NATIONAL RENEWABLE ENERGY LABORATORY

A Statistical Characterization of Solar Photovoltaic Power Variability at Small Timescales

Preprint

S. Shedd, B.-M. Hodge, A. Florita, and K. Orwig

To be presented at the 2nd Annual International Workshop on Integration of Solar Power into Power Systems Conference
Lisbon, Portugal
November 12–13, 2012

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-5500-56165
August 2012

Contract No. DE-AC36-08GO28308

Available electronically at http://www.osti.gov/bridge

Available for a processing fee to U.S. Department of Energy
and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

Cover Photos: (left to right) PIX 16416, PIX 17423, PIX 16560, PIX 17613, PIX 17436, PIX 17721

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

A Statistical Characterization of Solar Photovoltaic Power Variability at Small Timescales

Sandra Shedd, Bri-Mathias Hodge, Anthony Florita, and Kirsten Orwig
National Renewable Energy Laboratory
Golden, CO 80401; USA

Abstract—**Integrating large amounts of variable and uncertain solar photovoltaic (PV) power into the electricity grid is a growing concern for power system operators in a number of different regions. Power system operators typically accommodate variability—whether from load, wind, or solar—by carrying reserves that can quickly change their output to match the changes in the solar resource. At timescales in the seconds-to-minutes range, this is known as regulation reserve. Previous studies have shown that increasing the geographic diversity of solar resources can reduce the short-term variability of the power output. As the price of solar has decreased, the emergence of very large PV plants (greater than 10 MW) has become more common. These plants present an interesting case because they are large enough to exhibit some spatial smoothing by themselves. In this work, we examined the variability of solar PV output among different arrays in a large (~50 MW) PV plant in the western United States. We examined the correlation in power output changes between different arrays as well as the aggregated plant output at timescales ranging from 1 sec to 5 min.**

***Keywords*—photovoltaic power generation, stochastic processes, power system operation**

I. INTRODUCTION

Worldwide interest in higher penetration of solar photovoltaics (PV) into power systems is increasing. A consequence of increased levels of solar penetration is increased variability and uncertainty in power generation within the system. An important consideration in the integration of increased amounts of solar PV generation is the characterization of how the power output changes at multiple timescales and in different atmospheric conditions. Power system operators accommodate variability because of load, wind, or solar through systems of reserves that adjust power output levels of dispatchable plants to meet changes in demand or changes in variable generation. At short timescales (ranging from seconds to minutes in duration), this is known as regulation reserve, and it is used to maintain system frequency during short-term fluctuations in power generation. A better understanding of solar PV generation variability assists power system operators in making decisions regarding optimum levels of system reserves.

One important factor in solar PV power integration studies is the impact of varying geographic dispersion of PV generating units on overall PV output variability. Studies have shown that increased geographic diversity in the solar resource reduces the variability of power output at short timescales [1–2]. This is because solar irradiance is not highly correlated between even close locations at very short timescales. Mills and Wiser [1] found in their analysis of 23 time-synchronized solar PV plant sites in the southern United States that even for five geographically close plants the variability of aggregated power output was greatly reduced relative to that of an individual plant at sub-hourly timescales. Thus, the geographic distribution of the solar PV resource has an impact on the regulation reserves necessary. This is critical because a key impediment to large-scale integration of solar resources is the high cost that is theorized to be a consequence of the necessity of increased reserve levels to accommodate increased variability. Some PV integration studies that analyze (simulated) large-scale solar resources have worked under the assumption that increased grid penetration of solar PV generation leads to increased operating costs because of higher levels of expensive fast-acting reserves [3–8]. This paper addresses this assumption via statistical analyses of short-term variability of a large-scale solar PV plant.

In this work, we examined the variability of solar PV output among different arrays in a single large (~50 MW) PV plant in the western United States. We examined the correlation in power output ramps between arrays as well as aggregated plant output at varying timescales within the regulation timeframe (from 1 sec to 5 min). These analyses were undertaken with the aim of assessing the ramp and output smoothing that can be attributed to geographic diversity within a single very large solar PV plant.

II. METHODS AND DATA

This section outlines some of the important methods involved in the study. Section II-A contains information regarding the data sets analyzed in this paper. Section II-B provides some relevant background information regarding statistical methodology that may aid the reader in understanding the results that follow.

A. Data Utilized

In this paper, we examined solar irradiance and solar PV array output data from a PV plant located in the southwestern United States. The data sets consisted of four months (September to December 2011) of PV power output at the 1-sec resolution level. In addition, data describing local weather and solar irradiance conditions for the solar PV plant and surrounding area were included.

The authors are with the National Renewable Energy Laboratory, Golden, CO 80401; USA (email: bri-mathias.hodge@nrel.gov, sandra.shedd@nrel.gov, anthony.florita@nrel.gov, kirsten.orwig@nrel.gov)

B. Statistical Methodology

Histograms and probability density functions are used to illustrate both the range of values that a given random variable can take and the likelihood of a sampled random variable falling in a specific range [9]. In this paper, we used them to characterize the ramps in PV power output at varying timescales.

Often when analyzing data we want to know if two variables are correlated. Correlation measures the linear dependence between two variables, and correlation values fall within the range of -1 to 1. A value of 1 indicates that one variable is a positive linear function of the other, -1 means one variable is a negative linear function of the other, and 0 indicates a lack of correlation entirely. In this paper, we used the Spearman [10] correlation statistic, ρ, defined in (1):

$$\rho = \frac{n(\sum_{i=1}^{n} x_i y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{n(\sum_{i=1}^{n} x_i^2) - (\sum_{i=1}^{n} x_i)^2}\sqrt{n(\sum_{i=1}^{n} y_i^2) - (\sum_{i=1}^{n} y_i)^2}} \qquad (1)$$

where x and y are the random variables in question (as vectors) and n is the length of the random variables [9–11]. In this paper, correlations between power output and changes in output levels between individual PV arrays were calculated for timescales ranging from 1 sec to 5 min.

In Section III-B, the correlation coefficients calculated using (1) are displayed using heat maps, which is a graphical representation of data wherein the individual values in a matrix are represented as colors [12]. In the heat maps that follow, individual PV arrays were examined for cross-correlation; thus, the heat maps display darker colors for a pair of arrays that were relatively uncorrelated and lighter colors for correlations of greater magnitude.

III. Results

This paper has thus far discussed the importance of solar PV generation variability and introduced relevant statistical information. We next characterized the data statistically. In Section III-A, we assessed the variability of solar PV output in terms of observable ramps in individual PV arrays, in aggregate conglomerations of separate arrays, and throughout the entire plant. Section III-B incorporates the additional variable of timescales: correlation between ramps were considered at varying timescales ranging from 1 sec to 5 min.

A. Correlation of Power Output Fluctuations

Figures 1–7 show the distributions of changes in instantaneous solar PV power output aggregated throughout the entire plant at timescales ranging from 1 sec to 5 min.

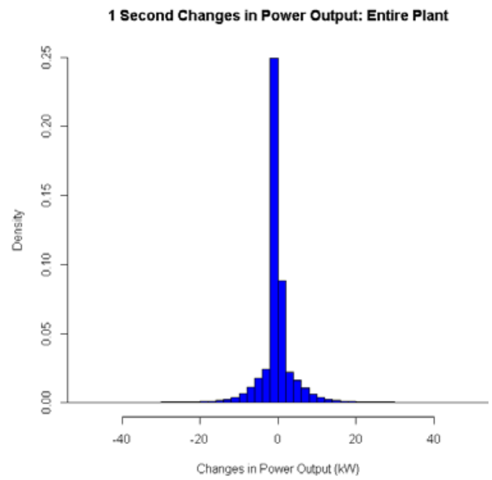

Figure 1. Distribution of changes in aggregated solar power output for the entire plant at 1-sec timescale

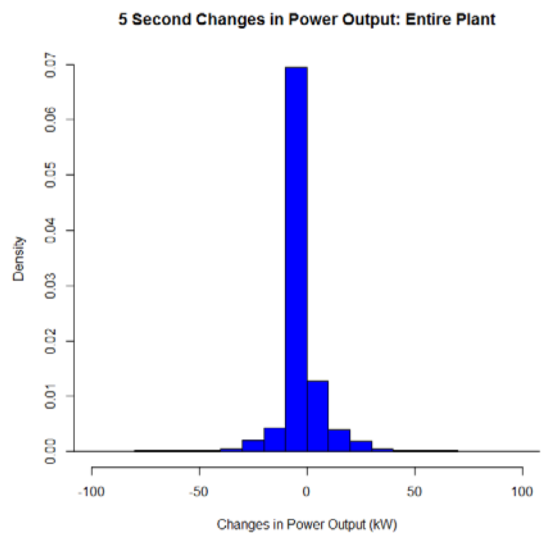

Figure 2. Distribution of changes in aggregated solar power output for the entire plant at 5-sec timescale

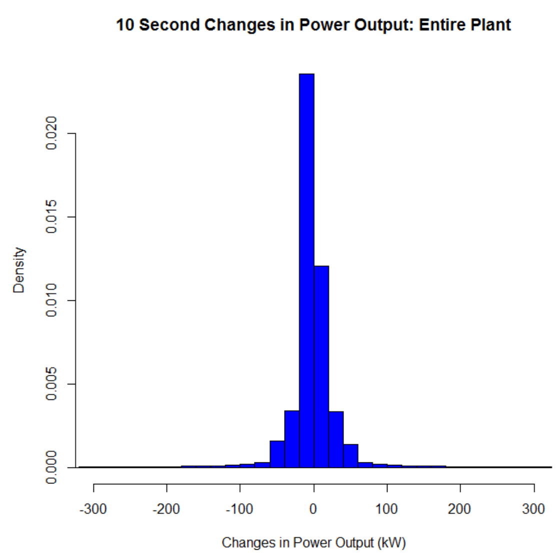

Figure 3. Distribution of changes in aggregated solar power output for the entire plant at 10-sec timescale

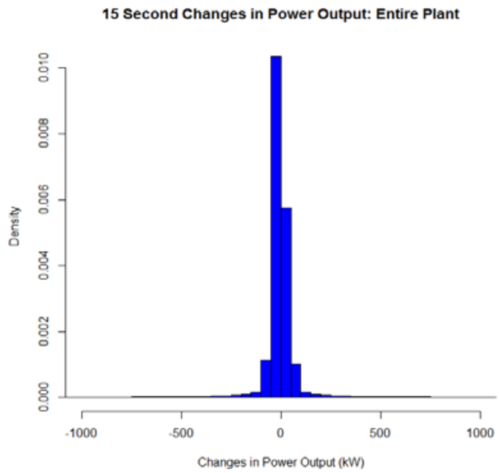

Figure 4. Distribution of changes in aggregated solar power output for the entire plant at 15-sec timescale

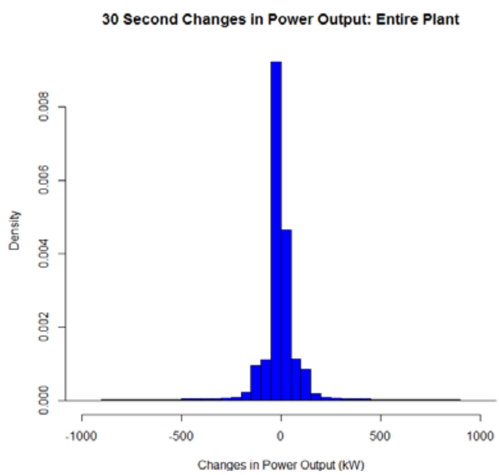

Figure 5. Distribution of changes in aggregated solar power output for the entire plant at 30-sec timescale

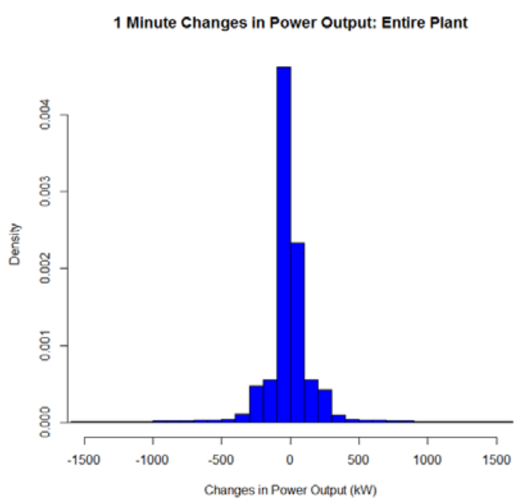

Figure 6. Distribution of changes in aggregated solar power output for the entire plant at 1-min timescale

Note that the scope of the histograms was limited to clarify visual appearance: the 1-sec, 5-sec, 10-sec, 15-sec, 30-sec, 1-min, and 5-min data sets contained ~182000, ~46800, ~29800, ~8100, ~8450, ~4900, ~1350 data points of greater magnitude than the displayed x-axes, respectively. At all timescales this represented very small percentages of total data.

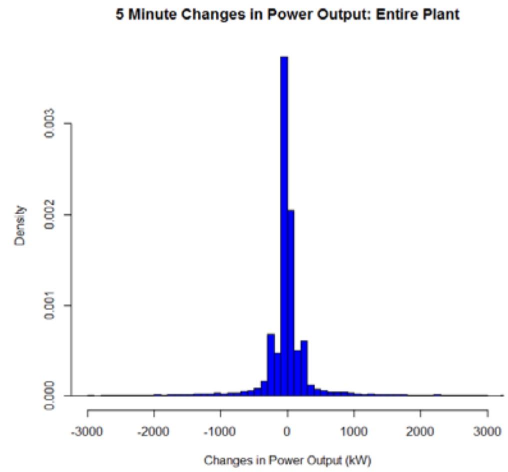

Figure 7. Distribution of changes in aggregated solar power output for the entire plant at 5-min timescale

As shown in Figures 1–7, in all cases the distributions were centered near 0, with minimal skewness. All seven distributions exhibited fat tails that indicated significant instances of relatively high-magnitude ramps. These were the changes in power output that are most problematic for power systems operations. We address these ramps further in Section III-B.

Table I summarizes an analysis of distributions of measured ramps in power output for various geographic spreads of PV arrays, including a single inverter, the entire plant, six inverters clustered within a short geographical distance, and six inverters scattered throughout the plant.

TABLE I. DISTRIBUTION STATISTICS FOR VARYING SPATIAL ARRANGEMENTS OF SOLAR PV ARRAYS[A]

Timescale	**Single Inverter**		**Entire Plant**	
	Standard Deviation	**Kurtosis[B]**	**Standard Deviation**	**Kurtosis**
1 sec	0.9	29294.3	29.0	1047485
5 sec	3.4	1280.2	95.4	64726.2
10 sec	5.7	476.6	199.7	7960.3
15 sec	7.5	335.7	282.8	2851.3
30 sec	10.8	221.0	506.5	814.0
1 min	14.2	160.4	838.3	269.0
5 min	21.4	25.2	1731.6	29.2
Timescale	**Clustered Inverters**		**Scattered Inverters**	
	Standard Deviation	**Kurtosis**	**Standard Deviation**	**Kurtosis**
1 sec	3.4	91511.2	2.6	244050
5 sec	14.1	2098.9	9.7	9235.9
10 sec	25.6	591.8	16.4	2891.8
15 sec	35.4	338.5	22.7	1225.2
30 sec	56.2	213.6	36.6	527.6
1 min	79.3	145.2	55.8	237.3
5 min	119.3	25.4	111.5	28.5

A. Instantaneous power output was calculated at each timescale and used to compute changes in power output at given intervals

B. Kurtosis: statistical value measuring the relative magnitude of the peak of a distribution

Two observable trends in Table I were both the increase in standard deviation and decrease in kurtosis with increasing timescale. The increase in standard deviation indicated the increased spread of the data set around the mean (which was zero in all cases), implying greater variability in ramps at the "longer" timescales within the regulation timeframe.

B. *Ramp Correlations at Varying Timescales*

Figures 8–14 show the correlation of solar PV power output (left) and changes in output (right) between the 96 different inverters in the PV plant. Table II lists the average Spearman's ρ values for power output and changes at each timescale:

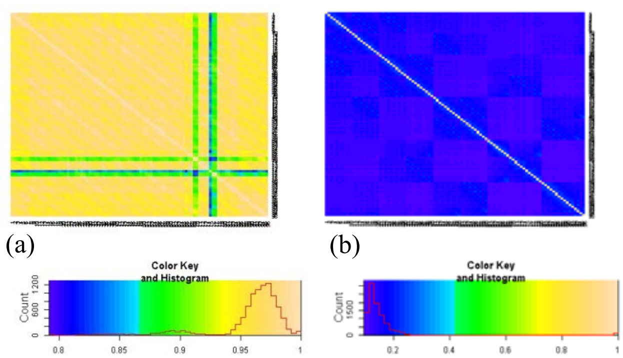

Figure 8. Heat maps of the correlation coefficients of (a) power output and (b) changes in output at 1-sec timescale for 96 inverters in the same PV plant measured during four months in 2011

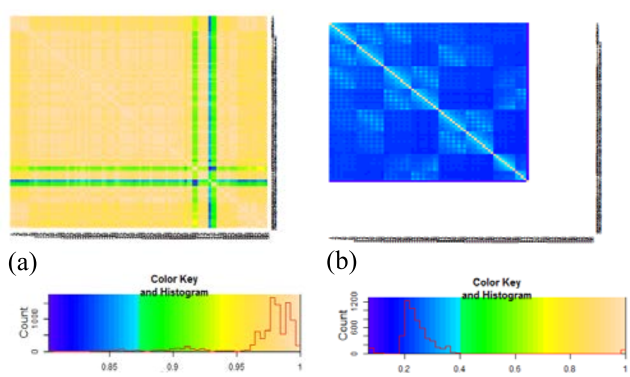

Figure 9. Heat maps of the correlation coefficients of (a) power output and (b) changes in output at 5-sec timescale for 96 inverters in the same PV plant measured during four months in 2011

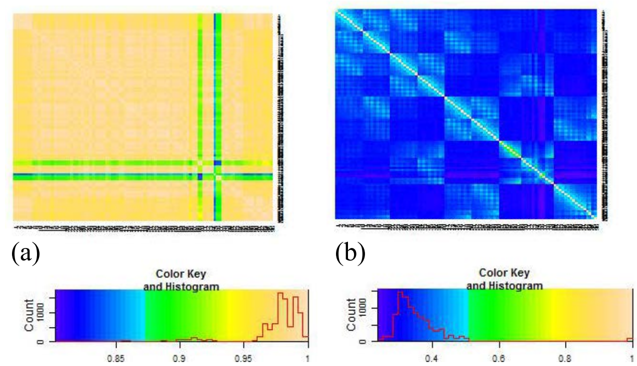

Figure 10. Heat maps of the correlation coefficients of (a) power output and (b) changes in output at 10-sec timescale for 96 inverters in the same PV plant measured during four months in 2011

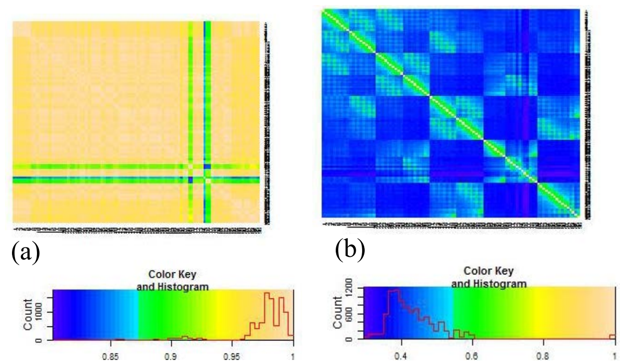

Figure 11. Heat maps of the correlation coefficients of (a) power output and (b) changes in output at 15-sec timescale for 96 inverters in the same PV plant measured during four months in 2011

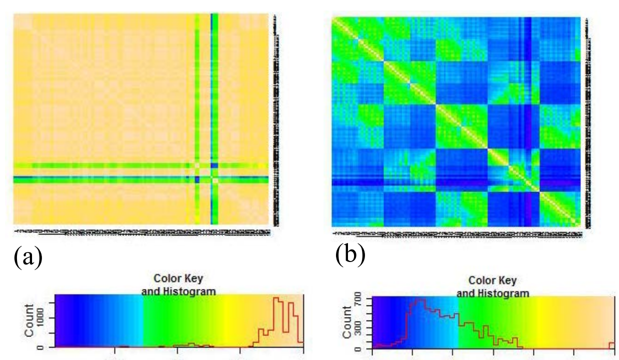

Figure 12. Heat maps of the correlation coefficients of (a) power output and (b) changes in output at 30-sec timescale for 96 inverters in the same PV plant measured during four months in 2011

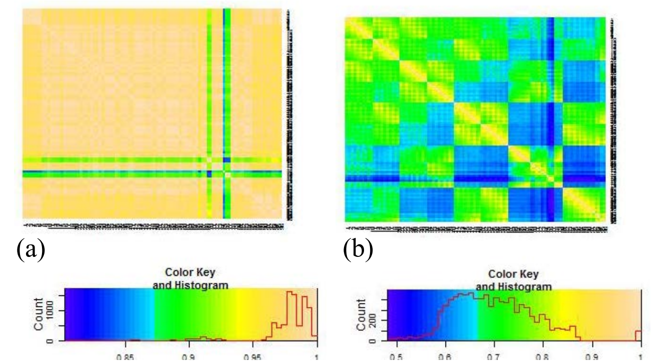

Figure 13. Heat maps of the correlation coefficients of (a) power output and (b) changes in output at 1-min timescale for 96 inverters in the same PV plant measured during four months in 2011

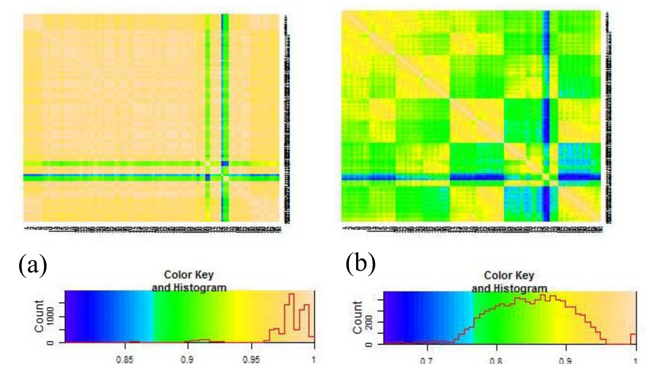

Figure 14. Heat maps of the correlation coefficients of (a) power output and (b) changes in output at 5-min timescale for 96 inverters in the same PV plant measured during four months in 2011

TABLE II. MEAN CORRELATION VALUES FOR POWER OUTPUT AND CHANGES IN POWER OUTPUT[C]

Timescale	Mean Value of Correlation for Power Output	Mean Value of Correlation for Changes in Power Output
1 Sec	0.96	0.14
5 Sec	0.98	0.26
10 Sec	0.98	0.37
15 Sec	0.98	0.45
30 Sec	0.98	0.59
1 Min	0.97	0.67
5 Min	0.97	0.84

C. Instantaneous power output was calculated at each timescale and used to compute changes in power output at given intervals

It is clear from Figures 8–14 and Table II that instantaneous power output between pairs of inverters was well correlated at each regulation reserve timescale considered. The average Spearman's ρ values for power output ranged from 0.96 to 0.98 for regulation timeframes, indicating strong correlation in power output.

Although understanding correlation of solar PV power output is important for integration of solar power into electricity systems, more critical for the determination of necessary regulation reserve levels is the correlation of changes in power output between individual solar PV power generating units. The dark blue colors of the heat maps that illustrate correlation in changes indicated the low correlation of power output fluctuations between individual PV units. Table II shows that the correlation of changes in power output was weak until the 5-min timescale (the longest timescale within the regulation reserve timeframe), and Spearman's ρ values at all shorter timescales were all of too low magnitude to claim strong correlation. These low-correlation coefficients indicated that the factors that caused variability in solar PV output at the regulation timescale were very localized weather patterns, such as a cloud passing over the PV arrays.

A clear trend, however, was the steadily increasing correlation of ramps with increasing timescale. This can be explained by the nature of solar irradiance at very short timescales. In the regulation timeframe (seconds to minutes), the primary source of variability in power output is cloud movement. A cloud passing over a field of PV generating units will cause a ramp in power output for only a single panel or a very small fraction of the PV plant in 1 sec. Therefore, at the 1-sec timescale, this ramp in power will not be experienced throughout the remainder of the plant, and thus correlation of changes in power output at the 1-sec timescale is lower. During 5 min, in contrast, the same cloud may be able to move across the entire plant, causing all panels to experience power output ramps, yielding higher correlation of changes in power output among individual PV arrays.

C. *Intra-Plant Effects of Geographic Distribution*

It is clear, then, that geographic diversity of individual solar PV power generation units has an effect on the variability of power output within a single large-scale plant. This can be understood as an extension of the reduction in variability throughout an interconnected grid of multiple widely dispersed (i.e., several kilometers apart) solar PV plants, as described in previous studies [1–8]. Further, the influence of geographic distribution can be detected even in smaller groups of PV power generating units within the plant. This is evidenced by the distinct "checkerboard" pattern of higher/lower correlation coefficients inside the heat maps describing correlation of changes in power output. The inverters that provided the measured power output data were numbered by location/arrangement, yielding the patterned correlation maps in Figures 8–14. To investigate this further, Figure 15 presents a graph of the standard deviations of distributions of changes in power output for four different geographic arrangements of PV generating units within the large-scale plant:

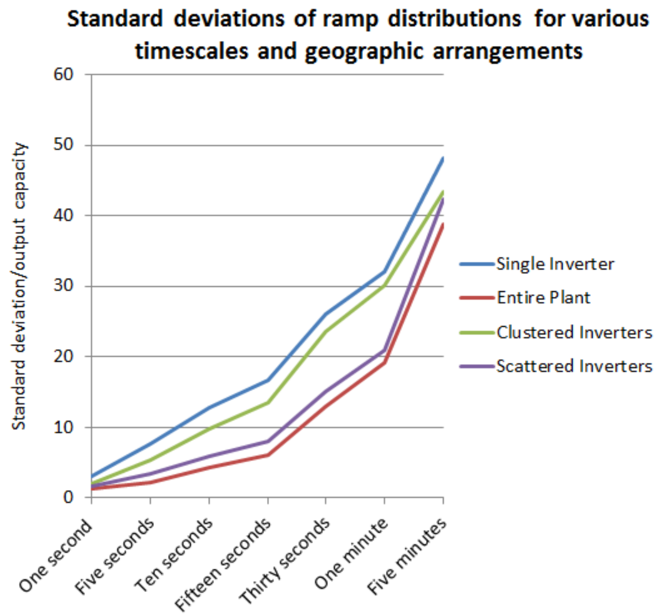

Figure 15. Graph of the standard deviations of distributions of ramps in the regulation timeframe for four different geographic configurations of solar PV generating units in a single large-scale PV plant. The configurations were a single generating unit (blue), the entire plant (red), six units clustered together in a contiguous area (green), and six units scattered throughout the total plant area (purple).

As shown in Figure 15, the reduction in power output fluctuation variability between a single PV generating unit and the entire plant was immediately apparent. Even more interesting was the reduction in variability attributed to scattering even a small production capacity (in this case, six individual PV generating units) throughout the plant area rather than clustering them in a close group. The six units in the "scattered" subset were selected based on unit number to ensure their individual locations were widely dispersed. The units in the "clustered" subset were selected by considering six adjacent unit numbers to ensure they were geographically adjacent. Figure 15 shows that the standard deviations for the distributions of ramps for six scattered inverters (purple line) were consistently lower than those for the distributions of ramps for six clustered inverters throughout the regulation timeframe. Similarly, the kurtosis values reported in Table 1 revealed that the relative magnitudes of the peaks of the clustered-unit histograms were smaller than the peaks of the histograms for the scattered arrays, for both timescales considered.

IV. CONCLUSIONS

In this paper, we examined the variability of solar PV output in the regulation reserve timeframe among various arrays within a large-scale (~50 MW) solar PV plant in the southwestern United States. Although the distributions of changes in aggregate power output throughout all timescales considered were clustered around a strong peak at zero, the distributions at all timescales exhibited significant instances of higher magnitude ramps in the tails of the histograms. To further characterize these ramps, which are the most problematic for systems operators, we considered the correlation of changes in power output in regulation timescales for individual solar PV generating units throughout the plant. Understanding the correlation of changes in output is more critical for determining the regulation reserves necessary to maintain system stability than the correlation of instantaneous power output. Although power output levels for individual generating units were significantly correlated at all timescales within the regulation timeframe, correlation of changes in power output was weak for all timescales shorter than 5 min. The steady increase in correlation with increasing timescale can be explained by the lack of correlation of solar irradiance at very short timescales, and is evidence of smoothing of variability because of geographic distribution of individual generating units within a single PV plant. The lack of strong correlation at short timescales indicates that the impact on the levels of regulation reserve required to accommodate variability could be less than previously theorized. Further study could investigate the influence of intra-plant geographic distribution on reserve levels and the associated costs of dispatch processes and systems operations.

REFERENCES

[1] A. Mills and R. Wiser, Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power, LBNL-38884E, Berkeley, CA: Lawrence Berkeley National Laboratory, 2010.

[2] J. Marcos et al., "Power output fluctuations in large-scale PV plants: One-year observations with one-second resolution and a derived analytic model," Progress in Photovoltaics: Research and Applications, vol. 19, pp. 218–227, 2011.

[3] R. Piwko, K. Clark, L. Freeman, G. Jordan, and N. Miller, Western Wind and Solar Integration Study: Executive Summary, NREL/SR-550-47781, Work performed by GE Energy, Schenectady, NY. Golden, CO: National Renewable Energy Laboratory, 2010, 40 pp.

[4] Navigant, Sandia National Laboratories, et al., Large-Scale PV Integration Study, PNNL-20677, Washington, D.C.: United States Department of Energy, 2011, pp. 1–179.

[5] A. Mills et al., "Dark shadows: Understanding variability and uncertainty of photovoltaics for integration with the electric power system," IEEE Power and Energy Magazine vol. 9 (3), pp. 33–41, May/June 2011.

[6] J. Ma et al., "Impact of wind and solar generation on the California ISO's intra-hour balancing needs," IEEE Power and Energy Society General Meeting Proceedings, Detroit, Michigan, July 24–29, 2011.

[7] T. E. Hoff and R. Perez, "Predicting short-term variability of high-penetration PV," Proceedings of the American Solar Energy Society Annual Conference, Denver, Colorado, May 13–17, 2012, pp. 1–6.

[8] M. Lave and J. Kleissl, "Solar variability of four sites across the state of Colorado," Renewable Energy, vol. 35, pp. 2,867–2,873, 2010.

[9] "R: A language and environment for statistical computing," Vienna, Austria: R Foundation for Statistical Computing, 2010.

[10] C. Spearman, "The proof and measurement of association between two things," American Journal of Psychology, vol. 15 (1), 72–101, 1904.

[11] D. Wuertz and D. Rmetrics core team members, uses code built-in from the following R contributed packages: gmm from Pierre Chauss, gld from Robert King, gss from Chong Gu, nortest from Juergen Gross, HyperbolicDist from David Scott, sandwich from Thomas Lumley, Achim Zeileis and fortran/C code from Kersti Aas. 2012. fBasics: Rmetrics - Markets and Basic Statistics. R package version 2160.81. http://CRAN.R-project.org/package=fBasics.

[12] G. R. Warnes, includes R source code and/or documentation contributed by Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw, Thomas Lumley, Martin Maechler, Arni Magnusson, Steffen Moeller, Marc Schwartz, and Bill Venables, 2012, gplots: Various R programming tools for plotting data. R package version 2.11.0, http://CRAN.R-project.org/package=gplo