Available online at www.sciencedirect.com

SciVerse ScienceDirect

Solar Energy 86 (2012) 2177–2189

SOLAR ENERGY

www.elsevier.com/locate/solener

Review

Modeling PV fleet output variability

Thomas E. Hoff[a], Richard Perez[b,*]

[a] Clean Power Research, 10 Glen Court, Napa, CA 94558, United States
[b] SRC, The University at Albany, 251 Fuller Rd., Albany, NY 12203, United States

Available online 10 December 2011

Communicated by: Associate Editor Frank Vignola

Abstract

This paper introduces a novel approach to estimate the maximum short-term output variability that an arbitrary fleet of PV systems places on any considered power grid. The paper begins with a model that demonstrates that the maximum possible variability for N identical, uncorrelated PV systems equals the total installed capacity divided by $\sqrt{2N}$. The paper then describes a general methodology that is applicable to arbitrary PV fleets. A key input to this generalized approach is the correlation, or absence thereof, existing between individual installations in the fleet at the considered variability time scale. In this respect, the article includes a presentation of new experimental evidence from hourly satellite-derived irradiances relating distance and fluctuation time scales in three geographic regions in the United States (Southwest, Southern Great Plains, and Hawaii) and from recent high density network measurements that both confirm and extend conclusions from previous studies, namely: (1) correlation coefficients decrease predictably with increasing distance, (2) correlation coefficients decrease at a similar rate when evaluated versus distance divided by the considered variability time scale, and (3) the accuracy of results is improved by including an implied cloud speed term.

Keywords: Variability; Photovoltaics; Solar resource; Fluctuations; Power grid

1. Introduction

PV capacity is increasing on utility systems. As a result, utility planners and grid operators are growing more concerned about potential impacts of power supply variability caused by transient clouds. Utilities and control system operators need to adapt their planning, scheduling, and operating strategies to accommodate this variability while at the same time maintaining existing standards of reliability.

It is impossible to effectively manage these systems, however, without a clear understanding of PV output variability or the methods to quantify it. Whether forecasting loads and scheduling capacity several hours ahead or planning for reserve resources years into the future, the industry needs to be able to quantify expected output variability for fleets of up to hundreds of thousands of PV systems spread across large geographical territories. Underestimating reserve requirements may result in a failure to meet reliability standards and an unstable power system. Overestimating reserve requirements may result in an unnecessary expenditure of capital and higher operating costs.

This issue has been the subject of considerable new work in the last few years – first in Japan and Germany where significant level of PV penetration started to occur (Wiemken et al., 2001, and Murata et al., 2009) and more recently in the US where this issue was brought to the forefront by utilities in the Southwest US (US Department of Energy, 2009). Numerous contributions to the field by the authors and others (Hoff and Perez, 2010; Perez et al., 2011a,b, have brought a better understanding of the temporal, spatial and site-specific characteristics of short-term variability (Sengupta, 2011; Lave and Kleissl, 2010, 2011; Frank

* Corresponding author.
E-mail addresses: tomhoff@cleanpower.com (T.E. Hoff), perez@asrc.albany.edu (R. Perez).

0038-092X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.solener.2011.11.005

et al., 2011; Kankiewicz et al., 2011; Hinkelman et al., 2011; Stein et al., 2011).

The objective of this article is to present the development of analytical methods and tools designed to quantify any arbitrary PV fleet's output variability, ranging from one single point system to multiple arbitrarily dispersed single point and extended PV systems. The methods and tools are designed to accommodate an arbitrary PV fleet because the composition of the PV fleet is defined by the electrical system configuration. For example, suppose utility planners are evaluating variability within a particular transmission or distribution system. The PV fleet of interest consists of PV systems that are connected to the transmission or distribution system, not simply PV systems that are in the same geographic location. Variability in time intervals ranging from a few seconds to a few minutes is of primary interest since control area reserves are dispatched over these time intervals.

Variability of a PV fleet is thus a measure of the magnitude of changes in its aggregate power output corresponding to the defined time interval and taken over a representative study period. Note that it is the *change* in output, rather than the output itself, that is desired. Also note that, for each time interval the change in output may vary in both magnitude and sign (positive and negative). A statistical metric is therefore employed in order to quantify variability: the *standard deviation of the change in fleet power output* $\sigma_{\Delta t}^{\text{fleet}}$ (Hoff and Perez, 2010).

$$\sigma_{\Delta t}^{\text{fleet}} = \sqrt{\text{Var}\left[\sum_{n=1}^{N} \Delta P_{\Delta t}^{n}\right]} \tag{1}$$

where N is the number of PV systems and $\Delta P_{\Delta t}^{n}$ is the time series of changes in power at the nth system occurring over a time interval of Δt.

It is helpful to graphically illustrate what is meant by output variability. The left side of Fig. 1 presents 10-s irradiance data (PV power output is almost directly proportional to irradiance) and the right side of the figure presents the change in irradiance using a 10-s time interval for a network of 25 weather monitoring stations in a 400-m by 400-m grid located at Cordelia Junction, CA on November 7, 2010. The light gray lines correspond to irradiance and variability for a single location and the dark red lines correspond to average irradiance distributed across 25 locations. This suggests that spreading capacity across 25 locations rather than concentrating it at a single location reduces variability. It is of course understood that the variability at any one of the single systems and the local transient-related interconnection issues such as output voltage and inverter tripping are not eliminated; however the transients of the ensemble and their related issues at a feeder or substation level are reduced by resource spreading.

A "fleet computation" approach can be taken to calculate output variability for a fleet of PV systems as follows: identify the PV systems that constitute the fleet to be studied; select the time interval and time period of concern (e.g., 1-min changes evaluated over a 1-year period); obtain time-synchronized solar irradiance data for each location where a PV system is to be sited; simulate output for each PV system using standard modeling tools; sum the output from each individual system to obtain the combined fleet output; calculate the change in fleet output for each time interval; and finally calculate the resulting statistical output variability from the stream of values.

A "fleet computation" approach, while technically valid, is difficult to implement in practice for several reasons. First multiple system calculations are highly computation intensive, and thus are not suitable for real-time operations particularly if the required time frequency is high. Second, solar irradiance data are not always available in sufficient time/space resolution – while commercial services such as SolarAnywhere (2011) are starting to offer products with a 1-min/1-km resolution, it may not be sufficient per se to address all questions down to scales of seconds and meters. Third, PV variability studies determined using the fleet computation approach would have to be re-commissioned whenever additional PV systems came on-line.

A more viable approach is to streamline the calculations through the use of a general-purpose PV output variability methodology. The method needs to quantify short-term fleet power output variability based on the premises that sky clearness and sun position drive the changes in the short-term output for individual PV systems and that technical specifications (i.e., dimensions, plant spacing, number of plants, etc.) determine overall fleet variability.

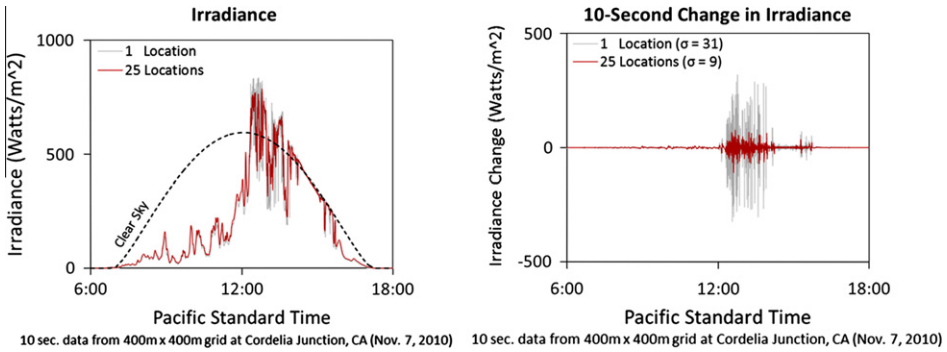


Fig. 1. Twenty-five location network reduces 10-s variability by more than 70% in a 400 m × 400 m grid at Cordelia Junction, CA on November 7, 2010.

Hoff and Perez (2010) developed a simplified model as a first step towards a general method to quantify the output variability resulting from an ensemble of equally-spaced, identical PV systems.

The simplified model covered the special case when the change in output between locations is uncorrelated (i.e., cloud impacts at one site are too distant to have predictable effects at another for the considered time scale), fleet capacity is equally distributed, and the variance at each location is the same. Under these conditions, Hoff and Perez showed that fleet output variability equals the output variability at any one location divided by the square root of the number of locations:[1]

$$\sigma_{\Delta t}^{\text{Fleet}} = \frac{\sigma_{\Delta t}^{1}}{\sqrt{N}} \tag{2}$$

where $\sigma_{\Delta t}^{1}$ is the standard deviation of the change in output of the fleet concentrated at one single location, and N is the number of uncorrelated locations. Mills and Wiser (2010) have derived a similar result that relates variability to the square root of the number of systems when the locations are uncorrelated.

2. Maximum output variability model

Eq. (2) has important implications for utility planners. It allows them to determine reserve capacity requirements to mitigate worst case fleet variability at any time scale of interest. For example, suppose that the variability of a single system was 10 kW per minute and there were 100 uncorrelated identical systems in the fleet. Total fleet variability equals 0.1 MW ($\frac{100\times 10\text{ kW}}{\sqrt{100}}$) per minute. The planner could then apply the desired confidence level (e.g., they may choose 3 standard deviations) to determine the required reserve capacity (e.g., $3 \times 0.1\text{ MW} = 0.3\text{ MW}$).

This calculation is applicable when two fundamental conditions are satisfied: (1) the output variability at a single location can be quantified and (2) the change in output variability between locations is uncorrelated.

Consider the first condition. One approach to determining single location variability ($\sigma_{\Delta t}^{1}$) is to analyze historical solar resource data for the location of interest. The data would need to have been collected at a rate that accommodates the time interval of interest (perhaps down to a few seconds) over a substantial and representative period of time (perhaps over several years). Such high-speed, high-resolution data are not generally available.[2]

An alternative approach is to construct a data set that simulates worst case variability conditions. The theoretically worst case variability of a single PV plant would be that it cycles alternately between 0% and 100% of its rated output every time interval. For example, suppose that the PV plant is rated at 1 MW and the time interval of interest is 1 min. As illustrated in Table 1 maximum variability occurs when the PV plant is at full power at 12:00, zero power at 12:01, full power at 12:02, etc. The corresponding change in power fluctuates between −1 and 1 MW. The standard deviation[3] of the change in power output equals 1 MW per minute. That is, a 1 MW PV plant that is exhibiting maximum variability over a 1 min time interval has a 1 MW per minute standard deviation. This would imply that 1 MW of reserve capacity is required to compensate for the output variability for a single plant.

Table 1
Maximum change in power output at one location.

Time	Power (MW)	Change (MW/min)
12:00	1	−1
12:01	0	+1
12:02	1	−1
12:03	0	+1
12:04	1	

Table 2
Maximum change in power output at two locations (scenario 1).

Time	Power (MW)			Change (MW/min)
	Plant 1	Plant 2	Fleet (1 + 2)	
12:00	0.5	0.5	1	−1
12:01	0	0	0	+1
12:02	0.5	0.5	1	−1
12:03	0	0	0	+1
12:04	0.5	0.5	1	

Table 3
Maximum change in power output at two locations (scenario 2).

Time	Power (MW)			Change (MW/min)
	Plant 1	Plant 2	Fleet 1 + 2	
12:00	0.5	0	0.5	0
12:01	0	0.5	0.5	0
12:02	0.5	0	0.5	0
12:03	0	0.5	0.5	0
12:04	0.5	0	0.5	

Suppose that the PV "fleet" capacity was split between two locations and each were to exhibit maximum output variability. Two possible scenarios exist. The first scenario, illustrated in Table 2, assumes that both plants turn on and off simultaneously. As was the case where all capacity is concentrated at a single location, the change in output fluctuates between −1 and 1 MW and the standard deviation for this scenario is 1 MW per minute.

The second scenario, illustrated in Table 3, assumes that the plants cycle on and off alternately with a time shift of

[1] See Eq. (8) in Hoff and Perez (2010).

[2] One of the few examples of this sort of data is provided by Kuszamaul et al. (2010).

[3] The standard deviation of a random variable X equals the square root of the expected value of X squared minus the square of the expected value of X. $\sigma = \sqrt{E[X^2] - E[X]^2}$.

Table 4
Maximum change in power output assuming random output.

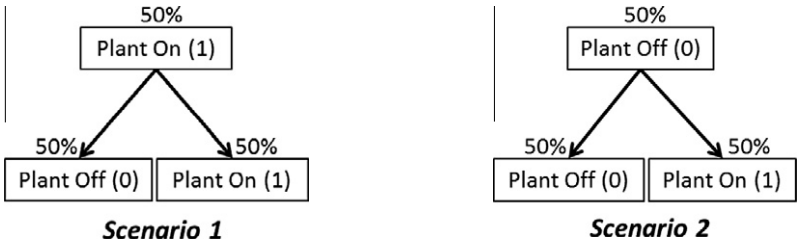
Time	Power (MW)	Change (MW/min)
12:00		25% chance if 1 50% chance if 0 25% chance if −1
12:01		

Table 5
Summary of input data.

Region	Southwest	Southern Great Plains	Hawaii
Location #1	Latitude: 32°–42° Longitude: −125° to−109° Grid Size: 2.0°	Latitude: 35°–38° Longitude: −99° to −96° Grid Size: 1.0°	Latitude: 19°–20° Longitude: −156° to −155° Grid Size: 0.5°
Location #2	0.1°, 0.3°, . . . , 1.9° from #1	0.1°, 0.3°, . . . , 2.9° from #1	0.1°, 0.2°, . . . , 1.0° from #1
Time intervals	1, 2, 3, and 4 h	1, 2, 3, and 4 h	1, 2, 3, and 4 h
Clear sky irradiance	10 irradiance bins in intervals of 0.1 kW/m^2	10 irradiance bins in increments of 0.1 kW/m^2	10 irradiance bins in increments of 0.1 kW/m^2

1 min. In this case, the change in output from the first location cancels the change in output at the second location. The result of this scenario is a standard deviation of 0 MW per minute.

It is incorrect to conclude, however, that the upper bound of output variability for 1 MW of PV is 1 MW per minute. This is because each of the two scenarios violates the assumed condition that the locations are uncorrelated. Specifically, the change in output between the two locations has perfect positive correlation in the first scenario (i.e., correlation coefficient equals 1) and perfect negative correlation in the second scenario (i.e., correlation coefficient equals −1).

2.1. Feasible maximum output variability

These scenarios demonstrate that it is impossible for two systems to exhibit the behavior of worst case variance individually (by cycling on and off at each interval) without having either perfect positive or perfect negative correlation. Indeed, for each system to exhibit its maximum variance, its output changes must be exactly in tempo with the time interval, loosely analogous to each member of an orchestra following in time to its conductor, in which case the systems would by definition have perfect correlation (whether positive or negative). By this reasoning, the maximum output variability scenario described above (1 MW of variability for each 1 MW of fleet capacity) is impossible. When the systems have less than perfect correlation, as must be the case for any real-world fleet, the variability of the combined fleet must be less than the total fleet capacity.

To correct the worst case scenario, retain the assumption that each power change is either a transition from zero output to full output or from full output to zero output. This assumption in itself is highly conservative since the impacts of cloud transients on PV systems will almost never produce changes with magnitudes as high as 100% of rated output and will generally produce changes much less than 100%. As for timing, rather than being synchronized, each system is assumed to cycle on and off in a random fashion, representing fleets of PV systems with outputs that are uncorrelated.

Random timing of power output changes is illustrated for a single location in Table 4 for a 1 MW PV system. Suppose that it is 12:00 and the time interval is 1 min. There is a 50% chance that the plant is on and a 50% chance that the plant is off at 12:00. If the plant is on at 12:00, then there is a 50% chance it will turn off and a 50% chance it will remain on at 12:01. If the plant is off at 12:00, then there is a 50% chance it will stay off and a 50% chance it will turn on at 12:01. The right column in Table 4 presents the probability distribution of the change in power. At each time interval, there is a 25% chance of a 1 MW per minute decrease in power, a 50% chance of no change in output, and a 25% chance of a 1 MW per minute increase in power.

Note that while this is the maximum possible change, it is extremely unlikely that such a distribution would

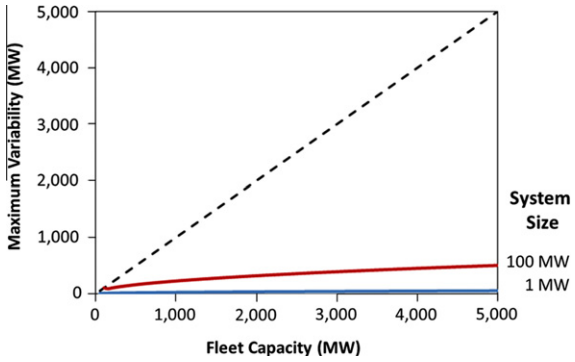


Fig. 2. Maximum variability for 1 MW and 100 MW system sizes with uncorrelated changes.

actually exist. First, weather conditions would have to be exceptionally erratic. Second, clouds would need to be so dark that there would be no output when covering a PV system. Third, the entire system would have to turn on and off, rather than a subset of the arrays. Fourth, each PV system would need to operate as a "point source" of output; Kuszamaul et al. (2010) and Mills et al. (2009) have demonstrated that, in fact, a smoothing effect occurs as system size increases.[4]

With these caveats, the above distribution represents an upper bound of worst case conditions that is conservative from a grid operator standpoint. This distribution has a standard deviation of $\frac{1}{\sqrt{2}}$ times 1 MW.[5] If the entire fleet of PV systems were concentrated at a single point, and the fleet had a capacity of C^{Fleet}, then the maximum standard deviation of change in output equals:

$$\text{Maximum } \sigma^{1}_{\Delta t} = \frac{C^{\text{Fleet}}}{\sqrt{2}} \tag{3}$$

The maximum output variability for a fleet of uncorrelated locations can be calculated using this numerical definition of the maximum output variability for a single system by substituting Eq. (3) into Eq. (2). The result is that maximum output variability equals fleet capacity divided by the square root of 2 times the number of uncorrelated locations.

$$\text{Maximum } \sigma^{\text{Fleet}}_{\Delta t} = \frac{C^{\text{Fleet}}}{\sqrt{2N}} \tag{4}$$

Eq. (4) places an upper bound on the maximum output variability for any time interval as long as the change in output between locations is uncorrelated. Actual results are likely to be lower – this practical upper bound on single point output is substantiated by a wealth of empirical evidence (see Perez et al., 2011).

[4] See Fig. 13 in Kuszamaul et al. (2010) and Fig. 7 in Mills et al. (2009).

[5] $\sigma = \sqrt{[(0.25)(-1)^2 + (0.50)(0)^2 + (0.25)(1)^2] - [(0.25)(-1) + (0.50)(0) + (0.25)(1)]^2} = \frac{1}{\sqrt{2}}$.

2.2. Example

Suppose that a utility system plans to incorporate 5000 MW of PV. Fig. 2 presents the maximum output variability calculated using Eq. (4) for PV fleets with capacities ranging from 0 to 5000 MW based on two fleet composition strategies. The blue line is the variability when the fleet is composed of uncorrelated 1 MW systems. The red line is the variability when the fleet is composed of uncorrelated 100 MW systems. As illustrated in the figure at the 5000 MW level, if 100 MW systems are installed at 50 locations ($N = 50$) with uncorrelated changes in output, maximum output variability is 500 MW per time interval, or 10% of fleet capacity. However, if 1 MW PV systems are installed at 5000 locations ($N = 5000$) with uncorrelated changes in output, maximum output variability is 50 MW, or 1% of fleet capacity.

This example illustrates the potential benefit of dividing the PV capacity into small systems, and spreading them apart geographically so that output changes are uncorrelated. More importantly, it also illustrates the unnecessary potential cost that could be incurred if system planners were to procure reserves without adequate tools for quantifying PV variability. The dotted line in Fig. 2 represents the reserve resources that would be procured when each MW of PV was fully "backed up" with a MW of fossil, battery, or other dispatchable resource intended to remove variability at the time scale of interest. In the $N = 5000$ example, such a planning practice — at least for fleets made up of uncorrelated systems— would result in capital expenditures 99 times the required amounts.

2.3. General model

The preceding section assumed that changes in the output from the different plants are uncorrelated. This section develops a model that considers what happens when the output between the various plants is correlated.

Eq. (1) stated that the standard deviation of the change in fleet output equals the square root of the variance of the sum of the change in output from each of the systems individual. The variance of the sum, however, equals the sum of the covariance of all possible combinations.

$$\sigma^{\text{fleet}}_{\Delta t} = \sqrt{\text{Var}\left[\sum_{n=1}^{N} \Delta P^{n}_{\Delta t}\right]} = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} \text{COV}(\Delta P^{i}_{\Delta t}, \Delta P^{j}_{\Delta t})} \tag{5}$$

The covariance between any two plants equals the standard deviations of each of the locations times the correlation coefficient between the two locations (i.e., $\text{COV}(\Delta P^{i}_{\Delta t}, \Delta P^{j}_{\Delta t}) = \sigma^{i}_{\Delta t}\sigma^{j}_{\Delta t}\rho^{i,j}_{\Delta t}$). As a result,

$$\sigma^{\text{fleet}}_{\Delta t} = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} \sigma^{i}_{\Delta t}\sigma^{j}_{\Delta t}\rho^{i,j}_{\Delta t}} \tag{6}$$

The critical observation to be made about Eq. (6) is that the standard deviation of the changes in fleet output is based entirely on the standard deviation of the change in plant output at each location and the correlation between the locations.

The result is that it is crucial to understand and quantify correlation between the various plants.

3. Correlation versus distance

3.1. Background: critical factors affecting correlation

The critical factors that affect output variability are the clearness of the sky, sun position, and PV fleet technical specs (i.e., dimensions, plant spacing, number of plants, etc.). Hoff and Perez (2010) introduced a parameter called the Dispersion Factor. The Dispersion Factor is a parameter that incorporates the layout of a fleet of PV systems, the time scales of concern, and the motion of cloud interferences over the PV fleet. Hoff and Perez showed that relative output variability resulting from the deployment of multiple plants decreased quasi-exponentially as a function of the generating resource's Dispersion Factor. Their results demonstrated that: (1) relative output variability decreases as the distance between sites increases; (2) the decrease with distance weakens as the time interval increases; and (3) the decrease weakens as well as the cloud transit speed increases.

Mills and Wiser (2010) analyzed measured 1-min insolation data over an extended period of time for 23 time-synchronized sites in the Southern Great Plains network of the Atmospheric Radiation Measurement (ARM) program (Stokes and Schwartz, 1994). Their results demonstrated[6] that the correlation of the change in the global clear-sky index: (1) decreases as the distance between sites increases and (2) decreases more slowly as the time interval increases.

Perez et al. (2011b) analyzed the correlation between the variability observed at two neighboring sites as a function of their distance and of the considered variability time scale. They used 20-s to 1-min data to construct virtual networks at 24 US locations from the ARM network and the SURFRAD Network and cloud speed derived from SolarAnywhere (2011) to calculate the station pair correlations for distances ranging from 100 m to 100 km and from variability time scales ranging from 20 s to 15 min. Their results confirmed that the correlation of the change in global clear-sky index: (1) decreases predictably as the distance between sites increases; and (2) decreases more slowly as the time interval increases.

The consistent conclusions[7] of these studies are that correlation: (1) decreases as the distance between sites increases and (2) decreases more slowly as the time interval increases. Hoff and Perez (2010) add that the correlation decreases more slowly as the speed of the clouds increases. New experimentally-based evidence tends to confirm qualitatively these general trends (e.g., Sengupta, 2011; Lave and Kleissl, 2011; Hinkelman et al., 2011) noting (Hinkelman et al., 2011) that other factors, such as direction with respect to cloud speed also could modulate observed relationships.

3.2. Determination of station pair correlation

New evidence is brought forth in this article to quantify the station-pair correlations dependence upon distance and time interval. This evidence includes: (1) a ***macro scale analysis*** of regional satellite-derived irradiances with time scales ranging from one to 4 h and distances ranging from 10 km and up, and (2) a ***micro scale view*** analyzing 10 s data from a high density 25-station network.

The analysis is focused on the clear-sky index Kt^* that equals the measured global horizontal insolation (GHI) divided by the clear-sky insolation, thereby removing much of the predictable solar geometry-induced variability (Mills and Wiser, 2010; Perez et al., 2011b). Specifically the change in the clear-sky index between two points in time is referred to as ΔKt^*. Since the change occurs over some specified time interval, Δt, at some specific location n, the variable is fully qualified as $\Delta Kt^{*n}_{t,\Delta t}$. This only represents one pair of points in time. A set of values is identified by convention by bolding the variable. Thus, $\Delta Kt^{*n}_{\Delta t}$ is the set of changes in the clear-sky indices at a specific location using a specific time interval over a specific time period.

$$\Delta Kt^{*n}_{\Delta t} = \left\{ (t_1, \Delta Kt^{*n}_{t_1,\Delta t}), (t_2, \Delta Kt^{*n}_{t_2,\Delta t}), \ldots, (t_T, \Delta Kt^{*n}_{t_T,\Delta t}) \right\} \quad (7)$$

Let $\Delta Kt^{*1}_{\Delta t}$ and $\Delta Kt^{*2}_{\Delta t}$ represent two sets of observed data values for the change in the clear-sky index that have a mean of 0 and standard deviations, σ_1 and σ_2.[8]

Pearson's product-moment correlation coefficient (typically referred to simply as the correlation coefficient) equals the expected value of $\Delta Kt^{*1}_{\Delta t}$ times $\Delta Kt^{*2}_{\Delta t}$ divided by the corresponding standard deviations.

$$\rho^{1,2} = \frac{E[\Delta Kt^{*1}_{\Delta t} \Delta Kt^{*2}_{\Delta t}]}{\sigma_1 \sigma_2} \quad (8)$$

3.3. Macro scale satellite data analysis

The correlation coefficients were computed for station pairs extracted from the gridded satellite-derived SolarAnywhere dataset (2011). This dataset uses a model that evolved from the model of Perez et al. (2004) which had been applied to produce the US National Solar Resource Data Bases (1998–2005). Whereas satellite-derived data

[6] See Fig. 5 in Mills and Wiser (2010).

[7] The results apply to either changes in PV output directly or changes in the clear sky index.

[8] The expected value of ΔKt^* equals 0 as long as the starting and ending GHI values are the same. This condition is satisfied when the time period of the analysis is performed over one day because the starting and ending GHI both equal 0. It will also be approximately true when the analysis encompasses many data points (as would be the case, for example, of an analysis of 1 h of data using a 1-min time interval).

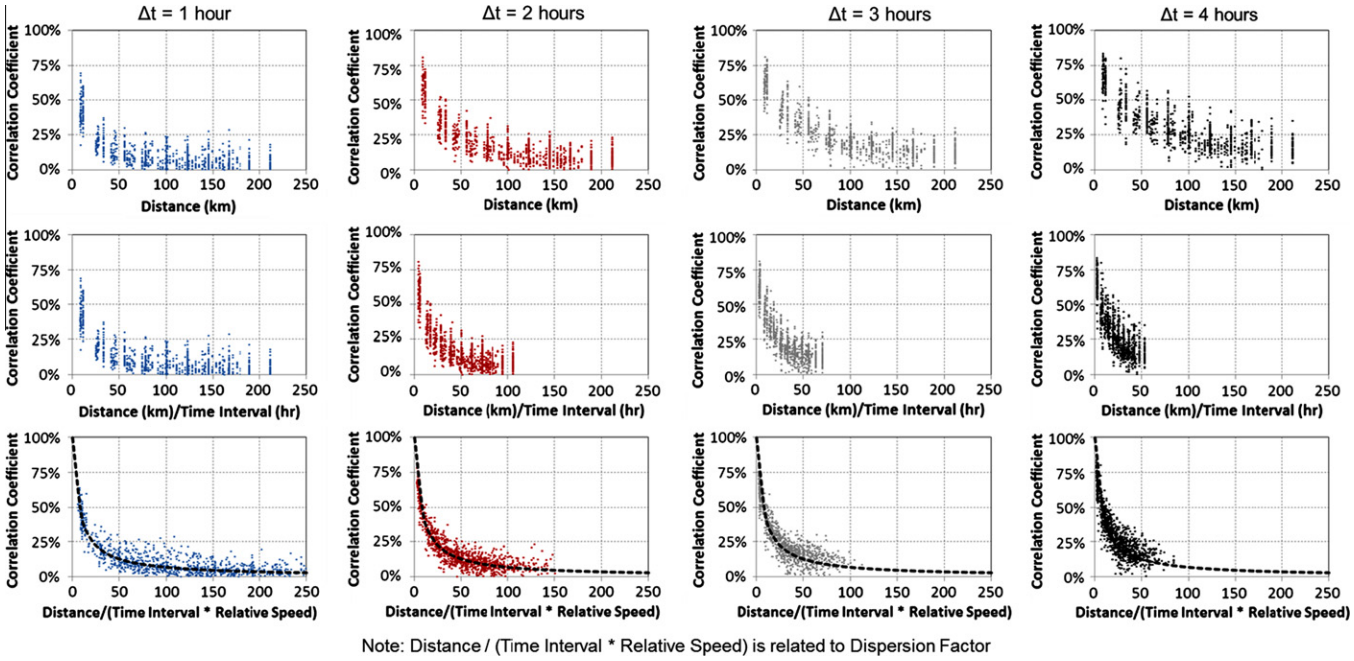


Fig. 3. Correlation coefficients presented by time interval for Southwest.

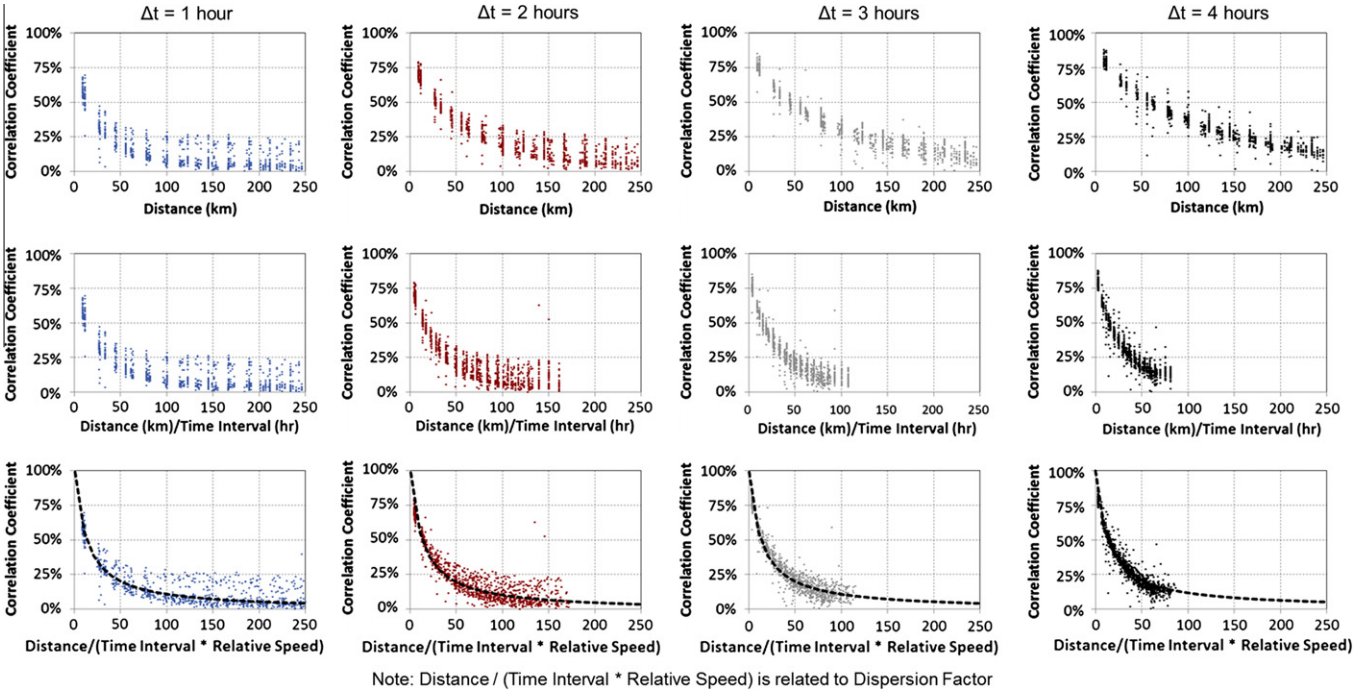


Fig. 4. Correlation coefficients presented by time interval for Great Plains.

fall short of site-specific measurements in terms of absolute accuracy, their strength lie in relative accuracy, hence are well suited to determine point to point differences and correlations. As summarized in Table 5, three separate geographic regions in the United States were selected for analysis: Southwest, Southern Great Plains, and Hawaii. For each pair of points analyzed, the first location, was selected using a grid size of 2.0°, 1.0°, or 0.5° for the Southwest, Southern Great Plains, and Hawaii, correspondingly, and the second location was selected between 0.1° and 2.9° (about 10–300 km) from the first location in north–south and east–west directions. Many

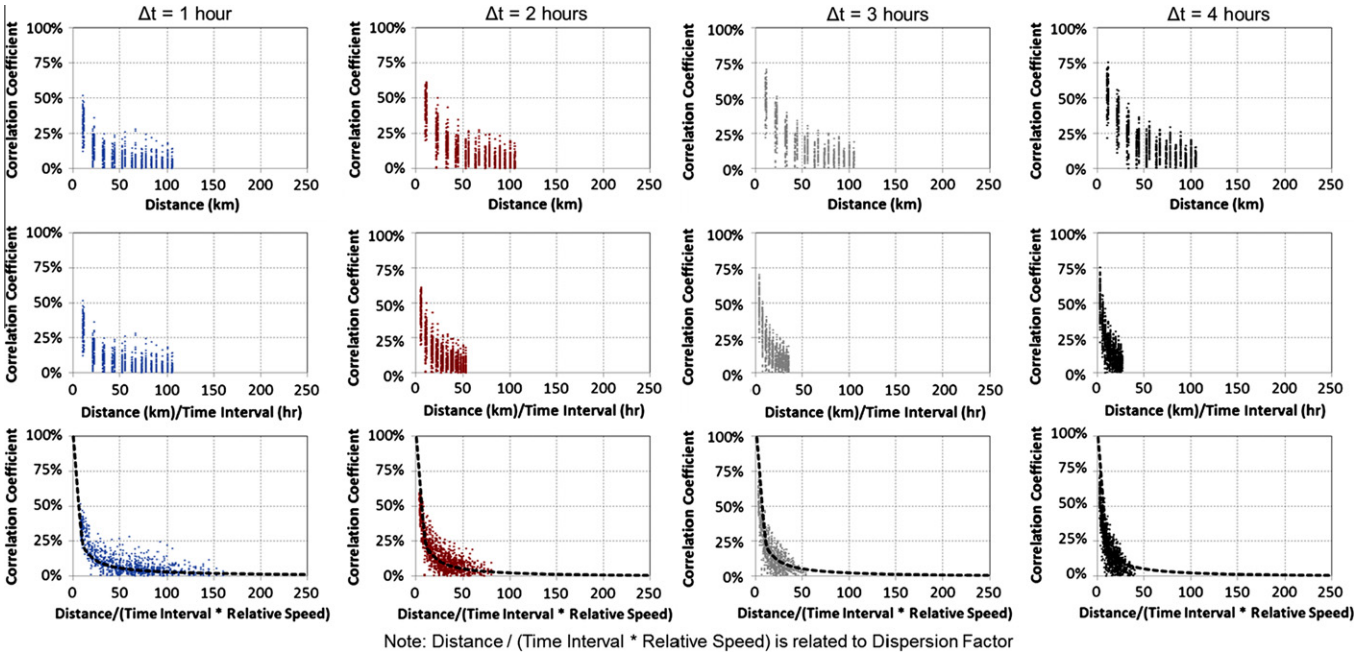


Fig. 5. Correlation coefficients presented by time interval for Hawaii.

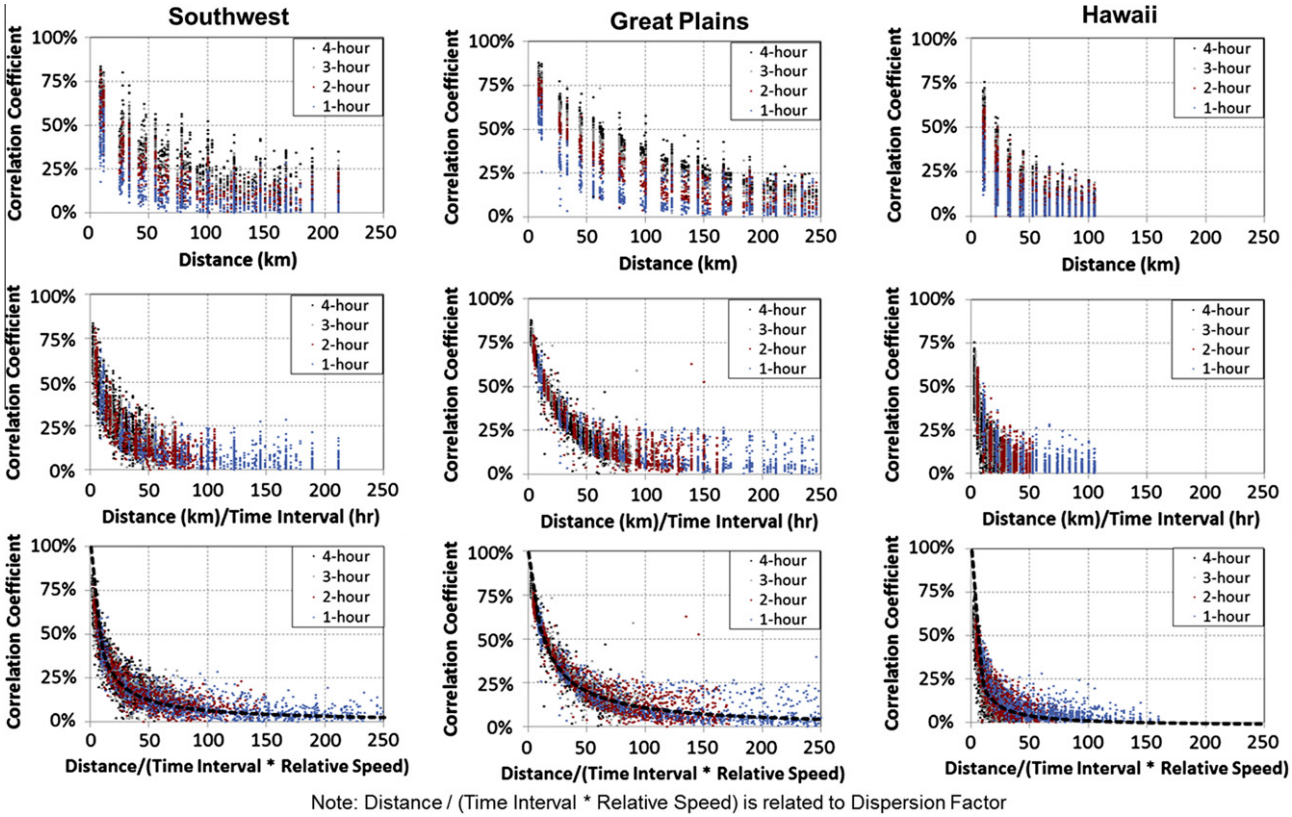


Fig. 6. Correlation coefficients for all locations and time intervals.

other map coordinates and possible pair of points were available on the SolarAnywhere $0.1° \times 0.1°$ grid, but selected points provided samples: over 70,000 pairs in total. Hourly insolation data were obtained for each selected pair of locations from January 1, 1998 through September 30, 2010. The site-pair correlation analysis was then performed as described above for time intervals (Δt) of 1, 2, 3, and 4 h.

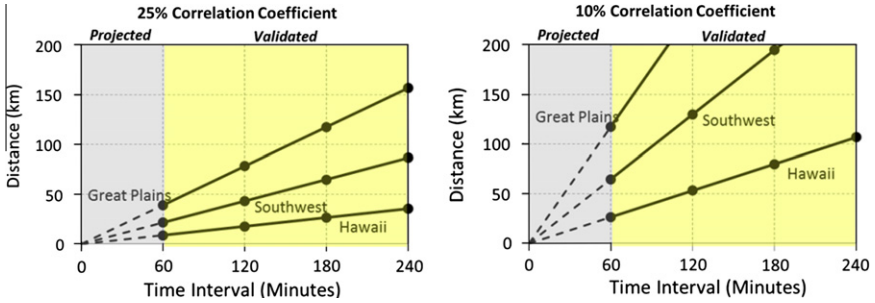


Fig. 7. Results scale linearly with the time interval for a fixed correlation coefficient.

3.4. Results

Fig. 3 presents the correlation coefficients for the Southwest.[9] The columns summarize the results for each time interval and the rows present the measured correlation coefficients versus several alternative candidate sets of variables. The first column summarizes results for a time interval of 1 h. The second, third, and fourth columns plot the same results using time intervals of 2, 3, and 4 h. Results in the top row present correlation coefficients versus the distance between the two locations. Results in the middle row present correlation coefficients versus distance divided by time interval. Results in the bottom row present correlation coefficients versus distance divided by time interval multiplied by an implied relative cloud speed;[10] this estimated cloud speed is related to the Dispersion Factor introduced by Hoff and Perez (2010). The dashed line in the bottom figures represents the results of a generalized method, proposed in this paper for use in future tools, that will be validated in the present analysis. Results are calculated using parameters obtained from SolarAnywhere. Figs. 4 and 5 present comparative results for the Great Plains and Hawaii. The patterns presented in the figures are similar across all time intervals in the three geographic locations. Fig. 6 compresses the results for each location and presents results where all time intervals are combined into the same figure.

3.5. Discussion

The analysis provides several key findings. First, consistent with previous studies, the correlation coefficients decrease with increasing distance (top row of Fig. 6). Second, also consistent with previous studies, this decrease occurs more slowly with longer time intervals (top row of Fig. 6). An alternative way of viewing this result is that correlation coefficients decrease at a similar rate when plotted versus distance divided by time interval (middle row of Fig. 6). Third, the scatter in results is further decreased when an implied relative speed[10] is introduced for the first location in the pair of locations (bottom row of Fig. 6). Finally, a model, shown by the dashed black line in the bottom row of Fig. 6, fits the empirical data well when calibrated using the location-specific derived input parameters, where

$$\rho = \frac{1}{1 + \frac{\text{Distance}}{(\Delta t)(\text{Relative speed})}} \tag{9}$$

3.6. Microscale analysis

3.6.1. Projection to shorter time intervals

An encouraging result of the foregoing analysis is the ability of the proposed general method, validated directly with several empirical data sets, to predict correlation coefficients with such accuracy. Even more encouraging is that the method is shown to be valid regardless of the selected time interval. While input data to produce Eq. (9) was taken from the SolarAnywhere data set with a 1-h time interval, the method is shown to produce accurate correlation coefficients for 1-h, 2-h, 3-h, and 4-h time intervals. This finding prompted the authors to evaluate the potential of using the method based on parameters derived from the SolarAnywhere data set to project results to time intervals shorter than 1 h.

While the desired objective is to demonstrate that the method accurately determines correlation coefficients (and therefore variability) as a function of PV spacing, a mathematically equivalent objective is to show that, for a given correlation coefficient, it is possible to accurately determine spacing between PV systems.

The circles in Fig. 7 correspond to the method results taken from the dotted curve in the bottom row of Fig. 6. For example, Fig. 6 implies that PV systems need to be spaced 40 km apart in the Great Plains in order to achieve a 25% correlation coefficient using a 60 min time interval. Triple the time interval to 180 min and plants need to be spaced triple the distance (120 km apart) to achieve the same 25% correlation coefficient.

The solid lines connecting the four time interval observations for each location in Fig. 7 illustrate that the relationship is linearly related to the time interval. The figure

[9] There were a large number of results for this analysis. Plotting all of the results made it difficult to read the figure, particularly when the different time intervals were overlaid on the same figure. Thus, the figures present randomly selected samples of the results to make the figures more readable.

[10] Implied relative cloud speed equals the implied speed derived for the specific location from SolarAnywhere data by the average implied speed across the entire geographic region. Note that this implied cloud speed is solely used for presentation purposes for the benefit of the reader so that the scale of the x-axis remains constant.

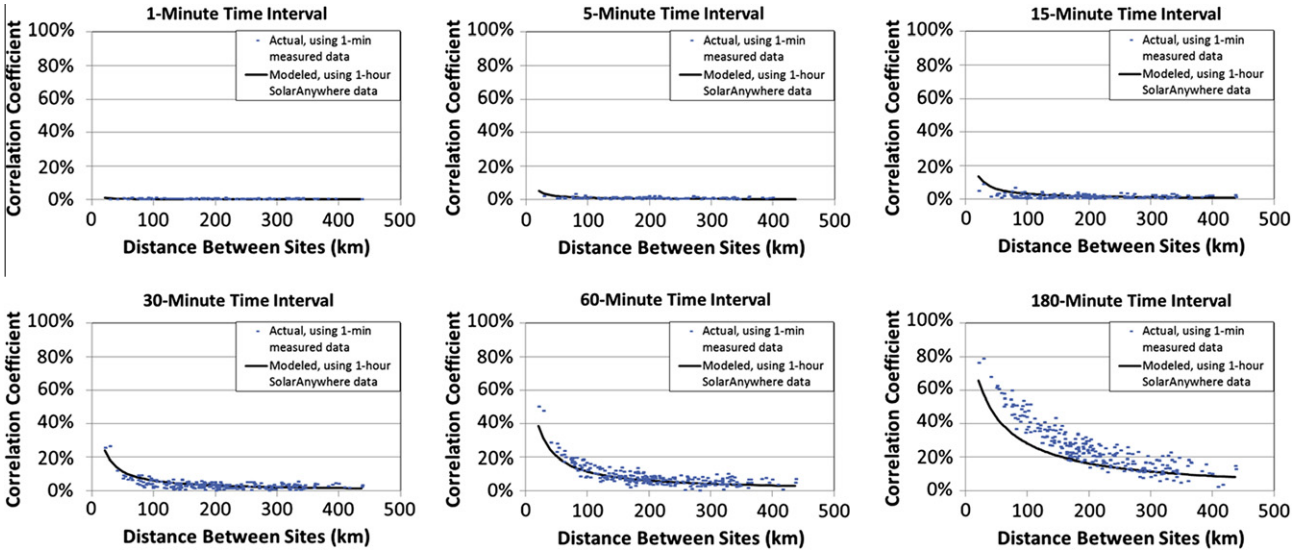


Fig. 8. Comparison of results to geographic diversity study (Mills and Wiser, 2010).

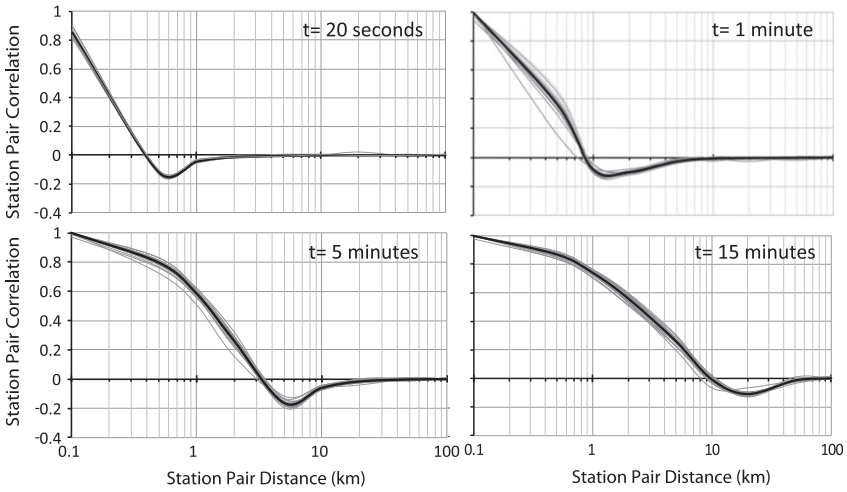


Fig. 9. Key results from virtual network study (Perez et al., 2011b).

begs the question as to whether the results can be projected in the region with shorter time intervals (i.e., the gray sections of the figures).

3.6.2. Evaluation of time-independence claim

The above linear relationship suggests that the method is independent of selected time interval, even down to the very short time intervals (several seconds to several minutes) that are of primary interest to utilities particularly for issues pertaining to PV integration within their distribution network. This section provides an initial validation of time-independence by comparing results calculated from the 1-h SolarAnywhere data set against results from independent studies that used 10-s, 20-s, and 1-min data sets.

3.6.2.1. Geographic diversity study. Mills and Wiser (2010) used measured 1-min insolation data for 23 time-synchronized sites in the Southern Great Plains network of the Atmospheric Radiation Measurement (ARM) program to characterize the variability of PV with different degrees of geographic diversity. That report presented[11] the correlation of changes in global clear-sky index between these geographically dispersed sites. Mills and Wiser provided an electronic version of their results and these were used to compare against the general method proposed here. While the 1-h SolarAnywhere data set was used as input to the general method, correlation coefficients were calculated that corresponded to much shorter time intervals in the Mills and Wiser study. The results, presented in Fig. 8, are comparable to the Mills and Wiser study even down to 1-min time intervals.[12]

[11] Fig. 5 in Mills and Wiser (2010).

[12] The differences in 180-min time intervals are due to methodological differences between the two studies in how the interval results were calculated.

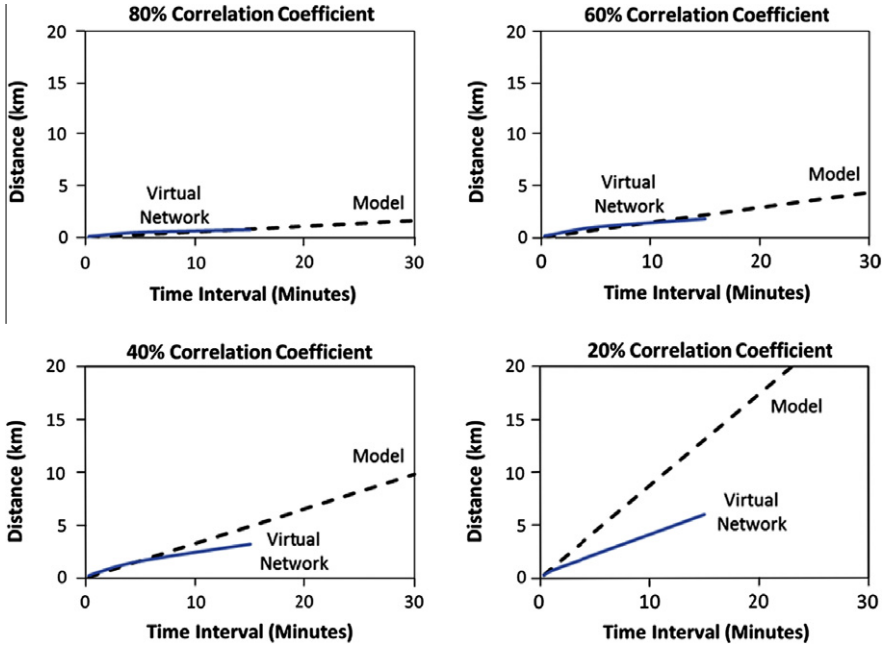


Fig. 10. Comparison of results to virtual network study.

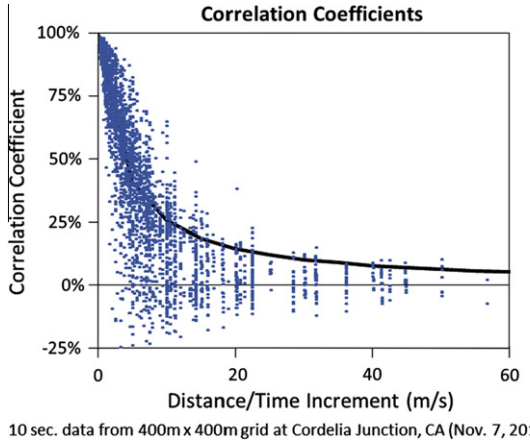


Fig. 11. Correlation coefficients for high-density, 25 unit network at Cordelia Junction, CA on November 7, 2010 for time intervals from 10 s to 5 min.

3.6.2.2. Virtual network study. Perez et al. (2011b) obtained 20-s to 1-min insolation data for 24 measuring stations, including 17 stations in the ARM network and 7 stations in the SURFRAD network. They constructed 1-dimensional virtual networks[13] using satellite-derived cloud speeds to translate time measurements into space measurements. They then calculated correlation coefficients between the change in clear-sky index for various time intervals and distances. Fig. 9 presents some of the key results from that study. Fig. 10 re-plots the data from the virtual network study along with corresponding projections from Eq. (9). Results compare well to virtual network study down to correlation coefficients of 40% for time intervals between 20 s and 15 min. Results from the virtual network study correlation coefficients below 40% may be lower as a result of the negative correlation arising from locations that are very close together and because of the one dimensional nature of the virtual networks.

3.6.2.3. High density weather station network. A third data set was provided based on a data set from a network of 25 weather collection devices. This network is interesting from several perspectives. First, it is one of the few known high-density networks providing high speed data (see Kuszamaul et al., 2010 for a network of 24 sensors in Lanai, HI). Second, it is designed to be deployed to multiple locations for short durations of time and thus is mobile.

The clocks on the 25 data recording devices were set using a single computer to ensure that time synchronized results would be obtained. This network was then deployed at Cordelia Junction, CA in a 400-m by 400-m configuration (a square composed of 100 m between stations). Fig. 11 presents the correlation coefficients for November 7, 2010. Since there are 25 locations, there are 625 possible combinations, 300 of which are unique. Each of these combinations was evaluated using nine different time intervals (10, 20, 30, 40, 50, 60, 90, 120, and 300 s). Thus, there are 2700 unique scenarios.

The black line in the figure represents the relationship proposed in Eq. (9). These independently measured data fit the proposed method fairly well. It is interesting to note that this data set exhibits some of the negative correlation

[13] See Hoff and Perez (2010) for a discussion of virtual network construction.

effects identified by Hoff and Perez (2010) and Perez et al. (2011b) using the virtual network approach.

4. Conclusions

The objective of this paper was to lay the foundation for a new method that could be employed in future utility tools to enable the calculation of PV fleet variability for planning and operational purposes.

A maximum output variability model was introduced as a practical tool for utilities to size reserve capacity requirements applicable to arbitrary time scales. One of the key inputs to this model is the correlation coefficients between the variability of individual plants composing a PV fleet at any considered time scale.

A method was proposed to extract such coefficients based upon station distance and implied cloud speed. Hourly global horizontal insolation data from SolarAnywhere were used to validate the method by calculating correlation coefficients for 70,000 pairs of points across three separate geographic regions in the United States (Southwest, Southern Great Plains, and Hawaii), while varying distance, time interval, insolation bin, and other parameters. These empirical correlation coefficients compared favorably with those derived by the method. The method was then shown to be independent of selected time interval, such that hourly satellite data could be used to calculate correlation coefficients for very short time intervals (several seconds to several minutes). These extrapolated results were validated using results from studies that are based on 20-s to 1-min insolation data and using high density network data with time scales of 10 s to 5 min.

The article presented and confirmed important findings. First, correlation coefficients decrease with increasing distance. Second, correlation coefficients decrease at a similar rate when plotted versus distance divided by time interval. Third, the accuracy of results is further improved when an implied speed term is introduced into the analysis. Together, these results provide the basis for validating the proposed site-pair correlation method. The method, derived with input parameters from hourly SolarAnywhere data, can produce correlation coefficients for short time intervals (seconds to minutes) that compare quite well to results from independent studies that used 10-s, 20-s, and 1-min data sets.

Acknowledgements

Portions of this study were funded under a California Solar Initiative (CSI) Grant Agreement titled "Advanced Modeling and Verification for High Penetration PV." The California Public Utilities Commission is the Funding Approver, Itron is the Program Manager, and the California IOUs are the Funding Distributors. Thanks to Ben Norris (Clean Power Research) who designed, implemented, and operated the mobile irradiance network and provided valuable comments on the paper. Thanks to Jay Apt (Carnegie Mellon University), Andrew Mills (Lawrence Berkeley Labs), Sophie Pelland (Natural Resources Canada), Jeff Ressler (Clean Power Research), and Ken Zweibel (George Washington University) for their comments. Opinions expressed herein are those of the authors only.

References

Frank, J., Freedman, J., Brower, M., Schnitzer, M., 2011. Development of high frequency solar data. In: Proc. Solar 2011, American Solar Energy Society Conf., Raleigh, NC.

Hinkelman, L., George, R., Sengupta, M., 2011. Differences between along-wind and cross-wind solar variability. In: Proc. Solar 2011, American Solar Energy Society Conf., Raleigh, NC.

Hoff, T.E., Perez, R., 2010. Quantifying PV power output variability. Solar Energy 84, 1782–1793.

Kankiewicz, A., Sengupta, M., Li, J., 2011. Cloud meteorology and utility scale variability. In: Proc. Solar 2011, American Solar Energy Society Conf., Raleigh, NC.

Kuszamaul, S., Ellis, A., Stein, J., Johnson, L., 2010. Lanai high-density irradiance sensor network for characterizing solar resource variability of mw-scale PV system. In: 35th Photovoltaic Specialists Conference, Honolulu, HI. June 20–25, 2010.

Lave, M., Kleissl, J., 2010. Solar intermittency of four sites across the state of colorado. Renewable Energy 35, 2867–2873.

Lave, M., Kleissl, J., 2011. Solar Variability over various timescales using the top hat wavelet. In: Proc. Solar 2011, American Solar Energy Society Conf., Raleigh, NC.

Mills, A., Wiser, R., 2010. Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power. Lawrence Berkeley National Laboratory Technical Report LBNL-3884E.

Mills, A., Alstrom, M., Brower, M., Ellis, A., George, R., Hoff, T., Kroposki, B., Lenox, C., Miller, N., Stein, J., Wan, Y., 2009. Understanding Variability and Uncertainty of Photovoltaics for Integration with the Electric Power System. Lawrence Berkeley National Laboratory Technical Report LBNL-2855E.

Murata, A., Yamaguchi, H., Otani, K., 2009. 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.

NSRDB, 1998–2005. <http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/>.

Perez, R., Ineichen, P., Kmiecik, M., Moore, K., George, R., Renné, D., 2004. Producing satellite-derived irradiances in complex arid terrain. Solar Energy 77 (4), 363–370.

Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr., C., Hoff, T.E., 2011a. Parameterization of site-specific short-term irradiance variability. Solar Energy 85 (7), 1343–1353.

Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr., C., Hoff, T.E., 2011b. Short-term irradiance variability correlation as a function of distance. In: Proc. Solar 2011, American Solar Energy Society Conference, Raleigh, NC.

Sengupta, M., 2011. Measurement and modeling of solar and PV output variability. In: Proc. Solar 2011, American Solar Energy Society Conf., Raleigh, NC.

SolarAnywhere. 2011. Web-Based Service that Provides Hourly, Satellite-Derived Solar Irradiance Data Forecasted 7 days Ahead and Archival Data back to January 1, 1998. <www.SolarAnywhere.com>.

Stein, J., Ellis, A., Hansen, C., Chadliev, V., 2011. Simulation of 1-minute power output from utility-scale photovoltaic generation systems. In: Proc. Solar 2011, American Solar Energy Society Conf., Raleigh, NC.

Stokes, G.M., Schwartz, S.E., 1994. The atmospheric radiation measurement (ARM) program: programmatic background and design of the

cloud and radiation test bed. Bulletin of American Meteorological Society 75, 1201–1221.

US Department of Energy, 2009. Workshop Report on High Penetration of Photovoltaic (PV) Systems into the Distribution Grid, Ontario, CA, February 24–25, 2009.

Wiemken, E., Beyer, H.G., Heydenreich, W., Kiefer, K., 2001. Power characteristics of PV ensembles: experience from the combined power productivity of 100 grid-connected systems distributed over Germany. Solar Energy 70, 513–519.