

Interim Report for Task 4.2: PV Variability Analysis

Michaelangelo Tabone, Duncan S. Callaway and Daniel M. Kammen
University of California, Berkeley

February 27, 2011

This is one of three interim reports completed by a UC Berkeley team of students and faculty in the Energy and Resources Group. The overall objective of the project is to determine the impact and value of coupling distributed storage with photovoltaic systems. Our specific focus is on understanding distribution system impacts and the opportunity for creating value by incorporating storage into the CAISO's dispatch process.

The tasks of the project are

- 4.2 PV Variability Analysis
- 4.3 - CPP Tariff
- 4.4 - Aggregate control
- 4.5 - CAISO Product

This particular interim report documents our efforts to study the impact of distributed PV portfolios on grid operations and planning. The impacts of distributed PV will be assessed both for existing SolarCity installations then predicted for an increased build out of distributed PV. Thus far, a method for assessing the existing impact, and possible methods for predicting additional impacts have been established using data from 16 SolarCity installations in the Los Angeles area taken at roughly one minute intervals collected for one week. The final results of this study will use one minute interval data from 100 installations, collected for one year, in the vicinity of San Jose and the Fresno. The method for assessing existing impact quantifies variability impacts on short term system balancing, distribution voltage control, and the likelihood of backwards flow on a feeder. Predictive methods for increased distributed PV will use a relationship of correlation and distance and a predicted distribution of the aggregate generation signal.

1 Background

The power systems literature has addressed the integration of intermittent renewable energy since the 1970's. Early papers were concerned with reliability problems that would result from rapid fluctuations in large solar and wind installations, akin to transmission losses or unscheduled plant outages. However the effect of geographic diversity (even within the same wind farm) was shown to dramatically reduce the magnitude of fluctuations from large installations. For a comprehensive review of early renewable integration studies, see Mills and Wiser 2010 [1].

1.1 System-wide integration studies

Most recent integration studies have focused on the operational impacts of intermittent generation. Specifically, the work assesses the impact on system balancing authorities, which will need to secure fast responding dispatchable resources in order to respond to significant fluctuations resulting from additional intermittent resources. A major contribution of these studies is the distinction between variability and uncertainty. Uncertainty refers to disparities between predicted and actual output from renewable generators while variability refers to rapid fluctuations in output that require flexibility even if they are accurately predicted.

The Western Wind Integration Study and its California specific predecessor, the Intermittency Analysis Project, both predict the impact of increasing renewable electricity on the power grid [2, 3]. These studies assess the distribution of hourly changes in overall net load at varying timescales and different amounts of installed wind and solar generation. Hourly ramps are separated by hour of day and by season to outline systematic ramps that are correlated due to use patterns or daily weather patterns. Annual hourly deltas of more than 3 standard deviations from the mean are counted as a measure of extreme events. A similar analysis is completed for 10 min deltas. In the California study, the impact of sub-hourly variability is measured using a 15 minute moving average which represents total dispatched generation beyond which automated generation control (AGC) is needed to adjust total generation to real time load. Changes over 5 minutes in this moving average are used as a likely ramping requirement for generation on the 5 minute market. Deviations between the 15 minute moving average and real-time load are used to simulate the needs of AGC on a minute-to-minute basis. Figure 1 is reproduced from Bai et al. to show how metrics are drawn from the 15 minute moving average [3]. In the Western wind integration study, the impact of sub-hourly variability is assessed using short time-scale dispatch simulations and minute-by-minute wind data obtained from NREL.

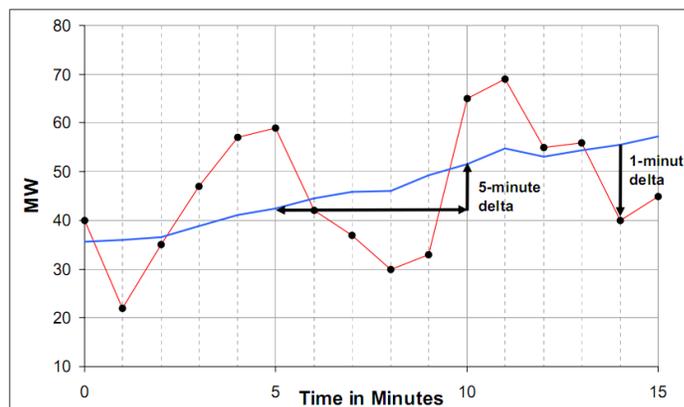


Figure 1: Figure depicting the calculation of 1 min and 5 min deltas in the California Intermittency Analysis Project. 1 min deltas represent those used for AGC, 5 min deltas represent a 5min market.

More recent work in California has attempted to isolate the operational requirements that result from variability and uncertainty in renewable generation [4, 5, 6]. Makarov 2009 outlines the basic method for this analysis. First, the net load (ignoring transmission and distribution constraints) is projected for each minute over the course of a year by subtracting projected wind generation from projected load. Load is determined by applying an annual growth rate to the 2006 aggregate load profile. Predicted wind generation was developed by AWS truewind. The impact of variability and uncertainty on power capacity needs in the 5 minute ahead real time market and regulation capacity markets are calculated by subtracting the total dispatched generation from the net load. Total dispatched generation is calculated

as described in Equations 1 and 2 where D represents the total dispatched generation at either the hour ahead, ha , point or the 5min market, $5min$; $avg_T(L)$ refers to the interval average of net load, $L - W$, at intervals of T ; \mathfrak{R}_{T_i} is an operator which adds ramps of length T_i between hourly averages; and ϵ represents forecast error.

$$D_{ha} = \mathfrak{R}_{20min}\{avg_{60min}(L - W) + \epsilon_{L,ha} - \epsilon_{W,ha}\} \quad (1)$$

$$D_{5min} = \mathfrak{R}_{5min}\{avg_{5min}(L - W) + \epsilon_{L,5min} - \epsilon_{W,ha}\} \quad (2)$$

The required signal for the real time market is the difference between the hour ahead schedule and the real time market schedule, $S_{rtm} = D_{ha} - D_{5min}$. The required signal to be followed by AGC is the difference between the net load and the generation dispatched in the real time market, $S_{agc} = L - W - D_{5min}$. The maximum and minimum of S_{rtm} and S_{agc} are presented for each hour of the day in each season. Stratifying by hour of day and season controls for systematic effects due to diurnal or seasonal patterns. Figure 3, explained in the method section, provides a graphical representation of Makarov's method.

Makarov 2009 also isolated the likely ramps of extreme change and duration in order to identify requirements for generators participating in the real time market or AGC. In the swinging window algorithm a ramp in the net load signal is considered to continue between two subsequent readings if every measurement between the minutes is within some error ϵ of a linear interpolation between the readings at each minute. This method is explained in detail in Section 3.2.

The California integration work builds on previous work by done by Kirby Consulting for the Oak Ridge National Laboratory that assessed customer contribution to ancillary service requirements for AGC and the real time market [7]. This report uses a the deviation of customer load from a moving average to estimate total ancillary service (AS) requirements if only that customer's load was being followed. The system then uses the correlation between each customer's AS signal to determine to what extent each customer contributed to the *total* AS signal, where more correlated loads contribute a larger percent of their overall signal. When assessing the regulation contribution of one signal among a group of signals, the signal must be considered in isolation against the aggregation of all other signals.

1.2 Geographic diversity

The effect of geographic diversity has been well documented in literature, though many definitions of variability are used: unfiltered PV generation signals are used in many papers including [8, 9]; the maximum fluctuation during a moving window is used in Murata 2009 [10], differenced signals are used in Mills and Wiser 2010 [1]; an autoregressive moving average model is used in Choudhury and Rahman [11]; and frequency domain periodic analyses are used in other sources [12, 13, 14]. Regardless of the metric used, many studies find a relationship between the correlation of variability between sites, the distance between sites, and the timescale of the variability measured, shown in Figure 2 where the shorter distance between sites and longer timescales of variability lead to greater correlation of variability between sites. [1].

1.3 Impacts of variability on small geographic areas

Aggregate variability of renewable generation will have a much different impact on smaller geographic areas than on larger ones due to the decreasing effect of geographic diversity, i.e., solar generation located within a radius of a few kilometers will be much more correlated than solar generation located over the whole of California. Kirby and Milligan show that balancing authorities with smaller footprints

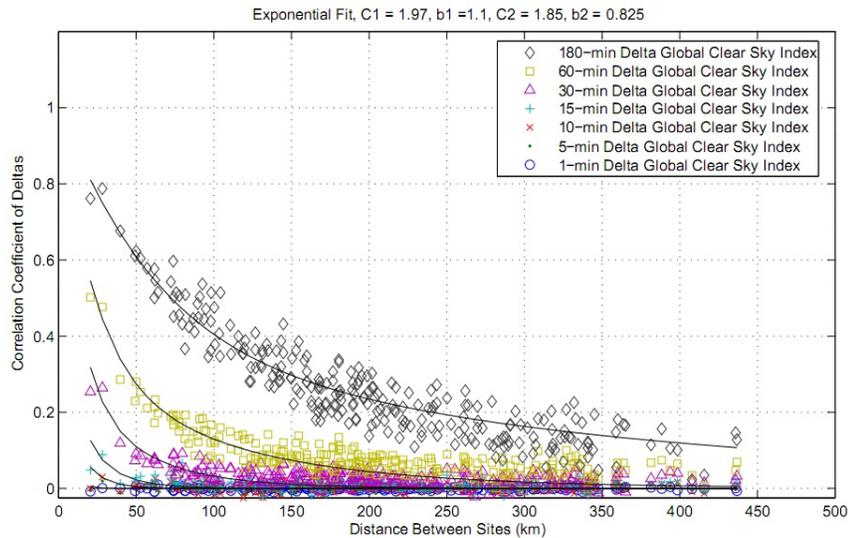


Figure 2: Correlation of variability between pairs of sites as related to the distance between sites and the timescale of variability. Reproduced from Mills and Wiser 2010

will likely have more difficulty managing renewable variability [15]. Small area geographic diversity will also affect the impact of distributed PV on the distribution system.

Studies on the impact of PV in the distribution system have been limited, though there is significant worry that rapid fluctuations in consumer load (resulting from solar PV) will lead to insignificant voltage control, increased short circuits, and destabilization of protection measures. Liu and Bebic perform steady state voltage analysis on a radial distribution circuit and finds that all voltage issues can be managed through additional installed voltage support as well as possibly with support from PV inverters [16]. Teleke et al., perform a dynamic analysis of a distribution circuit under an extremely rapid and large fluctuation in load: approximately a 250kW to a 1MW change over 1 second. The study finds that voltage fluctuations are manageable in the modeled circuit (that contains secondary circuits) however multiple changes are required of tap changing transformers over a time period of 1-2 seconds.

2 Method

In this section we will describe methods for (1) measuring the impact of a variable generation signal, and (2) predicting the properties of aggregated PV generation signals under higher levels of penetration. Variability metrics are intended to be proxies for the impact that an additional variable signal will have on grid operations and planning. This paper outlines metrics related to the impacts of short time scale variability (1 to 30 minutes in duration) in 4 distinct areas:

1. The capacity requirements for regulation (i.e. automatic generation control or AGC) and real time markets (load-following).
2. The ramping requirements for power generators participating in AGC or load-following.
3. The potential for out-of-bounds voltage swings due large load fluctuations and insufficient voltage support.

2.1 Capacity requirements for system balancing

The impact of an additional signal on the existing capacity requirements is evaluated using a combination of the method described in Makarov *et al* (2009) for evaluating requirements and the method in Kirby (2000) for relating the AS requirements of independent signals to the AS requirements of an aggregate signal [5, 7].

Makarov *et al*'s analysis is performed on the generation signal from each individual PV site, as shown in Figure 3. In Figure 3, Part A displays the cumulative dispatched generation after the hour ahead schedule (red line), the 5 min ahead schedule (black line), and the AGC response (blue dots). The hour ahead schedule and the 5min ahead schedule are calculated as described in Equations 1 and 2 respectively, where forecast errors are set to 0 in order to isolate variability without forecast uncertainty. The cumulative generation following the AGC signal is equal to the PV generation signal exactly. Parts B and C of Figure 3 display the resulting signal required of generation in the load-following market and generation participating in AGC respectively.

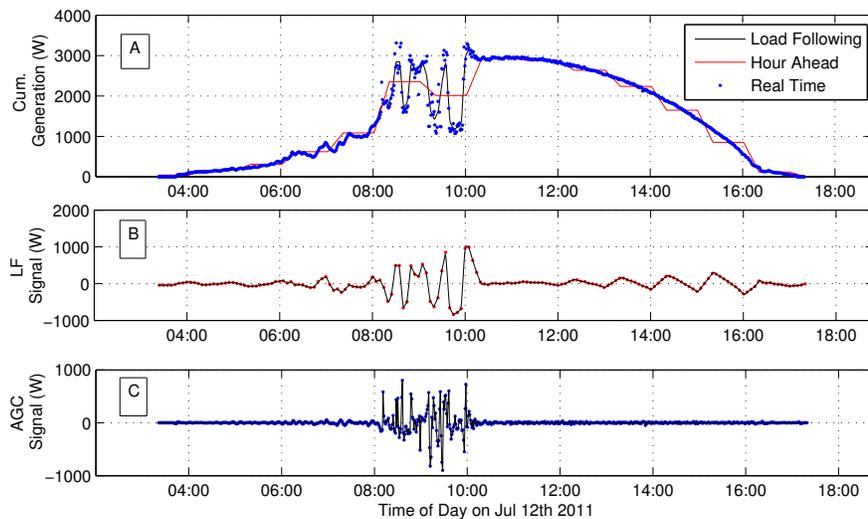


Figure 3: Makarov Example, method for calculating the AGC and the Load-following signal.

The maximum, minimum and the standard deviation of the load-following and AGC signal at each hour will be used to evaluate the system balancing requirements of the aggregate signal on the grid. This will be sufficient for measuring how the marginal system balancing requirements change as more distributed PV is added to an area, however the actual requirements have to be related to total system load before it can be used to calculate the costs of balancing additional PV.

2.2 Ramping requirements for system balancing and the probability of voltage fluctuations

The swinging window algorithm from Makarov *et al* (2009) is used to assess the sustained ramp in a signal at durations intervals of 1 minutes to 12 minutes. Figure 4, reproduced from Makarov 2009, graphically displays the algorithm. The average rate of generation change between the points 1,3; 1,4; and 2,4 are all valid, sustained, ramps; because each point between the two bounding points falls within

a margin of error of a linear interpolation (known as the swinging window). However, the interpolation between point 1 and 5 is not a sustained ramp, because point 4 lies outside of the window [5].

Sustained ramps in the regulation and the load-following are measured to test the physical requirements of generators participating in Ancillary Service programs, where the sum of the ramping ability of each available generator should match the likely ramping requirements of the ancillary service signal.

Sustained ramps in the overall signal are measured to test the likelihood of a voltage fluctuation on a feeder, where large isolated ramps on circuits and sub-circuits can require additional equipment (or more rapid equipment) for reactive power compensation and voltage control.

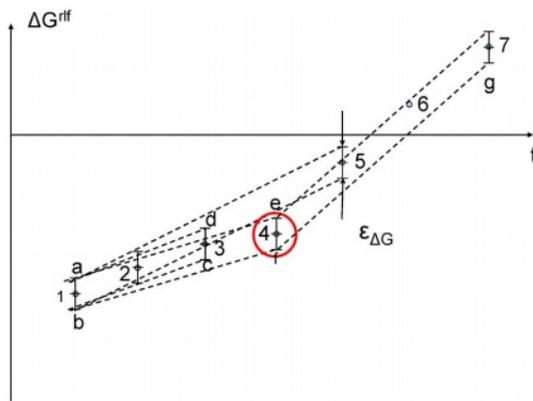


Fig. 7. Concept of the “swinging window” algorithm.

Figure 4: Swinging window algorithm (see [5] for full details)

2.3 Predictive Methods

Two methods are identified for predicting the impacts of increased penetration of distributed generation:

1. Use a relationship between the effective number of sites, n_{eff} , and the normalized impact in a given area to find the impact of any fleet of PV within the studied area.
2. Use relationship of signal correlation to distance, summation of variances, and an expected distribution shape from a fleet of PV sites, to estimate the distribution of impact in a fleet of PV installations.

Method 1 benefits from empirical simplicity however will not generalize to areas larger than (or outside of) the studied area. Method 2 allows for the analysis of any size area, however relies on strong assumptions that need to be thoroughly examined. Preliminary results from both methods using the sample data are included in Section 4.

2.4 Method 1: Effective number of sites

For the purposes of this method, the amount of distributed generation on the system is considered without regard to the geographic location; this is appropriate for a small spatial area of installations,

where the placement of studied PV arrays and the expected placement of new arrays is uniformly random over a well defined geographic area.

The normalized variability metrics of an aggregate system are related to the effective number of solar installations providing power, n_{eff} . n_{eff} accounts for the varying size of systems, e.g., if a 1kW system and a 1MW system are placed in the same network, overall variability will be driven by the larger system as its energy dwarfs that of the smaller one. The effective number of sites is calculated by the inverse Herfindahl-Hirschman Index, HHI^{-1} , shown in Equation 3 where S_i is the share of total network capacity provided by site i , C_i is the capacity installation i , and N is the set of installations included in the network. n_{eff} will equal the actual number of installations, n , if each installation has equal capacity, and n_{eff} will approach 1 if the capacity of one system is much greater than all others. C_i is calculated using the maximum of an empirically derived clear sky index, however the methodology will be updated to use the rated capacity of each installation as recorded in the installation data. Variability metrics of the aggregate system are normalized to the total capacity in the network, $\sum_{i \in N} C_i$.

Networks of solar installations are sampled by using unique combinations of the 15 systems in the sample data. For each possible number of installations (1,2,...15) samples of up to 1000 unique combinations of sites were generated. Each metric is calculated on each unique combination of installations.

$$S_i = \frac{C_i}{\sum_{i \in N} C_i}$$

$$n_{eff} = HHI^{-1} = \frac{1}{\sum_{i \in N} \frac{1}{S_i}}, \quad (3)$$

2.5 Method 2. Prediction of aggregate signal distribution

Each measure of variability is represented as a signal from individual installations, the ramping signal, the AGC signal, and the load-following signal, as shown in Figure 3. The distribution of values in each signal represents the overall impact of variability on the system planning and operation of the grid. To obtain the variability measure of a aggregate system of distributed PV, signals from multiple installations can be summed directly. The variance of this aggregate signal can be calculated from characteristics in the individual signals as shown in Equation 4, where ρ_{ij} is the correlation coefficient between sites i and j , σ_I is the standard deviation of the signal from site I , and N is the set of all sites in the portfolio of PV sites.

$$\sigma_{tot}^2 = \sum_{i \in N} \sum_{j \in N} \rho_{ij} \sigma_i \sigma_j \quad (4)$$

Following two assumptions, the variance of any set of systems is derived from a general standard deviation for any individual site, σ , the distance between pairs of hypothetical sites, d_{ij} , and the capacity of each hypothetical site C_i . These two assumptions are: Equation 5, any installation (within a given area, season, and possibly weather regime) exhibits a standard deviation in signal directly related to the capacity of the installation; and Equation 6, there exists a predictive relationship based on the correlation between signals from a pair of sites and the distance between those sites. Evidence for the first assumption is shown through the distribution of standard deviations from signals in the sample data; evidence for the second assumption is shown through historic relationships of correlation between generation signals and distance between PV arrays, referenced in the background section. Equation 7 shows the resulting relationship used to find the standard deviation of aggregated variability from a distributed PV fleet.

$$\sigma_i = C_i \sigma \quad \forall i \quad (5)$$

$$\rho_{ij} = f(d_{ij}) \quad (6)$$

$$\sigma_{tot}^2 = \sum_{i \in N} \sum_{j \in N} f(d_{ij}) C_i C_j \sigma^2$$

$$\sigma_{tot} = \sigma \sqrt{\sum_{i \in N} \sum_{j \in N} f(d_{ij}) C_i C_j} \quad (7)$$

The overall ability of variance estimation is tested by using a mean normalized standard deviation for the general standard deviation (σ in Equation 5), and exponential fit relationships between distance and correlation.

The shape of the distribution in the aggregate signal is needed, along with the mean and variance, in order to properly characterize the impact of the total fleet. The shape of a distribution is often described by its kurtosis (or "peakiness") where high kurtosis relates to a sharp peak in the distribution and a low kurtosis has a dull peak or no peak. Distributions of impacts from individual sites typically have a high peak at 0, resulting from non-cloudy times. When signals are summed, these times do not always overlap and the kurtosis of the resulting distribution is decreased. Kurtosis is commonly defined as the fourth standardized moment, shown in Equation 8, however the effect is also estimated by the number of standard deviations to a percentile, shown for the 95th percentile in Equation 9. Changes in the shape of each signal's distribution are plotted as the number of installations aggregated into the total signal increases.

$$\text{kurtosis} = \beta_2 = E \left[\left(\frac{X - \mu}{\sigma} \right)^4 \right] \quad (8)$$

$$\kappa_{95} = \frac{P_{95} - \mu}{\sigma} \quad (9)$$

3 Results and Analysis

3.1 Ancillary Service Capacity Needs

Figure 5 displays the normalized AGC capacity requirements versus the effective number of installations providing power. Part A of the figure displays the capacity needs for HE 12, where at midday variability is driven by clouds which are likely to be less correlated between sites. Part B of the figure shows hour ending (HE) 16, where systematic ramping down of solar generation may cause signals to be more highly correlated. For AGC, the relationship between morning and midday hours is not pronounced. There is a distinct decreasing relationship between the effective number of sites and the normalized metric; however it seems the most of the effect is seen through a change in the variance, particularly in HE 16 where the minimum seems to remain constant regardless of n_{eff} .

Figure 6 displays the normalized load-following capacity needs for hours ending 12 and 16. The systematic diurnal ramp has significantly greater impact on load following capacity needs than on AGC

capacity needs. The decrease in requirement with n_{eff} is less pronounced however the overall difference between morning and midday hours is significant.

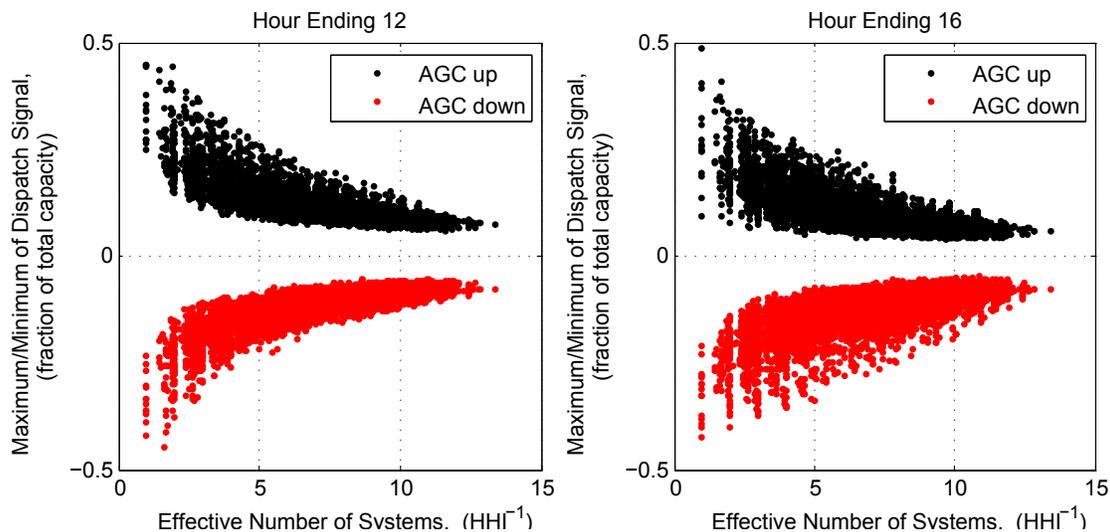


Figure 5: Normalized AGC capacity needs for following an aggregation of solar generation signals during hour ending 12 (A) and hour ending 16 (B).

It is likely that the large variance seen at low numbers of sites is also a result of the geographic autocorrelation effect. Where n_{eff} is low and the relative impact is very high, a few highly correlated sites are being added together; Where n_{eff} is low and the relative impact is low, a few poorly correlated sites are likely being added together. Thus if n_{eff} were able to also account for the distance between sites, which is related to the correlation of output, it is expected that the overall variance will decrease at low n_{eff} and that the decreasing trend will become more pronounced.

The asymptotic value approached by the normalized metric is also affected by geographic autocorrelation of variability and the small spatial scale from which sites are sampled. All arrays sampled for this study are contained within a 5km radius area. Thus, as arrays are added to an already densely populated area (e.g., for large n_{eff}), the average distance between arrays will remain small (only a few kilometers). Thus, for geographically autocorrelated variability, even as system diversity increases, average correlation will plateau to a value based on the geographic area encompassed by the network. For an increase in network area, it is expected that the asymptotically approached normalized metric will decrease.

The flat slope in Figure 6-B for load following requirements versus n_{eff} is likely related to systematic diurnal effects. As seen in Figure 3, most of the diurnal ramps are met by load following capacity. Because these diurnal effects are systematic, they are expected to be highly correlated between sites within the same region. Increasing system diversity is expected to have a smaller effect for highly correlated variability than for less correlated variability from cloud transients.

Figure 7 shows changes in both AGC and load following needs as n_{eff} increases from an n_{eff} range of 1 to 2 to a range of 11 to 12 installations. The geographic diversity effect is more pronounced for AGC (which encompasses shorter timescale variability).

The relationship between capacity needs and n_{eff} is stronger for AGC capacity than for load following capacity, regardless of hour of day. This effect is also attributable to correlated variability. Larger clouds are more likely to pass over multiple sites at once and take a longer time to pass over each site. Thus,

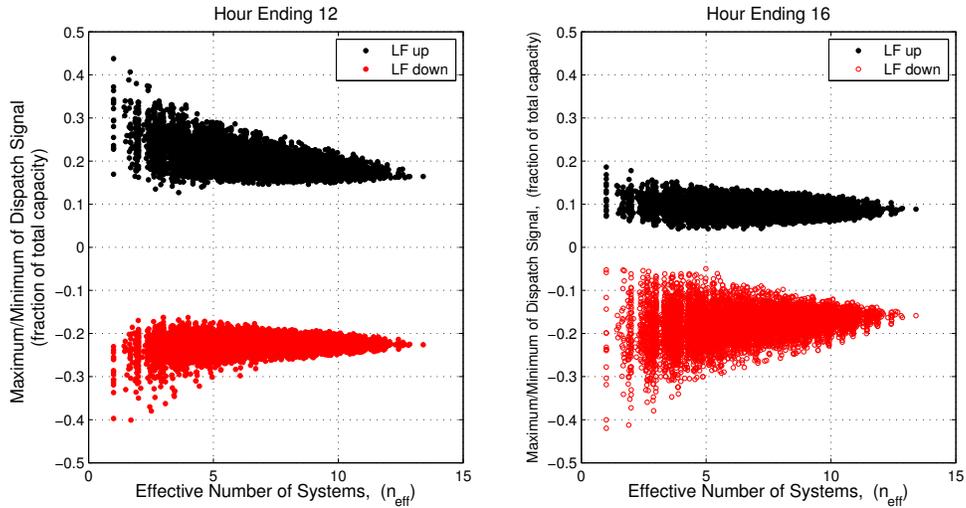


Figure 6: Normalized load-following capacity needs for following an aggregation of solar generation signals during hour ending 11 (A) and hour ending 16 (B).

for cloud transient induced variability, longer time-scale variability is expected to be more geographically autocorrelated than shorter timescale variability. This effect corroborated by Mills and Wisner 2010 as shown in Figure 2.

3.2 Ancillary service ramping requirements

Figure 8 show the maximum normalized ramp in the AGC signal as related to the effective number of sites. Sustained ramps reduce rapidly, even among sites within a small geographic radius. In fact, geographic diversity decreases normalized ramp so quickly that adding an additional site may have a negative effect on the overall likely ramp, not just the normalized likely ramp. These results are promising for reducing fears of increasing costs of PV interconnection related to voltage control on distribution systems; specifically, it appears that none of these ramps are likely to approach the dramatic ramps used by Teleke 2011 [17]. However, it will still be interesting to take a closer look at this effect at higher levels of penetration and smaller spatial scales to isolate potential ramps is isolated portions of feeders.

3.3 Geographic autocorrelation of impact signals

A relationship between correlation and distance for pairs of sites will be used to predict the distribution of impacts of additional PV installations using the method described in Section 3.5. The geographic autocorrelations of ancillary service signals were tested and the 1-min ramping signals were tested and are shown in Figures 9 and 10. From Figure 9, it appears that the geographic autocorrelation falls to 0 for the AGC signal at distances of less than 2 kilometers. For load following the signals appear to plateau at 0.4, this is likely due to systematic effects resulting from the morning and evening ramps. Hourly averages are highly correlated, as expected. Figure 10 shows the geographic autocorrelation of individual ramps in the sample data. This correlation drops to zero at distances less than 1km; too quickly to be captured by the sample data.

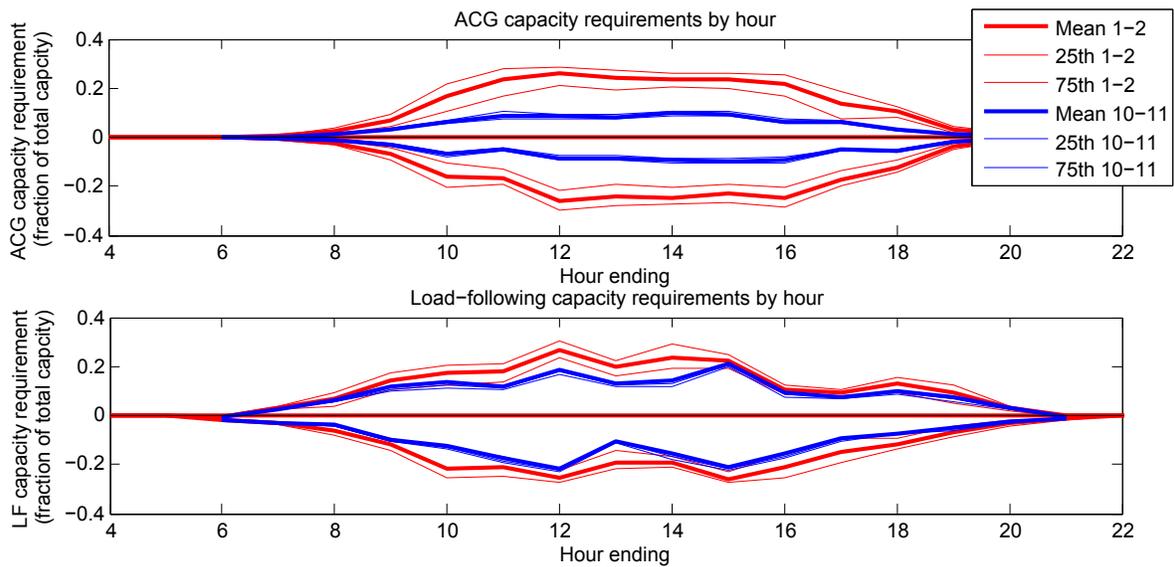


Figure 7: Change in normalized AC needs with increasing effective number of sites. The red lines represent the mean, 25th and 75th percentiles of normalized impact for systems with n_{eff} between 1 and 2, the blue lines represent systems with n_{eff} between 10 and 11

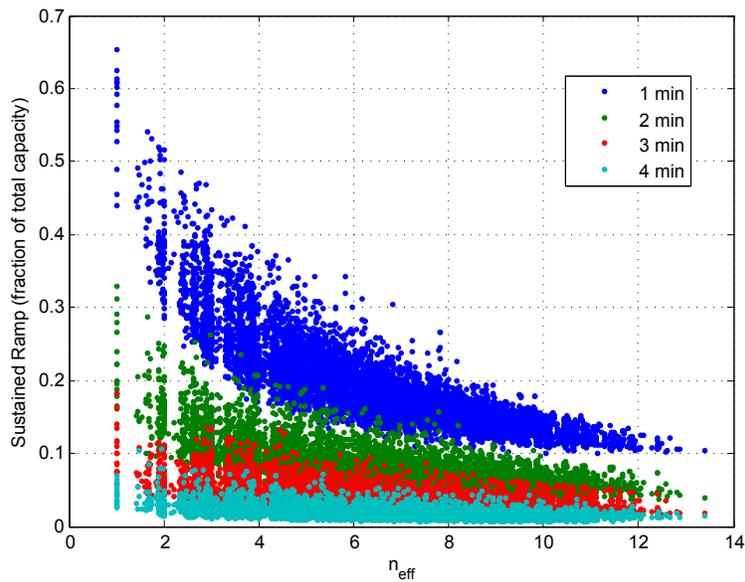


Figure 8: Maximum sustained ramps of duration 1, 2, 3, and 4 minutes versus the effective number of arrays.

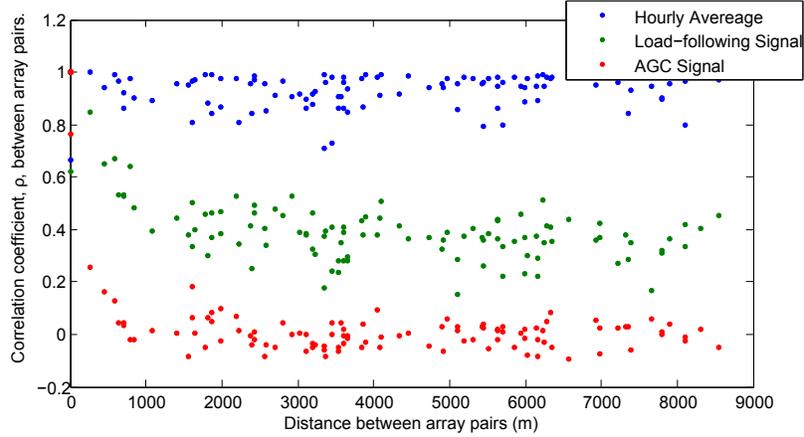


Figure 9: Geographic autocorrelation of AS requirements for each pair of sites in the sample data

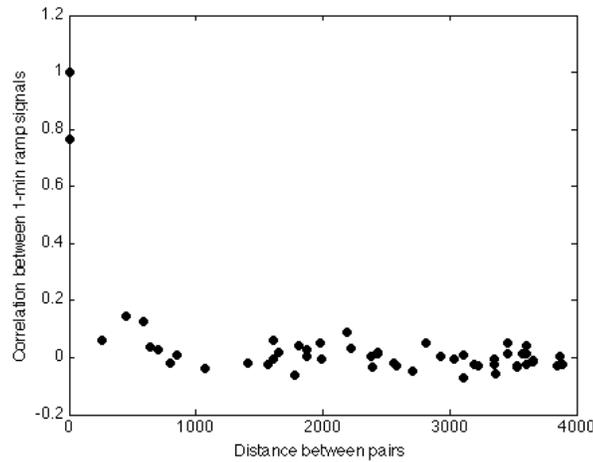


Figure 10: Geographic autocorrelation of the signals for 1-min ramps for each pair of sites in the sample data

3.4 Distribution shape and variance assumptions

Another major assumption for the method described in Section 3.5 is that the shape of the distribution for all sites in an area will be similar, and that the variance will be directly related to the capacity of a site. Figure 11 shows "Q-Q plots" that plot the quantile values of a distribution at one sample site against those of another sample site. A linear relationship in a q-q plot is evidence that both distributions have the same shape. For the purposes of clarity, q-q plots of all site pairs are plotted on the same axis. The left, center, and right axes represent distributions in the ramping, AGC, and load-following signals respectively. All data is normalized by the standard deviation and differenced to the mean; i.e., values on the x and y axes represent standard deviations away from the mean. Vertical lines represent the average values of percentiles among all sample sites (5^{th} , 25^{th} , 50^{th} , 75^{th} , and 95^{th}).

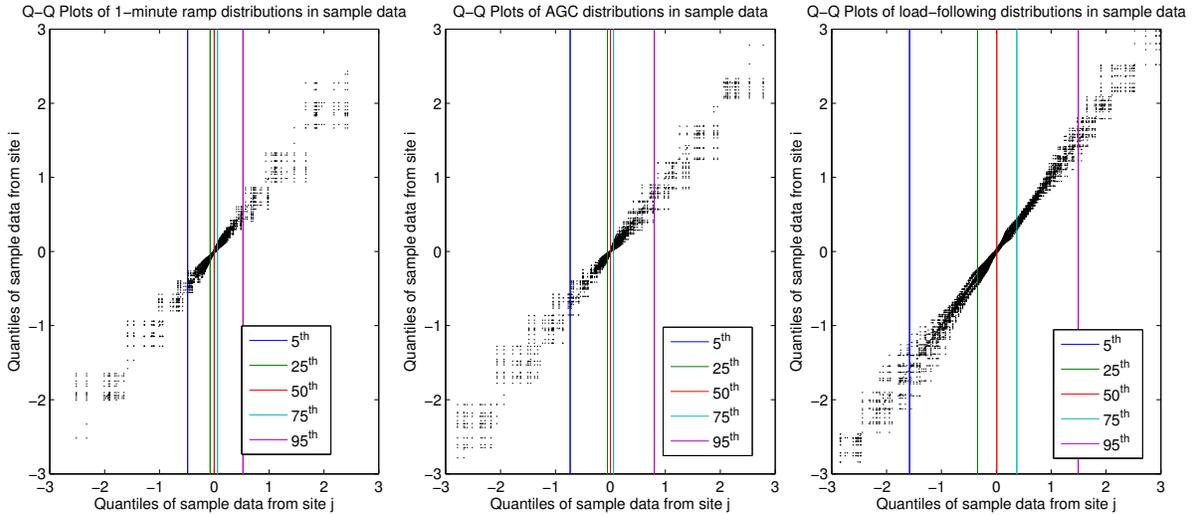


Figure 11: q-q plots for the ramping, AGC, and load-following signals plotted in the left, center, and right panels respectively. Comparisons for all pairings of sites are plotted on the same axis for clarity. Vertical lines represent average percentile values among all sites.

Figure 11 shows that for all pairs of signals in the sample data, the q-q lines follow a linear relationship at least out to the 95th percentile (where most data exists). It also shows the high kurtosis of each of the individual signals, the 95th percentile is less than 1 standard deviation away from the mean for both the AGC signal and the ramping signal.

Figure 12 shows box plots of the normalized mean and standard deviation for the ramping, AGC, and load-following signals. For the preliminary analysis, the mean of the standard deviation is used in the predictive estimates. However, sensitivity analysis should be completed to address the spread of the distribution shown in Figure 12.

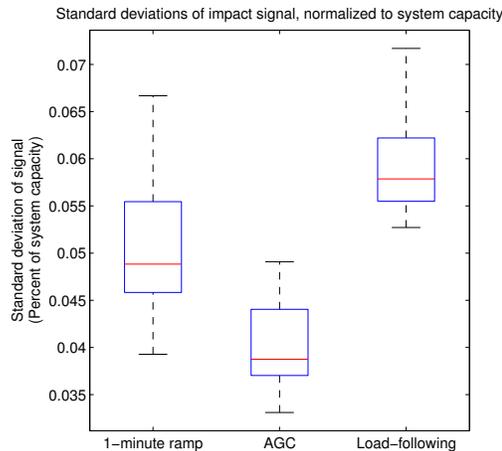


Figure 12: Boxplots for standard deviations of the individual ramping, AGC, and load-following signals

The variance prediction method, shown in Equation 7 is tested using sample data, results are shown in

Figure 13. The mean standard deviation from the sample sites is used as the general standard deviation for each signal. AGC and ramping signals are assumed to be completely uncorrelated, load-following correlation is assumed to follow the empirical relationship described in Equation 10 which is derived from the relationship shown in Figure 9.

$$\rho_{ij} = 0.6 \exp\left(\frac{d_{ij}}{1000m}\right) + 0.4 \quad (10)$$

Figure 13 shows that the variance estimation methods works well for a small number of sites, with 10-20% error; and that the estimation accuracy increases with the number of sites aggregated, leading to errors of less than 10% in 13 aggregated sites. It should be noted that these results are not predictive; the standard deviations and correlations used to find σ , and $f(d_{ij})$ result from the same data for which the estimate is derived.

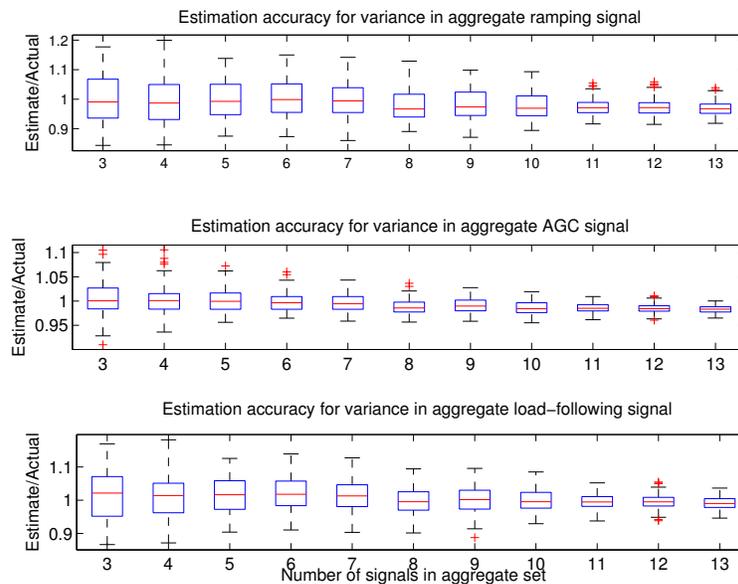


Figure 13: Boxplots representing the accuracy of the variance prediction method (Eq. 7) using the mean standard deviations from the sample data.

The shape of the distribution of an aggregated system of PV installations is needed, in conjunction with variance, to find the overall impact. Figure 14 plots the number of standard deviations to different percentiles of the aggregate distribution against the number of sites that are aggregated. The flattening peak is shown through the increasing number of standard deviations out to the 95th percentile (shown in red), though the resulting distribution still has a significantly high kurtosis. The distribution tails flattening is shown through the 99th percentile decreasing in number of standard deviations away from the mean. This same effect is exhibited in the maximum and minimum, though these values are much more variable and much further from the mean.

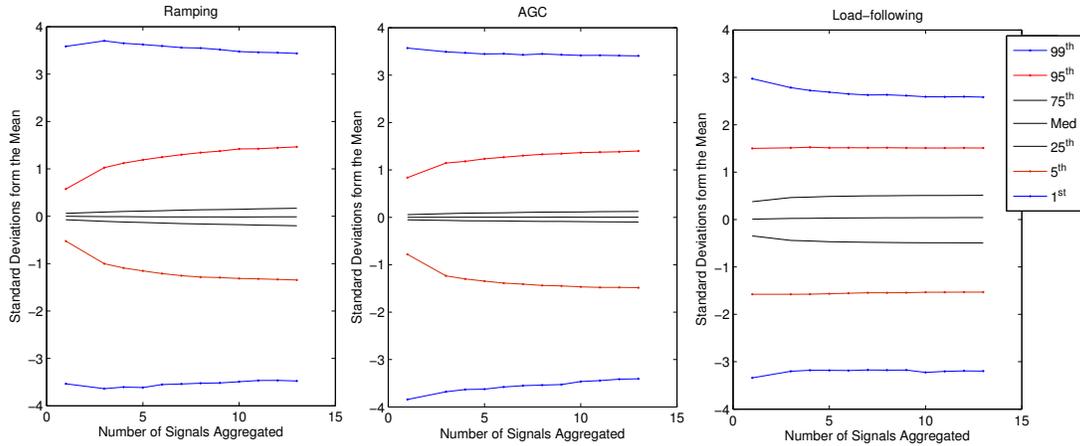


Figure 14: Shape of aggregate signal distribution versus the number of sites aggregated. The distance from the mean of 7 percentile values are plotted, normalized to the standard deviation of the entire signal. Legend applies to each panel

4 Discussion and Future Work

For method 1, the normalized impact metric appears to asymptotically approach a single value as n_{eff} increases; however neither the asymptotic metric value, nor the n_{eff} are likely generalizable to networks over larger areas. As the size of the network increases, so will the distribution of distances between each new site to every existing site. Specifically, the average distance between array will increase with the radius of the network. Thus the asymptotically approached value of any variability metric is expected to decrease as the geographic size of the network increases. This effect will be better characterized with additional data over larger geographic areas and longer time-periods.

Method 2 outlines a more general approach that can be applied to areas of any size, and preliminary results from sample data are promising that this approach will yield significant results. However, the preliminary estimates are not predictive, and their accuracy may be dependent on consistent weather patterns. Specifically, the general standard deviations were characterized in the same time-period (and thus the same weather regime) as the signals they were used to predict. Future work will stratify the analysis based on weather regimes, allowing for scenario or stochastic analysis.

Another limitation of Method 2 is that it is not as accurate for maximum or minimum values. The smaller the acceptable confidence interval, the more accurate that the prediction is likely to be. Thus even a 99% confidence interval is likely to be a much more accurate prediction that a likely maximum value in an aggregate signal. This is more significant for a small aggregation of systems, where distribution tails are wider. Future work will include a comparison of the two methods' ability to account for maximum and minimum fluctuations.

Finally, Method 2 currently does not address systematic effects dependent on the time of day. It is likely that cloud transient impacts will be smaller in the mornings and evenings, however diurnal effects have large impacts in these hours. In future work, the analysis will be performed separately for each hour of the day do control for these effects.

4.1 Future deliverables for this task

1. Stratified statistical analysis to account for hour of day, season, and weather regime.
2. Comparison of results from each empirical prediction method to account for the geographic area of the portfolio and the maximum and minimum of impact signals.
3. Scenario analysis for impacts over a feeder footprint (low, medium, and high penetration)
4. Scenario analysis for impacts over a Sub-Load Aggregation Point.
5. Comparison of small distributed generation (<500 kW) to centralized/community generation (>500 kW).

References

- [1] A. Mills and R. Wiser. Implications of Wide-Area geographic diversity for Short-Term variability of solar power. Technical report, Lawrence Berkeley National Laboratory, 2010.
- [2] R. Piwko, K. Clark, L. Freeman, G. Jordan, and N. Miller. Western wind and solar integration study: Executive Summary,(WWSIS) may 2010. Technical report, National Renewable Energy Laboratory (NREL), Golden, CO., 2010.
- [3] Xinggang Bai, Kara Clark, Gary A. Jordan, Nicholas W. Miller, and Richard J. Piwko. Intermittency analysis project: Appendix b impact of intermittent generation on operation of california power grid. Technical report, California Energy Commission, 2007.
- [4] Y.V. Makarov, P.V. Etingov, Z. Huang, J. Ma, and K. Subbarao. Incorporating wind generation forecast uncertainty into power system operation, dispatch, and unit commitment procedures. Technical report, Pacific Northwest National Laboratory (PNNL), Richland, WA (US), 2010.
- [5] Y. V Makarov, C. Loutan, Jian Ma, and P. de Mello. Operational impacts of wind generation on california power systems. *IEEE Transactions on Power Systems*, 24(2):1039–1050, May 2009.
- [6] Mark Rothleder. Track i direct testimony of mark rothleder on behalf of the california independent system operator corporation.
- [7] Brendan Kirby and Eric Hirst. Customer-Specific metrics for the regulation and Load-Following ancillary services. Technical Report ORNL/CON-474, January 2000.
- [8] Thomas E. Hoff and Richard Perez. Quantifying PV power output variability. *Solar Energy*, 84(10):1782–1793, October 2010.
- [9] E. Wiemken, H. G. Beyer, W. Heydenreich, and K. Kiefer. Power characteristics of PV ensembles: experiences from the combined power production of 100 grid connected PV systems distributed over the area of germany. *Solar Energy*, 70(6):513–518, 2001.
- [10] Akinobu Murata, Hiroshi Yamaguchi, and Kenji Otani. A method of estimating the output fluctuation of many photovoltaic power generation systems dispersed in a wide area. *Electrical Engineering in Japan*, 166(4):9–19, March 2009.

- [11] B.H. Chowdhury and S. Rahman. Is central station photovoltaic power dispatchable? *Energy Conversion, IEEE Transactions on*, 3(4):747–754, 1988.
- [12] Norihiro Kawasaki, Takashi Oozeki, Kenji Otani, and Kosuke Kurokawa. An evaluation method of the fluctuation characteristics of photovoltaic systems by using frequency analysis. *Solar Energy Materials and Solar Cells*, 90(18-19):3356–3363, November 2006.
- [13] Toshiya Nanahara, Masahiro Asari, Takamitsu Sato, Koji Yamaguchi, Masaaki Shibata, and Tsutomu Maejima. Smoothing effects of distributed wind turbines. part 1. coherence and smoothing effects at a wind farm. *Wind Energy*, 7(2):61–74, April 2004.
- [14] Jay Apt. The spectrum of power from wind turbines. *Journal of Power Sources*, 169(2):369–374, June 2007.
- [15] B. Kirby. Impact of balancing areas size, obligation sharing, and ramping capability on wind integration. Los Angeles, California, June 2007.
- [16] E. Liu and J. Bebic. Distribution system voltage performance analysis for high-penetration photovoltaics. Subcontract Report NREL/SR-581-42298, National Renewable Energy Laboratory, 2008.
- [17] Sercan Teleke, Farbod Jahanbakhsh, Farid Katiraei, and Julio Romero Aguero. Analysis of interconnection of photovoltaic distributed generation. In *2011 IEEE Industry Applications Society Annual Meeting (IAS)*, pages 1–6. IEEE, October 2011.