SAND2017-6197C

Intelligent Sampling of Periods for Reduced Computational Time of Time Series Analysis of PV Impacts on the Distribution System

Jason Galtieri, Matthew J. Reno

Sandia National Laboratories, Albuquerque, NM, 87185, USA

***Abstract* — In this work, a sampling method, known as intelligent sampling (IS), is presented to reduce simulation time in Quasi Static Time Series (QSTS) analysis on electrics grids with distributed PV. The sampling method decomposes a year's worth of input solar and load data into six hour intervals and bins the intervals according to irradiance and load metrics. Representative samples are chosen from the bins and simulated using standard power flow solvers. We show that when using the IS method, only a fraction of the total entries in the year need to be simulated. An example test circuit is used and the IS method achieves a 57% reduction in simulation time while meeting acceptable error margins.**

I. Introduction

The addition of renewable and distributed generation on the electric power system has altered traditional control techniques and grid analysis methods. Historically, in distribution feeders, power flows from the substation to the various loads along the feeder length with the voltage regulators reacting to changes in load and self-correcting to maintain normal operating voltages. Increasing penetrations of distributed PV can create significant power output fluctuations and reverse power flows along specific segments of the feeder, causing voltage limit violations and additional wear and tear on voltage regulation equipment [1].

In order to model the variability of distributed PV, high-resolution quasi-static time-series (QSTS) simulations are required to simulate the grid impact at different times of year and to determine any interactions between PV and existing voltage regulation equipment [2]. To capture the interactions and seasonal variations, accurate QSTS simulation should be performed at high-resolution (<5 second time-step) and for the duration of a year [3]. Certain QSTS metrics such as extreme voltages and line losses can be approximated using relatively large time steps, but voltage regulators and capacitor switching require time-steps on the order of one to a few seconds. These types of QSTS simulations are computationally intensive and can take days to perform for large distribution system models. The computational burden limits the practicality of QSTS for parametric analysis [4] or for the hundreds of PV interconnection requests that a utility receives.

There has been limited research into improving the speed of QSTS simulations [5]. Due to the unbalanced nonlinear nature of the distribution system power flow equations, it can be quite challenging to decrease the computational time [6]. Additionally, the voltage regulation devices have time delays, deadbands, and hysteresis that require each power flow to be solved sequentially in order [6].

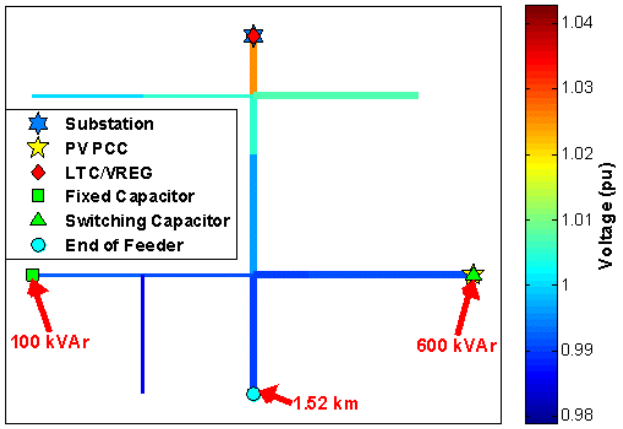


Figure 1. Diagram of the modified IEEE 13-node feeder colored by voltage.

This paper proposes a novel method to intelligently perform QSTS simulation for part of the year while accurately modelling the PV impacts over the entire year. The intelligent sampling (IS) algorithm works by analyzing the simulation input data and selecting a representative sample of inputs throughout the year. This reduced input list is then simulated using a powerflow solver and results are scaled to estimate the yearly simulation results. In general, [3] demonstrated that solving the QSTS for half the days randomly sampled results in situations with potentially very high error compared to running the entire year. This is explained by the logic that randomly sampling could sometimes miss certain common situations (e.g. sampling only clear days, or sampling entirely from winter months). The purpose of proposed intelligent sampling method is to ensure the full range of days throughout the year are analyzed in detail with the QSTS simulation, and then the results for the non-simulated days will be inferred by the simulation results for similar type days. The intelligent sampling method is explained in Section III, and Section IV provides a demonstration of the errors introduced by the method and the reduction in computational time.

II. Test Setup

The modified IEEE 13-bus circuit shown in Figure 1 is used in this work. A 2 MW PV plant is located at the end of the feeder and accounts for up to 40% of the peak load. Measured 1-second resolution solar irradiance data and 5-minute measured load data are applied as the inputs to the simulation. The effectiveness of intelligent sampling will be determined by comparing the algorithm's outputs with the actual results attained from the yearlong QSTS simulation. The brute-force

simulation is performed at 1-second resolution using OpenDSS. Due to the significant impact variable PV generation can have on distribution system voltage regulators, the number of tap changes predicted for the year is used as the simulation accuracy evaluation metric. Based on feedback from distribution system engineers, the expected accuracy of number of regulator tap changes in a year should be within 10% of the detailed brute-force QSTS simulation and will be the focus of this work [3].

III. Intelligent Sampling Methodology

The brute-force QSTS simulation results are saved into 6-hour periods (e.g. the number of tap changes per 6-hours), resulting in 1460 6-hour periods in a year. The objective of the intelligent sample selection is to select which of those 1460 6-hour periods are the most effective to simulate with QSTS to estimate the yearly impacts.

A. Categorizing Input Data

The input data time-series profiles are analyzed using irradiance variability metrics found in the literature and simple load metrics such as the maximum, minimum, mean, and median over a given time period. The irradiance variability index (VI) [7] and irradiance variability score [8] were used for this work. Variability metrics were chosen, as opposed to static metrics, from the intuition that voltage regulator operations are caused by grid dynamics. Based on the simulated data, a stepwise linear regression was performed to identify the two statistics that are the mostly highly correlated with the number of regulator tap changes. For this work, it was determined that the VI and the median of the load for each period were the most highly correlated statistics from the input data timeseries.

Intelligent sampling (IS) decomposes the year into the smaller 6-hour time periods. Each of these time periods is then categorized and binned according to the calculated statistics from the input time-series during this period. The intuition is that "similar" time periods (i.e inside the same bin with the same input time-series statistics) will have closely correlated QSTS results. Due to the focus on variability, the length of the time period interval plays an important role. Too large a time interval will average out periods of high variability with idle periods, while very short intervals lack enough data to be able to differentiate points of interest. A 6-hour interval is chosen to split the day into quarters. The middle two quarters encompass most of the daylight hours while the first and fourth quarters cover morning and nighttime, respectively.

Several binning techniques are available to group data with two input metrics for sampling. The simplest binning technique is to lay a grid across the 2D-plane and treat the resulting, equally sized, rectangles as individual bins, where samples are pulled per bin. This method is called stratified sampling and has been previously demonstrated for sampling representative

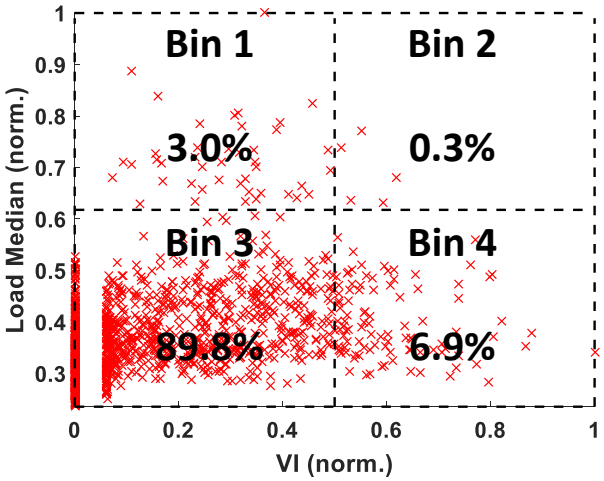


Figure 2. Four bin grid example using VI and load median as input metrics

days for simulation [9]. Empty bins, with zero entries, are possible with this approach but only "filled" bins are sampled.

Another binning strategy, known as K-means clustering, employs an iterative approach to create clusters where the intra-cluster Euclidean distances between samples are minimized. Resulting clusters will have unique shapes and sizes. Many other binning techniques are possible by considering different geometric bin shapes and sizes.

A grid-based stratified sampling approach is used for this work. Being the simplest approach, it is easily adaptable to different circuit topologies, whereas K-means approaches suffer replicability and tuning issues, as the initial clusters locations are typically randomly placed. Clusters are also heavily influenced by outliers, which can be mitigated by removing outliers, but we do not want to make any assumptions beforehand on the designation or significance of outliers. Ultimately the grid-based stratified sampling approach requires the least circuit tuning, with robust and repeatable results.

After the bins are compiled, a determined number of sample periods are randomly chosen from each bin and simulated with the QSTS power flow solver. These simulation results are used to establish an estimated average bin output ($\hat{B}_i$) for each bin (i). The estimated total yearly tap changes ($\hat{y}$) is calculated according to (1), where n_i is the total number of entries in bin i, s are individual bin samples, and v_i is the number of samples drawn from B_i.

$$\hat{y} = \sum n_i \hat{B}_i \tag{1}$$

$$\hat{B}_i = \frac{s_1 + \cdots + s_v}{v_i}, v_i \leq n_i \tag{2}$$

The accuracy of the yearly estimations will be determined by how close the estimated bin mean, $\hat{B}_i$, is to the true bin mean, $\bar{B}_i$. Simulating more bin samples will always increase $\hat{B}_i$'s accuracy, but at the expense of simulation time. From Figure 2, it is apparent samples are not spread evenly among the bins with most falling in bin 3. For an accurate representation,

sampling also may not be even among the bins. One sample from bin 2 may suffice to estimate $\bar{B}_2$, but estimating $\bar{B}_3$may require more samples. The challenge then becomes determining the number of samples to draw from each bin to establish a good estimate of the bin means, especially as the number of bins increases.

Figure 3 shows a binning example using a 21x21 grid. Figure 3a shows the average of the number of tap changes that occurred during the 6-hour periods for that bin. Figure 3b displays the number of 6-hour samples with the statistics that correlate to that bin. Although there is some correlation between input values and bin means, it is quite apparent that the relation is not linear due to the nonlinear power flow equations and piecewise discontinuities from the discrete operating states of the voltage regulators. We cannot apply a least squares fit to the input/output data, and any fit would be circuit dependent, requiring a large amount of simulation data to create the fit model. Figure 3a shows that the input metrics by themselves cannot accurately predict the tap change results, but the metrics can be used for intelligent stratified sampling.

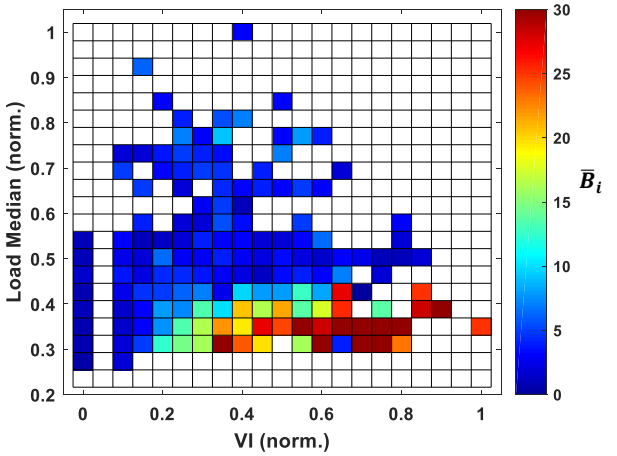
B. Sampling from Bins

Several sampling methods were tested to try to reduce the simulation time as much as possible while keeping errors within the 10% tolerance. A key focus was to limit the number of assumptions that may or not hold across multiple circuits. This was also the reason we chose a simple binning technique as opposed to a complex one. The complex binning processes could be tuned extensively for our given circuit and yield good results, however, there was no certainty the tuning parameters were universal.

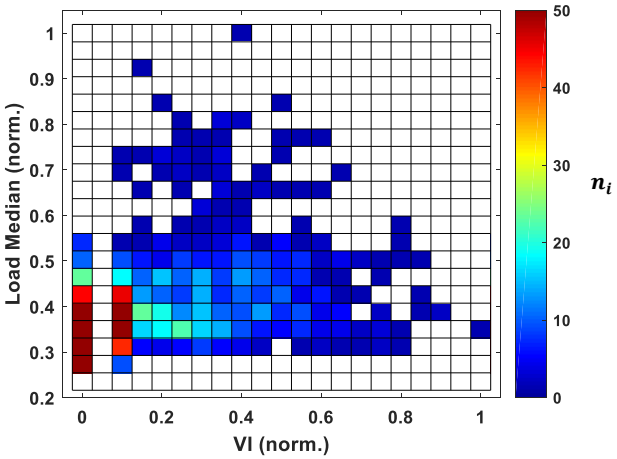
On the same note, we did not make too many assumptions about the impact of outliers on sampling. Visually, outliers are easily identified in Figure 2 at high load and VI metrics. When grid size increases, outliers are typically the only sample in their bin and their $\bar{B}_i$ is attained by simulating the one sample. However, (1) small bin sizes (low n_i) may or not may contribute much significance to the overall year metric.

Even with the relatively small grid size in Figure 3, bin counts (number of samples in the bin) vary unpredictably across the range of inputs metrics. The high concentration when VI is zero is due to the dark time intervals, when the sun is not up and the solar variability is zero. Comparing Figure 3a and 3b, there is little correlation between bin mean and bin count. Outlier bins with few samples on the edge display the largest range in the average number of tap changes. Techniques such as outlier removal, undersampling or oversampling outlier bins, or combining outlier bins would have to make assumptions on the importance of outliers, which will not be consistent for different distribution systems being analyzed with QSTS.

Another idea considered was to undersample the bins with low intra-bin variability between samples in the bin. Of course, this is a very limited approach since it assumes the true means



(a) Bin Mean Regulator Tap Changes



(b) Bin Counts

Figure 3. 21x21 grid bin example: (a) Bin Sample Counts (b) Bin Mean Regulator Tap Changes. White squares are empty bins

are known beforehand, but some intuition narrows the focus. Bins where the VI index is zero, corresponding to dark time periods in early morning or late at night should have low PV induced variability in their output metrics. These time intervals are also numerous and take up about half the year. Therefore, simulation time can be greatly reduced by selecting only a couple samples from these bins to estimate the bin mean. However, as given by (1), bins with high sample counts get a larger weighting to estimate the yearly averages. Even small variations in the bin estimate are amplified by the high bin count.

Using the same 21x21 grid as before, Figure 4 shows the individual bin's standard deviation (σ_B) of the number of tap changes recorded for all samples in the bin multiplied by the bin count. The product term gives an approximate error margin on sampling a single or low number of entries from each bin. Large error margins are seen across the spectrum of bins, independent of bin size. For the above reasons, we concluded that an accurate representative sample needs to span the entire range of samples in the year.

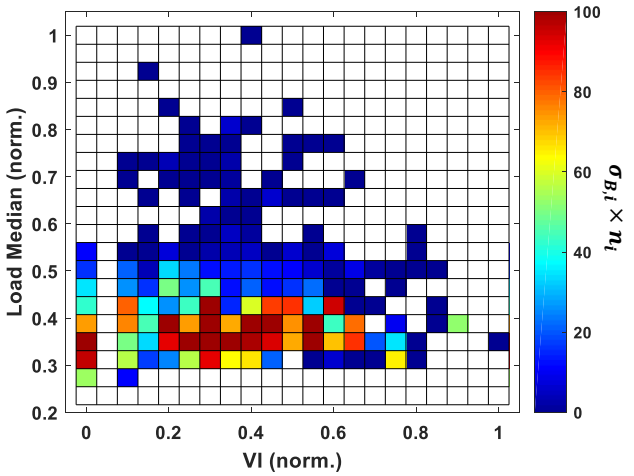


Figure 4: 21x21 grid bin example for number of regulator tap changes. Color code shows each bin's $\sigma_{B,i} \times n_i$ as an approximate error margin on the bin. White squares are empty bins

A considerable effort was spent trying to develop convergence methods to test whether the individual $\hat{B}_i$ were within the error tolerance of their true values. Rather than running simulations for a fixed sample size, the sample number would increase until convergence was detected. Some iterations would converge quickly while others needed additional samples. The intuition was that fixed simulation times had to be sized to bound the worse-case scenarios, while a variable method would display a shorter mean time to convergence. Some of the convergence methods considered were: tracking changes in $\hat{B}_i$ when adding additional samples, tracking changes in $\hat{y}$ yearly estimates when adding additional samples, and performing post-processing on samples to decrease sample variability. However, we were unable to find a 100% reliable convergence method and leave this to future work.

In the chosen IS method, the number of samples chosen from the bins is determined by the bin count and the total number of samples desired. For example, if 50% simulation time reduction is desired, then the number sampled from the bin is simply the bin count divided by two. We determine the number of samples (τ_i) taken from the *i*th bin by

$$\tau_i = \left\lceil \frac{n_i}{x} \right\rceil \qquad (3)$$

where the ceiling function guarantees at least one entry from each nonempty bin. The variable *x* is a tuning parameter which is used to get the sum of $\tau's$ as close to the target sample size as possible. The optimal value of *x* is dependent on the number of non-empty bins, as well as the target sample size, and is found iteratively.

For simplicity, the bins are always kept as a square grid of equal proportions. As a result, the number of bins is always equal to the square of the grid size. Grid size plays an important role in intelligent sampling. If the grid includes a large number of bins, then there are too many single entry bins, each of which must be included based on the previous outlier discussion, and too many total samples will be included. However, with few bins, there will be too much variability for the entries inside the bin and our "similar" samples hypothesis fails. The algorithm tries to find the largest grid size that can reach the target sample size.

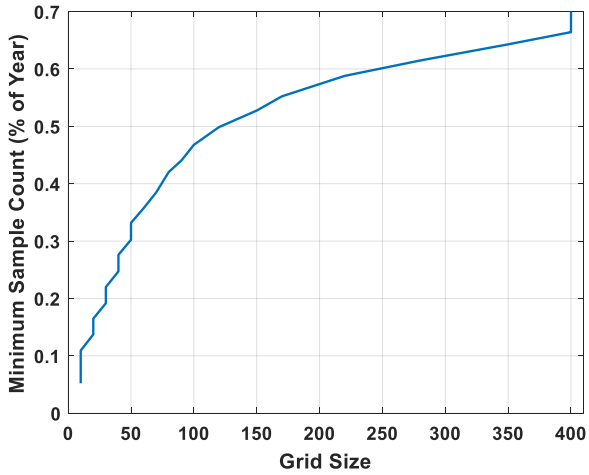


Figure 5: Minimum attainable sample count as a function of grid size

With the sampling constraint in (3), the grid size becomes closely tied to the minimum samples drawn, as illustrated in Figure 5. Due to the relatively small 1460 sample count (number of 6-hour periods in 365 days), even the smallest grid size at 10x10 has at least 60 filled bins, corresponding to about 4% of the year. Smaller grid sizes introduce more intra-bin variability and weaken our initial binning assumption that intra-bin variability is minimal. With more intra-bin variability, additional samples needs to be drawn from the bin for $\hat{B}_i$ to accurately estimate $\bar{B}_i$. As a result, sampling smaller grid sizes becomes difficult to accurately estimate bin sizes while keeping the total sample size low. On a note, we tried a variety of sampling methods without the "one sample per bin" requirement. However, in these methods the unsampled filled bins had to be accounted for with a multiplier variable, which introduced too much error.

IV. Results

The effectiveness of the intelligent sampling method is analyzed using a Monte Carlo (MC) simulation. For each MC simulation, the mean absolute error (MAE) is calculated between the actual yearly regulator tap changes ($\hat{y}$) and the estimated yearly regulator tap changes ($\bar{y}$) averaged between the three regulators.

$$\text{MAE} = \frac{1}{3}\left(abs(\hat{y}_1 - \bar{y}_1) + abs(\hat{y}_2 - \bar{y}_2) + abs(\hat{y}_3 - \bar{y}_3)\right) \qquad (4)$$

The goal is to find the smallest number of days necessary where the MAE is less than 10%. Additionally, we want to always perform within the error threshold, so the maximum of the MAE from all MC simulations is calculated as well. If the worst-case MAE from the MC is not bounded, then we cannot

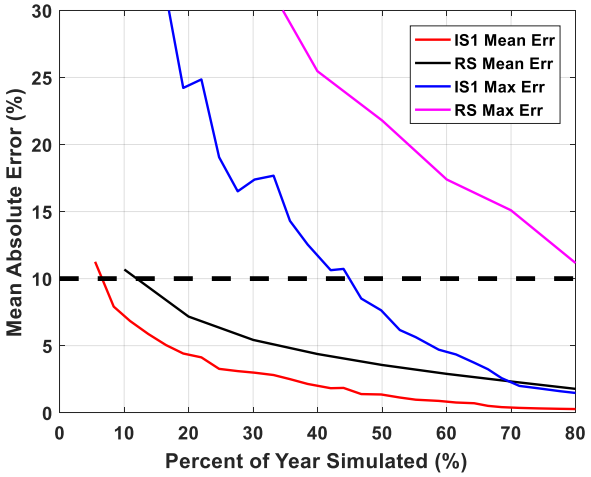


Figure 6. Mean absolute error for voltage regulators's yearly tap changes. Intelligent sampling method (IS) is compared with random sampling (RS)

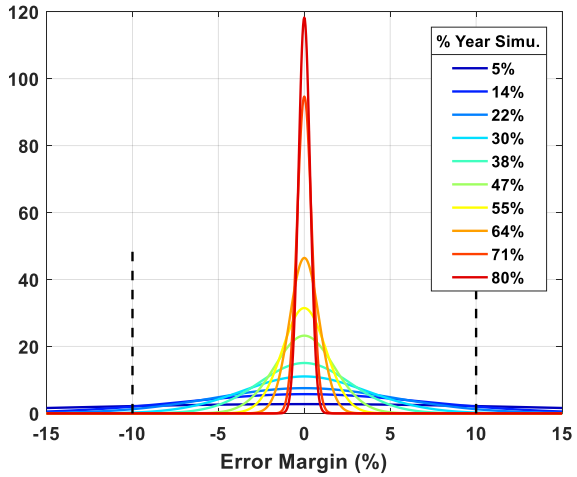


Figure 7. Error margins for regulator one tap changes using IS algorithm

have confidence that results from a single intelligent sampling selection are not outliers with extreme error.

The intelligent sampling results are compared to randomly sampling (RS) time periods out of the entries to ensure there is merit to binning inputs. Random sampling was tried in [2] and found to have too of large error margins to be effective.

The MC simulation is run for 100,000 iterations at different target sample sizes and the results for the sampling method is displayed in Figure 6. The MAE are calculated for each MC iteration and the means of the iterations are shown for the IS and RS algorithms. The maximum MAE among the MC simulations are also displayed for two methods. The goal is for the mean and max MAE to be below 10% which is based on feedback from distribution system engineers [2].

The mean MAE are relatively low and similar for the IS and RS algorithms. The means for IS and RS cross the 10% error tolerance at 7% and 12 %, respectively, of the year simulated. Above 30% of the year simulating, further increases in sample size has only marginal effects on the mean MAE. On the other hand, the maximum error is closely tied to sample size. The maximum error for the IS algorithm crosses the 10% threshold when approximately 43% of the time periods are selected to be simulated with QSTS. The RS algorithm's maximum error crosses the 10% threshold around 85% of the year sampled and is more heavily influenced by outliers. To limit the potential errors, RS requires nearly twice as much data, which shows the IS algorithm has significant benefit compared to simple random sampling.

Although Figure 6 shows the maximum error, it does not indicate just how many iterations are falling outside the 10% error margin. For this we plot the distribution of errors for the 100,000 MC simulations using IS in Figure 7. The low samples sizes (<20% of year) have relatively flat error distributions with large parts of the tail falling outside the 10% tolerance. However, the 20-30% samples sizes fall mainly inside the margins with a small fraction actually violating the tolerance. Although previous constraints called for error to be entirely bounded, the improvement in simulation time may be worth the slight relaxation in error constraints. For example, a large parametric QSTS, with long simulation times, may trade some accuracy for increased speed.

V. Estimating Sample Errors

The very small percentage of MC simulations that are outside the error tolerances in Figure 7 suggests that lower samples sizes (20-30%) are possible if that small fraction of erroneous sampling scenarios is somehow detectable. Several convergence methods were experimented with, but ultimately it is difficult to accurately determine convergence for such a limited sample size. Sampling 30% of the year is approximately 400 samples, but due to the intelligent stratified sampling, each sample represents a unique set of sampling conditions based on that bin. Most bins end up with only two or three samples in them, so it is not possible to detect the difference in accuracy or convergence between two or three samples. This may be a limitation of the binning where, at low sample size and grid size, there is always a particular subset of samples that give high error. For example, all the bins have some intra-bin variability and if the chosen samples are all skewed the same way (high or low), the overall output metric will be skewed as well.

One convergence strategy tried was to post-process the samples drawn using bootstrapping to determine the variability of the sample population. The idea was that the width of the distribution of yearly predictions from bootstrapping the samples would give an indication on the convergence to the true population mean. A fixed number of samples were initially drawn with IS, with at least one sample per bin. Next, using that sample population, smaller subsets were randomly drawn from the initial sample and used to calculate bin estimates and corresponding $\hat{y}$. Bootstrapping to repeat this secondary sampling multiple times gave us a distribution for each $\hat{y}$ for

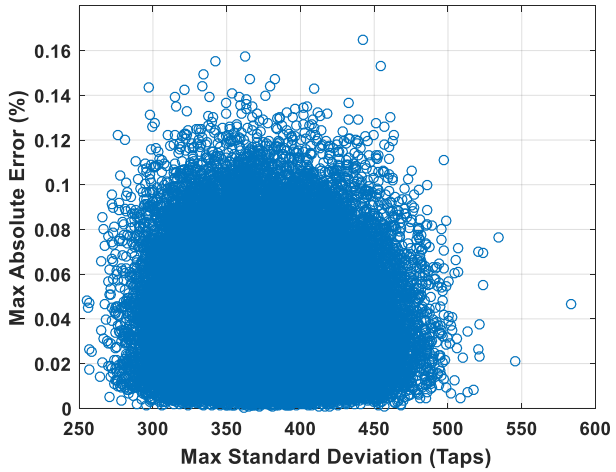


Figure 8: Bootstrapping convergence method showing sample standard deviation versus max error.

which we could find the mean and standard deviation. The goal was to use the standard deviation of the bootstrapping as the convergence metric. Results for a MC simulation of the above method are shown in Figure 8. Interestingly, the MC simulations where the true error fell outside the 10% margin do not typically exhibit very large standard deviations using bootstrapping. This seems to suggest that particular sample subsets are precise but not accurate and cluster tightly around an incorrect sample mean. For this reason, the convergence test is not reliable. Future work may involve trying to detect when a skewed sample is drawn from the bins.

VI. Conclusion

A new intelligent sampling algorithm has been presented to reduce simulation time for yearlong QSTS distribution system analysis with high penetrations of distributed PV. The algorithm analyzes the input irradiance and load data and selects representative time periods to simulate using QSTS. By simulating these representative samples, which accounts for a fraction of the total days of the year, the simulation time is decreased. The goal for this work was to estimate the number of regulator tap changes in a year within 10% error margins, and to achieve a 50% reduction in simulation time. With intelligent sampling, we demonstrate a 57% reduction in simulation time that meets error tolerances with the test circuit. The algorithm is also compared with a random sampling algorithm to demonstrate a significant improvement in the maximum possible sampling errors.

Future work will continue to develop ideas to calculate stopping conditions for sampling during the simulation by determining statistical convergence in the confidence interval around the true answer. While intelligent sampling only reduces the computational time of QSTS simulations to around 50% of the brute-force yearlong simulation, IS methods can easily be incorporated into other algorithms. For example, IS can select the days in the year to simulate, and then QSTS simulation of those days can use a variable time-step method [5] for additional speed. We will also investigate using the intelligently selected sample periods to train machine learning algorithms to model the correlation to the number of tap changes [10].

References

[1]. Palmintier, R. Broderick, B. Mather, et al., "On the Path to SunShot: Emerging Issues and Challenges in Integrating Solar with the Distribution System," National Renewable Energy Laboratory, NREL/TP-5D00-65331, 2016.

[2] R. J. Broderick, J. E. Quiroz, M. J. Reno, A. Ellis, J. Smith, and R. Dugan, "Time Series Power Flow Analysis for Distributed Connected PV Generation," Sandia National Laboratories, SAND2013-0537, 2013.

[3] M. J. Reno, J. Deboever, and B. Mather, "Motivation and Requirements for Quasi-Static Time Series (QSTS) for Distribution System Analysis." *IEEE PES General Meeting*, 2017.

[4] J. Seuss, M. J. Reno, R. J. Broderick, and S. Grijalva, "Analysis of PV Advanced Inverter Functions and Setpoints under Time Series Simulation," Sandia National Laboratories, SAND2016-4856, 2016.

[5] M. J. Reno and R. J. Broderick, "Predetermined Time-Step Solver for Rapid Quasi-Static Time Series (QSTS) of Distribution Systems," IEEE Innovative Smart Grid Technologies (ISGT), 2017.

[6] J. Deboever, X. Zhang, M. J. Reno, R. J. Broderick, S. Grijalva, and F. Therrien "Challenges in reducing the computational time of QSTS simulations for distribution system analysis," Sandia National Laboratories, SAND2017-5743, 2017.

[7] J. S. Stein, C. W. Hansen, and M. J. Reno, "The variability index: A new and novel metric for quantifying irradiance and PV output variability." World Renewable Energy Forum. 2012.

[8] M. Lave, M. J. Reno, and R. J. Broderick, "Characterizing local high-frequency solar variability and its impact to distribution studies." *Solar Energy* 118 (2015): 327-337.

[9] B. Palmintier, J. Giraldez, K. Gruchalla, et al., "Feeder Voltage Regulation with High-Penetration PV Using Advanced Inverters and a Distribution Management System: A Duke Energy Case Study," National Renewable Energy Laboratory, NREL/ TP-5D00-65551, 2016

[10] M. J. Reno, R. J. Broderick, and L. Blakely, "Machine Learning for Rapid QSTS Simulations using Neural Networks," IEEE Photovoltaic Specialists Conference (PVSC), 2017.

This research was supported by the DOE SunShot Initiative, under agreement 30691. 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.