaussian, Weibull, Gamma, etc., Gaussian mixture model appears to be the most flexible and accurate. Indeed, Gaussian mixtures have been used to model load distributions in [6]. From historical data, the maximum likelihood method is used to estimate parameters of the Gaussian mixture model. Performance of this approach is shown in Section V.

B. Training the Neural Network

To train the neural network, we need to generate state and measurement training samples $\mathcal{S} = \{(x[k], z[k])\}$. To this end, we draw net-injection samples from the learned net-injection distributions. In particular, given an injection sample $S[k]$, the power flow equations (2)-(3) are used to solve for system state $x[k]$ and measurement $z[k]$. Two additional sets of samples are generated independently, $\check{\mathcal{S}}$ for testing performance, $\tilde{\mathcal{S}}$ for validation.

To approximate the MMSE state estimator, the weight of the neural network is chosen to minimize the empirical risk

$$\begin{aligned} w^* &= \arg\min_w L(w; \mathcal{S}) \\ &= \arg\min_w \frac{1}{|\mathcal{S}|} \sum_{k:(x[k], z[k]) \in \mathcal{S}} ||x[k] - \mathcal{K}(z[k]; w)||^2 \end{aligned} \tag{5}$$

The empirical risk minimization problem above is well studied for deep learning problems. For the state estimation at hand, the stochastic gradient descent algorithm [18] appears to offer the training-generalization tradeoff. In particular, the adaptive moment (ADAM) technique [19] designed for nonstationary objectives and noisy measurements appear to be most appropriate for the considered application.

A characteristics of deep learning is over-parameterization, which means that the number of neurons (weight parameters) tends to be comparable or smaller than the available training samples. A key component in deep learning, therefore, is a way to regularize the optimization of (5). Standard techniques include L_1 regularization, dropout regularization [20], and early stopping [21]. The early stopping technique, for example, uses the validation data set $\tilde{\mathbb{S}}$ to determine the stopping time for the gradient descent, thus in a way to regularize the optimization. These techniques are tested in our numerical study.

V. Simulations Results and Discussions

A. Network, data and benchmarks

The simulations were performed in a 85 and 141 bus systems defined in the MATPOWER toolbox [22]* The results of the two systems were similar; only the results for the 85 bus system are reported here. To model the relatively high penetration of DER, two-thirds of the buses were chosen to have solar PV attached to the load.

Smart meters were used to measure energy consumption of every bus with 3 types of measurement accuracy: 0.015, 0.0175 and 0.02 pu. Current magnitude meters were placed in 20% of distribution system branches to measure the current magnitude and a SCADA meter was placed at the slack bus to measure complex power injection. Both with a measurement accuracy of 0.01 pu. Smart meters had a sampling rate 30 times slower than the SCADA and current magnitude measurements.

We used the data sets from the Pecan Street collection† for distribution learning and testing. The data set was split into 21st May to 21st September 2015 for training and same dates of 2016 for test, which coincided with a summer patterned net consumption.

The proposed approach was compared with WLS methods with two kinds of pseudo-measurements of net power injections in the literature [4], [5] : i) averaging the last energy consumption measurement over the number of samples; ii) using a neural networks whose inputs are the last energy consumption vector and output the net power injection measurements for each sample.

The performance was evaluated based on the per-node average squared error (ASE) of state estimate using test data set $\check{\mathbb{S}}$ defined as:

$$\text{ASE} = \frac{1}{MN} \sum_k ||\hat{x}[k] - x[k]||^2 \tag{6}$$

where M is number of cases, N is number of nodes.

B. Learning Distributions

Gaussian Mixture Model of 3 components was selected as the best fit for both random variables, active power consumption U_i and solar power generation V_i, where i represents each node. Fig 3 exemplifies the distribution fitting for training samples of V_i, where GMM of 3 components resulted to be the best fit. On the right, it presents the variation between the distribution learned from training samples of solar generation and the actual test samples. An estimated distribution of the test samples was included for clarity in the comparison.

*The 85 and 141 bus systems are single phase systems in MATPOWER.
†http://www.pecanstreet.org/

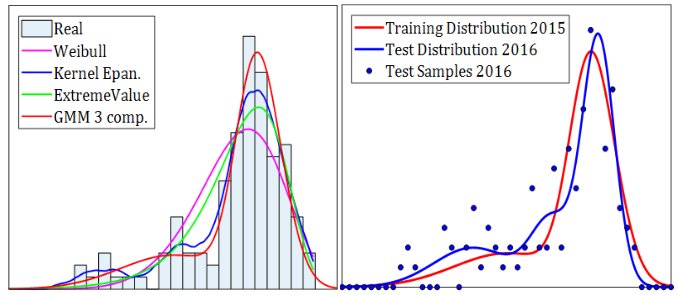

Fig. 3: Distribution fitting from solar power training samples (left) and Comparison between fitted training distribution and test samples (right)

From these distributions we generated the training and validation set comprised of 600 and 300 cases respectively. Validation set is used for early stopping. We assumed a power factor of 0.86 giving a ratio of 0.6 between active and reactive power injections. Therefore, each element of the net complex power injections vector $S[k]$ was generated as $S_i[k] = (u_i[k] - v_i[k]) + j(0.6u_i[k])$where $u_i[k]$ and $v_i[k]$ are realizations of U_i and V_i.

C. Neural Network Architecture

The first experiment consisted of exploring what network architectures would be more suitable for the Distribution System State Estimation problem. Therefore, two strategies were followed. In Fig 4 the total number of neurons was fixed and the ASE was calculated at different number of hidden layers. Three total number of neurons were specified.

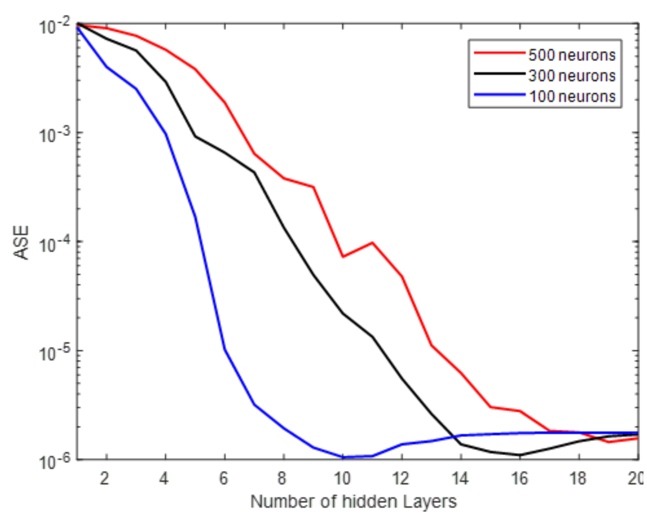

Fig. 4: NN Architecture - Total number of neurons fixed.

The results suggest a 'deeper' neural network architecture performs better than another with fewer number of hidden layers, as long as the different methods presented previously are followed to prevent overfitting.

D. Performance

We conducted simulation on the performance of the proposed state estimation technique for each hour of the day. The mean net consumption is presented as a reference, although it is shown unitless to focus on the ASE magnitude. For every hour a different neural network was trained. Fig 5 shows that the ASE of deep neural networks was clearly better than the first pseudo-measurements method.Comparison between the second pseudo-measurements method shows that training deep neural network to estimate the states directly has more efficient than estimating pseudo-measurements.

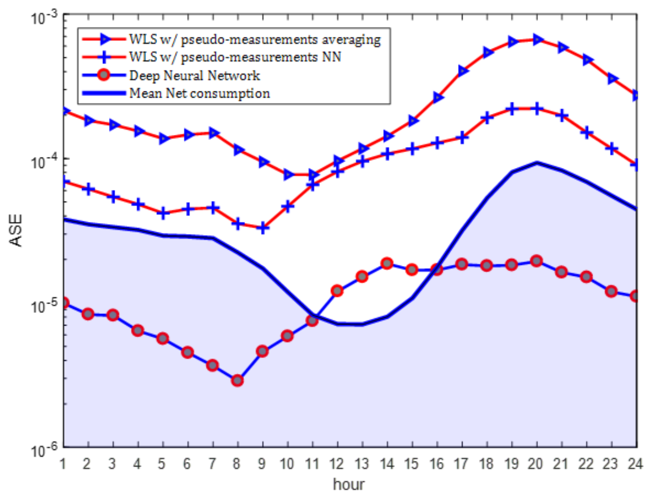

Fig. 5: Hourly simulation

For testing the robustness of our method against bad data, a generalized likelihood ratio test with $\alpha = \beta = 0.05$ was implemented.Bad data measurements are assumed to have five times higher error variance than that of the uncorrupted. In the simulation the percentage of bad data is increased in steps of 10% from 0 to 50%. The outcome of the experiment shows that the DNN ASE stayed within 5% when the number of bad data measurements increased, while PM methods' ASE increased 50% from the initial uncorrupted case. We have also conducted comparison studies between the proposed technique and WLS methods with increasing SCADA complex power measurements. With additional complex power measurements, the network became observable. As expected, the WLS technique gradually outperform the Bayesian methods. Details of these results will be reported in separate work.

VI. Conclusion

This paper presents a Bayesian approach using deep neural networks for state estimation in unobservable distribution systems. To benefit from deep neural network architectures, this approach learns net power injection distributions from historical measurements and generates more training samples. Furthermore, stochastic optimization methods and regularization methods are used to avoid overfitting. This method is computationally efficient and robust against bad data, variation of net consumption distributions and high penetration of DERs; which makes it suitable for real-time operation in the distribution system.

References

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

[2] A. Primadianto and C. N. Lu, "A review on distribution system state estimation," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3875–3883, Sept 2017.

[3] F. C. Schweppe, J. Wildes, and D. P. Rom, "Power system static state estimation, Parts I, II, III," *IEEE Tran. on Power Appar. & Syst.*, vol. PAS-89, pp. 120–135, 1970.

[4] A. Bernieri, G. Betta, C. Liguori, and A. Losi, "Neural networks and pseudo- measurements for real-time monitoring of distribution systems," *IEEE Transactions on Instrumentation on Measurements*, 1996.

[5] E. Manitsas, R. Singh, B. Pal, and G. Strbac, "Modelling of pseudo-measurements for distribution system state estimation," *SmartGrids for Distribution*, 2008.

[6] R. Singh, B. C. Pal, and R. A. Jabr, "Distribution system state estimation through gaussian mixture model of the load as pseudo-measurement," *IET Generation, Transmission Distribution*, vol. 4, no. 1, pp. 50–59, January 2010.

[7] G. Pieri, M. Asprou, and E. Kyriakides, "Load pseudomeasurements in distribution system state estimation," *IEEE PowerTech*, 2015.

[8] Y. Hu, A. Kuh, T. Yang, and A. Kavcic, "A belief propagation based power distribution system state estimator," *IEEE Computational Intelligence Magazine*, vol. 6, no. 3, pp. 36–46, Aug 2011.

[9] P. Chavali and A. Nehorai, "Distributed power system state estimation using factor graphs," *IEEE Transactions on Signal Processing*, vol. 63, no. 11, pp. 2864–2876, June 2015.

[10] L. Schenato, G. Barchi, D. Macii, R. Arghandeh, K. Poolla, and A. V. Meier, "Bayesian linear state estimation using smart metersand pmus measurements in distribution grids," *IEEE International Conference on Smart Grid Communications*, 2014.

[11] K. Emami, T. Fernando, H. H.-C. Iu, H. Trinh, and K. P. Wong, "Particle filter approach to dynamic stateestimation of generators in power systems," *IEEE Transactions on Power Systems*, 2015.

[12] P. A. Pegoraro, A. Angioni, M. Pau, A. Monti, C. Muscas, F. Ponci, and S. Sulis, "Bayesian approach for distribution system stateestimation with non-gaussian uncertainty models," *IEEE Transactions on Instrumentation and Measurement*, 2017.

[13] A. M. Kettner and M. Paolone, "Sequential discrete kalman filter for real-timestate estimation in power distributionsystems: Theory and implementation," *IEEE Transactions on Instrumentation and Measurement*, 2017.

[14] T. Nakagawa, Y. Hayashi, and S. Iwamoto, "Neural network application to state estimation computation," in *Proceedings of the First International Forum on Applications of Neural Networks to Power Systems*, Jul 1991, pp. 188–192.

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

[16] P. Barbeiroa, H. Teixeira, J. Krstulovica, J. Pereiraa, and F. J. Soares, "Exploiting autoencoders for three-phase state estimation in unbalanced distributions grids," *Electric Power Systems Research*, 2015.

[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] L. Deng, G. Hinton, and B. Kingsbury, "New types of deep neural network learning for speech recognition and related applications: an overview," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 8599–8603.

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

[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[21] L. Prechelt, *Neural Networks: Tricks of the Trade*. Springer, 2012.

[22] 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, no. 1, pp. 12–19, Feb 2011.