USING LV REAL-TIME DATA FOR PSEUDO-MEASUREMENTS GENERATION IN MV DISTRIBUTION NETWORKS

Henrique TEIXEIRA
INESC TEC – Portugal
henrique.s.teixeira@inesctec.pt

Pedro N. PEREIRA BARBEIRO
INESC TEC – Portugal
pedro.p.barbeiro@inesctec.pt

Jorge PEREIRA
INESC TEC – Portugal
jpereira@inesctec.pt

António A. MESSIAS
EDP Distribuição – Portugal
aires.messias@edp.pt

Diogo A. LOPES
EDP Distribuição – Portugal
diogo.alveslopes@edp.pt

Pedro G. MATOS
EDP Distribuição – Portugal
pedro.godinhomatos@edp.pt

ABSTRACT

This paper presents a new method to generate pseudo-measurements of different electrical quantities for distribution secondary substations. The idea is to take advantage of the stronger correlation that exists between electrical variables in a given MV/LV substation and the correspondent downstream network. Autoencoders trained with historical data are integrated on an optimisation algorithm, together with real-time metering information of the LV network, for accurately estimate pseudo-measurements at the correspondent MV/LV substation. The proposed methodology is particular suitable either for MV/LV substations not being telemetered in real-time or in case of failures related to telemetry equipment existing in the substation.

INTRODUCTION

The concept of smart distribution grids relies on an active management and operation of the grid. Under this paradigm, an accurate Distribution State Estimation (DSE) module capable of giving a real-time snapshot of distribution networks has a crucial importance. Such tool allows distribution system operators to take the best decisions for operating power systems in a secure and economic way.

Currently, the large majority of the DSE techniques proposed for distribution networks are based on enhanced versions of the conventional state estimation algorithms (e.g. Weighted Least Squares) [1]. Although conventional methods have been successfully applied in transmission networks, their application at the distribution level is a big challenge. One of the main reasons has to do with the number of nodes and branches, which is significantly higher in distribution than in transmission networks, thus a much larger number of real-time measurements are required to guarantee system observability. However, the number of real-time telemetry devices is still very limited in the majority of the today's distribution networks, what can be especially critical when it comes for the execution of the conventional DSE methods. In order to overcome this issue, the use of pseudo-measurements has been the most common practice. This topic is an important area of research where several different techniques have been proposed [2-6], being normally based on historical data, load curve assessment or short-term load forecasts. In [2] pseudo-measurements are achieved using typical daily load diagrams from different consumers which are computed using historical data. A fuzzy load allocation model in a Distribution Management System (DMS) environment is presented in [3]. In [4], the authors present two distinct approaches to model pseudo-measurements, one based on correlation and other based on load probability density functions. The modelling of pseudo-measurements using Artificial Neural Networks (ANN) is addressed in [5]. In a recent study [6], pseudo-measurements are obtained through an approach that combines the use of ANN and typical load profiles. In this study, the authors modelled ANN error through a Gaussian mixture model algorithm in order compute the variance of the pseudo-measurements.

In accordance with the above views, becomes clear the importance of methods to generate pseudo-measurements. Although these type of measurements are always less accurate than real-time measurements, they are essential to run state estimation. Therefore, improving their quality can be the key for increase accuracy in the existing DSE algorithms, particularly in networks with few telemetry devices.

In this paper a new method for the generation of pseudo-measurements for the Medium Voltage (MV) side of the MV/LV substations is presented. Differently from the large majority of the existing methods, which only use information related to the network area where the pseudo-measurements are generated, i.e. MV level, the proposed method uses electrical data gathered from the telemetry equipment located downstream in the Low Voltage (LV) network to provide pseudo-measurements for the MV side of the respective MV/LV substation. In this way, this method takes advantage of a very stronger correlation that exists between electrical quantities of the LV network and the ones obtained at the correspondent MV/LV substation. Therefore, on opposition to other methods that only use LV aggregated information gathered from the MV side, the influence of the downstream LV network is not neglected.

In order to meet the purpose of this study, an expert system is trained with historical data collected from an advanced metering infrastructure which is expected to be

massively deployed at LV level in the future, e.g. Smart Meters (SM) located at consumers' place and/or other telemetry equipment dispersed among the grid. The pseudo-measurement generation tool can be located at DMS level or at Distribution Transformer Controller (DTC) level of the correspondent MV/LV substation and relies on a specific type of artificial neural networks – the Autoencoders (AE). The proposed methodology will be tested in a typical LV network using a set of real data saved from a real MV/LV substation load in EDP Évora Pilot Site.

METHODOLOGY

In general the methodology presented in this paper can be divided in three independent main processes:

a) Building of a synchronized historical dataset;
b) Training the AE;
c) Generation of pseudo-measurements.

As stated before, the AE is the principal element of the pseudo-measurements generation methodology presented in this paper. AE are feedforward neural networks that are built to mirror the input space in their output, being the size of its output layer always the same as the size of its input layer. Therefore, an AE is trained to display an output equal to its input. With adequate training, it can learn the dataset pattern and saves in its weights information about the training data manifold. Once the AE is trained, if an incomplete pattern is presented, it may be replaced by random values producing a significant mismatch between input and output. Different approaches can be followed in order to find the missing values on the way to minimise this error. However, according to some related works in the state estimation area [7-9], the one denoted as constrained search appears to the most suitable method to search a missing signal. This process is based on an optimisation algorithm in order to discover the values that should be introduced in the missing components such that the input-output error for all variables becomes minimised (see Figure 1).

In the previous paragraphs was given a general overview of the basic concepts behind the proposed methodology. In the next sections it will be detailed each one of the processes referred in the points a), b) and c) as well as the interactions between them.

Build the Historical Dataset

The achievement of accurate pseudo-measurements depends on the existence of a large historical database, which is used during the training stage of the AE. It is crucial that such database contain data about the variables that will be passed to the AE (both for the missing signals and for the measurements that will be available in real-time). At least, it is mandatory the existence of historical data for the pseudo-measurements to be generated and a few number of measurements being transmitted in a real-time mode. Additionally, all data must be synchronized and in sufficient number for each time instant/operating point, otherwise the AE will not learn effectively the patterns/correlations between the electrical variables of a given network.

There is no rule of thumb regarding the quantity of data in the historical database. However, it is known that few or too much data will lead to an inaccurate AE. A trial and error approach can be followed to identify the optimal quantity of data in the historical database to be passed to the AE whenever a large amount of historical is available. Off course, if the data records are scarce (e.g. less than a week) all data should be considered.

Training Stage

The first step of the training stage comprises the selection and execution of an adequate training algorithm in order to get an AE properly trained. Secondly, it is necessary to define the AE architecture with respect to the number of hidden layers responsible for coding the input information. According to [10], neural networks with more hidden layers have proved to be difficult to train although could lead to better accuracy in some particular cases. In this sense, an AE with a one hidden layer was the architecture adopted (see Figure 1). The most adequate number and type of variables (from the historical dataset) to be used as the input data of the AE is also performed at this stage. It is important to state that whenever the quantity and type of measurements present in the input dataset is changed, a new training must be performed. For instance, in a real world implementation, a new training is required to take advantage of new telemetry devices that have been added into the network.

Regarding the training algorithm, an adaptive gradient-based algorithm called Resilient Back-Propagation algorithm was chosen [11]. This algorithm belongs to the most widely used class of algorithms for supervised learning of neural networks and it is a variation of the classical Back-Propagation algorithm. Experimental training tests were carried out in order to select the most appropriate activation function for the hidden and output layers. For the specific problem under analysis, results showed that a non-linear activation function, namely a symmetric sigmoid, was the best option for both layers.

Other important parameters which have to be defined are the number of training epochs and number of neurons in the hidden layer. In order to have the AE properly tuned, several preliminary tests were performed. It involved testing several combinations of these parameters, as well as variations on the quantity of historical data. After this preliminary analysis, it was considered the parametrisation that lead to a good compromise between time performance (both in training and running stages) and accuracy. The parameters selected were: 300 training epochs, value of 0.4 for the hidden layer reduction rate (ratio between the number of neurons in the hidden layer and the number of neurons in the input/output layer) and 30 days of historical data for training the AE (in time steps of 15 minutes).

It is important to stress that before the execution of the

training algorithm all data must go through a standardisation procedure with the goal of pre-treating the input and output training dataset. In this scale adjustment process, the input and output variables are adjusted to the range of the activation function in the interval of [-1,1].

Pseudo-Measurements Generation

In accordance to the mentioned at the begging of the current section, an optimisation algorithm was used to find the missing values, meaning the pseudo-measurements. Therefore, after having an AE properly trained, the LV measurements available in real-time (e.g. coming from SM installed at consumers' place) are used as input of the AE for guiding the optimization algorithm on the achievement of the pseudo-measurements. The concept is illustrated in Figure 1 where an Evolutionary Particle Swarm Optimisation (EPSO) was the algorithm chosen in this work for finding the missing signals. It has been successfully applied in several problems of the power systems area [12].

In the context of this approach, the pseudo-measurements that are intended to be generated are values of injected active and reactive power and value of voltage magnitude.

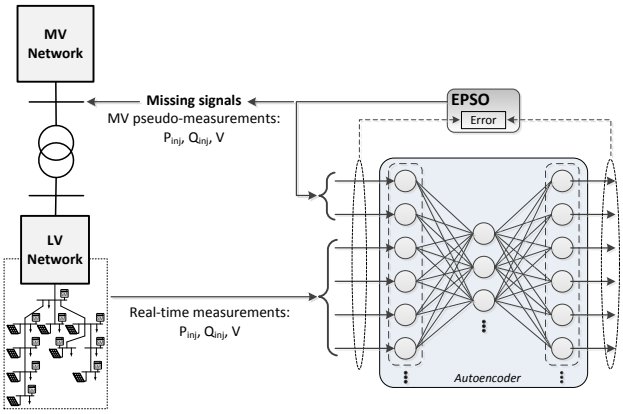

Figure 1: Illustration of the methodology proposed for pseudo-measurements generation.

CASE STUDY AND RESULTS

Low Voltage Network and Data Modelling

A small typical Portuguese LV network was used with the purpose of testing the proposed methodology [9]. The single-line diagram of the network is present in Figure 2. The network has a total of 33 nodes (containing 57 consumers, some single-phase and other three-phase) and an annual peak load of 62.6 kW, registered in the winter season.

Several microgeneration units (photovoltaic panels) were added and randomly distributed through the network clients. Each microgeneration unit represents 50% of the contracted power of the correspondent consumer. In order to represent different days (e.g. sunny, cloudy, rainy, etc.) and consequently different power generations, 5 distinct profiles obtained from a real meteorological station [13] were randomly distributed by existing units according to their probability of occurrence in a typical Portuguese winter.

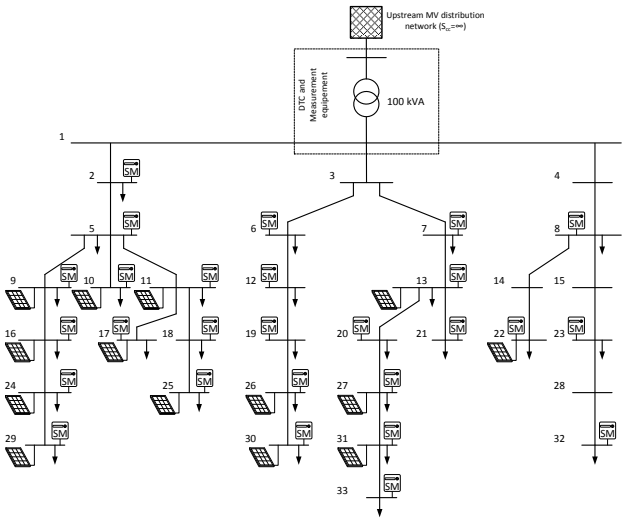

Figure 2: Single-line diagram of the studied network.

Regarding the data related to the load, the only available for this grid were average values at the MV/LV substation level (data saved from a real MV/LV substation load in EDP Évora Pilot Site). The data include records of the active and reactive power values as well as the voltage magnitude values for an entire month in time steps of 15 minutes. In order to represent the behaviour of the individual loads dispersed among the grid for each time instant, a load allocation technique was required to be performed. The approach consisted on the distribution of the active and reactive power values measured at the referred substation proportionally to the contracted power of the clients existing in the grid. This task was done in such a way that, at the end, the load flow results for both values of voltage magnitude and injected powers at the reference bus (bus 1 in Figure 2) match the historical data of the MV/LV substation.

Telemetry Scenarios Characterisation

With respect to the measurement equipment, it was considered that the MV/LV substation holds a DTC as well as the associated measurement equipment with the capability of saving information in about the following variables: active and reactive power flows in the transformer and the voltage magnitude values at the medium and low voltage sides of the transformer. It was also assumed that each consumer owns a SM to monitor his consumption and an additional SM for measuring its power generation whenever a microgeneration unit is also installed. It was assumed that all of these devices are capable of monitoring synchronously the following measurements: active and reactive power of each consumer (consumed or produced) and voltage magnitude values. Phase angles were assumed not to be measured since for the majority of SM foreseen to be

deployed in LV grids this capability is not expected.

As it happens in the already existing smart grid pilot test sites, is expected that only some of the SM installed will have the capability of transmitting data in real-time (for instance, using GPRS technology). Thus, in order to test the proposed methodology, five different scenarios for the LV customers that own SM with the capability of transmitting data in real-time (SM_r) were defined. The first three were created by following an approach similar to the one presented in [14]. This approach uses Information Theory concepts to determine the best locations for a given number of SM_r. The best locations mean those with stronger influence on the accuracy of the pseudo-measurements. The other two scenarios were defined as follows: in scenario 4 SM_r were placed at the customers located on the farthest buses from the secondary substation, whereas in scenario 5 was assumed that all customers with microgeneration have a SM_r. In Table I is presented the SM_r bus location in each scenario in accordance to Figure 2.

Table I: Number and location of SM_r in each scenario.

Scenario (Scn)	Number of SM_r	Location of SM (bus)
1	1	30
2	4	24-25-30-32
3	8	2-7-23-24-25-30-31-32
4	8	17-21-22-25-29-30-32-33
5	14	9-10-11-13-16-17-22-24-25-26-27-29-30-31

Results

The results obtained for the pseudo-measurements generation methodology are presented below. For each previously created scenario, results were computed for an evaluation set of seven days (last seven days from the historical database). In Figure 3 are depicted boxplots with the absolute error for the injected active and reactive power and also for the voltage magnitude. In addition, it is presented in Table II the Mean Absolute Error (MAE) of the obtained pseudo-measurements.

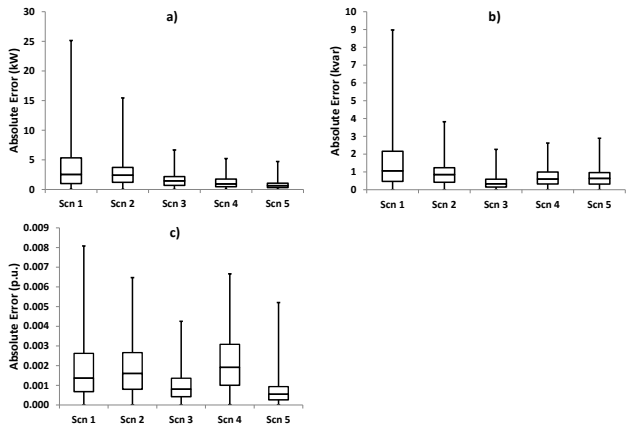

Figure 3: Absolute error of: a) injected active power, b) injected reactive power and c) voltage magnitude.

As it can be seen by observing Figure 3 and Table I, scenarios 3 and 5 are those that present generally pseudo-measurements with better accuracy. Nevertheless, scenario 3 requires a smaller number of SM_r meaning a better trade-off between the number of real-time measurements required and the accuracy of the pseudo-measurements obtained. Looking to the same results, it is clear that the scenarios 1 and 2 accounts for the less accurate pseudo-measurements. This was expected because these are the scenarios with less real-time measurements (only one SM_r exists in scenario 1 and four in scenario 2). Nevertheless, the error obtained for the pseudo-measurements remains satisfactory. For instance, in scenario 1, the absolute error for the pseudo-measurement of voltage magnitude is below 0.003 p.u. for 75% of the samples analysed (7 days).

Table II: Pseudo-measurements MAE.

Pseudo-Measurement	Scn 1	Scn 2	Scn 3	Scn 4	Scn 5
P (kW)	3.50	2.78	1.56	1.19	0.76
Q (kvar)	1.48	0.89	0.42	0.69	0.68
|V| (p.u.)	0.00179	0.00183	0.00095	0.00210	0.00069

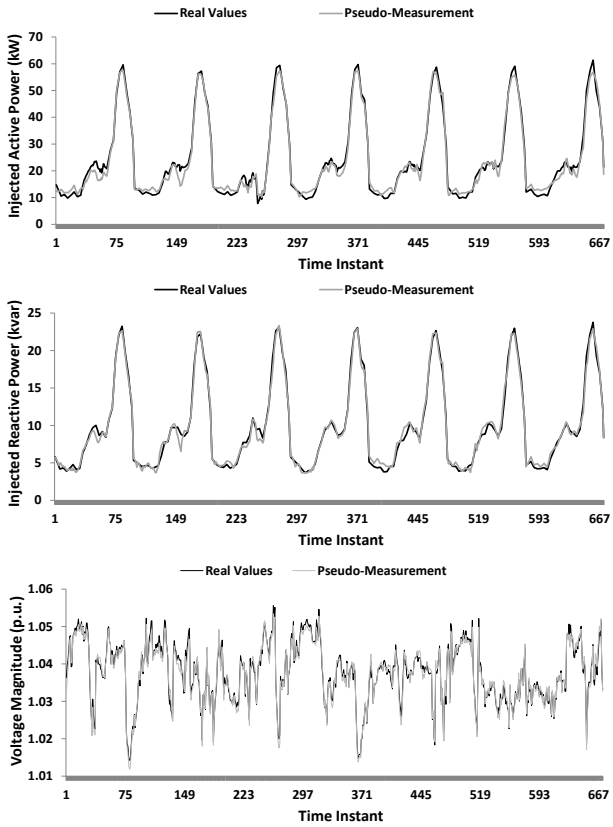

Figure 4: Real values and the respective pseudo-measurements for scenario 3.

Although scenario 4 is identical to scenario 3, with respect to the total number of SM_r, the pseudo-measurements achieved in scenario 4 are less accurate than in scenario 3 (except for the active power values). These results can be explained because the SM_r belonging to scenario 4 do not follow any optimisation procedure for its location as in scenario 3.

In Figure 4, it is depicted for scenario 3 a graphical

representation of the real values and of the pseudo-measurements obtained for the entire evaluation set, one week. As it can be seen, each pseudo-measurement analysed fits nearly its real value curve in all time instants. This result highlights the performance of the proposed methodology regarding accuracy.

CONCLUSION

A new method for the generation of pseudo-measurements was presented in this paper. The main advantage of the proposed method is that it only uses information coming from LV assets to provide accurate pseudo-measurement at the correspondent MV/LV substation. Therefore, the method can take advantage of the stronger correlation that exists between the pseudo-measurements generated and the electrical variables of the correspondent downstream LV network.

The obtained results indicate that with an AE properly trained, accurate pseudo-measurements can be achieved, even when a low number of real-time measurements are available. As expected, the addition of real-time measurements lead in general terms to a decrease in the error of the pseudo-measurements generated. However, the trade-off between a better accuracy and an increased cost should be carefully analysed. The results attained evidence the importance on the use of methodologies to found the most suitable locations for the installation of metering devices with the capability of transmitting data in real-time. In fact, if this kind of approaches is executed, it could be possible to obtain with less number of telemetry devices a similar error for the pseudo-measurements than with more devices randomly located.

The proposed approach could be particularly useful either when substations are not being telemetered in real-time or in case of telemetry equipment failures. In these circumstances, the missing real-time measurements could be replaced by the pseudo-measurements generated and later used to guarantee observability in a DSE conventional algorithm (e.g. WLS).

ACKNOWLEDGMENTS

This work was co-financed by the European Community's 7th Framework Program (FP7/2007-2013) under grant agreement n°308755 – the SuSTAINABLE project. It was also made in the framework of the BEST CASE project ("NORTE-07-0124-FEDER-000056") financed by the North Portugal Regional Operational Programme (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF).

REFERENCES

[1] R. Singh, B. C. Pal, and R. A. Jabr, 2009, "Choice of estimator for distribution system state estimation", *IET Generation, Transmission & Distribution*, vol. 3, no. 7, 666-678.

[2] M. D. Nimrihter, R. M. Ciric, and A. Slijepcevic, 2001, "Measurements and pseudo-measurements in distribution networks", in *Proc. IEEE Instr. and Measurement Technology Conf.*, vol. 3, 1791-1796.

[3] V. Miranda, J. Pereira, and J. T. Saraiva, 2000, "Load allocation in DMS with a fuzzy state estimator", *IEEE Transactions on Power Systems*, vol. 15, no. 2, 529-534.

[4] E. Manitsas, R. Singh, B. Pal, and G. Strbac, 2008, "Modelling of pseudo-measurements for distribution system state estimation", in *IET-CIRED Seminar: SmartGrids for Distribution*, 1-4.

[5] A. Bernieri, G. Betta, C. Liguori, and A. Losi, 1996, "Neural networks and pseudo-measurements for real-time monitoring of distribution systems", *IEEE Transactions on Instrumentation and Measurement*, vol. 45, no. 2, 645-650.

[6] E. Manitsas, R. Singh, B. C. Pal, and G. Strbac, 2012, "Distribution System State Estimation Using an Artificial Neural Network Approach for Pseudo Measurement Modeling", *IEEE Transactions on Power Systems*, vol. 27, no. 4, 1888-1896.

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

[8] J. Krstulovic, V. Miranda, A. J. A. Simoes Costa, and J. Pereira, 2013, "Towards an Auto-Associative Topology State Estimator", *IEEE Transactions on Power Systems*, vol. 28, no. 3, 3311-3318.

[9] P. N. P. Barbeiro, J. Krstulovic, H. Teixeira, J. Pereira, F. J. Soares, and J. P. Iria, 2014, "State estimation in distribution smart grids using autoencoders", in *IEEE International Power Engineering and Optimization Conference (PEOCO)*, 358-363.

[10] G. E. Hinton, and R. R. Salakhutdinov, 2006, "Reducing the Dimensionality of Data with Neural Networks", *Science*, vol. 313, no. 5786, 504-507.

[11] M. Riedmiller, and H. Braun, 1993, "A direct adaptive method for faster backpropagation learning: the RPROP algorithm", in *IEEE International Conference on Neural Networks*, vol. 1, 586-591.

[12] V. Miranda, and N. Fonseca, 2002, "EPSO - best-of-two-worlds meta-heuristic applied to power system problems", in Congress on Evolutionary Computation, Honolulu, Hawaii, vol. 2, 1080-1085.

[13] C. Gouveia, D. Rua, F. Ribeiro, C. L. Moreira, and J. A. P. Lopes, 2013, "INESC Porto experimental Smart Grid: Enabling the deployment of EV and DER", in IEEE PowerTech, Grenoble, 1-6.

[14] A. A. Augusto, J. Pereira, V. Miranda, J. C. Stacchini de Souza, and M. B. Do Coutto Filho, 2014, "Most relevant measurements for state estimation according to information theoretic criteria", in *Int. Conf. on Probabilistic Methods Applied to Power Systems (PMAPS)*, 1-6.