PV Hosting Capacity Estimation: Experiences with Scalable Framework

Preprint

Wenbo Wang, Daniel Thom, Kwami Senam Sedzro, Sherin Ann Abraham, Yiyun Yao, Jianli Gu, and Shibani Ghosh

National Renewable Energy Laboratory

Presented at the 49th IEEE Photovoltaic Specialists Conference (PVSC49)
Philadelphia, Pennsylvania
June 5–10, 2022

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

Conference Paper
NREL/CP-6A40-81851
June 2022

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

PV Hosting Capacity Estimation: Experiences with Scalable Framework

Preprint

Wenbo Wang, Daniel Thom, Kwami Senam Sedzro, Sherin Ann Abraham, Yiyun Yao, Jianli Gu, and Shibani Ghosh

National Renewable Energy Laboratory

Suggested Citation
Wang, Wenbo, Daniel Thom, Kwami Senam Sedzro, Sherin Ann Abraham, Yiyun Yao, Jianli Gu, and Shibani Ghosh. 2022. *PV Hosting Capacity Estimation: Experiences with Scalable Framework; Preprint*. Golden, CO: National Renewable Energy Laboratory. NREL/CP-6A40-81851. https://www.nrel.gov/docs/fy22osti/81851.pdf.

NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Contract No. DE-AC36-08GO28308

Conference Paper
NREL/CP-6A40-81851
June 2022

National Renewable Energy Laboratory
15013 Denver West Parkway
Golden, CO 80401
303-275-3000 • www.nrel.gov

NOTICE

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Renewable Energy and Energy Efficiency Solar Energy Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via www.OSTI.gov.

Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.

NREL prints on paper that contains recycled content.

PV Hosting Capacity Estimation: Experiences with Scalable Framework

Wenbo Wang, Daniel Thom, Kwami Senam Sedzro, Sherin Ann Abraham, Yiyun Yao, Jianli Gu, Shibani Ghosh
Energy Systems Integration
National Renewable Energy Laboratory
Golden, CO, USA
{Wenbo.Wang, Daniel.Thom, Kwami.Sedzro, SherinAnn.Abraham, Yiyun.Yao, Jianli.Gu, Shibani.Ghosh}@nrel.gov

***Abstract*—Hosting capacity is an indication of the amount of solar photovoltaics (PV) that can be hosted in a distribution system without additional changes to infrastructure or operations. This paper presents a framework for estimating the PV hosting capacity at scale. First, we analyze computational, modeling and other key challenges of performing relevant, large-scale simulations, provided along with the experiences and lessons learned. Then, we develop two open-source Python-based software tools to conduct repeatable distribution analyses: the Distribution Integration Solution Cost Options (DISCO) for configuring and analyzing simulations and the Job Automation and Deployment Engine (JADE) for parallelizing jobs on high-performance computing clusters. A case study of hosting capacity estimation for the SMART-DS San Francisco (SFO) 2000+ synthetic feeders, is used to demonstrate the capability of the developed DISCO+JADE framework and tools. The framework and tools can help utilities assess the overall hosting capacity of their service territory, which can help them better plan for the overall upgrade costs to integrate more PV in the future. The experiences are shared to aid the tool users and researchers to conduct relevant studies and research.**

***Index Terms*—Distributed energy resources, Hosting capacity, Large-scale simulation**

I. Introduction

The fast deployment of distributed solar photovoltaics (PV) stretches the electric grid toward limitations and creates operational concerns for utilities. The grid's ability to accommodate PV is typically estimated through hosting capacity. In the field, PV interconnection screening processes are often evaluated based on the understanding of the feeder hosting capacity [1]. The concept of hosting capacity is defined as the total PV capacity that can be accommodated on a given feeder without violating operational constraints [2]. Note that in this paper, we focus on distributed PV (DPV); thus PV and DPV are used interchangeably.

Estimating hosting capacity normally involves simulating many scenarios of different locations and sizes of PV to evaluate the impact and identify boundary scenarios from operational violations. This can be done through steady-state or time-series simulations, and the corresponding results are often called static or dynamic hosting capacity (SHC or DHC), respectively. Using steady-state analysis, [3] considers a limited number of scenarios, including PV distributed evenly, aggregated near the beginning, and aggregated near the end of feeders; [4] proposes to estimate hosting capacity more comprehensively based on stochastic analysis, generating many scenarios and penetration levels for PV deployments. Reference [5] proposes using year-long time-series simulations to estimate hosting capacity, where the duration of violations and the movement of the legacy controllers can be captured. However, because of the large number of scenarios that need to be considered, past research mostly focused on a limited number of feeders. It would be beneficial to have a flexible and scalable framework for estimating feeder hosting capacity, e.g., with thousands of feeders and potentially millions of scenarios. This will help utilities better plan for the overall upgrade costs to integrate more PV [6], and it will facilitate data analytics of the interconnection process [7]. To estimate hosting capacity at scale, one can (but is not limited to) use a smaller number of scenarios for more feeders [7], speed up each simulation [8], optimize the computing execution, and use more computing power [9].

Setting this work apart from previous hosting capacity research, this paper develops a scalable hosting capacity solution through optimizing models and the computing execution. The main contributions can be summarized as follows:

- A framework is developed for scalable hosting capacity estimation. As a byproduct, two open-source Python-based software tools are developed for conducting repeatable distribution analyses and simulation job submission: namely, the Distribution Integration Solution Cost Options (DISCO) [10] and the Job Automation and Deployment Engine (JADE) [11], respectively.
- The challenges and experiences of scalable hosting capacity estimation are analyzed and discussed, including computational and modeling challenges.
- The capability of the developed framework and tools are demonstrated through the SMART-DS San Francisco (SFO) region 2000+ feeders [12].

II. Hosting Capacity Estimation

A. Methodology

This paper uses a Monte Carlo-based stochastic approach [4], [13] to estimate hosting capacity. Fig. 1 shows the analysis flow; both SHC and DHC [5] are shown. Starting with the feeder models and weather data, the stochastic approach is used to generate the PV deployment scenarios at different

TABLE I
SHC Metrics

Metric	Threshold
Voltage	± 5% deviation from the nominal value
Thermal	100% asset loading
Power quality	Voltage unbalance 3%, etc.
Protections	Coordination, set points (false/miss detection)

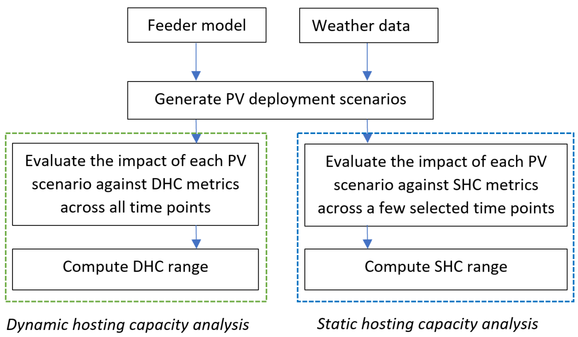

Fig. 1. Hosting capacity analysis flowchart

penetration levels and for a diverse spatial distribution according to the technique introduced in [14]. Next, the impact of each PV scenario is independently assessed with regard to operational metrics and thresholds. Examples of SHC metrics and thresholds are shown in Table I [2], [13]. Based on the impact assessment, the hosting capacities of the system under study are determined.

In addition, because of the stochastic nature of the PV deployment (location, size), the analysis typically results in a range of hosting capacities for each system, which are characterized by minimum and maximum hosting capacities [4].

1) Generating PV Deployments: Because it is nearly impossible to perfectly predict the adoption pattern of PV in terms of location and size distribution, we develop several adoption patterns or deployment samples to capture diversified and realistic PV scenarios [14]. The developed process considers three spatial placement types: close, random, and far. In each spatial placement type, the deployment of PV is incremental from one penetration level to next; therefore, each PV scenario is uniquely identified by its placement, sample, and penetration level.

2) Multi-Time-Point SHC Analysis: Instead of a single snapshot, the multi-time-point SHC analysis considers several grid conditions that are often used in grid planning studies. The most common four conditions include minimum daytime load, maximum PV output, minimum net load, and maximum load, which are extracted from load and solar irradiance profiles. The thermal and voltage impacts of integrating each PV scenario into the distribution grid are assessed for these conditions. In the end, the worst-case results from the four conditions are selected.

3) DHC Analysis: Unlike the multi-time-point SHC analysis, which evaluates the grid impact at a few selected time points, the DHC analysis assesses the impact of each PV scenario across year-long time series. DHC analysis allows violations for a short duration, and it can track moving averages and device operation counts. More details on the DHC metrics and suggested thresholds can be found in [5].

B. Challenges and experiences of Estimating Hosting Capacity at Scale

This subsection lists the challenges and considerations of estimating hosting capacity at scale. In each listed challenge or consideration, it is also provided alongside with a description of how the developed framework (i.e. DISCO, JADE tools) approaches them.

1) Computational Challenge: The main challenge is how to use computational resources efficiently to manage large numbers of jobs under the constraints of central processing unit (CPU) cores, memory, and storage space. Here we consider the use of high-performance computing, with access to multiple compute nodes simultaneously. Ideally, one can use as many as compute nodes as are available; however, not only are the resources limited, but also, in many cases, the benefit of using more nodes is outpaced by the burden of the communication among nodes [9]. Conveniently, in the case of estimating hosting capacity, the PV deployment scenarios can be run independently (naturally partitioned) with little communication required (data dependence and synchronization) among nodes.

2) Modeling Challenge: This is related to the standardization of the data and models. Models can take many forms, and it is impractical to support all models. DISCO [10] defines standard models and then provides transformations for specific formats. For example, DISCO can run simulation both at the feeder level and the substation transformer level. Also, note that the actual power flow is conducted through OpenDSS [15]; DISCO leverages PyDSS [16], an OpenDSS Python wrapper that provides PV control functions with enhanced convergence (i.e., volt-var, volt-watt) and many other functions.

3) Other Practical Considerations:

- Computational burden load balancing refers to the practice of distributing approximately equal amounts of work among processors so that all processors are kept busy all the time [17]; otherwise, the slowest task will determine the overall run time. In the case of estimating hosting capacity, different feeders with different numbers of circuit elements that create challenges to computational load balancing (not to be confused with the load in customer demand in kilowatts). In DISCO, a linear regression model is developed to predict the job run times with the predictor variables that include the numbers of PV units and circuit elements. Based on the predicted run time, the jobs can be batched and allocated roughly evenly to the processors. This linear regression model builds its estimates based on a training dataset created by dry run jobs.

- Often, not all jobs will successfully run the first time due to issues including model errors, computational limits, or convergence challenges. The capability to test run, debug, and rerun failed or missing jobs is critical for managing large numbers of jobs. Both DISCO and JADE have comprehensive logging functionalities for each steps of the simulation providing meaningful debug information to the users and developers. In addition, JADE records status for each simulation job (i.e. pass or fail), and has a function that directly re-submit the failed jobs.
- Job monitoring and reporting are important because the execution information can be used for tuning the simulation parameters such as required computational nodes on HPC, simulation wall time, job batch size per computational node, etc. The reported metrics for job execution in JADE include individual job status, errors and events, job execution times, and compute resource utilization statistics such as CPU and memory usage, and networked communication related metrics[1] (e.g. time consumed transmitting packets from CPU to hard drives, hard drives to CPU).
- Care is required in data architecture and formats when working with large quantities of input and output data. Data storing, query, sharing are critical to data management and analytics. DISCO has build-in function to ingest raw output of the simulation results into a `SQLite` database. The current database schema are designed for distribution impact analysis, e.g. hosting capacity.

III. Developed Open-Source Tools

A. Hosting Capacity Estimation with DISCO

DISCO—an NREL-developed, open-source tool—is a collection of integrated functions that can be used to automate a wide range of electric distribution analyses at scale. For instance in the LA100 distribution analysis effort, it was used to conduct impact analysis and estimate upgrade costs for thousands of feeders with hundreds of scenarios each [18]. Here we focus on its use for distributed PV hosting capacity estimates. Fig. 2 shows the flowchart for the main steps to run the hosting capacity estimation. These blocks are briefly described as follows, and more details about the implementation and examples can be found in [10]:

- Prepare the OpenDSS models and directory structures according to the data sources defined by DISCO, then provide the input path to DISCO. Four types of data sources are currently supported.
- Transform the source of the OpenDSS models into DISCO models. In the case of hosting capacity, the transformed DISCO models include PV deployment scenarios and OpenDSS instances through PyDSS with PV control enabled as well as functions such as selectively saving simulation results. These functions and the files are described by a `JSON` file, which is the output of this step.
- Configure the JADE jobs based on the DISCO models with customized execution requirements. Execution on a high-performance computer (HPC) is highly configurable depending on the job resource requirements, e.g., the number of computational nodes to use, the number of jobs to run in parallel on each node. The output is an updated `JSON` file from the previous step with the added entries including all the job configuration information.
- Submit the jobs with JADE based on the `JSON` file. Underneath, JADE uses subprocess management [19] to parallelize the execution of the jobs on either HPC clusters (with Slurm [20]) or stand-alone computers. The submitted jobs will be run once the requested resources become available.
- After the jobs are complete, JADE can assist with the execution analysis by showing summaries of the individual job status, errors and events, job execution times, and compute resource utilization statistics. DISCO provides simulation results analysis and certain visualizations.

B. JADE for Submitting Jobs

JADE [11] automates the parallelized execution of jobs. It has specific support for distributing work on HPC compute nodes, but it can also be executed on stand-alone computers. Some important features are described as follows; for more information, see [11]:

- Maximizing the number of jobs that can be completed on a given node in a specific time duration is critical to optimize jobs on HPC systems, even more so if the HPC systems are managed such that the computational nodes are typically allocated for a limited period of time and are not always available. JADE constructs per-node job batches by accounting for job duration, number of required and available CPUs, and allocation time to maximize the use of each node. JADE allows customization of all parameters.
- For job monitoring and reporting, for example, after the simulation jobs are submitted, JADE provides ways to monitor the simulation status and results, find failed jobs, and restart them.
- For pipeline capability, JADE allows users to specify inter-job dependencies and pipeline stages to submit all work in one step. JADE implements a distributed submitter protocol[2] whereby a node can submit new jobs once dependent jobs are complete. This obviates the need to monitor jobs from a software application that must remain running for the duration of the work, which can take multiple days or weeks, depending on node availability.

IV. Case Studies

The capability of the developed framework is demonstrated on the SMART-DS synthetic SFO 2000+ feeders [12]. This

[1]https://nrel.github.io/jade/tutorial.html#resource-monitoring

[2]https://nrel.github.io/jade/distributed_submission.html

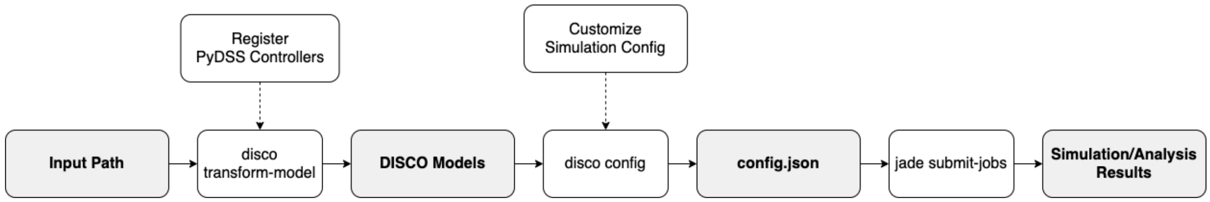

Fig. 2. DISCO workflow: bold fonts with grey boxes indicate data input and output; and dashed arrows indicate customized configurations and features.

section analyzes the large-scale simulations and results. We ran simulations on NREL's Eagle HPC. The HPC contains 2500+ compute nodes, where each node has 36 cores and at least 90 GB of memory. Only a fraction of these nodes were used at any given time for this analysis. The synthetic SFO system is built in the geographic area of the extended San Francisco Bay Area, California. The data set contains 40 subregions that span both urban and rural geographies. It covers a total of 2,236 feeders, 4.3 million consumers, 9.8 million electrical nodes, 632 primary substations, and 559,151 distribution transformers and is publicly available at [21].

A. Large-Scale Hosting Capacity Results

The large-scale hosting capacity estimation result is visualized as the hosting capacity map shown in Fig. 3. The map color codes feeders based on the percentage of PV hosting capacity to peak loads. It shows a diverse hosting capacity results for all 2000+ feeders. Fig. 4 provides a zoom-in example results for 3 feeder near San Mateo area, which shows the example feeders can host relatively high PV penetration. Fig. 5 gives the distribution of the hosting capacity results in terms of the number the feeders, it roughly follows a normal distribution except the extreme 0 and 200 percent results.

B. Computational Efficiency

In addition to a large number of feeders under study, the stochastic approach for estimating hosting capacity requires running many power flows with different PV scenarios for each feeder. In this study, there are a total of 849,719 jobs to run, and each job contains 8 snapshot power flows, including 2 control modes for PV (unity power factor and volt-var) and 4 time points (see Section II.A.2). All the jobs are packed in batches, and each batch is assigned to compute nodes based on an estimated run time (see Section II.B.3). Fig. 6 gives the job simulation time distribution. The average job simulation time is 12 minutes, with a standard deviation of 8 minutes. The total simulation time is 10,012,654 minutes, which is equivalent to 19 years. This is the amount of time needed if all the simulations were run in a serial program. Using HPC with the developed framework, we required 1000 computational nodes, and the total simulation was done in approximately 35 hours, plus 5 more hours to post-process the results.

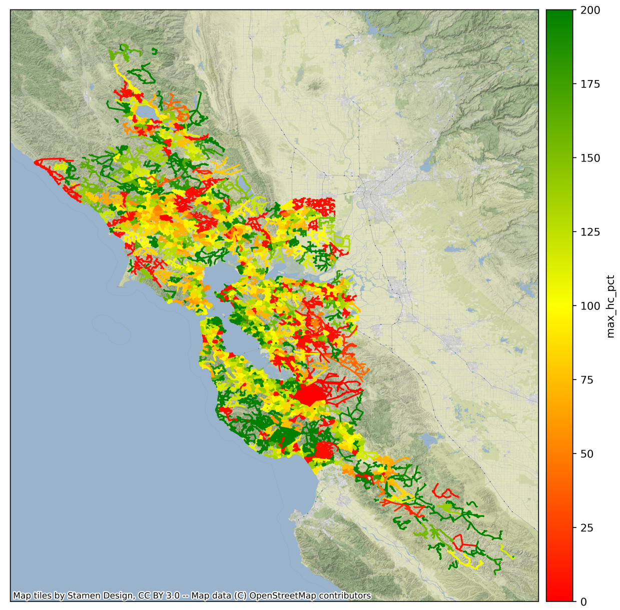

Fig. 3. Hosting capacity map for synthetic SFO feeders, where the maximum hosting capacity percentage relative to peak loads are displayed. Note that this figure serves as a demonstration only, the values of the hosting capacity might be different when the analysis assumptions are different, e.g. PV DC-AC ratio, load and PV values, legacy controls, etc.

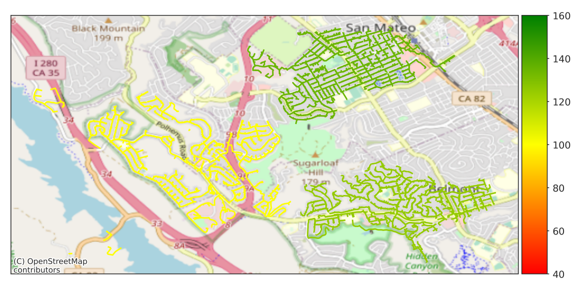

Fig. 4. Zoomed-in hosting capacity map for 3 feeders, other adjacent feeders are not displayed for better readability.

V. Conclusions

This paper has described our experiences estimating PV hosting capacity at scale for distribution systems. First, we analyze the key challenges of performing relevant, large-scale simulation, including computational and modeling challenges, our experiences and lesson learned are also provided. Then, two Python-based, open-source tools are created for modeling and running the simulations. The case study using the SMART-DS SFO 2000+ feeders demonstrates the capability of the developed framework and tools. Our experience shows that estimating hosting capacity at scale requires large number of power flow simulations, it is critical to efficiently manage the limited computational resources.

The outcomes from this research can help utilities better plan for the overall costs of integration, and it can help enable

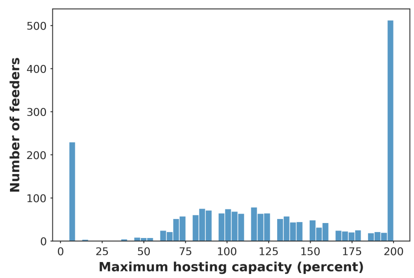

Fig. 5. Maximum hosting capacity (percent) distribution. Note that the maximum PV deployment tested was 200% of peak load, such that the right most histogram bar indicates 200% or higher hosting capacity.

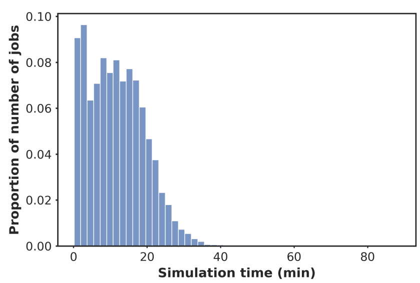

Fig. 6. Simulation time histogram

data analytics of the interconnection process. Future work will include large-scale simulations of DHC and cost-benefits analysis of traditional upgrade and non-wire alternatives.

Acknowledgments

This work was authored by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

This research was performed using computational resources sponsored by the Department of Energy's Office of Energy Efficiency and Renewable Energy and located at the National Renewable Energy Laboratory.

References

[1] W. Wang, J. Keen, J. Giraldez, K. Baranko, B. Lunghino, D. Morris, F. Bell, A. Shumavon, K. Levine, T. Ward, A. Dave, and J. Bank, "Supervised learning for distribution secondary systems modeling: Improving solar interconnection processes," *IEEE Transactions on Sustainable Energy*, 2022, early access.

[2] S. M. Ismael, S. H. Abdel Aleem, A. Y. Abdelaziz, and A. F. Zobaa, "State-of-the-art of hosting capacity in modern power systems with distributed generation," *Renewable Energy*, vol. 130, pp. 1002–1020, 2019.

[3] R. A. Kordkheili, B. Bak-Jensen, J. R-Pillai, and P. Mahat, "Determining maximum photovoltaic penetration in a distribution grid considering grid operation limits," in *2014 IEEE PES General Meeting — Conference Exposition*, 2014, pp. 1–5.

[4] J. Smith, "Stochastic analysis to determine feeder hosting capacity for distributed solar pv," Elect. Power Res. Inst., Palo Alto, CA, USA, Tech. Rep. 1026640, 2012.

[5] A. K. Jain, K. Horowitz, F. Ding, K. S. Sedzro, B. Palmintier, B. Mather, and H. Jain, "Dynamic hosting capacity analysis for distributed photovoltaic resources—framework and case study," *Applied Energy*, vol. 280, p. 115633, 2020.

[6] K. A. Horowitz, B. Palmintier, B. Mather, and P. Denholm, "Distribution system costs associated with the deployment of photovoltaic systems," *Renewable and Sustainable Energy Reviews*, vol. 90, pp. 420–433, 2018.

[7] M. J. Reno and R. J. Broderick, "Statistical analysis of feeder and locational pv hosting capacity for 216 feeders," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1–5.

[8] R. J. Broderick *et al.*, "Rapid QSTS simulations for high-resolution comprehensive assessment of distributed PV." Sandia National Lab, Albuquerque, NM, USA, Tech. Rep. 2021-2660, 2021.

[9] Q. Mu, L. Niu, and Y. Cheng, "The communication methodology on the large-scale power system realtime simulation," in *The 16th IET International Conference on AC and DC Power Transmission (ACDC 2020)*, vol. 2020, 2020, pp. 497–504.

[10] D. Thom. DISCO. Accessed: Dec. 10, 2021. [Online]. Available: https://github.com/NREL/disco

[11] ——. JADE. Accessed: Dec. 10, 2021. [Online]. Available: https://github.com/NREL/jade

[12] C. Mateo, F. Postigo, F. D. Cuadra, T. G. S. Roman, T. Elgindy, P. Duenas, B. M. Hodge, V. Krishnan, and B. Palmintier, "Building large-scale u.s. synthetic electric distribution system models," *IEEE Transactions on Smart Grid*, vol. 11, pp. 5301–5313, 11 2020.

[13] F. Ding and B. Mather, "On distributed PV hosting capacity estimation, sensitivity study, and improvement," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 3, pp. 1010–1020, 2017.

[14] K. S. A. Sedzro, M. Emmanuel, and S. A. Abraham, "Generating sequential PV deployment scenarios for high renewable distribution grid planning ," *submitted for publication*, 2022.

[15] R. C. Dugan and D. Montenegro, *The Open Distribution System Simulator (OpenDSS Manual)*, 2020. [Online]. Available: https://smartgrid.epri.com/SimulationTool.aspx

[16] A. Latif, M. Ikechi, I. Hawaiian Electric Company, and S. E. Industries, "NREL/PyDSS," 2018-12. [Online]. Available: https://www.osti.gov/biblio/1512458

[17] B. Barney. Introduction to parallel computing. [Online]. Available: https://hpc.llnl.gov/training/tutorials/introduction-parallel-computing-tutorial

[18] B. Palmintier, K. Horowitz, S. Abraham, T. Elgindy, K. S. Sedzro, B. Sigrin, J. Lockshin, B. Cowiestoll, and P. Denholm, "Chapter 7: Distribution System Analysis," in *The Los Angeles 100% Renewable Energy Study (LA100)*, J. Cochran and P. Denholm, Eds. Golden, CO: National Renewable Energy Lab, Mar. 2021. [Online]. Available: https://www.nrel.gov/docs/fy21osti/79444-7.pdf

[19] Subprocess management. Accessed: Dec. 10, 2021. [Online]. Available: https://docs.python.org/3/library/subprocess.html

[20] SchedMD. Slurm workload manager. [Online]. Available: https://slurm.schedmd.com/

[21] OEDI. SMART-DS. Accessed: Dec. 10, 2021. [Online]. Available: https://data.openei.org/submissions/2981