

Generative neural networks for providing pseudo-measurements in electric power distribution systems.

Luiz Phillip Quintanilha da Silva   [Fluminense Federal University | luizphillip@id.uff.br]

Julio Cesar Stacchini de Souza  [Fluminense Federal University | julio@ic.uff.br]

Milton Brown Do Coutto Filho  [Fluminense Federal University | mbrown@ic.uff.br]

 Institute of Computing, Universidade Federal Fluminense, Av. Gal. Milton Tavares de Souza, s/n, São Domingos, Niterói, RJ, 24210-590, Brazil.

Received: 19 March 2023 • Accepted: 09 November 2023 • Published: 12 July 2024

Abstract The success of automation and control functions envisioned for smart distribution networks depends on reliable real-time network supervision. This task is performed by the distribution state estimator, responsible for processing a set of measurements received by the supervisory control and data acquisition (SCADA) system. In smart grids, the advanced measurement infrastructure (AMI) allows to collect regular readings of consumer voltage and power measurements—this can complement the few measurements (coming from the SCADA system) usually available for monitoring the distribution network and benefit the state estimation process. However, due to communication bottlenecks, such measurements are available only on an hourly basis. In order to circumvent the lack of real-time measurements this paper investigates the application of different neural network models—AutoEncoder, Contractive AutoEncoder, and Variational AutoEncoder—and proposes a methodology to generate AMI pseudo-measurements to complement SCADA measurements when only the latter are available for processing. Simulations performed with a 34-bus distribution system illustrate the proposed methodology, and the results obtained confirm its potential for pseudo-measurement provision.

Keywords: Smart Grids, Generative Models, Redundancy, Real-Time Monitoring, State Estimation

1 Introduction

The real-time monitoring of electric power distribution networks, aided by the state estimation (SE) implemented in distribution management systems (DMS), is becoming an increasingly important requirement. Lefebvre *et al.* [2014] articulate the SE essentiality for the success of modern automation/control functions by providing helpful information for controlling voltage profile, optimal feeder reconfiguration, and detection of islanding in networks with distributed generation, among others. However, for such monitoring to be effective, various challenges need to be overcome, many of which are different from those encountered in transmission systems due to characteristics such as:

- Few measurements are available for processing, and network observability is not achieved without pseudo-measurements [Mestav *et al.*, 2019];
- The presence of imbalances in the network and loads of distribution systems;
- Relationships between resistances and reactances are distinct from those found in transmission systems, especially in underground distribution networks;
- Communication systems have limitations for transmitting measurements provided by smart meters.

Pseudo-measurements are artificial values obtained, for instance, through system operation historical data, load curve assessment, and state/measurement forecasting models, generally accepted as actual measurements when insufficient data feed SE, i.e., the available measurements do not allow

network observability. Pseudo-measurements typically have larger errors than real-time measurements.

The smart distribution network paradigm has leveraged advances in measurement systems and communication infrastructure, allowing the acquisition and storage of large data volumes through digital relays, phasor measurement units (PMUs), intelligent electronic devices (IEDs), and smart meters.

Through the advanced metering infrastructure (AMI), it is possible to make available for processing measured quantities that contribute to improving the quality of the SE process. Despite efforts to develop adequate communication infrastructure, the data transmission capacity has limited the progress of advanced analysis, automation, and control functions, particularly those related to real-time operation, SE included [Li *et al.*, 2011; Gaspar *et al.*, 2023]. Therefore, most of the time the DMS available measurements are still insufficient for complete observability and reliable real-time network monitoring. In this case, the use of pseudo-measurements presents itself as an alternative to supplement the data to be processed. Due to the usual unobservability of distribution grids, alternative solutions have been analyzed to mitigate this condition.

From the methodological perspective, the methods for reconstructing missing data can be classified into two broad groups: statistical analysis-based and machine learning-based. The former is prevalent (e.g., regression/mean imputation), being relatively straightforward but somewhat impractical to face situations of high dimensionality/loss rates; reconstructed data accuracy is weak. The latter uses a learning

process to reconstruct effectively missing values, overcoming the limitations of statistical analysis-based methods.

In recent years, with the remarkable development of artificial intelligence, scholars have increasingly concentrated on its application in data recovery. An autoencoder-based approach to providing data in the context of a DMS was proposed in Miranda *et al.* [2011], focused on recovering information related to network topology. When properly trained, autoencoders perform well in recomposing missing measurements (the input pattern coherent with the real system produces a similar output with a negligible error). Dehghanpour *et al.* [2019] present how relevance vector machine (RVM) models can be used, in conjunction with game theory, to generate both pseudo-measurements and their uncertainties. Dahale *et al.* [2020] show how linear power flow and linear programming can be combined to obtain the system state. On the other hand, Mestav *et al.* [2019] present Monte Carlo simulations and deep neural network models to achieve better results regarding measurement error detection and the SE overall performance.

This paper presents a methodology for providing pseudo-measurements to complement, whenever necessary, SCADA measurements. For this purpose, three neural network models—namely AutoEncoders, Contractive AutoEncoder, and Variational AutoEncoder—are tested, and their performance is assessed regarding the provision of pseudo-measurements. The database adopted is formed with AMI and SCADA measurements. The generated pseudo-measurements allow for an increase in the frequency of instances where a redundant data set is available for processing, ensuring observability and supervision of the network. Simulations conducted with a 34-bus distribution system illustrate the application of the proposed methodology, and the results obtained show the potential of the models adopted for generating high-quality pseudo-measurements for SE.

2 AutoEncoder

AutoEncoders (AE) are a class of unsupervised deep learning models composed of two structures: an encoder and a decoder [Liu *et al.*, 2017]. The encoding structure is able to extract relevant features from input data [Liu *et al.*, 2017; Hinton and Zemel, 1994]. The decoder is responsible for reconstructing the input data from the extracted features [Liu *et al.*, 2017]. This is one of the most popular general-purpose models among those in which both structures (encoder and decoder) are formed by neural networks [Jabbar *et al.*, 2021].

Neural network models can adapt to a dataset through a cost function, which aims to measure how far/close the model is to reproduce the dataset faithfully. In the case of AEs, the most commonly used cost function is the average of the Euclidean distances between the expected output and the reconstruction produced by the model [Jabbar *et al.*, 2021], according to:

$$L_{AE} = \frac{1}{n} \sum_{i=0}^n \|x - f_{\theta}(g_{\theta}(x))\| \quad (1)$$

where f_{θ} represents the decoder function and g_{θ} the encoder function.

Each of these functions ($f(\theta)$ and $g(\theta)$) represents a different model structure. The number of layers and neurons depends on the problem and should be defined by the user. After applying g_{θ} to the input data, one obtains a new representation of the input data (Z), known as the latent space.

2.1 Contractive and Variational AutoEncoder

The AutoEncoders' traditional objective function has limitations regarding the encoding function rate of variation and the sampling process from the model's hidden dimension. Two classes of models have been developed to address these needs—Contractive (CAE) and Variational AutoEncoder (VAE). Both models modify the typical structure of the AE and add a regularization term to the cost function.

The Jacobian matrix Frobenius norm of the encoding process is added in the Contractive model. In the Variational model, the Kullback-Leibler divergence is included. Additionally, the latter model has a sampling process in the encoding stage. All of these models can be adopted in various categories of problems, such as in anomaly detection [Zhang *et al.*, 2022; Ko *et al.*, 2021], computer vision for information retrieval and noise removal in images [Yilmaz *et al.*, 2022; Xu *et al.*, 2022; He *et al.*, 2022; Vijayalakshmi and Shanthakumar, 2019], where the model receives a set of inputs in which there are faults/missing information in the images (the goal is to recreate/remove noise from them). The reconstruction process can be used for generating pseudo-measurements when there is little information loss (about 20%) and also for reconstructing the topology of the electrical system [Miranda *et al.*, 2011; Krstulovic *et al.*, 2013].

3 Bayesian Optimization

Evaluating the cost functions of an optimization process is not always easy. Many times, the functions are so complex that assessing a single value can be extremely costly [Bergstra *et al.*, 2011]. Given this scenario, Sequential Model-Based Bayesian Optimization (SMBO) models are valuable allies for solving such problems ([Bergstra *et al.*, 2011]. This method of optimizing cost functions depends on two key factors. The first is to obtain a probabilistic model that, given a set of parameters, can predict the possible outcome of the cost function [Candelieri *et al.*, 2020]. The second factor is a criterion for deciding what should be the next value of the parameters to be tested [Candelieri *et al.*, 2020]. SMBO optimization works as follows:

1. A random set of parameters is tested;
2. Based on the tested parameters and the obtained response when evaluating the cost function, a model is trained to predict the value of the cost function for a parameter X ;
3. A new set of parameters is selected for testing in order to optimize the established criterion;
4. Return to step 2 until the established stopping criterion is reached;

The criterion established for SMBO optimization is determined by the user. However, the case of Expected Im-

provement (EI) is one of the most commonly used criteria [Bergstra *et al.*, 2011]. This criterion is presented in Equation(2):

$$EI(x) := \int_{-\infty}^{y^*} (y^* - y)p(y|x)dy \quad (2)$$

where y^* is an established threshold (usually based on the performance achieved up to the current iteration), and p is the outcome of the probabilistic model.

The most commonly used model with SMBO is Gaussian Process Regressors (GPR) [Candelieri *et al.*, 2020]. However, this type of model has some issues associated with categorical parameters and parameter dimensionality [Candelieri *et al.*, 2020; Bergstra *et al.*, 2011]. Given this scenario, the Tree-Structured Parzen Estimator (TPE) emerges as an excellent option. In the case of TPE, the change made is in the probabilistic model. Instead of using GPR, the Parzen Window method is used to obtain probability distributions [Bergstra *et al.*, 2011]. Furthermore, as there is no prediction model, two probability distributions are fitted: the first with the parameters that have achieved the best performance with respect to the cost function up to the current iteration, and the second with the remaining parameters [Bergstra *et al.*, 2011]. The remaining process remains the same as established for SMBO, with only the necessary adjustments made in Equation (2) for the case with two probability distributions.

4 Proposed Methodology

The AMI in smart grids enables regular readings of voltage and power measurements from consumers, which can benefit network observability and the SE process as a whole. Measurements from smart meters at the low-voltage level are generally recorded every 15 minutes or at larger time intervals. However, due to limitations in the communication system, AMI measurements are transmitted only on an hourly basis. On the other hand, the time interval for acquiring SCADA measurements is a few seconds. Therefore, over long time intervals, only SCADA measurements are available, which is insufficient to ensure system observability. Although AMI measurements may be abundant, their integration for processing with SCADA measurements requires an appropriate strategy [Huang *et al.*, 2015; Zhao *et al.*, 2020].

The proposed methodology establishes that pseudo-measurements complement the available SCADA measurements to achieve network observability while AMI measurements are absent. Based on the time instances in which AMI and SCADA measurements are available, a database is constructed, serving as a training set for the three neural network models described in Section 2. During the training phase, each model captures the relationships between SCADA and AMI measurements. After that, the idea is to reconstruct AMI measurements when they are unavailable in the DMS, based on the observed set of SCADA measurements. Thus, one can increase the frequency at which a redundant data set is available for processing by SE.

This method is divided into two stages: (i) a model construction (offline), with higher computational cost, and (ii) pseudo-measurements provision (online), which can be per-

formed in the real-time environment. The following subsections present each of these stages.

4.1 Model Construction Stage

The dataset used at this stage contains historical records of both SCADA and AMI measurements, observed when those measurements are simultaneously available at the DMS during system operation (hourly). The model should be trained to reconstruct all measurements (SCADA and AMI) as output data having only SCADA measurements available at the input layer. In such layer, the inputs corresponding to AMI measurements are set to zero. The idea is to provide the model with capacity to reconstruct missing data (AMI measurements) in the time intervals they are not available.

Several parameters can be modified to obtain the best possible fit for the available data when building a model. In addition, depending on the model, the user should adjust a series of hyperparameters, affecting the entire learning process. Therefore, it is necessary to carefully adjust the hyperparameters to extract the models' maximum capacity.

In the offline stage, depicted in Figure (1), the hyperparameters capable of obtaining the best performance for each model type are selected. Initially, it is necessary to preset the number of tests to be performed and the probability distribution of the hyperparameters. After this definition, the models are submitted to the training and validation datasets. Using an adequate metric, the model that presented the best responses to the validation dataset is selected. From the perspective of this work, such model is the most suitable to reconstruct the missing AMI measurements in a DMS.

4.2 Obtaining Pseudo-measurements

The training phase deals with determining the model that brings more benefits to the distribution system monitoring through pseudo-measurement provision. The selected model can then be used in real-time to reconstruct AMI measurements. In Figure (2), the system is considered observable at the gray time instants since both SCADA and AMI measurements are available. However, in the other time instants, the DMS receives only the SCADA measurements, and executing SE is impossible due to the lack of observability. In such situations, the selected model should reconstruct the AMI measurements, which serve as pseudo-measurements that complement the SCADA ones and guarantee system observability. In such scenario, SE and other DMS advanced functions that depend on observability can be executed. The pseudo-measurements are obtained every five minutes, widely accepted as an adequate time interval to track and monitor system operation condition. When both SCADA and AMI measurements are available (hourly) it is not necessary to generate the pseudo-measurements.

Figure (3) illustrates the process of measurement recovery, showing the model operation when there are no available AMI measurements. As in the model construction stage, the input vector contains the SCADA measurements that are observed in real time, with the inputs corresponding to AMI measurements being set to zero. Using this input dataset, the model performs the encoding and decoding processes. At the

Figure 1. Model Construction Stage.



Figure 2. Example of input/output availability

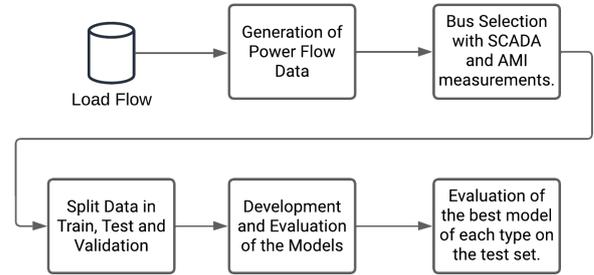


Figure 4. Flowchart of the employed methodology.

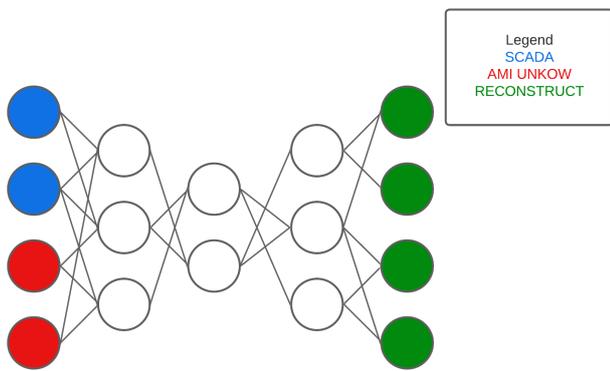


Figure 3. Measurement recovery.

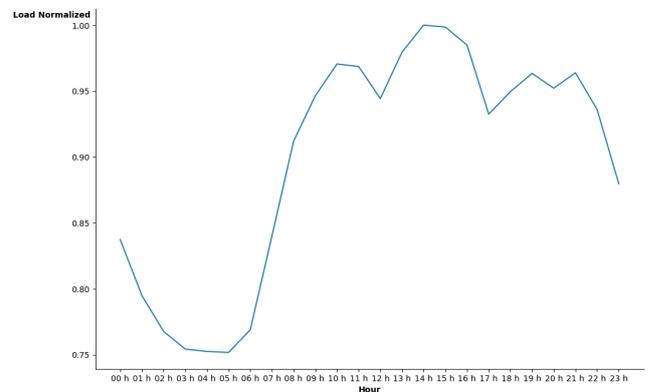


Figure 5. ONS Load Flow Normalized

output layer the reconstructed AMI measurements are the outputs of interest, to be used as pseudo-measurements. Such pseudo-measurements complement the real-time SCADA measurements and guarantee system observability.

5 Simulations and Results

The methodology proposed in this paper is constructed from a set of measurements used to monitor a distribution system. After the construction and selection stages, the best model is tested and its performance is evaluated using unseen data. Tests were carried out on a 34-bus distribution network Soares *et al.* [2019], for which training and validation data were generated. These data correspond to measured quantities that are obtained through a power flow program and the subsequent addition of a random disturbance. The Figure (4) presents a flowchart with the data generation, model development, and validation steps. The following subsections explain how each of these steps was developed.

5.1 DataBase Generation

The data used in the tests were generated from simulations using power flow for three-phase distribution networks, considering imbalances in both network topology and system loads. In order to encompass a wider range of operating conditions, it was assumed that both the power factor and the participation of the buses in the system consumption could vary by $\pm 5\%$ of the original values. This resulted in the generation of a larger amount of data, with greater diversity, representing

more realistic situations to be considered by the model in the training and testing stages.

For the simulation of power flow a load flow available in the webpage of Brazilian National Operator [ONS - Operador Nacional do Sistema, 2023], was utilized. It represents the intra-day consumption of the Brazilian Power System (Sistema Interligado Nacional - SIN). The use of this database was carried out as follows:

1. Data normalization by the maximum value;
2. Linear interpolation for the discretization of the data in smaller 5-minute intervals;
3. Multiplication the original load value of each bar by the value obtained in the previous step;

Figure (5) show the load flow available in ONS webpage.

After executing the power flow, it was necessary to transform the exact data generated (true values) into noisy measurements. It was assumed that the measurements are random variables, normally distributed, with a mean equal to the value obtained in the power flow and standard deviation represented as $\sigma = \frac{z_{flow}}{100}$ for power and current measurements, or $\sigma = \frac{z_{flow}}{300}$ for voltage measurements. It is worth noting that z_{flow} is the value obtained when executing the power flow.

It was considered that there are SCADA measurements (active/reactive power flows) in branches incident to 10 of the network buses. It is worth noting that observability cannot be guaranteed with only such SCADA measurements, and the provision of additional data is necessary. Buses with-

out SCADA data are those with load connected and have AMI measurements: bus voltages and active/reactive power injections. As mentioned earlier, AMI measurements are not available for processing with the same frequency as SCADA measurements, and when unavailable, they must be generated using the proposed methodology.

Once a database with SCADA and AMI measurements available for different time intervals is generated, the next step is to split the database into training, testing, and validation sets. The training and validation sets are for model training and selection, while the testing set is used to assess the overall performance of the model selected by the previous sets. It is worth noting that those sets were separated according to the following criterion: 64% of the data generated for training, 16% for validation, and 20% for the final testing. In addition, the model constantly receives 94 SCADA measurements, being responsible for reconstructing 165 AMI measurements, representing the provision of more than 63% of the total number of measurements to the DMS.

5.2 Off-line Stage

The selection of the model structure was performed in this stage, with several tests being carried out by varying the number of hidden layers, the number of neurons in each layer, the dropout rate for each hidden layer of the network, and the learning rate. The objective was to test different hyperparameter combinations and test the model with the lowest reconstruction error on the validation data. Table (1) presents the range of the tested hyperparameters.

The selection of the best model can be conducted in several ways. As neural network models consume more computational time during training than other models, testing all possible architectures for AEs becomes impractical. For this reason, the framework of Akiba *et al.* [2019] was used to select the best architecture among a pre-selected set.

Table 1. range of tested hyperparameters

Hyperparameters	Minimum Value	Maximum Value
Quantity of Layers	3	12
Number of Neurons	15	240
Dropout Rate	0.01	0.5
Learning Rate	10^{-5}	10^{-1}

The TPE method for optimization of the hyperparameters [Bergstra *et al.*, 2011] was selected in the library presented by Akiba *et al.* [2019]. As a Bayesian method, not all values are tested and the region with the most significant contribution is selected. It should be noted that all hyperparameters were tested as log-uniform distributions. In addition, the activation function in the hidden layers was kept as GELU, while the activation function in the last layer was the Hyperbolic Tangent.

The best model selection was based on minimizing the mean absolute percentage error of the reconstruction according to Equation (3).

5.3 Performance Indicator

The mean absolute percentage error (MAPE) is computed by:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{x_i^{real} - x_i^{reconstructed}}{x_i^{real}} \quad (3)$$

where x_i^{real} is the value of the i -th measurement and $x_i^{reconstructed}$ is the reconstructed value.

A MAPE lower than 3% for power measurements or lower than 1% for voltage measurements are considered satisfactory since these values represent $\pm 3\sigma$ of such random variables.

MAPE was used in both stages of the proposed method. In the offline stage, this metric helps to select the best model hyperparameters. In the second stage, MAPE demonstrates how the model behaves with data not used during the training stage. Thus, the error made in the last stage can be understood as the expected error when applying the methodology in DMS. It is worth noting that this metric presents issues when real values are zero or very close to zero since MAPE tends to $\pm\infty$, although it does not necessarily mean that the model is inadequate.

5.4 Results

The developed method was applied to the 34-bus distribution system [Soares *et al.*, 2019]. Buses 1, 3, 6, 7, 10, 12, 16, 27, 30, and 34 were selected to contain the SCADA measurements, while the remaining buses have AMI measurements, which most of the time are not available in real-time for SE. The following subsections present the results of the architecture selection for each model and the performance of the measurement reconstruction.

5.4.1 Architecture Selection

In this stage, the framework presented by Akiba *et al.* [2019] was used with TPE as the optimizer to find the best possible architecture for the neural network. It is worth noting that the goal was to minimize the MAPE of the previously separated validation data.

For all models, the best result was obtained with 3 hidden layers and the learning rates were around 10^{-4} . The number of neurons in each layer and the dropout rate are presented in Table (2). It is worth noting that the layer with the lowest number of neurons was selected as the last layer of the encoding process.

Table 2. Structure selected by the optimization process

Model	Layer	N. of Neurons	Dropout Rate
AE	1	224	≈ 0.025
AE	2	67	≈ 0.019
AE	3	79	≈ 0.010
CAE	1	195	≈ 0.002
CAE	2	172	≈ 0.002
CAE	3	66	≈ 0.005
VAE	1	189	≈ 0.1801
VAE	2	190	≈ 0.036
VAE	3	194	≈ 0.013

5.4.2 Model Application

After selecting the architecture for each model, their application was simulated to evaluate their performance on the test dataset.

The test set comprises 20% of the total amount of generated data. This set was not used in any other stage of training. Therefore, its results are believed to represent the model’s behavior.

The statistics related to the models are shown in Table 3. The columns STD, Q1, Q3, P90, and P99 represent the standard deviation, first quartile, third quartile, 90th percentile, and 99th percentile, respectively. They refer to the absolute percentage error made by the model. The Mean column represents the value obtained by applying (3).

From Table 3, it is possible to observe some characteristics of the models:

- The voltage measurement is the most reliable in all models;
- The reconstruction with the Variational AutoEncoder was not satisfactory when compared to the other models;
- The Contractive AutoEncoder model shows a slight superiority over the AutoEncoder;
- The Reactive Power proves to be more difficult to predict, despite its low robust statistics (median and quartiles);

As the error values obtained for the CAE and AE models were similar, a Kolmogorov-Smirnov test was used to check for a significant difference in the error committed by the models. The test results are presented in Table (4).

With this result, one can conclude that the Contractive AutoEncoder model was the best for this dataset, possibly due to adding the Jacobian matrix Frobenius norm in the cost function, considering that the model reduces significant variations in the encoding function.

The values presented in Table 3 refer to the statistics of the percentage errors for the tested models. Larger values are observed for the errors in reactive power measurements and this is because some of those measurements present high variability and/or values very close to zero. These are common characteristics of reactive power quantities and not only observed in the dataset employed in this paper. The difficulty to predict reactive power is also reported in related works in the technical literature [de Souza *et al.*, 2017]. Reactive power measurements that are difficult to predict are associated with particular load consumption profiles that are known in advance. Regarding the proposed approach, such reconstructed reactive measurements can be discarded and the remaining ones will be sufficient to provide measurement redundancy that guarantees state estimation observability.

A relevant analysis regarding the model’s performance is to understand which measurements have the highest error. Thus, as there are more measurements than necessary to ensure observability, the end-user can remove these measurements from the process to guarantee SE reliability. Figures (6, 7, and 8) show the 5 measurements with the highest average errors obtained with the models. Such measurements correspond to reactive power flows and injections. One ob-

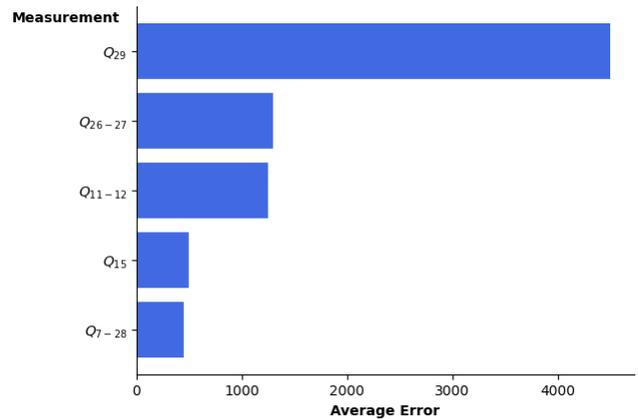


Figure 6. AE highest errors

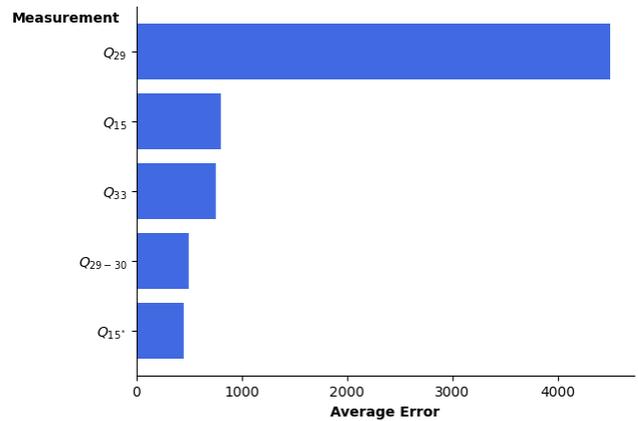


Figure 7. CAE highest errors

serves that the AE and CAE models have the reactive power injection at bus 29 (Q29) as the measurement with the highest error in the reconstruction. It means that the reactive power consumption at bus 29 is tricky to predict by these models with the available data. Therefore, it is possible to exclude the active/reactive power measurements from this bus, without compromising observability, i.e., being still able to perform the SE process.

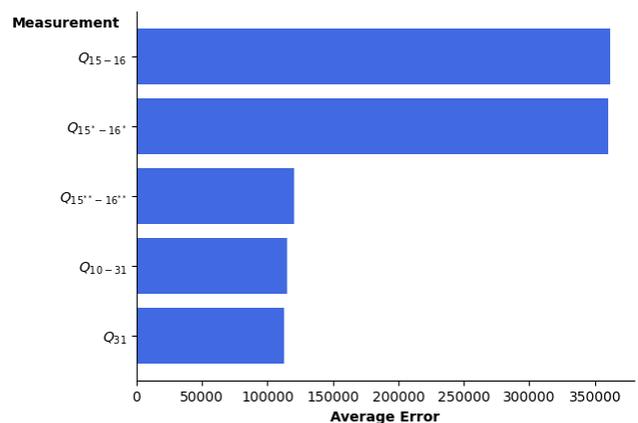


Figure 8. VAE highest errors Error

Table 3. Statistics of the mean absolute percentage error for each model

Model	Measure Type	Mean	STD	Q1	Median	Q3	P90	P99
AE	Voltage - AMI	0.005	0.007	0.001	0.003	0.006	0.011	0.035
AE	Active P. - AMI	2.142	1.693	0.777	1.810	3.170	4.419	7.049
AE	Active P. - SCADA	0.621	0.852	0.163	0.362	0.711	1.374	4.648
AE	Reactive P. - AMI	124.033	5073.754	1.383	3.800	7.515	11.651	28.182
AE	Reactive P. - SCADA	108.252	5518.151	0.208	0.460	0.879	1.604	9.816
CAE	Voltage - AMI	0.003	0.005	0.001	0.002	0.004	0.008	0.024
CAE	Active P. - AMI	2.147	1.695	0.768	1.844	3.199	4.399	7.019
CAE	Active P. - SCADA	0.557	0.852	0.130	0.293	0.609	1.266	4.542
CAE	Reactive P. - AMI	157.650	7219.111	1.358	3.921	7.652	11.719	28.073
CAE	Reactive P. - SCADA	88.451	6367.784	0.159	0.356	0.679	1.178	4.781
VAE	Voltage - AMI	0.097	0.117	0.022	0.057	0.127	0.231	0.560
VAE	Active P. - AMI	6.709	5.925	2.465	5.259	9.298	13.987	28.326
VAE	Active P. - SCADA	6.326	5.615	2.228	4.911	8.818	13.367	27.259
VAE	Reactive P. - AMI	5823	173535	3.541	7.516	12.953	19.459	4691.135
VAE	Reactive P. - SCADA	26882	704357	2.887	6.329	11.718	19.826	119162

Table 4. Kolmogorov-Smirnov Test

Measure Type	D	P-Value
Active P. - AMI	0.008233	$1.18 \cdot 10^{-34}$
Active P. - SCADA	0.078901	≈ 0.0
Active P. - AMI	0.014880	$3.06 \cdot 10^{-112}$
Active P. - SCADA	0.101475	≈ 0.0
Voltage - AMI	0.136935	≈ 0.0

6 Conclusion

Distribution networks suffer from a lack of available measurements. The use of advanced metering infrastructure becomes indispensable to guarantee network observability; although hindered by the low frequency at which measurements are made available for processing by the communication system. This paper compares three neural network models— AutoEncoder, Contractive AutoEncoder, and Variational AutoEncoder—aiming to generate pseudo-measurements to complement SCADA data whenever AMI measurements are unavailable. Such measurement provision can be obtained in real time by the neural network previously trained in an offline environment. Tests were performed on a 34-bus network, and the results showed that bus voltage and active branch power measurements were reconstructed with adequate quality, contrasting with reactive power measurement reconstruction. The proposed models were able to generate good pseudo-measurements that can be used for SE when only SCADA measurements are available. Among them, the Contractive AutoEncoder model presented the best results.

Declarations

Authors' Contributions

All authors contributed to the writing of this article, read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests

Availability of data and materials

Data can be made available upon request

References

- Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. DOI: 10.1145/3292500.3330701.
- Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for hyper-parameter optimization. In Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc. Available at: <https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf>.
- Candelieri, A., Galuzzi, B., Giordani, I., and Archetti, F. (2020). Learning optimal control of water distribution networks through sequential model-based optimization. In *International Conference on Learning and Intelligent Optimization*, pages 303–315. Springer. DOI: 10.1007/978-3-030-53552-0_28.
- Dahale, S., Karimi, H. S., Lai, K., and Natarajan, B. (2020). Sparsity based approaches for distribution grid state estimation—a comparative study. *IEEE Access*, 8:198317–198327. DOI: 10.1109/ACCESS.2020.3035378.
- de Souza, J. C. S., Assis, T. M. L., and Pal, B. C. (2017). Data compression in smart distribution systems via singular value decomposition. *IEEE Transactions on Smart Grid*, 8(1):275–284. DOI: 10.1109/TSG.2015.2456979.
- Dehghanpour, K., Yuan, Y., Wang, Z., and Bu, F. (2019). A game-theoretic data-driven approach for pseudo-measurement generation in distribution system state esti-

- mation. *IEEE Transactions on Smart Grid*, 10(6):5942–5951. DOI: 10.1109/TSG.2019.2893818.
- Gaspar, J., Cruz, T., Lam, C.-T., and Simões, P. (2023). Smart substation communications and cybersecurity: A comprehensive survey. *IEEE Communications Surveys & Tutorials*. DOI: 10.1109/COMST.2023.3305468.
- He, Z., Hong, K., Zhou, J., Liang, D., Wang, Y., and Liu, Q. (2022). Deep frequency-recurrent priors for inverse imaging reconstruction. *Signal Processing*, 190:108320. DOI: 10.1016/j.sigpro.2021.108320.
- Hinton, G. E. and Zemel, R. (1994). Autoencoders, minimum description length and helmholtz free energy. In Cowan, J., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems*, volume 6. Morgan-Kaufmann. Available at: <https://proceedings.neurips.cc/paper/1993/file/9e3cfc48eccf81a0d57663e129aef3cb-Paper.pdf>.
- Huang, S.-C., Lu, C.-N., and Lo, Y.-L. (2015). Evaluation of ami and scada data synergy for distribution feeder modeling. *IEEE Transactions on Smart Grid*, 6(4):1639–1647. DOI: 10.1109/TSG.2015.2408111.
- Jabbar, A., Li, X., and Omar, B. (2021). A survey on generative adversarial networks: Variants, applications, and training. *ACM Computing Surveys (CSUR)*, 54(8):1–49. DOI: 10.1145/3463475.
- Ko, J. U., Na, K., Oh, J.-S., Kim, J., and Youn, B. D. (2021). A new auto-encoder-based dynamic threshold to reduce false alarm rates for anomaly detection of steam turbines. *Expert Systems with Applications*, page 116094. DOI: 10.1016/j.eswa.2021.116094.
- Krstulovic, J., Miranda, V., Costa, A. J. S., and Pereira, J. (2013). Towards an auto-associative topology state estimator. *IEEE transactions on power systems*, 28(3):3311–3318. DOI: 10.1109/TPWRS.2012.2236656.
- Lefebvre, S., Prévost, J., and Lenoir, L. (2014). Distribution state estimation: A necessary requirement for the smart grid. In *2014 IEEE PES General Meeting | Conference Exposition*, pages 1–5. DOI: 10.1109/PESGM.2014.6939030.
- Li, H., Lai, L., and Zhang, W. (2011). Communication requirement for reliable and secure state estimation and control in smart grid. *IEEE Transactions on Smart Grid*, 2(3):476–486. DOI: 10.1109/TSG.2011.2159817.
- Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26. DOI: 10.1016/j.neucom.2016.12.038.
- Mestav, K. R., Luengo-Rozas, J., and Tong, L. (2019). Bayesian state estimation for unobservable distribution systems via deep learning. *IEEE Transactions on Power Systems*, 34(6):4910–4920. DOI: 10.1109/TPWRS.2019.2919157.
- Miranda, V., Krstulovic, J., Keko, H., Moreira, C., and Pereira, J. (2011). Reconstructing missing data in state estimation with autoencoders. *IEEE Transactions on power systems*, 27(2):604–611. DOI: 10.1109/TPWRS.2011.2174810.
- ONS - Operador Nacional do Sistema (2023). Curva de carga horária. Available at: https://www.ons.org.br/Paginas/resultados-da-operacao/historico-da-operacao/curva_carga_horaria.aspx. Accessed in: January de 2023.
- Soares, W., de Souza, J. C. S., Do Coutto Filho, M. B., and Augusto, A. A. (2019). Distribution system state estimation with real-time pseudo-measurements. In *2019 IEEE PES Innovative Smart Grid Technologies Conference-Latin America (ISGT Latin America)*, pages 1–5. IEEE. DOI: 10.1109/ISGT-LA.2019.8895379.
- Vijayalakshmi, G. and Shanthakumar, M. (2019). Image restoration on fusion of mammographs and mri breast images using dual tree complex wavelet transform. *International Journal of Advanced Science and Technology*, 28(17):842–857. Available at: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85081187832&partnerID=40&md5=178acb64427a5e10eb12295b44c97275>.
- Xu, W., Zhang, J., Li, X., Yuan, S., Ma, G., Xue, Z., Jing, X., and Cao, J. (2022). Intelligent denoise laser ultrasonic imaging for inspection of selective laser melting components with rough surface. *NDT & E International*, 125:102548. DOI: 10.1016/j.ndteint.2021.102548.
- Yilmaz, C. S., Yilmaz, V., and Gungor, O. (2022). A theoretical and practical survey of image fusion methods for multispectral pansharpening. *Information Fusion*, 79:1–43. DOI: 10.1016/j.inffus.2021.10.001.
- Zhang, X., Mu, J., Zhang, X., Liu, H., Zong, L., and Li, Y. (2022). Deep anomaly detection with self-supervised learning and adversarial training. *Pattern Recognition*, 121:108234. DOI: 10.1016/j.patcog.2021.108234.
- Zhao, J., Huang, C., Mili, L., Zhang, Y., and Min, L. (2020). Robust medium-voltage distribution system state estimation using multi-source data. In *2020 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. DOI: 10.1109/ISGT45199.2020.9087787.