Phasor measurements estimation on distribution networks using machine learning

Silviu Nistor, Aftab Khan and Mahesh Sooriyabandara
Telecommunications Research Laboratory, Toshiba Research Europe Limited
32 Queen Square, Bristol, BS1 4SL, United Kingdom
Email: {silviu.nistor,aftab.khan,mahesh}@toshiba-trel.com

***Abstract*—The uptake of distribution generation on electricity distribution networks imposes the operators to install new measurement devices such as phasor measurement units to achieve network observability. In this paper, we propose a framework for estimating synchronized phasor measurements for a virtual node using the measurements from the other nodes in the network. This system uses a machine learning method, in particular supervised regression models, to provide estimates. We show the performance of the proposed framework comparing two widely used regression methods i.e., Generalized Linear Models and Artificial Neural Networks. We extensively evaluate the proposed approach utilizing a real-world dataset collected from a medium voltage ring feeder. Our results indicate very low error rates; the average error for voltage magnitude was approx. $0.2V$ while for phase angle was $0.7mrad$. Such low errors indicate the potential for reducing the scale of the measuring infrastructure required on distribution networks and increasing their reliability.**

I. INTRODUCTION

Conventionally, distribution networks transported electricity from the transmission substations to the end consumers. The one way flow of electricity and radial topology meant that conservative dimensioning of the network was sufficient to ensure the correct operation, without too many real-time measuring points. However, over the last decade, more consumers, communities and businesses have installed distributed generators. The integration of the electricity system with the transport (e.g. electric vehicles) and heating (e.g. fuel cell co-generation units) infrastructures is also taking place at the medium and low voltage levels of the electricity grid. With these technologies come a series of challenges which require the network operators to have complete network observability, similar to the transmission system operators. However the distribution network requires significantly more measuring devices than the transmission network to achieve this.

In this paper, we investigate the performance of a machine learning (ML) driven engine to replace physical measurement devices on the electricity networks. A dataset of real measurements from the same network are utilized to train models for estimating measurements. These trained models are capable of providing pseudo measurements based only on the real measurements from the other nodes. The applications for the ML engine include the reduction of the number of physical measurement devices installed or, in case of a temporary failure of the measurement device, fill in for the lack of measurements. To evaluate this, we use syncro-phasor measurements from 7 nodes within a real distribution network (measurements collected during an innovation project). We compare the performance of the proposed approach against the current standards for Phasor Measurement Units (PMUs). We employ two supervised ML approaches namely, Generalized Linear Models and Artificial Neural Networks.

Wide Area Monitoring Systems with PMUs are used to provide phasor measurements (current or voltage) syncronised to the same time reference. They can offer an accurate snapshot of the state of the power network they are connected to. Phasor measurements have been used on the transmission network for observability and for stability applications. Recently, such measurement devices have also been installed on the distribution networks. An extensive list of potential applications is given in [1] including unintentional islanding detection and a state estimation (SE) algorithm.

SE algorithm is a mathematical method to output a description of the power system by computing the best estimate of the state variables (V, θ) based on measurements from PMUs and SCADA system, pseudo-measurements from smart meters and network topology. The rest of the secondary variables (real and reactive power flows through the lines and nodal power injections) can be calculated from the state variables. Because the PMUs bring direct measurements of state variables and as they are synchronized they are regarded as accurate and are expected be assigned higher weights in the SE algorithm.

Different machine learning algorithms applied in power system sector have been described in the literature to extract information from measurements either to estimate pseudo-measurements or network topology. These approaches can be separated into two categories: *network information driven* and *network information independent*. In the former category, the formulation of the problem to be solved with ML includes at least one equation from AC or DC circuit analysis. An example of this type of algorithm is described in [2] where the topology of a distribution network is detected using the correlations between voltage measurements and a sparse Markov random field. In [3] the state variables (V, θ) of a section of a distribution network are found using a Bayesian linear estimator based on a linear approximation of the power flow equations considering smart meter and PMU type of measurements.

The network information independent methods employ a black box approach, where the variables that need to be

978-1-5090-4000-1/17/$31.00 ©2017 IEEE

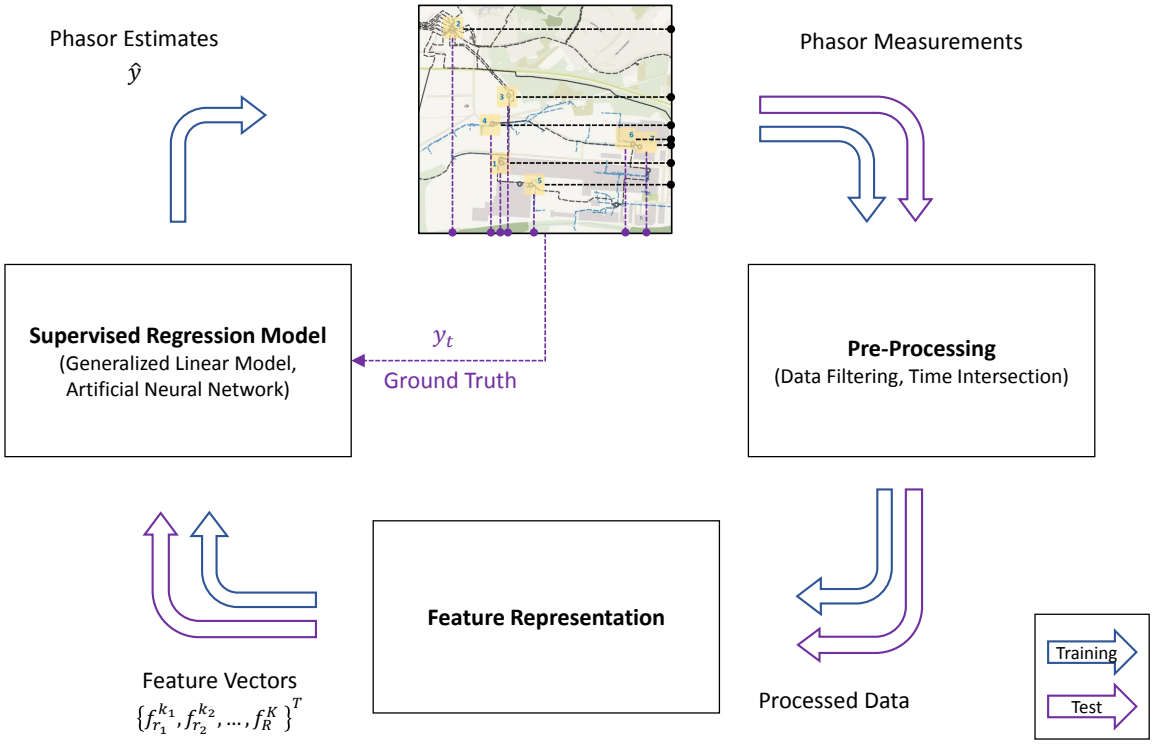


Fig. 1. Block diagram for estimating phasor measurements using supervised regression models. Training and test phases are separately highlighted.

estimated by the algorithm are not connected to the measurement available through physical equations. These estimated variables are then used in SE or power flow analysis. In [4] an ML approach was used for load forecasting. Repeating patterns hidden inside the load time series are used to define rules governing the load variation. These were found using a parallel distributed processing model. In [5] support vector machine (SVM) and Kohonen network (SOM) have been used to detect pre-emergency operating points of the network starting from the voltage measurements, which can help in preventing network voltage collapses. Fault detection considering substation measurements and weather data using Neural Networks (NN) and Naïve Bayes (NB) was tested in [6].

The ML method introduced in this study with applications in power systems can be categorized as a network information independent approach. Our study contributes to the state of the art with an extensive investigation for estimating directly the state variables (V, θ), rather than pseudo-measurements. We argue that, given enough data points for training, the physical measurement equipment can be replaced by the virtual measurements based on our approach. Virtual measurements are obtained by an ML engine using regression models.

The proposed method can be used as a solution to the multi-stage optimal PMU placement problem [7] in which a gradual deployment of PMUs across the distribution network is required because of the high number of nodes. Our ML engine can create virtual nodes capable of providing estimates until the physical equipment is installed. Further, when they are installed and network observability is achieved, the ML algorithm can offer redundancy and reliability of the solution in the case of failures in measurement devices.

II. Methodology

A mathematical representation for the instant voltage is:

$$v(t) = \sqrt{2} \cdot V \cdot cos(2\pi ft + \theta) \quad (1)$$

Where V is the RMS value of the voltage amplitude, f is the instantaneous frequency and θ is the angular starting point for the waveform. A PMU reports all the three measurements needed to reconstruct the instant voltage for a certain time.

The current standard for PMUs, IEEE C37.118 [8], specifies that the Total Vector Error (TVE) should be kept below 1%. TVE aggregates the errors from the three sources of error (V, θ, f). If taken separately, the error for the voltage magnitude limit would be 1%, while the error for the voltage angle is $10mrad$ [9].

There are several phasor measurement devices on the market, suitable for the use in transmission networks, but suitable devices for the distribution part of the network are still at development stage. Because of the particularities of the distribution network (the short length between nodes and the small reactance of the lines) the accuracy of the

Fig. 2. SUNSEED Map: Electricity distribution network with PMU locations.

measurement nodes must be higher than of those installed on the transmission network. The PMU in the SUNSEED[1] project offers an accuracy for the voltage magnitude of 0.1%, while for the voltage angle $0.34mrad$. Another issue which needs to be highlighted is the requirement of a high number of PMUs to achieve network observability. According to [10], between $1/3$ and $1/5$ of nodes must be fitted with PMUs in order to achieve complete observability. This could mean a significant investment for distribution network operators (DNOs). Our approach is to reduce this number by using machine learning to substitute physical equipment with virtual measurement points.

As can be seen in Figure 1 phasor measurements are initially pre-processed. In this step, pre-processing techniques such as data interpolation for missing data or filters for noisy measurements can be used. For evaluation purposes, we extract an intersection of timestamps where data from all the nodes is available. This can be utilized at the training phase only where synchronised data is required to train such a framework. This step results in a processed data of nodes within a network with measurements from the same timestamps. This data is then used for establishing features to train for estimating target variables.

In our feature representation stage (cf Figure 1), synchronised raw measurements are directly used with minimal addition of extra features. Due to this, the feature extraction stage is significantly faster than traditional approaches where several statistical or geometric features are extracted over a time window of measurements. They also have a significant limitation for this application as the output target in such scenarios is an aggregated estimate whereas the proposed system has the capability of producing estimates for a given time stamp based on the real direct measurements. Also note,

[1]https://sunseed-fp7.eu/

TABLE I
SUMMARY OF VARIOUS ATTRIBUTES RELATED TO THE COLLECTED DATASET.

Dataset Overview	
Total Nodes (N)	7
Total Samples	174213
Sampling Rate	$1Hz$
Measurements (M)	9 $((V_1,...,V_3),(\theta_1,...,\theta_3),h,psp_v,psp_\theta)$
Total Features	7 nodes $\times$ 9 measurements $=$ 63
Train/Test Split	$75/25$

that there is no explicit *segmentation* stage as each set of measurements from all nodes at a given timestamp are directly used (i.e., the total number of segments are the same as the total number of data points).

In total we use 9 measurements for every node in the network, as seen in Table I, which include $\{(V_1,V_2,V_3),(\theta_1,\theta_2,\theta_3),hour,psp_v,psp_\theta\}$ where (V_1,V_2,V_3) are the voltage magnitudes for each of the three phases, $(\theta_1,\theta_2,\theta_3)$ are the voltage phase angles, $hour$ is the hour of the day feature (computed using the associated time stamps). psp_v and psp_θ are the positive sequence voltage magnitude and phase respectively.

Feature data from the previous step is then used to train individual models for separate target variables.Note, in our feature representation, no measurements from the test node are used reflecting a real-world scenario in which a real node when replaced with a virtual node will have no measurements from the virtual node.

In this work, we utilize a supervised machine learning framework in which labeled data is provided in the form of true measurements at the training stage. In particular, we use regression models that are appropriate for estimating numeric measurements (as opposed to using classification models in which the target is categorical in nature):

a) Generalized Linear Model (GLM): : Input features f are used to fit GLMs [11] for estimating the target data individually. The GLM model assigns coefficients to each of the input feature in the form of a linear equation capable of estimating the target variable. This linear model is of the form:

$$y_p^q \sim 1 + [f_{r_1}^{k_1} + f_{r_1}^{k_2} + ... + f_R^K] \tag{2}$$

where, $M = \{k_1,k_2,...,K\}$, $N = \{n_1,n_2,...,L\}$, $q \in M$, $p \in N$, $r = \{\{r_1,r_2,...,R\} : r \in N \text{ and } r \notin p\}$.

b) Neural Network (NN): : We also use the same set of input features to train multiple Neural Networks [12] for all of the target variables. In particular, we use Bayesian regularization [13] for training the models with 100 nodes.

III. SUNSEED USE CASE

The algorithm described in the previous section was evaluated using real data collected in the SUNSEED project [14]. This project investigate converging communication infrastructures of DNO and telecommunication companies for future smart grids applications. The field trial, conducted in Slovenia,

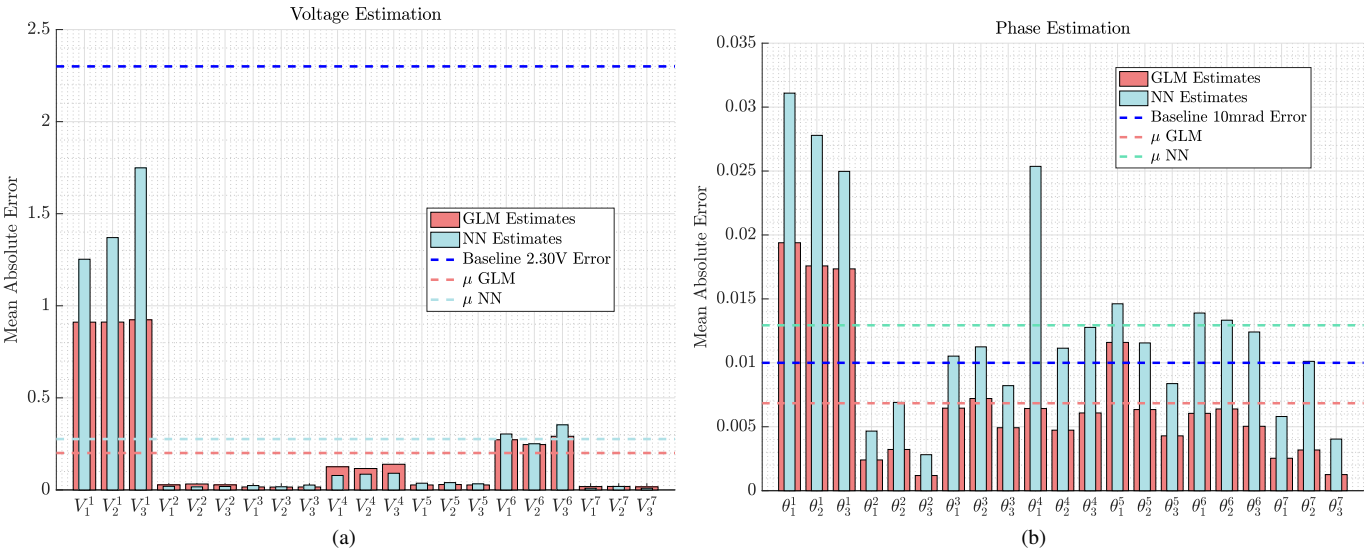


Fig. 3. Voltage and Phase estimation results are shown using both GLM and NN regression models. Comparisons are made against respective baselines. Average error rates are also indicated.

involved deploying smart meters, power measurement devices and PMUs in several locations. Figure 2 shows the GIS map with nodes labeled by number.

Syncronised voltage phasor measurements are collected using PMUs installed on two 20kV urban feeders from the Kromberk area. Apart from the loads, the feeder connect photovoltaic (PV) panels and a co-generation unit therefore reverse power flow is common. The feeders are part of a new strong network with underground cabling. The PMUs take 3 phase measurements from the secondary side of the 20/0.4kV transformers. The measurements are sent securely to a Mongodb database and form part of the input of a distribution state estimation application developed during the SUNSEED project.

As shown in Figure 1, we evaluate the proposed framework using the nodes within the same network which are filtered out and an intersection of timestamps is established. Features, as described in Section II are then used to train the two types of supervised regression models. We then compare the two methodologies against the baseline represented by the hardware accuracy specified in IEEE C37.118.

A. Experiments and Results

In order to evaluate the performance of the proposed framework using the two regression models, we perform several sets of experiments. For evaluation purposes, a consistent metric of *Mean Absolute Error* is utilised calculated as below:

$$e = \frac{1}{s_t}\sum_{i=1}^{s_t} |\hat{y}_i - y_i| \tag{3}$$

where s_t represent the total number of test samples, y_i is the true value of a target variable (for example Voltage on a certain node) and $\hat{y}_i$ represents the estimated value for a target variable.

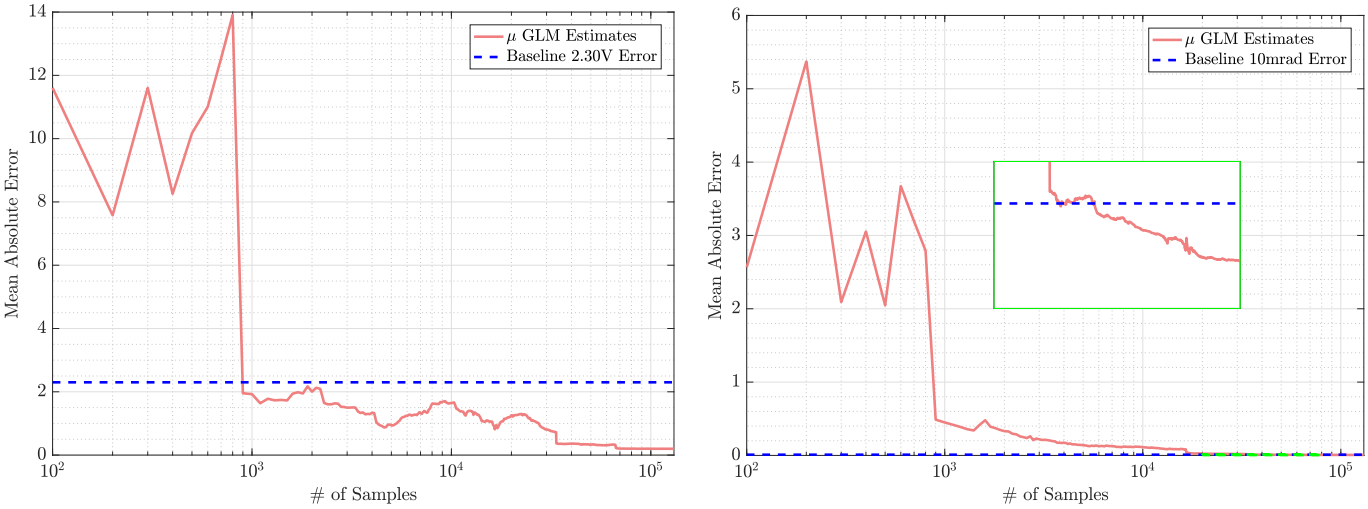
Figure 3a shows the voltage estimation results using both GLM and NN. Baseline of $2.30V$ is used [9]. It can be seen that in all cases, the estimation results have low errors compared against the baseline. In general GLM estimation performs better compared against the NN estimation. Similarly in Figure 3b, phase estimation results are shown using both GLM and NN. A baseline of $10mrad$ is used [9]. In this case, GLM outperforms both the baseline and the NN estimation results. NN estimation on average has a higher error rate compared against the baseline.

Based on these results, it can be inferred that a certain linear relationship exists between phasor measurements from different nodes and therefore can be more accurately modeled using a linear model. However, with more training data and higher complexity neural networks (such as deep learning methods), these results can further be improved.

In Figures 4a and 4b), voltage and phase estimation results are respectively shown using GLM with respect to the number of samples in the training data. Models are trained using the total number of samples (x-axis; incremented within the 75% of training data) and errors are reported on the test set (remaining 25% of data as defined in the previous experiment). In the case of voltage, estimation results outperform the baseline at around 1000 samples (representing 1000s of data) whilst for phase, this is achieved at 44600 samples. This means that for phase estimation, a substantial amount of data, compared with the voltage, is required in order to make precise estimations.

Figure 5 shows an example node and its associated true and predicted voltage estimation results. Several points in the time line are zoomed in. It can be seen that error rates are very low in majority of the cases (i.e., predicted results very closely match the true measurements on this node).

Figure 6 shows the effect of added number of nodes to



(a) Voltage Estimation (b) Phase Estimation

Fig. 4. Voltage and Phase estimation results are shown with respect to the number of samples. All of these results are obtained using GLM with a fixed test set.

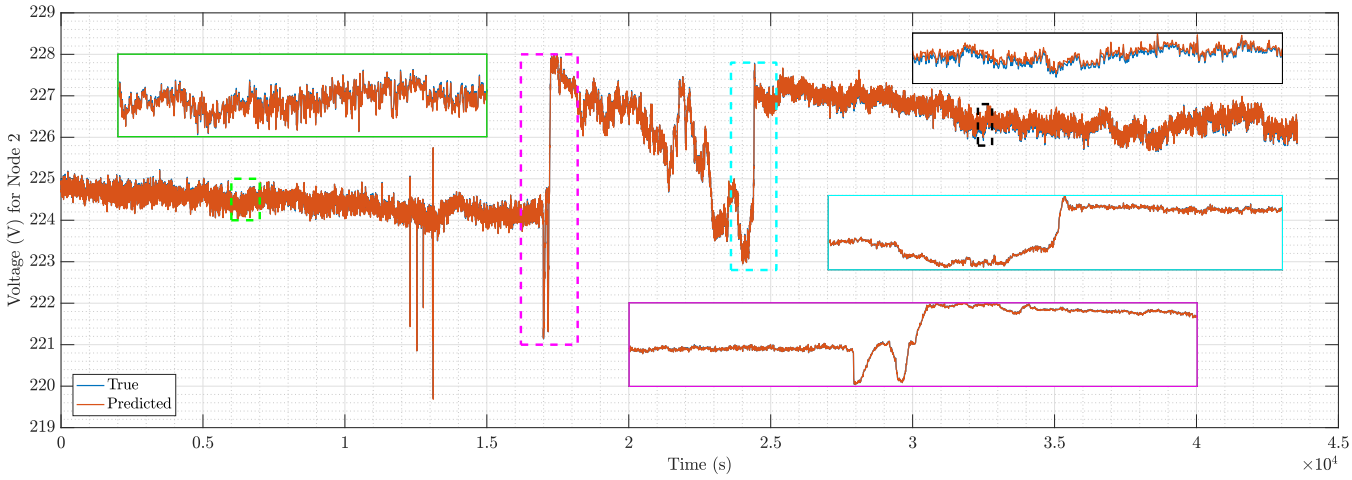


Fig. 5. True and predicted voltage estimation results for Node 2 of the SUNSEED network. Several parts of the results are zoomed in to illustrate low error rates even when the true signal has high variance.

both voltage and phase estimation results. In each case, a single node is tested and training is performed using all the combinations of the remaining nodes. For example, for 3 number of training nodes, separate models are trained using all combinations of 3 nodes out of 6 nodes and using the 7th node for test. This is repeated for all nodes to be tested separately. Average results are then shown in Figure 6. It can be seen that as more nodes are added, estimation results improve with errors reducing. For Voltage, 3 nodes onwards, the results are stable and in all cases the error is lower than the baseline of $2.30V$ [9]. For Phase, all 6 nodes are needed for the errors to be below the baseline of $10mrad$ [9].

In summary, these results indicate a very high level of accuracy and shows that data-driven virtual nodes (trained using machine learning) are capable of replacing physical measurement equipment used for power system applications.

B. Discussions

Because both voltage magnitude and phase angle have small variations on the MV and LV networks in comparison to the transmission system, the estimations of any state variables must be accurate as they can introduce large errors in the SE or load flow softwares. The high precision measurements taken with the PMUs installed in the SUNSEED project are essential for achieving the performance reported in Section III-A. The results show that the GLM method described here performs very well in estimating the voltage magnitude at a certain network node based on the measurements from other nodes of the network. Both the error estimated for individual nodes and the average error are below the precision required in the C37.118 standard as can also be seen in Figure 3a. Although they are still below the limit of the standard, the

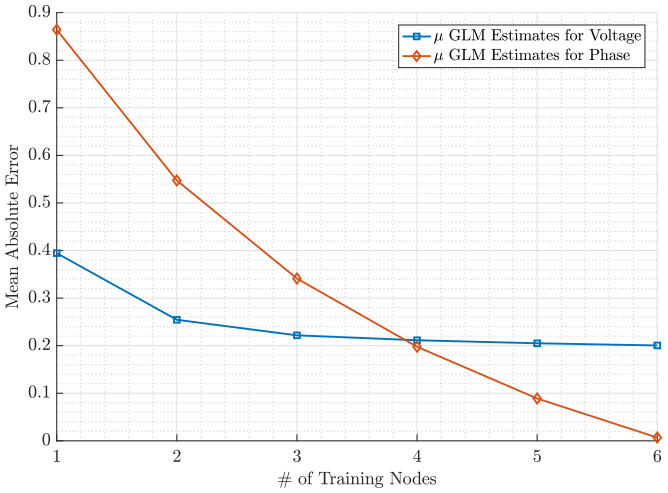


Fig. 6. Estimation results with respect to the total number of nodes.

estimates of node 1 are less accurate than the rest of the nodes. For justification we need to observe the network configuration in Figure 2. Node 1, apart from being connected to the ring feeder, is also connected to another feeder (left of the map). This means that the voltage at node 1 is also influenced by the demand and generation of another feeder where no measurements were taken, making it harder to estimate.

For phase angle estimation, Figure 3b, the GLM method outputs an average error below the value specified in the standard. Node 1 has a higher value due to the network topology, as described in the previous paragraph.

The minimum number of samples required for GLM to obtain below 1% error for magnitude was approximately 1000 while for phase angle 44600 samples are required. Therefore, for an accurate estimation system, higher of the two numbers must be used, which corresponds to 5.5 hours of data. This analysis is important as it enables setting the minimum time for the measuring campaign when installing an ML based solution for network observability. In Figure 5 we have showed that the performance of the GLM method does not deteriorate and stays reliable even with significant network voltage variations.

The results displayed in Figure 6 show that the error in the target node is heavily dependent on the number of nodes used for training. For voltage magnitude the largest reduction in estimation is realized from the first node to the second. Further on, from 2 to 3 and so on, the gain diminishes significantly. For phase angle, estimation accuracy increases almost linearly as more nodes are used in the training data. Different applications tolerate different error rates, therefore establishing the number of nodes required to achieve that accuracy level can be performed with such an analysis. This can lead to significant cost savings in implementing such solutions.

IV. Conclusion

In this paper, we proposed a method that uses machine learning to estimate synchronized phasor measurements. We validated our approach using voltage phasor measurements from a field trial on an electricity distribution ring feeder. The results showed that the estimation framework is highly reliable and capable of replacing the hardware measurement equipment. The average error for voltage magnitude and phase angle was $0.2V$ and and $0.7mrad$, respectively. In particular, GLMs performed better than the NN model and the baseline. The estimation error decreased with the increase in the total number of nodes considered for training, however we found that the rate of decrease is dependent on the type of measurement.

For future work we aim to quantify the benefits of using such an ML-driven approach when applied to a state estimation algorithm.

Acknowledgment

This work is partially funded by the European Commission, under Grant agreement no. 619437 “SUNSEED”. The SUNSEED project is a joint undertaking of 9 partner institutions and their contributions are fully acknowledged.

References

[1] A. von Meier, D. Culler, and R. Arghandeh, “Micro-synchrophasors for distribution systems,” in *ISGT 2014*, Feb 2014, pp. 1–5.

[2] S. Bolognani, N. Bof, D. Michelotti, R. Muraro, and L. Schenato, “Identification of power distribution network topology via voltage correlation analysis,” in *52nd IEEE Conference on Decision and Control*, Dec 2013, pp. 1659–1664.

[3] L. Schenato, G. Barchi, D. Macii, R. Arghandeh, K. Poolla, and A. Von Meier, “Bayesian linear state estimation using smart meters and PMUs measurements in distribution grids,” in *2014 IEEE International Conference on Smart Grid Communications, SmartGridComm 2014*, 2015, pp. 572–577.

[4] J. Wu, Y. He, and N. Jenkins, “A robust state estimator for medium voltage distribution networks,” *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1008–1016, May 2013.

[5] V. Kurbatsky and N. Tomin, “Identification of pre-emergency states in the electric power system on the basis of machine learning technologies,” in *2016 12th World Congress on Intelligent Control and Automation (WCICA)*, June 2016, pp. 378–383.

[6] J. Zheng and A. Dagnino, “An initial study of predictive machine learning analytics on large volumes of historical data for power system applications,” in *2014 IEEE International Conference on Big Data (Big Data)*, Oct 2014, pp. 952–959.

[7] F. Aminifar, M. Fotuhi-Firuzabad, A. Safdarian, A. Davoudi, and M. Shahidehpour, “Synchrophasor measurement technology in power systems: Panorama and state-of-the-art,” *IEEE Access*, vol. 2, pp. 1607–1628, 2014.

[8] I. Power and E. Society, *C37.118.1-2011 - IEEE Standard for Synchrophasor Measurements for Power Systems*, 2011, vol. 2011, no. December.

[9] M. Lixia, C. Muscas, and S. Sulis, “On the accuracy specifications of Phasor Measurement Units,” pp. 1435–1440, 2010.

[10] P. M. Ashton, G. A. Taylor, M. R. Irving, I. Pisica, A. M. Carter, and M. E. Bradley, “Novel application of detrended fluctuation analysis for state estimation using synchrophasor measurements,” *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 1930–1938, 2013.

[11] P. McCullagh and J. A. Nelder, *Generalized linear models (Second edition)*. London: Chapman & Hall, 1989.

[12] H. B. Demuth, M. H. Beale, O. De Jess, and M. T. Hagan, *Neural Network Design*, 2nd ed. USA: Martin Hagan, 2014.

[13] F. D. Foresee and M. T. Hagan, “Gauss-newton approximation to bayesian learning,” in *Neural Networks,1997., International Conference on*, vol. 3, Jun 1997, pp. 1930–1935 vol.3.

[14] J. Nielsen and et. all, “Secure Real-Time Monitoring and Management of Smart Distribution Grid using Shared Cellular Networks,” *IEEE Wireless Communications Magazine*, 2017. [Online]. Available: arxiv: 1701.03666v1