Calculating PV Hosting Capacity in Low-Voltage Secondary Networks Using Only Smart Meter Data

Joseph A. Azzolini, Matthew J. Reno, and Jubair Yusuf
Electric Power Systems Research
Sandia National Laboratories
Albuquerque, NM, USA

Samuel Talkington and Santiago Grijalva
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA, USA

***Abstract*—Residential solar photovoltaic (PV) systems are interconnected with the distribution grid at low-voltage secondary network locations. However, computational models of these networks are often over-simplified or non-existent, which makes it challenging to determine the operational impacts of new PV installations at those locations. In this work, a model-free locational hosting capacity analysis algorithm is proposed that requires only smart meter measurements at a given location to calculate the maximum PV size that can be accommodated without exceeding voltage constraints. The proposed algorithm was evaluated on two different smart meter datasets measuring over 2,700 total customer locations and was compared against results obtained from conventional model-based methods for the same smart meter datasets. Compared to the model-based results, the model-free algorithm had a mean absolute error (MAE) of less than 0.30 kW, was equally sensitive to measurement noise, and required much less computation time.**

***Index Terms*—advanced metering infrastructure (AMI), data-driven analysis, hosting capacity analysis, smart meter data.**

I. Introduction

Residential solar photovoltaic (PV) systems are becoming increasingly prevalent in distribution systems. To ensure that new PV installations can be safely and reliably accommodated by the grid, a locational PV hosting capacity (HC) analysis can be conducted to determine the maximum PV capacity that can be installed at various locations on the grid before operational constraints are violated or upgrades are required [1]. This type of "locational" HC analysis (sometimes referred to as integration capacity analysis, or ICA) resembles the detailed studies required by some PV interconnection requests. In contrast, streamlined [2] and stochastic HC methods [3] evaluate the feeder-level impacts of various PV deployment scenarios on the medium-voltage (MV) networks, which are more useful for planning tasks than facilitating interconnection procedures. Residential PV systems are interconnected with the distribution grid at low-voltage (LV) secondary network locations, and models of these networks are often over-simplified or non-existent; loads may be lumped together and connected directly to the MV primary network, meaning the service transformers and secondary network conductors are missing. This lack of detail in LV network modeling can lead to significant errors in HC results [4]. Even when detailed distribution grid models are available and free of other common errors that impact HC results [4], model-based locational HC are computationally intensive and time-consuming [5].

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number 38426. The views expressed herein do not necessarily represent the views of the U.S.Department of Energy or the United States Government. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

With the widespread adoption of advanced metering infrastructure (AMI), including smart meters installed at LV network locations, many data-driven methods have since been proposed to improve grid models [6]. Geographic Information Systems (GIS) data has also been leveraged to improve secondary modeling for PV HC analysis by applying supervised machine learning and logistic regression to predict network topologies and conductor types [7]. Data-driven distribution analysis methods have also been proposed as faster, model-free alternatives. Specifically, some methods leverage smart meter data to estimate the feeder-level PV HC using Bayesian statistical inference [8] or supervised univariate regression modeling [9]. To determine the HC of LV secondary networks with model-free methods, the voltage sensitivities to active power injections must be extracted. In [10], these sensitivities are determined from smart meter data using a combination of linear regression and clustering methods, while [11] trains a Deep Neural Network (DNN) model to learn the sensitivities from smart meter data and predict voltage impacts from PV injections for HC analyses.

However, existing methods for model-free PV HC analysis require measurements from multiple smart meters and information about the network topology or knowledge of which customers are fed through the same service transformer. These inputs may not always be available and are susceptible to the same common errors that impact the accuracy of model-based HC results [4]. Yet, measurements from just a single LV network location are often sufficient to approximate the linearized voltage sensitivity at that location [12], which suggests that PV HC could be calculated for any residential customer location with smart meter data available.

To evaluate this hypothesis, a model-free HC algorithm was developed in this paper to extract a linearized voltage sensitivity for any LV secondary location with smart meter data available and to predict PV impacts that determine HC. Specifically, the contributions of this paper include: 1) a novel model-free algorithm to calculate the voltage-constrained PV HC at any LV network location, 2) validation against conventional model-based HC results, and 3) a noise sensitivity analysis for each of the two methods.

II. METHODS

A novel algorithm was developed to calculate the PV HC for any residential behind-the-meter (BTM) PV location on a LV secondary network using only measurements of the voltage, real power, and reactive power at that location as inputs. While this algorithm can be iteratively applied to evaluate multiple locations, each location is handled independently, meaning it is not intended to determine how much total PV can be installed inside a given secondary network or across an entire feeder. Rather, the intent is that the proposed algorithm could facilitate residential PV interconnection requests by determining ahead of time the maximum PV size that can be accommodated at any potential BTM PV location.

Experimentally, the proposed algorithm was evaluated on two different utility datasets and the results were compared to benchmark model-based locational HC results obtained using the same smart meter data as inputs. The analyses were then repeated for both methods under increasing levels of measurement noise to assess their robustness.

A. *Model-Free HC Algorithm*

The proposed algorithm aims to: 1) estimate the sensitivity of voltage magnitude to changes in real power at a customer location from its smart meter data, and 2) apply that sensitivity to predict the potential voltage impacts of PV power injections and determine the HC. These two components of the proposed algorithm are presented in Figure 1 and Figure 2, respectively. In Figure 1, the algorithm starts by taking in historical real power, reactive power, and voltage measurements (i.e., P, Q, and V) from the smart meter and calculating the apparent power consumption, S, and the power factor, PF. Next, the approximate derivatives are computed for each variable by taking the difference between adjacent data points in time.

The algorithm then applies two distinct methods to characterize the voltage sensitivity, each with uniquely tuned pre-processing steps to adjust for the potential multicollinearity of real and reactive power impacts on voltage when power factor remains relatively constant. In the left branch of Figure 1, the algorithm applies the linear surface fit in (1) to extract separate voltage sensitivity coefficients for changes in real power (σ_P) and reactive power (σ_Q). Since HC analysis considers PV systems that operate at unity PF, only the σ_P coefficient is of interest. To focus on time points where σ_P would be most apparent, the fit is applied after the input data are filtered to remove the bottom 25% of $|\Delta P|$ values, the bottom 10% of $|\Delta PF|$ values, and the top 1% of $|\Delta V|$ values. In the right branch of Figure 1, the algorithm applies the linear curve fit in (2) to extract the voltage sensitivity coefficient for changes in apparent power (σ_S). The fit is applied after the input data are filtered to remove the bottom 50% of $|\Delta P|$ values and the top 5% of $|\Delta V|$ values. Note that the threshold percentages for the filtering steps were manually tuned for each method.

$$\Delta V = c_1 + (\sigma_P * \Delta P) + (\sigma_Q * \Delta Q) \quad (1)$$

$$\Delta V = c_2 + (\sigma_S * \Delta S) \quad (2)$$

After the sensitivity has been estimated by each method, the algorithm sets σ_{Final} to either σ_P or σ_S. In theory, the σ_P is more precise, but in practice, constant power factor load may cause ΔP and ΔQ to be too correlated for (1) to accurately parse out σ_P and σ_Q from the smart meter data available. Thus, σ_S can be selected as a hedge for those scenarios, since the impacts of ΔP and ΔQ are captured in ΔS; essentially, this method assumes the external impedance at the LV network location has an X/R ratio of 1, such that $\Delta S \approx \Delta P \approx \Delta Q$ in (2). Since X/R ratios in distribution systems are typically low, this method provides a reasonable calibration check. In this paper, σ_S is only selected when σ_P disagrees by more than 30%, since σ_P was prone to extreme outliers.

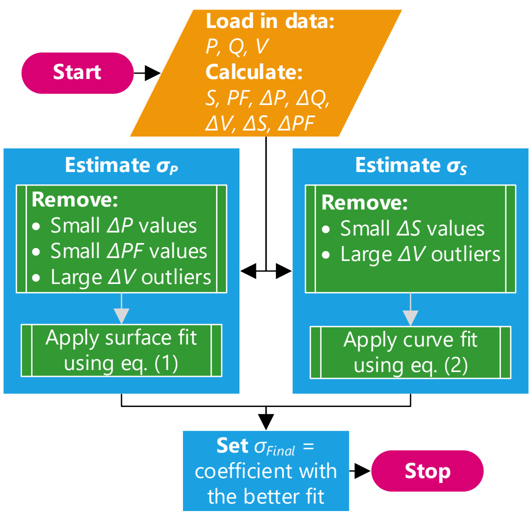

Figure 1. Procedure for estimating the voltage sensitivity at an LV secondary network location from its smart meter data.

Once σ_{Final} is set, the algorithm moves on to the procedure described in Figure 2, where σ_{Final} and the original smart meter voltage measurements are used to calculate the maximum real power injection, kW_{max}, that could have been accommodated at each time point, t, before exceeding the upper bound voltage limit, V_{limit}, using (3):

$$kW_{max}(t) = (V_{limit} - V(t))/\sigma_{Final} \quad (3)$$

Here, V_{limit} was set to 1.05 V per unit.

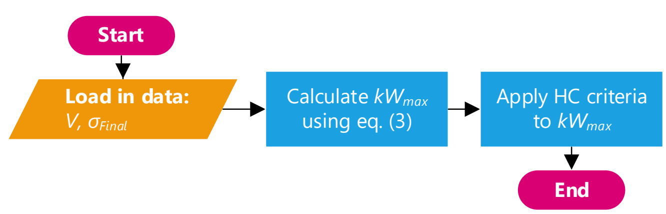

Figure 2. Procedure for applying the voltage sensitivity coefficient to determine HC.

The final step of the algorithm in Figure 2 determines the PV HC from kW_{max}. The phrase "HC criteria" was intended to be general, as a variety of HC definitions could be accommodated. For example, HC is often calculated based on some set of worst-case PV conditions, so one option would be to simply take the minimum value of kW_{max} such that if a PV system was outputting that kW value at all time points, the limits would never be exceeded. However, if that minimum value occurred at night, it would not make sense to call that the PV HC since real power is only generated during the day. For this paper, HC is defined as the daytime minimum of kW_{max}, where "daytime" is between 09:00 to 15:00 each day to focus on expected peak PV production hours.

B. *Model-Based HC Algorithm*

The model-based HC algorithm implemented in this work is similar to conventional methods, such as the CYME ICA module [13], but contains modifications for calculating the locational HC as a time-series to provide a more appropriate benchmark for evaluating the model-free algorithm results. Instead of analyzing just a handful of worst-case scenarios, the model-based HC algorithm in this work applies yearlong quasi-static time-series (QSTS) simulations to calculate kW_{max} at every time point for each customer location, then applies the same daytime minimum criteria as the model-free algorithm to determine the PV HC.

To address the additional computational burden of this modification, a time-wise linear interpolation approach was adopted. For each location, full yearlong QSTS simulations were conducted for several different values of constant PV power injections (0, 5, 10, 20, and 40 kW). Then, when the QSTS simulations concluded, kW_{max} was interpolated from the results for every time point. An example of this approach is presented in Figure 3 for customer location 1 at *t=12:00 p.m.* on a randomly selected day. The QSTS results for each PV injection are represented by the blue diamond markers and V_{limit} is the horizontal black dotted line at 1.05 per unit. Since the response is highly linear, interpolation can be applied to determine the value of $kW_{max}(t)$ = 27.6 kW. This process would then be repeated to calculate kW_{max} for all time points.

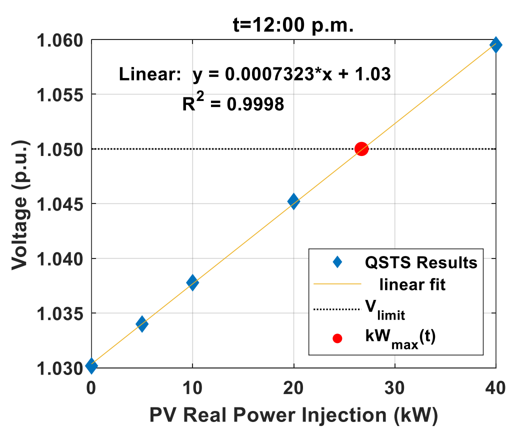

Figure 3. Example of linear interpolation approach used in the model-based HC analysis.

C. *Test Circuit and Smart Meter Datasets*

A modified version of the EPRI Ckt5 test circuit, shown in Figure 4, was utilized in this work. This test circuit has 1,379 customer locations, denoted as "potential BTM PV" locations in Figure 4, for which the HC will be evaluated. Each potential BTM PV location is a residential single-phase 240V connection at the end of service line that varies from 40 to 180 ft from the service transformer. Each location was assigned a real power and reactive power time-series profile selected from one of two anonymous utility smart meter datasets: Dataset 1 and Dataset 2. Note that the model-free HC algorithm was developed using Dataset 1 and tested on both datasets, without seeing Dataset 2 ahead of testing. In each dataset, the smart meters recorded measurements at 15-minute intervals for a full year (35,040 time points per measurement). Once the profiles were assigned, a baseline QSTS simulation was conducted for Dataset 1 and Dataset 2 to generate semi-synthetic voltages for each location. This step ensures that the input data for the model-free and model-based HC algorithms are identical.

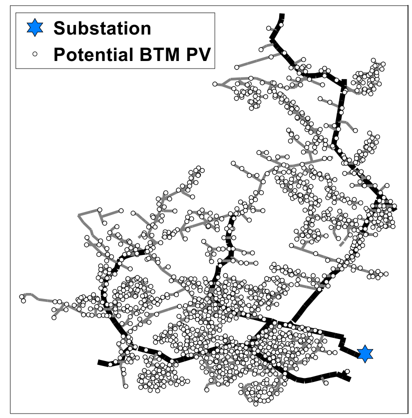

Figure 4. Circuit plot of the modified EPRI Ckt5 test feeder with markers at all potential PV locations.

III. RESULTS

While the proposed model-free algorithm relies on a linear approximation of voltage sensitivity, the true voltage sensitivity at a location varies over time. To assess the potential impact of this variation on the accuracy of the proposed algorithm, the model-based HC results were post-processed to calculate the σ_{Pmodel} values for the whole year at each customer location. As an example, the σ_{Pmodel} time-series for location 1 (Dataset 1) is shown in the top subplot of Figure 5, calculated using (4):

$$\sigma_{Pmodel} = (V_{limit} - V)/kW_{max} \qquad (4)$$

For reference, this model-based sensitivity calculation method is similar to the "perturb and observe" method in [14]. The average σ_{Pmodel} value of 7.46e-4 was then used in (3) instead of σ_{Final} to calculate kW_{max}, and the results are shown in the bottom subplot of Figure 5. Compared to the actual kW_{max} values, the average σ_{Pmodel} resulted in a mean absolute error over all time points (MAE_t) of 0.40 kW for location 1. After repeating this process for all locations, the average MAE_t was found to be 0.33 kW.

Next, a similar analysis was implemented to assess the accuracy of the two sensitivity estimation methods used in the model-free HC algorithm in Figure 1. For each location, kW_{max} was calculated separately using the σ_P and σ_S, as well as using the final combined σ_{Final} selection. These kW_{max} results were then compared to the model-based kW_{max} results, and the MAE_t was calculated for each location. The distributions of MAE_t values per location associated with each model-free sensitivity approximation are presented as boxplots in Figure 6. As anticipated, using σ_P resulted in much lower errors than σ_S for most locations but had several extreme outliers. The combination of the two methods, σ_{Final}, performed the best by correctly selecting σ_S instead of σ_P when necessary.

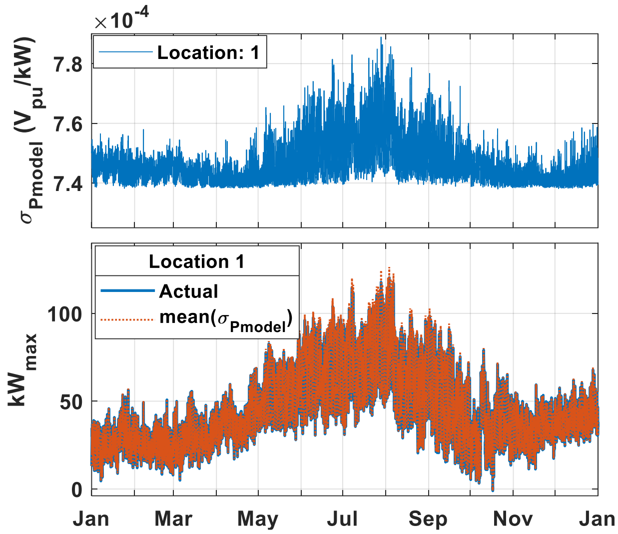

Figure 5. Model-derived σ_{Pmodel} variation through time for location 1 (top) and impact of using mean(σ_{Pmodel}) to calculate kW_{max}.

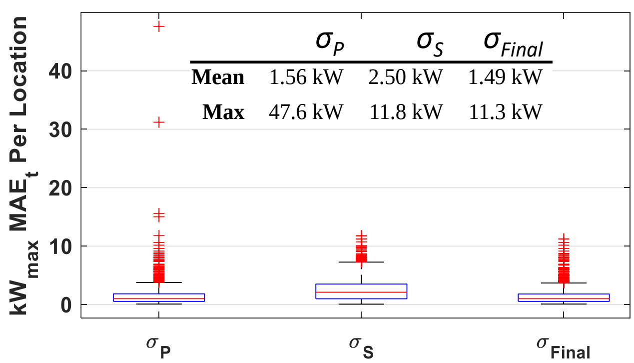

Figure 6. Effect of model-free sensitivity estimations on kW_{max}.

Figure 7 presents a comparison of results between the model-free and model-based HC algorithms for each smart meter dataset. In each plot, the blue circles represent the benchmark model-based (x-coordinate) and model-free (y-coordinate) HC result for each of the 1,379 customer locations. The orange diagonal line represents the ideal outcome in which the model-free algorithm returns the same value as the model-based algorithm. In general, many of the results for both datasets appear to be near the diagonal, which suggests the model-free algorithm provided an accurate result for those locations. It should also be noted that the computational time required for the model-free algorithm is significantly less than that of the model-based algorithm; the model-free HC results were generated within minutes, whereas the model-based results required multiple days of simulations.

Table 1 provides a quantitative overview of the model-free algorithm's performance. For Dataset 1 and 2, the MAE_{HC} of all locations was 0.26 kW and 0.29 kW, respectively. In other words, the proposed algorithm was able to calculate HC within 300 W—or roughly one PV module—of conventional model-based methods, on average. Similarly, the model-free algorithm results were within 1 kW of the model-based results for 96.6% and 95.8% of the customer locations, respectively, for Dataset 1 and 2. The model-free algorithm did struggle to determine an accurate result for some locations, resulting in errors up to 2.84 kW and 7.65 kW, respectively, for Dataset 1 and 2. Since the model-free algorithm utilizes statistical regressions, confidence intervals can also be calculated for each result that can be used to flag locations with poor fits.

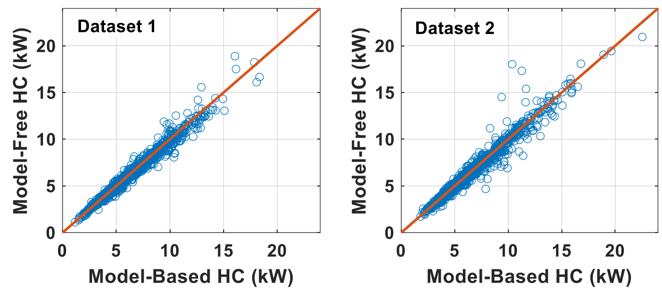

Figure 7. Comparison of model-free vs. model-based HC results.

TABLE 1. SUMMARY OF MODEL-FREE HC PERFORMANCE FOR ALL LOCATIONS IN FIGURE 4

HC Metric	Dataset 1	Dataset 2
MAE_{HC}	0.26 kW	0.29 kW
Max. Error	2.84 kW	7.65 kW
Locations <1kW Error	96.6%	95.8%

Since the proposed algorithm and the model-based HC algorithms both require smart meter data as inputs, it is also important to evaluate the impact of measurement noise on the accuracy of their results. Therefore, the HC results for each dataset were re-calculated after various levels of noise were added to the smart meter measurements. For the voltage measurements, Gaussian noise masks were generated for each location and each dataset using the nominal voltage base as the mean, and the standard deviation was set according to various meter accuracy classes. For example, a revenue meter with an accuracy class of 0.2 requires all measurements to be accurate within ±0.2% of the true value [15]. So, since Gaussian noise follows a normal distribution, setting the standard deviation of noise to 1/3 of the meter accuracy rating (0.067% for the 0.2 class) ensures that nearly all (99.7%) of the generated noise is within the accuracy range of that meter. For clarity, the noise sensitivity results are presented in terms of equivalent meter class accuracy, or simply "Meter Class".

For the model-based HC algorithm, the voltage noise masks were added to the QSTS results before the interpolation step was initiated. For the model-free HC algorithm, the voltage noise was added before any of the data filter steps were initiated. In addition, the model-free HC algorithm was also evaluated after the noise was added to the real and reactive power measurements. In this case, the same noise generation procedure was used but the mean value of the noise was set equal to the original measurement values. The results from each of these three cases in terms of MAE_{HC} are presented in Figure 8. Figure 9 shows the impact of noise on the ability of each HC algorithm to remain within 1 kW accuracy.

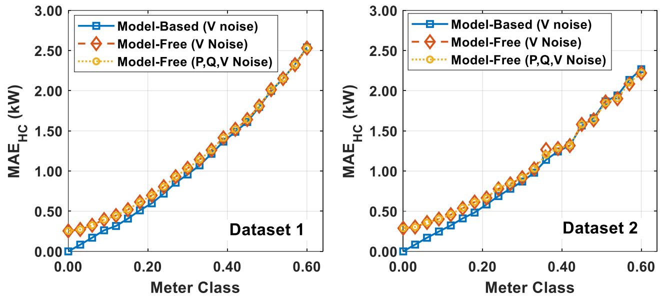

Figure 8. Impact of smart meter accuracy on MAE_{HC}.

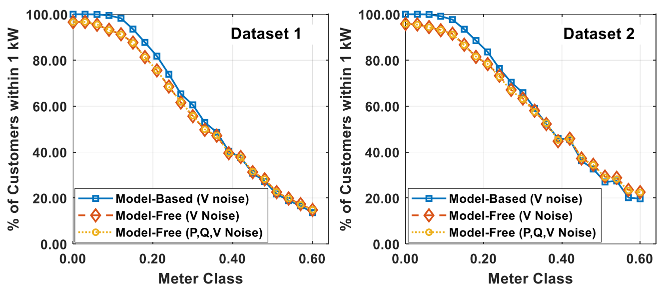

Figure 9. Impact of smart meter accuracy on the HC accuracy metric.

The results in Figure 8 and Figure 9 suggest that both HC methods have a high degree of sensitivity to measurement noise. However, since model-based HC analysis is already widely accepted, the fact that the proposed model-free algorithm responds so similarly to measurement noise provides validation, nonetheless. Thus, the differences between the model-free results and model-based results (ΔHC) at each location were calculated at each noise level, but the distribution of differences was very similar at each level; Figure 10 presents these distributions as histograms at two noise levels for Dataset 1 and Dataset 2.

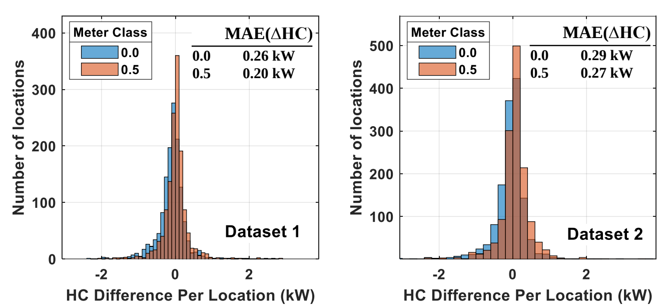

Figure 10. Histograms of the differences between model-free and model-based locational HC results with and without measurement noise.

In practice, even the most robust model-free or model-based analysis methods are only ever as accurate as the input data allows. There is also evidence that suggests that the noise levels analyzed in this paper may have been overly conservative, where the standard deviation of noise for class 0.5 meters was found to be 0.07% [16] compared to 0.167% used here.

IV. Conclusion

In this paper, a novel algorithm was presented that calculates the local voltage-constrained PV HC for any LV secondary network location and requires only smart meter measurements for that location. The algorithm was tested on actual smart meter datasets from two different utilities and was found to be accurate within 0.30 kW of conventional model-based HC results, on average, while requiring just a fraction of the computational time. The proposed algorithm was also observed to be equally sensitive to measurement noise as the benchmark model-based HC algorithm. Even as the level of noise injections increased, errors between the model-free and model-based HC results remained consistent. Overall, these results suggest that the proposed model-free algorithm provides a suitable alternative to existing model-based methods for calculating the voltage-constrained PV HC on LV secondary networks.

References

[1] S. Stanfield, Y. Zackai, and M. McKerley, "Key Decisions for Hosting Capacity Analyses," IREC, 2021.

[2] M. Rylander, J. Smith, and W. Sunderman, "Streamlined Method for Determining Distribution System Hosting Capacity," *IEEE Transactions on Industry Applications*, vol. 52, no. 1, pp. 105-111, 2016.

[3] EPRI, "Stochastic Analysis to Determine Feeder Hosting Capacity for Distributed Solar PV," EPRI, Technical Report 1026640, 2012.

[4] J. Azzolini *et al.*, "Improving Behind-the-Meter PV Impact Studies with Data-Driven Modeling and Analysis," presented at the IEEE Photovoltaic Specialists Conference (PVSC), 2022.

[5] J. Baranowski *et al.*, "Electric Program Investment Charge (EPIC) Final Report," SDGE, 2017.

[6] M. Reno *et al.*, "IMoFi - Intelligent Model Fidelity: Physics-Based Data-Driven Grid Modeling to Accelerate Accurate PV Integration Final Report," United States, 2022.

[7] W. Wang *et al.*, "Supervised Learning for Distribution Secondary Systems Modeling: Improving Solar Interconnection Processes," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 948-956, 2022.

[8] R. Lliuyacc-Blas, S. O. Nyberg, M. A. S. T. Ireshika, M. L. Kolhe, and P. Kepplinger, "PV Hosting Capacity Estimation in Low Voltage Feeders Through Bayesian Statistical Inference," in *2022 12th International Conference on Power, Energy and Electrical Engineering (CPEEE)*, 25-27 Feb. 2022 2022, pp. 250-255, doi: 10.1109/CPEEE54404.2022.9738661.

[9] A. T. Procopiou, M. Z. Liu, L. F. Ochoa, T. Langstaff, and J. Harding, "Smart meter-driven estimation of PV hosting capacity," in *CIRED 2020 Berlin Workshop (CIRED 2020)*, 22-23 Sept. 2020 2020, vol. 2020, pp. 128-131, doi: 10.1049/oap-cired.2021.0287.

[10] G. Valverde, T. Zufferey, S. Karagiannopoulos, and G. Hug, "Estimation of voltage sensitivities to power injections using smart meter data," in *2018 IEEE International Energy Conference (ENERGYCON)*, 3-7 June 2018 2018, pp. 1-6, doi: 10.1109/ENERGYCON.2018.8398841.

[11] V. Bassi, L. Ochoa, and T. Alpcan, "Model-Free Voltage Calculations for PV-Rich LV Networks: Smart Meter Data and Deep Neural Networks," in *2021 IEEE Madrid PowerTech*, 28 June-2 July 2021 2021, pp. 1-6, doi: 10.1109/PowerTech46648.2021.9494847.

[12] S. Weckx, R. D'Hulst, and J. Driesen, "Voltage Sensitivity Analysis of a Laboratory Distribution Grid With Incomplete Data," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1271-1280, 2015.

[13] Eaton. "Integration Capacity Analysis." https://www.cyme.com/software/cymeica/BR917066EN-ICA.pdf.

[14] F. Tamp and P. Ciufo, "A Sensitivity Analysis Toolkit for the Simplification of MV Distribution Network Voltage Management," *IEEE Transactions on Smart Grid*, vol. 5, no. 2, pp. 559-568, 2014.

[15] *ANSI C12.20-2015 American National Standard for Electricity Meters - 0.1, 0.2, and 0.5 Accuracy Classes*, ANSI, 2015.

[16] R. Steiner, M. Farrell, S. Edwards, T. Nelson, J. Ford, and S. Sarwat, "A NIST Testbed for Examining the Accuracy of Smart Meters under High Harmonic Waveform Loads," National Institute of Standards and Technology, 2019. doi: https://doi.org/10.6028/NIST.IR.8248.