Modeling PV Fleet Output Variability

Thomas E. Hoff and Richard Perez

ABSTRACT

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

INTRODUCTION

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

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

The present objective is to develop analytical methods and tools to quantify PV fleet output variability. Variability in time intervals ranging from a few seconds to a few minutes is of primary interest since control area reserves are dispatched over these time intervals. For example, regulation reserves might be dispatched by an ISO (Independent System Operator) every five seconds through a broadcast signal. Knowledge about PV fleet variability in five-second intervals could be used to determine the resources necessary to provide frequency regulation service in response to power fluctuations.

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

$$\sigma_{\Delta t}^{fleet} = \sqrt{Var\left[\sum_{n=1}^{N} \Delta \boldsymbol{P}_{\Delta t}^{n}\right]} \qquad (1)$$

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

It is helpful to graphically illustrate what is meant by output variability. The left side of Figure 1 presents 10-second irradiance data (PV power output is almost directly proportional to irradiance) and the right side of the figure presents the change in irradiance using a 10-second time interval for a network of 25 weather monitoring stations in a 400-meter by 400-meter grid located at Cordelia Junction, CA on November 7, 2010. The light gray lines correspond to irradiance and variability for a single location and the dark red lines correspond to average irradiance distributed across 25 locations. This suggests that spreading capacity across 25 locations rather than concentrating it at a single location reduces variability by more than 70 percent in this particular instance.

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

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

A more viable approach is to streamline the calculations through the use of a general-purpose PV output variability methodology. The method needs to quantify short-term fleet power output variability based on the premises that sky clearness and sun position drive the changes in the short-term output for

individual PV systems and that technical specifications (i.e., dimensions, plant spacing, number of plants, etc.) determine overall fleet variability.

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

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

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

(2)

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

MAXIMUM OUTPUT VARIABILITY MODEL

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

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

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

An alternative approach is to construct a data set that simulates worst case variability conditions. The theoretically worst case variability of a single PV plant would be that it cycles alternately between 0 and

---

[1] See Equation (8) in Hoff and Perez (2010).
[2] One of the few examples of this sort of data is provided by Kuszamaul, et. al. (2010).

100 percent of its rated output every time interval. For example, suppose that the PV plant is rated at 1 MW and the time interval of interest is 1 minute. As illustrated in Table 1, maximum variability occurs when the PV plant is at full power at 12:00, zero power at 12:01, full power at 12:02, etc. The corresponding change in power fluctuates between -1 and 1 MW. The standard deviation[3] of the change in power output equals 1 MW per minute. That is, a 1 MW PV plant that is exhibiting maximum variability over a 1 minute time interval has a 1 MW per minute standard deviation. This would imply that 1 MW of reserve capacity is required to compensate for the output variability for a single plant.

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

The second scenario, illustrated in Table 3, assumes that the plants cycle on and off alternately with a time shift of 1 minute. In this case, the change in output from the first location cancels the change in output at the second location. The result of this scenario is a standard deviation of 0 MW per minute.

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

Feasible Maximum Output Variability

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

To correct the worst case scenario, retain the assumption that each power change is either a transition from zero output to full output or from full output to zero output. This assumption in itself is highly conservative since the impacts of cloud transients on PV systems will almost never produce changes with magnitudes as high as 100 percent of rated output and will generally produce changes much less

---

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

than 100 percent. As for timing, rather than being synchronized, each system is assumed to cycle on and off in a random fashion, representing fleets of PV systems with outputs that are uncorrelated.

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

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

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

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

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

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

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

---

[4] See Figure 13 in Kuszamaul et. al. (2010) and Figure 7 in Mills et. al. (2009).
[5] $\sigma = \sqrt{[(0.25)(-1)^2 + (0.50)(0)^2 + (0.25)(1)^2] - [(0.25)(-1) + (0.50)(0) + (0.25)(1)]^2} = \frac{1}{\sqrt{2}}$

Example

Suppose that a utility system plans to incorporate 5,000 MW of PV. Figure 2 presents the maximum output variability calculated using Equation (4) for PV fleets with capacities ranging from 0 to 5,000 MW based on two fleet composition strategies. The blue line is the variability when the fleet is composed of uncorrelated 1 MW systems. The red line is the variability when the fleet is composed of uncorrelated 100 MW systems. As illustrated in the figure at the 5,000 MW level, if 100 MW systems are installed at 50 locations (N=50) with uncorrelated changes in output, maximum output variability is 500 MW per time interval, or 10 percent of fleet capacity. However, if 1 MW PV systems are installed at 5,000 locations (N=5,000) with uncorrelated changes in output, maximum output variability is 50 MW, or 1 percent of fleet capacity.[6]

This example illustrates the potential benefit of dividing the PV capacity into small systems, and spreading them apart geographically so that output changes are uncorrelated. More importantly, it also illustrates the unnecessary potential cost that could be incurred if system planners were to procure reserves without adequate tools for quantifying PV variability. The dotted line in figure 2 represents the reserve resources that would be procured when each MW of PV was fully "backed up" with a MW of fossil, battery, or other dispatchable resource. In the N=5,000 example, such a planning practice — at least for fleets made up of uncorrelated systems— would result in capital expenditures 99 times the required amounts.

General Model

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

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

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

The covariance between any two plants equals the standard deviations of each of the locations times the correlation coefficient between the two locations (i.e., $COV\left(\Delta P_{\Delta t}^{i}, \Delta P_{\Delta t}^{j}\right) = \sigma_{\Delta t}^{i}\sigma_{\Delta t}^{j}\rho_{\Delta t}^{i,j}$). As a result,

---

[6] Appendix A illustrates how to verify these results using an Excel spreadsheet.

$$\sigma_{\Delta t}^{fleet} = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} \sigma_{\Delta t}^{i}\sigma_{\Delta t}^{j}\rho_{\Delta t}^{i,j}} \qquad (6)$$

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

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

CORRELATION VERSUS DISTANCE

Background: Critical Factors Affecting Correlation

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

Mills and Wiser (2010) analyzed measured one-minute insolation data over an extended period of time for 23 time-synchronized sites in the Southern Great Plains network of the Atmospheric Radiation Measurement (ARM) program. Their results demonstrated[7] that the correlation of the change in the global clear sky index (1) decreases as the distance between sites increases and (2) decreases more slowly as the time interval increases.

Perez et. al. (2010b) analyzed the correlation between the variability observed at two neighboring sites as a function of their distance and of the considered variability time scale. They used 20-second to one-minute data to construct virtual networks at 24 US locations from the ARM network and the SURFRAD Network and cloud speed derived from SolarAnywhere (2011) to calculate the station pair correlations for distances ranging from 100 meters to 100 km and from variability time scales ranging from 20 seconds to 15 minutes. Their results demonstrated that the correlation of the change in global clear sky index (1) decreases as the distance between sites increases and (2) decreases more slowly as the time interval increases.

---

[7] See Figure 5 in Mills and Wiser (2010).

The consistent conclusions[8] of these studies are that correlation: (1) decreases as the distance between sites increases and (2) decreases more slowly as the time interval increases. Hoff and Perez (2010) add that the correlation decreases more slowly as the speed of the clouds increases.

Determination of station pair correlation

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

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

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

Let $\boldsymbol{\Delta Kt^{*1}_{\Delta t}}$ and $\boldsymbol{\Delta Kt^{*2}_{\Delta t}}$ represent two sets of observed data values for the change in the clearness index that have a mean of 0 and standard deviations, σ_1 and σ_2.[9]

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

$$\rho^{1,2} = \frac{E\left[\boldsymbol{\Delta Kt^{*1}_{\Delta t}} \boldsymbol{\Delta Kt^{*2}_{\Delta t}}\right]}{\sigma_1 \sigma_2} \tag{8}$$

Macro scale satellite data analysis

The correlation coefficient was computed for station pairs extracted from the gridded satellite-derived SolarAnywhere dataset (2011). As summarized in Table 5, three separate geographic regions in the United States were selected for analysis: Southwest, Southern Great Plains, and Hawaii. The first location, was selected using a grid size of 2.0°, 1.0°, or 0.5° for the Southwest, Southern Great Plains, and Hawaii, correspondingly. The second location was selected between 0.1° and 2.9° (about 10 to 300

---

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

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

km) from the first location (other map coordinates were available but the illustrated points provided sufficient data and ease of analysis, so were ignored). Hourly insolation data were obtained for each of the two locations covering the period January 1, 1998 through September 30, 2010, The site-pair correlation analysis was then performed as described above for time intervals (Δt) of 1, 2, 3, and 4 hours. This analysis resulted in more than 70,000 correlation coefficients.

Results:

Figure 3 presents the correlation coefficients for the Southwest.[10] The columns summarize the results for each time interval and the rows present the measured correlation coefficients versus several alternative candidate sets of variables. The first column summarizes results for a time interval of 1 hour. The second, third, and fourth columns plot the same results using time intervals of 2, 3, and 4 hours. Results in the top row present correlation coefficients versus the distance between the two locations. Results in the middle row present correlation coefficients versus distance divided by time interval. Results in the bottom row present correlation coefficients versus distance divided by time interval multiplied by an implied relative cloud speed;[11] this estimated cloud speed is related to the Dispersion Factor introduced by Hoff and Perez (2010). The dashed line in the bottom figures represents the results of a generalized method, proposed in this paper for use in future tools, that will be validated in the present analysis. Results are calculated using parameters obtained from SolarAnywhere. Figure 4 and Figure 5 present comparative results for the Great Plains and Hawaii. The patterns presented in the figures are similar across all time intervals in the three geographic locations. Figure 6 compresses the results for each location and presents results where all time intervals are combined into the same figure.

Discussion

The analysis provides several key findings. First, consistent with previous studies, the correlation coefficients decrease with increasing distance (top row of Figure 6). Second, also consistent with previous studies, this decrease occurs more slowly with longer time intervals (top row of Figure 6). An alternative way of viewing this result is that correlation coefficients decrease at a similar rate when plotted versus distance divided by time interval (middle row of Figure 6). Third, the scatter in results is further decreased when an implied relative speed[11] is introduced for the first location in the pair of locations (bottom row of Figure 6). Finally, a model, shown by the dashed black line in the bottom row of Figure 6, fits the empirical data well when calibrated using the location-specific derived input parameters, where

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

---

[10] The data in all of the figures represent randomly selected samples of points in order to make the results more readable.

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

Microscale Analysis

Projection to shorter time intervals

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

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

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

The solid lines connecting the four time interval observations for each location in Figure 7 illustrate that the relationship is linearly related to the time interval. The figure begs the question as to whether the results can be projected in the region with shorter time intervals (i.e., the gray sections of the figures).

Evaluation of Time-Independence Claim

The above linear relationship suggests that the method is independent of selected time interval, even down to the very short time intervals (several seconds to several minutes) that are of primary interest to the utilities. This section provides an initial validation of time-independence by comparing results calculated from the one-hour SolarAnywhere data set against results from independent studies that used 10-second, 20-second, and one-minute data sets.

Geographic Diversity Study: Mills and Wiser (2010) used measured one-minute insolation data for 23 time-synchronized sites in the Southern Great Plains network of the Atmospheric Radiation Measurement (ARM) program to characterize the variability of PV with different degrees of geographic diversity. That report presented[12] the correlation of changes in global clear sky index between these geographically dispersed sites. Mills and Wiser provided an electronic version of their results and these were used to compare against the general method proposed here. While the one-hour SolarAnywhere data set was used as input to the general method, correlation coefficients were calculated that

---

[12] Figure 5 in Mills and Wiser (2010).

corresponded to much shorter time intervals in the Mills and Wiser study. The results, presented in Figure 8, are comparable to the Mills and Wiser study even down to one-minute time intervals.[13]

Virtual Network Study: Perez et. al. (2010) obtained 20-second to one-minute insolation data for 24 measuring stations, including 17 stations in the ARM network and 7 stations in the SURFRAD network. They constructed one-dimensional virtual networks[14] using satellite-derived cloud speeds to translate time measurements into space measurements. They then calculated correlation coefficients between the change in clearness index for various time intervals and distances. Figure 9 presents some of the key results from that study. Figure 10 re-plots the data from the virtual network study along with corresponding projections from Equation (9). Results compare well to virtual network study down to correlation coefficients of 40 percent for time intervals between 20 seconds to 15 minutes. Results from the virtual network study correlation coefficients below 40 percent may be lower as a result of the negative correlation arising from locations that are very close together and because of the one dimensional nature of the virtual networks.

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

This network was deployed at Cordelia Junction, CA in a 400-meter by 400-meter configuration (a square composed of 100 meters between stations). Figure 11 presents the correlation coefficients for November 7, 2010. Since there are 25 locations, there are 625 possible combinations, 300 of which are unique. Each of these combinations was evaluated using nine different time intervals (10, 20, 30, 40, 50, 60, 90, 120, and 300 seconds). Thus, there are 2,700 unique scenarios.

The black line in the figure represents the relationship proposed in Equation (9). These independently measured data fit the proposed method fairly well. It is interesting to note that this data set exhibits some of the negative correlation effects identified by Hoff and Perez (2010) and Perez, et. al. (2010b) using the virtual network approach.

CONCLUSIONS

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

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

---

[13] The minor differences in 180-minute time intervals are due to methodological differences between the two studies.
[14] See Hoff and Perez (2010) for a discussion of virtual network construction.

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

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

Acknowledgements

References

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

Kuszamaul, S., Ellis, A., Stein, J., Johnson, L. 2010. Lanai High-Density Irradiance Sensor Network for Characterizing Solar Resource Variability of MW-Scale PV System. 35th Photovoltaic Specialists Conference, Honolulu, HI. June 20-25, 2010.

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

Mills, A., Alstrom, M., Brower, M., Ellis, A., George, R., Hoff, T., Kroposki, B., Lenox, C., Miller, N., Stein, J., Wan, Y., 2009. Understanding variability and uncertainty of photovoltaics for integration with the electric power system. Lawrence Berkeley National Laboratory Technical Report LBNL-2855E.

Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr., C. , Hoff, T. E. 2011. Parameterization of site-specific short-term irradiance variability, Solar Energy – in press.

Perez, R., Kivalov, S., Schlemmer, J., Hemker Jr., C. , Hoff, T. E. 2010b. Short-term irradiance variability correlation as a function of distance. Submitted to Solar Energy.

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

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.

Wikipedia. 2010. http://en.wikipedia.org/wiki/Correlation_and_dependence.

Woyte, A., Belmans, R., Nijs, J. 2007. Fluctuations in instantaneous clearness index: Analysis and statistics. Solar Energy 81 (2), 195-206.

Table 1. Maximum change in power output at one location.

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

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

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

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

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

Table 4. Maximum change in power output assuming random output.

Time	Power (MW)		Change (MW/min)
12:00	50% Plant On (1)	50% Plant Off (0)	25% chance of 1 50% chance of 0 25% chance of -1
12:01	50% Plant Off (0) 50% Plant On (1)	50% Plant Off (0) 50% Plant On (1)	
	Scenario 1	*Scenario 2*	

Table 5. Summary of input data.

Region	Southwest	Southern Great Plains	Hawaii
Location #1	Latitude: 32° to 42° Longitude: -125° to -109° Grid Size: 2.0°	Latitude: 35° to 38° Longitude: -99° to -96° Grid Size: 1.0°	Latitude: 19° to 20° Longitude: -156° to -155° Grid Size: 0.5°
Location #2	0.1°, 0.3°, …, 1.9° from #1	0.1°, 0.3°, …, 2.9° from #1	0.1°, 0.2°, …, 1.0° from #1
Time Intervals	1, 2, 3, and 4 hours	1, 2, 3, and 4 hours	1, 2, 3, and 4 hours
Clear Sky Irradiance	10 irradiance bins in intervals of 0.1 kW/m^2	10 irradiance bins in increments of 0.1 kW/m^2	10 irradiance bins in increments of 0.1 kW/m^2
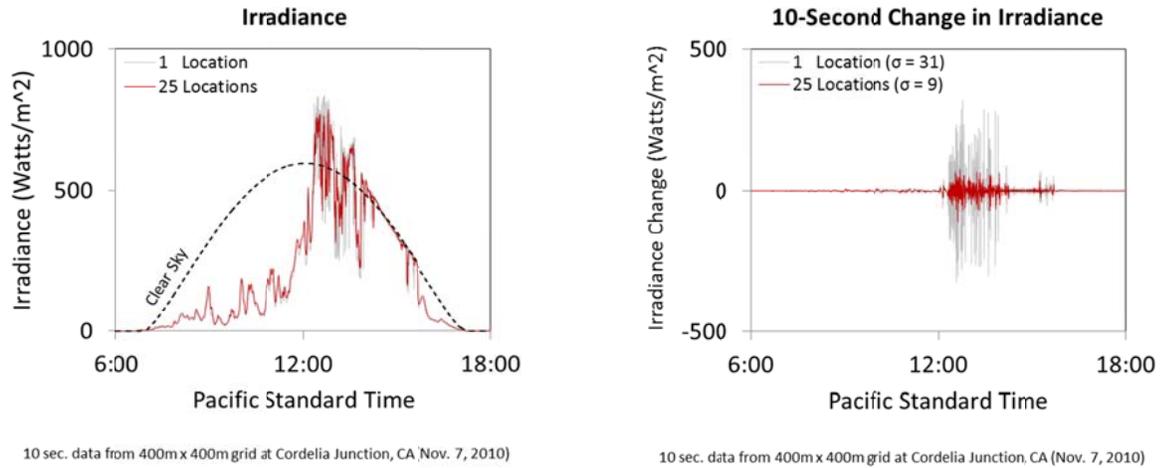
Figure 1. Twenty-five location network reduces 10-second variability by more than 70 percent in a 400 meter x 400 meter grid at Cordelia Junction, CA on November 7, 2010.
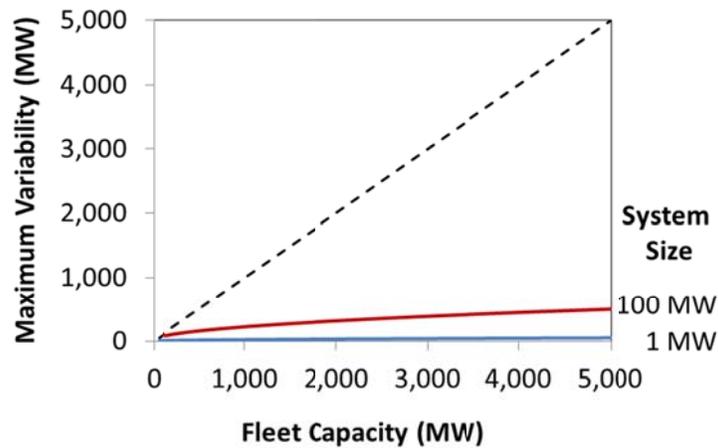
Figure 2. Maximum variability for 1 MW and 100 MW system sizes with uncorrelated changes.
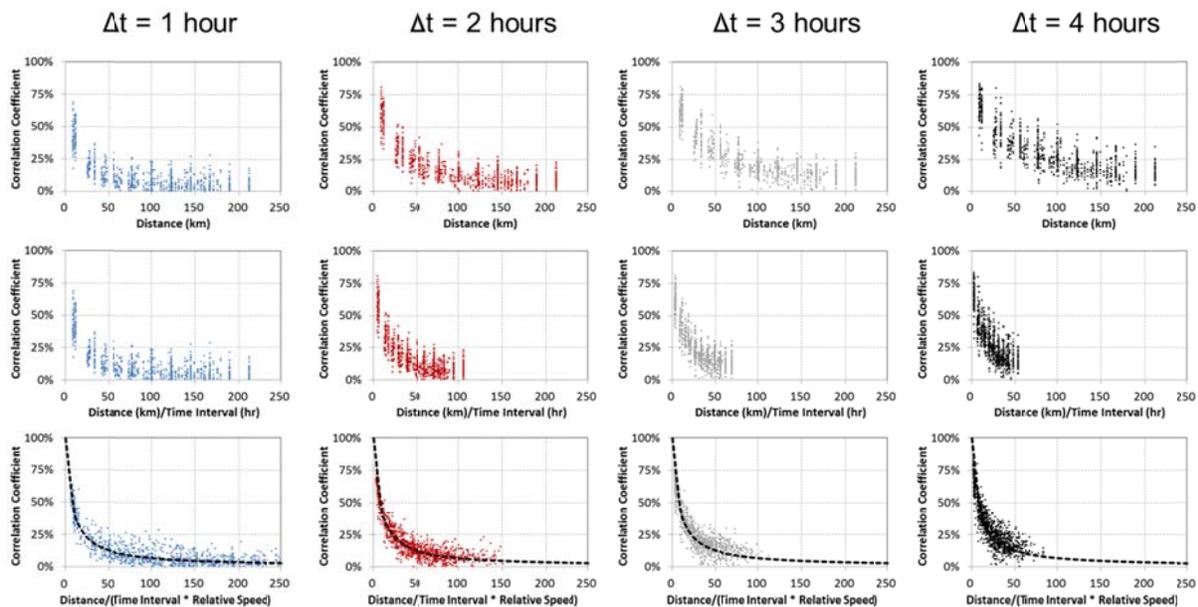
Note: Distance / (Time Interval * Relative Speed) is related to Dispersion Factor

Figure 3. Correlation coefficients presented by time interval for Southwest.
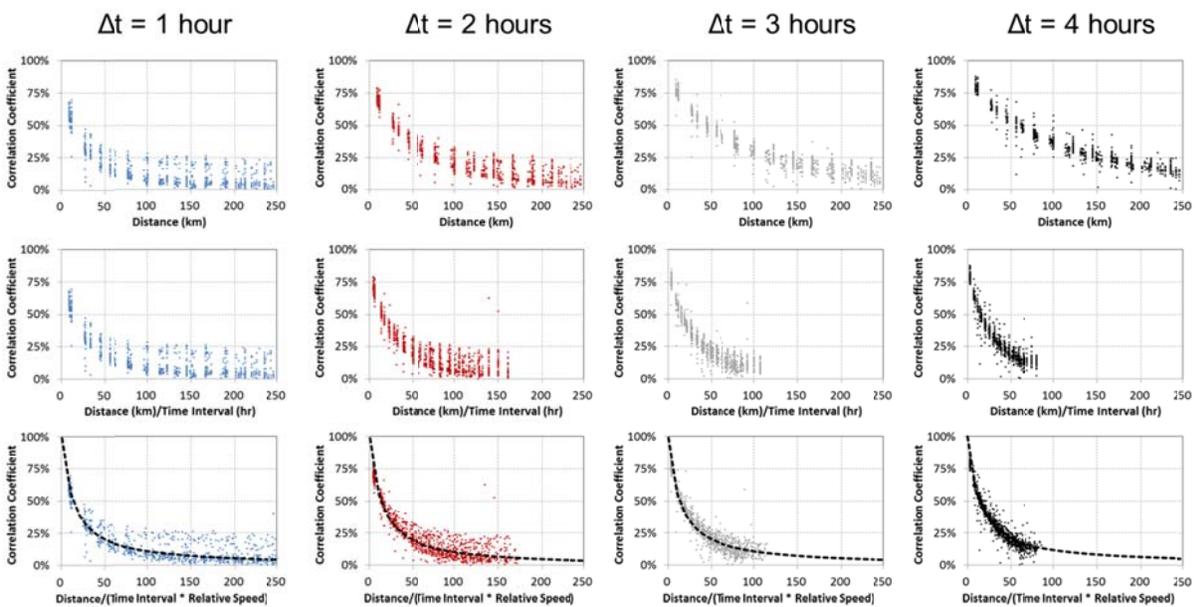
Note: Distance / (Time Interval * Relative Speed) is related to Dispersion Factor

Figure 4. Correlation coefficients presented by time interval for Great Plains.
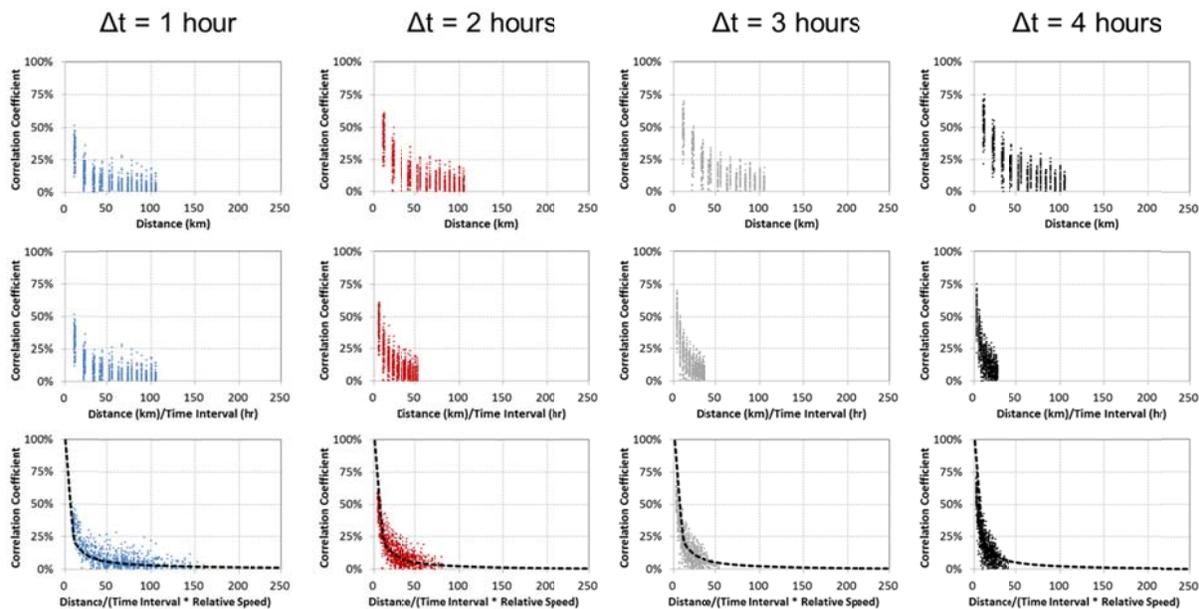
Note: Distance / (Time Interval * Relative Speed) is related to Dispersion Factor

Figure 5. Correlation coefficients presented by time interval for Hawaii.
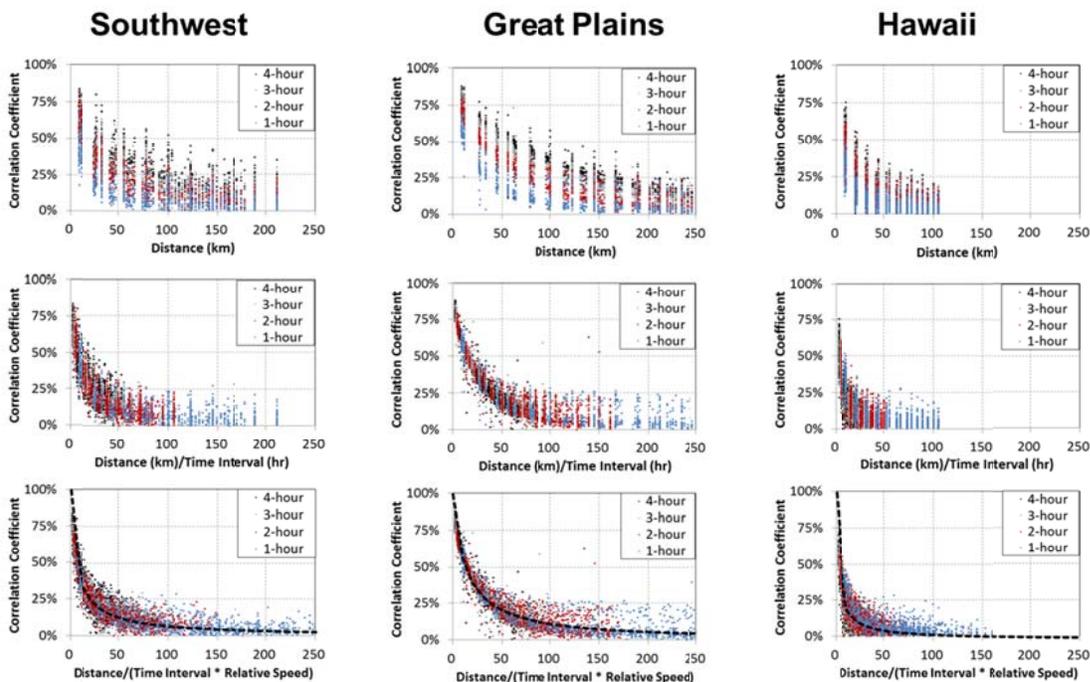
Note: Distance / (Time Interval * Relative Speed) is related to Dispersion Factor

Figure 6. Correlation coefficients for all locations and time intervals.
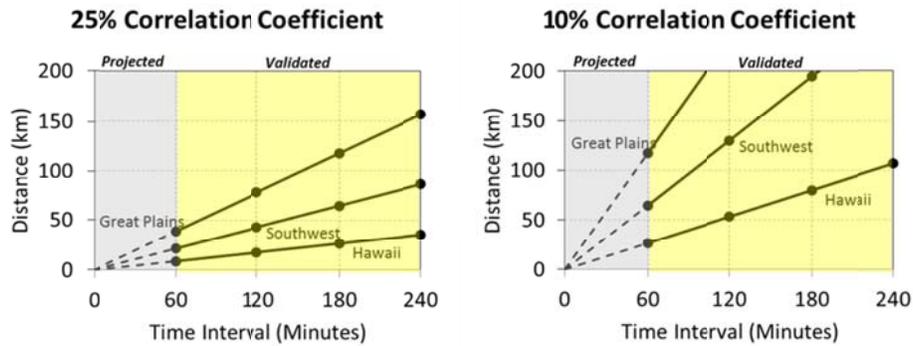
Figure 7. Results scale linearly with the time interval for a fixed correlation coefficient.
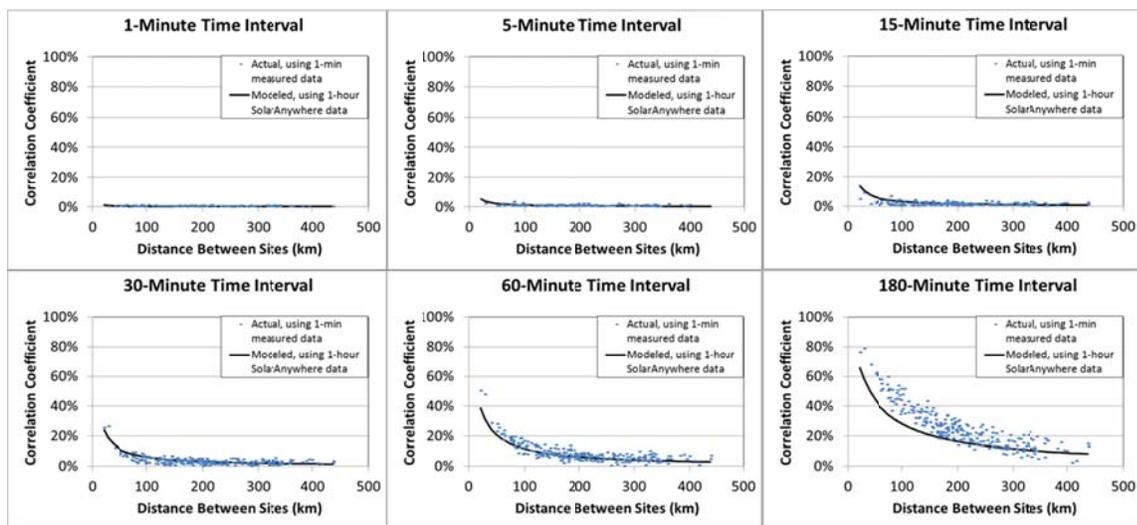
Figure 8. Comparison of results to geographic diversity study.
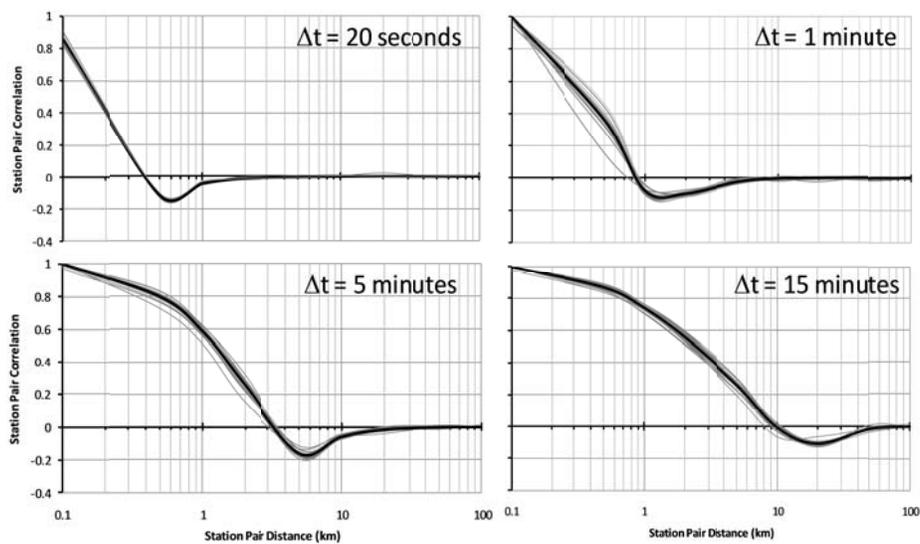
Figure 9. Key results from virtual network study.
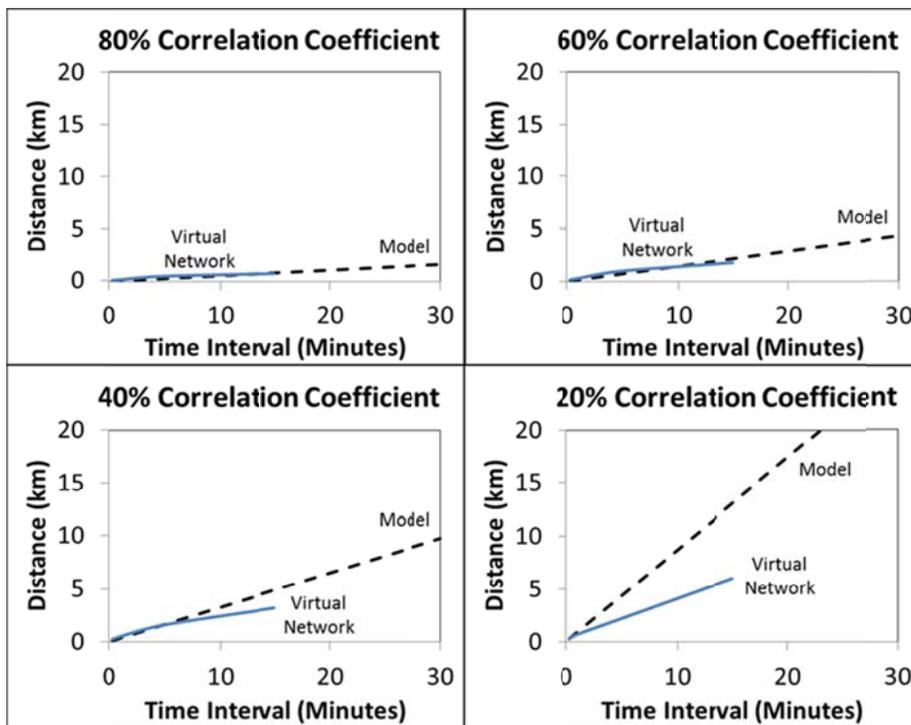
Figure 10. Comparison of results to virtual network study.

10 sec. data from 400m x 400m grid at Cordelia Junction, CA (Nov. 7, 2010)
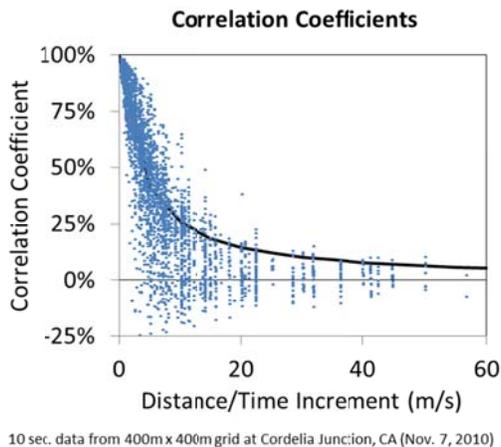
Figure 11. Correlation coefficients for high-density, 25 unit network at Cordelia Junction, CA on November 7, 2010 for time intervals from 10 seconds to 5 minutes.