UNIVERSITY OF CALIFORNIA, SAN DIEGO

Analyzing and Simulating the Variability of Solar Irradiance and Solar PV Powerplants

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Engineering Sciences (Aerospace Engineering)

by

Matthew S. Lave

Committee in charge:

Professor Jan Kleissl, Chair
Professor Ery Arias-Castro
Professor Raymond de Callafon
Professor Joseph Ford
Professor Kraig Winters

2012

The Dissertation of Matthew S. Lave is approved, and it is acceptable in quality and form of publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2012

EPIGRAPH

We are like tenant farmers chopping down the fence around our house for fuel when we should be using Nature's inexhaustible sources of energy – sun, wind and tide. ... I'd put my money on the sun and solar energy. What a source of power! I hope we don't have to wait until oil and coal run out before we tackle that.

Thomas Edison, 1931

I have no doubt that we will be successful in harnessing the sun's energy. If sunbeams were weapons of war, we would have had solar energy centuries ago.

George Porter (Nobel Prize winner in Chemistry), 1973

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I very much appreciate all that Professor Jan Kleissl has done to help me through my work. His guidance and feedback have been extremely valuable, and I can't image a better PhD advisor. I also appreciate all the help and guidance that I have gotten from co-workers at Sandia National Laboratories. Especially, Joshua Stein, Cliff Hansen, Abraham Ellis, and Matthew Reno have been very helpful in providing feedback, suggestions, and promoting my work. Finally, fellow graduate students here at UCSD have been excellent as both academic and social companions. They have been invaluable resources both in discussing my work and in sharing their work to broaden my understanding of solar energy and meteorology topics.

Chapter 2, in full, is a reprint of the material as it appears in *Renewable Energy* volume 35, issue 12, December, 2010. Lave, Matthew; Kleissl, Jan, 2010. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in *Renewable Energy* volume 36, issue 3, March, 2011. Lave, Matthew; Kleissl, Jan, 2011. The dissertation author was the primary investigator and author of this paper.

Chapter 4, in full, is a reprint of the material as it appears in *Solar Energy* volume 86, issue 8, August 2012. Lave, Matthew; Kleissl, Jan; Arias-Castro, Ery, 2012. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, is a reprint of the material as it appears in the *IEEE Transactions on Sustainable Energy* (preprint) 2012. Lave, Matthew; Kleissl, Jan; Stein, Joshua, 2012. The dissertation author was the primary investigator and author of this paper.

Chapter 6, in full, is a reprint of the material as has been submitted to *Solar Energy* 2012. Lave, Matthew; Kleissl, Jan, 2012. The dissertation author was the primary investigator and author of this paper.

VITA

2008	Bachelor of Arts in Physics, Occidental College
2009-2012	Research Assistant, University of California, San Diego
2010	Master of Science, University of California San Diego
2011-2012	Graduate Student Intern, Sandia National Laboratories
2012	Doctor of Philosophy, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Analyzing and Simulating the Variability of Solar Irradiance and Solar PV Powerplants

by

Matthew S. Lave

Doctor of Philosophy in Engineering Sciences (Aerospace Engineering)

University of California, San Diego, 2012

Professor Jan Kleissl, Chair

Solar energy is a promising renewable resource, but the variability of solar photovoltaic (PV) power due to cloud-caused fluctuations is a concern for electric grid operators. Fortunately, though, the relative variability of a PV powerplant will be reduced due to the geographic diversity within the plant. The amount of the variability reduction (VR) will depend on the geometry of the plant, the timescale of interest, and the local meteorological conditions. This work focuses on quantifying and modeling the VR in scaling up from an irradiance point sensor to a PV powerplant. The dependence of VR on timescale is demonstrated using high-frequency data and the wavelet transform. The wavelet variability model (WVM), which simulates the variability of a PV powerplant by estimating the VR at each timescale, is developed and validated. Inputs to the WVM are a point sensor irradiance timeseries, the powerplant layout, and the cloud speed. As an example application, the WVM is used to simulate the numbers of ramps larger than 10% of capacity per minute at various sizes of PV powerplants in Puerto Rico.

.

1. Introduction

1.1. **Motivation**

The sun is a colossal energy source. In one hour, enough energy is incident on the Earth though solar radiation as is consumed by humans in one year [1]. Capturing even a small fraction of this solar radiation can lead to substantial energy production. Solar photovoltaic (PV) modules are a simple way of converting solar radiation into electricity. Other than cost, which is being lowered every year and is expected to reach $1/watt by 2020 [2], the main downside to PV compared to conventional electric generation methods (such as coal, nuclear, or natural gas powerplants) is the variable nature of the power output. When clouds pass over a PV module, they temporarily reduce the power output. On longer timescales, additional variability is caused due to the sun's movement through the sky: maximum power is produced at solar noon, and no power is produced at night.

Cloud-caused variability is especially an issue for grid-connected solar PV powerplants as unanticipated changes in PV plant power output can strain the electric grid. At short timescales (seconds), sharp changes in power output from a PV powerplant can cause local voltage flicker issues. At longer timescales (minutes), producing less PV power than expected can cause balancing issues, where load can exceed generation. PV powerplant variability can be counteracted by other, fast ramping generation sources (e.g., gas turbines) and by storage systems (e.g., batteries), but both are quite expensive and substantially increase the cost of PV powerplants. Fortunately, though, geographic diversity leads to a reduction in the variability of PV powerplant output. The amount of the reduction in variability due to geographic smoothing depends on plant layout, timescale of interest, and meteorological conditions, but can be quite significant.

1.2. **Background**

Solar radiation reaching the Earth's surface can be broken down into 3 components: direct or beam radiation (B) which is radiation directly from the sun, diffuse radiation (D) which is scattered by clouds and atmospheric gases and particles, and reflected radiation (R_g) which is reflected off of the ground. The sum of all three forms of radiation is the global irradiance, $GI = B + D + R_g$. GI is a function of tilt and azimuth angle, and so GI reaching a PV panel or irradiance sensor depends on its orientation. GI is typically measured in units of W m^{-2}. If an irradiance sensor is mounted horizontal, then it is said to be measuring global horizontal irradiance (GHI). If the sensor is tilted, then it is said to be measuring plane of array (POA) irradiance. Typically, POA irradiance measurements exist at solar PV powerplants and at the same orientation as the PV modules.

For variability applications, the power output of the areal aggregate of an entire PV plant, $P(t)$, is the quantity of interest. Power output measurements are difficult to obtain because most PV plant operators consider this information proprietary. Irradiance measurements are much more common and easily available. Power output of PV modules can be approximated from the GI incident on those modules. POA irradiance measured at the 1.2 MW PV plant in Lanai, HI was found to be nearly linearly proportional to power output [3]. The slight non-linearity is mostly due to temperature effects, as conversion efficiency slightly decreases with increasing temperature: approximately 0.5% per ºC change in temperature for silicon modules. Highly detailed irradiance to power models which account for all the non-linear effects (such as [4]) can be used, but are generally unnecessary for variability applications. At timescales of interest to variability studies, changes in power output are dominated by changes in irradiance, and hence the variability of solar irradiance can be used as a proxy for the variability of solar power.

Inherent in all irradiance and power output timeseries are seasonal and diurnal cycles relating to changes in the Earth-Sun position and atmospheric effects. However, these occur over

long timescales and can be counteracted by slower ramping and hence cheaper ancillary power sources. In order to isolate short term, cloud-caused effects, a normalized irradiance called the clear-sky index (kt) is used. kt is the GI (either GHI or POA) divided by the clear-sky GI expected at the Earth's surface, taking into account the climatological atmospheric transmissivity and turbidity. Clear-sky values can exceed 1 due to cloud enhancement (Fig. 1.1). The clear-sky index is essential when applying the wavelet transform (Chapter 4), and is a non-dimensional value that allows for easy comparison and conversion between irradiance and power (Chapter 5).

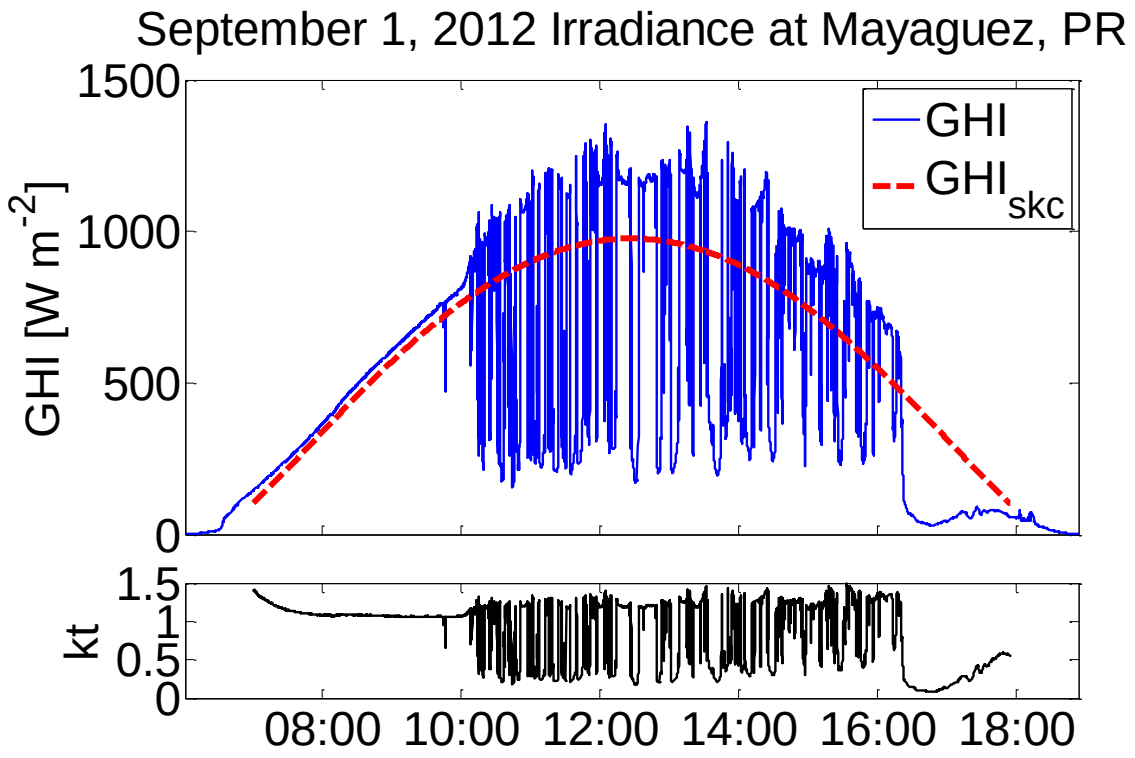

Fig. 1.1: Measured GHI and clear-sky modeled GHI_{skc} (top) and clear-sky index kt (bottom) on September 1, 2012 at the University of Puerto Rico, Mayaguez. The clear-sky GHI and clear-sky index were only calculated for times when the sun was more than 10° above the horizon.

For quantifying solar variability, ramp rate (RR) statistics are the most common and practically relevant quantities. RRs are of interest to PV plant and electric grid operators as extreme changes in power output impact electric grid operations disproportionately. RRs are

calculated by differencing values of a solar timeseries and dividing by the timescale. For example, GHI RRs are:

$$\mathrm{RR}_{\mathrm{GHI}}^{\Delta t}(t) = \frac{1}{\Delta t}\left(\sum_{t}^{t+\Delta t}\mathrm{GHI} - \sum_{t-\Delta t}^{t}\mathrm{GHI}\right). \quad 1.1$$

The timestep, Δt, is important to define. RRs calculated at short timesteps will, on average, be smaller than RRs at longer timesteps, as they have had less time to deviate from the previous value. Many works have presented RRs at 5-minute, 15-minute, or 1-hour timesteps, often because those were the highest temporal resolution data available at the time (e.g., Chapter 2). Recently, higher-resolution data become available and showed that 1-second RRs will have quite different statistics than 1-minute or 1-hour RRs (Chapter 4).

RRs are also often calculated for the power output of a PV system. We substitute in power, P, for GHI in Eq. 1.1 to calculate RRs of power output:

$$\mathrm{RR}_{\mathrm{P}}^{\Delta t}(t) = \frac{1}{\Delta t}\left(\sum_{t}^{t+\Delta t}\mathrm{P} - \sum_{t-\Delta t}^{t}\mathrm{P}\right). \quad 1.2$$

For comparison between PV systems, power RRs can be normalized by the capacity of installed PV, creating relative RRs (rRRs) that are the changes relative to fraction of installed capacity:

$$rRR_P^{\Delta t} = \frac{RR_P^{\Delta t}}{capacity\ of\ installed\ PV}. \quad 1.3$$

Due to geographic smoothing, rRRs are usually smaller at larger PV plants than at small ones. However, RRs are almost always larger for large PV plants, even if they benefit from a greater amount of geographic smoothing than smaller plants. For example, a 5MW plant experiencing a 1MW RR and a 10MW experiencing a 2MW RR will both have the same $rRR = 0.1$. For a further discussion of rRRs, see [5].

To quantify the reduction in variability achieved over a PV system, we define the variability reduction (VR) as the ratio of variability at a point to variability of the aggregate PV system:

$$VR(\bar{t}) = \frac{variability_{point\ sensor}}{variability_{PV\ system}}. \quad 1.4$$

Variability in this case can refer to any relative variability metric: maximum rRRs, average rRRs, or wavelet power contents. Since the variability at the point sensor will always be equal to or greater than the variability of the aggregate system, $VR \geq 1$. Larger values of VR indicate more geographic smoothing. At short timescales, we expect VR to be large, as short-timescale fluctuations are strongly damped by geographic smoothing. At longer timescales, though, VR will approach 1. On fully clear days, the VR has no meaning since both the point sensor and the whole PV system will have a clear, smooth profile.

1.3. **Summary and Discussion of Chapters**

The chapters in this thesis focus on examining and modeling the variability reduction (VR) in PV powerplants due to geographic smoothing. The chapters follow the progression of the PhD research: early chapters focus on analyzing and modeling factors that affect solar variability while the later chapters present a model for simulating PV powerplant variability.

In Chapter 2, the VR due to geographic smoothing is quantified through the aggregation of 5-minute global horizontal irradiance (GHI) data of four sites spread across the state of Colorado. This chapter presented a suite of tools required to analyze solar radiation timeseries. It focuses on ramp rate (RR) statistics and the geographic smoothing of four sites spread far apart (100s of km). Coherence spectra are presented and allow for the estimation of the timescale of decorrelation, and Fourier decompositions show the decrease in high-frequency power content for the average of all four sites. While this paper serves as solid evidence of geographic smoothing, the analysis tools were improved in future works. Most importantly, the wavelet decomposition was adopted (e.g., in Chapter 4) in place of the Fourier decomposition and coherence spectra. The

wavelet decomposition allows for temporal resolution of fluctuations and a wavelet shape can be chosen that better fits solar fluctuations than the sine wave used in the Fourier decomposition. Future works focused on powerplant or feeder-scale variability (a few kilometers), and shorter timescales (1-second), which are more applicable to current electric grid concerns.

Chapter 3 models the relationship between GHI and plane of array (POA) irradiance for a tilted PV module. This relationship is important when simulating the power output of PV powerplants, and is used in the wavelet variability model (Chapter 5). Power produced by a PV module is almost linearly proportional to the GI incident on that module. PV powerplants almost always have south-tilted PV modules and sometimes have tracking modules, yet irradiance sensors are often mounted horizontal and measure GHI. To simulate power output of a PV module, these GHI measurements much be converted to POA using the model described in this chapter. An additional interesting use of this GHI to POA model is that it can be used to determine the optimum orientation of a PV module by comparing the incident irradiance at all possible orientations. These optimal orientations, as well as the benefits of tracking over optimal orientation, are shown in the figures of Chapter 3.

Wavelet analysis is applied to high-frequency irradiance data in Chapter 4. The top hat wavelet is chosen because its shape closely resembles a fluctuation in solar irradiance due to a cloud passing over. 1-second GHI data measured on the UCSD campus is used to compute the power content of wavelet fluctuations at each timescale. While this is fundamentally similar to the Fourier transform, the wavelet transform is both a better fit to solar fluctuations due to the top hat shape and allows for a temporal resolution of fluctuations (i.e., when they occur during the day). The power contents show that the VR is a function of timescale, which is especially important at short timescales where the VR can change significantly. The command of the wavelet transform for solar variability applications developed in this chapter served as a major stepping stone towards creation of the wavelet variability model presented in Chapter 5.

Chapter 5 presents the wavelet variability model (WVM), which directly addresses the issue of estimating the impact of PV powerplants to the electric grid. The WVM is a model for simulating solar PV powerplant output given a single irradiance point sensor as input. The WVM takes advantage of the temporal resolution of wavelet fluctuations seen in Chapter 4. The wavelet transform of the input irradiance point sensor is found, and then scaled by the VR. Through an inverse wavelet transform, the PV powerplant output is simulated. The VR is modeled using an equation that estimates correlations between pairs of PV modules in the powerplant, taking into account the distance between the modules, the timescale, and the meteorological conditions. The WVM is advanced compared to most other models in its ability to account for different smoothing at different timescales (through use of the wavelet transform) and in its ability to account for varying meteorological conditions day-by-day. In Chapter 6 the meteorological impacts are further studied, and it is found that cloud speed is the dominant daily-changing factor in scaling the VR over a powerplant.

Chapter 6 builds upon the WVM presented in Chapter 5 by adding cloud speeds into the model. The WVM as described in Chapter 5 required an irradiance sensor network to resolve the daily meteorological conditions at the PV powerplant location to be simulated, but the cloud speed method eliminates that requirement. A simple cloud simulator is used to determine the effect of cloud size and cloud speed and to further test the correlation model used in the WVM. Cloud speed is found to be the dominant scaling factor, and cloud size does not have a noticeable influence except in extreme cases of very small clouds. By coupling cloud speeds obtained from numerical weather prediction with the WVM, the WVM can now be run for any location within the numerical weather forecast domain with only a single irradiance point sensor as input (no further meteorological conditions are required). This should have a profound impact in making the WVM easier to use.

The WVM has already found many applications. Through an internship with Sandia National Laboratories, the WVM has been to create realistic inputs to grid integration studies, mostly focusing on feeder-level effects. For example, the impact of adding PV to three different feeders in Utah was studied to determine the effect on voltage and tap changer use [6, 7], and the WVM was used to convert the available point sensor irradiance into simulated PV plant output. The WVM has also been used to simulate yet-to-be built PV plants for an estimation of their variability to allow for proper storage procurement and compliance with variability requirements. This has been especially important in Puerto Rico, where the Puerto Rico Electric Power Authority (PREPA) has instituted a requirement that all PV plants on the island must not produce changes in power output (RRs) greater than 10% per minute. With only PV plant layout, power capacity, and an input irradiance timeseries as inputs, the WVM can be used to estimate the number of times per day that 10% per minute RRs will occur. This can be of great help for prospective PV plants.

References

[1] O. Morton, Solar energy: A new day dawning?: Silicon Valley sunrise, Nature, 443 (2006) 19-22.

[2] SunShot Vision Study, United States Department of Energy, http://www1.eere.energy.gov/solar/sunshot/vision_study.html, 2012.

[3] C. Hansen, J. Stein, A. Ellis, Simulation of One-Minute Power Output from Utility-Scale Photovoltaic Generation System, 2011.

[4] D. King, W. Boyson, J. Kratochvil, Photovoltaic Array Performance Model, 2004.

[5] M. Lave, J. Stein, A. Ellis, Analyzing and Simulating the Reduction in PV Powerplant Variability due to Geographic Smoothing in Ota City, Japan and Alamos, CO, Photovoltaics Specialists Conference, Austin, TX, 2012.

[6] M.J. Reno, A. Ellis, J. Quiroz, S. Grijalva, Modeling Distribution System Impacts of Solar Variabiltiy and Interconnection Locations, World Renewable Energy Forum, Denver, CO, 2012.

[7] J. Quiroz, M. Reno, Detailed Grid Integration Analysis of Distributed PV, American Solar Energy Society, Denver, CO, 2012.

2. Solar Variability of Four Sites across the State of Colorado

Reprinted from *Renewable Energy*, Volume 35, Issue 12, Matthew Lave, Jan Kleissl, "Solar variability of four sites across the state of Colorado," Copyright 2010, with permission from Elsevier

Abstract

Solar Global Horizontal Irradiance (GHI) fluctuates on both short (seconds to hours) and long (days to months) time scales leading to variability of power produced by solar photovoltaic (PV) systems. Under a high PV penetration scenario, fluctuations on short time scales may require a supplementary spinning power source that can be ramped quickly, adding significant external cost to PV operation. In order to examine the smoothing effect of geographically distributed PV sites, GHI timeseries at 5 minute resolution at four sites across the state of Colorado were analyzed. GHI at the four sites was found to be correlated due to synchronous changes in the solar zenith angle. However, coherence analysis showed that the sites became uncorrelated on time scales shorter than 3 hours, resulting in smoother average output at short time scales. Likewise, extreme ramp rates were eliminated and the spread in ramp rate magnitude was significantly reduced when all four sites were averaged. Nevertheless, even for the averaged output, high frequency fluctuations in PV power output are relatively larger in magnitude than fluctuations expected from wind turbines. Our results allow estimation of the ancillary services required to operate distributed PV sites.

2.1. **Introduction**

Over the past few years, there has been an increasing interest in harnessing renewable energy sources such as wind and solar power as a supplement to, or replacement for, current carbon-based power sources. However, at high grid penetration, the variability of these renewable sources has the potential to affect grid reliability and energy cost. Wind power has thus far been the more popular technology for large scale implementation, with about 121 GW of wind power

installed across the world at the end of 2008 [1]. In areas where a large percentage of the power is provided by wind, fast-ramping of other power sources has been used to counteract wind variability [2].

In comparison to wind power, installed solar photovoltaic (PV) power capacity was relatively small, with only about 13 GW at the end of 2008 [1]. However, PV installation has increased rapidly over the past few years. Since 2002, PV power capacity has increased 48% per year, on average, and is expected to continue to be the fastest growing energy technology in the world [3]. Since PV is growing so quickly, it is pertinent to study high-penetration scenarios.

Geographically dispersing wind power sites is an effective way of reducing wind variability, as power production at different sites typically becomes uncorrelated over a few 100 km [4-6]. In Northern Europe wind power supply from sites more than 1500 km apart is uncorrelated [7]. When aggregated, the output of 1496 widely spread wind turbines in Germany showed maximum variations of 60% in 4 hours. Similarly, one would expect that geographic dispersion of solar energy production sites could mitigate solar variability caused by atmospheric transmissivity changes in short timescales (clouds), while being largely ineffective in mitigation of the day-night-solar variability.

Analyzing a month of 1 minute radiation data from 11 sites over 75 x 75 km in Wisconsin (a mid-latitude frontal weather regime), Long and Ackerman (1995) determined the correlation of Global Horizontal Irradiance (GHI) and GHI normalized by clear sky radiation [8]. As expected, the correlation coefficients were smaller for the normalized value, as the synchronized occurrence of rising and setting sun at all stations contributes significantly to a high correlation. Large day-to-day differences in correlations were observed indicating limitations for average statistics in describing or modeling insolation. Moreover, for individual days – especially overcast days – there was significant scatter in the correlation versus distance plots for all stations pairs indicating that atmospheric transmissivity is not an isotropic process. Barnett et al. (1998)

used Oklahoma Mesonet data from 111 GHI sensors to define spatial correlograms [9]. Subtracting out the diurnal signal, they found that characteristic length and time scales (i.e. the distances and time differences at which correlation goes to zero) were 300 km and 60 minutes, respectively. Curtright and Apt (2008) examined three PV sites spread across hundreds of kilometers in the state of Arizona and found a reduction in average 10 minute step size magnitude and in standard deviation for the sum of all three sites [10]. However, they also found that short timescale variability of large-scale PV power was still significant and that the geographical diversity did not dampen PV variability enough to eliminate the need for substantial supplemental power sources. Wiemken et al. (2001) studied 100 PV systems spread across Germany, and also found a decrease in average step size magnitude and standard deviation for the sum of all systems, but did not present timescale variability analysis [11].

In this paper we study the variability of measured GHI at four different sites across the state of Colorado. This choice is motivated by the fact that the greatest spatial density of 1 to 5 minute resolution irradiance data exists in Colorado. Topographical and meteorological differences between Wisconsin, Oklahoma, Arizona and Colorado also warrant the analysis of variability over different regions. The sharper terrain difference across these Colorado sites may lead to more varied weather patterns and increased geographic variation. Moreover, we extend the existing literature by analyzing shorter time scales (5 minute) and examining coherence between the sites and its effect on smoothing average output at different time scales through spectral analysis.

2.2. **Data**

While ultimately PV array power output is the relevant variable for variability analysis, 90% of the variability in PV output is explained by variability in GHI. Consequently, here we assume that solar radiation is proportional to PV power output and we use radiant flux density (W m^{-2}) rather than power for this analysis, neglecting the influence of PV panel temperature on

panel efficiency. Furthermore, while variability analysis of PV output is more practically relevant, these studies are not as representative, since they depend on the system specifications, and there is generally less publicly available time-resolved data for PV output.

GHI data were obtained from the National Renewable Energy Laboratory's (NREL) Measurement and Instrumentation Data Center (MIDC) [12]. Sites were chosen which were within a few hundred kilometers of one another such that they would typically feed into one utility grid, and had complete data for Jan 1, 2008 – Dec 31, 2008. Four sites fit these criteria: the National Wind Technology Center (NWTC), the NREL Solar Radiation Research Laboratory (SRRL), the South Park Mountain Data (SPMD), and the Xcel Energy Comache Station (XCEL, Fig. 2.1). The distances between sites are shown in Table 2.1. The NWTC site has an Eppley Laboratory, Inc. Precision Spectral Pyranometer, SRRL data were collected using a Kipp and Zonen CMP 22 pyranometer, and the XCEL site uses a LICOR LI-200 silicon Pyranometer. All sites had data at 1 minute resolution except for SPMD where a LI-200 was operated at five minute temporal resolution.

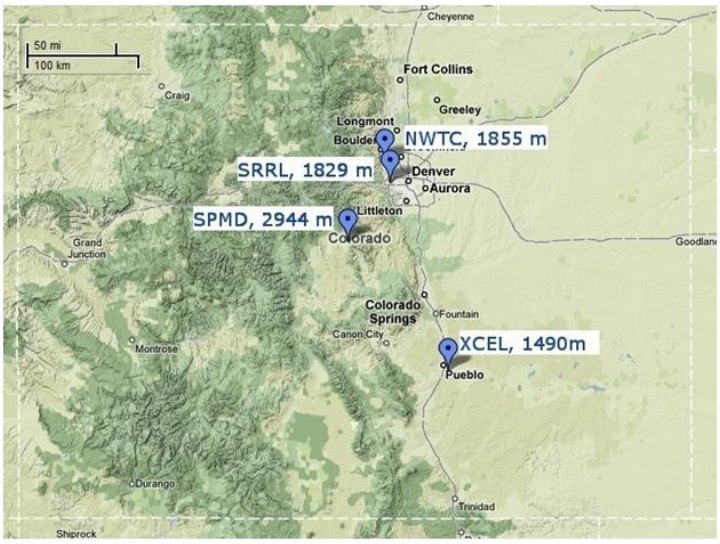

Fig. 2.1: The sites used for this study on a terrain map with elevations in meters: National Wind Technology Center (NWTC), Solar Radiation Research Laboratory (SRRL), South Park Mountain Data (SPMD), and Xcel Energy Comanche Station (XCEL). Map © 2010 Google - Map Data © 2010 Google

Table 2.1: Distance between sites.

From	To	Distance
NWTC	SRRL	19 km
NWTC	SPMD	78 km
NWTC	XCEL	197 km
SRRL	SPMD	65 km
SRRL	XCEL	178 km
SPMD	XCEL	149 km

Although a greater geographical wealth of solar radiation data is available through NREL's National Solar Radiation Database (NSRDB), these data are only recorded once per hour, and much of it is not based on GHI measurements. Using one hour resolution, even if it is an average of data collected at shorter intervals, will filter out the shorter time-scale intermittencies that produce the largest ramp rates (RRs). For example, large RRs caused by clouds occur on scales of seconds to ten minutes. Fig. 2.2 illustrates the difference in RRs between hourly data and one or five minute data, which is the motivation for using the unique collocation of highly time resolved data in Colorado in this paper.

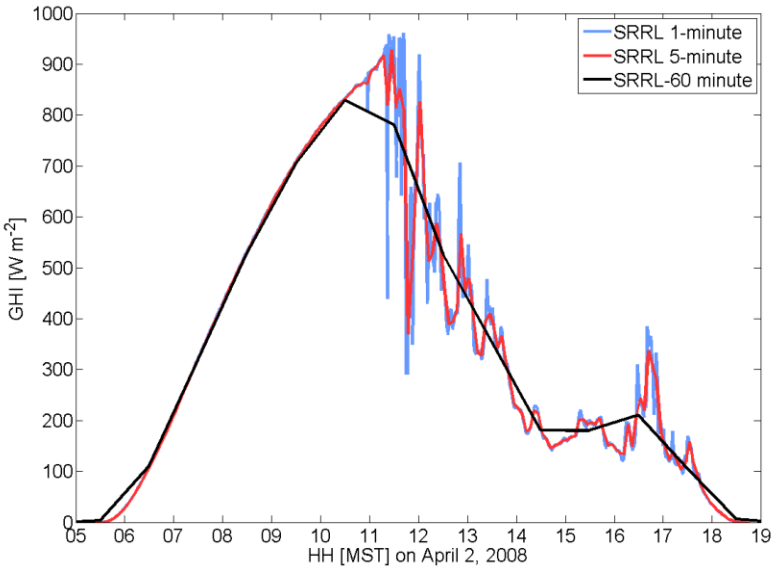

Fig. 2.2: Comparison of original 1 minute GHI data, with 5 minute and 60 minute averages for SRRL on April 2, 2008.

Visual examination of the timeseries revealed that SPMD tends to be shaded in the morning due to high surrounding terrain. This variability is naturally occurring (i.e. it would be

the same for a PV array at the same site), but was eliminated from the dataset by filtering out data at large solar zenith angles (SZAs). We also found that NWTC seems to be shaded at several times of the day, especially around 1400 MST. This unexplained shading in NWTC was difficult to correct for and the realization of the shading will be further discussed in the results section.

2.3. **Methods**

2.3.1. Data Quality Control

Since this paper does not attempt an assessment of the mean solar resource, but an analysis of the variability in GHI, slight sensor differences in offset and/or gain will not have a significant effect on our results. Nevertheless, we calibrated the sites against each other on clear days in the region, when they are expected to be similar given the small variability of atmospheric composition over short distances. The SRRL site is maintained daily by trained NREL staff and is considered to have the best data. Therefore, the NWTC and XCEL data were corrected using a linear regression against SRRL on nine clear days (Jan 13, Mar 3, Apr 14, Jun 14, Jul 13, Aug 28, Sep 16, Nov 19, Dec 25). The SPMD site was not corrected since higher clear sky atmospheric transmissivity associated with its high elevation (expected transmissivity of about 81% versus 79% for the other sites) would result in a different clear sky GHI.

A linear regression of GHI(SRRL) = A GHI(site X) + B was applied. The regression constants A and B for the XCEL site were nearly constant throughout the year, so an overall linear fit of all nine clear days was applied. The regression constants for NWTC showed a seasonal variation. Consequently, interpolated (time dependent) slopes and intercepts based on regressions from the nine clear days were used. The variability in the regression coefficients is an indication of differences in the cosine response of the sensors or of a sensor that is not leveled properly. While this will affect the average and seasonality of the mean solar resource, it has little influence on short-term RRs analyzed in this study.

The largest intercept from all regressions was B = 59.79 W m^{-2}, so all GHI less than 60 W m^{-2} were eliminated in the dataset. In this way, data near sunrise and sunset, which are prone to topographic shading effects and have little relevance in practice since power output is small and large RRs do not occur, are removed from the dataset. We stress that more advanced corrections (such as dependent on SZA) could be applied, but natural variability in airmass between sites and the objective of quantifying *variability* means the added value of such corrections is small.

2.3.2. Spatial and Temporal Correlation

To test the smoothing effect of geographic dispersion, a fifth virtual site, called average (AVG4), was defined as the average GHI of all four sites. Due to SPMD only having five minute temporal resolution, the average is comprised of the 5 minute averages of NWTC, SRRL, and XCEL, combined with the SPMD data.

For utilities and independent system operators, frequency of occurrence and magnitude of RRs of renewable power sources are the critical quantity of interest. A probability distribution function (pdf) can be used to compare the probabilities of different RR for the individual sites and for the average of all four sites. Five minute, daytime only (defined as GHI greater than 60 W m^{-2}) data were used for all sites. Ramps occur naturally throughout the day with changes in the SZA. However, since these ramps are predictable they are of lesser concern and were removed from the analysis. Expected clear-sky irradiance (SKC) was calculated using standard astronomical formulae and assuming that clear sky atmospheric transmissivity is $(0.75 + 2 \times 10^{-5} m^{-1} \times Z)$, where Z is station elevation in meters. SKC was subtracted from the GHI at each site, such that the remaining value was variation from expected irradiance. The RRs of these variations were calculated as the difference between successive data points over five minutes, using the equation RR = [GHI(t) - SKC(t) – GHI(t-5min) + SKC(t-5min)] / 5 min, given in units of W m^{-2} min^{-1}.

Power Spectral Density (PSD) analysis provides insight into how much of the variability in GHI can be explained by events at different frequencies such as daily and seasonal cycles in SZA, as well as variations in atmospheric transmissivity due to meteorological events such as cloud cover which occur over a range of frequencies. PSD analysis gives an indication of the amount of complementary ancillary services necessary to counteract variability at different frequencies. The larger the PSD, the larger the variance in power output and the larger the required rampable (conventional) power sources to make up for the difference. Low frequency oscillations, which may be due to daily or even seasonal cycles, can be supplemented with rather constant power sources such as traditional large power plants. High frequency oscillations are more challenging and must be supplemented with faster ramping power sources.

To test for overall correlation the Pearson correlation coefficient was calculated between the GHI time series at different sites. Due to the synchronized SZA variations at different sites in the same region, the correlation coefficients are usually close to one. Nevertheless, the smaller the correlation coefficient, the more smoothing would be expected (a correlation coefficient of -1 would indicate perfect smoothing of the aggregate power output of two sites). A better measure for correlation on shorter timescales, which is more important for smoothing the extreme ramp rates, is the coherence spectrum [13]. The coherence spectrum provides a normalized covariance at each frequency, allowing analysis of correlation at short timescales. Since long timescales (seasonal cycles, synoptic weather patterns, and daily cycles) affect all our sites in the same way, the coherence spectrum between any two sites is expected to have a value close to 1. On shorter timescales, the coherence spectrum indicates how correlated intra-day events such as transient clouds or mesoscale weather systems are between two sites. The timescale at which sites become uncorrelated is an indication of the longest timescale on which they will dampen aggregate variability.

2.4. **Results**

2.4.1. Ramp Rates

In the tables and figures referenced in this section the values for AVG4 are given throughout. However, we will wait to discuss these in a coherent fashion at the end of this section. The average and maxima of the magnitude of RRs for each site are shown in Table 2.2. With an average RR magnitude of 9.9 W m^{-2} min^{-1} SPMD had the largest, while XCEL had the smallest at 6.2 W m^{-2} min^{-1}. The maximum RRs are very similar for all sites. Since the mean absolute value of the RR is already a measure of the RR variability, the standard deviation of the RR (not the absolute value) is expected to give qualitatively similar results as is confirmed in Table 2.2. However, the kurtosis (the 4^{th} moment normalized by the 2^{nd} moment squared) weights extreme RR events higher. The fact that SPMD had the smallest kurtosis and largest standard deviation suggests that medium sized RRs were common, but extreme RRs were rarer than at the other three sites. The XCEL site had the lowest standard deviation but the highest kurtosis. Sensor shading would produce more extreme RRs, so the fact that the NWTC kurtosis is similar to SRRL and the standard deviations were equal indicates that shading likely was not a significant contributor to the variability at the site. Generally the kurtosis is much larger than 3 (the value expected for a Gaussian distribution) indicating the prevalence of extreme events (or large RR) in the distribution.

Table 2.2: Five minute ramp rate statistics for daytime GHI in 2008

	Mean(|RR|) [W m^{-2} min^{-1}]	Max(|RR|) [W m^{-2} min^{-1}]	Std(RR) [W m^{-2} min^{-1}]	Kurtosis(RR) [–]
NWTC	7.4	178.6	16.8	21.3
SRRL	7.4	160.8	16.8	20.6
SPMD	9.9	188.6	21.8	17.7
XCEL	6.2	163.7	14.9	23.8
AVG4	5.6	111.8	10.8	15.4

The probability density (pdf) of RRs is shown in Fig. 2.3. Consistent with the standard deviations found previously, the pdf of SPMD is the widest, indicating higher probabilities of

large RRs. The XCEL site has the most narrow pdf, meaning it is the least likely to have a large RR.

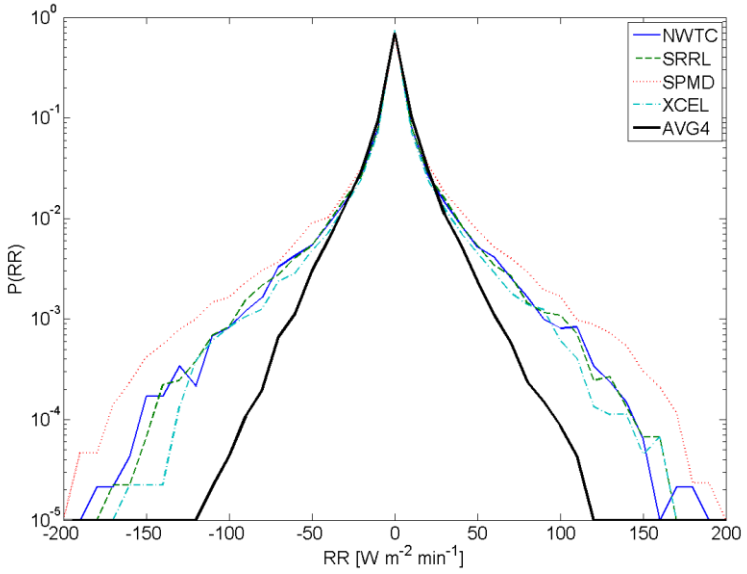

Fig. 2.3: Probability density function (pdf) of the ramp rates of NWTC, SRRL, SPMD, XCEL sites and the average of all four sites (AVG4) for 5 minute GHI-SKC data.

The cumulative density (cdf) of the absolute value of RR for each site is shown in Fig. 2.4. From the cdf, one can read the probability of RRs larger than a threshold. For the XCEL site, there is a 5% chance that RRs will be larger in magnitude than 31 W m^{-2} min^{-1}. At both NWTC and SRRL, there is a 5% chance of a RR with magnitude 36 W m^{-2} min^{-1} or larger. Again, SPMD shows the highest propensity to large RRs. There is a 5% chance that a RR at SPMD is larger than 49 W m^{-2} min^{-1} in magnitude. Given our 5 minute timestep and an average length of day of 10.7 hours, an event with a 5% chance occurs on average about 6 times per day.

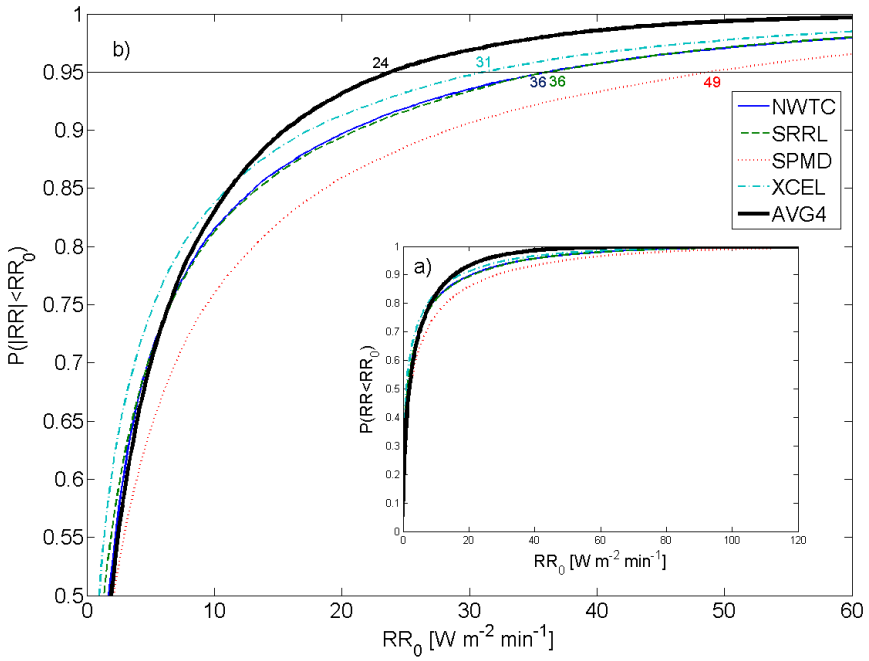

Fig. 2.4 : Cumulative probability distribution function (cdf) of the absolute value of GHI-SKC ramp rate over five minutes for NWTC, SRRL, SPMD, XCEL and their average. (a) Entire cdf; (b) Zoom in to the 'knee' of the graph showing the probability of medium to large RRs. The black horizontal line shows P=0.95 and intersects with this line are labeled with the RR magnitude which is exceeded 5% of the time.

2.4.2. Power spectral analysis

The power spectrum for all sites calculated for the entire year 2008 is shown in Fig. 2.5. In all four spectra, the largest peak appears at a frequency of 1.16×10^{-5} Hz, which corresponds to a period of 24 hours. This is expected due to cyclic daily availability of the SZA. Longer period cycles corresponding to variability over days to months also show a large PSD due to weather patterns and the seasonal variability of the SZA [10]. The higher frequency ($f > 2 \times 10^{-3}$ Hz) variations are dominated by atmospheric transmissivity changes by clouds. Generally, the PSD decreases with increasing frequency, but the rate of decrease varies from site to site.

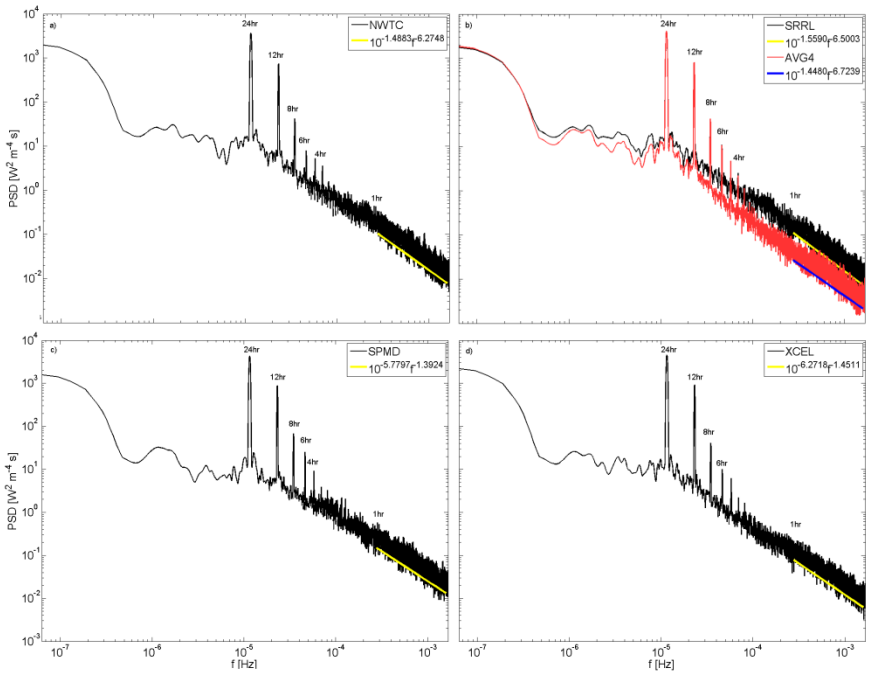

Fig. 2.5: Power spectral density of (a) NWTC (b) SRRL and AVG4 (c) SPMD, and (d) XCEL for 2008 using 5 minute averaged GHI data. The yellow line is the linear best fit line for time scales less than 1-hour. (b) The blue line is the linear best fit line for time scales less than 1-hour for AVG4.

To quantify this decrease in variability as a function of frequency at each site, linear fits for periods of less than 1 hour, 1-3 hours, and 3-11 hours are shown in Table 2.3. The SRRL site consistently had a steeper slope compared to the other sites. For periods less than one hour, the slopes ranged from -1.559 (SRRL) to -1.392 (SPMD). At 1-3 hours, SRRL again had the steepest slope of -1.477, but all the others had flatter slopes than for periods less than 1 hour. For 3-11 hours, the SRRL, NWTC, and XCEL slopes were -1.432, -1.452, and -1.509, respectively, but the SPMD slope was even flatter than in the other period ranges, at only -0.981.The ratios of the integrals of the PSD over periods shorter than one hour to the total integral over all periods (i.e., the fractions of total variance contained in the high frequencies) are also shown in Table 2.3 (2nd to last column). XCEL had the smallest ratio of high frequency oscillations, which suggests it has relatively more variability on long time scales. NWTC and SRRL had smaller integral ratios than SPMD, suggesting more high frequency forcing such as clouds at SPMD than at the other sites.

This is also consistent with the flatter slope at SPMD for $f > 1 / 3\ \mathrm{hr}^{-1}$ and the width of the pdf in Fig. 2.3.

Table 2.3: Statistics on the PSDs in Fig. 5. Columns 2-5 give linear regressions in a log-log plot, where the exponent of f is the slope. Column 6 is the fraction of the total variance contained in time scales less than 1 hour. Column 7 is the mean PSD intensity for timescales less than 1 hour.

	Linear regression for periods less than 1-hour	Linear regression for periods of 1-3 hours	Linear regression for periods of 3-11 hours	Linear regression for all periods	$\frac{\int_{f>1\mathrm{hr}^{-1}} \mathrm{PSD}}{\int_{\mathrm{all}\ f} \mathrm{PSD}}$	$< \mathrm{PSD} \mid f > 1\mathrm{hr}^{-1} >$ [$\mathrm{W^2\ m^{-4}\ s}$]
NWTC	$10^{-6.275}f^{-1.488}$	$10^{-4.918}f^{-1.115}$	$10^{-6.297}f^{-1.452}$	$10^{-5.857}f^{-1.352}$	0.0244	0.0448
SRRL	$10^{-6.500}f^{-1.559}$	$10^{-6.301}f^{-1.477}$	$10^{-6.186}f^{-1.432}$	$10^{-5.954}f^{-1.380}$	0.0224	0.0437
SPMD	$10^{-5.780}f^{-1.392}$	$10^{-5.323}f^{-1.256}$	$10^{-4.182}f^{-0.981}$	$10^{-5.333}f^{-1.247}$	0.0325	0.0684
XCEL	$10^{-6.272}f^{-1.451}$	$10^{-4.888}f^{-1.057}$	$10^{-6.640}f^{-1.509}$	$10^{-5.896}f^{-1.328}$	0.0160	0.0337
AVG4	$10^{-6.724}f^{-1.448}$	$10^{-6.617}f^{-1.408}$	$10^{-8.002}f^{-1.757}$	$10^{-6.674}f^{-1.432}$	0.0064	0.0117

2.4.3. Correlation coefficients and coherence spectrum

After discussing variability at each site, we will now compare the sites to one another to see if geographic dispersion can mitigate solar variability. The Pearson correlation of GHI and GHI-SKC at each of the sites increases with geographic proximity of the sites (Table 2.4). NWTC and SRRL show the strongest correlation due to their geographical proximity. The XCEL and SPMD sites show smaller correlations to the other two sites probably due to their large geographic distance (XCEL) and higher altitude (SPMD). Mountain ranges may act as natural barriers to cloud motion. GHI-SKC indicates the deviation of GHI from an expected value (SKC), so correlation of GHI-SKC is a better measure of ability to dampen variability on short time scales. As seasonal and diurnal cycles are taken out, the correlation coefficients for GHI-SKC decrease substantially (especially for distant sites) and become similar to the coherence at small time scales (Fig. 6). The weaker correlation of XCEL and SPMD to the other sites implies that integrating the XCEL and SPMD sites into the same grid as SRRL and NWTC will dampen fluctuations in solar power output of the average of all sites.

Table 2.4: Correlation coefficient between the sites for both GHI and GHI-SKC for daylight hours over the year 2008.

	NWTC		SRRL		SPMD		XCEL	
	GHI	GHI-SKC	GHI	GHI-SKC	GHI	GHI-SKC	GHI	GHI-SKC
NWTC	X	X	0.827	0.636	0.659	0.361	0.706	0.300
SRRL	0.827	0.636	X	X	0.658	0.365	0.706	0.306
SPMD	0.659	0.361	0.658	0.365	X	X	0.658	0.273
XCEL	0.706	0.300	0.706	0.306	0.658	0.273	X	X

The coherence of each of the four sites with the other sites is shown in Fig. 2.6. Generally the correlation between sites is highest for seasonal changes (about 0.8). Then, the correlation decreases with decreasing time scale but remains large including peaks at periods of 24 and 12 hours. A reduction in coherence is observed at a period of 2 days, which may indicate a decorrelation due to atmospheric transmissivity differences caused by different weather patterns or aerosol absorption. NWTC and SRRL again show the highest correlation, but only for timescales longer than three hours. For periods shorter than 12 hours (three hours for SRRL-NWTC), all sites have similar coherences of about 0.2, showing less correlated variation over short time scales. Consequently, when combined, the four sites are expected to smooth the averaged output on time scales shorter than 12 hours.

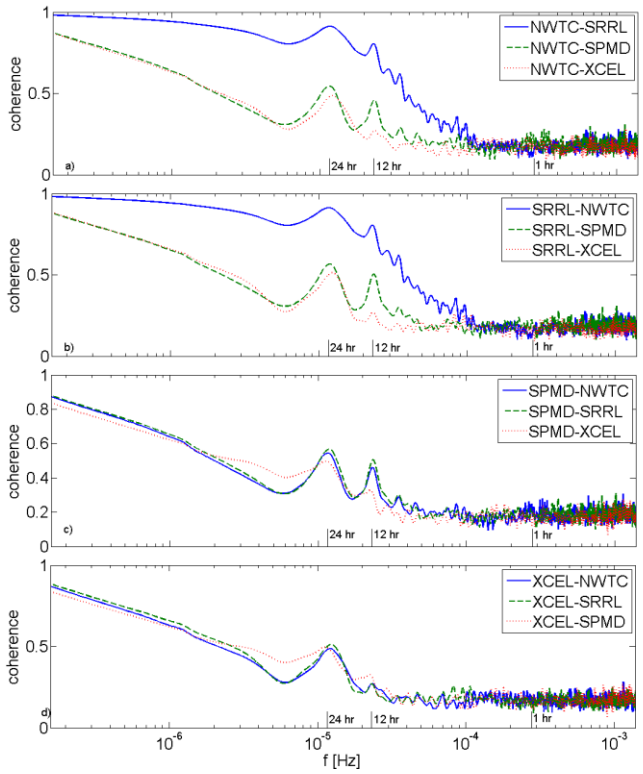

Fig. 2.6 : Coherence spectrum between GHI for each pair of sites for 2008.

2.4.4. Averaged output from all sites

The GHI timeseries sample (Fig. 2.7) showed that while each individual site fluctuates significantly over these two days, the average of all four sites has much smaller fluctuations. This confirms anecdotally that averaging geographically separated sites will lead to a smoother output.

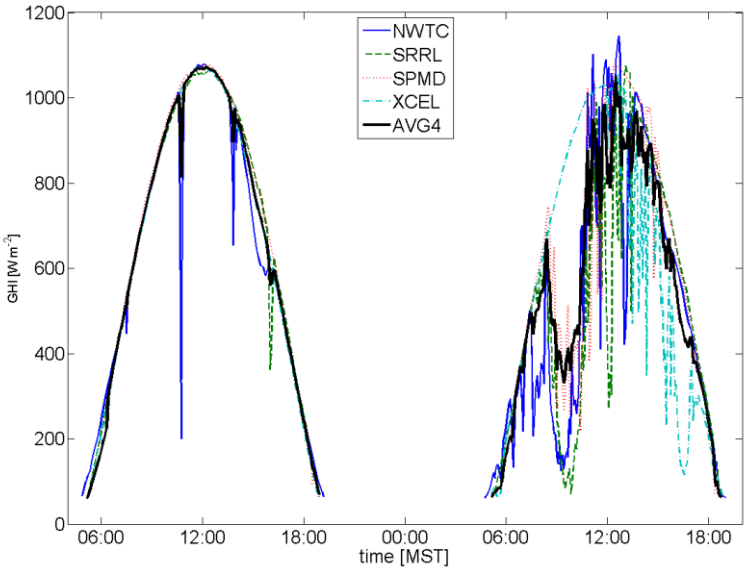

Fig. 2.7: Five minute GHI data at NWTC, SRRL, SPMD, XCEL and their average over June 10th and 11th, 2008, with values less than 60 W m-2 filtered out.

RR analysis of the average site confirms this effect (Table 2.2). There is a decrease in mean magnitude, maximum magnitude, standard deviation, and kurtosis for the AVG4 site over each of the individual sites. Particularly of note are the standard deviation and kurtosis, suggesting few extreme RRs. Fig. 2.3 confirms that the tails of the AVG4 pdf go to zero at a lower RR than the individual sites, indicating a lower probability of extreme RRs for AVG4. Based on the AVG4 cdf there was a 5% probability that the magnitude of the RR would be greater than 24 W m^{-2} min^{-1} (Fig. 2.4). This was less than half the RR found for the SPMD site alone.

While RR analysis has demonstrated a reduction in magnitude of RR of averaged irradiances from four sites, a power spectral analysis allows quantifying the time scales over which most of the reduction in RR occurs (Fig. 2.5b). The amplitude of the AVG4 PSD was smaller than the SRRL PSD for nearly all time scales. The PSD of AVG4 became visibly smaller than the individual sites for $f > 2 \times 10^{-6}$ Hz (5.8 days) and the ratio (as indicated by the vertical

distance on the log scale) became largest and remained constant for $f > 1 \times 10^{-4}$ Hz (2.8 h). This was consistent with the shortest period of high correlation observed for SRRL and NWTC in Fig. 2.6. Linear best fit lines and integral ratios of the PSD at the AVG4 site had consistent, steep slopes over all three period ranges that were close to the steepest slopes of any of the individual sites, indicating a strong reduction of high-frequency variability for AVG4. It should also be noted that the offset of the linear fit for the AVG4 site was the smallest for all three period ranges, which made the amplitude of variation of the AVG4 site on short timescales smaller than any of the other sites. For variability at periods shorter than 1 hour, the relative variance was less than half the relative variance at XCEL (the site with the least variability) and the mean PSD was less than a third the XCEL mean, indicating a significant decrease in high frequency variability when all four sites were averaged together.

2.5. **Discussion**

Overall, a significant smoothing effect was observed when the averaged solar irradiance at four solar sites across Colorado is compared to the individual sites. RR analysis showed a significant decrease in the mean RR magnitude, maximum RR magnitude, standard deviation, and kurtosis of the average compared to each site individually, consistent with previous work [10, 11]. Both the pdf and cdf of RRs indicated that the average of all four sites is less likely to have large fluctuations than each of the other sites individually. There was a 23-51% decrease in the RR that has a 5% probability of occurring for AVG4. This will mean smaller RRs and less uncertainty in operating the grid resulting in a reduced need for the procurement of expensive ancillary services or spinning reserve.

The variability of solar radiation over short timescales also decreased significantly for the averaged irradiance. Power spectral density analysis showed an overall slope of $f^{-1.43}$ for the average, while the individual sites ranged from $f^{-1.38}$ to $f^{-1.25}$ These were all consistent with

slopes for solar spectra, such as those found by Curtright and Apt of $f^{-1.3}$ for periods down to 10 minutes. However, these slopes were larger than the Kolmogorov spectrum ($f^{-5/3}$) expected for wind turbines and found by Apt (2007) and others [14]. PSD analysis showed significant decreases in the mean PSD intensity at frequencies corresponding to periods shorter than one hour and the ratio of high frequency oscillations to all oscillations. This drop-off of high frequency oscillations was consistent with Nanahara et al.'s (2004) findings for distributed wind turbine sites [15]. The average reduced high frequency fluctuations, but did not eliminate them. High frequency fluctuations in PV power output could be a problem because they require another power source to change its output at high frequencies in order to compensate.

High frequency variability could be reduced even further by increasing the number of and geographic dispersion between sites. For example, high correlation was found in GHI between NWTC and SRRL which are only 19 km apart. Coherence analysis showed that NWTC and SRRL were highly correlated in GHI for timescales longer than three hours, but become nearly uncorrelated for timescales less than three hours. Consequently while variability on large time scales will not decrease significantly by adding more sites, additional sites even if only a few km apart will decrease the short term variability.

As PV penetration into electricity grids increases, it is important to consider the variability, capacity factors, and peak shaving potential of PV, including the effect energy storage can have on the power output from PV systems. The appendix lists a simple analysis performed for the state of Colorado of the potential of PV to match the load.

Acknowledgements

We thank Tom Stoffel of NREL for his advice on Colorado weather and the choice and quality control of the solar monitoring sites. Funding was provided by DOE award DE-EE0002055.

Appendix: PV Power Production vs. Load

Another important consideration of solar energy is how the power produced will correspond to the load. DOE electric energy use data for the state of Colorado for 2007 shows that Colorado's net generation was 53,907 GWh, or 1.94×10^{17} J [16]. The AVG4 site would have produced 6.4×10^{9} Wh m^{-2}. Considering a virtual solar array with a solar conversion efficiency of 0.15, we find that it would take roughly 2.0×10^{8} m^2 or about 77 mi^2 of PV panels to produce enough electricity for Colorado during 2007. For reference, a very large PV array at Nellis Air Force Base in Nevada covers only 0.22 mi^2 [17]. In addition to land requirements, a storage device that could store this massive amount of energy for up to 6 months would be required. It has been demonstrated for wind turbines that by using storage devices, such as compressed air, the capacity factor can be raised to a baseload level (greater than 70%) [18]. However, since the average PSD slope found in this study is flatter than for wind turbines, PV will likely require more supplementary power than wind power.

These excessive land area and storage requirements show that it is very unlikely that solar power alone would ever power the entire state of Colorado, especially in the winter when evening heating demands are high and solar radiation is low. A better solution, at least in Colorado, appears to be using PV to supplement baseload power sources to cover the summer peak demand that is more synchronous with GHI [2, 19].

References

[1] REN21: Renewables Global Status Report, 2009.

[2] B.K. Sovacool, The intermittency of wind, solar, and renewable electricity generators: Technical barrier or rhetorical excuse?, Utilities Policy, 17 (2009) 288-296.

[3] R. Kropp, Solar expected to maintain its status as the world's fastest-growing energy technology., Sustainability Investment News, 2009.

[4] H.G. Beyer, J. Luther, R. Steinberger-Willms, Power fluctuations in spatially dispersed wind turbine systems, Solar energy., 50 (1993) 297-305.

[5] U. Focken, M. Lange, K. Mönnich, H.-P. Waldl, H.G. Beyer, A. Luig, Short-term prediction of the aggregated power output of wind farms—a statistical analysis of the reduction of the prediction error by spatial smoothing effects, Journal of Wind Engineering and Industrial Aerodynamics, 90 (2002) 231-246.

[6] C.L. Archer, M.Z. Jacobson, Supplying Baseload Power and Reducing Transmission Requirements by Interconnecting Wind Farms, Journal of Applied Meteorology and Climatology, 46 (2007) 1701-1717.

[7] I.f.S.E. Fordergesellschaft Windenergie, Increasing the penetration of wind energy in the European electricity network, 2000.

[8] L.C. N., A.T. P., Surface measurements of solar irradiance: a study of the spatial correlation between simultaneous measurements at separated sites, American Meteorological Society, Boston, MA, ETATS-UNIS, 1995.

[9] T.P. Barnett, J. Ritchie, J. Foat, G. Stokes, On the Space–Time Scales of the Surface Solar Radiation Field, Journal of Climate, 11 (1998) 88-96.

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

[11] E. Wiemken, H.G. Beyer, W. Heydenreich, K. Kiefer, Power characteristics of PV ensembles: experiences from the combined power production of 100 grid connected PV systems distributed over the area of Germany, Solar Energy, 70 (2001) 513-518.

[12] N.R.E. Laboratory, Measurement and Instrumentation Data Center (MIDC), in.

[13] T. Nanahara, M. Asari, T. Maejima, T. Sato, K. Yamaguchi, M. Shibata, Smoothing effects of distributed wind turbines. Part 2. Coherence among power output of distant wind turbines, Wind Energy, 7 (2004) 75-85.

[14] J. Apt, The spectrum of power from wind turbines, Journal of Power Sources, 169 (2007) 369-374.

[15] T. Nanahara, M. Asari, T. Sato, K. Yamaguchi, M. Shibata, T. Maejima, Smoothing effects of distributed wind turbines. Part 1. Coherence and smoothing effects at a wind farm, Wind Energy, 7 (2004) 61-74.

[16] DOE/EIA, DOE/EIA-0348(01)/2, Available at http://www.eia.doe.gov/cneaf/electricity/st_profiles/colorado.html, 2009.

[17] C. News, Green Tech. Air force base in Nevada goes solar with 14-megawatt array, Available at http://news.cnet.com/8301-11128_3-9829328-54.html [accessed 04.11.09].

[18] P. Denholm, G.L. Kulcinski, T. Holloway, Emissions and Energy Efficiency Assessment of Baseload Wind Energy Systems, Environmental Science & Technology, 39 (2005) 1903-1911.

[19] P. Denholm, R.M. Margolis, Evaluating the limits of solar photovoltaics (PV) in traditional electric power systems, Energy Policy, 35 (2007) 2852-2861.

Chapter 2, in full, is a reprint of the material as it appears in *Renewable Energy* volume 35, issue 12, December, 2010. Lave, Matthew; Kleissl, Jan, 2010. The dissertation author was the primary investigator and author of this paper.

3. Optimum Fixed Orientations and Benefits of Tracking for Capturing Solar Radiation in the Continental United States

Reprinted from *Renewable Energy*, Volume 36, Issue 3, Matthew Lave, Jan Kleissl, "Optimum fixed orientations and benefits of tracking for capturing solar radiation in the continental United States," Copyright 2011, with permission from Elsevier

Abstract

Optimum tilt and azimuth angles for solar panels were calculated for a grid of 0.1^0 by 0.1^0 National Solar Radiation Data Base (NSRDB-SUNY) cells covering the continental United States. The average global irradiation incident on a panel at this optimum orientation over one year was also calculated, and was compared to the solar radiation received by a flat horizontal panel and a 2-axis tracking panel. Optimum tilt and azimuth angles varied by up to 10^0 from the rule of thumb of latitude tilt and due south azimuth, especially in coastal areas, Florida, Texas, New Mexico, and Colorado. Compared to global horizontal irradiation, irradiation at optimum fixed tilt increased with increasing latitude and by 10% to 25% per year. Irradiation incident on a 2-axis tracking panel in one year was 25% to 45% higher than irradiation received by a panel at optimum fixed orientation. The highest increases in tracking irradiation were seen in the southwestern states, where irradiation was already large, leading to annual irradiation of over 3.4 MWh m^{-2}.

3.1. **Introduction**

Solar photovoltaic (PV) systems are quickly gaining popularity in the United States (U.S.), thanks to incentive programs and enhanced interest in environmental sustainability and energy independence. As more PV systems are installed across the U.S., it becomes increasingly important to maximize their power output. Aside from increasing a panel's solar conversion efficiency, power output can be increased by considering the solar geometry as well as the seasonal and daily variation of atmospheric transmissivity at a particular site. Specifically, it is important to know what the optimum tilt and azimuth angles are at which to mount a fixed tilt

panel on a flat roof or on the ground such that it receives maximum irradiation. In addition, knowing the increase in solar radiation incident on a two-axis tracking panel will allow analysis of the economics of tracking PV systems, which are more expensive to install and maintain.

Since the power production of a PV panel is close to linearly proportional to the amount of solar radiation (photons) reaching the panel surface, incident irradiation is an excellent proxy for power output. To maximize absorption of solar radiation in clear skies, the normal to the plane of the PV panel should be pointing towards the sun such that the solar direct beam is perpendicular to the panel surface. While a fixed tilt panel can only be normal to the incident sunlight once a day, a two-axis tracking panel improves over a fixed tilt panel by following the sun through the sky such that the plane of array normal is always parallel to the incident sunlight. However, when the majority of global irradiance is diffuse, horizontal alignment often provides the maximum global irradiance [1].

Some previous studies used modeled extraterrestrial radiation incident on the top of the atmosphere to find equations for optimum tilt over a large area[2, 3]. This method accounts for the deterministic (celestial) variables which affect solar radiation, but it does not consider the stochastic (clouds and other weather) variables which also affect the optimum angles. Using an extraterrestrial radiation model, the optimum azimuth is always due south (or north in the southern hemisphere), since solar radiation will be symmetric about solar noon. These studies [2, 3] confirmed the simple rule of thumb that tilt angle, β, equal to latitude, φ, is optimal for a clear year (e.g., panel at 40^0N should have 40^0 tilt from horizontal). For different seasons of the year, the optimum tilt was found to differ by up to 15° from latitude (more in the winter, less in the summer).

Other studies have used measured solar radiation data instead of clear-sky models to account for both the celestial and weather changes. Measured global horizontal irradiation (GHI) at four sites in the U.S. state of Alabama was used to find the optimum yearly tilt angle, $\beta_{opt} =$

$\varphi \pm 8^0$ [4]. Using a fixed tilt PV panel at different tilts and a tracking PV panel mounted on a roof in Sanliurfa, Turkey (37^0 N) optimum tilt and the effect of tracking was quantified[5]. Optimum tilts ranged from 13^0 in July to 61^0 in December. Solar irradiation received on a tracking panel was 29% larger compared to a panel at optimum tilt for one day in July. Using measured GHI and diffuse horizontal irradiation (DHI) values the optimum yearly tilt was found to be 3.3^0 for Brunei Darussalam (4.9^0 N)[6], and 30.3^0 for Izmir, Turkey (38.5^0N)[7].

One comprehensive study computed optimum tilt for a large area using measured irradiation[8]. GHI and DHI from 566 ground meteorological stations across Europe, turbidity data from 611 sites, and a digital elevation model were used to derive expected radiation over a 1 x 1 km grid covering Europe. The optimum yearly panel tilt is less than latitude tilt for Europe and is not solely a function of latitude (as concluded in the clear sky studies), but is also a function of cloudiness.

To the best of our knowledge, maps of optimum tilt and azimuth angles based on measured radiation have not been published for the U.S. In this paper, we present solar maps of the continental United States (CONUS) showing the optimum panel tilt and azimuth, the radiation incident on a panel at optimum tilt and azimuth, and the radiation received by a tracking panel. While maps showing irradiation at latitude tilt facing due south and tracking irradiation have been published by the National Renewable Energy Laboratory (NREL)[9], the NREL maps do not show optimum tilt and azimuth. Furthermore, with their resolution of 40 x 40 km, the maps presented here will have sixteen grid points for every one grid point in the NREL maps. The higher spatial resolution is especially important in areas with strong gradients in radiation, such as coastal or mountainous areas. We will describe the data source (Section 2), the model used to compute irradiation on a tilted plane (Section 3), and the derivation of optimum angles (Section 4). Section 5 presents a validation of the algorithm, and sections 6 and 7 describe the resulting maps and conclusions, respectively.

3.2. **Data**

Satellite derived GHI and DHI obtained from the National Solar Radiation Database (NSRDB-SUNY) were used for this study[10]. The NSRDB-SUNY dataset contains hourly GHI, DHI, and direct normal irradiation (DNI) values for the entire United States on a 0.1^0 node registered grid, corresponding to a grid spacing of about 10 km, for 1998-2005. NSRDB-SUNY was created by applying the model developed at the State University of New York (SUNY) – Albany [11] to satellite imagery of the U.S. from Geostationary Operational Environmental Satellites (GOES). A cloud index was derived for each pixel and was used along with atmospheric turbidity, site elevation, ground snow cover, ground specular reflectance characteristics and individual pixel sun-satellite angle to derive surface irradiations. Atmospheric turbidity is quantified in terms of the air mass independent Linke Turbidity coefficient[12], which is a function of monthly average atmospheric aerosol content, water vapor and ozone. This Linke Turbidity coefficient was used to compute clear sky DNI and DHI. Clear sky DNI was multiplied by a ratio of DNIs calculated using the DIRINT model [13] to find DNI for the actual sky condition. DHI was found by finding the vertical component of DNI[14].

The SUNY gridded data comes from two satellites: GOES-East and GOES-West, which produce snapshot images at 15 minutes past the hour and on the hour, respectively. Although GOES satellite data has a resolution of 1 x 1 km, the data are down-sampled to a 10 km grid to reduce computation time of the SUNY model[14]. For consistency, the SUNY gridded data is shifted and interpolated to model the sum of irradiance incident on each grid point for the previous hour (‘Sglo’ column). This results in each hourly irradiation value having the units of Wh m^{-2}. Hourly uncertainties of the SUNY gridded data range from 8% under optimal conditions to up to 25%[14], though the mean bias error for long periods of time – such as the 8 years used in this study – is expected to be much lower than these values [15].

We chose to use the SUNY gridded data because of the long time period, the high spatial resolution, the consistency in its derivation for a large area, and because DHI and DNI are provided. While the NSRDB and other sources contain measurements from grounds stations which have smaller errors than SUNY measurements, the spatial resolution is poor. The NSRDB, for example, only contains high quality ground irradiation measurements from 221 sites in 40 states [14] as compared to the 97,305 NSRDB-SUNY grid points covering the CONUS.

3.3. **Global Irradiation on a Tilted Plane**

3.3.1. Direct Irradiation

For this study, SUNY irradiations on a horizontal surface had to be converted to irradiations at an arbitrary tilt and azimuth. We chose to use the algorithms described by Page [16] due to the deterministic functional dependence upon location which is desirable in processing data for the entire CONUS. Other models (e.g. Perez[17]) require empirical coefficients which must be determined at each location using ground measurements.

The Page Model takes GHI, DHI, time, latitude, and longitude as inputs and outputs global hourly irradiation (GI) for a panel of any tilt and azimuth as $GI = B + D + R_g$. Direct beam (B), diffuse (D), and ground reflected irradiation (R_g) on the tilted surface are calculated using astronomical variables, ground surface albedo, and an empirical function relating diffuse and GI.

Direct (beam) irradiation, B, on the tilted surface is a function of tilt β and azimuth α ($\alpha = 0$ is due south) as

$$B(\beta, \alpha) = B_n \cos v(\beta, \alpha), \tag{3.1}$$

where $v(\beta, \alpha)$ is the solar incidence angle on the tilted panel. B_n is beam normal irradiation at the tilted panel surface, which is calculated from the SUNY data as

$$B_n = (G_h - D_h)/\sin \gamma_s. \tag{3.2}$$

For computational efficiency, the SUNY DNI was not used , but was found to be similar to B_n from Eq. 3.2. G_h and D_h represent GHI and DHI, respectively. The mid-hour solar altitude angle, γ_s, is the angular elevation of the sun above the horizontal plane, and is a function of the solar declination angle, the solar hour angle, and the latitude at which the panel is installed. Since the NSRDB-SUNY presents the average of the irradiation over the previous hour, mid-hour values of the solar altitude angle are used since they are nearly an average of the solar altitude angle over the previous hour. For example, the irradiation with time stamp 1100 LST will be the average of irradiation between 1000 LST and 1100 LST and the solar altitude angle will be calculated at 1030 LST.

3.3.2. Diffuse Irradiation

The diffuse component of irradiation on a tilted surface is significantly more complicated to model. Page [16] computes the ratio of diffuse radiation on the tilted panel to diffuse horizontal radiation as

$$\frac{D(\beta,\alpha)}{D_h} = f(\beta)(1-K_b) + \frac{K_b \cos\nu(\beta,\alpha)}{\sin\gamma_s} \tag{3.3}$$

To account for cloud cover, a modulating function in the form of a clearness index, K_b, is used: $K_b = (G_h - D_h)/[\epsilon * (1367\,\mathrm{Wh\ m^{-2}}) \sin\gamma_s]$, where ϵ is the correction to the mean solar distance from earth. We used the empirical function $f(\beta)$ found by Page for Southern Europe to relate the directionality of diffuse irradiation to the panel tilt angle and clearness index

$$f(\beta) = cos^2\left(\frac{\beta}{2}\right) + \left(0.00263 - 0.7120K_b - 0.6883{K_b}^2\right)[\sin\beta - \beta\cos\beta - \pi\, sin^2\left(\frac{\beta}{2}\right)] \tag{3.4}$$

which was accurate for our validation site in Golden, Colorado (see Section 3.5 later).

3.3.3. Reflected Irradiation

Reflected irradiation, R_g, is modeled by

$$R_g = r_\beta \rho_g G_h, \tag{3.5}$$

where the reflection coefficient r_β is solely a function of panel tilt, $r_\beta = \frac{1-\cos\beta}{2}$. We assumed a ground surface albedo of $\rho_g = 0.2$ as an average for land. Use of albedo maps from remote sensing would be more accurate, but spatial heterogeneity of albedos within the typical 100 km^2 grid cell, especially in urban areas, would still cause a large margin of error. At the orientations used in this study, the sensitivity to albedo is small. In San Diego, California (117.25^{0}W, 32.85^{0}N), changing the albedo by ±0.1 from the assumed 0.2 leads to a ±0.71% change in the average annual irradiation reaching a panel at optimum orientation and a ±1.08% change for a tracking panel. Similarly, for Albany, New York (73.85^{0}W, 42.85^{0}N), altering the albedo by ±0.1 leads to a ±0.87% difference for a panel at optimum orientation and a ±1.20% difference in a tracking panel. These albedo changes led to 2° changes in optimum tilt for both San Diego and Albany.

3.4. **Optimum panel angles, tracking, and resulting irradiation**

3.4.1. Optimum tilt and azimuth angles for a fixed panel

To determine optimum tilt and azimuth angles, the Page Model was written in function form with inputs of panel tilt, panel azimuth, latitude, longitude, time, G_h, and D_h. The output of this function is the sum of GI on a panel over the 8 years contained in the SUNY gridded data. Then, for each SUNY grid point (fixed latitude, longitude, time, G_h, and D_h), the local maximum GI as a function of panel tilt and panel azimuth was found using unconstrained nonlinear optimization (function 'fminsearch' in MATLAB, The Mathworks, Inc.). The optimum tilt and azimuth angle as well as the maximum annual irradiation reaching a panel at optimum fixed tilt were recorded for comparison to horizontal and tracking panels.

3.4.2. Irradiation onto a Tracking Panel

For concentrating systems, DNI is the relevant metric and corresponding maps already exist [9]. This study focuses on fixed (typically PV) systems and tracking results are only shown

for reference. To determine the average annual GI reaching a two axis time-position (or chronological) tracking panel, the same function described in Section 3.4.1 was used, but the panel tilt angle was set equal to the solar altitude angle while the panel azimuth was set equal to the solar azimuth. While other tracking technologies such as single axis, active or passive tracking exist, we chose to use time-position two-axis tracking due to its simplicity and generality. Other tracking techniques may outperform time-position tracking in high diffuse radiation conditions as then a flat orientation is usually optimal [1].

3.5. **Validation**

3.5.1. Comparison to measurements at SRRL

The model was validated against measured irradiation from the NREL Solar Radiation Research Laboratory (SRRL) located at Golden, Colorado (39.74^0N, 105.18^0W). SRRL was selected due to its high data quality and because hourly measurements of GHI, DHI, GI on a surface tilted 40^0 due south, and GI on a tracking panel all taken at the same location are available. A Kipp and Zonen CM 22 pyranometer measures GHI, another CM 22 pyranometer with a diffuse shading disk measures DHI, an Eppley Laboratory, Inc. Precision Spectral pyranometer measures GI on the tilted surface, and GI for a two axis tracking panel was measured using a Kipp and Zonen CM 21 pyranometer[18].

3.5.1.1. Tilted Panel

Inputs of hourly GHI and DHI for January 1 to December 31 2009 were obtained from the NREL Measurement and Instrumentation Data Center (MIDC)[18]. Although 1-minute resolution is available from the MIDC, hourly resolution was chosen to be representative of the hourly SUNY data. We input albedo (0.2), tilt angle (40^0) and azimuth of the tilted panel (0^0) to obtain direct, diffuse, and reflected radiation on the tilted surface. These were summed to create an estimated GI on the tilted panel (Section 3.3), which was compared to the measured GI on the tilted panel for times when solar altitude angle > 10^0 (Fig. 3.1a).

For this comparison, the mean absolute error (MAE), mean bias error (MBE), root mean squared error (RMSE), relative MBE (rMBE), and relative RMSE (rRMSE) were computed from the instantaneous error, e, by

$$e = GI_{tilt,calc} - GI_{tilt,meas} \quad 3.6$$

$$MAE = \frac{\sum |e|}{n} \quad 3.7$$

$$MBE = \frac{\sum e}{n} \quad 3.8$$

$$RMSE = \sqrt{\frac{\sum (e^2)}{n}} \quad 3.9$$

$$rMBE = \frac{\frac{\sum e}{n}}{\sum GI_{tilt,meas}} \quad 3.10$$

$$rRMSE = \frac{\sqrt{\frac{\sum (e^2)}{n}}}{\sum GI_{tilt,meas}} \quad 3.11$$

The Pearson correlation coefficient and other error metrics (Table 1) show the strong correlation between the estimated and measured GI. The RMSE for GI on the tilted surface was found to be 5%, which is smaller than errors reported for GHI and DNI in the SUNY-gridded data [14]. Since the 8 year sum is the only value used in creating the maps presented in this paper, the small rMBE value validates the Page Model for our application.

3.5.1.2. Tracking Panel

The Page Model was also applied for a tracking panel at SRRL, and compared to measured values (Fig. 3.1b). Again, a high correlation is observed between measured and modeled GI (Table 1). All statistics show larger errors than for the fixed tilt case, but the RMSE (7%) is still smaller than errors reported in the SUNY data[14]. In addition, the rMBE remains very small.

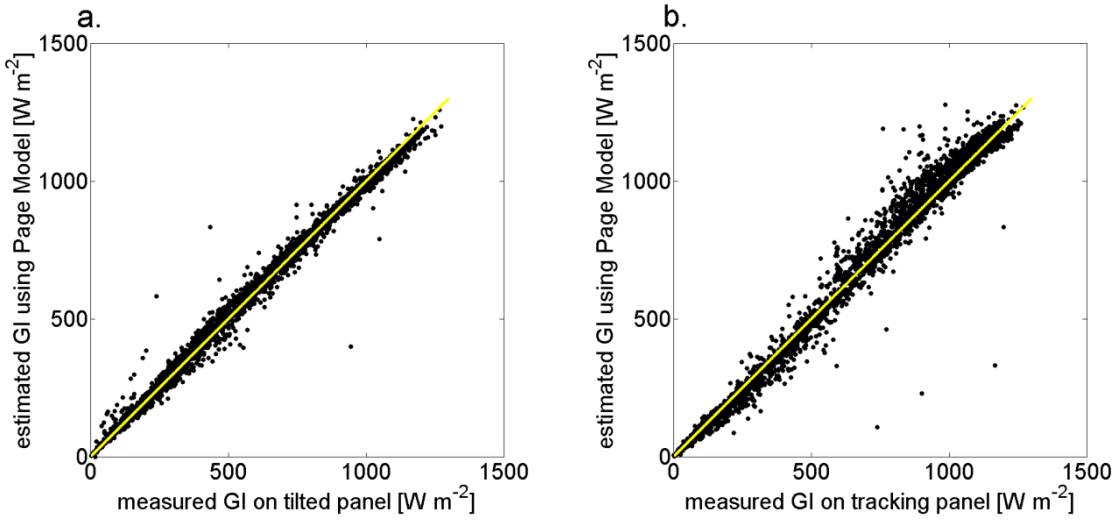

Fig. 3.1: Scatter plot of (a) the measured GI on a surface at 40^0 tilt, facing due south, and (b) the measured GI on a tracking panel, plotted versus values calculated using the Page Model. Measurements were taken at SRRL in Golden, Colorado. The values shown are hourly daytime values (solar elevation angle $>10^0$) for the entire year 2009. The solid lines are 1:1 lines for reference.

Table 3.1: Daytime (solar altitude angle>10^0) statistics for errors between the measured GI at the SRRL panel tilted 40^0 or the SRRL tracking panel and the calculated GI using the Page Model for 1998-2005.

	Mean Measured GI [Wh m^{-2}]	Pearson Correlation Coefficient [-]	MAE [Wh m^{-2}]	MBE [Wh m^{-2}]	RMSE [Wh m^{-2}]	rMBE [-]	rRMSE [-]
40^o tilt	541.2	0.997	17.1	2.37	27.5	0.44%	5.0%
tracking	704.4	0.992	29.2	4.84	50.6	0.69%	7.2%

While the Page Model shows larger MAE and RMSE errors resulting from differences on an hour-by-hour basis, bias errors of GI received on either a fixed tilt or a tracking panel are small. The analysis in this paper relies on 8 year averages of the SUNY-gridded data, so the rMBE is the most important statistic presented in Table 1.

3.5.2. Comparison to PVWatts

PVWatts2 [19] is a tool published by NREL "to permit non-experts to quickly obtain performance estimates for grid-connected PV systems" [20]. It allows for the calculation of

irradiation on a panel of any fixed tilt provided by the user, as well as for a 1- or 2-axis tracking panel at any location in the U.S. using gridded irradiation data at 40 km resolution. Comparisons between PVWatts2 and our algorithm for 5 cities in the U.S. are shown in Table 2. The centers of PVWatts2 grid points do not correspond to the SUNY 10km data, so a distance weighted average of the 4 closest SUNY sites to the center of the PVWatts2 grid point was used to determine the values presented in Table 2. Some deviations between our results and PVWatts2 are expected due to the differences in spatial resolution, especially in areas with large gradients in irradiation. For this reason, the PVWatts2 grid points chosen for Table 2 are far away from coasts, except for Los Angeles, where the grid cells are at least 25 km from the ocean. Overall, the absolute differences between our results and PVWatts2 range from 0.5% to 5.1%. These are smaller than the expected error of PVWatts2 of 10-12%[19], suggesting that our implementation of the Page Model compares well to established models.

Table 3.2: Comparison and relative mean bias error of irradiation calculated from PVWatts2 and Page Model for panels at optimum fixed tilt and for tracking panels at selected sites.

Location	Optimum Orientation (tilt/azimuth)	GI at Optimum Fixed Orientation [kWh m^{-2} day^{-1}]			GI for Tracking [kWh m^{-2} day^{-1}]		
		SUNY + Page	PVWatts2	rMBE	SUNY + Page	PVWatts2	rMBE
Orlando, FL (81.35^0W, 28.49^0N)	$29.1^0/7.6^0$E	5.22	5.29	-1.3%	6.77	6.81	-0.6%
Dallas, TX (96.84^0W, 32.72^0N)	$30.5^0/5.9^0$E	5.09	5.24	-2.9%	6.76	6.92	-2.3%
Phoenix, AZ (112.21^0W, 33.50^0N)	$33.4^0/0.3^0$ W	6.50	6.29	3.3%	9.09	8.65	5.1%
Los Angeles, CA (117.95^0W, 34.08^0N)	$32.4^0/3.8^0$ W	5.79	5.82	-0.5%	7.73	7.52	2.8%
St. Louis, MO (90.28^0W, 38.49^0N)	$34.8^0/1.0^0$ W	4.76	4.82	-1.2%	6.25	6.31	1.0%

3.6. **Maps**

The Page Model was applied to produce solar radiation maps of the CONUS. Figures 2a and 2b show the optimum annual tilt from horizontal and the optimum tilt minus latitude,

respectively. The values for optimum tilt were calculated coupled with the optimum azimuth. Since there is an interdependence of optimum azimuth and tilt, optimum tilts at an azimuth of 0^0 were also calculated, but the mean absolute deviation from Figure 2a was less than 0.1^0 and the maximum deviation was 1.4^0. Therefore, results for optimum tilt angels in Figure 2a apply to both optimum azimuth and due south azimuth. Figure 2b can be used to investigate the rule of thumb that panels should be installed at latitude tilt. If latitude tilt were indeed the optimum tilt, then Figure 4 would show zero differences. Indeed, the differences shown in Figure 4 are small for low latitudes, but they become significant at higher latitudes. Some areas exhibit distinct differences in optimum tilt from points at the same latitude, showing that optimum tilt is not solely a function of latitude. Latitude tilt might be accurate for clear sky conditions, but if a site shows seasonal variations in cloudiness, its optimum tilt will be altered. For example, California's Central Valley experiences Tule fog during winter. Consequently, the optimum tilt there is weighted towards the best tilt in the (clearer) summer months when more radiation can be collected.

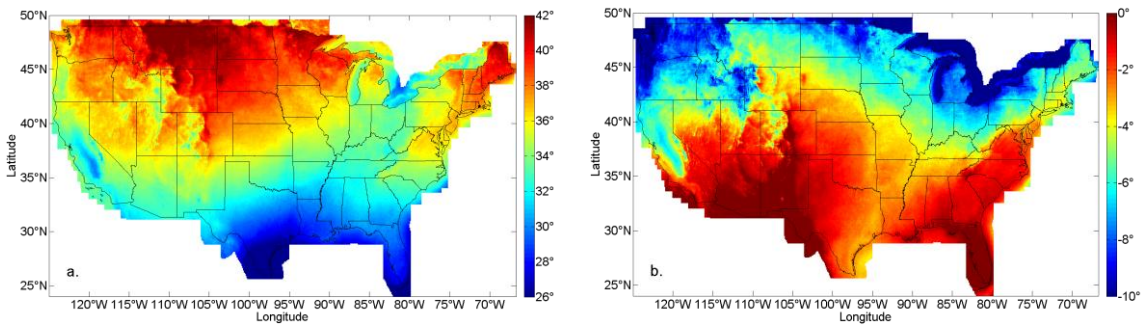

Fig. 3.2: (a) Map of the optimum tilt from horizontal to maximize annual incident GI. (b) Map of the optimum tilt from horizontal (a) minus the latitude for each location. This shows the difference in degrees between the optimum tilt and the rule of thumb suggesting latitude tilt.

Figure 3 displays the optimum annual azimuth for a solar panel, and can be combined with Figure 2a to determine the optimum annual orientation (tilt and azimuth) for a fixed solar panel. Figure 3 can also be used to investigate whether a due south azimuth results in the maximum irradiation. Due south is optimal for many parts of the country, but there are notable

exceptions in Florida, Central Texas, the centers of Wyoming, Colorado, and New Mexico, and along the Pacific Coastline. A due south azimuth would suggest that equal amounts of solar radiation are received before and after solar noon. A non-zero azimuth therefore suggests that solar radiation at a given site was not symmetric. For example, many parts of the Pacific Coastline are subject to summer morning fog that evaporates in the late morning, leading to more afternoon irradiation and thus an optimum azimuth facing towards the west. Large parts of Florida and New Mexico, on the other hand, are often subject to afternoon convective clouds, leading to an optimum azimuth facing east. Figure 3 shows a discontinuity in optimum azimuth angle around a latitude of 107.5^{o} which marks the border between GOES-East and GOES-West satellite data in the SUNY dataset. An error in the time shift of SUNY satellite irradiances to hourly irradiation for the evening hours related to the timing of GOES imagery is the likely explanation [15]. This would indicate that our azimuth angles for the area west of 107.5^{o} are biased towards the west.

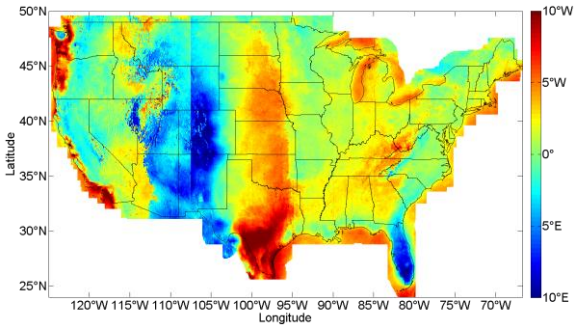

Fig. 3.3: Map showing the optimum azimuth to maximize incident GI. Due south is 0^{0}, and values shown on the map are east or west of due south. When coupled with Figure 2a, the optimum orientation (tilt and azimuth) at any location on the map can be determined.

The average annual GI reaching a panel at optimum fixed tilt and azimuth is mapped in Figure 4a (a very similar map of irradiation reaching a panel at latitude tilt and south azimuth is presented on the NREL website[9]). This map is useful in determining where it would be best to install fixed solar panels, since incident radiation is nearly linearly proportional to power output of a PV panel. Areas of highest annual solar radiation at optimum tilt are located in the

southwestern U.S. including southeastern California, southern Nevada, Arizona, southern Utah, New Mexico, southern Colorado, and western Texas. Most of the rest of the CONUS receives much lower amounts of annual solar radiation at optimum tilt. For example, Florida and the southern tip of Texas are both at lower latitudes than the southwestern states, yet receive around 0.5 MWh m^{-2} (about 20%) less GI per year.

To determine the importance of installing a solar panel at the optimum orientation, the percentage increase in GI reaching an optimally oriented panel versus GHI is shown in Figure 4b. At almost every location in the CONUS the irradiation increases by at least 10% at optimum orientation over flat. The gain increases with increasing latitude, with a maximum of 25% increase observed in parts of Montana. This is consistent with the increase in optimum tilt shown in Fig. 3.2a. The further north a site is, the larger the difference between flat and optimum orientation and the larger the increase in irradiation received at optimum orientation.

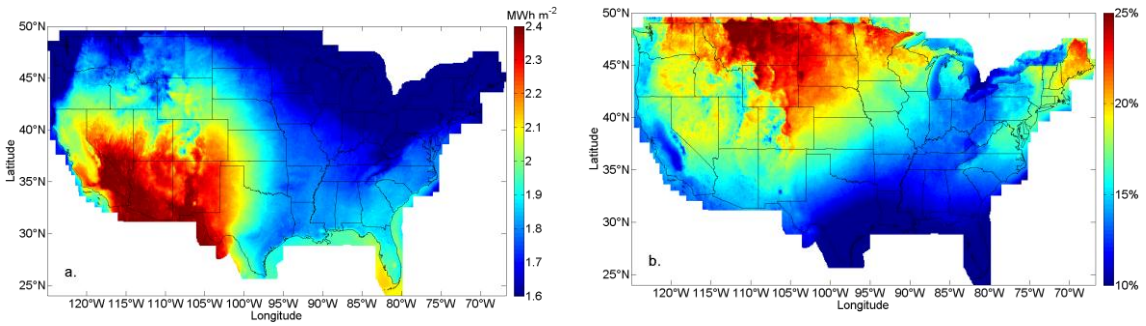

Fig. 3.4: (a) Map showing the average annual GI reaching a panel at optimum tilt and azimuth. (b) Map showing the percentage increase in GI incident on a PV panel at optimum tilt and azimuth versus a flat horizontal panel.

Fig. 3.5a shows the average annual GI reaching a tracking panel at each location. This map is similar to Fig. 3.4a which shows the GI at optimum orientation. This is expected since both depend on the GHI reaching each location. However, the solar irradiation reaching tracking panels is significantly larger than the irradiation reaching optimally oriented panels. To quantify this increase, Figure 5b shows the percentage increase in GI reaching a tracking solar panel over a panel at fixed optimal orientation, which ranges from 25% to 45%. The largest percent increases

occur in the southwestern U.S., while the smallest increases occur in the eastern U.S. and on the Pacific Coastline. The large increases are related to clear skies when irradiance is more variable as a function of view angle and tracking shows greater benefits. Smaller increases are related to cloudy conditions which cause more uniform diffuse irradiation patterns. In cloudy conditions other tracking strategies would result in a slightly better performance (Section 3.4.2). The highest percentage increases occur in areas where irradiation at optimum tilt was already large (i.e., less cloudy areas), making tracking panels very attractive for areas such as the southwestern states.

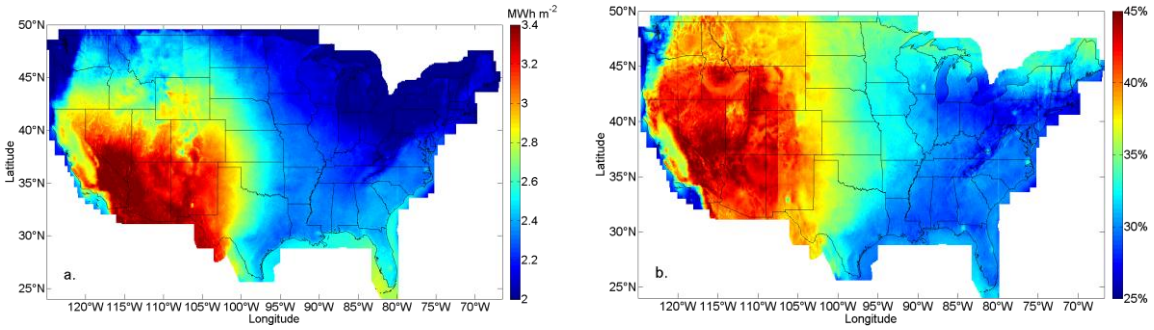

Fig. 3.5: (a) Map of the annual GI reaching a two-axis tracking solar panel. (b) Map of the percentage increase in GI reaching a tracking panel over GI reaching a panel at fixed optimum orientation.

3.7. **Conclusion**

The optimum tilt and azimuth angles to collect global solar irradiation (GI) in the CONUS were determined using the Page Model applied to the SUNY 10 km gridded data. While rules of thumb suggest that maximum GI is obtained at latitude tilt with an azimuth facing due south, it was found for most locations in the CONUS that higher GI could be obtained by deviating from this rule. The optimum tilt was never found to be greater than latitude tilt, but it was found to be up to 10^0 less than latitude tilt. On average, the deviation from latitude tilt increased at higher latitudes, but optimum tilt was not found to simply be a function of latitude. Seasonal weather patterns such as winter clouds led to changes in the optimum tilt. Azimuths deviating up to 10^0 west or east of due south were found for areas with typical daily cloud patterns such as morning fog or afternoon thunderstorms.

Areas of high GI on an optimal fixed orientation panel occur in the southwestern U.S., with up to 2.4 MWh m^{-2} per year. Compared to global horizontal irradiation, irradiation at optimum fixed tilt increased with increasing latitude and by 10% to 25% per year. These increases are significant considering they require no more active work than determining the optimum orientation during panel installation. However, the sensitivity of annual irradiation to inclination is small near the optimum point. While small increases in power production may have a significant impact on economic payback time, other factors such as aesthetics and mounting considerations may dictate a near-optimal angle that would result in a relatively small loss in annual power production.

GI reaching a tracking surface shows very similar geographic patterns to GI at optimum fixed orientation, but a tracking panel can receive over 3.4 MWh m^{-2} per year. The increase from using a tracking panel was strongest where fixed orientation irradiation was large, corresponding to relatively clear skies on average. This suggests using tracking panels in areas of high irradiation so long as the increases are enough to balance the higher initial costs, maintenance costs, and energy lost to the tracking mechanism.

Overall, we found that the rule of thumb for orientation of a solar panel was up to 10^0 off for tilt, azimuth, or both. We recommend using optimum tilt and azimuth angles presented here to increase irradiation received at any site. Our analysis does not consider the temperature effect on PV efficiency. Given that the panel temperatures are larger in summer than in winter and larger in the afternoon than in the morning, consideration of this effect would result in larger tilts and an azimuth facing more east of south, but the changes are expected to be small and depend on the PV temperature coefficient. Moreover this analysis optimizes for annual irradiation, but does not consider the seasonality and diurnal pattern of electricity prices. If the PV array output displaces consumption from the local facility with time-of-use pricing or if the electricity generated is bid into the market, the irradiations would have to be weighted by the electricity price at the time to

determine economic effects. Since electricity prices vary by region and time, accumulating such a database from 1997-2005 was beyond the scope of this project. Generally, since electricity is more expensive in the summer and during the afternoon peak demand the optimum azimuth would be further west of south and the optimum tilt would be closer to zero if economic consideration were taken into account.

Acknowledgements

We appreciate funding from the DOE High Solar PV Penetration grant 10DE-EE002055.

The maps presented in this paper are also available as supplementary material for use with Google Earth.

References

[1] N.A. Kelly, T.L. Gibson, Improved photovoltaic energy output for cloudy conditions with a solar tracking system, Solar Energy, 83 (2009) 2092-2102.

[2] J.A. Duffie, W.A. Beckman, Solar Engineering of Thermal Processes, Wiley, New York, 1980.

[3] G. H.P., Treatise on Solar Energy: Volume1: Fundamentals of Solar Energy, Wiley, New York, 1982.

[4] G. Lewis, Optimum tilt of a solar collector, Solar & Wind Technology, 4 (1987) 407-410.

[5] M. Kacira, M. Simsek, Y. Babur, S. Demirkol, Determining optimum tilt angles and orientations of photovoltaic panels in Sanliurfa, Turkey, Renewable Energy, 29 (2004) 1265-1275.

[6] M.A.M. Yakup, A.Q. Malik, Optimum tilt angle and orientation for solar collector in Brunei Darussalam, Renewable Energy, 24 (2001) 223-234.

[7] K. Ulgen, Optimum tilt angle for solar collectors., Energy Sources, 28 (2006) 1171-1180.

[8] M. Suri, T.A. Huld, E.D. Dunlop, PV-GIS: a web-based solar radiation database for the calculation of PV potential in Europe, International Journal of Sustainable Energy, 24 (2005) 55-67.

[9] NREL, Dynamic maps, GIS data, and analysis tools - solar maps., 2010.

[10] NREL, National Solar Radiation Database, 1991-2005 Update, 2007.

[11] R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, F. Vignola, A new operational model for satellite-derieved irradiances: descritpion and verification, Solar Energy, 73 (2002) 307-317.

[12] P. Ineichen, R. Perez, A new airmass independent formulation for the Linke turbidity coefficent, Solar Energy, 73 (2002) 151-157.

[13] R. Perez, P. Ineichen, E. Maxwell, R. Seals, A. Zelenka, Dynamic Global-to-Direct Irradiance Conversion Models., ASHRAE Transactions-Research Series, (1992) 354-369.

[14] S. Wilcox, National solar radiation database 1991-2005 update: user's manual, 2007.

[15] A. Nottrott, J. Kleissl, Validation of the NSRDB–SUNY global horizontal irradiance in California, Solar Energy, 84 (2010) 1816-1827.

[16] J. Page, The Role of Solar Radiation Climatology in the Design of Photovoltaic Systems, Practical Handbook of Photovoltaics: Fundamentals and Applications, Elsevier, Oxford, 2003, pp. 5-66.

[17] R. Perez, R. Stewart, C. Arbogast, R. Seals, J. Scott, An anisotropic hourly diffuse radiation model for sloping surfaces: description, performance validation, site dependency errors, Solar Energy, 36 (1986) 481-498.

[18] NREL, Measurement and Instrumentation Data Center (MIDC), 2010.

[19] B. Marion, M. Anderberg, R. George, P. Gray-Hann, D. Heimiller, PVWATTS version 2 - enhanced spatial resolution for calculating grid-connected PV-performance., 2001.

[20] NREL, PVWatts: a performance calculator for grid-connected PV systems, version 2, 2010.

Chapter 3, in full, is a reprint of the material as it appears in *Renewable Energy* volume 36, issue 3, March, 2011. Lave, Matthew; Kleissl, Jan, 2011. The dissertation author was the primary investigator and author of this paper.

4. High Frequency Irradiance Fluctuations and Geographic Smoothing

Reprinted from *Solar Energy*, Volume 86, Issue 8, Matthew Lave, Jan Kleissl, Ery Arias-Castro, "High-frequency irradiance fluctuations and geographic smoothing," Copyright 2012, with permission from Elsevier.

Abstract

Using six San Diego solar resource stations, clear-sky indices at 1-sec resolution were computed for one site and for the average of six sites separated by less than 3 km to estimate the smoothing of aggregated power output due to geographic dispersion in a distribution feeder. Ramp rate (RR) analysis was conducted on the 1-sec timeseries, including moving averages to simulate a large PV plant with energy storage. Annual maximum RRs of up to 60% per second were observed, and the largest 1-sec ramp rates were enhanced over 40% by cloud reflection. However, 5% per second ramps never occurred for a simulated 10 MW power plant. Applying a wavelet transform to both the clear-sky index at one site and the average of six sites showed a strong reduction in variability at timescales shorter than 5-min, with a lesser decrease at longer timescales. Comparing these variability reductions to the Hoff and Perez [1] model, good agreement was observed at high dispersion factors (short timescales), but our analysis shows larger reductions in variability than the model at smaller dispersion factors (long timescales).

4.1. Introduction

The variable nature of solar radiation is a concern in realizing high penetrations of solar photovoltaics (PV) into an electric grid. High frequency fluctuations of irradiance caused by fast moving clouds can lead to unpredictable variations in power output on short timescales. Short-term irradiance fluctuations can cause voltage flicker and voltage fluctuations that can trigger automated line equipment (e.g. tap changers) on distribution feeders leading to larger maintenance costs for utilities. Given constant load, counteracting such fluctuations would require dynamic inverter VAR control or a secondary power source (e.g. energy storage) that

could ramp up or down at high frequencies to provide load following services. Such ancillary services are costly to operate, so reducing short-term variation is essential. Longer scale variations caused by cloud groups or weather fronts are also problematic as they lead to a large reduction in power generation over a large area. These long-term fluctuations are easier to forecast and can be mitigated by slower ramping (but larger) supplementary power sources, but the ramping and scheduling of power plants also adds costs to the operation of the electric grid. Grid operators are often concerned with worst-case scenarios, and it is important to understand the behavior of PV power output fluctuations over various timescales.

Many previous studies have shown the benefit of high-frequency irradiance data. Suehrcke and McCormick [2] and Gansler et al. (1995) [3] found 1-min data to have different statistics from lower-frequency data, including a much more bi-modal distribution than 1-hour or 1-day data. Gansler et al.[3] mention that while using 1-hour data may be acceptable for space and water heating systems, where the thermal capacitance effects dampen out short-term variations, the time response of PV systems is much faster and using 1-hour data will likely lead to errors.

Understanding that high-frequency fluctuations are important, further studies have looked to characterize these fluctuations, often by comparing fluctuations at one site to fluctuations at the average of multiple sites. Otani et al. [4] use a fluctuation factor defined as the root mean squared (RMS) value of a high-pass filtered 1-min time series of solar irradiance to demonstrate a 2-5 times reduction in variability when considering 9 sites located within a 4 km by 4 km grid. Curtright and Apt [5] and Lave and Kleissl [6] used 1-min timeseries to show reductions in the mean, maximum, and standard deviation of ramp rates (RRs) when considering the average of three or four sites versus only one site. Power spectral densities (PSDs) presented in Otani et al. [4], Curtright and Apt [5], and Lave and Kleissl [6] all show strong reductions in power content of fluctuations of the average of multiple sites versus the power content of fluctuations at one site.

Lave and Kleissl [6] also present coherence spectra which show that the sites in Colorado which were 60 km or more apart were uncorrelated on timescales shorter than 12-hours. Two sites that were only 19 km apart were uncorrelated on timescales shorter than 3-hours.

Wiemken et al. [7] used 5-min normalized output from 100 PV sites spread throughout Germany. They found the standard deviation of the average of 100 sites to be 0.61 that of 1 site for the month of June, and that 5-min fluctuations of ±5% of power output at nameplate capacity are virtually nonexistent in the average, yet single sites have fluctuations larger than ±50%. Also included in that paper is a figure from Beyer et al. [8], which shows exponential decay of cross-correlation as a function of distance for hourly irradiance data from six sites in Germany which can be used to estimate the reduction in standard deviation when averaging sites. Murata et al. [9] analyzed 1-min data from 52 PV systems spread across Japan to determine the "smoothing effect" of aggregating multiple systems. The authors introduce a fluctuation index, which is the maximum difference in aggregated power output over a given time interval. They found that over 1-min, sites more than about 50-100 km apart were uncorrelated and thus that there was a limit reached whereby adding more PV sites had no effect on reducing variability, since the variability introduced by the diurnal cycle eventually becomes larger than the cloud-induced variability. For times greater than 10-min, however, they reject the hypothesis that sites within 1000 km are independent, though some of the dependence may be due to diurnal solar cycles and could be eliminated by using a normalized solar radiation.

Hoff and Perez [1] (hereafter HP10) present a framework to estimate the decrease in standard deviation of irradiance achieved by aggregating PV sites. The reduction in standard deviation is a function of the number of PV sites and a dispersion factor, D, defined as the number of time intervals it takes for a cloud to pass over all PV sites across the region being considered. The dispersion factor is useful in determining when the transition from PV sites being uncorrelated to correlated occurs. They predict a factor of $\sqrt{N}$ reduction in standard deviation of

the average of N sites compared to the standard deviation of one site for the "spacious region," where the number of sites is much less than the dispersion factor, $N << D$. This corresponds to the sites being fully independent of one another, and is a known result from statistics on independent random variables. At an "optimal point" where the number of sites equals the dispersion factor, $N = D$, they derive a factor of N reduction in standard deviation. At this point, sites would be perfectly correlated when shifted by an appropriate timestep (the dispersion factor divided by the number of sites), and the standard deviation of the average of all sites will be reduced more than would be expected if the sites were entirely independent. HP10 also define a "limited region," between the "spacious region" and the "optimal point," $N < D$. In the "crowded region" where the number of sites is larger than the dispersion factor, $N > D$, they propose that the standard deviation will be reduced by a factor of D, since the sites are at least partially dependent, and adding additional sites will not reduce the standard deviation of the average since the reduction is only a function of the dispersion factor, and not the number of sites. HP10 perform a limited model validation by simulating a fleet of PV systems based on measured irradiance at only one site. In the simulated system, the irradiance at the non-measured sites was found by shifting the timestamp but otherwise maintaining the measured data, thus simulating an equally spaced system in which frozen clouds move at a constant speed along a line containing all sites.

Woyte et al. [10] present a unique study in that they use very high frequency data (1-sec, 5-sec, or 1-min depending on the site) collected for up to 2-years, instantaneous clearness index, and a wavelet transform to analyze fluctuations of all scales in time, from very short to very long. The Haar wavelet was applied to each clearness index dataset to detect fluctuations over various timescales. They introduce a fluctuation power index, which is the sum of the square of the wavelet mode at each timescale, and is used to quantify the amplitude and frequency of occurrence of fluctuations on a specific timescale. In other wavelet studies, Kawasaki et. al [11]

applied the Daubechies 4 wavelet to 1-min irradiance 2-year timeseries from nine sites in a 4x4 km grid, and Perpinan and Lorenzo [12] applied the MODWT wavelet to 1-sec solar irradiance timeseries from a few days in October, 2009.

This paper builds on these previous works by using 1-sec clear-sky index (K_c) data from 6 sites on a microgrid similar to urban distribution feeders (Section 4.2) to quantify extreme ramp rates (RRs). Methods are described in Section 4.3. RRs were analyzed by computing statistics at different time steps and by using varying moving average intervals to represent large PV plants or storage (Section 4.4.1). Coherence spectra are employed in Section 4.4.2 to analyze the correlation between six sites at different time scales. We apply a wavelet to detect variability over various timescales relevant to the operation of a power grid (Section 4.4.3). Wavelet analysis allowed for a localized study of the power content of variations over various timescales. The power content of variations at one site was compared to the power content of variations at the average of six sites in close proximity to study the reduction in variability over various timescales achieved by using multiple site locations (Section 4.4.4) and to test the model of HP10 (Section 4.4.5). Conclusions are presented in Section 4.5.

4.2. **Data**

Global Horizontal Irradiance (GHI) was recorded once per second at sites throughout the University of California, San Diego (UCSD) campus as part of the UCSD Decision Making using Real-time Observations for Environmental Sustainability (DEMROES) network of sensors [13] (Fig. 1). All sites employ a LICOR Li-200SZ silicon pyranometer sampling at 1Hz. The collection of 1-sec data proved to be a challenge of both data storage on the datalogger and sensor reliability, and so data availability is inconsistent. While there are 8 sites maintained as part of the DEMROES network, at any given time a maximum of 6 sites recorded 1-sec data.

The main site used in this paper was the Engineering Building II (EBU2, 32.8813^0N, 117.2329^0W), for which data was available for all of 2009 except for May 23 through June 4. We

do not expect for the 13 days of missing data to lead to a strong seasonal or other bias, and therefore will refer to this as 1-year of data. Five other sites also recorded data from July 31 to August 25, 2009, and are used to study the benefits of aggregating sites. These five sites ranged from Hubbs Hall (HUBB, 32.8670^{0}N, -117.2533^{0}W) which is 0.1 km from the Pacific Ocean to Moores Cancer Center (MOCC, 32.8782^{0}N,-117.2229^{0}W) 3km to the east. The Biomedical Science Building (BMSB, 32.8758^{0}N, 117.2362^{0}W), RIMAC Arena (RIMC, 32.8852^{0}N, 117.2402^{0}W), and Tioga Residence Hall (TIOG, 32.8790^{0}N 117.2434^{0}W) are more centrally located (Fig. 4.1). Since these data covers nearly the entire month of August, we will refer to this as a 1-month dataset.

After applying the factory calibration, clear days were used (assuming identical atmospheric composition) to create linear fits against RIMC, and each site was cross-calibrated by this linear fit. In addition, careful quality control was carried out by visually examining each site for shading and other errors. We noticed 5 occurrences at EBU2 and 15 at RIMC of greater than 400 W m^{-2} decreases lasting less than 3-sec during otherwise clear periods. We suspect that these were due to birds or airplanes instantaneously shading our sensors, and not cloud effects which were the focus of this study. Therefore, such dips were removed from the data using a linear interpolation plus characteristic variance to maintain statistics, which should be appropriate over such a short time. It is possible that such events also occurred during cloudy conditions, but is not clear how to detect such events then without possibly removing real variability.

To eliminate the deterministic effect of diurnal cycles, GHI measurements (in W m^{-2}) were converted into a dimensionless clear-sky index by dividing the measured GHI by the clear-sky irradiance. We used the Sunny Days model [14] based on Long and Ackerman [15], which uses input GHI and diffuse horizontal irradiance (DHI, measured by a Dynamax SPN1 pyranometer at EBU2) to calculate clear-sky irradiance. Since Sunny Days is locally calibrated day-by-day it was found to be more accurate than the Ineichen and Perez [16] climatological

clear-sky model especially for mornings and evenings. Times when the solar altitude angle was less than 10° were removed to eliminate both nighttime values when the clearness index is expected to always be zero and early morning and late evening periods when the pyranometer is subject to errors in cosine response.

Fig. 4.1: Map of the UCSD solar resource sites, showing proximity to the Pacific Ocean (left), and Interstate 5 (center). From EBU2, distances and headings are: BMSB(0.69km, 205^0), HUBB (2.47km, 230^0), MOCC (0.95km, 105^0), RIMC (0.80km, 300^0), and TIOG (1.00km, 255^0). Map © 2010 Google – Image © 2010 TerraMetrics.

4.3. **Methods**

4.3.1. 1-Year Analysis at One Site: Ramp Rate Analysis

The frequency of occurrence and magnitude of RRs of solar PV are of critical interest to power system operators. From the 1-sec clear-sky indices, we can extract two different averages which have different practical relevance.

First, block averages were taken on time intervals varying from 1-sec to 1-hour, which shows the difference in statistics over various data averaging intervals. Typically irradiance or power output data are averaged over longer periods and our analysis allows comparison to such data. The block average method produces fewer data points as the block size increases.

Second, *moving* averages over intervals of $T = 2^j$ sec ($j = 1,2, \ldots 12$ corresponding to $T = 2,4, \ldots 4096$ sec) were computed at time steps of 1-sec such that the average at any given time, *t*, is the average of values $(2^j - 1)$ seconds before and $2j$ seconds after *t*. Intervals of 2^j seconds were chosen to be consistent with wavelet analysis presented later. No moving average was computed when $TOD < 2^j - 1$ *or* $\max(TOD) - TOD < 2^j$, where TOD is the time of day, such that the moving average would not be complete. Since this clipping removes more data at the beginning and end of the day for larger j reducing the available data, no analysis was performed for timescales larger than 4096-sec. Moving averages at different T are representative of power sampled every second, but averaged spatially over the dimensions of a solar power plant or by using energy storage.

From either the block average or the moving average, RRs were computed as the difference between successive clear-sky indices. Cumulative distribution functions (cdf) of RRs show the statistical distributions and extreme values. Additionally, we computed the averages of 1-sec ramps with magnitude greater than 0.25 to show the typical behavior before large positive or negative ramps. RRs of clear sky indices give the percent change (as a fraction of clear-sky irradiance) over one timestep, regardless of the TOD when that change occurred. The clear sky index provides the best measure to compare cloud induced solar variability analyses between different sites. If the occurrence of clouds is independent of TOD, it also provides the most relevant measure to characterize solar energy variability at a site, especially for 2D tracking power plants (whose output fluctuates less over a clear day). However, if clouds occur preferentially over certain TODs and a fixed-tilt plant is considered, then clear sky index variability does not translate directly to power output variability of a PV plant.

4.3.2. Geographic smoothing at six sites over one month

4.3.2.1. Coherence spectrum

As a measure of spatial correlation of the clear-sky index over various time scales, we calculated the coherence spectrum between EBU2 and the other 5 sites. The coherence spectrum provides normalized covariance at each frequency, allowing for visualization of correlation over various timescales. The coherence is expected to be large at long timescales as large weather systems will lead to similar clear-sky indices for all the sites. Note, however, that solar cycles have been removed by using the clear-sky index and thus the coherence will not be as large as if irradiances had been used. The timescale at which sites become weakly correlated is an indication of the longest timescale on which the sites are nearly independent and will dampen aggregate variability. Although negative correlation would reduce variability more than zero correlation, negative correlation is not expected physically.

4.3.2.2. Wavelet Analysis

The stationary or dyadic wavelet transform, *W*, of a signal $x(t)$ is (Mallat [17]):

$$W_{2^j}^{\tau} = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right) dt \qquad 4.1$$

where t is time, τ is the time offset from the beginning of the day, ψ is the wave used to produced the wavelet transform, and s is the scaling factor. Since we used a real wavelet and a discrete transform, we required that j be a positive integer. Although Woyte et al [10] used the Haar wavelet to detect dips and compute power content at each timescale, we found the Haar wavelet to be lacking in that large wavelet coefficients exist only at sharp signal transitions. This means that changes from one state to another (e.g. a step from cloudy to clear) are detected by the Haar wavelet rather than the duration of an up or down fluctuation (a top hat).

We instead chose to employ the top hat wavelet as the basis function of our analysis of clear-sky index timeseries. The top hat wavelet is defined as

$$\psi(t) = \begin{cases} 1, & \frac{1}{4} < t < 3/4 \\ -1, & 0 < t < \frac{1}{4} \quad || \quad \frac{3}{4} < t < 1 \\ 0, & else \end{cases} \tag{4.2}$$

and is shown in Fig. 4.2. Although typically in wavelet analysis $s = 2^j$ is used in Eq. 4.1, we chose to define $s \equiv 2^{j+1}$ instead, so Eq. 4.1 becomes

$$W_{2^j}^{\tau} = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{2^{j+1}}} \psi\left(\frac{t-\tau}{2^{j+1}}\right) dt. \tag{4.3}$$

This altered definition allows for the timescale, 2^j seconds, to describe the duration of the clear or cloudy period of interest rather than the duration of the entire wavelet. Substituting the clear-sky index, $x(t) = K_c(t)$, into Eq. 4.3 will result in a separate timeseries $w_j(\tau)$ for each j value (mode), where $w_j(\tau)$ is defined such that

$$w_j(\tau) = \int_{-\infty}^{\infty} K_c(t) \frac{1}{\sqrt{2^{j+1}}} \psi\left(\frac{t-\tau}{2^{j+1}}\right) dt\,. \tag{4.4}$$

Just as for the moving averages, we chose to limit our analysis to $j \leq 12$ (corresponding to 1.1 hours or less), and only τ values for which data were available over the entire interval of size 2^{j+1} around τ were retained. As such, early morning and late evening periods are not resolved at the longer modes. It is common in wavelet transforms to extend the original timeseries using either a periodic extension or zero padding, but we feel that neither is appropriate in this situation as they will both introduce effects that were not present in the original timeseries.

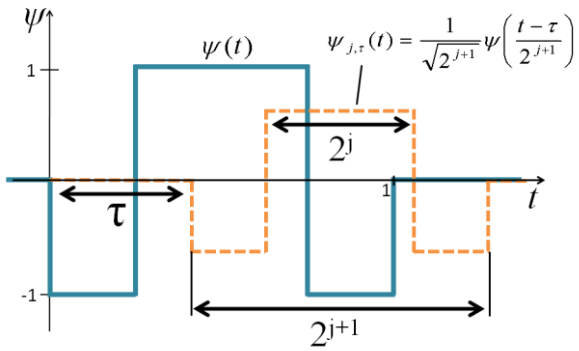

Fig. 4.2: Top hat wavelet $\psi(t)$ (solid line) and the scaled and translated wavelet $\psi_{j,\tau}(t)$ (dashed line). This scaled wavelet would capture a clear period of duration 2^j bordered by cloudy periods.

The power content of each timeseries $w_j(\tau)$, can be found by calculating the wavelet periodogram *I*. Following the definition of the Fourier periodogram, the wavelet periodogram is the square of the coefficients of the wavelet transform, normalized by the length over which the wavelet was applied, which in this case is 2^{j+1}:

$$I_j(\tau) = \frac{1}{2^{j+1}} \left| w_j(\tau) \right|^2 \quad 4.5$$

4.3.2.3. Application of Wavelet Analysis to Determine Reduction in Variability from Averaging 6 Sites

The wavelet periodograms are still timeseries, and are difficult to examine visually for periods longer than one day. Therefore, we use the 'fluctuation power index,' as described by Woyte et al. (2007) to quantify the power contained in fluctuations at each timescale. The fluctuation power index, fpi, is:

$$\text{fpi}(j) = \frac{1}{T_j} \int_0^{T_j} I_j(\tau) d\tau, \quad 4.6$$

where T_j is the length of the timeseries $I_j(\tau)$, which decreases as j increases due to unresolved periods of the higher modes. Using T_j instead of a constant value based on the length of the original $K_c(t)$ timeseries means that fpi(j) is an average value, which allows for comparison of fpi at different *j* values.

The fpi is essentially variance at each timescale (is variance if $\text{mean}(w_j(\tau)) = 0$), so we used fpi to evaluate the reduction in variability achieved by averaging six sites versus the variability at EBU2 alone and to compare our results to the HP10 model. HP10 define a dispersion factor $D = \frac{L}{V\,T}$, where L is the length of the region with the PV sites, V is the cloud velocity, and T is the relevant timescale. Although L and V remain constant for a given area and time, varying the timescale changes D. Since variability at multiple timescales was calculated through the fpi, we were able to test the HP10 model over various dispersion factors for the 1-month data.

4.4. **Results**

4.4.1. Ramp Rate Analysis

The cdf of the absolute value of step sizes (SS) for K_c averaged over blocks of 1-sec, 10-sec, 1-min, 10-min, and 1-hr simulating data averaged over and sampled at those intervals are shown in Fig. 4.3. The probability of occurrence of SSs greater than 5%, 10%, and 25% are shown in Table 4.1. Both Fig. 4.3 and Table 4.1 show SS statistics vary significantly over all timescales, which is consistent with previous findings that 1-min and 1-hr data have different statistics (i.e., Suehrcke and McCormick [2], Gansler [3]et al). These variations in statistics of SSs down to 1-sec show the importance of sampling data as frequently as possible when studying irradiance fluctuations. Large step sizes have a much greater probability of occurring when using 1-hr averages than when using 1-sec averages. However, due to the nature of block averaging, at longer time intervals, the sample size is small and events with high probabilities of occurrence do not happen very often in a day (Table 4.1). Still, the cdf of SSs shows a trend toward SS magnitude decreasing as the averaging time decreases – short-time steps will not be as extreme as long-time steps.

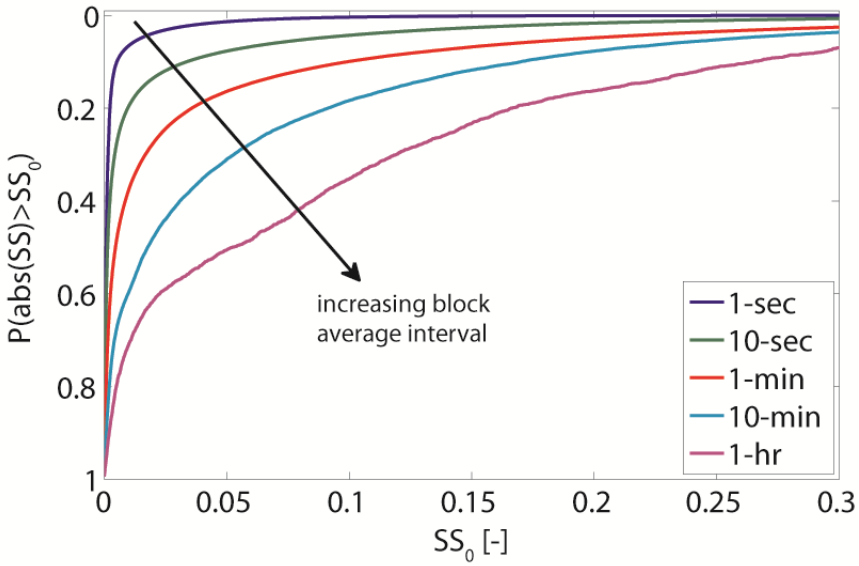

Fig. 4.3: Cumulative distribution function of SSs for block averages over 1-sec to 1-h at EBU2 for 2009. The probability of occurrence of a certain SS (or larger SSs) can be determined by locating the SS on the x-axis and going up to intercept the line of the desired block averages. The y-value at that point provides the probability. For example, a 25% SS for 1-h block averages occurs 11% of the time or about once per day, on average. The 1-hr curve is based on a smaller sample (the number of 1-hr blocks contained in the 1-year of data) compared to the other curves. For example, the 10-min curve is based on a sample that is 6x larger. This explains why the 1-hr curve is slightly different in shape and slightly more irregular than the other curves.

Table 4.1: Probabilities of SSs larger than 10%, 25% or 50% at each timescale of block averages along with approximate number of occurrences per day. Occurrences per day were found using an estimated annual average of 10-hours per day when solar altitude angle is greater than 10^0.

Block average interval	abs(SS)>0.10		abs(SS)>0.25		abs(SS)>0.50	
	P(abs(SS)>0.10)	#/day	P(abs(SS)>0.25)	#/day	P(abs(SS)>0.50)	#/day
1-sec	0.37%	132	0.02%	6.3	0.0002%	0.1
10-sec	4.29%	155	1.07%	38.4	0.10%	3.5
1-min	9.96%	59.8	3.48%	20.9	0.63%	3.8
10-min	18.39%	11.0	5.26%	3.2	0.85%	0.5
1-hr	35.22%	3.5	11.23%	1.1	0.91%	0.1

While block averages represent sampling data at certain periods where the actual variability is unaffected, moving averages can be used to simulate the effects of fast-ramping energy storage (e.g. flywheels). If the length of the moving average is equal to the time over which energy storage has the capacity to eliminate fluctuations through charging or discharging, then a moving average timeseries will be representative of the PV + storage output timeseries. Moving averages are also relevant to simulating power output of a large PV array or a fleet of PV

sites that all sit along the cloud motion vector and are spaced evenly, as used in HP10. In this case, a longer moving average interval will simulate the output of a larger PV plant, since large systems will ideally average over a timescale of $A^{1/2} / V$, where $A^{1/2}$ is the square root of the area of the array and V is the cloud velocity. Moving averages at various timescales are shown in Fig. 4.4 for August 22, 2009.

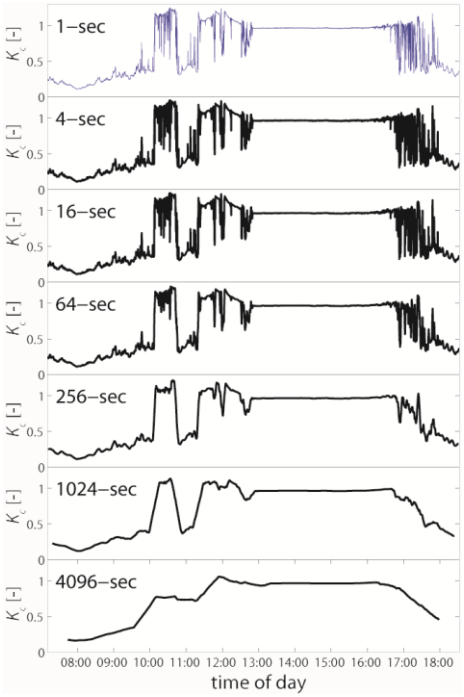

Fig. 4.4: Moving averages of the clear-sky index, K_c, over various averaging intervals for EBU2 on August 22, 2009.

The cdf of RRs for various moving averages is shown in Fig. 4.5, and specific values are shown in Table 4.2. For the moving averages, increasing the averaging time decreases the probability of a large ramp. For example, for a 4096-sec (about 1-hour) moving average, the probability of a ramp larger than 0.1% s^{-1} is zero. This is intuitive, since the change in the moving average is the change in the step size divided by the averaging interval. Since a 1-sec average under both the block and moving averages simply represents the original timeseries, the 1-sec cdf which appears in both Figs. 4.3 and 4.5 and Tables 4.1 and 4.2 serves as a reference for comparison between the two averaging methods. To create the power plant size column Table 4.2,

we assumed that the 1-sec data was representative of the fluctuations of a typical household PV installation of 2.5kW. Then, using the $A^{1/2}\ /\ V$ relation mentioned earlier, we determined the relationship between moving average intervals and PV plant sizes. This assumed a frozen cloud field traveling at a constant speed, V, over the entire PV plant. While this is unlikely physically, it givens an indication of the best-case scenario and allows for a comparison of fluctuations over various PV plant sizes.

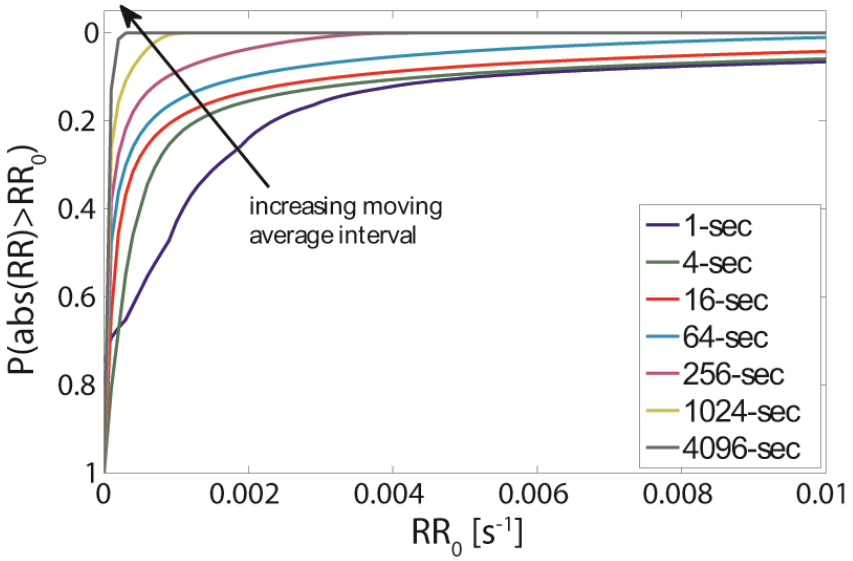

Fig. 4.5: Cumulative distribution function of 1-sec RRs and RRs of moving averages over various timescales (representing large PV plants or plants with energy storage) at EBU2 for 2009. The 1-sec value at $RR_0 = 0$ is 0.75 and not 1.0 due to the very small changes that can occur over 1-sec resulting in $RR < 0.0001$. For all other timescales, $RRs < 0.0001$ never occur.

Table 4.2: Probabilities of RRs exceeding 0.1%, 1%, or 5% s^{-1} at moving average timescales along with approximate number of occurrences per day. Occurrences per day are based on a 10 sunlight-hour day.

Moving average interval	Power plant size	RR>0.001 s^{-1}		RR>0.01 s^{-1}		RR>0.05 s^{-1}	
		P(RR>0.001)	#/day	P(RR>0.01)	#/day	P(RR>0.05)	#/day
1-sec	2.5 kW	42.98%	15,472	6.55%	2,359	1.35%	486
4-sec	40 kW	23.57%	8,486	5.90%	2,125	0.81%	292
16-sec	640 kW	19.53%	7,031	42.98%	1,511	0.04%	15
64-sec	10.2 MW	15.46%	5,564	1.03%	370	0%	0
256-sec	164 MW	8.84%	3,181	0%	0	0%	0

In order to examine the typical behavior leading up to and after the largest 1-sec ramps, Fig. 4.6 displays the mean (or conditional average) of all 1-sec ramp events greater than 25%, separated into positive and negative ramps. An 'ideal' ramp would simply be a step function from a small K_c to a large K_c or vice versa. However, in practice K_c is variable before or after large ramps as the clear or cloudy period before or after the ramp is often shorter than one minute. For the negative (or clear to cloudy) ramp, there is successive enhancement in clear-sky index in the 1-min before the ramp. This is a manifestation of short clear periods but also of cloud edge enhancement; as a cloud nears the path between the sun and the sensor, some sunlight is reflected off the near edge of the cloud and down to the sensor, while the sun-sensor path is mostly unobstructed. Cloud enhancement leads to irradiances larger than the clear-sky model due to additional diffuse irradiance, resulting in a clear-sky index greater than 1 (Fig. 4.6). A similar but opposite behavior is observed for the up-ramp. The change in mean clear-sky index from one minute before a large negative ramp to one minute after is about 10%, which indicates a change of state from clear to cloudy. For large positive ramps, this change is only about 3%, and so represents a much smaller change in average state of the sky.

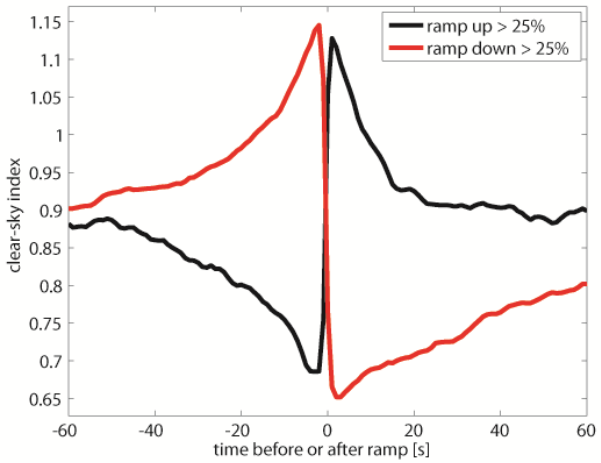

Fig. 4.6: Means of all ramps at EBU2 in 2009 that were greater than 25% s^{-1}, separated into positive and negative ramps. The red line shows the mean of 1006 timeseries starting 1-min before and ending 1-min after a ramp that was more than a 25% s^{-1} decrease in clear-sky index. The black line shows the mean of 511 such timeseries that were centered around a greater than 25% s^{-1} increase in clear-sky index.

Over the entire year, there were five 1-sec ramps up (probability of 2.1×10^{-7} s^{-1}) and 17 1-sec ramps down (probability of 7.3×10^{-7} s^{-1}) with magnitudes greater than 50%. The maximum up ramp was 58% s^{-1} and maximum down ramp was 59% s^{-1}. Thus, as an absolute worst case scenario, a maximum change of 60% over 1-sec can be assumed. The worst irradiance fluctuations were 432 W m^{-2} for an up ramp (June 5, 14:01:42) and 516 W m^{-2} for a down ramp (April 15, 13:33:42), which corresponded to 45% and 54% clear-sky index ramps, respectively. We emphasize, however, that this applies only for one point sensor, and when sites are averaged or PV arrays are considered, these maximum ramps are expected to be strongly reduced.

4.4.2. Coherence spectra

The coherence spectra over 1-month showing the coherence between EBU2 and the other 5 sites are shown in Fig. 4.7. At long timescales, the coherence spectra all approach 1. This is expected since hourly and longer weather phenomena such as changes in synoptic cloudiness and atmospheric composition changes affect all sites. Since the coherence spectra were calculated using clear-sky indices, the spectra do not approach 1 as quickly as would be expected with irradiances since the daily cycle of the sun rising and setting is (mostly) removed. The sites are uncorrelated for time scales shorter than 10 min. BMSB, RIMC, and TIOG have the highest coherence values against EBU2 at long timescales. HUBB and MOCC have consistently lower coherence values for timescales longer than 10-min. While it is expected that HUBB will have lower coherence due to it being at the coast and more than twice as far away from EBU2 than the other sites, it is somewhat surprising that MOCC also has such low coherence. MOCC (~1km ESE) and TIOG (~1km WSW) are almost the same distance away from EBU2, albeit in nearly opposite directions, and yet the coherence spectra for each is markedly different. This indicates different weather patterns to the west of EBU2 as to the east. Anecdotal sky observations have confirmed that clouds often evaporate as they move eastward which would result in a smaller coherence.

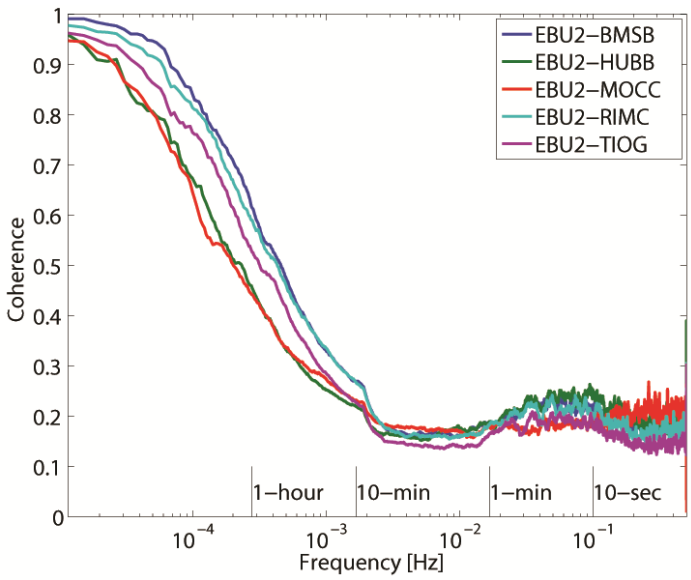

Fig. 4.7: Coherence spectra for EBU2 and each of the other 5 sites for July 31 through August 25, 2009. Each spectrum is smoothed by a moving average smoothing filter for clarity. The apparent vertical lines at the far right of the plot are fluctuations that were not dampened by the smoothing filter since they are too close to the boundary. Different time scales are marked through vertical lines.

4.4.3. Wavelet decomposition

Wavelet periodograms were computed from the clear-sky index for EBU2 as well as from the clear-sky index for the average of 6 sites for each timescale, $j = 1$ to 12 for the month when 6 sites were simultaneously available. The periodograms from August 22, 2009 over modes $j = 6$ (about 1-min) to $j = 12$ (about 1-hr) are shown in Fig. 4.8. August 22 was chosen because it has both cloudy and clear periods and because it has a distinct clear period followed by a distinct overcast period both lasting about 30-min. This serves as a validation of our application of wavelets, as we expect this period to produce two peaks at the $j = 11$ mode (34-min). Indeed, the most distinct peaks in the wavelet periodogram shown in Fig. 4.8 are on the $j = 11$ mode, and occur at about 10:30 and 11:00. We can also see from the periodogram that the dominant timescale of fluctuations between 16:30 and 18:00 was 256-sec ($j = 8$). This was not obvious by inspecting the original timeseries, but rather is a useful result found through wavelet decomposition.

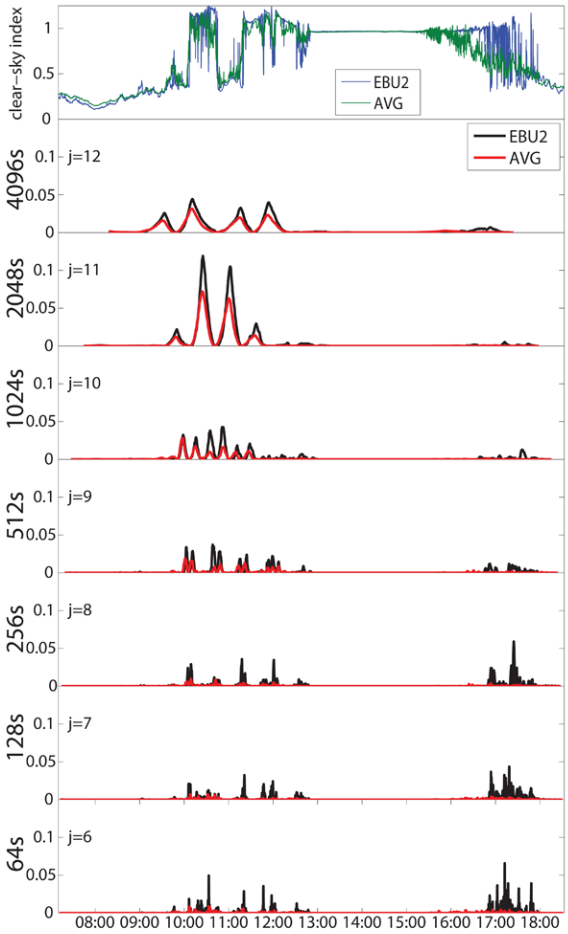

Fig. 4.8: Clear-sky index (blue and green thin lines) and wavelet periodogram (black and red thick lines) of modes $j = 6$ through $j = 11$ for EBU2 and the average of all 6 sites on August 22, 2009.

Inspection of the wavelet periodogram shows that the amplitude is only slightly reduced for the average versus EBU2 at high modes ($j \geq 10$), but the average amplitude is much smaller at modes corresponding to shorter timescales. Since the amplitude of the periodogram at each scale is the variance at that scale, this allows quantifying how averaging multiple sites will lead to a stronger reduction in variability at shorter timescales.

4.4.4. Fluctuation power index

The reduction in variability as a function of timescale due to averaging sites for the 1-month period is shown in Fig. 4.9, by plotting the fpi for each timescale. Fig. 4.9 also shows the

ratio fpi_{EBU2}/fpi_{AVG}, which we will call the variability ratio (VR), for each timescale. The VR is a measure of the reduction in the power (or variance) of fluctuations. A higher VR means a larger reduction in fluctuations, while a variability ratio of 1 means no reduction in variability compared to a single site. For timescales shorter than 256s (about 4-min), VR was close to 6 for the average of the 6 sites. This is consistent with the factor of 6 reduction in variance that we would expect for 6 sites spread far enough apart such that their clear-sky indices can be considered independent of one another (or uncorrelated). At timescales longer than 128s, the fpi ratio decreased in an exponential fashion as the sites become more and more correlated. Eventually, at 4096-sec, the VR was nearly one, indicating that on timescales longer than 1-hour, the clear-sky indices at these 6 sites are too correlated to cause significant reductions in variability.

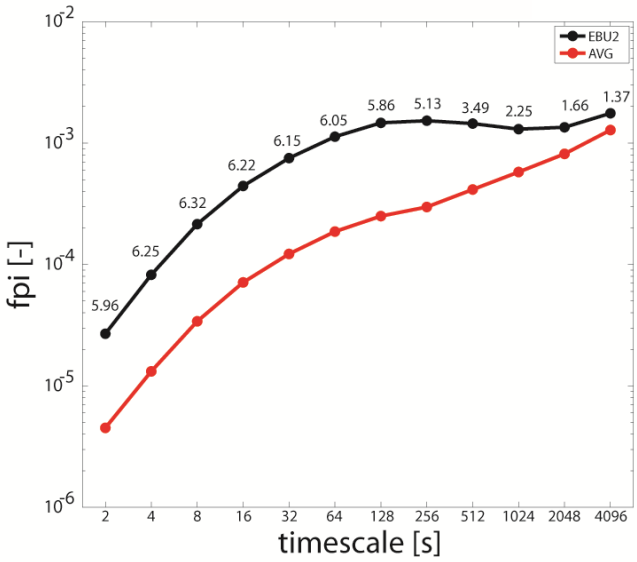

Fig. 4.9: Fluctuation power index for EBU2 and the average of 6 (AVG) sites over 1-month. The numbers above the EBU2 black line are the ratio of fpi_{EBU2}/fpi_{AVG} for each timescale.

4.4.5. Comparison to Hoff and Perez model

The VR at each timescale was used to test the Hoff and Perez [1] (hereafter HP10) theoretical model for the decrease in variability at various dispersion factors D for our six sites. To compute D, cloud speed V and the distance L that clouds must travel in the direction of cloud

movement to pass over all sites are required. We estimated distance by assuming that clouds typically travel from west to east over our campus, which was consistent with sky imagery analysis. The east-west distance between HUBB and MOCC, the furthest west and the furthest east sites, respectively, is 2.75 km. We assumed a 5 m s^{-1} cloud velocity given analysis from total sky imagery (Chow et al.[18]), such that $D = \frac{L}{V\,\Delta t} = \frac{550s}{\Delta t}$, with $\Delta t = 2^j$. Since HP10 modeled the ratio of *standard deviation* at the average of all sites to the standard deviation at one site ("relative output variability," ROV), for comparison we take the inverse of the square root of the VR (Fig. 4.10). In order to plot a single curve for the HP10 ROV, we assumed the lower bound ($ROV = \frac{1}{\sqrt{6}}$) for the reduction in variability in the "limited region." At long timescales (in the "crowded region", expected $ROV = \frac{1}{D}$), D became less than 1 and would have led to an increase in variability at the average versus just one site. This does not make physical sense, and thus the ROV was capped at 1.

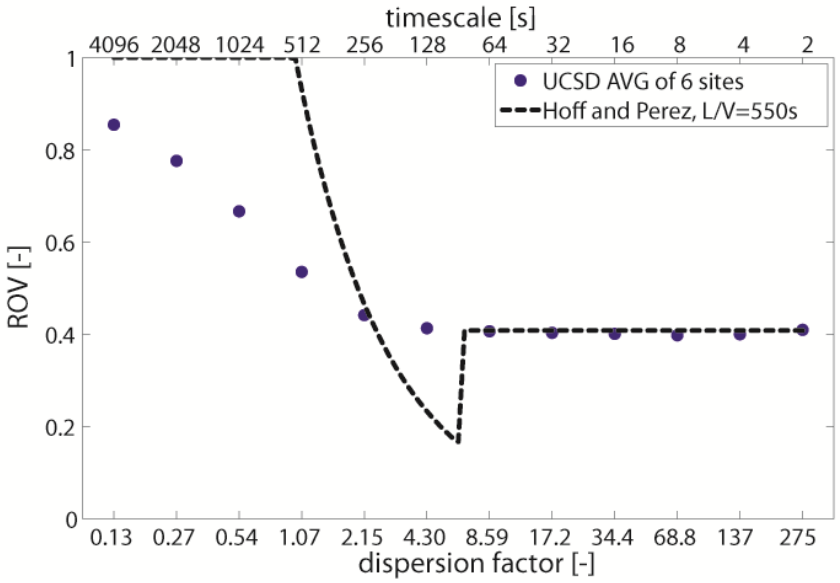

Fig. 4.10: Relative output variability (ROV) for various dispersion factors calculated at each of the timescales used in the wavelet decomposition compared to the Hoff and Perez (2010) model. Hoff and Perez model values greater than 1 were set equal to 1, and the lower limit of $1/\sqrt{6}$ for the "limited region" was used. AVG is the average of 6 sites.

At large D corresponding to short timescales, the sites are indeed independent, and our variability reductions fit the HP10 model. However, the agreement is worse at smaller D corresponding to longer timescales. Our data show no signs of an optimal point in the "limited region". In calculating the optimum point, HP10 assumed that each pair of sites was perfectly correlated (with some timeshift), but we suspect that small lateral dimensions of clouds and cloud condensation / evaporation cause each site to experience a different cloud pattern. Even for two sites only 10 m apart (not shown in Fig. 4.1), on a highly variable day (November 4, 2008) we calculated a maximum time-shifted correlation of only 0.94 (4-sec time shift). In the "crowded region" the ROV increases with increasing time scale, but the HP10 increase is much sharper than the increase found in our data, again presumably due to decorrelation in the cloud fields. Overall, our data indicate a model for the ROV that starts at 1 for very long timescales and then follows an exponential decay that asymptotically approaches $1/\sqrt{N}$ for very short timescales.

4.5. **Conclusion**

Clear-sky indices at 1-sec resolution from a solar resource station at UC San Diego were used to compute statistics over an entire year, showing the benefits of high-frequency data to obtain accurate irradiance statistics. Two types of averages were applied to the 1-year data: a block average representing different temporal resolutions of averaged solar radiation data, and a moving average. For the block averages, the probability of a large SS increases with increasing averaging time, but the number of occurrences per day decreases. This was consistent with Fig. 4.3 in Mills and Wiser [19] where 10-min and 1-hour block averages had increasingly higher probabilities at any given ramp than the 1-min block average. This same trend was also seen down to 1-sec resolution for one day in Fig. 4 in Mills et. al [20].

With moving averages, the probability and number of occurrences per day of large RRs both decrease with increasing averaging interval. Using storage with a certain capacity solely for

RR reduction will result in a PV power output similar to the moving average over an interval equivalent to the storage-time capacity, and will therefore lead to a reduction in both magnitude and occurrence of extreme RRs. Likewise large solar systems will ideally average irradiance over a time scale of $A^{1/2} / V$, where $A^{1/2}$ is the square root of the area of the array (which is a proxy for array length in the direction of the cloud velocity V) and a frozen cloud field with infinite lateral dimension is assumed (similar as in HP10). This leads to RR reductions at short time scales.

A new framework for quantifying geographic smoothing was presented through coherence spectra and wavelet analysis of 1-sec K_c data over one month at six sites within 3 km distance. Comparisons of clear-sky indices showed the reduction in variability when averaging 6 sites. Coherence spectra indicated strong correlations between sites up to timescales as short as 10-min and the correlation at longer timescales decreased with distance but not isotropically. A top hat wavelet transform was applied to both EBU2 and the average of the 6 sites. At timescales of 256-sec and shorter, there was approximately a factor of 6 reduction in variance for the average versus just the one site. This indicates that all 6 sites were independent at timescales shorter than about 5-min, reducing variability. At longer timescales, however, the reduction in variance became smaller, and there was almost no reduction in variance at 4096-sec. The variability reductions were compared to the theoretical model presented by HP10. While there was strong agreement between the model and our data at high dispersion factors (short timescales), there was good qualitative but poor quantitative agreement at smaller dispersion factors (long timescales). We suspect that the HP10 assumption that sites are perfectly correlated when time-shifted does not apply for real data. Instead, our data suggests an exponential decay of the relative output variability as a function of dispersion factor.

As PV gains higher and higher penetration, it is important to understand the typical fluctuations on various timescales, as well as the potential for storage, PV array size, and

geographic dispersion to dampen these fluctuations. The top hat wavelet transform is a novel approach to deconstruct clear-sky indices into separate timescale components, and was very useful in determining the benefits of storage and geographic averaging. However, cloud statistics depend on local meteorological conditions so more tests in different regions are required to ensure applicability of our results.

Acknowledgments

We appreciate funding from the DOE High Solar PV Penetration grant 10DE-EE002055.

References

[1] T.E. Hoff, R. Perez, Quantifying PV power Output Variability, Solar Energy, 84 (2010) 1782-1793.

[2] H. Suehrcke, P.G. McCormick, Solar radiation utilizability, Solar Energy, 43 (1989) 339-345.

[3] R.A. Gansler, S.A. Klein, W.A. Beckman, Investigation of minute solar radiation data, Solar Energy, 55 (1995) 21-27.

[4] K. Otani, J. Minowa, K. Kurokawa, Study on areal solar irradiance for analyzing areally-totalized PV systems, Sol Energ Mat Sol C, 47 (1997) 281-288.

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

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

[7] E. Wiemken, H.G. Beyer, W. Heydenreich, K. Kiefer, Power characteristics of PV ensembles: experiences from the combined power production of 100 grid connected PV systems distributed over the area of Germany, Solar Energy, 70 (2001) 513-518.

[8] H.G. Beyer, Reduction of fluctuations in lumped power output from distantly spaced PV-arrays., 1991 Solar World Congress Proceedings of the Biennial Congress of the ISES, Pergamon Press, Oxford, New York,, Denver, Colorado, USA, 1991, pp. 321-326.

[9] A. Murata, H. Yamaguchi, K. Otani, A method of estimating the output fluctuation of many photovoltaic power generation systems dispersed in a wide area, Electrical Engineering in Japan, 166 (2009) 9-19.

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

[11] N. Kawasaki, T. Oozeki, K. Otani, K. Kurokawa, An evaluation method of the fluctuation characteristics of photovoltaic systems by using frequency analysis, Sol Energ Mat Sol C, 90 (2006) 3356-3363.

[12] O. Perpiñán, E. Lorenzo, Analysis and synthesis of the variability of irradiance and PV power time series with the wavelet transform, Solar Energy, 85 (2011) 188-197.

[13] J. Kleissl, J. Harper, A. Dominguez, A solar resource measurement network at high spatio-temproal resolution,American Solar Energy Society 2010, Phoenix, Arizona, USA, 2010.

[14] S. Wilcox, National solar radiation database 1991-2005 update: user's manual, 2007.

[15] C.N. Long, T.P. Ackerman, Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects, J. Geophys. Res., 105 (2000) 15609-15626.

[16] P. Ineichen, R. Perez, A new airmass independent formulation for the Linke turbidity coefficient, Solar Energy, 73 (2002) 151-157.

[17] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, 2009.

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

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

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

Chapter 4, in full, is a reprint of the material as it appears in *Solar Energy* volume 86, issue 8, August 2012. Lave, Matthew; Kleissl, Jan; Arias-Castro, Ery, 2012. The dissertation author was the primary investigator and author of this paper.

5. A Wavelet-based Variability Model (WVM) for Solar PV Power Plants

Abstract

A wavelet variability model (WVM) for simulating solar photovoltaic (PV) power plant output given a single irradiance point sensor timeseries using spatio-temporal correlations is presented. The variability reduction (VR) that occurs in upscaling from the single point sensor to the entire PV plant at each timescale is simulated, then combined with the wavelet transform of the point sensor timeseries to produce a simulated power plant output. The WVM is validated against measurements at a 2MW residential rooftop distributed PV power plant in Ota City, Japan and at a 48MW utility-scale power plant in Copper Mountain, NV. The WVM simulation match the actual power output well for all variability timescales, and the WVM compares well against other simulation methods.

5.1. Introduction

High penetration of solar power is highly desirable from an environmental point of view, but the variability of solar photovoltaic (PV) power is considered an obstacle to managing the electrical distribution and transmission system. Solar PV power production is variable due to the rising and setting of the sun, cloud shadows, changes in composition of the clear atmosphere (e.g., dust, smoke, humidity), and system-specific variables such as inverter performance, module temperature, and soiling. Cloud-induced fluctuations have the highest potential to affect the electric grid since they introduce changes in power production at short timescales (<1-hr). The other causes of variability typically change over longer timescales (>1-hr) and are often more predictable than clouds. Fortunately, though, since clouds are not homogeneous, geographic smoothing reduces short timescale variability for a fleet of PV systems.

We define the variability reduction (VR) as the ratio of variance in a time-varying quantity (e.g., normalized irradiance or PV power output) at one site to the variance of the average of all sites in a network (i.e., a PV power plant). Defined this way, an increase in VR will indicate a decrease in the aggregate variability. Values of VR depend on the correlation between sites: $VR = 1$ for perfectly correlated sites and $VR = N$ for independent sites. VR values of 2.8 for 1-min irradiance timeseries of 9 sites in a 4 km by 4 km grid in Tsukuba, Japan [1], 1.7 to 3.3 for 1-min steps of power output of 3 sites in 100s of km apart in Arizona [2], 2.4 to 4.1 for 5-min clear-sky index timeseries of 4 sites 100s of km apart in Colorado [3], and 2.7 for 5-min power output of 100 sites spread throughout Germany [4] have been found.

Other studies have shown the decorrelation of PV system output as a function of distance [5-7]. Hoff and Perez [8] showed that the correlation values collapse onto a line when the distance is divided by timescale. Accounting for cloud speed as determined from satellite further decreased the scatter, suggesting a universal correlation law.

With a similar objective to the present study, Marcos et al. [9] used a solar irradiance point sensor timeseries to simulate variability of a larger power plant using a transfer function based on a low pass filter which is scaled by the power plant area. Here, we describe a wavelet variability model that will help system planners gain an understanding of the variability of a potential power plant (i.e., largest ramp rates and how often they occur) with only limited data required as input. We describe the steps of the model in Section 5.2, Section 5.3 demonstrates and validates use of the model at two test sites, and Section 5.4 presents the conclusions.

5.2. **Methods**

We propose a wavelet variability model (WVM) for simulating power plant output given (1) measurements from a single irradiance point sensor, (2) knowledge of the power plant footprint and PV density (Watts of installed capacity per m^2), and (3) a correlation scaling coefficient (Section 5.2.3) by determining the geographic smoothing that will occur over the area

of the plant (Fig. 5.1). The simulated power plant may have any density of PV coverage: it may be distributed generation (i.e., a neighborhood with rooftop PV) with low PV density, centrally located PV as in a utility-scale power plant with high PV density, or any combination of both. In the WVM, we assume a statistically invariant irradiance field both spatially and in time over the day (i.e., stationary), and we assume that correlations between sites are isotropic: they depend only on distance, not direction.

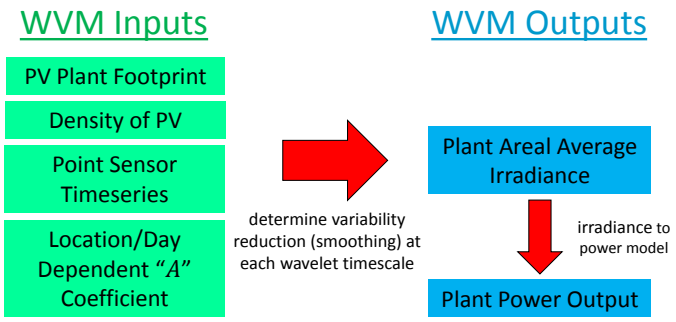

Fig. 5.1. Diagram showing the inputs and outputs for the WVM.

The WVM is designed to provide simulated power plant output to grid integration studies which test the effects of adding PV to existing electric feeders. These studies are done historically, after load and irradiance have been measured, and show the potential impacts of PV variability had PV been installed on the feeder being studied.

While the WVM is not a stand-alone forecasting model, it could be integrated into forecasting methods if spatially-distributed forecasted irradiances were available at high temporal resolution. However, satellite and numerical weather prediction forecast models typically only output at a temporal resolution of 30 minutes such that an upscaling method such as the WVM is not required since essentially no geographic smoothing occurs over power plant length scales (O~10 km) at such long timescales.

The main steps to the WVM are detailed in the sections below.

5.2.1. Wavelet Decomposition

We decompose the input irradiance point sensor timeseries into its components at various timescales by using a wavelet transform.

To obtain a stationary signal, the irradiance timeseries is normalized such that output during clear conditions is 1.

$$GHI_{norm}(t) = GHI(t)/GHI_{clr}(t), \tag{5.1}$$

where $GHI_{norm}(t)$ is the normalized signal, and $GHI_{clr}(t)$ is the clear-sky model (here the Ineichen model [10]). For simplicity of notation, we here assume that the point sensor is a global horizontal irradiance (GHI) sensor. If instead a plane of array (POA) sensor were used, a POA clear-sky model would be required.

The wavelet transform of the clear-sky index, $GHI_{norm}(t)$, is:

$$w_{\bar{t}}(t) = \int_{t_{start}}^{t_{end}} GHI_{norm}(t') \frac{1}{\sqrt{\bar{t}}} \psi\left(\frac{t'-t}{\bar{t}}\right) dt', \tag{5.2}$$

where the wavelet timescale (duration of fluctuations) is $\bar{t}$, t_{start} and t_{end} designate the start and end of the GHI timeseries, and t' is a variable of integration. For the discrete wavelet transform, $\bar{t}$ is increased by factors of 2, such that values of $\bar{t}$ are defined by $\bar{t} = 2^j$. We used the top hat wavelet, defined by:

$$\psi(T) = \begin{cases} 1, & \frac{1}{4} < T < 3/4 \\ -1, & 0 < T < \frac{1}{4} \quad || \quad \frac{3}{4} < T < 1 \\ 0, & else \end{cases}, \tag{5.3}$$

because of its simplicity and similarity to the shape of solar power fluctuations [11]. For 1-day at 1-sec resolution, we compute wavelet modes (timeseries) for $\bar{t}$ values ranging from 2-sec ($j = 1$) to 4096-sec ($j = 12$), thus decomposing the $GHI_{norm}(t)$ timeseries into 12 modes $w_{\bar{t}}(t)$ showing fluctuations at these various timescales. Symmetric signal extension is used to ensure resolution at endpoints. The largest timescale over which correlations are considered is 4096 sec (j = 12) because over the spatial scales of interest (O~10 km) the amount of smoothing that occurs at longer time scales is insignificant. In addition, modes for $j > 12$ require such significant signal extension that they are no longer representative of true

fluctuations in the irradiance timeseries. We adopt a special definition for the highest wavelet mode, defining $w_{\bar{t}=2^{12}}(t)$ to be the moving average with window 4096-sec. By doing so, we achieve the property that the sum of all wavelet modes equals the original input signal:

$$\sum_{j=1}^{12} w_{\bar{t}=2^j}(t) = GHI_{norm}(t). \quad 5.4$$

5.2.2. Distances

We next discretize the power plant into individual 'sites'. A single site is chosen to be an area over which $\rho(d_{m,n}, \bar{t}) \approx 1$ for the timescales of interest. For distributed plants, a single site is one rooftop PV system. For utility-scale plants, a single site is a small container of PV modules, as dictated by computational limitations. When using larger containers, a correction is applied for the in-container smoothing. Once discrete sites have been defined, the distance between each pair of sites is computed.

5.2.3. Correlations

To determine correlations between sites we assume that correlation is a function of distance divided by timescale [8]:

$$\rho(d_{m,n}, \bar{t}) = \exp(-\frac{1}{A}\frac{d_{m,n}}{\bar{t}}), \quad 5.5$$

where ρ is the correlation between sites, $d_{m,n}$ is the distance between sites m and n, $\bar{t}$ is the timescale, and A is a correlation scaling factor. The A value can be found using a small network of irradiance sensors (at least ~4-6 sites) where the correlations, distances, and timescales are known and A may be solved for using Eq. 5.5. The A value varies day-by-day and by location due to changing cloud speed. Smaller A values (1-3, typically observed at coastal sites with low, slow clouds) result in lower correlations between sites, while large A values (>4, typical of inland sites with high, fast-moving clouds) mean higher correlations. Through future work, we will create a closed form solution where A can be determined from geographic and meteorological variables and multiple irradiance sensors are not required.

5.2.4. Variability Reduction

The variability reduction as a function of timescale, $\mathrm{VR}(\bar{t})$, is defined as the variance of the point sensor divided by the variance of the entire PV power plant at each timescale. VR can be expressed as the inverse of the average of all correlations modeled in Eq. 5.5:

$$\mathrm{VR}(\bar{t}) = \frac{\mathrm{N}^2}{\sum_{\mathrm{m}=1}^{\mathrm{N}} \sum_{\mathrm{n}=1}^{\mathrm{N}} \rho(d_{m,n}, \bar{t})}, \tag{5.6}$$

where N is the total number of sites. Defined this way, $VR = N$ for entirely independent sites ($\rho = 0, \mathrm{m} \neq \mathrm{n}$), and $VR = 1$ for entirely dependent sites.

5.2.5. Simulate Wavelet Modes of Power Plant

By combining the wavelet modes $w_{\bar{t}}(t)$ found in Section 5.2.1 with the variability reductions $\mathrm{VR}(\bar{t})$ from Section 5.2.4, we simulate the wavelet modes of the power plant. The simulated wavelet modes of normalized power are reduced in magnitude by the square root of VR:

$$w_{\bar{t}}^{sim}(t) = \frac{w_{\bar{t}}(t)}{\sqrt{VR(\bar{t})}}, \tag{5.7}$$

where $w_{\bar{t}}^{sim}(t)$ are the simulated power plant wavelet modes. We can sum the simulated wavelet modes (inverse wavelet transform) to create a simulated clear-sky index of area-averaged GHI over the power plant:

$$< GHI_{norm}^{sim} >_{pp} (t) = \sum_{j=1}^{12} w_{\bar{t}=2^j}^{sim}(t). \tag{5.8}$$

5.2.6. Convert to Power Output

Power output is obtained by multiplying the spatially averaged irradiance (Section 5.2.6) by a clear-sky power output model, $P(t)_{clr}$.

$$P(t)^{sim} = < GHI_{norm}^{sim} >_{pp} (t) * P(t)_{clr} \tag{5.9}$$

$P(t)_{clr}$ is created by combining a plane of array irradiance clear-sky model with the plant's capacity, PC, and a constant conversion factor, C.

$$P(t)_{clr} = PC \times C \times POI_{clr}(t) \quad 5.10$$

To obtain $POI_{clr}(t)$, we apply the Page Model [12] to $GHI_{clr}(t)$. Since the Page Model requires GHI and diffuse irradiance as inputs, diffuse fraction was estimated as in [13]. The constant conversion factor, C, is determined based on the power plant's conversion efficiency.

Since PV plant power output is nearly linearly proportional to spatially averaged irradiance [14], using only a constant multiplier (C) is a reasonable approximation for this application. In practice, though, more sophisticated performance models [15] should be used that depend on ambient temperature, wind speed, and module specifications. The improvement in accuracy of power output achieved by using such a non-linear model depends on how far variables such as temperature deviate from standard test conditions (STC), but is expected to usually be less than 10%. Errors in estimating the variability at short timescales will be even smaller, since most of the non-linear irradiance to power effects occur over long timescales.

5.3. **Results/Application to Ota City and Copper Mountain Power plants**

To demonstrate the WVM model, we use the 2.13MW_p distributed generation (residential rooftop) plant in Ota City, Japan, and the Copper Mountain 48MW_p utility scale PV power plant in Boulder City, NV. Footprints of each plant are shown in Fig. 5.2. For both, the results of the WVM simulation were compared to the actual measured power output for the whole plant.

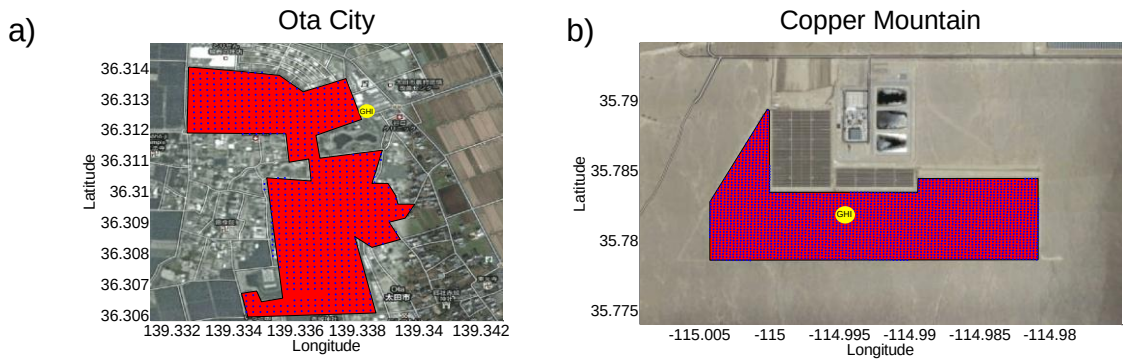

Fig. 5.2. Polygons showing the footprints of the (a) Ota City and (b) Copper Mountain power plants. The red shading shows the polygon footprints, while the blue dots show the simulation containers representing either houses (Ota City), or small groups of PV modules (Copper Mountain). The large yellow dots indicate the location of the GHI point sensors used as input. The Ota City map shows approximately 1 × 1 km, while the Copper Mountain map is approximately 3 × 2 km. Maps © Google Maps.

The Ota City (OC) plant consists of 550 houses, most with polycrystalline silicon PV systems ranging from 3-5kW, at varying tilts and azimuths. The average orientation of all PV modules producing power at OC on the test day was found to be 15° tilt from horizontal and 10° azimuth east of south.

GHI recorded once per second using an EKO instruments ML-020VM silicon pyranometer (expected uncertainty< 5%) was used as input to the WVM, and power output of the entire neighborhood also at 1-sec resolution was used for validation. The total power output was simply the sum of the output of each house, so ignored transmission losses. Results for a typically variable day (October 12, 2007) in OC (Fig.5.3a) will be presented.

The Sempra Generation Copper Mountain (CM) plant contains ground-mounted cadmium telluride (CdTe) thin-film PV at a fixed tilt of 25°. GHI at 1-sec resolution from a Kipp & Zonen CMP11 (expected daily uncertainty < 2%) was input to the WVM model, and the output was compared to power output of the entire plant also measured at 1-sec. Total power output was the sum of all inverter power outputs, so ignores AC transmission losses. October 1, 2011 was chosen as a typically variable test day at CM (Fig.5.3b).

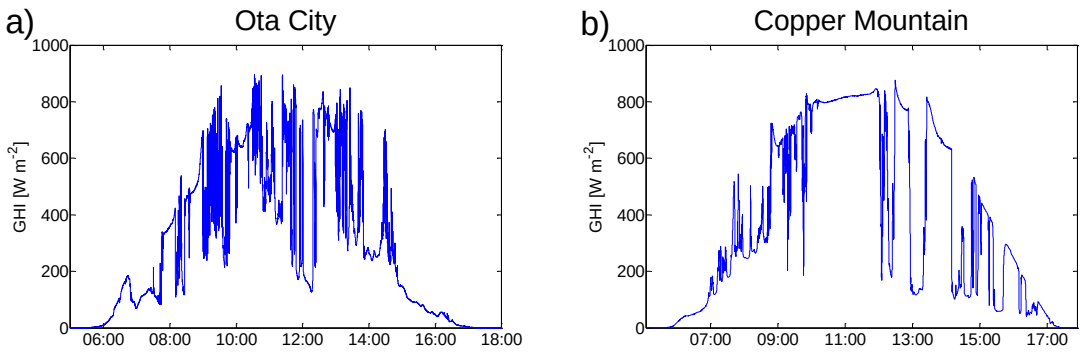

Fig.5.3. GHI at 1-sec resolution at (a) Ota City on October 12th, 2007, and at (b) Copper Mountain on October 1, 2011. X-axis is time in [HH:MM].

For purposes of validating the WVM, we will use irradiances at a point, area-averaged irradiances, and power output (for nomenclature see Table 5.1). For validation, the simulated power output (P^{sim}) is compared to the actual power output (P). The point sensor is included in some comparisons to show how the WVM output has improved from the input GHI.

Table 5.1: Nomenclature for GHI, simulated power output, and actual power output. Angle brackets denote averaging. GHI_{norm} is the clear-sky index, while $< GHI_{norm} >_{pp}$ is the ratio of actual to clear-sky output power. Since the later does not align with the definition of the clear-sky index, the GHI_{norm} notation was chosen.

	GHI units [W m^{-2}]	Clear-sky index [-]	Power units [MW]
measured point sensor	GHI	GHI_{norm}	
simulated power plant output	$< GHI^{sim} >_{pp}$	$< GHI^{sim}_{norm} >_{pp}$	P^{sim}
actual power plant output	$< GHI >_{pp}$	$< GHI_{norm} >_{pp}$	P

5.3.1. Inputs and Running the Model

The layouts of OC and CM were input and the WVM picked discrete points inside these footprints to use as "sites" in the simulation, as shown in Fig. 5.2. The spacing of sites was chosen based on the density of the plant. For OC, discrete sites were spaced roughly 20m apart such that each site represented a single house with rooftop PV. For CM, sites were closer together

(a few meters), with each site representing a sufficiently small group of PV modules that can be assumed to have a correlation coefficient of 1.

Also input to the WVM were the GHI measurement vectors, as well as latitude, longitude, and UTC offset (for creating the clear-sky model, GHI_{clr}). The test day at OC contains large cloud-induced irradiance fluctuations throughout the day. The test day at CM has some large irradiance fluctuations as well as a few clear periods (e.g., 10:00-12:00). Thus, the WVM model will be tested at two different sites and types of daily cloud conditions (highly variable and partly variable).

For OC on the test day, $A = 7.8$, as found from 6 GHI point sensors. Similarly, 15 plane-of-array reference cells at CM were used to determine $A = 4.9$ on October 1 (Fig.5.4). The small scatter of correlation points (black dots) around the best-fit curve, most noticeable at CM, is likely due to small anisotropic effects (i.e., pairs of sensors arranged in a certain direction may have higher correlation for all timescales).

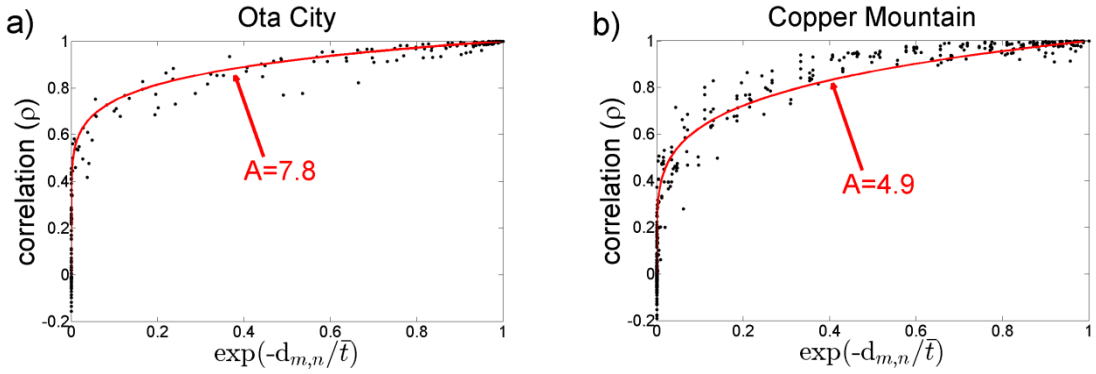

Fig.5.4. Correlations of wavelet modes for pairs of point sensors (a) at Ota City on October 12th, 2007 and (b) Copper Mountain on October 1, 2011. The x-axis is $\exp(-\frac{d_{m,n}}{\bar{t}})$ to show the exponential relationship between correlation and the quantity distance divided by timescale. Six GHI point sensors were used at Ota City, and 15 plane-of-array reference cells were used at Copper Mountain. Time scales range from 2-sec to 2048-sec. Dots to the far left have large $\frac{d_{m,n}}{\bar{t}}$ ratios meaning either very short timescales or very long distances, while dots to the far right have small $\frac{d_{m,n}}{\bar{t}}$ ratios so either very long timescales or very short distances. The red line is the line of best fit and corresponds to using $A = 7.8$ (Ota City) or $A = 4.9$ (Copper Mountain) in Eq. 5.5.

Following steps A-E (Section 5.2), a simulated normalized irradiance (similar to a clear-sky index, see Table 5.1) timeseries for the entire Ota City and Copper Mountain power plants was created, $< GHI_{norm}^{sim} >_{pp} (t)$. Fig.5.5a shows the effect of spatial averaging on the (simulated) reduction in variability. Fluctuations on timescales less than 8-sec are essentially zero and the fluctuations are reduced for timescales of 64-sec and shorter. At longer time scales, though, the wavelet modes between the GHI point sensor and the simulation are essentially identical. The power plant output 'clear-sky index' agrees with the simulation results across all timescales (Fig.5.5b).

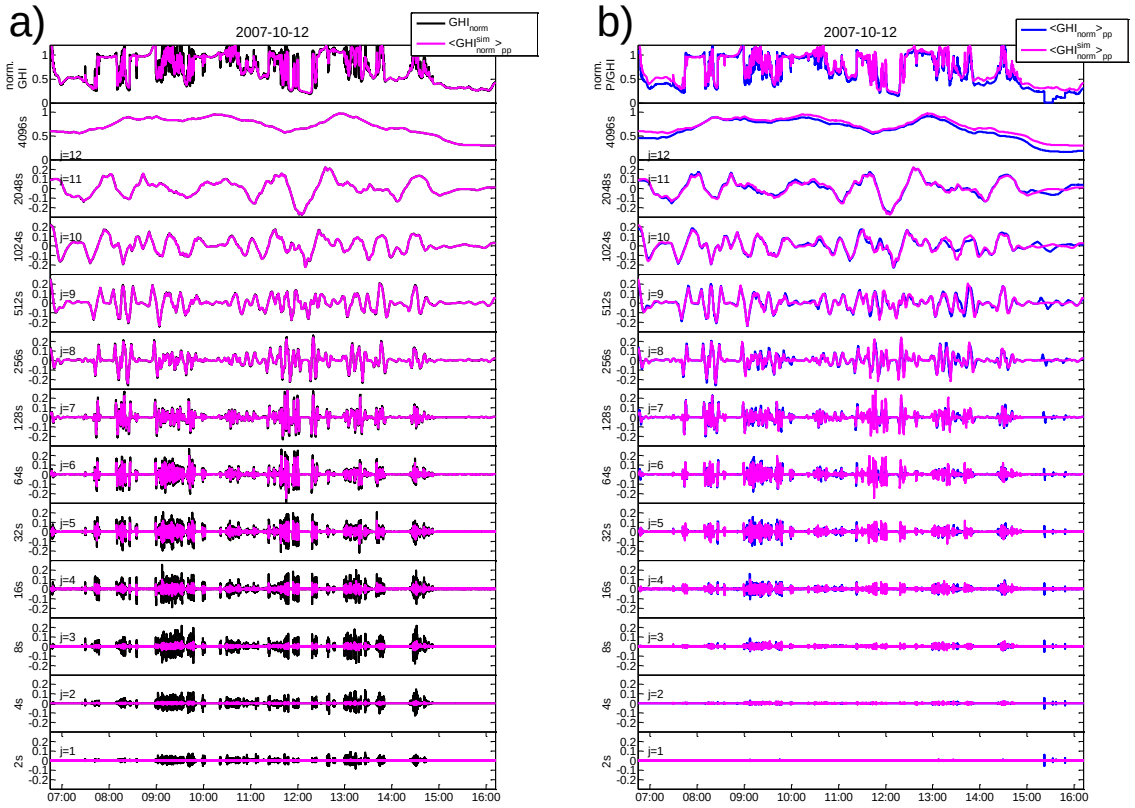

Fig.5.5. [top most plots] Clear-sky index timeseries, and [bottom 12 plots] wavelet modes for Ota City on the test day. [Left] Clear-sky index measured (GHI_{norm} , black) and simulated spatially averaged across the power plant ($< GHI_{norm}^{sim} >_{pp}$, magenta). [Right] Power plant output 'clear-sky index' measured ($< GHI_{norm} >_{pp}$, blue) and simulated ($< GHI_{norm}^{sim} >_{pp}$, magenta).

5.3.2. Validation of Simulated Power plant Output

Here we present comparisons between the simulated and actual power output using the fluctuation power index (fpi) and ramp rate (RR) distributions. The fpi, which is the power content of fluctuations in the wavelet modes at each timescale, is used as a variability metric to test the WVM. The goal of the WVM is to create simulated power output that statistically has the same variability distribution across timescales as the actual power output. The main output of the WVM is the VR as a function of timescale, but small errors in VR are irrelevant if the fpi is small (such as on clear days or at very short timescales when fluctuations have tiny magnitudes), as errors will also be very small. However, when the fpi is large (such as on cloudy days or at long timescales), errors in VR can lead to significant errors in fpi. Additionally, GHI and total power output can be slightly offset in time based on the direction of cloud movement and the location of the GHI sensor versus the centroid of the power plant. Since the fpi describes the variability content (and total variance) rather than the time of occurrence, it allows measuring the accuracy of the WVM independent of these geographic limitations.

Fig.5.6 shows the fpis of GHI_{norm}, $< GHI_{norm} >_{pp}$, and $< GHI_{norm}^{sim} >_{pp}$ for both OC and CM. As expected based on Fig.5.5, the fpi of the irradiance at a point sensor, GHI_{norm}, deviates from the fpi of the spatial average, $< GHI_{norm} >_{pp}$, at short timescales due to geographic smoothing. There is good agreement between simulated and actual fpis of the power output 'clear-sky index' at all timescales. The variability at CM at timescales of 2 to 8-sec is higher than estimated by the WVM, though this may be caused by differences in sensor response times between the GHI and power measurements. Since the power content is so small at these timescales ($< 10^{-8}$ compared to $> 10^{-4}$ at longer timescales), though, there is little effect on simulated power output. At both sites, the WVM ($< GHI_{norm}^{sim} >_{pp}$) significantly improved over its input (GHI_{norm}) at quantifying the fluctuation power content of the actual power output .

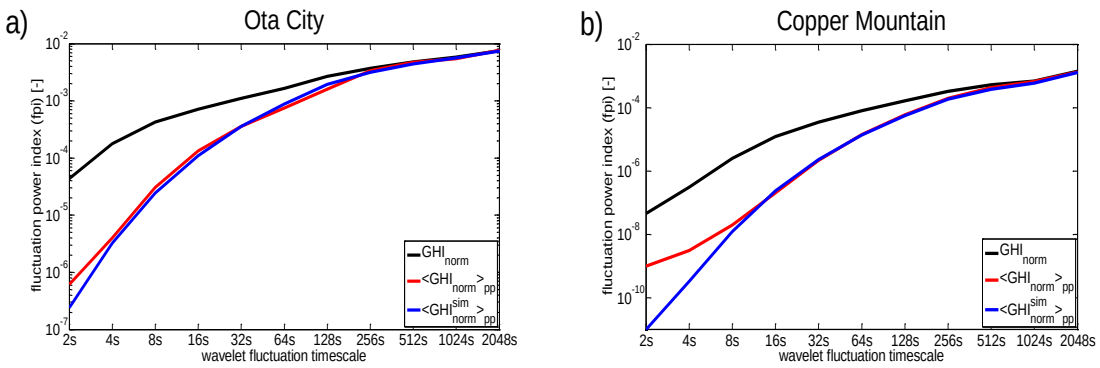

Fig.5.6. Fluctuation power index (fpi) for the GHI point sensor (black), actual power output of (red), and simulated power output (blue line) at (a) Ota City on October 12, 2007 and (b) Copper Mountain on October 1, 2011.

Another validation of the model is to compare the irradiance profile of the input GHI point sensor, WVM output areal averaged irradiance, and the 'actual' area-averaged irradiance derived from the power output. To obtain the actual area-averaged irradiance, we multiply the power output clear-sky index by the GHI clear-sky model:

$$< GHI(t) >_{pp} = < GHI(t)_{norm} >_{pp} * GHI(t)_{clr}. \quad 5.11$$

The simulated area-averaged irradiance clear-sky index must also be multiplied by the GHI clear-sky model to obtain simulated areal averaged irradiance:

$$< GHI^{sim}(t) >_{pp} = < GHI(t)^{sim}_{norm} >_{pp} * GHI(t)_{clr}. \quad 5.12$$

Fig.5.7 shows that $< GHI^{sim}(t) >_{pp}$ is essentially a filtered version of $GHI(t)$; during long clear or cloudy periods the area-averaged GHI approaches the point sensor while short fluctuations e.g., at 09:09, are more strongly reduced. The timeseries of fluctuations, $< GHI(t) >_{pp}$ and $< GHI^{sim}(t) >_{pp}$ are not expected to match perfectly, since only a single point sensor is used as input, but the statistics of the fluctuations are expected to agree. For the eight minute segment shown in Fig.5.7a, the 'upper envelope' of area-averaged irradiance is accurately simulated. However, especially during long cloud events (large timescales), the lower envelope of the power plant is as small as or even smaller than the envelope of the point sensor. Physically this could mean that the optical depth of the cloud at the point sensor was less than the average of the cloud

system over the power plant (spatial heterogeneity); it could also mean that the tilt and azimuth of the PV modules comprising the plant (versus the horizontal GHI) resulted in a smaller diffuse irradiance at the power plant during these morning hours.

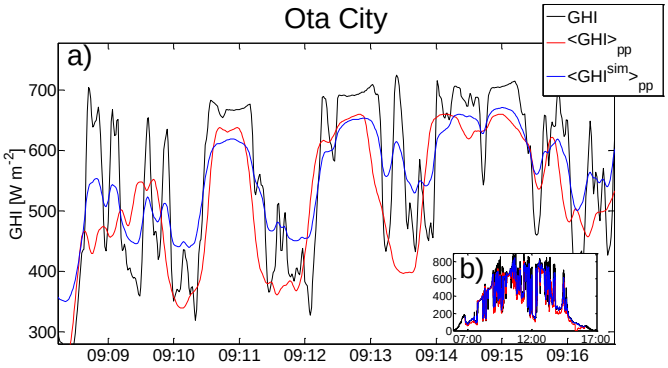

Fig.5.7. Point sensor GHI (black), power plant area-averaged GHI (red), and simulated area-averaged GHI (blue) for Ota City on October 12, 2007. (a) Zoomed in to eight minutes in the morning, and (b) the entire day.

While the WVM has been shown to match fpi (Fig.5.6) and smoothing of area-averaged irradiance (Fig.5.7), ramp rate statistics are often of greater interest to power plant and grid operators. Fig.5.8 shows the cdf of ramp rates at 1s, 10s, 30s, and 1-min, for both OC and CM. P and P^{sim} show similar RR distributions at all of these timescales. The match between RRs is slightly worse at CM, but this is expected since CM is a larger area than OC. Over the larger area, the irradiance statistics may be less homogenous and lead to error in the WVM output. For example, half of CM could be consistently experiencing partly cloudy conditions while the other half is under clear skies. If the sensor used as input to the WVM were in the clear-sky section, the WVM would under predict the variability of the total plant. This scenario is less likely to happen at OC, since the distance across the plant is shorter (~1km at OC vs. ~2km at CM).

5.3.3. Comparison to Other Upscaling Methods

To evaluate the performance of the WVM, we compare it to other methods for simulating power plant output. We chose to compare to 3 other methods that have been applied for upscaling. (1) A primitive method is to simply linearly scale the irradiance measured by a single point sensor as $GHI \times CF$, where CF

is a constant factor related to the plant area. This assumes identical relative variability between the point sensor and the whole plant, strongly overestimating variability for all but the smallest of PV plants. (2) Since geographic smoothing occurs over certain time scales (Fig.5.5), a moving average (or box filter) can be applied to an irradiance timeseries. The timescale of the moving average can be physically motivated from the relation $t_{avg} = \frac{A^{1/2}}{V}$, where A is the plant area and V is the cloud velocity. While fundamentally similar to the WVM, this moving average method is restricted to smoothing at a single timescale (t_{avg}), and so cannot take into effect different amounts of smoothing at different timescales. (3) A third method is to average the timeseries from all available irradiance sensors. This method becomes more accurate as more sensors are added and in the extreme case of e.g. one sensor per PV panel would be exact. It requires the PV sensors to be representative of the actual plant layout: for best results, sensors should be equally spaced and cover the entire plant footprint (but not extend past the footprint, or smoothing will be overestimated).

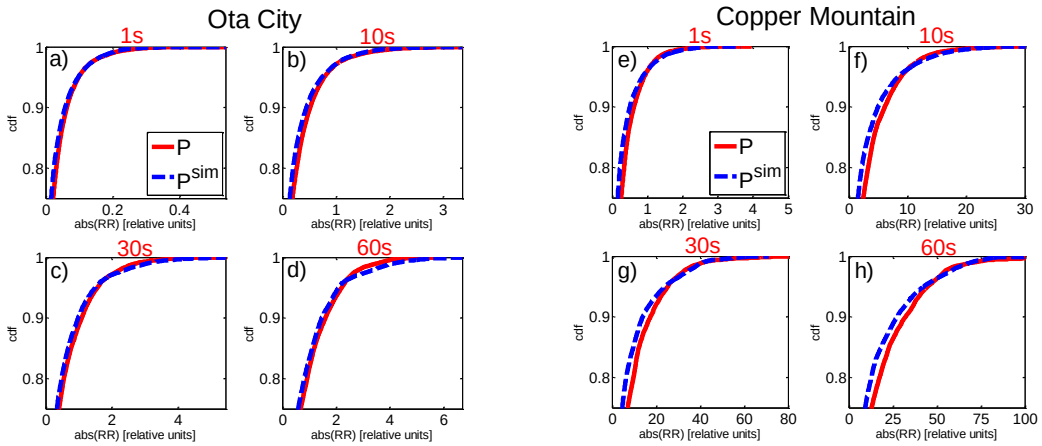

Fig.5.8. Extreme (> 75 percentile) ramp rate distributions at 1-sec (a, e), 10-sec (b, f), 30-sec (c, g), and 1-min (d, h) for P and P^{sim} for Ota City (a-d) and Copper Mountain (e-h). X-axis units are multiplied by an arbitrary scaling factor to protect the confidentiality of the power data.

In Fig.5.9, the extreme RRs of CM power output for these 3 methods are compared to the WVM on the test day. For method 2, the plant area is 1.33km, and a cloud speed of 10 m s^{-1} was assumed, resulting in a moving average timescale $t_{avg} = 115s$. For method 3, the 15 reference cells available at CM were averaged.

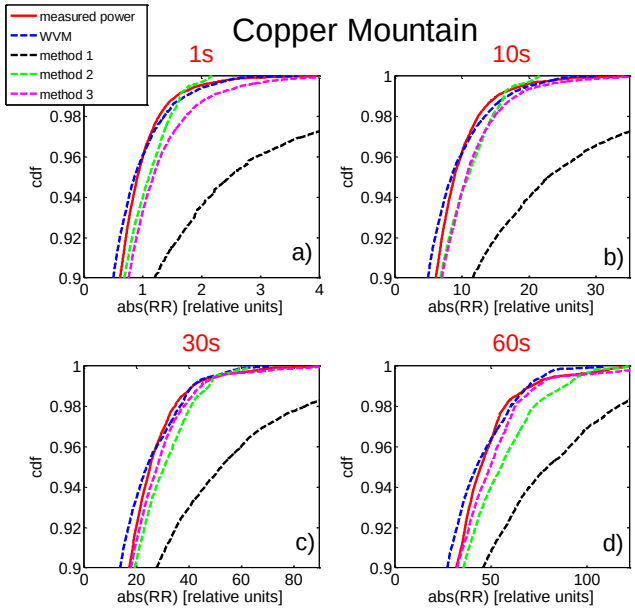

Fig.5.9. Extreme (>90th percentile) ramp rate cumulative distribution functions at CM on October 1, 2011 at 1-sec (a), 10-sec (b), 30-sec (c), and 1-min (d) for measured power output (solid red), and for different methods of simulating PV power plant output: WVM (dashed blue), method 1: linearly scaling from a point sensor (dashed black), method 2: a moving average of 115s corresponding to $t_{avg} = \frac{A^{1/2}}{V}$ (dashed green line), and method 3: averaging all 15 reference cells (dashed magenta line).

Method 1 overestimated the RRs at all timescales, since geographic smoothing is ignored. Method 2 also overestimated most RRs at all timescales (maximum RRs were underestimated), but performed markedly better than method 1. Method 2 could be improved by using measured cloud velocity instead of an assumed value. Method 3 overestimated RRs at shorter timescales (1s and 10s), but was more accurate at longer timescales (30s and 60s). At short timescales when short-distance changes are important, 15 sensors will not be enough to accurately represent all of the PV modules. As the timescale increases, the distances over which changes are important will increase, and this 15 sensor network will be more accurate in representing the entire plant.

Since grid operators are often concerned about worst-case scenarios, it is important that these simulation methods also produce reasonable maximum RRs. Table 5.2 shows the maximum RRs by timescale for each of the simulation methods. The WVM showed small percentage errors (-8% to -20%) at all timescales. Methods 1 and 3 will always overestimate the maximum RRs,

since at 1 and 15 point measurements, respectively, they both underrepresent the actual diversity of the plant. As expected, Method 3 outperformed Method 1 with a range of errors of 8% to 38%. Method 2, always overestimated the maximum RRs, and had moderate errors at most timescales (-0% to -45%).

Table 5.2: Maximum RRs for CM on the test day compared to maximum RRs simulated using other methods described in Section 5.3.3. Relative units are used for RRs to protect confidential data.

		P	WVM	method		
				1	2	3
1s	max RR	0.50	0.46	4.08	0.28	0.68
	% error		-8%	723%	-45%	38%
10s	max RR	4.29	3.79	21.80	2.70	4.70
	% error		-12%	408%	-37%	9%
30s	max RR	11.19	8.92	24.95	7.88	12.11
	% error		-20%	123%	-30%	8%
60s	max RR	15.29	13.70	28.17	15.28	18.02
	% error		-10%	84%	-0%	18%

It is also worth noting the data requirements of each method. Aside from method 1 (which requires only a single sensor), the WVM has the simplest input requirements, needing just a single irradiance sensor and an A value. The A value can either be determined from a small network of irradiance sensors (~4-6), or reasonably estimated based on observed trends in A values. Method 2 requires the cloud velocity, which is difficult to determine since ground wind speeds are not well correlated with cloud speeds, and only sparse measurements exist of cloud height and winds aloft. Method 3 performs best with a high-density sensor network. When such a network is not available, the accuracy of method 3 will decrease towards that of method 1. Overall, the WVM has simple input requirements and is best at simulating extreme and maximum RRs.

5.4. **Conclusion**

A wavelet-based variability model (WVM) for simulating the power output of a solar photovoltaic (PV) plant was presented and tested. The WVM uses the plant footprint, density of PV modules in the plant, and the timeseries measurements from a single point sensor to create a simulated power output timeseries. First, correlations between sites (i.e., houses or small groups of PV modules) within the power plant are determined using an equation based on the distance between sites, timescales, and a correlation scaling coefficient (A value). From these correlations, variability reductions (VRs), or the ratio of variability of a single point sensor to the variability of the entire PV plant, at each fluctuation timescale are found. Wavelet decomposition is then used to separate the normalized input point sensor timeseries by fluctuation timescale. By combining the wavelet modes at each timescale with the VRs at each timescale, the normalized plant power output is simulated. Actual power output (in MW) is then obtained by using a clear-sky model for power output.

The WVM was validated at the 2MW distributed residential rooftop plant in Ota City, Japan, and the Copper Mountain, NV 48MW central power plant. For both test cases, the WVM simulation matched the statistics of the actual power output well. Fluctuation power index (fpi) comparisons showed that the WVM accurately represented variability by timescale at both Ota City and Copper Mountain. Ramp rates (RRs) at 1-sec, 10-sec, 30-sec, and 60-sec were also compared between simulated and actual power, and again, the WVM fared well at both test sites.

Comparison between the WVM and other power plant simulation methods highlighted the benefits of using the WVM. The WVM has reasonable input requirements (single sensor and A value), while other methods required harder to obtain inputs such as cloud velocity or irradiance from a dense sensor network. The WVM was best at matching extreme and maximum RRs for the test day at Copper Mountain.

Future work will concentrate on characterizing the *A* values used in the correlation equation in Section 5.2.3. Currently, *A* values are determined form a small network of irradiance sensors. To allow for broader application of the WVM, we will determine how *A* values vary based on geographic region and meteorological condition. Then, a closed form WVM program will be created which will estimate the variability of potential PV plants for grid integration and siting studies without requiring a sensor network.

Acknowledgment

We appreciate the help of Yusuke Miyamoto and Eichi Nakashima from Kandenko in Ibaraki, Japan for providing and supporting the Ota City data, as well as David Jeon, Leslie Padilla, and Shiva Bahuman from Sempra Energy in San Diego, California Darryl Lopez from Eldorado Energy, and Bryan Urquhart from UCSD for providing and supporting the Copper Mountain data.

References

[1] K. Otani, J. Minowa, K. Kurokawa, Study on areal solar irradiance for analyzing areally-totalized PV systems, Sol Energ Mat Sol C, 47 (1997) 281-288.

[2] A.E. Curtright, J. Apt, The character of power output from utility-scale photovoltaic systems, Prog Photovoltaics, 16 (2008) 241-247.

[3] M. Lave, J. Kleissl, Solar variability of four sites across the state of Colorado, Renew Energ, 35 (2010) 2867-2873.

[4] E. Wiemken, H.G. Beyer, W. Heydenreich, K. Kiefer, Power characteristics of PV ensembles: experiences from the combined power production of 100 grid connected PV systems distributed over the area of Germany, Sol Energy, 70 (2001) 513-518.

[5] A. Murata, H. Yamaguchi, K. Otani, A Method of Estimating the Output Fluctuation of Many Photovoltaic Power Generation Systems Dispersed in a Wide Area, Electr Eng Jpn, 166 (2009) 9-19.

[6] A. Mills, R. Wiser, Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power, LBNL Report No. 3884E, (2010).

[7] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker, T. Hoff, Short-term irradiance variability: Station pair correlation as a function of distance, Submitted to Solar Energy, (2011).

[8] T. Hoff, R. Perez, Modeling PV Fleet Output Variability, Submitted to Solar Energy, (2011).

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

[10] P. Ineichen, R. Perez, A new airmass independent formulation for the Linke turbidity coefficient, Sol Energy, 73 (2002) 151-157.

[11] M. Lave, J. Kleissl, E. Arias-Castro, High-frequency irradiance fluctuations and geographic smoothing, Sol Energy.

[12] J. Page, The role of solar radiation climatology in the design of photovoltaic systems, in: T. Markvart, L. Castaner (Eds.) Practical handbook of photovoltaics: fundamentals and applications, Elsevier, Oxford, 2003, pp. 5-66.

[13] J. Boland, B. Ridley, B. Brown, Models of diffuse solar radiation, Renew Energ, 33 (2008) 575-584.

[14] C. Hansen, J. Stein, A. Ellis, Simulation of One-Minute Power Output from Utility-Scale Photovoltaic Generation System, 2011.

[15] D. King, W. Boyson, J. Kratochvil, Photovoltaic Array Performance Model, 2004.

Chapter 5, in full, is a reprint of the material as it appears in the *IEEE Transactions on Sustainable Energy* (preprint) 2012. Lave, Matthew; Kleissl, Jan; Stein, Joshua, 2012. The dissertation author was the primary investigator and author of this paper.

6. Cloud Speed Impact on Solar Variability Scaling – Application to the Wavelet Variability Model

Matthew Lave, Jan Kleissl, Cloud Speed Impact on Solar Variability Scaling, submitted to Solar Energy, 2012.

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 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.

6.1. 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 [1]. 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 [2-5]. Other investigators [6-8] 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 [9]. Correlation was also shown to depend on orientation relative to the direction of cloud motion [10].

Wavelet analysis has been used to decompose solar fluctuations into different timescales and measure the variability at that timescale [11-13]. 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 [14, 15]. In this paper, we build on the wavelet-variability model (WVM) presented in [15] by using cloud speeds to make the model more generally applicable. The WVM and its correlation equation are described in Section 6.2, the relationship between cloud speed and the correlation scaling coefficient (A value) is shown in Section 6.3, a method for determining cloud speeds and converting to A values is presented in Section 6.4, and in Section 6.5 we apply and validate the WVM with cloud-speed derived A values at a 48MW PV plant.

6.2. **The Wavelet Variability Model (WVM)**

6.2.1. 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 [15]. 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, $\rho(d_{m,n}, \bar{t})$, between the wavelet modes at different sites. A location/day dependent correlation scaling coefficient (A) is needed to account for cloud speed (Section 6.2.2).
4) Use $\rho(d_{m,n}, \bar{t})$ 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 areal-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 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. 6.1), 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.

6.2.2. 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 [6-8, 16], 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

$$\rho\left(d_{m,n}, \bar{t}\right) = \exp\left(-\frac{d_{m,n}}{A\,\bar{t}}\right). \qquad 6.1$$

Eq. 6.1 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.

6.2.3. Correlation Model Limitations

The assumption of isotropy in Eq. 6.1 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 [7] and in the along-wind direction in [10]. 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 $\psi(T)$ 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. 6.1 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 the cloud field.

In both limitations A and B, Eq. 6.1 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.

6.2.4. 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. 6.1. Distance, timescale, and correlation are known from measurements in the network and a best fit *A* value can be determined (Fig. 6.1). The slight variation of correlations above and below the *A* value best fit in Fig. 6.1 are likely due to limitations A and B.

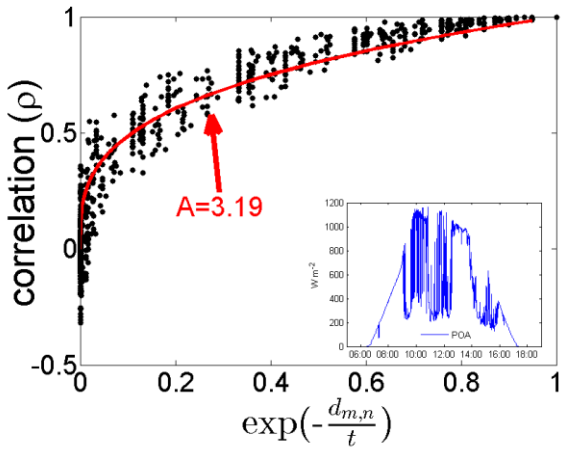

Fig. 6.1: Correlations between wavelet modes of clear sky indices measured in a sensor network. The x-axis is $\exp(-\frac{d_{m,n}}{\tilde{t}})$ to show the exponential decay of correlation as a function of distance and timescale. The red line is the correlation modeled using Eq. 6.1, where the A value (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.

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

6.3.1. 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. 6.2a. 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. 6.2b.

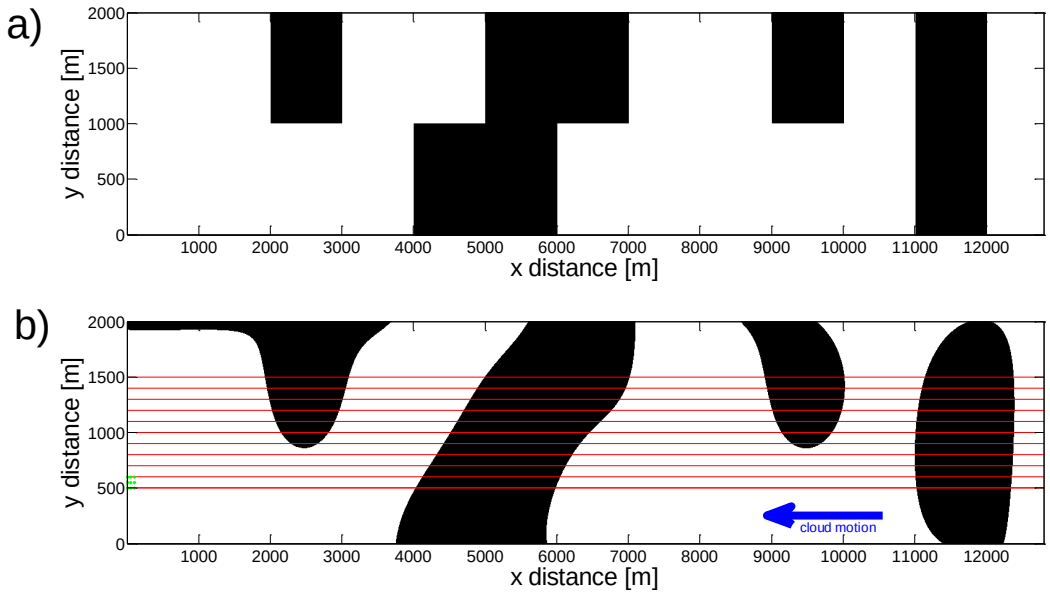

Fig. 6.2: Snapshot of the domain for (a) the coarse cloud field and (b) the fine cloud field. White represents clouds and black represent clear-sky. The blue arrow shows the direction of cloud motion, the green dots (at ~ x=0, y=500) show an example sensor network, and the red lines show the range of 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 6.2.4, the correlations of wavelet modes of $kt(t)$ for various sensor pairs were found, and Eq. 6.1 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 sensor positions 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 1/10th the cloud size (500 to 1500 m in Fig. 6.2). 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.

6.3.2. 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. 6.3a. Fig. 6.3b shows the A value as a function of cloud speed.

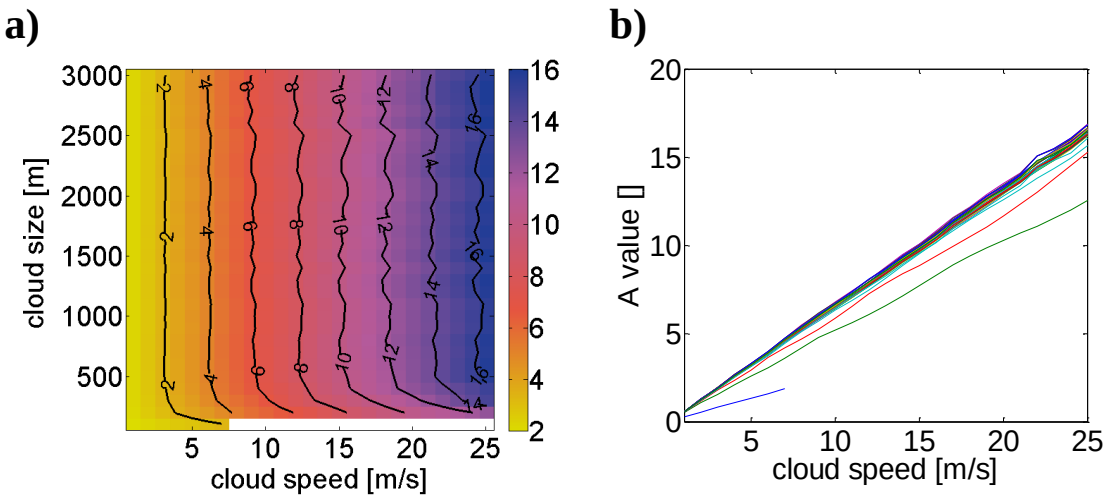

Fig. 6.3: A values determined through cloud field simulation: (a) on a 2D plot showing the A value as a function of cloud speed and cloud size, and (b) on a line graph showing the A value as a function only of cloud speed. Most of the A value fits for the 100 m cloud size had errors larger than the 15% 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. 6.3b), we define:

$$A = C_1 * CS. \tag{6.2}$$

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. 6.4 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 6.2.3 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.

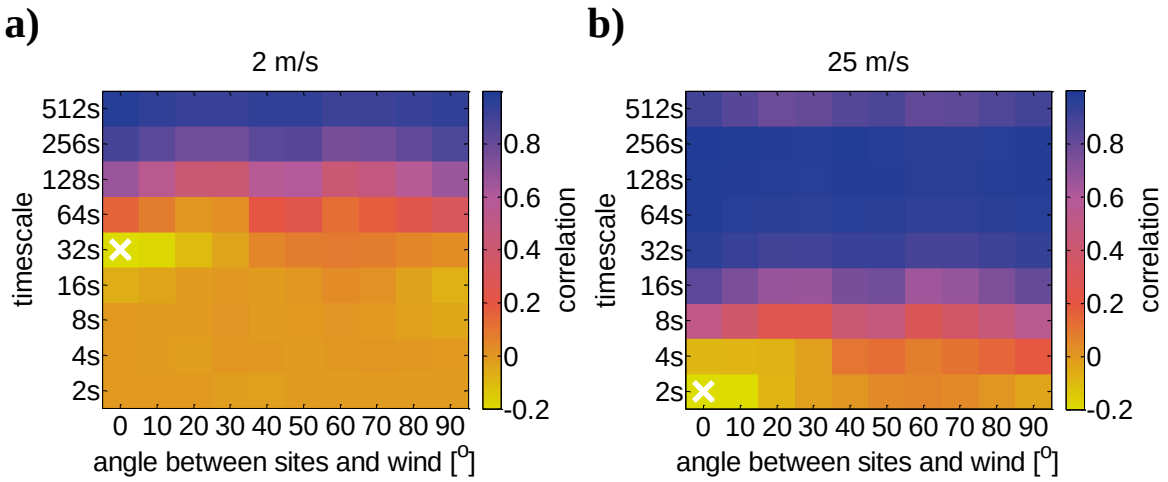

Fig. 6.4: Correlations from one run of the cloud simulator with 1000 m clouds and 0.7 cloud cover, averaged over all 10 sensor network offset positions. The plots show the dependence of correlations on the angle between the sites and the wind direction and on the timescale for (a) 2 m s^{-1} cloud speed 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), 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. 6.1), and hence a smaller C_1. As cloud size increases, though, the effect of limitation B vanishes, and C_1 becomes a constant.

Fig. 6.5 shows the effect of limitation B on the correlations. At 200 m cloud size, correlations in the along-wind direction (red dots in Fig. 6.5a) 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. 6.5b).

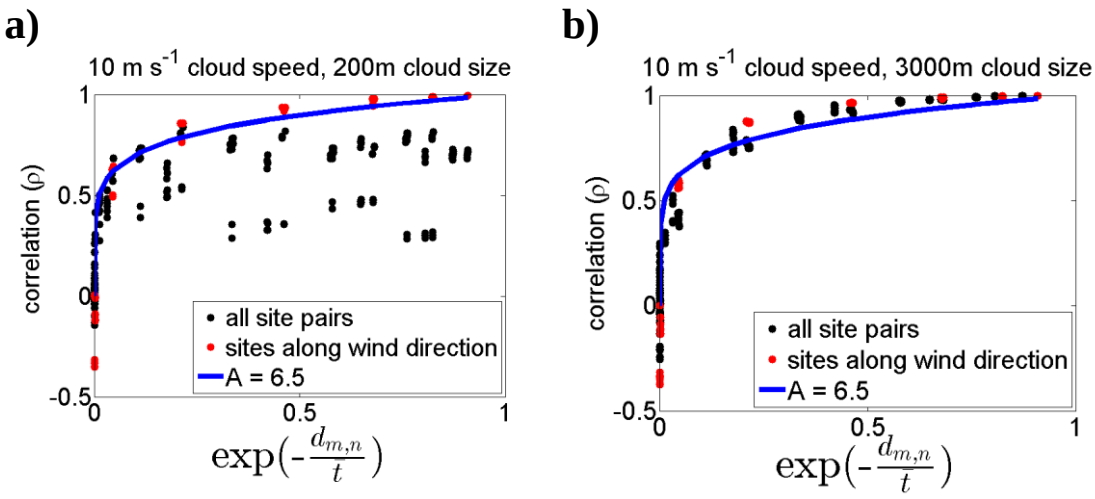

Fig. 6.5: Correlations from one run of the cloud simulator with 10 m s-1 cloud speed and 0.7 cloud fraction, averaged over all 10 sensor network offset positions, for (a) 200m cloud size and (b) 3000m cloud size. The black dots show all correlations, while the red dots show correlations between sites 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. 6.2 to hold, empirical evidence (e.g., Fig. 6.8a) suggests that $C_1 = 0.5$, and Eq. 6.2 becomes:

$$A = \frac{1}{2} * CS. \tag{6.3}$$

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

Eq. 6.3 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) [17] 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 [18]. Techniques for measuring cloud speed exist [19, 20], but only a limited amount of suitable sensors are deployed. Here, we use the NOAA North American Mesoscale (NAM) numerical weather

forecast [21], which has both a higher spatial (~12x12 km grids covering continental North America) and temporal (once per hour) resolution than radiosonde measurements.

6.4.1. 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. 6.6). 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% must 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.

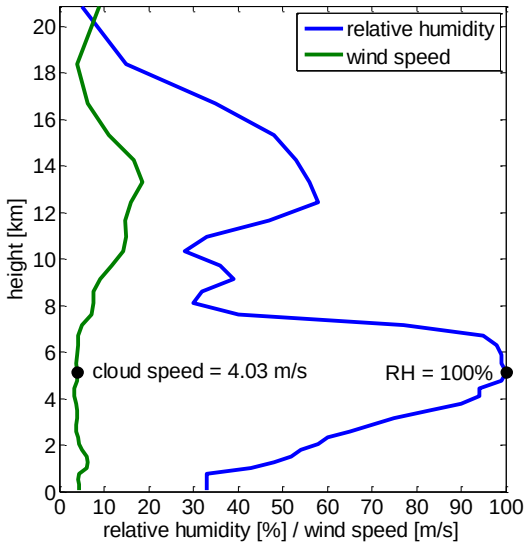

Fig. 6.6: Example plot of relative humidity and wind speed on July 26, 2011 at Copper Mountain, NV. In this case, the cloud was found to be at about 5km, with a speed of 4.03 m s-1.

6.4.2. *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. 6.3. 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. 6.7 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. 6.7. 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.

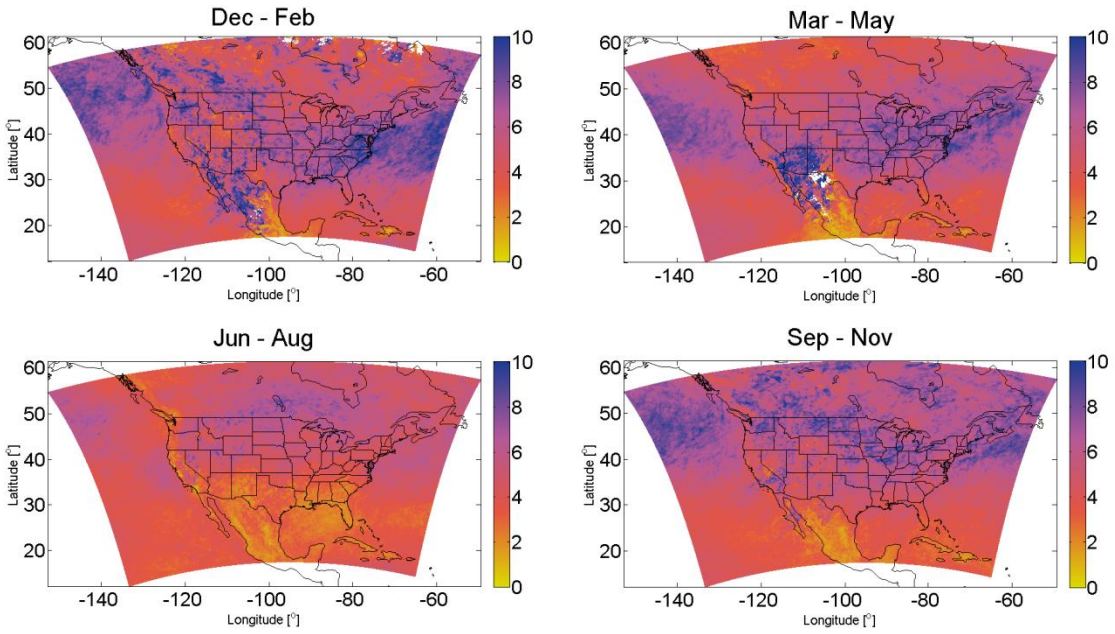

Fig. 6.7: GHI variance-weighted ***A*** values for the NAM coverage area by season.

6.4.3. 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 6.5). Fig. 6.8 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. 6.8b shows that the cumulative distribution of NAM *A* values is similar to the cumulative distribution of actual *A* values.

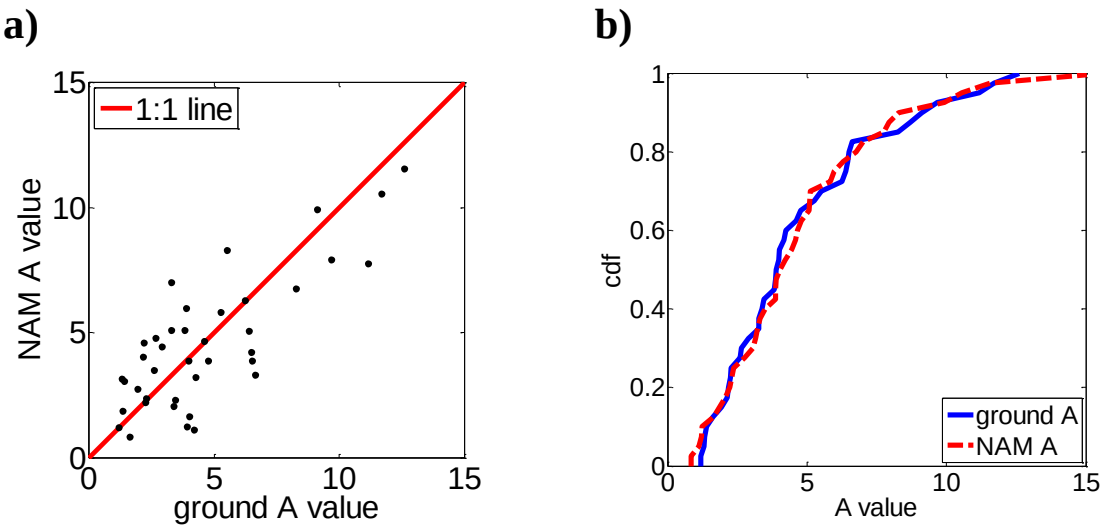

Fig. 6.8: Comparison of ground ***A*** and NAM ***A*** values during the time period August 1st, 2011 through July 31st, 2012. (a) scatter plot with 1:1 line, and (b) cumulative distribution functions (cdfs).

6.4.4. *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 6.4.4).

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.

6.5. **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. 6.1 as in Fig. 6.1. 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.

6.5.1. 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. 6.7 and 6.8.

As described in Section 6.4.4, 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. 6.9 shows the large (>90^{th} percentile) RRs of actual power output and the three WVM scenarios.

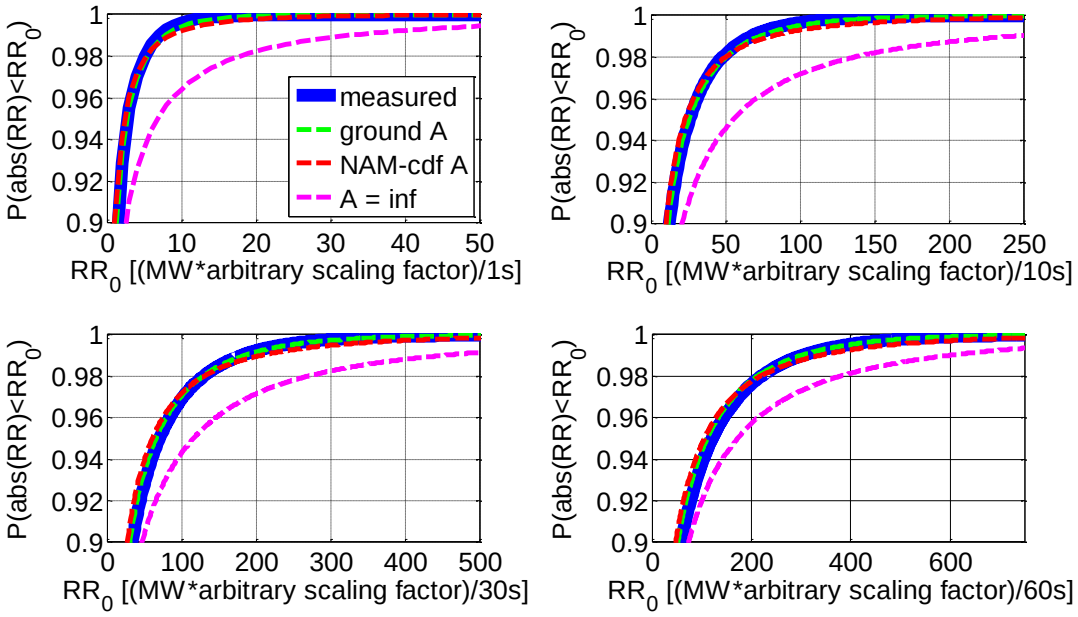

Fig. 6.9: Cumulative distribution of ramp rates in power output for the 1-year period from August 1, 2011 through July 31st, 2012. Ramp rates are shown at various timescales: 1s (top left), 10s (top right), 30s (bottom left), and 60s (bottom right). At each timescale, the ramp rates of measured power 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. 6.9. 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. 6.9 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.

6.5.2. 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 Cramer–von Mises criterion:

$$\omega^2 = \int_0^\infty |F_{WVM\ method}(RR_0) - F_{measured}(RR_0)|^2 dF_{WVM\ method}(RR_0). \quad 6.4$$

Values of ω^2 for each method at 1s, 10s, 30s, 1-min, 5-min, and 10-min are shown in Fig. 6.10. Due to different maximum RRs at each timescale, the Cramer–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 the powerplant have well-correlated output over long timescales, so the scaled point sensor becomes more accurate.

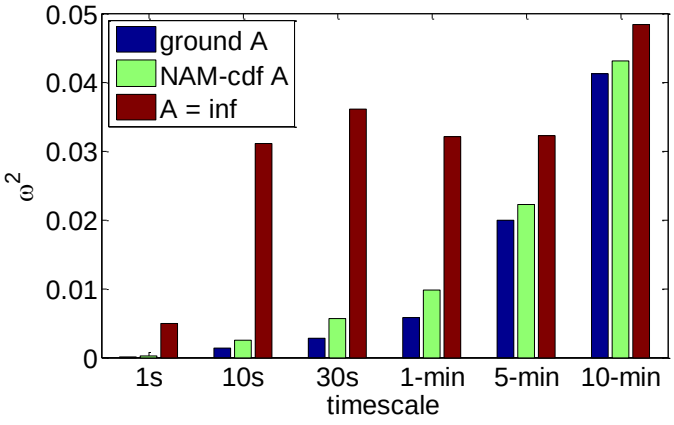

Fig. 6.10: Cramer-von Mises criterion (ω^2) showing the difference between the cumulative distribution of measured ramp rates and WVM ramp rates found using ground A values (blue), NAM A values (green), and the un-smoothed point sensor (A=inf, red).

6.6. **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 correlation model proposed by Perez et. al in [9] (Eq. 9): $\rho = \frac{1}{1+\frac{d_{m,n}}{CS\, \bar{t}}}$. As expected, both models predict nearly zero correlation at very long distances or very short timescales, and nearly perfect correlation at short distances or long timescales. In between these two limits, the differences between the two models are small; the exponential model presented here decays slightly faster than the fractional model proposed in [9]. In [9], 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 in both works. The agreement in correlation models between the two works shows that the nondimensional quantity $\frac{d_{m,n}}{CS\, \bar{t}}$ 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.

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. [10] 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 [10] 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 [10] support the assumption that cloud sizes are large relative to distances between sites, as no sharp dropoff 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

[1] Puerto Rico Electric Power Authority, Minimum Technical Requirements for Photovoltaic Generation (PV) Projects, http://www.fpsadvisorygroup.com/rso_request_for_quals/PREPA_Appendix_E_PV_Minimum_Technical_Requirements.pdf.

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

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

[4] K. Otani, J. Minowa, K. Kurokawa, Study on areal solar irradiance for analyzing areally-totalized PV systems, Sol Energ Mat Sol C, 47 (1997) 281-288.

[5] E. Wiemken, H.G. Beyer, W. Heydenreich, K. Kiefer, Power characteristics of PV ensembles: experiences from the combined power production of 100 grid connected PV systems distributed over the area of Germany, Solar Energy, 70 (2001) 513-518.

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

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

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

[9] T.E. Hoff, R. Perez, Modeling PV fleet output variability, Solar Energy, 86 (2012) 2177-2189.

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

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

[12] N. Kawasaki, T. Oozeki, K. Otani, K. Kurokawa, An evaluation method of the fluctuation characteristics of photovoltaic systems by using frequency analysis, Sol Energ Mat Sol C, 90 (2006) 3356-3363.

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

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

[15] M. Lave, J. Kleissl, J.S. Stein, A Wavelet-Based Variability Model (WVM) for Solar PV Power Plants, Sustainable Energy, IEEE Transactions on, PP (2012) 1-9.

[16] T.E. Hoff, R. Perez, Quantifying PV power Output Variability, Solar Energy, 84 (2010) 1782-1793.

[17] National Weather Service Radiosonde Observations, National Oceanic and Atmospherica Administration http://www.ua.nws.noaa.gov/factsheet.htm.

[18] R. Perez, Personal Communication, 2012.

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

[20] J. Bosch, J. Kleissl, Deriving cloud velocity from an array of solar radiation measurements, submitted to Solar Energy, (2012).

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

Chapter 6, in full, is a reprint of the material as has been submitted to *Solar Energy* 2012. Lave, Matthew; Kleissl, Jan, 2012. The dissertation author was the primary investigator and author of this paper.

7. Conclusion and Example of Future Work: WVM simulation in Puerto Rico

7.1. PREPA 10% RR Requirement

The enhancement of incorporating cloud speeds in the WVM (Chapter 6) has greatly increased the number of locations where the WVM can be run to simulate PV powerplant output. One such particularly important location is Puerto Rico. Because of the Puerto Rico Electric Power Authority (PREPA) requirement that all PV plants on Puerto Rico limit ramps to less than 10% of capacity per minute [1], there is strong interest in estimating the RRs that PV plants being installed or considered in Puerto Rico will produce.

7.2. Data Availability

As of summer 2012, no high-frequency irradiance or power data measured anywhere on the island of Puerto Rico was publically available. Because of this, in August 2012 the Kleissl Lab Group installed three irradiance sensor in close proximity (a few meters) on a rooftop at the University of Puerto Rico, Mayaguez [2]. These sensors not only give a high-frequency irradiance input to the WVM, but also allow for resolution of cloud speed based on methods described in [3]. By using the cloud-speed enhanced WVM presented in Chapter 6, the WVM can be run to simulate powerplant RRs in Mayaguez.

At the time of writing this dissertation, only data from the month of September, 2012 is available from the irradiance sensors in Mayaguez. The GHI for each day is shown in Fig. 7.1. Many days at Mayaguez are clear in the morning but become highly variable by midday (with changes in irradiance exceeding 50% in 1-minute). However, since only one month of data is available this may not be representative of yearly trends at Mayaguez. Additionally, the Mayaguez data may not accurately represent other locations in Puerto Rico. Mayaguez is on the western coast of Puerto Rico. Locations further inland or on different coasts may have different irradiance statistics due to different weather patterns. The analysis presented in this chapter is

meant to be illustrative and to give a broad understanding of the variability of PV plants in Puerto Rico.

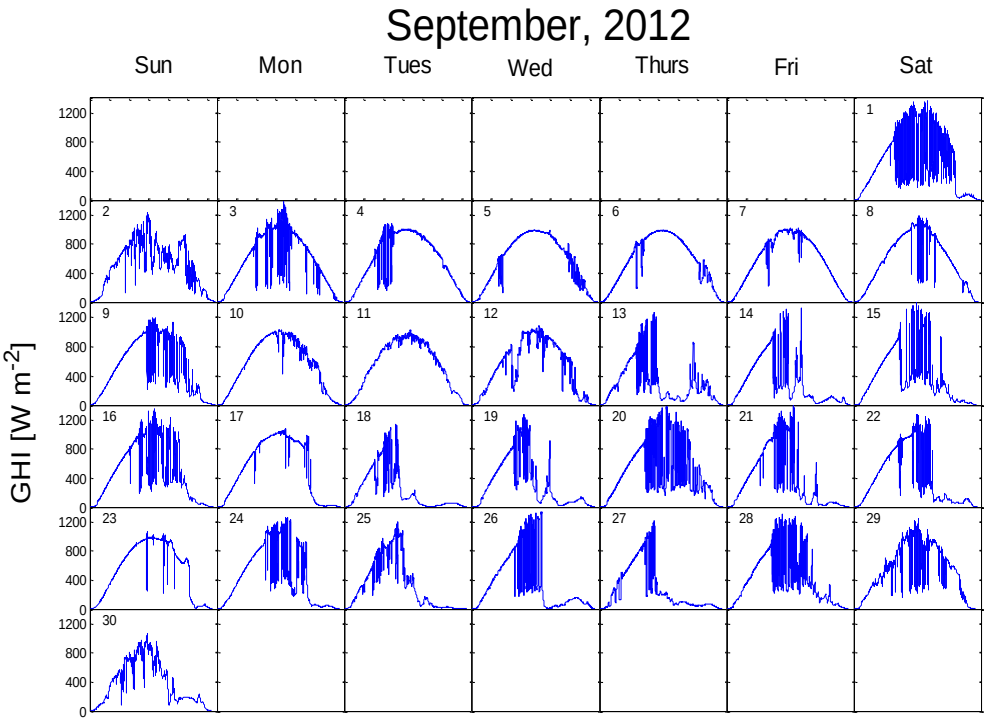

Fig. 7.1: Calendar showing the daily GHI profiles at Mayaguez, PR in September 2012.

7.3. **WVM Simulation**

For the one month of data, the WVM was used to simulate 5, 10, 20, 40, and 60MW square-shaped PV powerplants in Mayaguez with typical utility-scale PV density of 30 W m^{-2}. Particular attention was paid to the number of RRs greater than 10% of capacity ("violations"), due to the PREPA requirement. Table 7.1 shows the number of violations simulated by the WVM for each size of PV plant. By increasing plant size (and, hence, increasing geographic diversity), there is a noticeable decrease in violations: 737 for the 60MW plant versus 1322 for the 5MW plant. However, this decrease in violations does not scale linearly with increasing plant size. In all cases (5-60MW), the number of violations is significant, averaging at least 44 per day.

Table 7.1: RRs larger than 10% of capacity ("violations") in September 2012.

Plant Size	5MW	10MW	20MW	40MW	60MW
Violations in September 2012	1322	1192	1051	873	737

The number of violations per day changes depending on the variability of each day. Fig. 7.2 shows the timeseries of RRs and number of violations per day for the 60MW plant. The number of violations changes substantially by day, from a maximum of 110 violations on September 1st to having no violations on 6 other days. This shows the strong impact of meteorology.

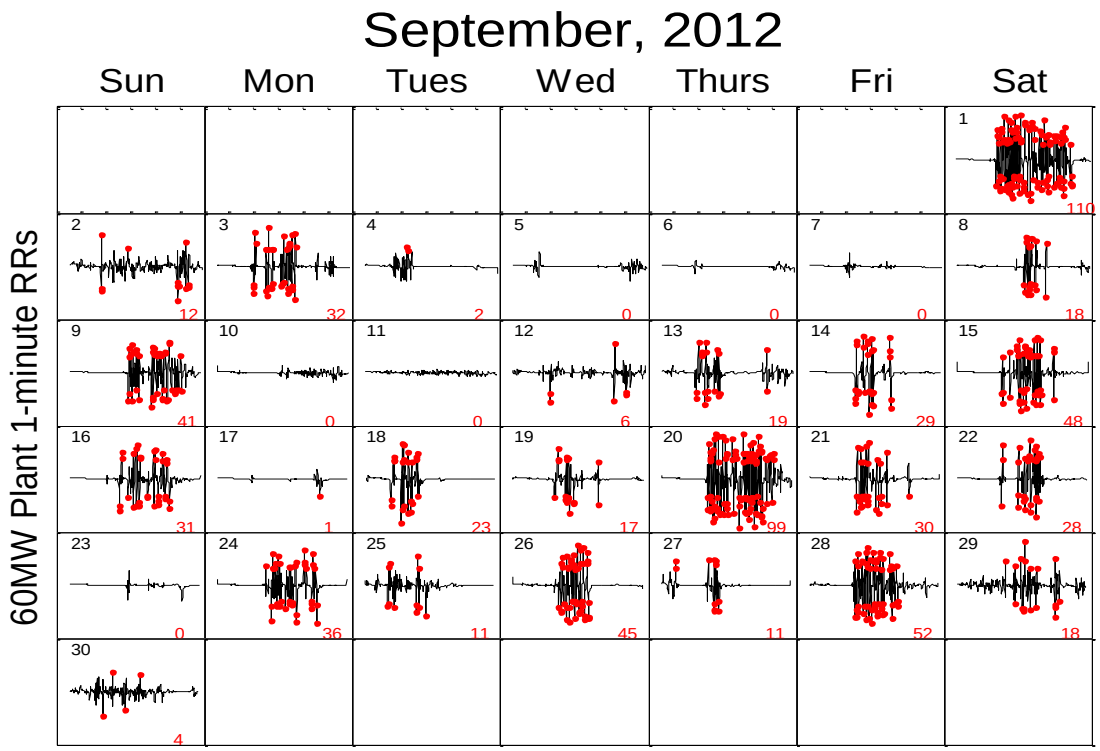

Fig. 7.2: RRs for 60MW plant. Violations are highlighted by red dots, and the total number of violations per day is shown in red at the bottom right of each plot.

To compare the different PV plant sizes, we can examine how many days a certain number of violations occurred per day. Fig. 7.3 shows these distributions. The 5MW plant had a maximum of 160 violations per day, while the 60MW plant had a maximum of only 110 violations. All but the 60MW plant had more than 5 days with more than 50 violations. The

number of violations on the most extreme days is likely the most important factor for sizing and controlling storage at these PV sites.

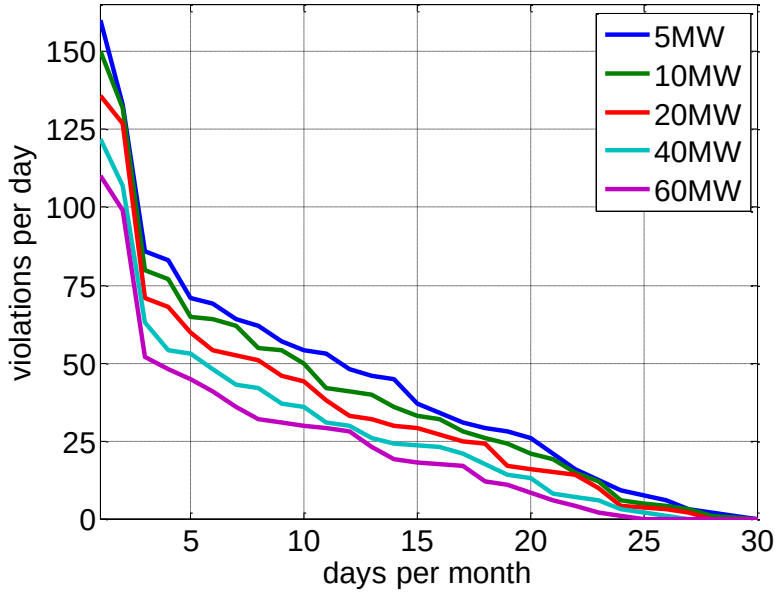

Fig. 7.3: Distributions showing how many days per month each number of violations per day will occur. For example, the 5MW plant had 5 days with 70 or more violations.

It is important to not just look at how many violations occur, but also how large the 1-minute RRs were, since that will influence the amount of storage and the control algorithms needed to comply with the PREPA requirement. Fig. 7.4 shows the number of occurrences of large 1-minute RRs. In the month, the 60MW plant had no RR larger than 30% of capacity, while the 5MW plant had 272. The maximum RR had the 5MW plant was over 50% of capacity. It is worth remembering, though, that in MWs, the maximum RR at the 60MW plant will still be much larger than the maximum RR at the 5MW plant: 18MW versus 2.5MW. As such, even though the 60MW plant will have less violations, and violations will tend to be less severe, it will still require a larger storage system (in terms of energy capacity) than the 5MW plant.

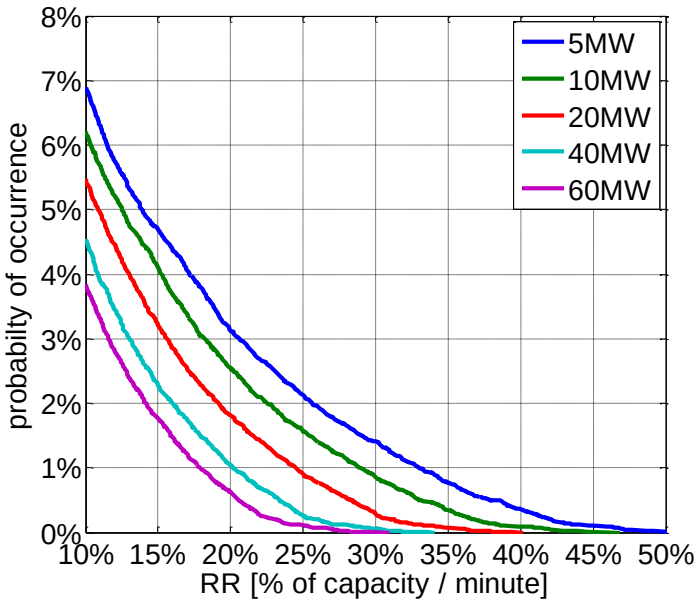

Fig. 7.4: Number of occurrence of large 1-minute RRs in September 2012.

The results presented here have important implications for Puerto Rico: PV plants in Mayaguez, PR will very often produce RRs larger than 10% of capacity. The number of violations per day can exceed 100, or nearly once per 5-minutes. In order to comply with the PREPA requirement, large amounts of batteries or other storage will be required, which will considerably increases the cost of installing PV systems in Puerto Rico.

The WVM developed in the dissertation research has proved very useful for PV plant and grid operators Solar developers can use the WVM to estimate the RRs that will occur at their plant, and module siting, plant sizing, and forecasting and storage requirements can be investigated and adjusted before the plant is installed. This has proven especially useful for PV plants that are given RR restrictions by electric authorities during their permitting.

7.4. **Future Work**

Future work should focus on both modifying the WVM to apply over larger areas and on running electric grid simulations with WVM simulated data to test the impact of powerplant variability to the electric grid.

Currently, the WVM assumes statistical homogeneity: that all sites simulated have the same irradiance statistics. This limits the application of the WVM to short lengthscales (e.g., a few kilometers). Over longer lengthscales, irradiance statistics will vary from site to site. To simulate the combination of different powerplants, the additional smoothing due to varying irradiance statistics over long distances should be further investigated.

While the WVM has been used to effectively estimate PV powerplant RRs, the effect of these RRs on the electric grid is not well known. Variables such as feeder or substation capacity, length, and customer locations as well as the location of the PV powerplant will all impact the effect that RRs have on voltage fluctuations and grid stability. Using grid simulation tools such as OpenDSS, the impact of PV RRs should be examined in detail (e.g., as presented in [4, 5]). In this way, the true impact of PV powerplant variability on the electric grid can be understood.

References

[1] Puerto Rico Electric Power Authority, Minimum Technical Requirements for Photovoltaic Generation (PV) Projects, http://www.fpsadvisorygroup.com/rso_request_for_quals/PREPA_Appendix_E_PV_Minimum_Technical_Requirements.pdf.

[2] DOE-Funded Solar Variability Model in High Demand in Puerto Rico, SunShot Initiative High Penetration Solar Portal, https://solarhighpen.energy.gov/article/doe_funded_solar_variability_model_in_high_demand_in_puerto_rico?print, 2012.

[3] J. Bosch, J. Kleissl, Deriving cloud velocity from an array of solar radiation measurements, submitted to Solar Energy, (2012).

[4] M.J. Reno, A. Ellis, J. Quiroz, S. Grijalva, Modeling Distribution System Impacts of Solar Variabiltiy and Interconnection Locations, World Renewable Energy Forum, Denver, CO, 2012.

[5] J. Quiroz, M. Reno, Detailed Grid Integration Analysis of Distributed PV, American Solar Energy Society, Denver, CO, 2012.