Towards Improved Real-Time Observability of Behind-Meter PhotoVoltaic Systems: A Data-Driven Approach

Chung Ming Cheung,
Sanmukh Rao Kuppannagari
chungmin@usc.edu, kuppanna@usc.edu
University of Southern California
Los Angeles, California 90089

Rajgopal Kannan
rajgopak@usc.edu
US ARL-West
Los Angeles, California 90094

Viktor K. Prasanna
prasanna@usc.edu
University of Southern California
Los Angeles, California 90089

ABSTRACT

The stochasticity in energy supply introduced by increased PV penetration in grid distribution systems can lead to frequent and sudden supply-demand imbalances, thereby causing voltage and frequency fluctuations. Enhanced observability, in real-time, of grid attributes such as real and reactive power flows, power injections/consumption and voltages is becoming increasingly essential to ensure smooth grid operation under these new large-scale PV integration scenarios. Two significant challenges prevent grid operators from achieving this objective: 1) Computational complexity of quasi-static timeseries simulations, and 2) Partial observability of behind-meter grid assets. In this work, we take a step towards this objective by enhancing the observability of smart inverter equipped PV systems located behind AMI meters. Smart inverters convert the DC output of PV systems into real and reactive power, the ratio of which is governed by a power factor setting. This in turn is governed by the real-time voltage at its terminal, as per a pre-set volt/var curve. While it is possible to infer the power factor setting - thereby, real and reactive power injection of the PV system - at each time interval by quasi-static time series simulations, such methods are not scalable to large grid sizes. We propose to reduce the reliance on simulations by developing a deep neural network based data-driven model which predicts the time dependent power factors of the PV systems in the grid. The model uses a small amount of simulated data to learn the dependence of power factor on 1) load and apparent PV power injection (root mean square of real and reactive power) data, and 2) AMI net-load data (for BTM PV assets). Our results show that our model can predict power factors with mean absolute error of less than 6.2% under full observability scenario and under 8.6% under the partial observability scenario. Thus, our data-driven model can reduce the amount of simulations required to compute real and reactive injections from PV systems in the distribution system with or without complete information.

CCS CONCEPTS

• **Hardware** → **Smart grid.**

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

e-Energy '19, June 25–28, 2019, Phoenix, AZ, USA

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6671-7/19/06...\$15.00

<https://doi.org/10.1145/3307772.3331019>

KEYWORDS

Data-Driven Approach, Real-Time Observability, Neural Networks, Disaggregation

ACM Reference Format:

Chung Ming Cheung, Sanmukh Rao Kuppannagari, Rajgopal Kannan, and Viktor K. Prasanna. 2019. Towards Improved Real-Time Observability of Behind-Meter PhotoVoltaic Systems: A Data-Driven Approach. In *Proceedings of the Tenth ACM International Conference on Future Energy Systems (e-Energy '19)*, June 25–28, 2019, Phoenix, AZ, USA. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3307772.3331019>

1 INTRODUCTION

The penetration of PhotoVoltaic (PV) systems has been increasing significantly in smart grids in recent years [7]. This has dramatically increased the complexity of grid operations as the stochastic nature of PV generation has the potential to cause frequent voltage and frequency violations.

Recent technological developments such as Smart Inverters allow real time controllability of PVs and other renewable energy based Distributed Energy Resources (DER) via protocols such as Active Network Management (ANM) [18]. Grid operators now have the tools at their disposal to dispatch DERs in real time and mitigate violations in voltage and frequency.

However, real time controllability requires in turn real-time full observability of various grid attributes. Observability refers to the knowledge of signal measurements for every component in the grid such as real and reactive power flows over the lines, power injections or consumption at homes or PV systems in the grid, voltages at transformers etc. If the economics of the system are discounted, real-time full observability can be obtained by extensive installation of metering infrastructure to monitor each attribute. This is however not feasible in budget-constrained real world scenarios, where remaining attributes must be inferred only using partial data from the limited metering system.

In this work, we focus on improving the real-time observability of PV systems installed in distribution systems. Typically, such systems consist of PV modules connected to smart inverters installed behind the AMI meter of a residential customer, thus the name Behind-The-Meter (BTM) PV system. The data available in the best case scenario is the *apparent* power output of the PV system while in the worst case only net-load information i.e. the apparent load consumption minus the PV generation information is available.

Data driven approaches have been developed to disaggregate and infer BTM PV generation power [4, 5, 15, 16]. Such models, once trained, can infer PV generation in a small amount of time (order of milliseconds), thus, providing real-time observability [20].

However, the works mentioned above focus only on real power. Thus, they cannot be applied to infer the real and reactive power injected from PV systems connected to smart inverters. Moreover, smart inverters are configured with Volt/Var curves which enables them to configure the amount of real and reactive power injected into the grid based on the voltage at its terminal [2]. This introduces inter dependencies between the various nodes (residential customers, PV systems, transformers etc.) further complicating the problem. Quasi-static timeseries simulations can be used to model these inter dependencies. However these simulations are based on solving a set of non-linear power flow equations, they have high computational requirements from the iterative nature of gradient-based methods used to solve these equations and are unsuitable for real-time observability.

In this work, we propose a real-time data driven method to predict the *real and reactive* output of PV systems connected to smart inverters. Using either historical data (if available) or a small amount of simulated data, our model learns the dependence of power factor - a smart inverter configuration setting which determines the real and reactive output at a given time - on the following:

- (1) Load and apparent PV power injection data for systems which have separate meters for both the load and the PV system.
- (2) AMI net-load data for BTM PV assets.

The remaining sections of the paper are organized as follows. In Section 2, background information on smart inverters and recurrent neural networks are explained briefly. Section 3 lists existing related work in the literature and explain how our paper differs from them. Section 4 gives a concise problem definition and the requirements and assumptions of it. Section 5 explains the methodology used in details. Section 6 describes the datasets used and the experimental setup, then provide the results and discuss the implications. Section 7 suggests future directions our work can pursue. Finally, Section 8 concludes the contributions our work has made.

2 BACKGROUND

2.1 Smart Inverters and Volt/Var Controls

Installation of PVs into distributed grids creates a reverse power flow being injected, this causes a voltage rise in the network, which may lead to voltage unstabilities. While this problem is alleviated partially by voltage regulation facilities like capacitor banks, they incur an operation cost. Thus, researchers have proposed to use the inverters connected to PVs for regulation of voltage through absorption or injection of reactive power [14] and reduce the cost of voltage regulation.

Given the voltage at the point of the connection of the PV denoted by V_{PV} and the available apparent power capacity of the inverter denoted by Q_{avail} , the amount of capacitive or inductive reactive power to be created by the inverter is determined by a Volt/Var control function $f(V_{PV}, Q_{avail})$ to be defined according to the needs of the user. Figure 1 shows a typical droop control curve used as the Volt/Var control curve. The user defines the voltage points V_2 to V_5 in reference to pu voltage to set the range of allowable voltage. If the voltage goes beyond these values, the inverter changes its power factor to try to bring the voltage back to the allowable range.

Figure 1: Example of a Volt/Var Curve

2.2 Recurrent Neural Networks

Recurrent Neural Networks are a specific kind of neural networks specialized in learning from time series or sequential data [9]. Its main feature is a feedback loop that takes the output weights of the current step as an input to the next step. This allows information to be propagated through the steps and makes it advantageous in learning the relationships in sequential data.

One problem with using such networks is the vanishing gradient problem. When the network is unrolled, it is equivalent to a very deep neural network, and during the backpropagation step, it is easy for the gradient to vanish and thus information cannot be propagated through many layers. Researchers have thus proposed the Long Short-Term Memory nodes to remedy this problem [10].

As LSTM RNNs are not the focus of this paper, we omit the detailed explanation.

3 RELATED WORK

Improving observability or prediction of PV output in a feeder network is a well researched area. This is because there is an increasing demand for installing distributed PV, but it introduces many challenges for the utility companies due to the volatility of PV generation making it difficult to plan ahead for these volatile distributed generation in the network [6].

One popular problem to be solved in this field of research is the forecasting of future PV generation based on historical data [21]. Recent research explore how various data can be processed by different approaches to acquire forecasting of higher accuracy. Some papers use modern machine learning architecture for the forecasting, for example, [3] uses a LSTM RNN network for short term PV generation forecasting, and [19] uses a Convolutional Neural Network (CNN) for discovering cloud coverage conditions from images of the sky for the forecasting. Statistical models can also be used in PV generation forecasting [22].

A different approach to the observability problem is to disaggregate the net load signal from AMI measurements into its components. Several papers have focused particularly on the disaggregation of solar generation behind-the-meter [4, 5, 15, 16]. These papers use supervised or unsupervised methods for training a model to disaggregate the AMI measurement into behind-the-meter consumption load and PV generation components. These methods do not require historical PV generation data of the households of interest to train the models, thus making them more widely applicable to existing infrastructure.

While these methods are able to achieve high accuracy to providing information on solar generation behind-the-meter for making decision processes for smart grid operation optimization such as curtailment decisions, they do not consider smart inverters connected to PVs which are playing an increasingly important role in voltage regulation in smart grids. As these smart inverters make local control decisions that creates an effect on the active and reactive power output of PVs which in turn affects the voltage in the sub-network, it is an important to investigate how to acquire information on the smart inverter control decisions with minimal changes to the existing infrastructure. Moreover, smart inverters allow injection of both real and reactive power, but the current methods are only able to predict real power. To the best of our knowledge, our work is the first to develop a data driven model for predicting both real and reactive output of PV systems.

4 PROBLEM DEFINITION

Our task is that given a feeder network topology with the load consumption and PV generation at each node, obtain the power factor settings taken by each PV node for a subset of the full period of interest by simulations, then predict the power factor settings in the remaining period of time with a regression model. We require that

- (1) Accuracy should not be compromised. The neural network predictions should have a small error compared to the simulation results. Accuracy can be measured in terms of Mean Squared Error (MSE) and Mean Average Percentage Error (MAPE).
- (2) The network is not fully observable. That is, we do not have measurements recorded for every signal in the network. Important signals for decision processes include active and reactive power of power consumption and distributed energy sources. This is important because current AMI does not capture every single detail in a network.

Assumptions we have made include:

- Active power is known but reactive power is unknown for PV generation. Most public datasets do not have reactive power recorded.
- All PV systems are equipped with smart inverters running Volt/Var control. This is a reasonable assumption because if there are inverters with a fixed power factor, it is still representable as a Volt/Var control curve.
- Power factor of consumption loads for each customer is a constant. We assume that this is the average power factor of the consumer and that power factor does not fluctuate heavily throughout a day. This is to simply the problem to a manageable version. For future work, we would like to investigate the case of varying load power factor.

4.1 Applications

By solving the proposed problem, it can facilitate the increase of PV penetration by aiding PV capacity analysis through simulations in various ways. In this paper, we will show promising results for applying our solution for development in the following directions:

- (1) Real time power factor analysis - Performing inference with a neural network model takes a short amount of time. Once a model is trained, it can be used to find the power factor of PV output given the real time inputs quickly.
- (2) Analysis for partially observed networks - The trained model would work for similar nodes as well, meaning the analysis can be applied to other nodes without retraining of the model. This may be useful when not the full topology is known.
- (3) Inferring missing information in a partially observed network - By combining our method with a disaggregation algorithm, we can disaggregate an AMI measurement that is the aggregated net load from a household into its consumption and solar generation component in addition to the PV output power factor. Thus, our method allows greater observability in the network without incurring higher cost in terms of upgrading existing AMI infrastructure.

5 METHODOLOGY

Ultimately, our method trains a model with data from a limited number of customers with full information access to predict PV smart inverter power factors for customers with only partial information (AMI measurements of aggregated net load only). Figure 2 illustrates the work flow of our methodology. First, a simulation is performed with OpenDSS [12] so that we can obtain PV smart inverter power factor data for training the RNN model. This step takes the network topology along with the consumption load and PV generation at each node as input, and records the power factor output of PVs controlled by smart inverters. Then, a LSTM RNN is trained to predict the power factor given the same inputs as the previous step. Finally, given AMI measurements of net load (aggregation of load and generation signals) from neighboring nodes, we can use the trained model to predict PV output power factors for these node.

5.1 Simulation

For simulation, due to the lack of publicly available real world data with detailed topology and load data, we perform our simulation on the widely used IEEE 34-node test feeder [1] with consumption and generation data from other sources.

In our simulation setup, each load in the topology is assigned a corresponding load shape according to an analysis done to classify nodes as industrial, commercial, domestic, or street light loads. Details of the simulation setup is covered in Section 6.3.1.

5.2 Power Factor Analysis

For finding the power factor of each PV, a LSTM RNN model is trained to find the relationship between consumption loads, PV apparent power output, and the power factor setting. We consider two scenarios to investigate, the first one is the complete observability case, where the measurements for active and reactive power of consumption, in addition to PV apparent power output are all known. The second one is the partial observability case, where only the AMI measurement and the consumption power factor, which is assumed to be constant, are known. The AMI measurement is a aggregated signal of the consumption and generation behind the meter.

Figure 2: Overall Architecture of Our Methodology

5.2.1 The Neural Network Model. The LSTM RNN we use is a 2 LSTM layer neural network model that takes the load time series as inputs and output the corresponding PV power factor. We model the problem as a single value output regression problem. First, a sliding window is shifted over the time series to create subsequences of the time series, and the value to be predicted from the subsequence is the power factor of PV output at the end of the subsequence. Mathematically, this is written as follows.

Given D time series input each of length L in a matrix $\mathbf{X}$ of size $D \times L$, a sliding window of size W defines the lookback window for creating the data points. The objective value is the time series of power factor $\mathbf{Y}$ of length L . The sliding window is slid over the input time series such that we take each $(\mathbf{X}_{t-W:t}, Y_t)$ pair as a data point for $t = [W..L]$. The neural network model then aims to predict Y_t from $\mathbf{X}_{t-W:t}$.

To preprocess the data for the RNN, each time series is normalized to values between $[0, 1]$, and after creating subsequences with the sliding window method, data points with nan are removed from the dataset. The size of the sliding window is denoted the lookback window and is a parameter that needs to be tuned in the experimental step.

5.2.2 Full Observability Case. In our simulations, there are several PV systems simulated. The neural network model will be trained on only data from 1 PV system. Data is split by 8 : 2 into training and testing data so that the two subsets does not overlap. The network model takes the following time series as input:

- Active power of consumption load of the analyzed node
- Reactive power of consumption load of the analyzed node
- Apparent power of PV generation of the analyzed node
- Active and reactive power of consumption load of the neighboring nodes

The model learns to output the power factor of the PV output at the end of the given time series. The trained model can then be used to predict power factor for unseen data from either the same node or neighboring nodes.

5.2.3 Partial Observability Case. Most current AMI measurements only provide the aggregated measurements of consumption load and PV generation. In this case, a full simulation to acquire the PV generation power factor is not possible. With our method, only complete observability is needed for one node for the training of the model, then we propose that for the remaining neighbor nodes, we can use an unsupervised disaggregation method [5] for disaggregating the AMI measurement into load and PV signals, assuming

that the consumption load power factor is a known constant, then the disaggregated signals can be used as input into trained model.

Such a use case is useful for limiting the cost needed to improve observability of the network. For example, if we wanted to know the PV output power factor for a sub-feeder network of 20 households, instead of upgrading the AMI in all 20 households, we only need to upgrade one AMI to record full information including active and reactive power of load and PV output. The rest can then be inferred with our method.

The disaggregation is done by solving the following optimization:

$$\min_{\theta_i, \phi_i} \sum_{t=0}^T (\theta_i^T g_t + \phi_i R_t - y_{it})^2 + \lambda |\theta_i|_1 \quad (1)$$

$$\text{subject to } \theta_{ij} \geq 0 \quad \forall i, j \quad (2)$$

$$\phi_i \leq 0 \quad \forall i \quad (3)$$

, where g_t represents centroids of major consumption patterns, R_t is the solar irradiance at time t , y_{it} is the AMI measurement for user i at time t .

Details of the methods can be found in [5].

6 EXPERIMENTS

6.1 Datasets

The load dataset provided by Southern California Edison [13] is used for acquiring industrial, commercial and street light load shapes. Static load profiles from 2015 is taken for our experiments, the load profile GS-1 is used as the commercial, GS-2 as the industrial, and St-Ltng as the street light load shape. These load shapes are available in 1-hour time intervals.

For domestic load shapes and PV generation load shapes, the data is taken from the dataset provided by Pecan Street Inc [11]. In order to acquire different domestic load shapes for each node, we extracted 150 customer load shape data and their PV generation data in 2015, then aggregated every 5 customers data to create 30 load shapes for consumption and PV generation. These load shapes are available in 1-min time intervals, and it was downsampled to match with the Southern California Edison dataset.

For topology of the network, we used the IEEE 34-node test feeder. [1]

Solar irradiance data is used for the disaggregation step of the partial observability case. We extract irradiance data from the National Solar Radiation Database [8] provided by the National Renewable Energy Laboratory. The dataset covers the whole US in resolution

of $4\text{km} \times 4\text{km}$ grids, we extract data from Austin, Texas to match with the data from the Pecan Street dataset.

6.2 Evaluation Metrics

We use Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Squared Area (MAPE) as the evaluation metrics. The power factor found by our model is compared to the values from the simulation. RMSE puts a heavier weight on outlier points while MAE measures the average error for all points equally. MAPE is a normalized metric that measures the percentage error and thus is not affected by the amplitude of the values. However, MAPE is very sensitive to small values, for instance, if the true value was close to 0, the percentage error would be close to infinite. Thus, we ignore the time intervals where the true value is below 0.2, this consists of less than 2.3% of the whole test data so that the results are still credible. The definitions of each metric are as follows, where $\mathbf{x}$ denotes the true values, $\hat{\mathbf{x}}$ denotes the predicted values, and N is the total number of values:

$$MAE = \frac{1}{N} \|\mathbf{x} - \hat{\mathbf{x}}\|_{l1} \quad (4)$$

$$RMSE = \sqrt{\frac{1}{N} \|\mathbf{x} - \hat{\mathbf{x}}\|_{l2}^2} \quad (5)$$

$$MAPE = \frac{1}{N} \left\| \frac{\mathbf{x} - \hat{\mathbf{x}}}{\mathbf{x}} \right\|_{l1} \quad (6)$$

6.3 Experimental Setup

For our experiments, we took 60 days of real world data from Jan 1st to March 1st 2015 from various sources to perform a quasi-static time series simulation to obtain ground truth data, then train a neural network model under various conditions to test its accuracy in finding the PV output power factor.

6.3.1 Simulation. To obtain the ground truth power factors of PV output, quasi-static time series analysis is used. OpenDSS is used for performing the simulation on the IEEE 34-node test feeder model. The feeder data only provides a snapshot of the network. We extend the model to a time series model with data from the sources described in Section 6.1.

We follow the analysis used by [17] to distinguish the loads on the topology to industrial, commercial, domestic loads and street lights by their apparent power value. For industrial, commercial and street light consumption load shapes, we take the provided by Southern California Edison [13] in 2015. For consumption and PV generation load shapes, we take individual load shapes from customers in provided by the Pecan Street dataset [11] in 2015, and aggregate 5 customers to produce a load shape for each node. We picked the 860-836 bus in the 34-node feeder model for installing PVs with smart inverters onto each of the 3 nodes on that bus. All 3 nodes are domestic loads. The smart inverters are all controlled by a Volt/Var control with the Volt/Var curve illustrated in Figure 1, where $V2 = 0.98\text{pu}$, $V3 = V4 = 1.0\text{pu}$, $V5 = 1.02\text{pu}$. We label the three nodes as Node 0, Node 1 and Node 2. Table 1 shows the configuration of each of the nodes.

Node	Load kW	Load kVar	Load Power Factor	PV kVA
0	30	15	0.667	5
1	10	6	0.625	5
2	42	22	0.656	10

Table 1: Configurations of Nodes in Simulation

	RMSE	MAE	MAPE
Without Neighbors	0.0835	0.054830101	9.57%
With Neighbors	0.0783	0.0434	7.20%

Table 2: Testing Errors for Model With or Without Neighbor Inputs for Complete Observability Case

After setting up the circuit, the simulation is ran in 10 second interval timesteps for 60 days. The short interval ensures control iterations are performed frequently enough that a convergence can be reached quickly every time it needs to be changed.

6.3.2 Power factor analysis. For the neural network model used in power factor analysis, we train a 2-layer LSTM RNN model for the regression. Hyperparameters that are tuned are

- RNN Layer size
- Drop out
- Sliding window size

5-fold cross-validation is used to pick the optimal hyperparameters, the best hyperparameters are used to train a RNN with early stopping.

For our first experiment of the complete observability case, the model trained with data from Node 0. The data is split by 8 : 2 into training and testing datasets, and the model is evaluated with the testing dataset. We first compare the performance of the model with using only data from Node 0 as input, to using data from all neighboring nodes as input.

For the second experiment of the complete observability case, we train the model for Node 0, then test the performance of the model on the other neighboring nodes Node 1 and Node 2.

Finally, we carry out the experiment for the partial observability case. Instead of assuming that separated signals for load and PV generation is available, we first perform a disaggregation of the AMI measurements using the method in [5] which only requires the AMI measurements and irradiance estimate of the region. These disaggregations are then used as inputs and the above two experiments are repeated.

In the following results sections, results are plotted at 10 or 15 minute intervals instead of the 10 second simulation time interval for ease of formatting and interpretation of results. The downsampling is done by taking the average of values.

6.4 Results: Complete Observability Case

Table 2 shows the prediction error for with or without neighbor inputs. Figure 3 and 4 compare the predicted power factor and the simulated power factor of Node 0 PV output on the testing dataset. We can see that the use of neighbor inputs improves the accuracy of the model, as using neighbor data lowers the error in terms of

Figure 3: Comparison between Neural Network Outputs and Simulation Outputs on Node 0 with no neighbor inputs for Complete Observability Case

Figure 4: Comparison between Neural Network Outputs and Simulation Outputs on Node 0 with neighbor inputs for Complete Observability Case

Node	RMSE	MAE	MAPE
1	0.0745	0.0433	7.82%
2	0.0873	0.0617	12.50%

Table 3: Testing Errors for Node 1 and 2 for Complete Observability Case

Node	RMSE	MAE	MAPE
0	0.1178	0.0856	16.14%
1	0.1016	0.0664	12.28%
2	0.1045	0.0635	12.62%

Table 4: Testing Errors for Node 0, 1 and 2 for Partial Observability Case

all metrics. This shows that the controls taken by smart inverters is dependent on behavior of both the node itself and neighboring nodes, thus giving all the inputs of local environment allows more accurate predictions.

Table 3 shows the prediction error for the two neighboring nodes, Node 1 and Node 2. The predictions are made with the model trained on Node 0's simulated power factor results along with Node 1 and Node 2 inputs. We can see that the errors for Node 2 is higher than the testing results on Node 0. Figure 5 and 6 show the power factor predictions by the model compared to the simulated values. We can see that even though the results are not as good as for Node 0, we still get good accuracy on the predictions of power factors.

In general, we see that in the complete observability case, power factors can be predicted at a high accuracy of MAE 0.434 using a RNN model trained with simulated data for a particular node, showing it can substitute Quase-static timeseries simulations for subsequent analysis on that node. Also, at the cost of a higher MAE 0.0617, the model can be used for neighboring nodes as well.

6.5 Results: Partial Observability Case

For the partial observability case, we first disaggregated the AMI measurements of Node 1 and Node 2 into their consumption load and solar generation components. Figure 7 and 8 illustrates a short subsequence showing how the disaggregated signals are close enough to the true values so that they can be used for the next step of the experiment.

For Node 0, the true load and solar data are available but the neighbor data inputs of Node 1 and 2 are disaggregation results. From Figure 9, 10, 11, we can see that some accuracy dropped compared to the complete observability case where the dotted line representing LSTM network prediction deviates from the solid line ground truth more. This can be seen numerically from Table 4 which has higher error values than Table 2 and 3. This is expected because we are using estimated load and solar values from unsupervised disaggregation instead of the true values, however, as seen from the figures, the predictions are still quite close to the true values. This shows that our methods when used together with a disaggregation

Figure 5: Comparison between Neural Network Outputs and Simulation Outputs on Node 1 for Complete Observability Case

Figure 6: Comparison between Neural Network Outputs and Simulation Outputs on Node 2 for Complete Observability Case

Figure 7: Subsequence of Disaggregated Load compared to True Values for Node 1

Figure 8: Subsequence of Disaggregated Solar compared to True Values for Node 1

algorithm can complete missing information in a partially observed network without needing extra instrumentation.

7 FUTURE WORK

This paper showed promising results for modelling the complex physical relationships underlying quasi-static time series simulations with a neural network. For the neural network model to better

aid in PV integration analysis, it would be useful for it to take PV configurations as an input instead of the model being trained for a specific configuration.

For example, a model can take the Volt/Var curve parameter as an input, and predict power factor behavior for different Volt/Var curve setups. This facilitates quick analysis for finding optimal Volt/Var curve configuration in comparison to using simulations.

Figure 9: Comparison between Neural Network Outputs and Simulation Outputs on Node 0 for Partial Observability Case

Figure 10: Comparison between Neural Network Outputs and Simulation Outputs on Node 1 for Partial Observability Case

Figure 11: Comparison between Neural Network Outputs and Simulation Outputs on Node 2 for Partial Observability Case

Another extension that can be made is to address the assumption of using a constant power factor for consumer load signals. A time-dependent function can be integrated into the model for modelling the power factor of the loads, this is because many consumption patterns exhibit daily seasonality, for example, lighting are used at nighttime, fridge are turned on consistently, etc.

8 CONCLUSION

In this paper, we have proposed a LSTM RNN model that can be trained with simulated data from quasi-static time series simulations to predict power factor of PV generation output controlled by smart inverters with Volt/Var Controls. The model takes the consumption load and PV generation of itself and neighboring nodes as input, or it can take disaggregated signals from net load of AMI measurements as inputs. Our model predicts power factors with a

mean absolute error under 6.2% for all 3 nodes in our experiments if full data is available, or under 8.6% if disaggregated data is used. The results show that when a network is only partially observed, meaning that not all consumption and generation of all nodes are measured, our model is still capable of replacing or complementing simulations to analyze PV power factors.

ACKNOWLEDGMENTS

This work has been supported in part by the U.S. National Science Foundation under EAGER award number CNS-1637372 and the U.S. Department of Energy (DoE) under award number DE-EE0008003.

REFERENCES

- [1] [n. d.]. IEEE PES AMPS DSAS Test Feeder Working Group Resources. <http://sites.ieee.org/pes-testfeeders/resources/>. Accessed: 2019-03-04.
- [2] 2012. Common Functions for Smart Inverter. Online: <http://www.epri.com/abstracts/Pages/ProductAbstract.aspx?ProductId=1026809>, <http://www.epri.com/abstracts/Pages/ProductAbstract.aspx?ProductId=1026809>
- [3] An Chen, WenZhan Song, Fangyu Li, and Javad Mohammadpour Velni. 2018. Distributed Cooperative Energy Management in Smart Microgrids with Solar Energy Prediction. In *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 1–6.
- [4] Dong Chen and David Irwin. 2017. Sundance: Black-box behind-the-meter solar disaggregation. In *Proceedings of the Eighth International Conference on Future Energy Systems*. ACM, 45–55.
- [5] Chung Ming Cheung, Wen Zhong, Chuanxiu Xiong, Ajitesh Srivastava, Rajgopal Kannan, and Viktor K Prasanna. 2018. Behind-the-Meter Solar Generation Disaggregation using Consumer Mixture Models. In *2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. IEEE, 1–6.
- [6] A. Dubey, S. Santoso, and A. Maitra. 2015. Understanding photovoltaic hosting capacity of distribution circuits. In *2015 IEEE Power Energy Society General Meeting*. 1–5. <https://doi.org/10.1109/PESGM.2015.7286510>
- [7] David J Feldman, Jack Hoskins, and Robert M Margolis. 2017. *Q2/Q3 2017 Solar Industry Update*. Technical Report. National Renewable Energy Lab.(NREL), Golden, CO (United States).
- [8] Aron Habte, Manajit Sengupta, and Anthony Lopez. 2017. *Evaluation of the National Solar Radiation Database (NSRDB): 1998-2015*. Technical Report. NREL (National Renewable Energy Laboratory (NREL), Golden, CO (United States)).
- [9] Simon Haykin. 1994. *Neural networks*. Vol. 2. Prentice hall New York.
- [10] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [11] Chris Holcomb. 2012. Pecan Street Inc.: A test-bed for NILM. In *International Workshop on Non-Intrusive Load Monitoring*, Pittsburgh, PA, USA.
- [12] Electric Power Research Institute. [n. d.]. OpenDSS electrical power system simulation tool. <http://smartgrid.epri.com/SimulationTool.aspx>. Accessed: 2019-03-04.
- [13] Electric Power Research Institute. [n. d.]. Southern California Edison 2015 Static Load Profiles. <https://www.sce.com/regulatory/load-profiles/2015-static-load-profiles>. Accessed: 2019-03-04.
- [14] P. Jahangiri and D. C. Aliprantis. 2013. Distributed Volt/VAR Control by PV Inverters. *IEEE Transactions on Power Systems* 28, 3 (Aug 2013), 3429–3439. <https://doi.org/10.1109/TPWRS.2013.2256375>
- [15] Emre C Kara, Ciaran M Roberts, Michaelangelo Tabone, Lilliana Alvarez, Duncan S Callaway, and Emma M Stewart. 2018. Disaggregating solar generation from feeder-level measurements. *Sustainable Energy, Grids and Networks* 13 (2018), 112–121.
- [16] Emre Can Kara, Michaelangelo Tabone, Ciaran Roberts, Sila Kiliccote, and Emma M Stewart. 2016. Estimating behind-the-meter solar generation with existing measurement infrastructure. In *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, 259–260.
- [17] Barry A Mather. 2012. Quasi-static time-series test feeder for PV integration analysis on distribution systems. In *2012 IEEE Power and Energy Society General Meeting*. IEEE, 1–8.
- [18] Milana Plečaš, Han Xu, and Ivana Kockar. 2017. Integration of energy storage to improve utilisation of distribution networks with active network management schemes. *CIREN-Open Access Proceedings Journal* 2017, 1 (2017), 1845–1848.
- [19] Yuchi Sun, Gergely Szűcs, and Adam R Brandt. 2018. Solar PV output prediction from video streams using convolutional neural networks. *Energy & Environmental Science* 11, 7 (2018), 1811–1818.
- [20] Michaelangelo Tabone, Sila Kiliccote, and Emre Can Kara. 2018. Disaggregating solar generation behind individual meters in real time. In *Proceedings of the 5th Conference on Systems for Built Environments*. ACM, 43–52.
- [21] Aidan Tuohy, John Zack, Sue Ellen Haupt, Justin Sharp, Mark Ahlstrom, Skip Dise, Eric Grimit, Corinna Mohrlen, Matthias Lange, Mayte Garcia Casado, et al. 2015. Solar forecasting: methods, challenges, and performance. *IEEE Power and Energy Magazine* 13, 6 (2015), 50–59.
- [22] Can Wan, Jin Lin, Yonghua Song, Zhao Xu, and Guangya Yang. 2017. Probabilistic forecasting of photovoltaic generation: An efficient statistical approach. *IEEE Transactions on Power Systems* 32, 3 (2017), 2471–2472.