PERFORMANCE EVALUATION OF AN AUTOENCODER STATE ESTIMATOR WITH REALISTIC LOW VOLTAGE GRIDS RECONSTRUCTED FROM OPEN DATA

Elio EL SEMAAN[1,2,3], Dat TIEN NGUYEN[1,2], Alessio IOVINE[2], Trung Dung LE[1], Philippe DESSANTE[1], Dominique CROTEAU[3], Keddy KAMGA[3], Mouna RIFI[3]

[1] GeePs, CentraleSupélec, CNRS, Université Paris-Saclay, France

[2] L2S, CentraleSupélec, CNRS, Université Paris-Saclay, France

[3] EDF R&D, SYSTEME department, France

elio.el-semaan@centralesupelec.fr

ABSTRACT

Distribution systems operators' (DSO) upcoming challenge is to monitor and control low voltage (LV) grids. Real time measurements' acquisition from LV grids has become possible thanks to the deployment of smart meters (SM) but it is still limited by technical constraints. Thus, a real time state estimator (SE) is needed to ensure the LV grid's observability. A machine learning model can be trained using the SM historical data and then run in real time using few available measurements to estimate the system's state. For the purpose of evaluating the performance of a machine learning based SE on a realistic LV grid, an auto-encoder based SE is presented in this paper. An accuracy enhancement using pseudo-measurements is proposed and a method for reconstructing realistic LV grid models is also described.

1. INTRODUCTION

Low voltage (LV) distribution grids are facing a major transition due to the high penetration of decentralized energy production and the introduction of new types of consumers such as electric vehicles and heat pumps. In order to deal with this new bidirectional use of LV grids and these new types of consumptions, distribution systems operators (DSO) target to use a real-time state estimator (SE) to evaluate the distribution system's current state (voltage magnitude notably).

In the recent literature, Weighted Least Square (WLS) based SE has been proposed to perform state estimation on LV grids [1]. Even if this technique is already successfully applied to high voltage and medium voltage (MV) grids [2], its application to LV grids faces many challenges [3]. Among them, the most complex to handle are the unbalance of LV grids and the lack of real-time transmitted measurements due to the communication constraints limiting the usage of smart meter (SM) data.

However, the availability of historical data provided by the deployment of SM opens to the possibility to train machine learning based SE and then execute the trained model in real time, even in case of a limited number of SM measurements. Several techniques are investigated in the literature: for example, feed-forward neural network, linear regression and support vector machine techniques [4], neural network based SE [5] and auto-encoder with particle swarm optimization (AE-PSO) based SE [6], [7].

Among the several methodologies, the AE-PSO based SE showed great flexibility in dealing with different missing measurements situations, which is often the case in LV grids, while still providing accurate results. In the present paper, we aim to increase the accuracy of an AE-PSO based SE in case of less available real-time measurements by adding pseudo-measurements. To the best of authors' knowledge, no work is present in the literature of AE-PSO based SE for LV grids that consider pseudo-measurements, and the main contribution of this paper lies in this research line. The methodology is tested on a realistic LV grid with realistic grid unbalance conditions.

Another contribution of the present paper concerns the method introduced to reconstruct realistic LV grids model while taking into account grids' topology and sizing, to be used for testing and verifying the performance of the considered SE. The tool developed on the introduced technique is based on open data, and allows to meet the growing need for data with the increasing tendency towards smart grids and artificial intelligence applications.

The remainder of this paper is organized as follows. In section 2, we recall the AE-PSO algorithm that is used to include pseudo-measurements. In section 3, we detail the method of reconstruction of LV grids models based on DSO's open data. Section 4 presents a case study on a reconstructed grid model and state estimation's results are discussed. And finally, conclusion is made in section 5.

2. AE-PSO STATE ESTIMATOR

Auto-encoder

The auto-encoder is an unsupervised machine learning model using a symmetrical neural network (NN) aiming at reconstructing the initial input values using an encoder-decoder process [6], [7].

The encoder is a fully connected NN aiming to compress the input vector x into a d-dimensional reduced latent space (Figure 1). By noting the compressed vector of the input, f the mapping function, w, b and σ respectively the weights, bias and activation functions of the encoder's NN, x_d is written as follows:

$$x_d = f_{w,B}(x) = \sigma(wx + b) \qquad (1)$$

The decoder is another fully connected NN aiming to map back from the vector x_d to an output vector x', that has same dimension as the input vector and is expected to be as similar as possible to the input vector (Figure 1). By noting f' the mapping function, w', b' and σ' respectively the weights, bias and activation functions of decoder's NN, x' is written as follows:

$$x' = f_{w',B'}(x_d) = \sigma'(w' x_d + b') \qquad (2)$$

To accomplish this task, the optimal parameters $\{w_{opt}, b_{opt}, w'_{opt}, b'_{opt}\}$ are learnt iteratively using a training dataset $\{x_i\}_{i=1}^{N}$, composed of N input vectors. The loss function J to minimize during the training process is:

$$\{\theta, \theta'\} = \min_{\theta,\theta'} \frac{1}{N} \sum_{i=1}^{N} J(x_i, x_i') \qquad (3)$$

where $\theta = \{w, b\}$ and $\theta' = \{w', b'\}$ are the trainable parameters of the encoder and decoder, respectively.

It should be noted that the dimension d of the latent space and the number of the encoder's (and the decoder's) hidden layers n_{hidden} are hyperparameters to be tuned. They depend on the dimension of the input vector, i.e., grid's size. In fact, the complexity of the model can be extended by increasing the number of hidden layers. Given a tuned option for the hyperparameters d and n_{hidden}, the number of neurons for each hidden layer is calculated in a way that they decrease linearly for the encoder and then increase linearly for the decoder (Figure 1).

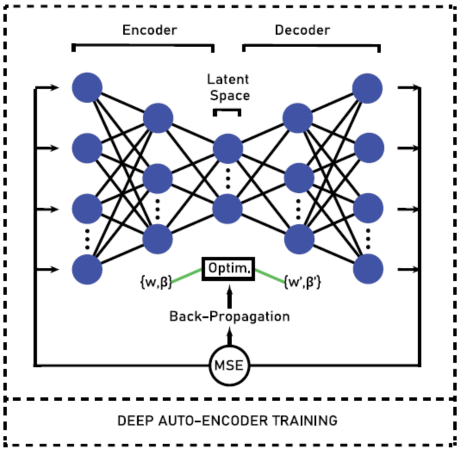

Figure 1. Auto-encoder training

For state estimation application, the learning dataset consists of a history of measurements of voltage magnitude, active power and reactive power of all nodes and all phases. Training the auto-encoder as described above leads to a model able to reconstruct all these parameters in the output vector.

A sample of the learning dataset is noted by x_i and can be written as follows:

$$x_i = [X_1^a, X_1^b, X_1^c, \ldots, X_n^a, X_n^b, X_n^c]_i \qquad (4)$$

where $i = \{1, \ldots N\}$, N being the number of samples, $X_j^{\varphi} = [P_j^{\varphi}, Q_j^{\varphi}, V_j^{\varphi}]$ the vector of the three measurements of a phase $\varphi \in \{a, b, c\}$ of a node $j \in \{1, \ldots n\}, n$ being the number of grid's nodes.

Particle Swarm Optimization

Once the training phase is over and we move to the real-time application, the amount of real-time measurements will be limited due to technical constraints. Thus, the input vector will present missing data, e.g., non-transmitted measurements of certain nodes. Among these measurements, there are the voltage magnitudes that are our final estimation goal.

The real time input vector x can be thus written as follows:

$$x = [x_{measured}, x_{missing}] \qquad (5)$$

A particle swarm optimization (PSO) technique is used to search for the values of the missing variables by using the pre-trained auto-encoder denoted by $AE_{trained}$ in the objective function as follows:

$$\min_{x_{missing}} MSE(x - AE_{trained}(x)) \qquad (6)$$

where MSE is the mean squared error.

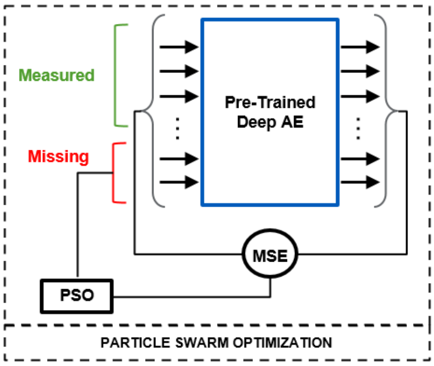

Figure 2. Particle Swarm Optimization for real time state estimation

Pseudo-measurements

When a phase φ of a node j is not measured, we suggest using pseudo-measurements for active and reactive power such that the only missing variables to be estimated would be the voltage magnitudes. The pseudo-measurements are generally derived from historical data or standard load profiles and present an error around 50% [8].

3. LV GRID RECONSTRUCTION

For an industrial application of a machine learning SE, the data needed to train the model is provided by the historical data of SM. However, the access to these data for research purposes is more complex. Thus, the creation of a database made up of simulated grids models is necessary.

In this section, we present a method consisting of different steps to reconstruct LV grids models from open data. We refer to "LV grid" as the set of transformers, lines and loads associated to a MV/LV substation. The developed tool is based on the distribution network geographical data available on Enedis' (the main French DSO) open data website [9].

The dataset in [9] provides the geographical positions (latitude and longitude) of the MV/LV substations, the LV

overhead lines and the LV underground lines. No information on the topology (connections between lines, attribution of lines to the MV/LV substations…), or electrotechnical dimensioning information (transformers' power, lines' sections, ...) are given.
Consequently, the proposed method consists in completing this missing information according to the sequential steps described below, while respecting the dimensioning rules of LV networks. It allows for the development of an automatic tool.

Topology identification

Given the geographical coordinates of the MV/LV substations and LV overhead and underground lines, a distance calculation between each two elements allows to identify the connections between elements.

Transformer sizing

Once the network topology is identified in the first step, the number of feeders is known and can be used to determine the rated power of the transformer according to the DSO's technical note [10], which indicates the maximum number of feeders for each transformer's rated power.

Load distribution

The transformer rated power fixed in the second step will be used to distribute the load powers to the different feeders, nodes and phases. However, since LV loads have very different load profiles, the consumption peaks of the different loads are spread over the day. A load diversity factor is then necessary to load the transformer properly.
The NF C14-100 standard specifies the diversity factor to be considered according to the number of users located downstream of the section considered.
Since the number of connected users increases with the transformer's rated power, we consider a diversity factor for each transformer rated power according to Table I.

Transformer rated power (kVA)	Diversity factor
50 - 100	0.63
160 - 250	0.44
400 – 630 - 1000	0.41

Table I. Diversity factor per transformer's rated power

By dividing the transformer rated power by the corresponding diversity factor, we get the total power of loads that can be connected to the transformer. This power is distributed in a random quasi-balanced way to the feeders, then to the nodes and finally to the phases. Finally, each phase of each node will be a random combination of LV clients.

Assigning load profiles

Once the clients' rated powers are set in the previous step, different load profiles should be assigned for each client in order to represent the individual behavior.
Enedis' open data provides an average profile coefficient for residential consumers [11].
A first approach is to add random noise to the average profile in order to create different profiles corresponding to different consumers. This approach leads to a synchronous behavior of all customers, thus eliminating the effect of the previously mentioned diversity factor.
A second considered approach consists in using available open-source data of Irish residential electricity consumption [12]. Since these data are real, they are a more realistic representation of individual behavior than the first approach.
After verifying the similarity between the Irish and French electricity consumption on a daily and annual basis, the Irish load profiles are selected to be randomly assigned to the consumers in our grid model generator tool.

Cables sizing

The section of a cable is determined according to the power of loads connected downstream the cable. In this step, for each cable, we add up the powers connected downstream and refer to the French DSO's technical note which specifies a cable section according to the power connected and the type of the cable (overhead or underground).

4. CASE STUDY AND RESULTS

Generated test grid

With the aim of testing the developed SE on a realistic grid, we selected a grid generated by our pre-described open data-based LV grid reconstruction tool. This grid has a 630 kVA substation and is composed of 25 nodes including the substation's node (node 1). It should be noted that since our tool is based on respecting grid's sizing rules, we ran it in an overloading mode to get significant voltage drops so the developed SE faces challenging situations. In order to consider distributed energy resources (DERs), we randomly distributed 5 photovoltaic (PVs) production per phase, of 30 kVA each.
The geographical topology of the considered grid is shown in Figure 3, while a schematic representation is given in Figure 4.

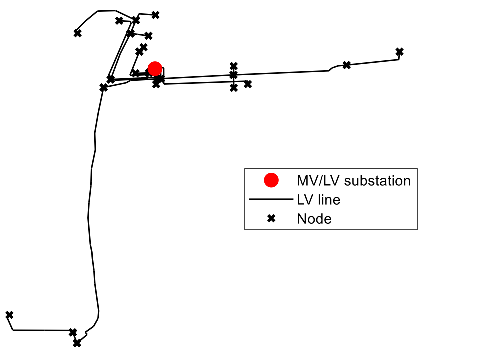

Figure 3. Geographical topology of the 25 nodes grid

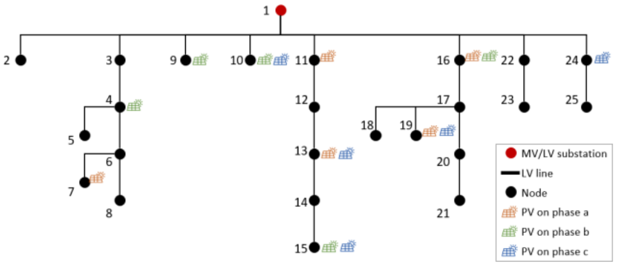

Figure 4. Representation of the topology of the 25 nodes grid

We consider 2000 samples with a time step of 30 minutes. Figure 5 shows voltage magnitude levels and voltage unbalance rate at each node. Clearly, we can notice a particular behavior for nodes 13, 14 and 15 in terms of voltage margin and unbalance rate. These are the farthest three nodes from the substation that are located in the lower part of Figure 3, which explains their critical conditions. We will refer to these nodes as "constrained nodes". We highlight that all the nodes present voltage values within the regulation limits and a realistic unbalance rate that remains below 2%.

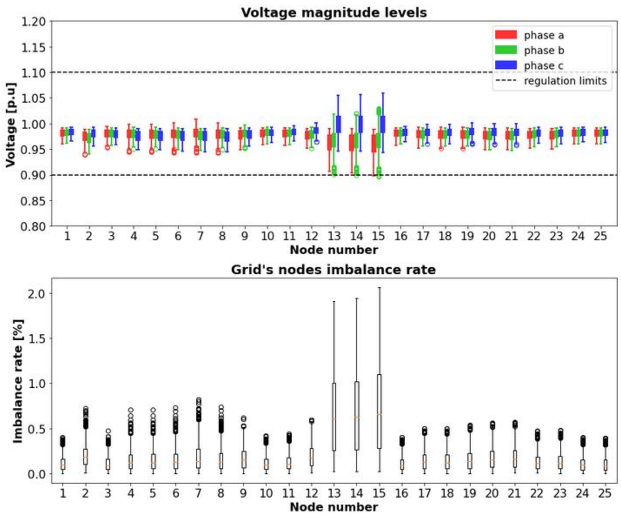

Figure 5. Voltage magnitudes levels and unbalance rate of the 25 nodes grid

Test description

According to standard procedures, we split our initial dataset into a training dataset and a test dataset. The training dataset is used to train the auto-encoder to reconstruct the input data as described above. The test dataset is used to test the AE-PSO state estimator under different cases of missing data.

We target an estimation error of 2% as desired performance for the SE of the LV grids.

Knowing that the voltage margin not to be exceeded is ±5% for MV grids against ±10% for LV grids, and that a 1% estimation error is acceptable for MV grids, the choice of having an error about 2% for LV is coherent.

When a node is not measured, then the voltage magnitude, active power, and reactive power of the three phases are all missing. We incrementally increased the ratio of missing measurements from 4% (1 node missing) to 96% (24 nodes missing) while always considering the node 1 that corresponds to the MV/LV substation as measured.

For each ratio of missing measurements, a group of missing nodes is randomly selected, and then the input data of the AE-PSO is constructed according to two case studies:

- no information for the missing measurements (red plot of Figure 6),
- pseudo-measurements for the missing active and reactive power measurements are considered, only the voltage magnitudes are missing (blue plot of figure 6).

Results

As shown in Figure 6, in the first case, where pseudo-measurements are not included, the SE performs well up to 64% of unmeasured nodes. When the ratio of missing measurements is further increased, the absolute estimation error increases significantly. In other words, without considering pseudo-measurements, a minimum of 36% of nodes should be measured and transmitted in real time in order to have satisfactory results. These observations are coherent with the literature.

In the second case, we investigate the effect of adding pseudo-measurements; we can note clearly a more stable performance against the increasing ratio of unmeasured notes. Actually, the estimation error remains within the acceptable margin up to 96% of missing measurements. This shows that an accurate state estimation is possible using only the substation's real measurements and pseudo-measurements for the remaining nodes.

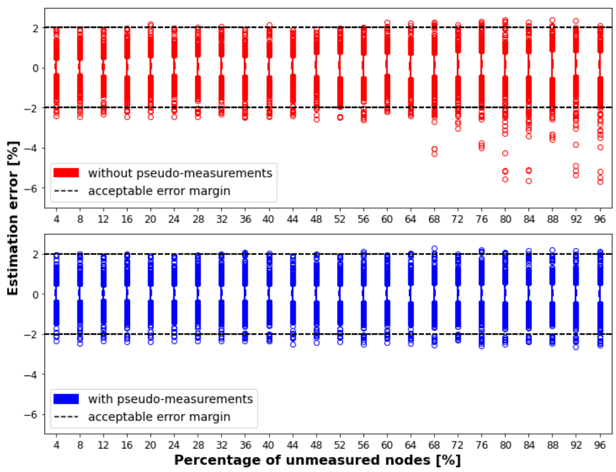

Figure 6. Pseudo-measurements' effect on estimation error with variable ratio of unmeasured nodes

Furthermore, we notice that a higher precision is obtained when estimating unconstrained nodes. It means that the acceptable error margin is always respected with two levels of accuracy: 1% for the unconstrained nodes and 2% for the constrained nodes, as shown in Figure 7.

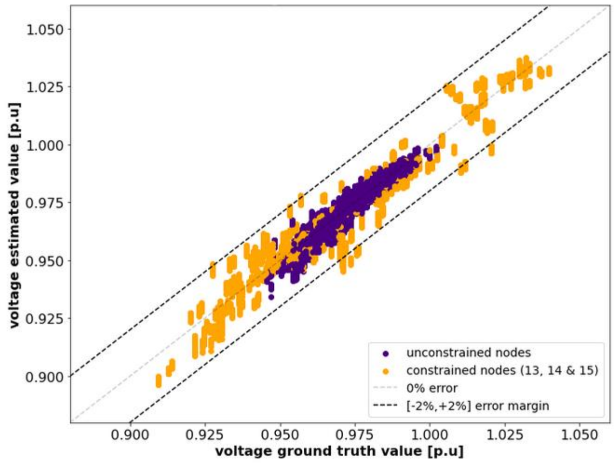

Figure 7. Voltage estimation scatter

Finally, we were interested in visualizing the estimation over 24 hours, i.e., 48 samples of an unconstrained node and a constraint node, e.g., nodes 7 and 13, respectively. The figure 8 shows that the global shape is well estimated in both cases, with a better estimation of voltage variation for the unconstrained node.

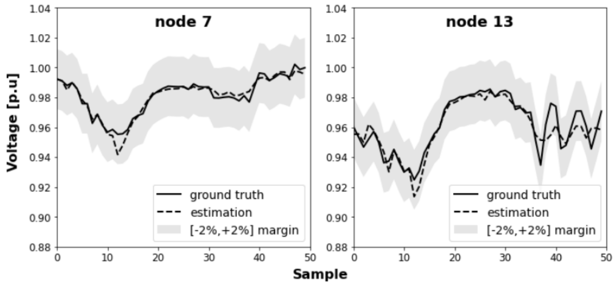

Figure 8. Voltage estimation of phase a of an unconstrained node (node 7) and a constrained node (node 13)

5. CONCLUSION

This paper evaluated the performance of an AE-PSO based SE on a realistic LV grid model when considering pseudo-measurements. Moreover, a method describing how to use available open data to reconstruct realistic LV grids models is presented. The resulting tool is of a great interest as it provides numerous realistic models and data that are necessary but often missing in the research field.

The AE-PSO was then tested on a realistic 25 nodes LV grid model with DER. The possibility to use pseudo-measurements allowed for a state estimation that needs only the real-time measurement from the MV/LV substation. Although several nodes present important voltage constraints and unbalance rates, the developed SE ensured an acceptable estimation with an error below 2%, with very high accuracy in case of unconstrained nodes, i.e., an error below 1%.

Future works aim to study the robustness of the developed model with respect to variations such as topology changes or the lack of knowledge about client's connection to phases.

Acknowledgments

This work has received partial funding from Institute for Sustainable Energy of the University Paris-Saclay.

REFERENCES

[1] F. Binot, T. D. Le, et M. Petit, "A Three-phase Four-wire State Estimator Algorithm for Low Voltage Network Management", Conference CIRED, June 2019.

[2] D. Croteau et O. Carre, "Distribution state estimation: outcomes from a field implementation aimed at tackling MV mastering in the presence of distributed energy resources (DER) ", CIRED - Open Access Proceedings Journal, vol. 2017, no 1, p. 1715-1717, 2017.

[3] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan and F. Bu, "A Survey on State Estimation Techniques and Challenges in Smart Distribution Systems," in IEEE Transactions on Smart Grid, vol. 10, no. 2, pp. 2312-2322, March 2019.

[4] G. Hong and Y. -S. Kim, "Supervised Learning Approach for State Estimation of Unmeasured Points of Distribution Network," in IEEE Access, vol. 8, pp. 113918-113931, 2020.

[5] A. S. Zamzam, X. Fu and N. D. Sidiropoulos, "Data-Driven Learning-Based Optimization for Distribution System State Estimation," in IEEE Transactions on Power Systems, vol. 34, no. 6, pp. 4796-4805, Nov. 2019.

[6] P. N. Pereira Barbeiro; H. Teixeira; Jorge Pereira; R. Bessa, "An ELM-AE State Estimator for real-time monitoring in poorly characterized distribution networks", IEEE PowerTech Conference, Eindhoven, 2015.

[7] V. Miranda, J. Krstulovic, H. Keko, C. Moreira and J. Pereira, "Reconstructing Missing Data in State Estimation With Autoencoders," in IEEE Transactions on Power Systems, vol. 27, no. 2, pp. 604-611, May 2012.

[8] A. Angioni, T. Schlösser, F. Ponci and A. Monti, "Impact of Pseudo-Measurements From New Power Profiles on State Estimation in Low-Voltage Grids," in IEEE Transactions on Instrumentation and Measurement, vol. 65, no. 1, pp. 70-77, Jan. 2016.

[9] https://data.enedis.fr/pages/cartographie-des-reseaux-contenu/ (access: 28 November 2022).

[10] https://www.enedis.fr/media/2168/download (French) (access: 28 November 2022).

[11] https://data.enedis.fr/pages/coefficients-des-profils-dynamiques-contenu/ (access: 28 November 2022).

[12] Commission for Energy Regulation (CER). (2012). *CER Smart Metering Project - Electricity Customer Behaviour Trial, 2009-2010* [dataset]. 1st Edition. Irish Social Science Data Archive. SN: 0012-00. https://www.ucd.ie/issda/data/commissionforenergyregulationcer/