Cloud Speed Impact on Solar Variability Scaling - Application to the Wavelet Variability Model

Matthew Lave
Jan Kleissl
University of California, San Diego
9500 Gilman Dr. #0411
La Jolla, CA 92093
mlave@ucsd.edu
jkleissl@ucsd.edu

Abstract

The wavelet variability model (WVM) for simulating solar photovoltaic (PV) powerplant output given a single irradiance sensor as input has been developed and validated previously. Central to the WVM method is a correlation scaling coefficient (A) that calibrates the decay of correlation of the clear sky index as a function of distance and timescale, and which varies by day and geographic location. Previously, a local irradiance sensor network was required to derive A. In this work, we determine A from cloud speeds. Cloud simulator results indicated that the A value is linearly proportional to the cloud speed (CS): $A = \frac{1}{2}CS$. Cloud speeds from a numerical weather model (NWM) were then used to create a database of daily A values for North America. For validation, the WVM was run to simulate a 48MW PV plant with both NWM A values and with ground A values found from a sensor network. Both WVM methods closely matched the distribution of ramp rates (RRs) of measured power, and were a strong improvement over linearly scaling up a point sensor. The incremental error in using NWM A values over ground A values was small. The ability to use NWM-derived A values means that the WVM can be used to simulate a PV plant anywhere a single high-frequency irradiance sensor exits. This can greatly assist in module siting, plant sizing, and storage decisions for prospective PV plants.

1. Introduction

The variable nature of power produced by PV power plants can be of concern to electric operators. For example, the Puerto Rico Electric Power Authority (PREPA) requires that utility-scale PV plants in Puerto Rico limit ramps (both up and down) to 10% of capacity per minute (PREPA). At short timescales such as 1-minute, the variability of solar PV power production is mostly caused by the movement of clouds across the PV plant. While a single PV module can produce highly variable output due to the instantaneous crossing of cloud edges, geographic diversity of modules within a PV plant will lead to smoothing of the total power output. Geographic diversity can be quantified through the correlation coefficients between the timeseries of power output of different PV modules within the plant. This correlation generally decreases with distance and increases with fluctuation timescale.

Irradiance and power measurements have been used to quantify the relative reduction in aggregate variability for a combination of sites. Sites a few to hundreds of kilometers apart were shown to lead to a smoothed aggregate output and the amount of smoothing varied based on the distances between sites and local meteorological conditions (Curtright and Apt, 2008, Lave and Kleissl, 2010, Otani, et al., 1997, Wiemken, et al., 2001). Other investigators (Mills and Wiser, 2010, Perez, et al., 2011, Perez, et al., 2012) calculated the correlation of irradiance fluctuations between sites and found decorrelation distances – the distances over which sites become independent of one another – to vary based on fluctuation timescale and distance between sites. Accounting for cloud speed further enhanced the accuracy of these correlation models (Hoff and Perez, 2012). Correlation was also shown to depend on orientation relative to the direction of cloud motion (Hinkelman, et al., 2011).

Wavelet analysis has been used to decompose solar fluctuations into different timescales and measure the variability at that timescale (Kawasaki, et al., 2006, Lave, et al., 2012, Woyte, et al., 2007). The wavelet transform and a smoothing factor that varies by timescale were then used to simulate the variability reduction in going from a single site to a powerplant (Lave, et al., 2012, Marcos, et al., 2011). In this paper, we build on the wavelet-variability model (WVM) presented in Lave, et al. (2012) by using cloud speeds to make the model more generally applicable. The WVM and its correlation equation are described in Section 2, the relationship between cloud speed and the correlation scaling coefficient (A value) is shown in Section 3, a method for determining cloud speeds and converting to A values is presented in section 4, and in section 5 we apply and validate the WVM with cloud-speed derived A values at a 48MW PV plant.

2. The Wavelet Variability Model (WVM)

2.1. WVM Description

To estimate the smoothing achieved over a solar power plant due to decorrelation within the plant, we have developed the WVM. The WVM is described in full in (Lave, et al., 2012). In short, the WVM takes as inputs a local irradiance point sensor, the PV plant footprint, PV plant capacity, and a correlation-scaling A value, and outputs a simulated plant power output timeseries. The simplified WVM procedure is:

1) Apply a wavelet transform to decompose the clear-sky index of the original irradiance timeseries into wavelet modes $w_{\bar{t}}(t)$ at various timescales, $\bar{t}$, which represent cloud-induced fluctuations at each timescale.
2) Determine the distances, $d_{m,n}$, between all pairs of 'sites' in the PV power plant; $m = 1, \dots, N$, $n = 1, \dots, N$. A site represents a single PV module.
3) Determine the correlations, $\rho\left(d_{m,n}, \bar{t}\right)$, between the wavelet modes at different sites. A location/day dependent correlation scaling coefficient (A) is needed to account for cloud speed (section 2.2).
4) Use $\rho\left(d_{m,n}, \bar{t}\right)$ to find the variability reduction, $\mathrm{VR}(\bar{t})$, at each timescale: $\mathrm{VR}(\bar{t}) = \frac{\mathrm{N}^2}{\sum_{\mathrm{m}=1}^{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \rho(d_{m,n}, \bar{t})}$
5) Divide each $w_{\bar{t}}(t)$ by the square root of the corresponding $\mathrm{VR}(\bar{t})$ to create simulated wavelet modes of the entire power plant. Apply an inverse wavelet transform to these scaled wavelet modes to yield the simulated clear-sky index of area-averaged irradiance over the whole power plant.
6) Convert this area-averaged clear-sky index into power output, $P(t)^{sim}$ by multiplying by a clear-sky power model e.g. determined from panel type, tilt and azimuth, and temperature. The temporal resolution of the simulated power output is the same as the temporal resolution of the original irradiance timeseries (e.g., 1s in, 1s out).

The WVM can be a valuable tool for simulating how often certain size ramp rates (RRs) will occur. For example, it can be used to estimate the number of RRs per day exceeding 10% of capacity (as is of special interest in Puerto Rico). The WVM has been limited, though, by the need for a local sensor network to derive A values (Step 3 above). In this paper, we present a method for determining A values based on cloud speeds from numerical weather forecasts, eliminating the need for a local sensor network. The motivation for introducing cloud speed is both mathematical and physical. Mathematically, dimensional analysis shows that the A value must have units of m s^{-1} (Eq. 1), and the only relevant speed is the cloud speed. Physically, the ratio of distance between sites to the distance traveled by the clouds is important for determining correlations. The distance traveled by the clouds is the cloud speed multiplied by the timescale.

2.2. Space-Time Correlations and A values

Estimating the correlations between the wavelet modes at different locations (step 3) is perhaps the most important step of the WVM. Based on previous works (Hoff and Perez, 2010, Mills and Wiser, 2010, Perez, et al., 2011, Perez, et al., 2012), it is clear that correlations between sites depend on both $d_{m,n}$ and t. The WVM models correlation between wavelet modes at different sites using the equation

$$\rho\left(d_{m,n}, \bar{t}\right) = \exp\left(-\frac{d_{m,n}}{A\,\bar{t}}\right). \quad 1$$

Eq. 1 assumes that correlations are isotropic, i.e. they do not depend on direction but merely the magnitude of distance between sites. Therefore it is meant to represent the average correlation for all site pairs in a powerplant.

2.3. Correlation Model Limitations

The assumption of isotropy in Eq. 1 is violated for two specific cases:

A. Along-wind sites can become negatively correlated when $d_{m,n} \approx CS\,\bar{t}$, where CS is the cloud speed; $CS\,\bar{t}$ is the distance traveled by the clouds. The correlation can even reach -1 if there is frozen cloud field advection and if clouds of length $d_{m,n}$ are spaced $d_{m,n}$ apart. Then, the same cloud or set of clouds that affects site m at time t affects site n at time $t + \bar{t}$. Negative correlation was demonstrated experimentally in the virtual network in Perez, et al. (2012) and in the along-wind direction in Hinkelman, et al. (2011). This negative correlation is caused by the time offset in "seeing" the cloud feature. At time $t + \bar{t}$, the top hat wavelet scaling function will have a value of -1 when multiplying the cloud feature at site m, and a value of 1 when multiplying that same cloud feature at site n. These opposite signs usually lead to negative correlation of wavelet modes, depending on the surrounding cloud features.

B. Sites can become uncorrelated when the cross-wind distance between them, $d_{m,n}^{crosswind}$, is greater than the cross-wind cloud feature size, CZ. A cloud feature does not have to be one single cloud, but can also be a cloud band or cloud front that is highly correlated along the cross-wind direction. For example, a cloud front may contain many individual clouds among which there is high correlation in the cross-wind direction. When $\frac{d_{m,n}^{crosswind}}{CZ} \ll 1$ sites m and n will see the same clouds features (but possibly at different onset times), and Eq. 1 applies. However, when $\frac{d_{m,n}^{crosswind}}{CZ} > 1$, the two sites are never affected by the same cloud feature. Thus, correlations for the cases where $\frac{d_{m,n}^{crosswind}}{CZ} > 1$ do not depend on distance or timescale, but only on the correlation between random pixels of the cloud field.

In both limitations A and B, Eq. 1 will tend to overestimate correlations. However, in simulating PV powerplants, we expect the errors to be small. Limitation A will only affect sites at the timescale $\bar{t} \approx \frac{d_{m,n}}{CS}$, and at that primarily affects sites aligned close to the along-wind direction. When many timescales and pairs of sites with different orientations are averaged over a powerplant, the error caused by limitation A will be minimal. Limitation B can have a significant effect if the distances across the powerplant are on the same order as the typical cross-wind cloud size. However, since powerplant length scales (meters to a few kilometers) are small relative to mesoscale weather effects, we expect the ratio $\frac{d_{m,n}^{crosswind}}{CZ}$ to be sufficiently small across the powerplant such that limitation B also has a minimal effect on correlations.

2.4. Calculating the A Value from a Sensor Network

When an irradiance sensor network exists near the powerplant to be simulated, the value of A can be found by back solving Eq. 1. Distance, timescale, and correlation are known from measurements in the network and a best fit A value can be determined (Fig. 1). The slight variation of correlations above and below the A value best fit in Fig. 1 are likely due to limitations A and B.

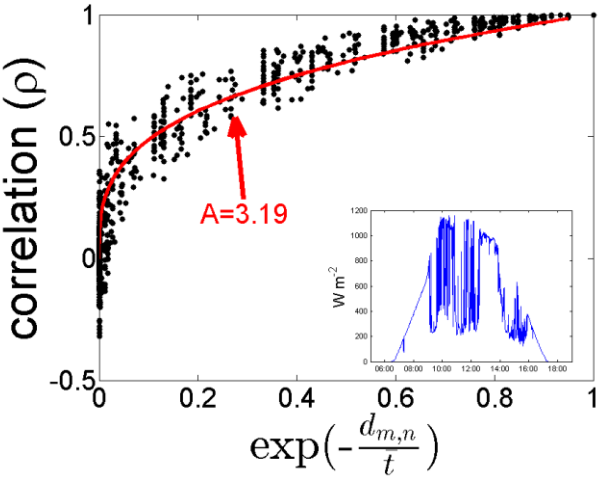

Fig. 1: Correlations between wavelet modes of clear sky indices measured in a sensor network. The x-axis is $\mathbf{exp}(-\frac{d_{m,n}}{t})$ to show the exponential decay of correlation as a function of distance and timescale. The red line is the correlation modeled using Eq. 1, where the $\boldsymbol{A}$ value (3.19) was fit. The plot in the bottom right shows the plane-of-array (POA) irradiance profile on this day.

This method requires a sensor network measuring at high frequency collocated with the powerplant to be simulated. This severely limits the application of the WVM as very few sensor networks currently exist at prospective sites for which WVM simulations are desired.

3. *A* Values and Cloud Speed from a Cloud Field Simulator

3.1. Cloud Field Simulation Method

To better understand the dependence of the *A* value on the physical variables of cloud size and cloud speed, we created a simple cloud simulator. Simulated cloud fields were created by first using a coarse grid, where one pixel was the size of one cloud. To simulate 1000 m clouds, for example, 1000 by 1000 m pixels were used. A uniformly distributed ([0 1]) random number was assigned to each pixel in the coarse grid. All coarse grid pixel values less than the specified cloud cover fraction were set to one to represent clouds, and all other pixels were set to zero to represent clear-sky. This procedure was used as a way to control the cloud size and reduced processing time over other methods such as randomly drawing circles in the field. In this way, a coarse cloud field was created, such as the one shown in Fig. 2a. The coarse field was then converted into a smooth cloud field at high resolution using a 2D spline interpolation. In the fine cloud field, each pixel was 1 by 1m, regardless of the cloud size. To maintain cloud cover fraction, after interpolation, values less than 0.5 were set to zero and greater than 0.5 were set to one. An example fine cloud field is shown in Fig. 2b.

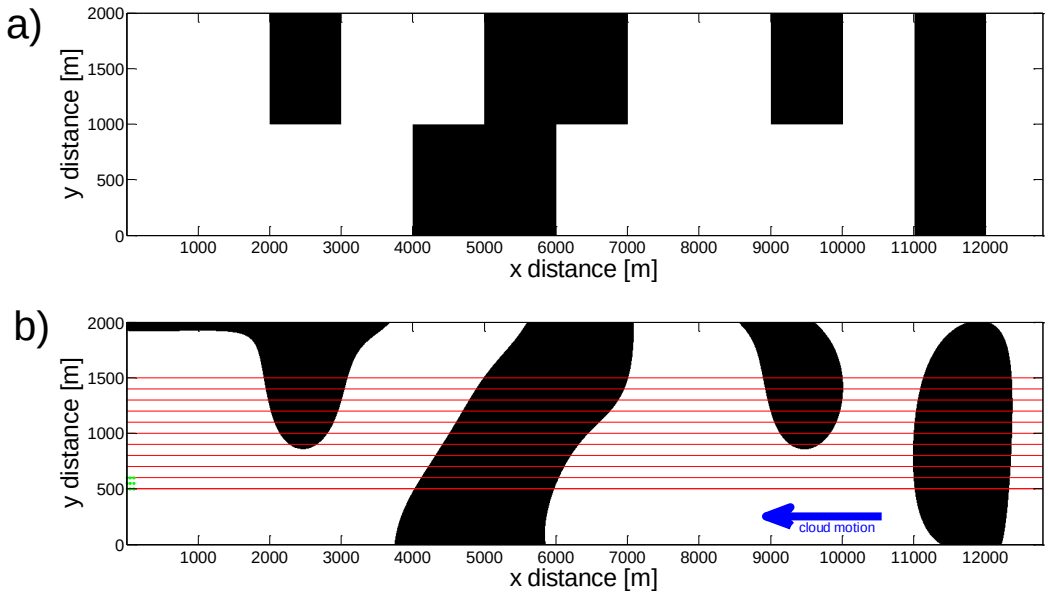

Fig. 2: Snapshot of the domain for (a) the coarse cloud field and (b) the fine cloud field. White represents clouds and black represent clear-sky. The blue arrow shows the direction of cloud motion, the green dots (at ~ x = 0, y = 500 m) show an example sensor network, and the red lines show the crosswind locations for the top edge of the sensor networks. These images were created for the case where the cloud size was 1000m, and the cloud cover fraction was 0.7.

To determine A values, a virtual sensor network was set up in the cloud field. To simulate cloud motion, the cloud field was advected in the negative x direction with the cloud speed. The simulated clear-sky index timeseries $kt(t)$ for a sensor at position (x, y) was then $kt(t) = (x + CS * t, y)$, where CS is the cloud speed. For simplicity, we assumed a kt value of 0 for cloudy pixels and 1 for clear pixels. As described in section 2.4, the correlations of wavelet modes of $kt(t)$ for various sensor pairs were found, and Eq. 1 was back solved to determine the A value for the specified cloud size and speed. Due to computational limitations, the length of the cloud field in the x-direction was always 12,800 m. This meant that the length of the timeseries kt was $(12{,}800\text{ m})/CS$. The shortest timeseries ($CS = 25\text{m s}^{-1}$) was $2^9 = 512$ s long, allowing for resolution of 9 wavelet modes. This is a sufficient range of timescales to produce accurate A value results.

The virtual sensor network was made up of 3 x 3 points spaced across a Cartesian grid with 50 m spacing. The location of the sensor network in the y-direction was varied to account for all possible sensor-cloud positions (i.e., sensor field in center of cloud, at edge of cloud, etc.). For each cloud size, 10 different cross-wind sensor y-offset were used. Assuming periodicity of the statistics with a period equal to the cloud size, the y-offset ranged from $\frac{1}{2}$ the cloud size to $\frac{3}{2}$ the cloud size in steps of 1/10$^{\text{th}}$ the cloud size (500 to 1500 m in Fig. 2). The offsets were started at $\frac{1}{2}$ the cloud size to eliminate edge effects caused by the spline interpolation at the boundary of the cloud field ($y = 0$). We ran this cloud simulation for cloud sizes of 100 to 3000 m and cloud speeds of 1 to 25 m s^{-1}, all with a cloud fraction of 0.7.

3.2. Relating *A* Value and Cloud Speed

The cloud simulator domain was setup 200 times for each cloud speed and cloud size pair to ensure that results were statistically representative. Simulations that yielded a relative root mean squared error (rRMSE) of the fit of the A value to the simulated correlations greater than 15% were not used. The mean of the remaining A values for each cloud speed and cloud size pair are shown in Fig. 3a. Fig. 3b shows the A value as a function of cloud speed.

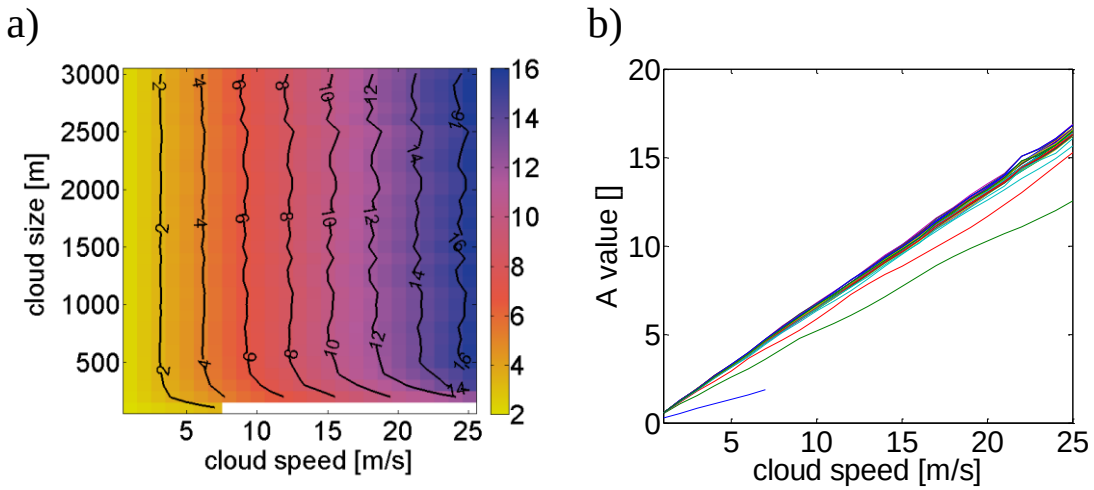

Fig. 3: ***A*** values determined through cloud field simulation: (a) on a 2D plot showing the ***A*** value as a function of cloud speed and cloud size, and (b) on a line graph showing the ***A*** value as a function only of cloud speed. Most of the ***A*** value fits for the 100 m cloud size had errors larger than the 15% threshold, resulting in the white unresolved area in (a) and short blue line in (b). The three lines in (b) that diverge from other lines are for the 100m (blue), 200m (green), and 300m (red) cloud sizes.

Since A is close to linearly proportional to cloud speed (Fig. 3b), we define:

$$A = C_1 * CS. \qquad 2$$

However, C_1 was not constant at slow cloud speeds or small cloud sizes. For example, the C_1 value at 2 m s^{-1} cloud speed was on average 20% lower than the C_1 value at 25 m s^{-1} cloud speed. However, as cloud speed increased, C_1 quickly converged; the difference between the C_1 value at 5 m s^{-1} cloud speed and the C_1 value at 25 m s^{-1} cloud speed was only 0.6%.

This decrease in C_1 at slow cloud speeds is caused by limitation A. Fig. 4 shows how correlations change as a function of the angle between the pair of sites and the wind direction, and as a function of timescale. When the cloud speed is slow, the timescale at which limitation A has the largest impact is longer: the 32s timescale for 2 m s^{-1} cloud speed versus the 2s timescale for 25 m s^{-1} cloud speed. This is consistent with the claim in section 2.3 that limitation A has the strongest effect when $d_{m,n} \approx CS\,\bar{t}$, since the average distance between site pairs oriented in the along-wind direction is 66 m. When limitation A affects a longer timescale (i.e. for smaller cloud speed), it more strongly reduces the average correlations across the entire sensor network. At fast cloud speeds, limitation A only affects the shortest timescales when correlations would have been nearly zero anyway, and its effect in determining A values is negligible.

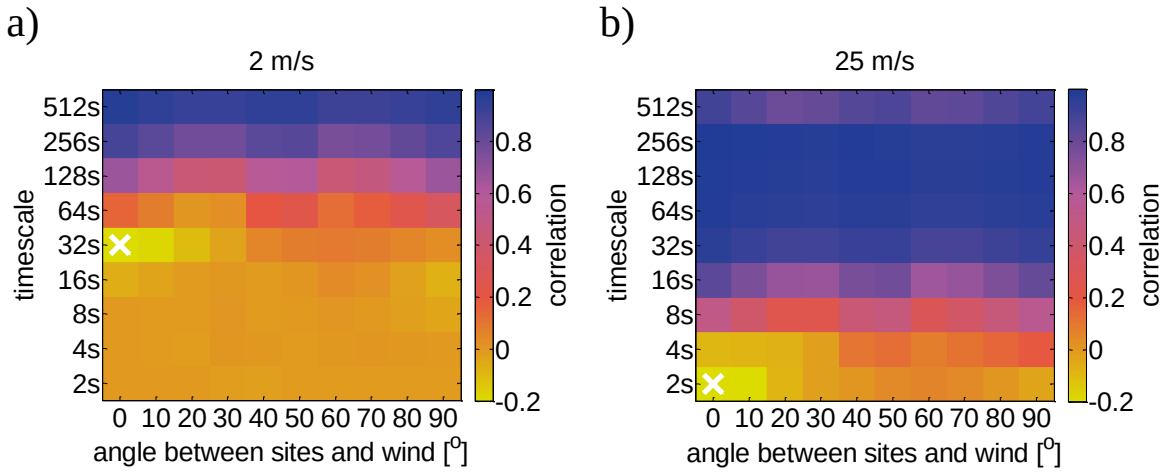

Fig. 4: Correlations from one run of the cloud simulator with 1000 m clouds and 0.7 cloud cover, averaged over all 10 sensor network offset positions. The plots show the dependence of correlations on the angle between the sites and the wind direction and on the timescale for (a) 2 m s^{-1} cloud speed and (b) 25 m s^{-1} cloud speed. The white x's mark locations of minimum correlation, showing where the effect of limitation A is strongest.

At small cloud sizes ($\frac{CS}{d_{m,n}}$<5; cloud size < 500m in Fig. 3a), C_1 increases significantly with cloud size due to limitation B. For small cloud sizes, some of the cross-wind distances between sites in the virtual sensor network are on the same order as the cloud size, and so correlations are smaller than at larger cloud sizes. Smaller correlations lead to smaller A values (Eq. 1), and hence a smaller C_1. As cloud size increases, though, the effect of limitation B vanishes, and C_1 becomes a constant.

Fig. 5 shows the effect of limitation B on the correlations. At 200 m cloud size, correlations in the along-wind direction (red dots in Fig. 5a) which limitation B has no effect on, are larger than correlations in all other directions. These along-wind correlations follow the scaling ($A = 6.5$) found for larger cloud sizes. Limitation B is responsible for the reduced correlations at all other site pair angles. As cloud size increases, though, the effect of limitation B goes away (Fig. 5b).

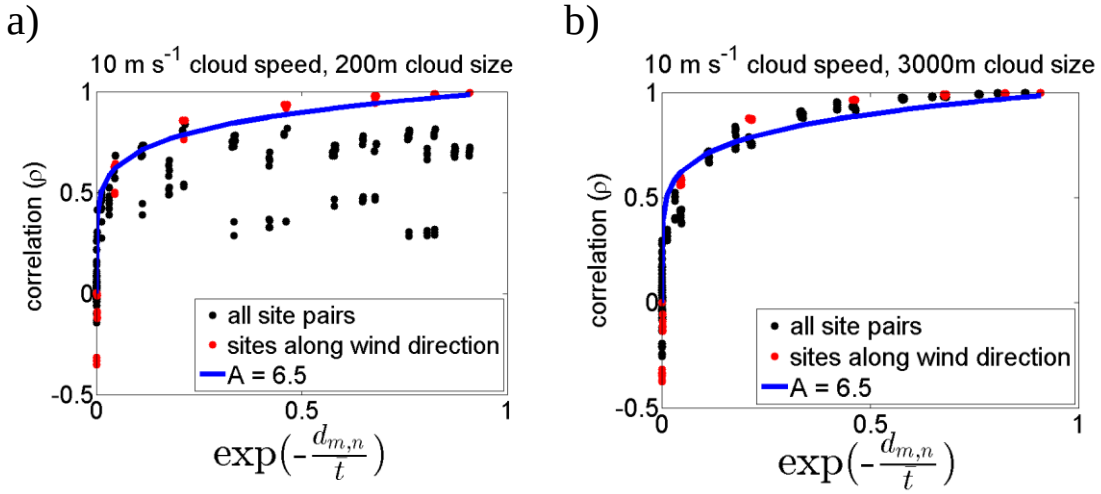

Fig. 5: Correlations from one run of the cloud simulator with 10 m s^{-1} cloud speed and 0.7 cloud fraction, averaged over all 10 sensor network offset positions, for (a) 200m cloud size and (b) 3000m cloud size. The black dots show all correlations, while the red dots show correlations between sites along the wind direction. The blue line is the best fit line $\boldsymbol{A = 6.5}$, which fits the 3000 m cloud size correlations well, and is included in (a) for reference.

For sufficiently fast cloud speeds and large cloud sizes (400-3000m cloud sizes and 5-25 m/s cloud speeds) $C_1 = 0.65$ resulted in the best fit to the data. Since the cloud simulator assumes uniform cloud optical depth in time and space $C_1 = 0.65$ is an upper limit. While we assume the linear relationship of Eq. 2 to hold, empirical evidence (e.g., Fig. 8a) suggests that $C_1 = 0.5$, and Eq. 2 becomes:

$$A = \frac{1}{2} * CS. \qquad 3$$

4. Determining Cloud Speeds and Converting to *A* values

Eq. 3 is only useful if cloud speeds are known. Unfortunately, cloud speeds are very difficult to measure. Atmospheric radiosonde measurements from the National Oceanic and Atmospheric Administration (NOAA) (Bosch, et al., 2013) can be used to determine cloud speeds, but locations are sparse and measurements are only taken at most twice per day. Satellite cloud speeds are often not representative of individual clouds but rather larger weather systems (Perez, 2012). Techniques for measuring cloud speed exist (Bosch, et al., 2013, Chow, et al., 2011), but only a limited amount of suitable sensors are deployed. Here, we use the NOAA North American Mesoscale (NAM) numerical weather forecast (Mathiesen and Kleissl, 2011), which has both a higher spatial (~12x12 km grids covering continental North America) and temporal (once per hour) resolution than radiosonde measurements.

4.1. Determining NAM Cloud Speed

NAM forecasts contain, among other variables, relative humidity and wind vectors at 39 different pressure levels. Cloud heights were determined from the relative humidity profiles (Fig. 6). We assumed that the dominant cloud layer was located at the height where the relative humidity was highest, provided that the relative humidity at that height was greater than 90%. A threshold lower than 100% was be used to resolve cloud layers that are smaller than the model vertical grid spacing. The wind speed at the cloud height was recorded as the cloud speed.

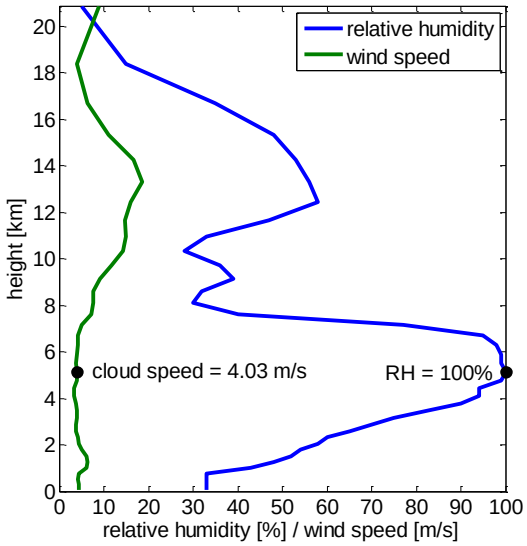

Fig. 6: Example plot of relative humidity and wind speed on July 26, 2011 at Copper Mountain, NV. In this case, the cloud was found to be at about 5km, with a speed of 4.03 m s^{-1}.

4.2. *A* Values for North America

Cloud speeds at each NAM grid point were computed every three hours. Although hourly data is available, using three hour data should be adequate for determining daily trends in cloud speeds and was a significant savings in both processing time and data storage requirements. The cloud speeds were converted into *A* values using Eq. 3. From the eight *A* values per day, we found the NAM GHI variance-weighted mean, which was recorded as the daily NAM *A* value. GHI was found hourly, and the variance of the clear-sky index during the 3-hour period containing the *A* value was used as the weight. This served to both eliminate nighttime *A* values which are meaningless in the context of solar radiation variability, and to place larger weight on *A* values during highly variable time periods that cause the largest ramp rates. Fig. 7 shows the seasonal averages of these daily *A* values. These seasonal averages were also found using a GHI daily variance-weighted mean.

We observe some interesting seasonal trends in Fig. 7. Nearly all locations within the NAM coverage area have small *A* values during the summer months (June-August). The other three seasons have areas with much larger *A* values, though the specific areas with large *A* values change by season. Large *A* values roughly follow the seasonal path of the Jet stream, since *A* values are derived from wind speeds. Some areas, such as coastal California and the Caribbean, have small *A* values ($A < 3$) year-round presumably due to low-level clouds in slower sea-breeze circulations. For otherwise identical conditions, small *A* values lead to small correlations and more smoothing across the plant. However, for the overall ramp rate distribution, the local variability in irradiance (i.e., the length of partly cloudy conditions), will usually dominate over smoothing due to a low *A* value.

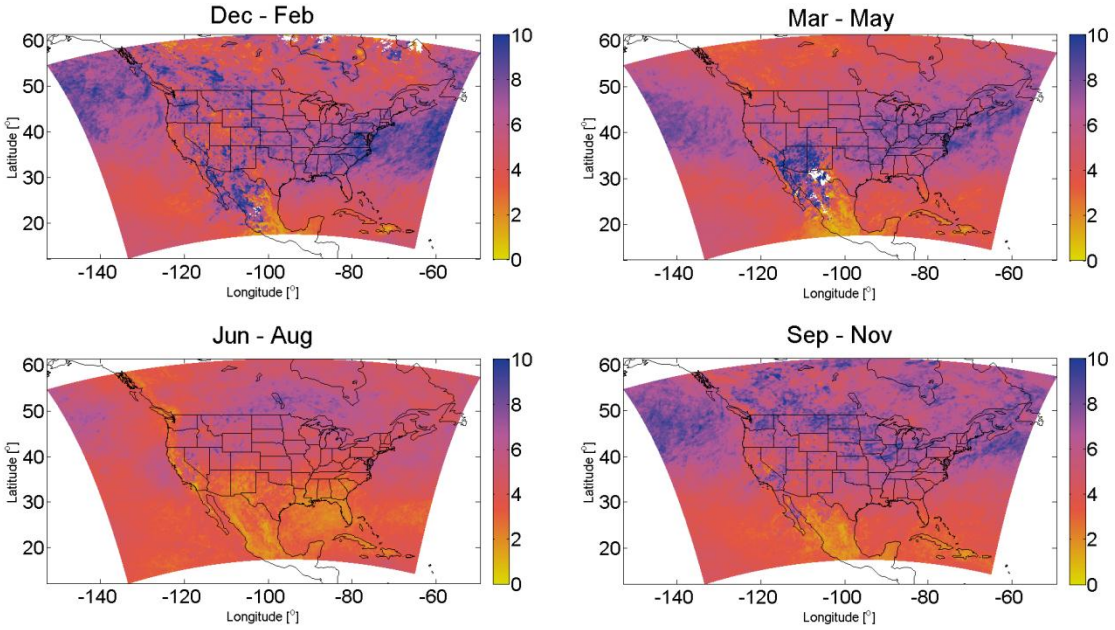

Fig. 7: GHI variance-weighted ***A*** values for the NAM coverage area by season.

4.3. Comparison of NAM *A* to Ground *A*

As a primary validation of the NAM *A* values, we compare NAM *A* values to ground *A* values from a network of irradiance sensors at the Copper Mountain site in southern Nevada (described in Section 5). Fig. 8 shows both a scatterplot and a comparison of cumulative distributions for the ground and NAM *A* values. These NAM *A* values were computed day-by-day for the time period during which CM reported data: August 2011 through July 2012. Since CM features an excellent solar resource, during this one year period there were only 40 days where clouds existed in the NAM, and hence there are 40 NAM *A* values to compare to ground *A* values. The scatterplot shows a high correlation ($\rho = 0.76$) between ground *A* and NAM *A* values. Fig. 8b shows that the cumulative distribution of NAM *A* values is similar to the cumulative distribution of actual *A* values

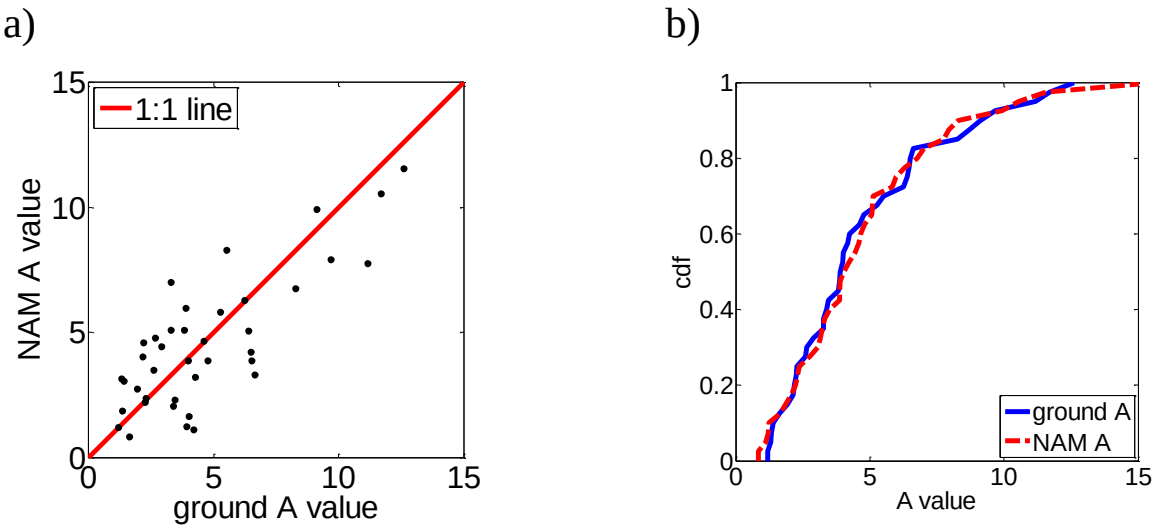

Fig. 8: Comparison of ground ***A*** and NAM ***A*** values during the time period August 1st, 2011 through July 31st, 2012. (a) scatter plot with 1:1 line, and (b) cumulative distribution functions (cdfs).

4.4. A Values for Any Day

The NAM does not simulate clouds at every site on every day, either because clouds did not exist or because the NAM failed to simulate them. On days when no clouds existed (and consequently no A value could be computed), both an irradiance point sensor and a powerplant will experience only minimal fluctuations away from their clear-sky profiles, and the A value is irrelevant. However, on days when clouds did exist but the NAM did not resolve them, it will be important to use a reasonable A value.

To allow for universal applicability, A values on any given day are determined by sampling from a distribution of NAM A values from the 90 days surrounding that day. We will refer to these as NAM-cdf A values to distinguish from the NAM A values found directly on the day of interest (Section 4.3).

To create NAM-cdf A values, cloud speeds for the year 2011 were processed. We assume that the year 2011 is representative of the seasonal distributions of A values, such that 2011 values may be applied to any other year. To allow for complete resolution of NAM-cdf A values, late 2011 values are included in the 90 day NAM-cdf window for early 2011 and vice versa. In this way, a complete, year-long record of NAM-cdf A values can be created for any site within the NAM coverage area. For example, to determine the NAM-cdf A value for February 14th 2012 in Copper Mountain, for example, we would sample from a distribution of the NAM A values in Copper Mountain from January 1st through March 31st 2011.

When using NAM-cdf A values to run the WVM, we expect larger errors on a day-by-day basis, due to the nature of A values being randomly selected from seasonal values. However, when daily results are aggregated over a season or a year, NAM-cdf WVM simulations should closely match distributions of seasonal or annual ramp rates in power output.

5. Application

In this section, we provide an example of how the WVM performs at simulating powerplant output given only a single irradiance sensor as input and using NAM-cdf A values. For this example, we use the 48MW Sempra US Gas & Power Copper Mountain (CM) utility scale PV power plant in Boulder City, NV. Irradiance measured once per second at an NREL calibrated plane-of-array reference cell was used as input to the WVM, and power output of the entire plant, also measured once per second, was compared to the output of the WVM. Additionally, CM contains a network of 15 reference cells, such that the ground A value can be determined through Eq. 1 as in Fig. 1. We analyze the year-long period of August 1st, 2011 through July 31st, 2012. 33 days from this period were eliminated due to errors in irradiance measurements, power measurements, or both. However, the 333 remaining days are well representative of annual trends.

5.1. Cumulative Distribution Functions of Ramp Rates

The WVM was run at CM for the 1-year period. The inputs to the WVM are: PV plant footprint, density of PV (in Watts of AC rated installed power per square meter), an irradiance point sensor timeseries, and a daily A value. The PV plant footprint and density of PV at CM are always fixed. The irradiance timeseries was from the same point sensor for all simulations. For each day, we ran three A values: the ground A value, the NAM-cdf A value, and $A = \infty$, which represents linearly scaling up a point sensor. This last scenario was included to show how the relative variability of the point sensor compared to the measured and simulated power outputs. Since the input irradiance timeseries was at 1-sec resolution, daily power output profiles at 1-sec resolution for each of the three scenarios were created. Yearly statistics are presented in Figs. 9 and 10.

As described in Section 4.4, A values were created based on 2011 NAM cdf; they were not NAM A values directly sampled on the days when CM recorded data. Better results would be obtained by using the NAM data from August 2011 through July 2012, but we chose to only use 2011 data to demonstrate

the ability of the NAM-cdf method to simulate a plant using the representative 2011 data when NAM data is not available during for the time of interest.

The goal of the WVM is to accurately simulate the variability of the actual plant power output. The exact timing of fluctuations will not be perfectly matched, because the point sensor will "see" clouds at different onset times than the total plant aggregate, but the statistics of variability should match. To test this, we use the cumulative distribution function (cdf) of RRs as a metric. Fig. 9 shows the large (>90th percentile) RRs of actual power output and the three WVM scenarios.

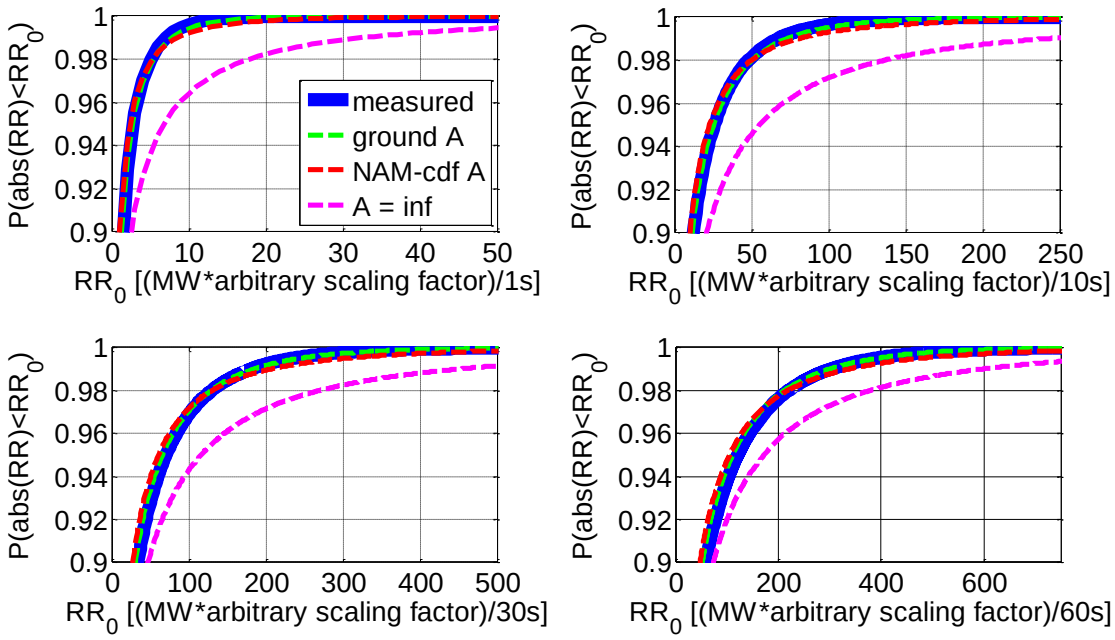

Fig. 9: Cumulative distribution of ramp rates in power output for the 1-year period from August 1, 2011 through July 31st, 2012. Ramp rates are shown at various timescales: 1s (top left), 10s (top right), 30s (bottom left), and 60s (bottom right). At each timescale, the ramp rates of measured power output (thick blue line), WVM run with ground *A* values (dashed green line), WVM run with NAM-cdf *A* values (dashed red line), and a point sensor with no smoothing (A = inf, dashed magenta line) are shown. The x-axis is the RR in [MW/timescale] multiplied by an arbitrary scaling factor to protect the confidentiality of the power data.

The cdfs of RRs match well between the measured power output and the ground *A* and NAM-cdf *A* WVM methods, as seen qualitatively in Fig. 9. Simply scaling up the point sensor ($A = \infty$) is inaccurate: Scaling up assumes that sites are always perfectly correlated, and so correlations and hence RRs will always be overestimated. The other two methods (ground and NAM-cdf) slightly overestimate the most extreme RRs (i.e., they are slightly shifted to the right in Fig. 9 for >98th percentile), meaning they slightly overestimate the correlations during the times when these RRs occur. Limitations A and B, both cause the WVM to overestimate correlation, and may explain part of the difference. Overall, though, errors in estimating the RR distributions of the actual power output are small when using either ground *A* or NAM-cdf *A* values.

5.2. Quantification of WVM Accuracy

To quantitatively compare the different methods, we find the sum of the squared errors, ω^2, between the cdfs of power RRs simulated with the various WVM methods and the cdfs of the measured power RRs using the Cramér–von Mises criterion:

$$\omega^2 = \int_0^\infty |F_{WVM\ method}(RR_0) - F_{measured}(RR_0)|^2 dF_{WVM\ method}(RR_0). \quad (4)$$

Values of ω^2 for each method at 1s, 10s, 30s, 1-min, 5-min, and 10-min are shown in Fig. 10. Due to different maximum RRs at each timescale, the Cramér–von Mises criterion is best used to compare errors between the different methods at the same timescale than to compare errors over different timescales (i.e., it is not normalized by timescale). The ground *A* WVM method is most accurate at simulating the RRs of measured power output at all timescales. The WVM using NAM-cdf *A*, though, causes only slightly larger errors. Scaling up the point sensor has the highest errors, especially at short timescales where the assumption that variability is the same at a point and over the whole powerplant is worst. At longer timescales (e.g., 10-min), errors between the three methods become comparable, as all PV modules within th powerplant have well-correlated output over long timescales, so the scaled point sensor becomes more accurate.

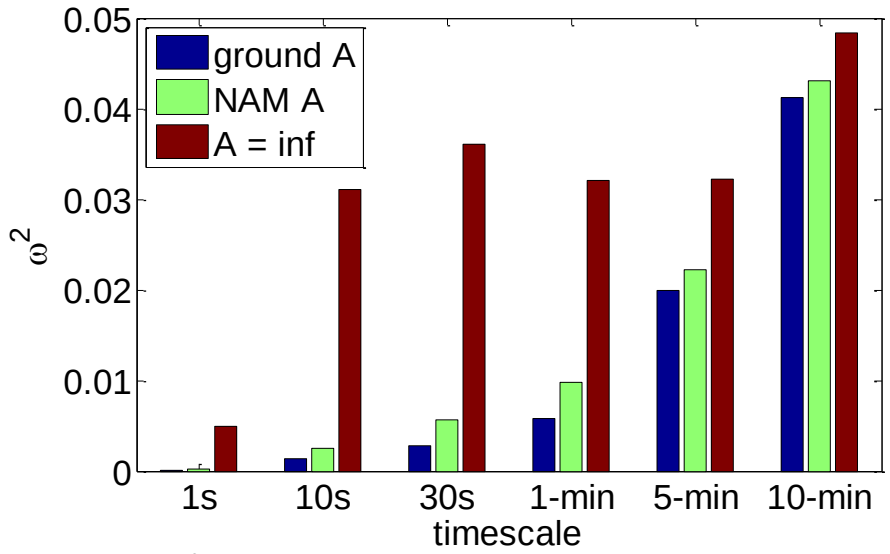

Fig. 10: Cramer-von Mises criterion (ω^2) showing the difference between the cumulative distribution of measured ramp rates and WVM ramp rates found using ground A values (blue), NAM A values (green), and the un-smoothed point sensor (A=inf, red).

6. Discussion and Conclusions

The Wavelet Variability Model (WVM) has shown great promise in simulating the output of PV powerplants. Previously, to apply the WVM an irradiance sensor network collocated with the PV plant to be simulated was required to determine the correlation scaling coefficient (A). In this work, we demonstrated that A is directly related to cloud speeds and present a method for determining cloud speeds from numerical weather forecasts. This allows for much broader application of the WVM.

By employing a cloud simulator, the dependence of correlation on cloud speeds, in addition to distance and timescale, was motivated. The correlation equation used in the WVM was reduced to: $\rho = \exp\left(-\frac{d_{m,n}}{\frac{1}{2} CS\, \bar{t}}\right)$. This is similar to the fractional equation $\rho = \frac{1}{1+\frac{d_{m,n}}{CS\, \bar{t}}}$ proposed in Hoff and Perez (2012) and is the same except for coefficient differences as the exponential equation $\rho = \exp\left(\frac{\ln(0.2)\, d_{m,n}}{1.5\, CS\, \bar{t}}\right)$ proposed by Perez, et al. (2011). All of these correlation models predict nearly zero correlation at very long distances or very short timescales, and nearly perfect correlation at short distances or long timescales, as expected physically. In between these two limits, the differences between the models are small; the Lave and Kleissl model presented here predicts slightly lower correlations between sites than the other two models, as shown in Fig. 11. In both Hoff and Perez (2012) and Perez, et al. (2011), distances between sites were on the scale of kilometers and timescales were on the order of hours. In this work, distances were on the scale of meters and timescales on the order of seconds, yet very

similar correlation models were found. The agreement in correlation models shows that the nondimensional quantity $\frac{d_{m,n}}{CS\,t}$ is the dominant quantity for determining correlations between sites, regardless of the magnitude of the distance or timescale being considered. In other words, the process is scale invariant.

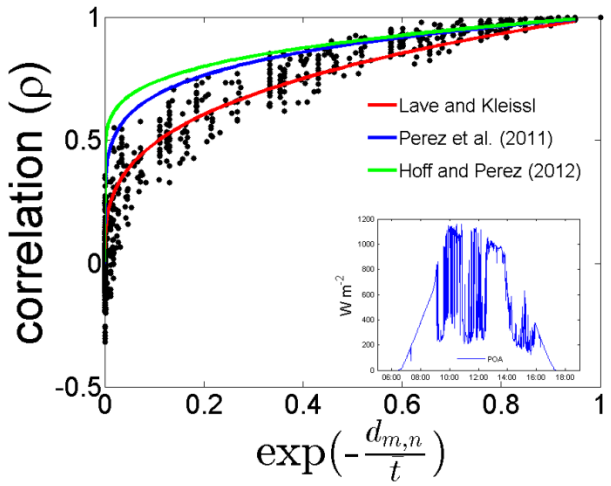

Fig. 11: Repeat of Fig. 1 with the Hoff and Perez (2012) and Perez et al. (2011) models added to the plot. The cloud speed on this day was determined from the ***A***-value to be 6.38 m s^{-1}.

These correlation models assume correlations are isotropic: they do not depend on direction between sites. In practice, though, there is a small directional dependence to correlation between two sites, which leads to slight limitations of isotropic correlation models. Hinkelman, et al. (2011) showed that correlations can vary depending on the orientation of sites relative to the cloud motion direction. Negative correlations were observed in the along-wind direction at certain timescales both in Hinkelman, et al. (2011) and in the cloud simulator used in this work. In the cloud simulator, we also noticed a decorrelation in the cross-wind direction when the distance between sites was comparable to the cloud size. For WVM applications, we assume that cloud speeds and sizes will be large enough that the effect of these limitations will be small. Results from the cross-wind correlations in Hinkelman, et al. (2011) support the assumption that cloud sizes are large relative to distances between sites, as no sharp drop off of correlations – which would occur if sites were further apart than the cloud size such that they were never covered by the same cloud – is seen in the cross-wind direction.

The use of cloud speeds to eliminate the need for a local sensor network when running the WVM opens many more possible applications. Solar developers who have high frequency irradiance point measurements on site can use the WVM to estimate the RRs that will occur at the plant. Module siting, plant sizing, and forecasting and storage requirements can be simulated before the plant is installed. This will be especially important for PV plants installed in locations (typically islands) that have RR restrictions (e.g., Puerto Rico).

Acknowledgment

We appreciate funding from the DOE High Solar PV Penetration grant 10DE-EE002055. We also appreciate the help of David Jeon, Leslie Padilla, and Shiva Bahuman from Sempra US Gas and Power and Bryan Urquhart from UCSD for providing and supporting the Copper Mountain data.

References

Bosch, J. L., Zheng, Y. , Kleissl, J., 2013. Deriving cloud velocity from an array of solar radiation measurements. Solar Energy. 87, 196-203.

Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J. , Washom, B., 2011. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. Solar Energy. 85, 2881-2893.

Curtright, A. E. , Apt, J., 2008. The character of power output from utility-scale photovoltaic systems. Progress in Photovoltaics: Research and Applications. 16, 241-247.

Hinkelman, L., George, R. , Sengupta, M., 2011. Differences between Along-Wind and Cross-Wind Solar Variability. American Solar Energy Society Conference.

Hoff, T. E. , Perez, R., 2010. Quantifying PV power Output Variability. Solar Energy. 84, 1782-1793.

Hoff, T. E. , Perez, R., 2012. Modeling PV fleet output variability. Solar Energy. 86, 2177-2189.

Kawasaki, N., Oozeki, T., Otani, K. , Kurokawa, K., 2006. An evaluation method of the fluctuation characteristics of photovoltaic systems by using frequency analysis. Solar Energy Materials and Solar Cells. 90, 3356-3363.

Lave, M. , Kleissl, J., 2010. Solar variability of four sites across the state of Colorado. Renewable Energy. 35, 2867-2873.

Lave, M., Kleissl, J. , Arias-Castro, E., 2012. High-frequency irradiance fluctuations and geographic smoothing. Solar Energy. 86, 2190-2199.

Lave, M., Kleissl, J. , Stein, J. S., 2012. A wavelet-based variability model (WVM) for solar PV power plants. Sustainable Energy, IEEE Transactions on. PP, 1-9.

Marcos, J., Marroyo, L., Lorenzo, E., Alvira, D. , Izco, E., 2011. From irradiance to output power fluctuations: the pv plant as a low pass filter. Progress in Photovoltaics. 19, 505-510.

Mathiesen, P. , Kleissl, J., 2011. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. Solar Energy. 85, 967-977.

Mills, A. , Wiser, R., 2010. Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power.

Otani, K., Minowa, J. , Kurokawa, K., 1997. Study on areal solar irradiance for analyzing areally-totalized PV systems. Solar Energy Materials and Solar Cells. 47, 281-288.

Perez, R., Kivalov, S. , Hoff, T. E., 2011. Spatial & Temporal Characteristics of Solar Radiation Variability. Proc. of International Solar Energy World Congress, Kassel, Germany.

Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr, K. , Hoff, T., 2011. Parameterization of site-specific short-term irradiance variability. Solar Energy. 85, 1343-1353.

Perez, R., 2012. personal communication.

Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr, K. , Hoff, T. E., 2012. Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance. Solar Energy. 86, 2170-2176.

PREPA. Puerto Rico Electric Power Authority Minimum Technical Requirements for Photovoltaic Generation (PV) Projects 2012.

Wiemken, E., Beyer, H. G., Heydenreich, W. , Kiefer, K., 2001. 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, 513-518.

Woyte, A., Belmans, R. , Nijs, J., 2007. Fluctuations in instantaneous clearness index: Analysis and statistics. Solar Energy. 81, 195-206.