

Improving Behind-the-Meter PV Impact Studies with Data-Driven Modeling and Analysis

Joseph A. Azzolini¹, Samuel Talkington², Matthew J. Reno¹, Santiago Grijalva², Logan Blakely¹, David Pinney³, and Stanley McHann³

¹Sandia National Laboratories, Albuquerque, NM, 87123, USA

²Georgia Institute of Technology, Atlanta, GA, 30318, USA

³National Rural Electric Cooperative Association (NRECA), Arlington, VA, 22203, USA

Abstract—Frequent changes in penetration levels of distributed energy resources (DERs) and grid control objectives have caused the maintenance of accurate and reliable grid models for behind-the-meter (BTM) photovoltaic (PV) system impact studies to become an increasingly challenging task. At the same time, high adoption rates of advanced metering infrastructure (AMI) devices have improved load modeling techniques and have enabled the application of machine learning algorithms to a wide variety of model calibration tasks. Therefore, we propose that these algorithms can be applied to improve the quality of the input data and grid models used for PV impact studies. In this paper, these potential improvements were assessed for their ability to improve the accuracy of locational BTM PV hosting capacity analysis (HCA). Specifically, the voltage- and thermal-constrained hosting capacities of every customer location on a distribution feeder (1,379 in total) were calculated every 15 minutes for an entire year before and after each calibration algorithm or load modeling technique was applied. Overall, the HCA results were found to be highly sensitive to the various modeling deficiencies under investigation, illustrating the opportunity for more data-centric/model-free approaches to PV impact studies.

Keywords—distribution system planning, hosting capacity analysis, machine learning, model calibration, time-series analysis

I. INTRODUCTION

Recent grid modernization efforts have led to dramatic changes in power distribution systems, such as increased penetration levels of behind-the-meter (BTM) distributed energy resources (DERs) like solar photovoltaic (PV) systems, advanced inverters with grid-support capabilities, and feeder-wide controls like DER management systems (DERMS) [1]. These changes have been accompanied by updated interconnection standards [2] that ensure new distributed PV installations can operate in autonomous grid-support modes and participate in DERMS objectives [3] or flexible interconnection schemes [4]. While these advancements offer opportunities to better regulate grid conditions and reduce losses [3], they have also led to added complexity in developing and maintaining *accurate* grid models that serve as the foundation for nearly all

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number 38426. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and 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.

distribution system planning and analysis tasks, including PV impact studies.

Fortunately, there has been a simultaneous increase in the adoption of advanced metering infrastructure (AMI), such as smart meters installed at customer premises and new grid assets like intelligent reclosers. Machine learning techniques and other data-driven approaches have since been developed to leverage the massive amounts of data available from these AMI devices to improve the fidelity of grid models [5] and the various analyses performed on them. For instance, data-driven methods have been developed for model calibration tasks such as identifying existing DERs, correcting customer phasing errors [6, 7] and service transformer pairings [8, 9], estimating low-voltage secondary network parameters [10, 11], and identifying errors in control device settings [12, 13]. However, the extent to which these data-driven methods improve the accuracy and reliability of PV impact studies remains unclear.

Hosting capacity analysis (HCA) is a useful type of PV impact study for distribution system planning, and it relies on distribution grid models. The term “hosting capacity,” or HC, refers to the maximum PV system size that can be deployed safely and reliably before operating violations occur or system upgrades are required [14]. In short, PV HCA utilizes power flow analysis on distribution grid models to analyze the potential impacts of PV at specific locations on the grid using deterministic methods [15]. Other methods calculate the total HC for an entire distribution feeder using stochastic methods [16-18] or streamlined methods [19]. Additional optimization-based methods [20] and data-driven methods [21, 22] have also been developed. However, these HCA methods all require grid models, which many smaller utilities do not have. There are also methods for discovering approximations of the power flow equations from AMI data without prior knowledge of the grid model [23, 24], which has created further opportunities to improve HCA or other PV impact studies [25, 26].

Increasingly, the outputs from PV HCA are being made accessible through public-facing HC maps [27], which can be used for facilitating PV siting and streamlining interconnection procedures [28]. However, the extent to which HCA can improve PV interconnection processes depends on the reliability of the results and the accuracy of the underlying grid models. Improving the accuracy of these models is an ongoing challenge for many utilities [29]. In this paper, the benefits of existing data-

driven modeling and calibration techniques were assessed through a series of conventional, model-based BTM PV HCAs conducted on a realistic test feeder equipped with actual utility AMI data. Overall, this paper highlights the importance of data-driven modeling and analysis techniques to ensure a safe and reliable transition to future high penetration PV scenarios.

II. METHODS

This paper utilized locational PV HCA as a means of quantifying the potential accuracy improvements to PV impact studies when data-driven model calibration algorithms and enhanced load modeling techniques are applied. Consecutive HCAs were conducted on a realistic feeder model before and after various modifications were applied using quasi-static time-series (QSTS) simulations, and this process was repeated for each algorithm or technique being tested. The following subsections describe the baseline circuit model, explain how the PV HCAs were conducted on it, and detail the modifications that were made to investigate each scenario.

A. Baseline Circuit Model

All HCAs were conducted on a modified version of the EPRI Ckt5 test feeder [30] depicted in Fig. 1, which represents an actual 12.47 kV, 3.2-mile-long distribution feeder that serves 1,379 residential loads. The modeled low-voltage secondary network included 591 service transformers and lines with realistic parameters. Each load was assigned a real and reactive power consumption profile for an entire year with a 15-minute resolution generated from an actual utility AMI dataset.

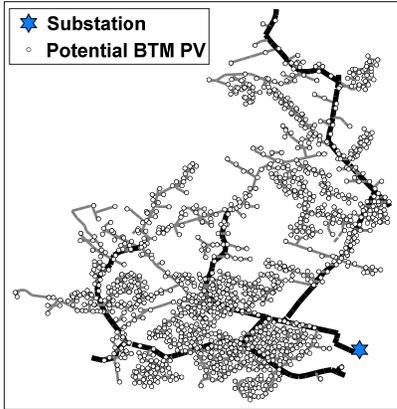


Fig. 1. Circuit plot of the baseline Ckt5 model with actual real and reactive power AMI data with 15-minute resolution at each low-voltage residential customer located at the end of varying lengths of 1/0 triplex.

This baseline circuit model, referred to as Ckt5, does not have any existing PV systems or voltage regulating equipment. The phase labels of all elements and all customer-to-transformer pairings in Ckt5 are assumed to be accurate. Additional details regarding the baseline Ckt5 model are summarized in Table I.

TABLE I. SUMMARY OF BASELINE MODELING AND ANALYSIS CONDITIONS

1	Distribution Model	No errors are present in the model—perfect knowledge of low-voltage service conductor lengths, transformer connections, phase information, PV information, and regulation equipment
2	Load Modeling	All loads modeled with real and reactive power AMI data at 15-minute resolutions
3	HCA Algorithm	QSTS analysis of every 15-minute time point

B. Locational Time-series Hosting Capacity Analysis

In this work, a *locational time-series* HCA framework was utilized. The term “locational” describes how the analysis sought to determine the maximum PV system size that can be installed at a particular *location* before operating violations occurred. The locations of interest for each HCA included the full set of customer buses across the low-voltage secondary networks on the feeder; these potential BTM PV locations are shown as white circles in Fig. 1. The “time-series” identifier informs that the HCA was conducted at regular intervals over a given time horizon. Unlike snapshot HCA methods that only consider a single set of worst-case conditions like during minimum feeder loading, the time-series HCA here was conducted for every 15-minute segment of an entire year (i.e., matching the resolution and time horizon of the AMI dataset).

Separate constraints for voltage and thermal metrics were selected to determine what constituted a violation. For the voltage constraint, a 1.05 per unit (pu) local voltage (V_l) threshold was selected. Thus, any real power injection from a PV system that resulted in $V_l > 1.05$ pu was considered a violation (similar to ANSI C84.1 Range A [31]). For the thermal constraint, a loading threshold of 120% of the upstream service transformer kVA rating (S_{rated}) was selected. Therefore, any real power injection from a PV system that caused the kVA at the nearest upstream service transformer (S_{xfmr}) to be $>1.2 \cdot S_{rated}$ was considered a violation (Dominion Energy uses a similar multiplier to set their normal loading capability limit [32]). Algorithm 1 provides additional details on how these thresholds were applied in the locational time-series HCA framework, where $n_{locations}$ was 1,379 for Ckt5, the discrete time step Δt was 15 minutes, and t_f was 525,600 minutes. This algorithm outputs two arrays of data for each BTM PV system location: the voltage- and thermal-constrained HCs at every 15-minute time point in the year ($HC_{VC,\ell}$ and $HC_{TC,\ell}$ in Algorithm 1).

ALGORITHM 1. LOCATIONAL TIME-SERIES HCA

```

Require: Compiled circuit model
1: for  $\ell = 1 : n_{locations}$  do
2:   for  $t = 1, \dots, t + \Delta t, \dots, t_f$  do
3:      $setAllLoads(t)$ 
4:      $P_{pv,\ell} \leftarrow 0$ 
5:     while  $(V_\ell \leq 1.05)$  or  $(S_{xfmr} \leq \frac{6}{5} S_{rated})$  do
6:        $V_\ell \leftarrow getVoltage(t)$ 
7:        $S_{xfmr} \leftarrow \sqrt{P_{xfmr}^2(t) + Q_{xfmr}^2(t)}$ 
8:       if  $V_\ell \leq 1.05$  then
9:          $HC_{VC,\ell}(t) \leftarrow P_{pv,\ell}$ 
10:      end if
11:      if  $S_{xfmr} \leq \frac{6}{5} S_{rated}$  then
12:         $HC_{TC,\ell}(t) \leftarrow P_{pv,\ell}$ 
13:      end if
14:       $P_{pv,\ell} \leftarrow P_{pv,\ell} + \Delta P_{size}$ 
15:    end while
16:  end for
17: end for

```

Additional filters were applied to further distill the outputs of Algorithm 1 down to a single HC value for each constraint at each location. First, a daytime filter was used to remove all the results outside of the time frame of 09:00 to 15:00; this filter focuses on time points where PV outputs generally reach their maximum daily value. Then, the time points are selected for which PV injections were most limited (i.e., the minimum values of $HC_{VC,\ell}$ and $HC_{TC,\ell}$ after the daytime filter was applied). This step is applied separately for each constraint and

each location, meaning that the most limiting time point is not necessarily the same from one location to the next, so the full spectrum of temporal and spatial variability is captured. Thus, this approach ensures that the final values selected for the voltage- and thermal-constrained HC at each location never result in a violation. For brevity, the acronyms “VC-HC” and “TC-HC” will be used hereafter to refer to the voltage- and thermal-constrained HC results, respectively. To summarize, the final VC-HC and TC-HC values represent the largest PV size that each location can support while ensuring the constraint is not violated at any time point through the year.

Each HCA was performed in OpenDSS [33] following Algorithm 1. First, the circuit model from Fig. 1 was loaded, or “compiled”, in OpenDSS and the simulation parameters were set to *duty* mode in which the real and reactive power consumption of the loads in the circuit would be adjusted at each time point according to their assigned time-series profile. Starting at the first customer location ($l = 1$), a PV system was placed and set to output at unity power factor (PF) with an initial size of 0 kW. Next, the power flow was solved iteratively as the size of the PV system is increased until the voltage and thermal constraints were both violated. The sizes of the PV injections that first caused each violation were recorded, and the algorithm moved to the next point in time. These iterative solutions were conducted for every time point ($t_f/\Delta t = 35,040$ in total) until the HCA had been solved for the entire year, at which point the PV system was moved to the next location ($l = 2$), and the entire process was repeated. Lastly, the results were post-processed to focus only on daytime conditions and the final VC- and TC-HC values were identified.

C. Model Calibration Scenarios

The first set of experiments was conducted to evaluate the potential accuracy improvements associated with various data-driven model calibration algorithms. Since Ckt5 was assumed to be completely accurate, unique errors were introduced to test each scenario such that the HCA could be applied to versions of the circuit model with and without the errors (i.e., before and after the errors were corrected for by each algorithm). The model calibration scenarios are summarized in Table II, where the abbreviation “Xfmr” is short for “Transformer”.

TABLE II. SUMMARY OF MODEL CALIBRATION SCENARIOS

#	Error Type	Experimental Implementation
A.1	Service Xfmr Size	De-rated 5 service transformers by 50%
A.2	Xfmr/Customer Pairing	Moved 3 customers to different service transformers, all selected at random
A.3	Missing Existing PV	Added an existing PV system with unity PF output to 5 random customer locations
A.4	Missing Existing PV w/ Volt-VAR	Added an existing PV system with Cat. B default Volt-VAR [2] to 5 random customer locations
A.5	Phase Labeling Errors	Mislabeled phases for 10% of customers and an entire single-phase lateral
A.6	Service Line Lengths	Set service lines to be 100 ft. of 1/0 triplex for all customers
A.7	Substation LTC Malfunction	Enabled a substation load tap changer (LTC) with a 3V bandwidth
A.8	Capacitor Malfunction	Enabled a 3-phase 600 kVAR capacitor connected to the medium-voltage network

For each scenario, it was assumed that the calibration algorithm successfully mitigated all corresponding modeling errors. Note that the model calibration scenarios were not

intended to represent an exhaustive list but rather to capture a subset of the most prevalent and pressing concerns faced by utilities for which existing solutions are available.

D. Load Modeling Scenarios

The next set of scenarios was investigated to quantify the potential improvements in HC results associated with data-driven load modeling techniques. The scenarios are summarized in Table III and represent cases where AMI datasets with high-resolution real and reactive power measurements were not available. In Scenario B.1, it was assumed that the AMI dataset only had real power measurements available, so the reactive power of the loads was modeled by assigning them all the same constant PF equal to the annual average PF measured at the feeder head. In Scenario B.2, it was assumed that there were no AMI data available and that the only time-series measurements available were from the substation. In this scenario, conventional load allocation was performed (as in the “business as usual case” in [30]) in which the allocation factors were determined during the peak time point and according to customer energy consumption data during that month. Once the allocation factors were assigned, a single normalized time-series profile was generated based on the total feeder power measured at the substation, while reactive power modeling was the same as Scenario B.1.

TABLE III. LOAD MODELING SCENARIOS

#	Modeling Technique	Experimental Implementation
B.1	Load P Modeling w/ AMI Data and Const. PF	Q profiles removed and replaced with constant PF (annual avg. PF at feeder head)
B.2	Conventional Load Allocation w/ Const. PF	Allocation factors calculated for feeder peak, one time-series profile from feeder head P measurements, constant PF (annual avg. PF at feeder head)

E. HCA Algorithm Scenarios

The last set of scenarios focused on the HCA algorithm itself. There have been many proposed HCA methods in literature and implemented in commercial distribution system analysis tools, but there are currently no standard algorithms for HCA. The scenarios in Table IV represent a subset of those proposed methods, representing varying degrees of input data requirements. Scenario C.1 was the same as Algorithm 1, except that all time points are considered (i.e., the daytime filter on line 13 was removed). Scenario C.2 also applied a similar algorithm to Algorithm 1 but included a relaxation component for each constraint that allowed them to be violated for up to 1% of the year. This scenario was more in line with distribution system operation standards (like ANSI C84.1 [31]) that allow for temporary violations to occur. While detailed metrics have been proposed to capture time-series HC constraints [34], in this work, a simplified relaxation approach was implemented here.

The remaining scenarios represented variations of conventional snapshot analysis methods that consider a single point in time. Scenarios C.3 and C.4 applied this snapshot method to the minimum loading time point on Ckt5, but C.4 considered only daytime points. Scenarios C.5 and C.6 were the same as C.3 and C.4, respectively, but were conducted on the version of Ckt5 created for Scenario B.2 in which no AMI data was available and load allocation was performed.

TABLE IV. HCA ALGORITHM SCENARIOS

#	HCA Type	Experimental Implementation
C.1	Annual Max. HC	Daytime filter not applied
C.2	Constraint Relaxation	Allow violations up to 1% of the year
C.3	Min. Load	Snapshot HC at minimum load time point
C.4	Min. Daytime Load	Snapshot HC at minimum daytime load time point
C.5	Min. Load (allocation case)	Snapshot HC at the minimum load time point (allocation case-B.2)
C.6	Min. Daytime Load (allocation case)	Snapshot HC at minimum daytime load time point (allocation case – B.2)

III. RESULTS

A. Baseline HCA Results

The HCA algorithm described in Section II. B. was first applied to the baseline Ckt5 feeder model. The results from this analysis, presented as histograms in Fig. 2, represented the “baseline” VC- and TC-HC values for each BTM PV system location in Ckt5 after the additional filters were applied. In general, there was a fairly wide range in the results for both constraints; the VC-HC results ranged from 2.1 to 21.6 kW with an average of 8.4 kW, and the TC-HC results ranged from 12.0 to 90.0 kW with an average of 39.4 kW. This range highlights the effects of different feeder characteristics like voltage drops throughout the low-voltage secondary networks and load consumption profiles. The results in Fig. 2 also show that the VC-HC results were more evenly distributed than the TC-HC results due to the fact that there were only a handful of different kVA ratings of the service transformers in Ckt5. Overall, these results are in line with expectations, given that there was no existing voltage regulation equipment on the feeder and that the thermal constraints were defined based on service transformer ratings. Also, while a number of locations had low VC-HC results, there were not any locations that had a VC-HC of 0 kW, meaning there were no existing violations during the daytime time points analyzed.

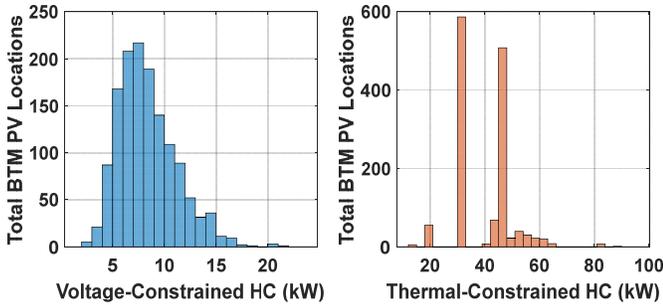


Fig. 2. Histograms of baseline voltage- and thermal-constrained HC results for each residential interconnection location on the feeder.

Detailed results for an average BTM PV location (#1241) on Ckt5 are presented in Fig. 3 to illustrate the time-series component of Algorithm 1, where the plots represent $HC_{VC,\ell}$ and $HC_{TC,\ell}$ after the daytime filter was applied but before the final VC- and TC-HC values were selected. Each time point in Fig. 3 show the maximum PV injection that could be tolerated at that time and location on the feeder before each constraint was violated. Thus, the final VC- and TC-HC values for location 1241 would be the minimum of each plot in Fig. 3 (ensuring that

the maximum PV sizes that cause no violations are selected), and would appear as a single data point in each plot of Fig. 2, respectively. The daily and seasonal changes shown in Fig. 3 corresponded to changes in loading conditions on the low-voltage secondary networks downstream of the service transformer and voltage variations on the medium-voltage network. Another interesting point illustrated in Fig. 3 is that there were many instances throughout the year when the thermal constraint was the most limiting factor (i.e., whenever the orange plot was below the blue plot), which was not apparent in the final HC results in Fig. 2.

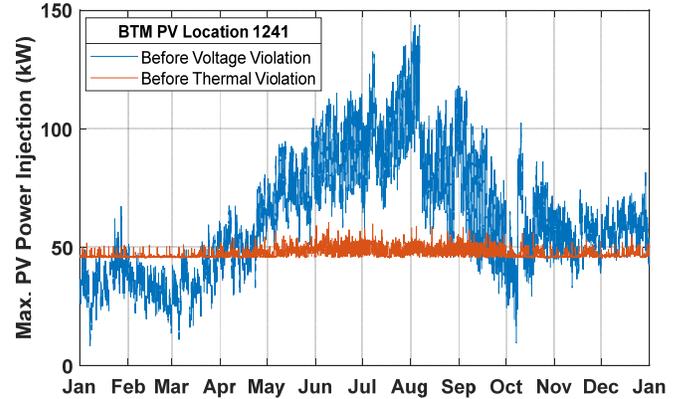


Fig. 3. Time-series of maximum active power injections for each 15-minute period during the daytime that an individual BTM PV interconnection can inject before causing a voltage or thermal violation for that location.

B. Model Calibration Scenarios

The locational time-series HCA was repeated after modifying Ckt5 according to the scenarios described in Table II. To compare the results of each scenario, the changes in VC-HC and TC-HC results from the baseline HCA were calculated following equations (1) and (2) for each BTM PV location (ℓ) and presented as boxplot distributions in Fig. 4.

Let ℓ be a proposed interconnection location. Let $VCHC_{\ell}^{base}$ and $TCHC_{\ell}^{base}$ be the baseline voltage- and thermal-constrained hosting capacity, respectively. Then the change in HC at ℓ due to scenario x is

$$\Delta VCHC_{\ell}(x) = VCHC_{\ell}(x) - VCHC_{\ell}^{base}, \quad (1)$$

$$\Delta TCHC_{\ell}(x) = TCHC_{\ell}(x) - TCHC_{\ell}^{base}. \quad (2)$$

The same VC-HC results are shown as a percentage of the baseline values in Fig. 5, according to equation (3).

$$\Delta VCHC_{\ell,\%}(x) = \frac{\Delta VCHC_{\ell}(x)}{VCHC_{\ell}^{base}} * 100 \quad (3)$$

In Scenario A.1, the HC results were impacted only at the locations downstream of the modified service transformers. Since the modifications in this case were related to the thermal constraint, the TC-HC results were impacted the most at those locations. The VC-HC results were affected to a lesser degree due to the changes in voltage drops across the de-rated transformers. In Scenario A.2, when 3 customer loads were moved to different service transformers, the HC results at locations downstream of all associated transformers were impacted. For the customers paired with a service transformer

that gained a customer, their locational TC-HC increased since more load was present during daytime conditions to offset PV injections, and the opposite was true for transformer-customer pairings that lost a customer. Similarly, the VC-HC results both increased and decreased depending on whether a customer was added or removed.

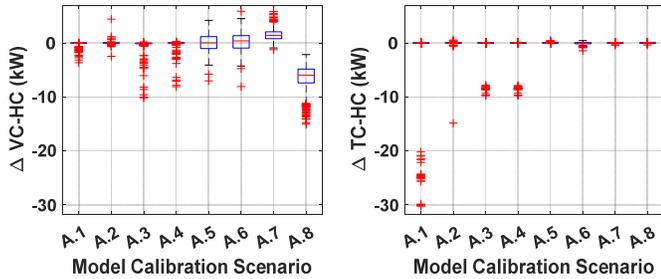


Fig. 4. Changes in VC-HC (left) and TC-HC (right) results associated with each model calibration scenario defined in Table II.

Scenario A.3 represents the case when PV systems had been installed on a feeder, but the model had not been updated. While most locations did not see much change, locations near those existing PV systems (e.g., other locations served by the same service transformer) did experience a significant change, where VC- and TC-HC results were reduced by up to 10 kW. Scenario A.4 had similar results, except that the impacts on the VC-HC results were less severe due to the voltage regulation provided by the existing PV system’s autonomous Volt-VAR function.

When the model contained mislabeled phases (Scenario A.5), the change in VC-HC results ranged from -7.0 kW to +4.2 kW, which is particularly noteworthy considering the average HC in Fig. 2 was 8.4 kW. Since distribution feeders often have notable voltage imbalances, the baseline VC-HC results were

either over-estimated or under-estimated depending on the voltage differences between the actual and mislabeled phase locations, but the baseline thermal results only experienced marginal changes. The middle 50% of VC-HC results in this scenario had a wider distribution than the previous 4 scenarios that can be attributed, in part, to the fact that more locations were directly modified, since phase labels are notoriously error-prone.

When service line lengths are unknown, they are often modeled with a length of 100 ft to include at least some of the expected voltage drops on the secondary networks (represented by Scenario A.6). This assumption over-estimated and underestimated the baseline VC-HC results (with a range of -8.1 kW to 5.9 kW) depending on whether the actual line was longer or shorter than the modeled line. The impacts to the TC-HC results in this case were also minimal since they were only associated with changes in power losses over the service lines.

Scenario A.7 included voltage regulation that the feeder head provided by an LTC enabled on the substation transformer. This LTC mitigated some of the extreme voltage fluctuations from the baseline model resulting in a significant increase in VC-HC for nearly all locations. Again, the TC-HC results were essentially unchanged. Scenario A.8 had the opposite effect on VC-HC results compared to A.7. This scenario represents the case when changes like retired or upgraded utility equipment are not captured in the model (e.g., a capacitor bank had been added as in A.8). In this case, the capacitor boosted voltages throughout the feeder, leading to reductions in VC-HC results at all locations. This modification did not impact load consumption, so the TC-HC results remained mostly the same.

To further illustrate some of the key potential impacts of the data-driven model calibration algorithms, the VC-HC results

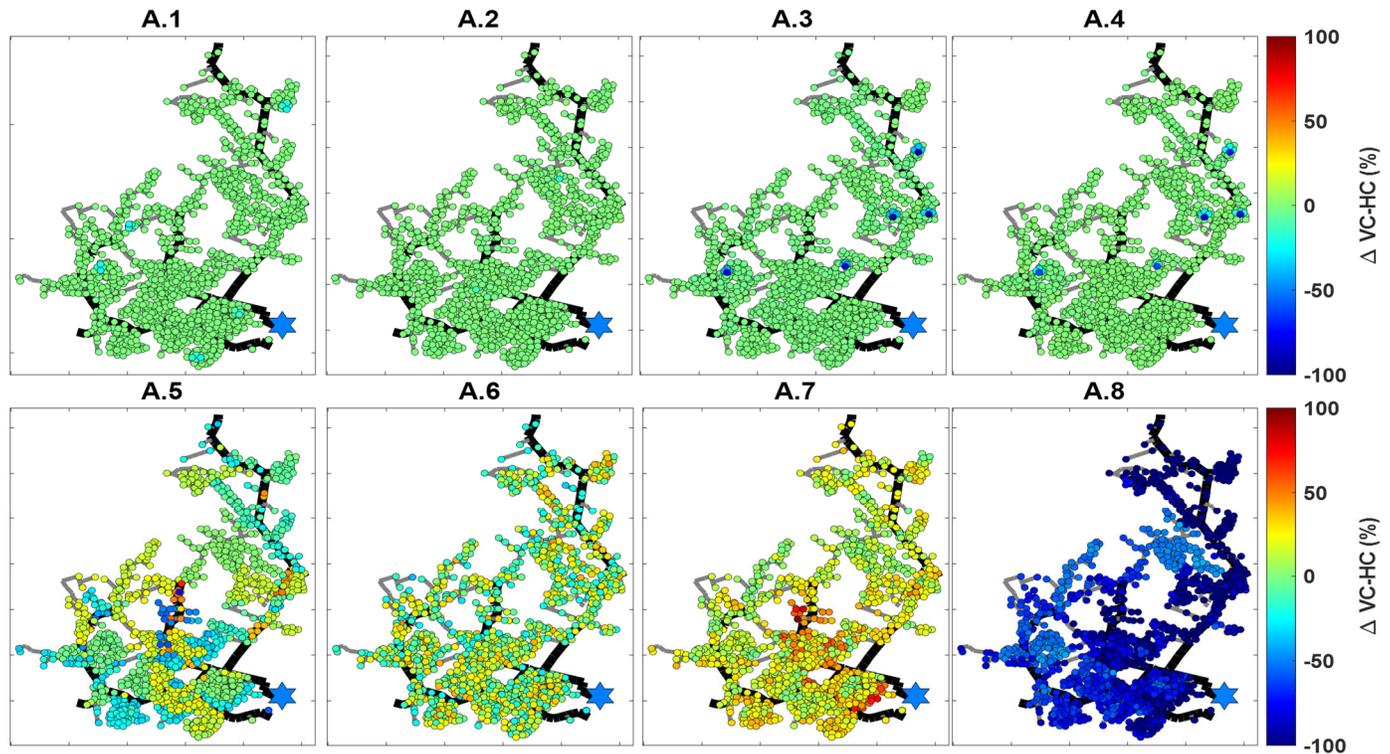


Fig. 5. Δ VC-HC results in % for Scenario A.1 to A.8 shown as a heatmap across the circuit plot of Ckt5.

from Fig. 4 were converted to percentages of the true VC-HC and displayed as a heatmap at their corresponding locations on Ckt5 in Fig. 5. In this figure, the green spectrum represents VC-HC results that did not deviate much from their baseline values, the red spectrum represents locations where VC-HC was overestimated, and the blue spectrum represents locations where VC-HC was underestimated.

C. Load Modeling Scenarios

The results from the two load modeling scenarios are presented in Fig. 6. In both cases, there were significant changes to the VC-HC and TC-HC results. Since both scenarios impacted the power consumption across the low-voltage secondary networks, it makes sense that the TC-HC results for all locations were affected. The load allocation led to more uniform power consumptions and reduced the worst-case conditions, resulting in more headroom for TC-HC. In both cases, the reactive power was modeled using a constant PF calculated from the feeder head annual average, which had a similar effect on reducing worst-case voltage conditions and lead to increased VC-HC for most locations.

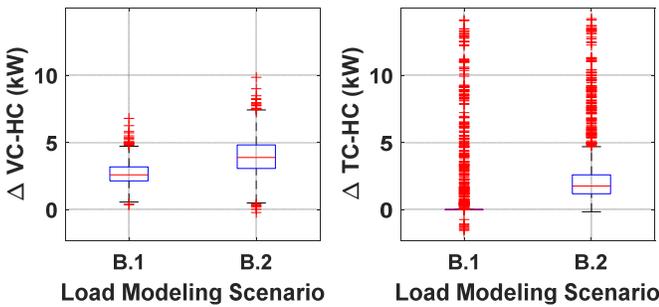


Fig. 6. Changes in HC results for each load modeling scenario from Table III.

D. HCA Algorithm Scenarios

The last set of scenarios focuses on how the HCA algorithm is implemented. Scenarios C.1 and C.2 were most similar to the locational time-series approach in Algorithm 1, so their results are presented together in Fig. 7. Scenario C.1 did not filter for only daytime conditions, meaning the HC results were only less than or equal to the baseline results. On the other hand, Scenario C.2 allowed for some violations to occur, so the HC results were always greater than or equal to the baseline results.

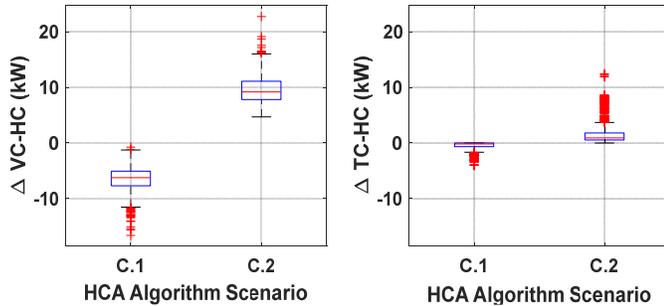


Fig. 7. Changes in HC results for the first two scenarios from Table IV.

The results from the remaining scenarios from Table IV are presented as estimated probability density functions (PDFs) in Fig. 8. The PDFs were estimated as normal distributions using

the *fitdist* function from the Statistics and Machine Learning Toolbox in MATLAB [35]. This figure shows that the impacts on TC-HC results were very similar across these four scenarios. For the VC-HC results, the load allocation cases had more of a positive bias and had higher peak densities (comparing C.5 vs C.3 and C.6 vs C.4, respectively), indicating that load allocation may be more predictable when snapshot methods are being used.

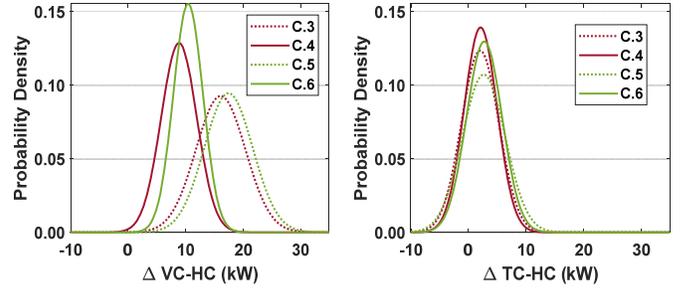


Fig. 8. Estimated probability density functions of the HC results for Scenarios C.3 to C.4.

IV. CONCLUSION

The accuracy and reliability of PV impact studies such as HCA are becoming increasingly important as PV penetration levels rise; under-estimations can lead to unnecessary denials of PV interconnection requests, while over-estimations can lead to safety concerns and unintended interactions with grid components. This paper quantified the improvements in accuracy for PV HC results when a variety of data-driven modeling and analysis techniques were applied. Overall, it was observed that the HCA voltage constraint was more sensitive to errors in the model than the thermal constraint, different load modeling techniques can significantly alter both VC- and TC-HC results, and that snapshot methods were not as accurate as time-series methods. If snapshot methods have to be used, applying load allocation may produce more repeatable results. Given that PV HCA was found to have a high degree of sensitivity to model errors and modeling techniques, there is an increasing need for data-driven model calibration algorithms and novel model-free PV impact analysis methods to ensure future high PV penetration scenarios can be properly accommodated.

REFERENCES

- [1] D. Narang, M. Ingram, S. Widergren, S. Gonzalez *et al.*, "GMLC Survey of Distributed Energy Resource Interconnection and Interoperability Standards," United States, 2021-09-23 2021.
- [2] *IEEE 1547 Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces*, IEEE, 2018.
- [3] K. Ardani, E. O'Shaughnessy, and P. Schwabe, "Coordinating Distributed Energy Resources for Grid Services: A Case Study of Pacific Gas and Electric," National Renewable Energy Laboratory, Golden, CO, 2018.
- [4] K. S. A. Sedzro, K. Horowitz, A. K. Jain, F. Ding, B. Palmintier, and B. Mather, "Evaluating the Curtailment Risk of Non-Firm Utility-Scale Solar Photovoltaic Plants under a Novel Last-In First-Out Principle of Access Interconnection Agreement," *Energies*, vol. 14, no. 5, p. 1463, 2021.
- [5] M. J. Reno, L. Blakely, R. D. Trevizan, B. Pena *et al.*, "IMoFi - Intelligent Model Fidelity: Physics-Based Data-Driven Grid Modeling to Accelerate Accurate PV Integration Final Report," Sandia National Laboratories, Albuquerque, NM, 2022.
- [6] L. Blakely and M. J. Reno, "Phase identification using co-association matrix ensemble clustering," *IET Smart Grid*, vol. 3, no. 4, pp. 490-499

- [7] F. Therrien, L. Blakely, and M. J. Reno, "Assessment of Measurement-Based Phase Identification Methods," *IEEE Open Access Journal of Power and Energy*, vol. 8, pp. 128-137, 2021.
- [8] L. Blakely and M. J. Reno, "Identifying Errors in Service Transformer Connections," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*, 2-6 Aug. 2020 2020, pp. 1-5.
- [9] L. Blakely and M. J. Reno, "Identification and Correction of Errors in Pairing AMI Meters and Transformers," in *2021 IEEE Power and Energy Conference at Illinois (PECI)*, 1-2 April 2021 2021, pp. 1-8.
- [10] M. Lave, M. J. Reno, and J. Peppanen, "Distribution System Parameter and Topology Estimation Applied to Resolve Low-Voltage Circuits on Three Real Distribution Feeders," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1585-1592, 2019.
- [11] K. Ashok, M. J. Reno, and D. Divan, "Secondary Network Parameter Estimation for Distribution Transformers," in *2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 17-20 Feb. 2020 2020, pp. 1-5.
- [12] J. Yusuf, J. A. Azzolini, and M. J. Reno, "Data-Driven Methods for Voltage Regulator Identification and Tap Estimation," presented at the 2022 IEEE Kansas Power & Energy Conference (KPEC), 2022.
- [13] R. D. Trevizan, C. Ruben, A. Rossoni, S. C. Dhulipala, A. Bretas, and N. G. Bretas, "µPMU-Based Temporal Decoupling of Parameter and Measurement Gross Error Processing in DSSE," *Electricity*, vol. 2, no. 4, pp. 423-438, 2021.
- [14] S. Stanfield, Y. Zakaï, and M. McKerley, "Key Decisions for Hosting Capacity Analysis," IREC, 2021.
- [15] E. Mulenga, M. H. J. Bollen, and N. Etherden, "A review of hosting capacity quantification methods for photovoltaics in low-voltage distribution grids," *International Journal of Electrical Power & Energy Systems*, vol. 115, p. 105445, 2020/02/01/ 2020.
- [16] D. Diaz, A. Kumar, J. Deboever, S. Grijalva *et al.*, "Scenario-Selection for Hosting Capacity Analysis of Distribution Feeders with Voltage Regulation Equipment," in *2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 18-21 Feb. 2019 2019, pp. 1-5.
- [17] J. Deboever, S. Grijalva, J. Peppanen, M. Rylander, and J. Smith, "Practical Data-Driven Methods to Improve the Accuracy and Detail of Hosting Capacity Analysis," in *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC)*, 10-15 June 2018 2018, pp. 3676-3681.
- [18] EPRI, "Stochastic Analysis to Determine Feeder Hosting Capacity for Distributed Solar PV," EPRI, Technical Report 1026640, 2012.
- [19] 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.
- [20] A. Dubey and S. Santoso, "On Estimation and Sensitivity Analysis of Distribution Circuit's Photovoltaic Hosting Capacity," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2779-2789, 2017.
- [21] X. Chen, W. Wu, B. Zhang, and C. Lin, "Data-Driven DG Capacity Assessment Method for Active Distribution Networks," *IEEE Transactions on Power Systems*, vol. 32, no. 5, pp. 3946-3957, 2017.
- [22] F. Ding and B. Mather, "On Distributed PV Hosting Capacity Estimation, Sensitivity Study, and Improvement," vol. 8, 2017-07-01 2017.
- [23] K. Christakou, J. LeBoudec, M. Paolone, and D. Tomozei, "Efficient Computation of Sensitivity Coefficients of Node Voltages and Line Currents in Unbalanced Radial Electrical Distribution Networks," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 741-750, 2013.
- [24] S. Bolognani and F. Dörfler, "Fast power system analysis via implicit linearization of the power flow manifold," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 29 Sept.-2 Oct. 2015 2015, pp. 402-409.
- [25] A. Procopiou, M. Liu, L. Ochoa, T. Langstaff, and J. Harding, *Smart meter-driven estimation of PV hosting capacity*. 2020.
- [26] V. Bassi, L. Ochoa, and T. Alpcan, *Calculating Voltages Without Electrical Models: Smart Meter Data and Neural Networks*. 2021.
- [27] S. Stanfield, Y. Zakaï, and M. McKerley, "Key Decisions for Hosting Capacity Analyses," IREC, 2021.
- [28] S. Stanfield and S. Safdi, "Optimizing the Grid: A Regulator's Guide to Hosting Capacity Analyses for Distributed Energy Resources," IREC.
- [29] A. Nagarajan and Y. Zakaï, "Data Validation for Hosting Capacity Analyses," National Renewable Energy Laboratory, NREL/TP-6A40-81811, 2022.
- [30] EPRI, "Enhanced Load Modeling: Leveraging Expanded Monitoring and Metering," Palo Alto, CA, 3002015283, 2019.
- [31] *ANSI C84.1-2020 American National Standard For Electric Power Systems and Equipment - Voltage Ratings (60 Hz)*, ANSI, 2020.
- [32] "Distribution Transformer Loading," Dominion Energy, 2019.
- [33] EPRI. "Open Distribution System Simulator (OpenDSS)." <http://sourceforge.net/projects/electricdss/>.
- [34] A. K. Jain, K. Horowitz, F. Ding, K. S. Sedzro *et al.*, "Dynamic hosting capacity analysis for distributed photovoltaic resources—Framework and case study," *Applied Energy*, vol. 280, p. 115633, 2020/12/15/ 2020.
- [35] *Statistics and Machine Learning Toolbox User's Guide*. (2019). MathWorks.