Available online at www.sciencedirect.com

SciVerse ScienceDirect

Solar Energy 86 (2012) 2170–2176

SOLAR ENERGY

www.elsevier.com/locate/solener

Review

Short-term irradiance variability: Preliminary estimation of station pair correlation as a function of distance

Richard Perez [a,*], Sergey Kivalov [a], Jim Schlemmer [a], Karl Hemker Jr. [a], Thomas E. Hoff [b]

[a] Atmospheric Sciences Research Center, The University at Albany, 251 Fuller Rd., Albany, NY 12203, USA
[b] Clean Power Research, 10 Glen Court, Napa, CA 94558, USA

Available online 18 April 2012

Communicated by: Associate Editor David Renne

Abstract

In this article, we report on the correlation between the irradiance variability observed at two neighboring sites as a function of their distance, and of the considered variability time scale. Correlation is the factor that determines whether the combined relative fluctuations of two solar systems add up when correlation is high, or attenuate when correlation is low.

Using one-dimensional virtual networks in 24 US locations and cloud motion derived from satellites as experimental evidence, we observe station pair correlations for distances ranging from 100 m to 100 km and from variability time scales ranging from 20 s to 15 min.

Within the limits of the assumptions from one-dimensional virtual networks, results show that the relationship between correlation, distance and time scale is predictable and largely independent of location and prevailing insolation conditions. Further, results indicate that the distance at which station pairs become uncorrelated is a quasi linear function of the considered time scale.

Keywords: Irradiance; Variability; High frequency; PV grid interaction

1. Introduction

The short-term variability of solar resource is perceived as a roadblock to the large scale deployment of solar power generation. This issue is the subject of several major research initiatives in the United States and internationally, (e.g., CSI, 2010; USDOE, 2009; BNL, 2010; SMUD, 2010; IEA, 2010).

In a recently published article, Hoff and Perez (2010a,b) advanced that the relative short-term variability of a fleet of identical PV generators decreases as the inverse of the square root of their number if the fluctuations of each system are uncorrelated. They defined relative short-term variability as the variability resulting from the fleet of systems, and quantified it by the standard deviation of the fleet's time series of changes in power output, normalized to the fleet's total capacity. More recently, Perez et al. (2011), building upon earlier work by Skartveit and Olseth (1992), showed that short-term variability for a single system at a given point in time could be estimated from hourly satellite-derived irradiances data such as Solar Anywhere (2010) or the NSRDB (1998–2005).

In this article we focus on station pairs, and investigate the correlation of their short-term variability as a function of their distance. A zero correlation would indicate that, per Hoff and Perez (2010a,b), their cumulative relative variability will be $1/\sqrt{2}$ times their individual relative variability. Further, the possible existence of negative correlation at some key distance would indicate that fluctuations tend to cancel out, as hypothesized in Hoff and Perez's optimum point.

* Corresponding author.
E-mail address: rperez@albany.edu (R. Perez).

0038-092X/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.solener.2012.02.027

2. Methods

2.1. Experimental data

Experimental measuring station pairs positioned at arbitrary distances and located in arbitrary climatic environments would constitute the ideal source of experimental data to undertake the present analysis. Unfortunately, this information is not fully available just yet. Although a few networks do exist where a partial validation of the present results will be possible (e.g., Kleissl, 2009), the necessary dense solar resource grids are on the drawing board or in the startup phase as of this writing – e.g., CSI (2010), SMUD (2010), BNL (2010).

However, as introduced by Hoff and Perez (2010a,b), there is an effective proxy to gridded networks of stations: ***virtual networks***. A virtual network consists of a single high-frequency measuring station, from which virtual stations can be inferred if the cloud speed aloft, V, is known. Letting $I_t^{station}$ be the irradiance measured at the station at time t, the irradiance at a neighboring virtual station, $I_t^{virtual}$ located at a distance L is given by:

$$I_t^{virtual} = I_{t_1}^{station} \tag{1}$$

with

$$t_1 = t - L/V \tag{2}$$

The virtual network concept makes two limiting assumptions:

(1) the virtual stations are located in the direction of the cloud motion vector; and
(2) the cloud fields stay nearly unchanged as they transit over the stations, although signal compression or extension is possible due to evolving cloud speeds.

The results presented in this paper are therefore valid within the framework of these two limiting assumptions. The second assumption may be considered as conservative because cloud field deformation over time would reduce the correlation of a pair of stations' fluctuations. The first limitation's impact will have to be evaluated in a planned follow-on study by analyzing high resolution network data when these become available, and by analyzing high resolution satellite images.

High frequency GHI and DNI data were obtained for 24 measuring stations, including 17 stations in the ARM network (Stokes and Schwartz, 1994) and seven stations in the SURFRAD network (SURFRAD, 2010). The ARM stations record data at a rate of three measurements per minute (20 s data) while the SURFRAD stations record data at a rate of one measurement per minute (1 min data). Fifteen months of data were analyzed for each station. A complete list of the stations is provided in Table 1.

Virtual networks were constructed around each station per Eqs. (1) and (2) by using the time/site specific cloud speeds produced as part of Solar Anywhere's operational cloud motion irradiance forecasts (Perez et al., 2010b). Site-time specific cloud speeds are derived by minimizing the RMSE of consecutive satellite images. This approach was first described and implemented by Lorenz et al. (2004). The cloud speeds have an operational frequency of one per hour. For the present analysis the cloud speed associated with each high frequency ground measurement was obtained via linear interpolation of the hourly speeds. Fig. 1 illustrates the hourly cloud speeds derived for a sample high variability day at the ARM central facility, superimposed upon the station's measured GHI.

2.2. Quantifying station pair variability correlation

Hoff and Perez (2010a,b) quantified the relative short-term variability of a fleet of N solar generators as:

$$\sigma_{\Delta t}^{\Sigma N} = \left(\frac{1}{C^{Fleet}}\right)\sqrt{Var\left[\sum_{n=1}^{N}\Delta \mathbf{P}_{\Delta t}^{n}\right]} \tag{3}$$

where C^{Fleet} is the total installed peak power of the fleet and $\Delta \boldsymbol{P}_{\Delta t}^{n}$ is the time series of changes in mean power output at the nth PV installation using a sampling time interval of Δt.

In this article, we focus our attention on the changes in the clearness index $\Delta \mathrm{Kt}^*$ instead of the changes in power output ΔP. Kt^* is the clear sky index (ratio between GHI and clear sky global irradiance, $\mathrm{GHI_{clear}}$). As such, Kt^* embodies the relative characteristics of flat plate PV systems' output fluctuations. For the present analysis, it offers two advantages: (1) normalizing variability to unity, and (2) removing the effect of solar geometry which is a source of variability, albeit fully predictable, as Δt increases. The $\Delta \mathrm{Kt}^*$ random variable time series is illustrated at the bottom of Fig. 1 for $\Delta t = 20$ s and 15 min.

The station pair correlation between the two random variables $\Delta Kt_{\Delta t}^{*station}$ and $\Delta Kt_{\Delta t}^{*virtual}$ is calculated independently for each day and each virtual network for virtual station distances ranging from 100 m to 100 km, and for sampling intervals of 20 s (only for the ARM-based networks), 1 min, 5 min and 15 min. Longer time intervals are not investigated since variability questions pertaining to longer time scales can already be addressed today by analyzing existing gridded satellite-derived data sets.

For a given pair of stations extracted from one of the 24 virtual networks, for a given day j, a given time interval Δt, and a given distance L, the station pair correlation, $Cor_{\Delta tj}^{L}$ is calculated from all high frequency data points in that day for solar elevations in excess of 10°. In effect the station-to-virtual is the station's time lag autocorrelation with the time lag depending both on the considered distance and the observed cloud speed.

For that network location, the prevailing station pair correlation, $Cor_{\Delta t}^{L}$, is derived as the weighted mean of each individual day's correlations. The weighting factor is the day's variability quantified by the daily variance of $\Delta Kt_{\Delta t,j}^{*}$.

Table 1
Experimental data.

	Station	Latitude	Longitude	Elevation (m)	Climate	Time span
Arm network	ARM-E27	35.27	96.74	386	Continental	1/09–4/10
	ARM-E19	35.56	98.02	421	Continental	1/09–4/10
	ARM-E20	35.56	96.99	309	Continental	1/09–4/10
	ARM-E21	35.62	96.07	240	Continental	1/09–4/10
	ARM-E15	36.43	98.28	418	Continental	1/09–4/10
	ARM-C1	36.61	97.49	318	Continental	1/09–4/10
	ARM-E13	36.61	97.49	318	Continental	1/09–4/10
	ARM-E12	36.84	96.43	331	Continental	1/09–4/10
	ARM-E16	36.06	99.13	602	Continental	1/09–4/10
	ARM-E11	36.88	98.29	360	Continental	1/09–4/10
	ARM-E10	37.07	95.79	248	Continental	1/09–4/10
	ARM-E9	37.13	97.27	386	Continental	1/09–4/10
	ARM-E7	37.38	96.18	283	Continental	1/09–4/10
	ARM-E6	37.84	97.02	409	Continental	1/09–4/10
	ARM-E4	37.95	98.33	513	Continental	1/09–4/10
	ARM-E1	38.20	99.32	632	Continental	1/09–4/10
	ARM-E2	38.31	97.30	450	Continental	1/09–4/10
Surfrad network	Goodwin creek	34.25	89.87	98	Subtropical	1/09–4/10
	Desert rock	36.63	116.02	1007	Arid	1/09–4/10
	Bondville	40.05	88.37	213	Continental	1/09–4/10
	Boulder	40.13	105.24	1689	Semi-arid	1/09–4/10
	Penn state	40.72	77.93	376	Humid continental	1/09–4/10
	Sioux falls	43.73	96.62	473	Continental	1/09–4/10
	Fort peck	48.31	105.10	634	Continental	1/09–4/10

$$Cor^{L}_{\Delta t} = \frac{\sum_{j=1}^{n}(Cor^{L}_{\Delta t j} * Var[\Delta Kt^{*}_{\Delta t,j}])}{\sum_{j=1}^{n} Var[\Delta Kt^{*}_{\Delta t,j}]} \quad (4)$$

where n is the total number of station-days analyzed.

The resulting correlation is thus influenced by the days with the highest variability and least influenced by days with little or no variability such as clear days when (1) variability is not an issue and (2) individual day correlations may depart from the trend because they reflect other than cloud induced transients such as solar geometry.[1]

3. Results

The objective is to understand how station pair correlation varies as a function of distance and the considered sampling interval.

Starting with the observation of the relationship obtained from one day's worth of observations at one of the virtual network, we proceed with analyzing the composite trend resulting from all the days analyzed at that same network location per Eq. (4), investigating how the single day's relationship evolves. Finally we observe how the relationship further evolves when all 24 virtual network locations are considered.

[1] Using kt^{*} removes most but not all effects of solar geometry because the reference clear sky is not exactly calibrated for each data point's conditions.

3.1. Single virtual network, single day example

For the single day example we selected a highly variable day for the ARM central facility, April 19, 2009, illustrated in Fig. 1.

The virtual station pair **correlation** for that day is plotted in Fig. 2 as a function of distance for each sampling interval.

Correlation reaches zero at respectively 350 m, 800 m, 2.7 km and 7.8 km for sampling intervals of 20 s, 1 min, 5 min and 15 min. Beyond the zero crossover, the considered station pairs do not exhibit any significant positive correlation. It is interesting to note that in each case, correlation becomes slightly negative beyond the zero crossover point. The negative correlation indicates that at some distance on that day (respectively 600 m, 1 km, 5 km and 10 km for Δt of 20 s, 1 min, 5 min and 15 min) the effect of the passing clouds resulted in a partial cancelation of fluctuations, when the succession of cloudy and sunny periods tended to be in opposition of phase for all time scales at the two sites. This negative correlation effect might be traceable to the one-dimensional nature of the virtual network and will have to be investigated further when two-dimensional real network data become available.

3.2. Single virtual network, all data

Fig. 3 is similar to Fig. 2, but the results are based on all 452 days analyzed at the ARM central facility. The line for each time scale is the weighted mean of individual days' correlations per Eq. (4). The dotted lines represent plus

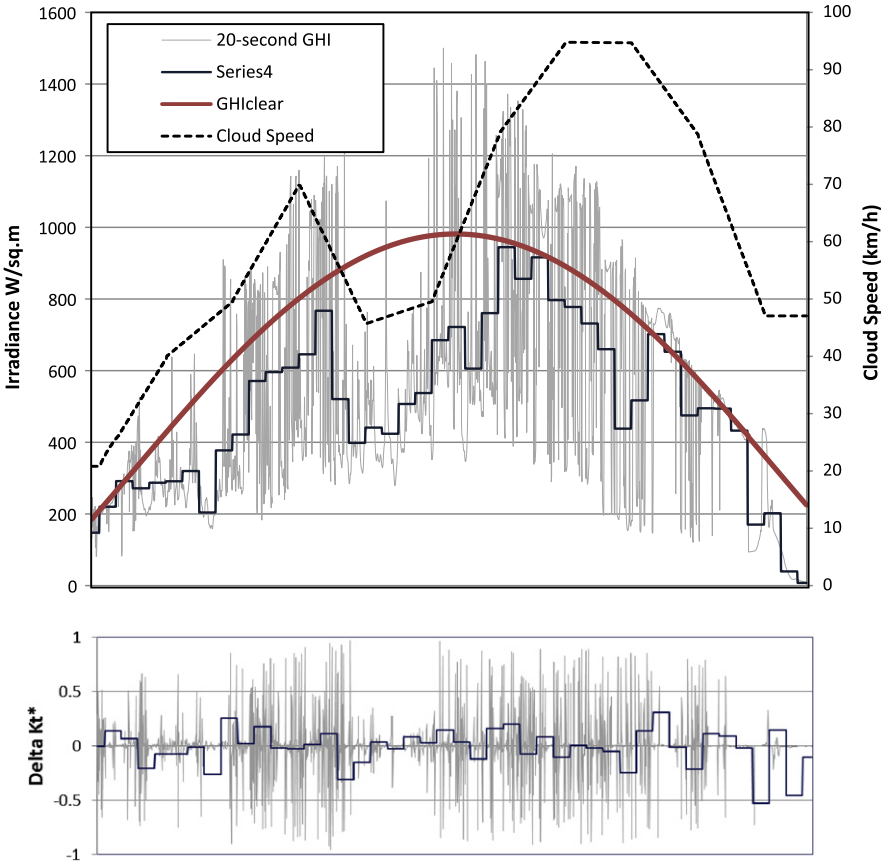

Fig. 1. Sample high-variability day in the ARM network, showing 20 s GHI, 15 min GHI and hourly interpolated cloud speeds.

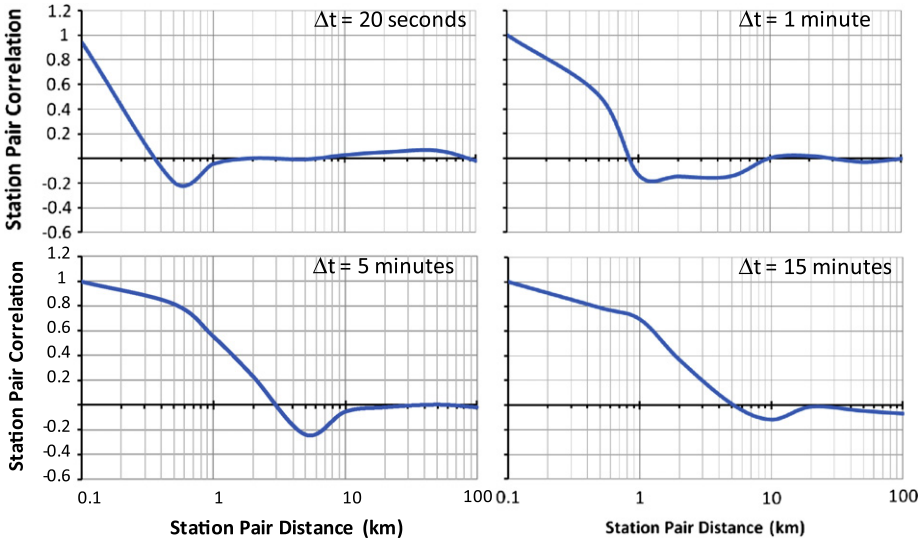

Fig. 2. Single virtual network, single day station pair correlation as a function of distance and time scale.

or minus one standard deviation around the weighted mean. The standard deviation is computed from the weighted sample variance.

The resulting traces cross the zero correlation threshold at respectively 400 m, 850 m, 3 km and 9 km for each considered sampling intervals, i.e., remarkably close to the single day example shown above. The tightness of the standard deviation indicates that individual days do not depart significantly from the trend, particularly for shorter sampling intervals (20 s and 1 min).

Interestingly, the resulting trend for the ARM site conserves the small negative correlation peak observed for the

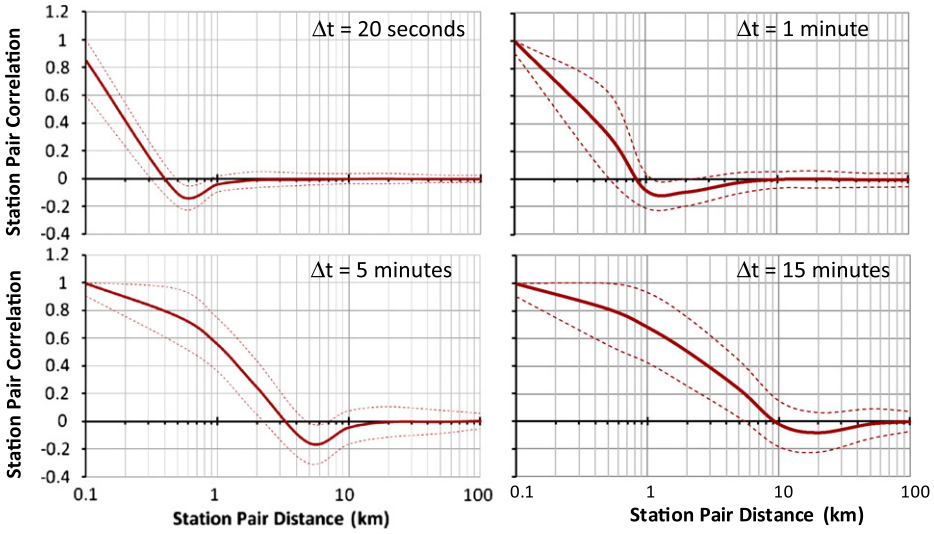

Fig. 3. Single virtual network's mean station pair correlation trends resulting from 15 months of data. The dotted line represent plus or minus one standard deviation around the mean trend.

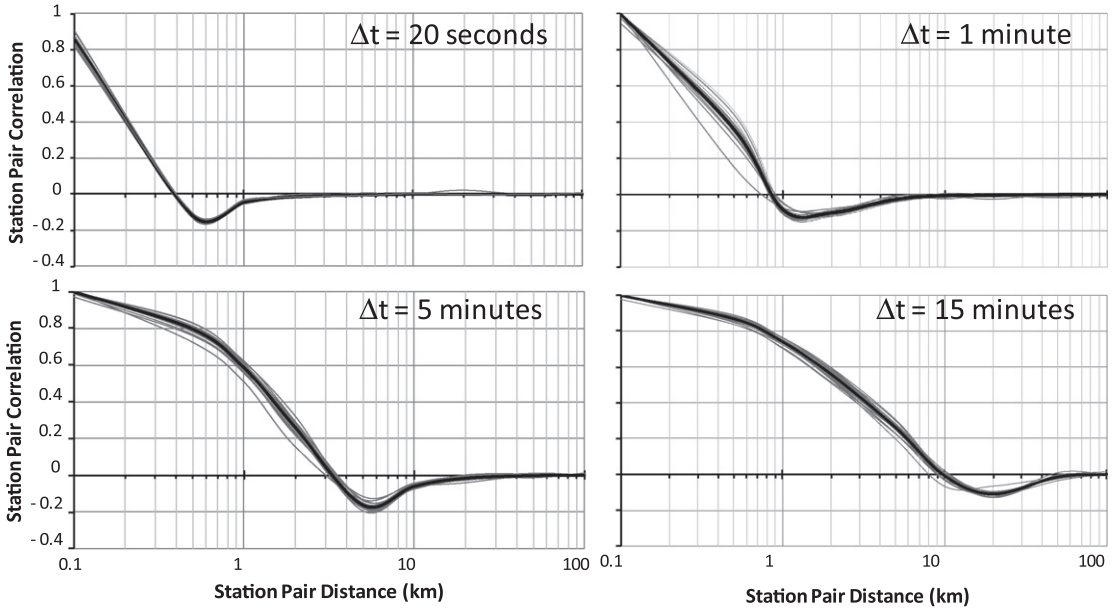

Fig. 4. Station pair correlation trends resulting from all virtual networks and 15 months of data. Thin gray lines represent individual networks.

individual day, despite the differing cloud speeds and cloud shapes that occurred each hour of each day.

3.3. All virtual networks, all data

Fig. 4 is similar to Figs. 2 and 3, but includes all data points analyzed at all locations (i.e., nearly 17 million 20 s data points). The resulting heavy line is the 24-network mean (only 17 networks for $\Delta T = 20$ s). Individual networks are represented by the thin gray lines.

The agreement between all networks, including nearly identical zero crossover points and negative correlation peaks, is remarkable given the diversity of possible weather conditions and cloud variability drivers in highly differing climatic environments.

4. Discussion

The virtual network analysis undertaken here on a large array of climatic environments, weather drivers, and seasonal conditions leads to a remarkably well defined set of trends linking distance, fluctuation frequency and station pair correlation.

The evidence from this exhaustive analysis suggests that 20 s fluctuations become uncorrelated positively at a distance of less than 500 m. The distances are respectively

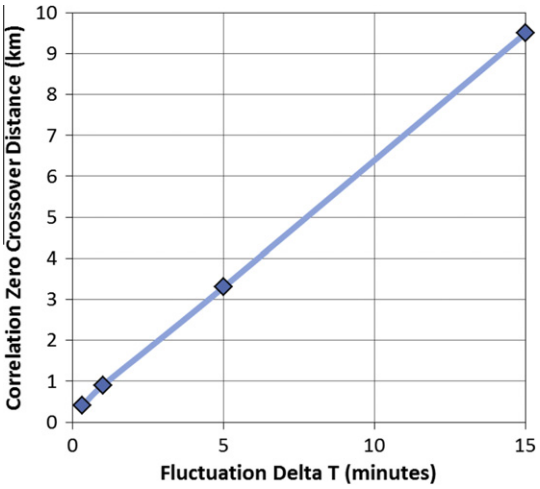

Fig. 5. Zero correlation crossover distance as a function of fluctuation time scale.

1 km, 4 km and 10 km for fluctuation time scales of 1, 5 and 15 min respectively.

The relationship between the zero correlation crossover distance and Δt is quasi linear as can be seen in Fig. 5. This quasi-linearity is consistent with the concept of the *dispersion factor* introduced earlier by Hoff and Perez (2010a,b) which embodies time scale, cloud speed and distance into one single parameter determining variability. Ongoing studies by Mills and Wiser (2010) and Hoff and Perez (2010a,b) **indicate that** the linear relationship could be extended at either end of the current Δt span to estimate station pair correlation for other time scales.

Extrapolating the present results to the case of a homogeneously dispersed solar resource in a metropolitan area such as the greater New York city area (40×40 km) suggests that the high frequency (20 s) variability experienced by a single small system should be reduced by a factor of

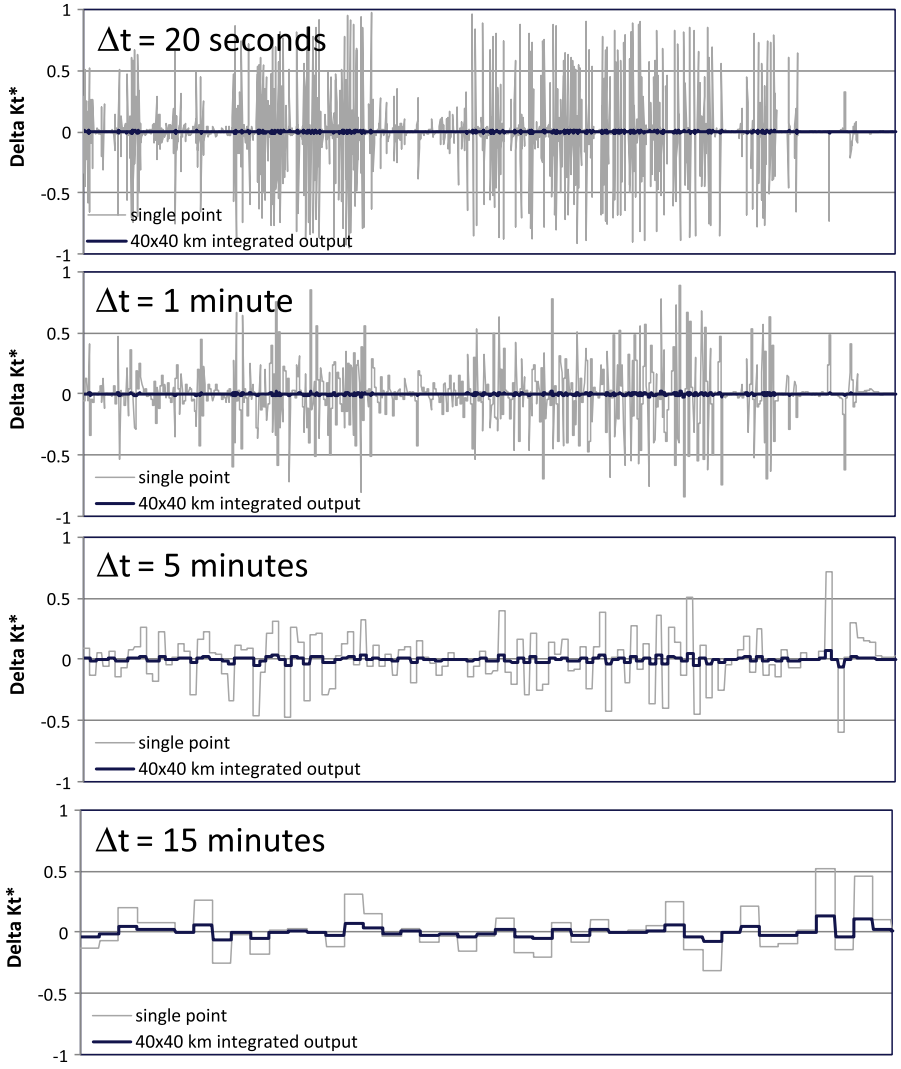

Fig. 6. Comparing single site and 40×40 km extended variability for different fluctuation time scales (at 20 s, the considered area includes ~64,000 uncorrelated locations and at 15 min it includes ~16 such locations).

80 when considering the entire metropolitan area.[2] The variability reduction would become smaller as the considered frequency increases: the metropolitan variability would be reduced by respectively 40, 10 and 4 for fluctuation time scales of 1, 5 and 15 min. This example is illustrated in Fig. 6, comparing the $\Delta Kt^*_{\Delta t}$ time series of a single installation to that of a homogeneous hypothetical deployment of systems over a 40 × 40 km area.

The findings of this study are consistent with limited evidence assembled from measurement station pairs – e.g., see Mills and Wiser, 2010; Soubdhan and Calif, 2010; Murata et al., 2009. However, these station pairs are too far apart to provide an exhaustive validation, hence the present findings will have to be substantiated by follow-on work. In particular, Clean Power Research recently deployed a low-cost 25-station modulable network in Central California designed to capture the relative ΔGHI data streams down to scales of 10 s (Hoff and Perez, 2010a,b). Further, as a planned next step of this work, we will analyze high resolution (1 km) satellite images and apply the cloud motion forecast algorithm (Perez et al., 2010b) to simulate high resolution, high frequency irradiance time series over any arbitrary extended area.

We will use these upcoming bottom-up and top-down experimental data to verify, and as needed adjust, the present virtual network-derived results. It is likely that these future two-dimensional network verifications will refine these preliminary results – in particular the negative correlation peak observed for all sites and time scales will probably be reduced or disappear. The negative correlation peaks are largely a product of the one-dimensionality of the virtual networks oriented in the direction of cloud movement. The decline should be more gradual in other directions. As a consequence, the distance vs. time scale relationships shown in Fig. 5 will have to be refined after analyzing two-dimensional experimental data from different climatic environments. However, the absence of correlation occurring consistently at all sites at distances beyond the negative correlation peak should hold.

Acknowledgements

This study was funded by Clean Power research under a California Solar Initiative (CSI) Grant Agreement titled "Advanced Modeling and Verification for High Penetration PV." The California Public Utilities Commission is the Funding Approver, Itron is the Program Manager, and the California IOUs are the Funding Distributors.

[2] i.e., considering the combined output of 64,000 uncorrelated systems and applying the $1/\sqrt{N}$ relationship from Hoff and Perez (2010).

References

BNL, Brookhaven National Laboratory, 2010. Personal Communication – Overview of BNL's Research Agenda on Solar Energy. BNL, Upton, NY.

CSI, California Solar Initiative R&D Program, 2010. California Public Utility Commission, San Francisco, CA.

Hoff, T., Perez, R., 2010a. Quantifying PV power output variability. Solar Energy 84 (10), 1782–1793.

Hoff, T., Perez, R., 2010. PV Power Output Variability: Correlation Coefficients. Technical Report to the California Solar Initiative, Grant Agreement for Advanced Modeling and Verification for High Penetration PV.

IEA Solar Heating and Cooling Programme, 2010. Task 46 – Solar Resource Assessment and Forecasting, Subtask A – Resource Applications for High Penetration of Solar Technologies. International Energy Agency, Paris, France.

Kleissl, J., 2009. <http://maeresearch.ucsd.edu/kleissl/demroes/map.html>.

Lorenz, E., Hammer, A., Heinemann, D., 2004. Short term forecasting of solar radiation based on satellite data. EUROSUN2004 (ISES Europe Solar Congress), Freiburg, Germany.

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

Murata, A., Yamaguchi, H., Otani, K., 2009. A method of estimating the output fluctuation of many photovoltaic power generation systems dispersed in a wide-area. Electrical Engineering in Japan 166 (4), 9–19.

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

Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr., K., Renne, D., Hoff, T., 2010b. Validation of short and medium term operational solar radiation forecasts in the US. Solar Energy 84 (12), 2161–2172.

Perez, R., Kivalov, S., Hoff, T., 2011. Spatial & temporal characteristics of solar radiation variability. In: Proceeding of International Solar Energy World Congress. Kassel, Germany.

Skartveit, A., Olseth, J.A., 1992. The probability density of autocorrelation of short-term global and beam irradiance. Solar Energy 46 (9), 477–488.

SMUD (ask Obadiah) — alternatively, use Mark Rawson, 2010. Solar and the smart grid session SPI 2010.

Solar Anywhere, (2010). <www.solaranywhere.com>.

Soubdhan, T., Calif, R., 2010. Spatio-temporal analysis of solar radiation measured at two Sites in Guadeloupe. In: EuroSun 2010 Conference. Graz, Austria.

Stokes, G.M., Schwartz, S.E., 1994. The atmospheric radiation measurement (ARM) program: programmatic background and design of the cloud and radiation test bed. Bulletin of American Meteorological Society 75, 1201–1221.

SURFRAD Network 2010. Monitoring Surface Radiation in the Continental United States. <www.srrb.noaa.gov/surfrad/>.

USDOE, US Department of Energy, 2009. Recovery Act: High Penetration Solar Deployment R&D Program. <http://www1.eere.energy.gov/solar/financial_opps_detail.html?sol_id=286>.