BNL-108386-2015-JA

V14

Manuscript Submission for Solar Energy Journal

For peer-review purposes

On the Spatial Decorrelation of Stochastic Solar Resource Variability at Long Timescales

Marc J.R. Perez[a]*,Vasilis M. Fthenakis[a,b]

[a]*Columbia University, New York, 10025, USA*

[b]*Brookhaven National Laboratory, Upton, New York, 11973, USA*

**mjp2167@columbia.edu*

Abstract

Understanding the spatial and temporal characteristics of solar resource variability is important because it helps inform the discussion surrounding the merits of geographic dispersion and subsequent electrical interconnection of photovoltaics as part of a portfolio of future solutions for coping with this variability. Unpredictable resource variability arising from the stochastic nature of meteorological phenomena (from the passage of clouds to the movement of weather systems) is of most concern for achieving high PV penetration because unlike the passage of seasons or the shift from day to night, the uncertainty makes planning a challenge. A suitable proxy for unpredictable solar resource variability at any given location is the series of variations in the clearness index from one time period to the next because the clearness index is largely independent of the predictable influence of solar geometry. At timescales shorter than one day, the correlation between these variations in clearness index at pairs of distinct geographic locations decreases with spatial extent and with timescale. As the aggregate variability across *N* decorrelated locations decreases as

$1/\sqrt{N}$, identifying the distance required to achieve this decorrelation is critical to quantifying the expected reduction in variability from geographic dispersion.

Using 10 years of satellite-derived daily-interval solar resource data across the world, we demonstrate that the spatiotemporal behavior of unpredictable solar resource variability is mirrored at longer timescales. We do so by examining over 1.4 million unique pairs of sites across the Eastern hemisphere and quantifying the influence each pair's geographic separation and bearing has on the correlation between the variability of each pair's clearness indices at timescales of one, two, four, seven, fifteen and thirty days. Expected pair-decorrelation distances are estimated by fitting exponential trends to the data using nonlinear least-squares regression and are presented as a function of timescale and pair orientation.

Reflecting the predominant direction in which meteorological phenomena propagate at each of these timescales, we find that pairs of sites require considerably shorter distances to decorrelate when they are oriented north to south versus when they are oriented east to west. As at shorter timescales, these decorrelation distances are shown to increase with both timescale and with geographic extent.

Keywords: correlation, variability, solar energy, photovoltaics

---

1. Introduction

With the rapid recent growth of the solar photovoltaic (PV) industry, it is of utmost importance to address the principal barrier to achieving its high-penetration across global electrical grids: the inherent variability of the solar resource. Variability of the solar resource is a result of largely unpredictable or stochastic meteorological phenomena and from the predictable rotation of the earth around the sun and about its own axis. To achieve very high PV penetration, the imbalance between the variable supply of sunlight and demand must be alleviated.

Overcoming the effects of unpredictable resource variability poses the greatest challenge as the uncertainty in PV output it drives makes it a challenge to plan for. This unpredictable variability occurs on every temporal scale: from second-to-second variations driven by individual passing clouds to year-to-year variations driven by periodic large-scale phenomena such as El-niño and variations in volcanic activity. If high penetration PV is to be achieved, a balanced portfolio of solutions for coping with these imbalances are required on both the supply and demand side.

The widespread geographically distributed deployment and electrical interconnection (geographic dispersion) of PV is one of three primary supply-side solutions to solar variability—which do not rely on support from other generation sources—that can minimize the cost of electricity generated therefrom. The other two primary supply-side solutions are storage (where excess solar generation is stored when it exceeds demand and is released when it does not meet demand), and smart curtailment (where solar capacity is oversized and excess generation is curtailed at key times to minimize the need for storage.)

While conventional electricity market structures and regulatory frameworks remain suboptimal for the optimized development of these aforementioned solutions, an understanding of the nature of solar resource variability is critical for an informed discussion about the relative merits of each. In these conventional electricity markets, variations in PV output introduced by fluctuations in the solar resource on the minute-to-minute and second-to-second timescales drive increased unit governor response and load frequency control requirements while influencing economic dispatch. Solar resource variability on the timescale of minutes to hours impacts load following requirements, while day-to-day variability and longer variations in the solar resource influences day-ahead requirements and long-term regional infrastructure planning, especially at higher penetrations. [1]

The object of this research is to understand the way in which unpredictable solar resource variability is affected by geographic distance, bearing and timescale for timescales greater than one day. In so doing, we

help frame and inform the discussion surrounding the relative effectiveness of geographic dispersion as a solution for achieving high penetration PV.

The series of variations in the clearness index from one time period to the next constitute a suitable proxy for unpredictable variations in the solar resource at any given location because the clearness index is largely independent of the predictable influence of latitude. At timescales shorter than one day, the correlation between these variations in clearness index at pairs of distinct geographic locations decreases with spatial extent and with timescale. [2-4] As the aggregate variability across N decorrelated locations decreases as $1/\sqrt{N}$ according to the central limit theorem, identifying the distance required to achieve this decorrelation is critical to quantifying the expected reduction in variability from geographic dispersion.

Herein, we examine 1.4 million unique pairs of sites across the Western hemisphere and quantify the influence of each pair's geographic separation and bearing on the correlation between the variability of each pair's clearness indices at timescales of one, two, four, seven, fifteen and thirty days. We then estimate the pair-decorrelation distances for each combination of timescale and pair orientation (pairs separated north-to-south or east-to-west) by fitting exponential trends to the data using locally weighted polynomial regression.

Much research has been performed over the past several decades surrounding the nature of solar resource variability, including its spatial and temporal characteristics [5-14], the effect of distance on its pair correlation [2-4], its spatially anisotropic nature and relation to cloud-speed [15-18], the nature and implications of spatial smoothing [1, 19-23] and the types of solutions it engenders [24, 25]. For the most part, these studies have focused on the nature and impacts of variability at short timescales (intra-day) and of those that have examined variability on timescales greater than day-to-day, none have explored the effect of timescale, distance *and* orientation on purely stochastic solar resource variability, hence the novelty of this present study.

The results we report further our understanding of the spatiotemporal nature of solar resource variability at longer timescales. Thereby, we inform discussion surrounding the relative merits of continental-scale electrical transmission grids such as that proposed by the Desertec foundation [26] and similar initiatives

as a viable solution to solar resource intermittency. We also highlight the fact that variability exhibits the same characteristics with respect to distance at timescales greater than one day, matching the effects observed at shorter timescales which are reported in [1-6].

2. Methodology

The object of this research is to quantify the effects of spatial extent, spatial orientation and timescale on stochastic solar resource variability at timescales greater than one day. We define stochastic variability of the solar resource as the series of variations (step-changes from one time interval to the next) which do not result from some deterministic process. At the time intervals of greater than one day that we investigate herein, the deterministic process whose influence we attempt to remove is the seasonal trend resulting from the Earth's rotation around the sun. We therefore use variations in the clearness index (Kt) as a proxy for variability, as the majority of latitudinal influence (which is responsible for the amplitude of this seasonal trend) is eliminated. This is because Kt is the ratio of surface radiation to radiation at the top of the atmosphere, which co-vary on a seasonal basis.

One way to test the way in which spatial extent and orientation influence solar resource variability is to calculate the correlation between the Kt variations at pairs of unique geographic locations and determine the way in which their correlation changes as a function of the geographic distance and bearing separating these locations. We use this approach to compute ~83 million correlations across ~1.4 million unique pairs of spatially discrete locations spanning the Western hemisphere and investigate trends between each pairs' geodesic separation, Cartesian bearing and the corresponding correlation coefficients.

By quantifying the decorrelation distances at which the bulk of coordinate pairs exhibit zero correlation between their respective Kt variations, we define a spatial extent across which the aggregate unpredictable variability of electrically interconnected solar PV will begin to decrease significantly. More precisely, if we have *N* locations within a region where each corresponding series of Kt variations (ΔKt_i) is independent (Eq.

1) and identically distributed (meaning they each share the same standard deviation, $\sigma_{\Delta Kt}$, Eq. 2), the standard deviation of the aggregate Kt variations across these *N* locations ($\sigma_{\overline{\Delta Kt}}$, Eqs. 3,4) is 1/√N times the standard deviation of Kt variations at a single one of these sites.

$if: \rho\left(\Delta Kt_i, \Delta Kt_j\right) = 0 \quad \forall i, j$	If all Kt variations in region are independent, the correlation between any two pairs of locations within the region is equal to zero.	Eq. 1
$if: \sigma_{\Delta Kt,i} = \sigma_{\Delta Kt,j} \equiv \sigma_{\Delta Kt} \quad \forall i, j$	If all Kt variations in region are identically distributed, the series of Kt variations at each location share the same standard deviation.	Eq. 2
$\overline{\Delta Kt} = \frac{1}{N}\sum_i \Delta Kt_i$	The mean series of Kt variations across a region of *N* locations.	Eq. 3
$\sigma_{\overline{\Delta Kt}} = \frac{1}{\sqrt{N}}\sigma_{\Delta Kt}$	The aggregate variability (standard deviation of the variations in clearness index) across *N* locations as a function of variability at single site if the conditions in Eqs. 1 and 2 are met.	Eq. 4

Therefore, by understanding the expected distances and spatial orientations required to achieve decorrelation between the changes in clearness index between pairs of locations, we can begin to estimate the size and shape of the region across which solar PV must be spread in order to reduce this unpredictable variability.. As unpredictable solar resource variability occurs across many different timescales, from day-to-day variations to month-to-month variations we quantify the expected decorrelation distances for Kt variations at 6 distinct time averaging intervals (LIST). By observing the individual decorrelation distances for each time-averaging intervals, we identify the interconnection distances required to begin reducing unpredictable variability resulting from regional-scale meteorological phenomena and seasonal climatic patterns.

2.1. Description and provenance of solar resource data

Daily-averages of horizontal-incidence irradiance at the top of the atmosphere (TOA)[*] and global horizontal-incidence irradiance at the earth's surface (Gh)[†] corresponding to every latitudinal and longitudinal

---

[*] Top-of-Atmosphere irradiance (TOA) refers to the amount of solar radiation received by a fixed plate tangent to the earth's surface at an altitude above the bulk of the atmosphere (approximately 90km or the top of the bottom of the thermosphere and top of the mesosphere)
[†]Global horizontal irradiance (Gh) refers to the amount of solar radiation received by a fixed plate tangent to the earth's surface and is the sum of two components: beam (collimated radiation that hits this surface directly from the sun) and diffuse (radiation hitting this surface which has been scattered across the sky hemisphere).

quadrangle on the planet[‡] at a 1°x1° spatial resolution from July 1983 through the end of 2007 were obtained from NASA-Langley's Surface Meteorology and Solar Energy group (SSE)[27]. As we seek to examine the correlations between pairs of coordinates, and there exist over 100 million unique coordinate pairs within this dataset of 14,400 coordinates, it is computationally necessary to extract a representative subsample, which is highlighted in figure 1.

Our representative subsample spans the North and South American continents and approximately 22% of the globe over the 10-year time horizon of 1997 – 2007. Geographically, this subsample is bounded by the 140th meridian West (140°E), the 20th meridian East (20°E), the 60th parallel north (60°N) and the 60th parallel south (60°S). Within this bounded region, we sample ~10% of the available coordinates (~1440), in a geometric grid spaced in 3° latitude/longitude increments. This region is displayed in figure 1 where we show the linearly interpolated mean Kt across the 10-year time-horizon and the sampled locations in light grey. We chose this region due to its large geographic and climatological range, which reflects a diversity of clearness index variation profiles. As the directionality and strength of mean wind patterns driving weather systems and therefore the fluctuations in Kt differ regionally, investigating their behavior across a large geographic and climatological range as we do in this study lets us identify common trends. We derive the series of clearness indices at each of the sampled locations throughout this region via the relation in Eq. 5 below:

$$Kt = \frac{Gh}{TOA}$$

Clearness index (Kt) is the ratio of surface global horizontal radiation (Gh) to top-of-atmosphere radiation (TOA) and is an indication of the degree of atmospheric attenuation as radiation travels towards the surface. Eq. 5

2.2. Time-averaging and variations of the clearness index

Fluctuations in the clearness index are apparent at every timescale, and are driven by meteorological phenomena of varying spatial scales: from passing clouds to frontal systems to the seasonally-shifting cloud band known as the inter-tropical convergence zone (ITCZ). As the timescale of induced fluctuations is linked

[‡] Reflecting a 1°x 1° spatial resolution.

to the spatial scale of meteorological phenomena, we want to study the effect of timescale on decorrelation distance. To do so, we compute time-averages ($\overline{Kt}^{\Delta t}$) of our regionally-sampled clearness index data across 6 different timescales (Δt): 1 day, 2 days, 4 days, 7 days, 15 days and 30 days as is demonstrated in Eqs. 6-7.

$$\overline{Kt}^{\Delta t}_{j} = \frac{1}{\Delta t}\sum_{i=j}^{j+\Delta t-1} Kt_i = \frac{1}{\Delta t}\left\{Kt_j + Kt_{j+1} + \cdots Kt_{j+\Delta t-1}\right\}$$

The average Kt across a period, *j,* of Δt days in length as a function of the mean Kt on each individual day, *i*, within that period. Eq. 6

$$where: \ell\left(\overline{Kt}^{\Delta t}\right) = \left\lfloor \Delta t^{-1} \cdot \ell\left(Kt\right)\right\rfloor$$

The length (in number of periods) of the time-averaged series as a function of the time-averaging interval and the length of the original daily-averaged series. Eq. 7

Equations 6 and 7 demonstrate the simplest case where one time-averaged series is generated per time-averaging interval with the averaging at index i = 1. Performing the analyses with a single series representing each time-averaging interval risks introducing sampling error because there exist as many indices, β, at which to start the averaging as the length of the time averaging interval (Δt). To account for this, we compile a set, $\overline{Kt}^{\Delta t}$ (Eq. 9), containing each unique time-averaged series, $\overline{Kt}^{\Delta t}_{\beta}$ (Eq. 8), that is possible when cycling through the potential starting indices for the corresponding time-averaging interval. For instance, if Δt = 30 days, the starting index β can range from 0 to 29 and the set, $\overline{Kt}^{30}$, therefore contains 30 unique individual series.

$$\overline{Kt}^{\Delta t}_{\beta,j} = \frac{1}{\Delta t}\sum_{i=j}^{j+\Delta t-1} Kt_{i+\beta} = \frac{1}{\Delta t}\left\{Kt_{j+\beta} + Kt_{j+1+\beta} + \cdots Kt_{j+\Delta t-1+\beta}\right\}$$

The average Kt across a period, *j*, of Δt days in length when the averaging period is shifted by *β* days. Eq. 8

$$\overline{Kt}^{\Delta t} = \left\{\overline{Kt}^{\Delta t}_{\beta}\right\}, where: \beta \in \left\{0,1\cdots(\Delta t-1)\right\}$$

The set of time-averaged Kt series reflecting every possible interval shift, β, for a given time-averaging interval, Δt. Eq. 9

Through this process, we generate 59 distinct series of Kt for each unique geographic location sampled from our region: 30 series of 30-day time averages (Δ t = 30), 15 series of 15-day time averages (Δ t = 15), 7 series of 7-day time averages (Δ t = 7), 4 series of 4-day time averages (Δ t = 4), 2 series of 2-day time averages (Δ t = 2) and 1 series of 1-day time averages (Δ t = 1).

Each of these distinct time-averaged series of Kt are *differenced* (Eqs. 10-11) in order to generate the series of step-changes or variations (ΔKt) between sequential periods of length Δt. These sets of clearness index variations reflect the stochastic variations in the solar resource for each corresponding timescale. For

instance, the series of variations in mean clearness index at a given location between sequential periods of 30 days reflects the variations in atmospheric attenuation or mean cloud cover between these same periods. These variations in cloud cover drive corresponding variations in the solar resource at the 30-day timescale: variations that are unpredictable in nature (unlike those variations driven by the solar system's geometry.)

$$\Delta \overline{Kt}_j^{\Delta t} = \overline{Kt}_j^{\Delta t} - \overline{Kt}_{j-1}^{\Delta t}$$

The variation in Kt between two subsequent periods of length Δt. Eq. 10

$$\ell\left(\Delta \overline{Kt}^{\Delta t}\right) = \ell\left(\overline{Kt}^{\Delta t}\right) - 1$$

The length (in number of periods) of the series of Kt variations expressed as a function of the length of the series of time-averaged Kt for a given time-averaging interval, Δt. Eq. 11

2.3. Calculation of correlation coefficients

$$\rho_{A,B} = \frac{\operatorname{cov}\left(\Delta Kt_A, \Delta Kt_B\right)}{\sigma_{\Delta Kt_A}\sigma_{\Delta Kt_B}} = \frac{\sum_i \left(\Delta Kt_{A,i} - \overline{\Delta Kt_A}\right)\left(\Delta Kt_{B,i} - \overline{\Delta Kt_B}\right)}{\sqrt{\sum_i \left(\Delta Kt_{A,i} - \overline{\Delta Kt_A}\right)^2}\sqrt{\sum_i \left(\Delta Kt_{B,i} - \overline{\Delta Kt_B}\right)^2}}$$

The Pearson product-moment correlation coefficient between the variations in Kt at location A and location B. Eq. 12

Slightly over 1.4 million unique pair combinations are possible within our geographic subsample of 1,440 distinct coordinates arrayed across the Americas. As 59 distinct series of Kt variations, $\Delta\overline{Kt}^{\Delta t}$, across 6 time-averaging intervals (Δt) and corresponding interval shifts (β) are generated at each geographic coordinate, slightly over 83 million unique correlations are performed.

For each pair of locations (A,B), we calculate the sample correlation coefficient, $\rho_{A,B}$, between each of the 59 representative series of clearness index variations corresponding to each location (ΔkT_A, ΔkT_B) as demonstrated in Eq. 12.

2.4. Bearings and distance calculations

For each pair of sites (A,B) for which we calculate a correlation coefficient, we also calculate the geodesic or shortest path distance ($d_{A\rightarrow B}$) separating them along the surface of the earth using Lambert's 1942 approximation [28], offering an accuracy of +- 12.5 meters across very long distances. In the algorithm

outlined in Eqs. 13-16 below, the equatorial radius of the ellipsoid representing the earth is represented by r_E, while f represents its flattening coefficient; both of which are part of the international cartography standard that comprises the latest revision of the World Geodetic System (WGS-84).[29] To simplify the geodesic calculation, reduced or parametric latitudes, β —projections of the geodetic latitudes, φ, onto the surface of the surrounding sphere—are calculated for each unique coordinate via the relation in Eq. 13. This permits the spherical law of cosines to be applied in Eq. 14 in order to determine the central angle between each pair of coordinates, $\theta_{A<B}$, as a function of their reduced latitudes, $\beta_{A,B}$, and the difference between their geodetic longitudes, $\Delta\lambda$. Lambert's approximation to correct the central angle calculated in spherical space to account for the earth's ellipticity is outlined in Eq. 15. The final geodesic distance between the pair of coordinates, $d_{A\rightarrow B}$, shown in Eq. 16, is the arc length between them along the surface of the earth as modified by Lambert's correction.

$\beta = arctan\left[\left(1-f\right)tan\left(\varphi\right)\right]$	Reduced latitude, β, as a fn. of earth flatness, f, and geodetic latitude, ϕ.	Eq. 13
$\theta_{A\angle B} = arccos\left[sin\left(\beta_A\right)sin\left(\beta_B\right)+cos\left(\beta_A\right)cos\left(\beta_B\right)cos\left(\Delta\lambda\right)\right]$	Central angle, $\theta_{A<B}$, between locations A, B as fn. of their reduced latitudes, $\beta_{A,B}$ and the difference in their geodetic longitudes, $\Delta\lambda$	Eq. 14
$X = sec\left(\frac{1}{2}\theta_{A\angle B}\right)\left[\theta_{A\angle B} - sin\left(\theta_{A\angle B}\right)\right]sin^2\left(\bar{\beta}\right)cos^2\left(\frac{1}{2}\Delta\beta\right)$ $Y = csc\left(\frac{1}{2}\theta_{A\angle B}\right)\left[\theta_{A\angle B} + sin\left(\theta_{A\angle B}\right)\right]cos^2\left(\bar{\beta}\right)sin^2\left(\frac{1}{2}\Delta\beta\right)$	Intermediate parameters derived by Lambert to correct the central angle for the Earth's ellipticity as a function of the central angle, $\theta_{A<B}$, mean latitude, $\bar{\beta}$, and the difference in latitude between locations A and B, $\Delta\beta$.	Eq. 15
$d_{A\rightarrow B} = r_E\left\|\theta_{A\angle B} - \frac{f}{2}\left(X+Y\right)\right\|$	Geodesic distance between locations A & B, $d_{A\rightarrow B}$, as a function of earth's equatorial radius, r_E, flatness, f, central angle, $\theta_{A<B}$, and intermediate parameters, X & Y	Eq. 16

Along with pair distances, we also calculate in Eq. 17 the Cartesian bearing, ω, of the line separating each pair of coordinates in order to investigate the way in which decorrelation distances change with pair orientation. Correlation coefficients for a given timescale are sorted into three piles, correlations which represent underlying pair orientations which are generally North-South ($-10° \leq \omega_{A,B} \leq 10°$), those which represent pair orientations that are generally East-West ($90° \leq \omega_{A,B} \leq 100°$) and those which represent all possible pair orientations ($\forall\omega_{A,B}$).

$$\omega_{A,B} = arctan\left|\frac{\Delta\lambda}{\Delta\varphi}\right|$$

The Cartesian bearing between locations A and B as a function of the difference between their longitudes, $\Delta\lambda$, and the difference between their latitudes, $\Delta\phi$. Eq. 17

2.5. Decorrelation distances

In order to infer the geographic extent at which Kt variations at pairs of sites are generally decorrelated at a given timescale, we fit a trend between the correlation coefficients ($\rho_{A,B}$) and their corresponding geographic separations ($d_{A\rightarrow B}$). A priori, we compute a moving average of correlation coefficients with respect to their corresponding distances in Eq. 18 in order to filter and more easily observe the central tendency across the millions of correlation coefficients.

$$\rho_d = \tfrac{1}{30}\sum_{d-30/2}^{d+30/2}\rho_i \quad for: d \in \tfrac{1}{2}\left\{30,60,90\cdots D\right\}$$

The moving average in 30 km increments of correlation coefficients, ρ_d, with respect to their corresponding separation distances, *d,* in km. D is the maximum observed pair separation distance. Eq. 18

This clarification allowed us to see that for nearly timescale, correlation coefficients between pairs of locations appeared to decrease exponentially with respect to their corresponding geodesic separation distances. An exponential decay function of the form outlined in Eq. 19 was thus fit to the data using nonlinear least-squares regression (nls).

$$\rho(d) \sim e^{-\beta d} + \gamma$$

The general exponential-decay relationship between the Kt correlation coefficients, ρ, and their corresponding separation distances, d where the parameters β and γ are determined through non-linear least squares regression. Eq. 19

When examining the influence of bearing on decorrelation distances, we found post-hoc that correlation coefficients which represent pairs of locations separated North to South ($-10° \leq \omega_{A,B} \leq 10°$) appear to decrease not as a pure exponential but as an exponentially-modulated sinusoid. This is due to the fact that at shorter timescales (<15 days), Kt variations at pairs of sites on who are on opposite sides of the equator from each other exhibit strong anti-correlation due to the Coriolis effect driving weather systems in opposite directions at each location. This strong anti-correlation occurs at a very precise distance of separation and at longer distances, the anti-correlation disappears thereby introducing a sinusoid into the trend. Therefore, a function

of the form outlined in Eq. 20 is fit to the trend relating N/S pair correlations to their corresponding separation distances, also using nls.

$$\rho(d) \sim cos(\alpha d)e^{-\beta d} + \gamma$$

The sinusoid-modulated exponential-decay relationship between the Kt correlation coefficients, ρ, and their corresponding separation distances, d when examining pairs of sites with North-South separations. The parameters α, β and γ are determined through non-linear least squares regression. Eq. 20

We define the decorrelation distance for each time-interval under investigation as the pair separation distance when the exponential trend (or cosine-modulated exponential trend, in the case of N/S pairs) first crosses a threshold of $\rho = 0.1$. A threshold value is required, as via the a priori rolling mean analysis, none of the exponential trends truly reach zero, although they do approach it.

3. Results

3.1. Distance, timescale and their effect on spatial smoothing

After performing ~83 million unique pair correlations across the Americas covering a comprehensive range of climatic zones, we are able to paint a clearer picture of the way in which stochastic solar resource variability at timescales greater than one day is mediated by both spatial extent and spatial orientation.

We demonstrate the results of sorting all of the correlations performed by timescale and distance in the scatterplots comprising figure 2 and summarized in Table 1. Each of the 6 subplots represents a particular time-averaging interval and each point represents the correlation between the clearness index variations at that timescale at two distinct locations versus their geographic separation. Fitted exponential trends establishing the relationship between correlation coefficients and separation distance are highlighted in light blue on each subplot. The equations of these relationships are printed the upper left-hand corner of each plot while the pertinent parameters underlying each are listed in table 1. Decorrelation distances for each timescale are demarcated by a dotted red line and in red text on each plot.

Several important points can be made from observing these results:

- Pair correlation decreases exponentially with increasing distance at every timescale.

- Pair correlation decreases more slowly with respect to pair separation as timescale is increased.
- Decorrelation distances increase with the length of the time-averaging interval.

Also of note is the fact that the number of points (representing distinct correlations performed) increases with each timescale. This is due to the necessity of phase-shifting the starting point when taking time-averages as discussed in more detail in section 2.2.

Expected decorrelation distances between two locations vary from 1123 km for day-to-day variations in the clearness index to 3117 km for month-to-month variations. The increase in expected decorrelation distance with increased timescale along with the exponential decrease in expected pair correlation with increased pair separation mirrors the patterns identified at shorter timescales. This similarity highlights the temporally and spatially fractal nature of meteorological phenomena. [4, 30]

3.2. The effect of geographic orientation on spatial smoothing

We now make use of the pair orientations calculated for each pair of coordinates and group the correlation coefficients into two distinct orientation categories: pairs that are oriented east-to-west and pairs that are oriented north-to-south (as described in section 2.4). The six subplots comprising figure 3 represent east-west orientations and show pair correlations as a function of their separation distance for the six individual time-averaging intervals under investigation. The six subplots comprising figure 4 show the same for north-south orientations. In each series of plots, vertical dotted red lines identify the expected decorrelation distances while solid lines demonstrate the fitted exponential or sinusoid-modified exponential trends between correlation coefficient and separation distance.

Expected decorrelation distances for east-west pairs range from 1224 km for day-to-day variations in the clearness index to over 10,000 km for month-to-month variations. These are much longer than the identified decorrelation distances than for north-south pairs at the same timescales; 900 km for day-to-day variations and 1643 km for month-to-month variations, respectively.

This dependence of the stochastic component of solar resource variability on spatial orientation as well as spatial extent is better observed by the series of polar plots comprising figure 5 where each of the six subplots represents a particular variability timescale. In each of these plots, each of the correlations performed for a given timescale are plotted according to the underlying pair separation (distance from the center) and pair orientation (rotation about the center). The degree of correlation between the Kt variations for each pair is linearly interpolated from the millions of individual values to display a smooth surface and indicated by the color scale to the left side of the plot. A dotted red line indicates the decorrelation isoline (where $\rho_{pair} = 0.1$). With these polar plots, one can clearly see how quickly pair correlation decreases with spatial extent, how the decorrelation distance increases with timescale, and how east/west oriented pairs are much more likely to be correlated than north-south pairs.

The reason why variations in the clearness index between two locations are much more strongly correlated if these locations are separated east to west than if they were separated by the same distance north to south is because the meteorological phenomena that drive these variations propagate predominantly from east to west.

A summary if important insights related to these plots:

- Pair correlation drops more rapidly with increasing pair separation if the pair orientation is North/South versus if the pair orientation is East/West.
- Expected decorrelation distances are correspondingly longer for east/west orientations than for north/south orientations.

4. Discussion

The results highlighted above demonstrate that even at longer timescales representing daily, weekly and monthly stochastic variations in the solar resource, the trends relating spatial extent to aggregate variability mirror the trends observed at shorter timescales. These trends include an exponential decrease in expected ΔKt pair correlation with increasing pair separation distance, and a corresponding longer expected

decorrelation distance with an increased variability timescale. In practical terms this means that spatial dispersion of solar generating facilities is an effective strategy for reducing the magnitude of unpredictable variations in their aggregate output even if these variations are of a very long temporal period. What these results also demonstrate is that the larger the spatial extent across which PV is spread and interconnected, the wider the temporal breadth of variations which can be reduced. For example: if *N* PV generation sites are spread across a region where the average separation between sites is 1100 km, the magnitude of aggregate variations in output from *day to day* across the region are reduced by $1/\sqrt{N}$. To achieve a $1/\sqrt{N}$ reduction in the magnitude of *month-to-month* variations in aggregate output, the average separation between individual sites would need to be ~3100km: a much larger region.

As ΔKt pair correlation is shown to decrease *exponentially* with pair separation, once decorrelation is achieved ($\rho = 0.1$), we can expect little benefit from further increasing this distance. If the aim is to reduce unpredictable day-to-day variations in aggregate solar output through geographic dispersion, limited benefit can be expected from increasing the average site separation across a given region past the decorrelation distance of 1100 km. Deterministic resource variability, particularly of the diurnal and seasonal kinds, can be further reduced through longer interconnections than those necessary to cope with the variability of the stochastic kind studied in this paper. It is important to note that the wider the geographic region across which these strategies are employed, the higher the required costs of interconnection become. A balance between the costs linked to this increased interconnection + geographic dispersion, energy storage, supply curtailment and demand-side management is what will be required to truly minimize integration costs of PV at high penetrations. [31,32]

What is also interesting is that the spread of points around the calculated trends showing expected pair correlation vs. pair separation (figures 2-4) increases with timescale. This is partially due to the fact that more correlations were performed for these longer timescales due to the increased number of ways to perform the time-averages (Eqs. 9-10). For North/South Pairs, (figure 4) this spread about the expected trend, particular at

longer time-averaging intervals, is likely reflective of the relative latitudes of each site. Future work could investigate the effect of mean pair latitude (in much the same way as we categorized pairs by Cartesian orientation) and in so doing, shed light on any potential effects.

Visible upon examination of figure 5 and when comparing figures 3 and 4, this directional dependence in the relationship between variability and spatial extent can be best explained by an example. Imagine a pair of sites separated by 1500 km and both lying at 15° N latitude. These sites will be affected simultaneously by the weather systems blown along by the Northeasterly trade winds (which blow primarily east to west) affecting the pattern of variations in their respective clearness indices in a similar fashion. By contrast, a pair of sites separated North to South by about 1500 km (where one site lies at 15° N latitude and the other site lies at 28° N latitude, for example will likely see a negligible or even negative correlation between the pattern of variations in their respective clearness indices. This is because while the site at 15° N experiences fluctuations induced by the east-to-west Northeasterly trade winds, the site at 28° N experiences fluctuations induced by primarily west-to-east anti-trade winds (a.k.a. westerlies). A pair of sites separated North to South by the same distance as a pair of sites separated East to West is more likely, therefore, to experience uncorrelated changes in cloudcover.

5. Conclusions

This study sheds light on the nature of unpredictable solar resource variability at timescales longer than one day over a geographical region covering the Americas and spanning a third of the Earth's non-polar zones. What these analyses show is that reducing unpredictable day-to-day or longer variations in the solar resource is possible through geographic dispersion, that the degree of geographic dispersion required to reduce aggregate variability by a targeted amount across a region is dependent on that region's orientation and these variations' time period.

In particular, we show that such variability decreases quasi-exponentially with increasing spatial area at every timescale investigated. We do this by showing that expected correlation between the variations in clearness indices (ΔKt) at pairs of sites decreases as the separation between them increases. This decrease in expected correlation with increased pair separation was shown to be less pronounced as the timescale was increased, leading the decorrelation distance (where pairs of sites are expected to show little to no correlation between the variations in their clearness indices) to correspondingly lengthen. Expected pair correlation for a given pair separation was also shown to be significantly lower if the geographic separation was oriented North to South as opposed to East to West. Correspondingly, expected decorrelation distances of North-South pairs were shown to be much shorter than the expected decorrelation distances of East-West pairs for all timescales longer than 1 day.

We can infer from these results that the propagation velocity of stochastic meteorological phenomena is more rapid in east-west directions than in north-south directions and that this propagation speed is inversely proportional to timescale, a relationship empirically demonstrated for timescales ranging from seconds to hours and now shown to be valid for timescales ranging from days to months.

These results better frame the nature of solar variability, and they highlight the way in which it behaves relative to distance and the way in which it depends on geographic orientation and timescale. While it is beyond the scope of this paper to estimate the interconnection costs necessary to take advantage of such effects, what it does provide is further understanding of the nature of unpredictable solar resource variability at longer timescales and a template for the direction future R&D into the spatial aspect of solar PV integration should take. For example, the geographic dispersion of 100 solar generating facilities of equal size along a generally north-south bearing and an average site separation of 900 km and electrical interconnection thereof can be expected to experience $1/10^{th}$ the unpredictable variations in aggregate output than if this distributed capacity were concentrated at a single location within this region.. By reducing this unpredictable intermittency component from the collective output of PV generation, we significantly reduce the amount of

unit governor response, load frequency control, economic dispatch, and load following otherwise necessary to deal with it. If the costs of building out distributed transmission capacity can be minimized, leveraging this geographic smoothing effect is a potential means by which to reduce the costs of economically achieving high penetration PV targets.

Acknowledgements

This work is based upon work supported by the National Science Foundation Graduate Research Fellowship Grant No. DGE 1144155. Resources from Columbia University's Center for Life Cycle Analysis and from the Columbia University Fu Foundation of Engineering and Applied Science Eggleston Doctoral Research Fellows' program were instrumental in the success of this research. In addition, solar radiation data used herein were obtained from the NASA Langley Research Center Atmospheric Science Data Center.

References

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

[2] 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.

[3] Perez R., S. Kivalov, J. Schlemmer, K. Hemker and T. Hoff, (2012): **Short-term irradiance variability: preliminary estimation of station pair correlation as a function of distance**. Solar Energy 86, 8, pp. 2170-2176

[4] Hoff, T.E., Perez, R. (2013): **Chapter 6: Solar Resource Variability.** Solar Energy Forecasting and Resource Assessment. Elsevier. pp 133 – 148

[5] Perez, R., S. Kivalov, T. Hoff, (2011): **Spatial & temporal characteristics of solar radiation variability**. Proc. of International Solar Energy World Congress, Kassel, Germany

[6] Hoff, T. and R. Perez, (2010): **Quantifying PV Power Output Variability.** Solar Energy **84** (10) , pp. 1782-1793

[7] Gueymard, C., Wilcox, S., (2011). **Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data**. Solar Energy 85 (5), 1068–1084.

[8] Lave, M., Kleissl, J., Arias-Castro, E., (2011). **High-frequency fluctuations in clear-sky index.** Solar Energy 86 (8), 2190–2199.

[9] Perez, R., Hoff, T.E., 2011**. Solar Resource Variability: Myth and Fact**. Solar Today. August/ September 2011.

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

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

[12] Perez, R., Hoff, T., Kivalov, S., (2011). **Spatial and temporal characteristics of solar radiation variability.** Proc. of International Solar Energy (ISES) World Congress, Kassel, Germany.

[13] Vignola, F., (2001). **Variability of Solar Radiation over Short Time Intervals.** Proc. Solar 2001, American Solar Energy Society Conf., Washington, D.C.

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

[15] Hinkelman, L., George, R., Sengupta, M., (2011). **Differences between Along-Wind and Cross- Wind Solar Variability**. Proc. Solar 2011, American Solar Energy Society Conf., Raleigh, NC.

[16] Hoff, T.E., Norris, B., (2010). **Mobile High-Density Irradiance Sensor Network: Cordelia Junction** Personal Communication.

[17] Kleissl, J. (2014) **A Poisson Model for Anisotropic Solar Ramp Rate Correlations**. *Solar Energy*, **101C**, pp. 192-202

[18] Lave, M., Kleissl, J., 2013. **Cloud speed impact on solar variability scaling – Application to the wavelet variability model.** Solar Energy 91, 11–21.

[19] Perez, M.J.R. & Fthenakis, V.M. (2012): **Quantifying the Impacts of Long-Timescale Solar Resource Variability at High PV Penetrations.** Proc. IEEE PVSC 38 Photovoltaic Specialists Conference.

[20] Perez, M.J.R. & Fthenakis, V.M. (2012): **Quantifying Long-Timescale Solar Resource Variability.** Proc. WREF World Renewable Energy Forum.

[21] Murata, H. Yamaguchi, K. Otani (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) , pp. 9–19

[22] Hoff, T. E. (2011). U.S. **Patent Application: Computer-Implemented System and Method for Determining Point-to-Point Correlation of Sky Clearness for Photovoltaic Power Generation Fleet Output Estimation** (Application Number 13/190,435)

[23] 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.

[24] Perez, M.J.R. & Fthenakis, V.M. (2013) **Long-distance interconnection as solar resource intermittency solution: Optimizing the use of energy storage and the geographic dispersion + interconnection of solar generating facilities.** Proc. IEEE PVSC 29th Photovoltaic Specialists Conference.

[25] Perez, M.J.R. & Fthenakis, V.M. (2013): **Optimizing the Mix of Energy Storage and Long-Distance Interconnection as Solutions to Solar Resource Intermittency at High Penetrations of PV on the Grid.** Proc. ASES National Solar Conference.

[26] DESERTEC (2009). **Clean Power from Deserts: The DESERTEC Concept for Energy, Water and Climate Security.** WhiteBook, 4th Edition. DESERTEC Foundation.

[27] NASA. (2013) **Surface Meteorology and Solar Energy Data.** Langley Research Center Atmospheric Science Data Center. Accessed, Fall, 2012.

[28] Lambert, W.D. (1942). **The distance between two widely separated points on the surface of the earth**. *J. Washington Academy of Sciences* **32** (5): 125–130.

[29] NIMA (2004) . **Department of Defense World Geodetic System 1984: It's Definition and Relationships with Local Geodetic Systems.** National Imagery and Mapping Agency (NIMA). Technical Report TR8350.2

[30] Mandelbrot, B., (1982). **The fractal Geometry of Nature.** WH Freeman and Co, New York, 1982.

[31] Perez, M.J.R. (2014) **Geographic Dispersion and Curtailment of VLS-PV Electricity.** IEA PVPS Task 8 report, Ch.4 Future Technical Options for the Entire Energy System, Forthcoming.

[32] Perez, M.J.R. (2014). **A model for optimizing the combination of solar electricity generation, supply curtailment, transmission and storage.** (Order No. 3621033, Columbia University). ProQuest Dissertations and Theses, 246 pages.

Nomenclature
Kt : *Also known as the clearness index.* T*he ratio of Global Horizontal solar radiation (downward shortwave radiative flux) to Top-of-Atmosphere solar radiation (downward shortwave radiative flux)*
Cartesian Bearing: *The bearing in degrees based on an equirectangular projection of the earth. Important because true geodesic bearing can change significantly along a given straight-line path, especially over long distances.*
TOA: *TOA irradiance stands for downward shortwave radiative flux at the Top of the Atmosphere.*
Gh: *Gh irradiance stands for Global horizontal radiation or the downward shortwave radiative flux at the surface of the earth.*

List of Tables

- Table 1: Decorrelation distances and fitted trend parameters for pair correlations regardless of their orientation, for east-west and north-south oriented pairs, respectively. Parameters fit into Eq.'s 19 and 20, respectively.

List of Figures

- Fig. 1. Mean over 10 years of daily clearness index for the region under investigation. Data is derived from dataset obtained via NASA-SSE and linearly interpolated in this plot. Sampled locations from which correlations are calculated are represented as small grey dots.
- Fig. 2. Pair correlations of Kt variations as a function of pair separation for 6 distinct timescales noted in upper right corner of each subplot. Blue lines show fitted exponential trends and red vertical lines indicate the identified decorrelation distance (where $\rho = 0.1$) at each timescale.
- Fig. 3: Pair correlations of Kt variations as a function of pair separation for 6 distinct timescales where pairs of locations are oriented East-West of each other. Blue lines show fitted exponential trends and red vertical lines indicate the identified decorrelation distance (where $\rho = 0.1$) at each timescale.
- Fig. 4 Pair correlations of Kt variations as a function of pair separation for 6 distinct timescales where pairs of locations are oriented North-South of each other. Blue lines show fitted exponential trends and red vertical lines indicate the identified decorrelation distance (where $\rho = 0.1$) at each timescale.
- Fig. 5. Linearly interpolated polar plot of ΔKt pair correlations (color scale) as a function of pair separation (radius) and pair orientation (angle) for 6 distinct timescales. Dotted red isolines indicate the identified decorrelation distance (where $\rho = 0.1$) at each timescale.

Figure 1.

Image available upon request (too large to attach to this file)

Figure 2.

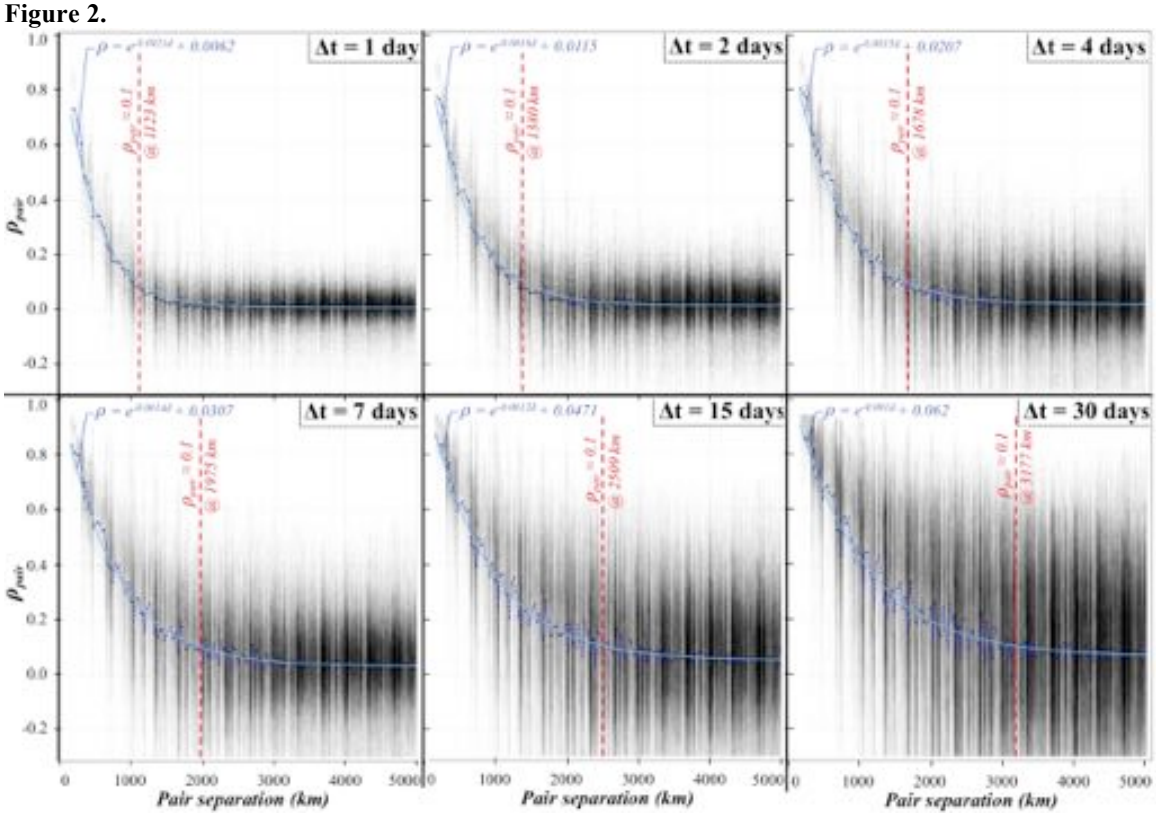

Figure 3.

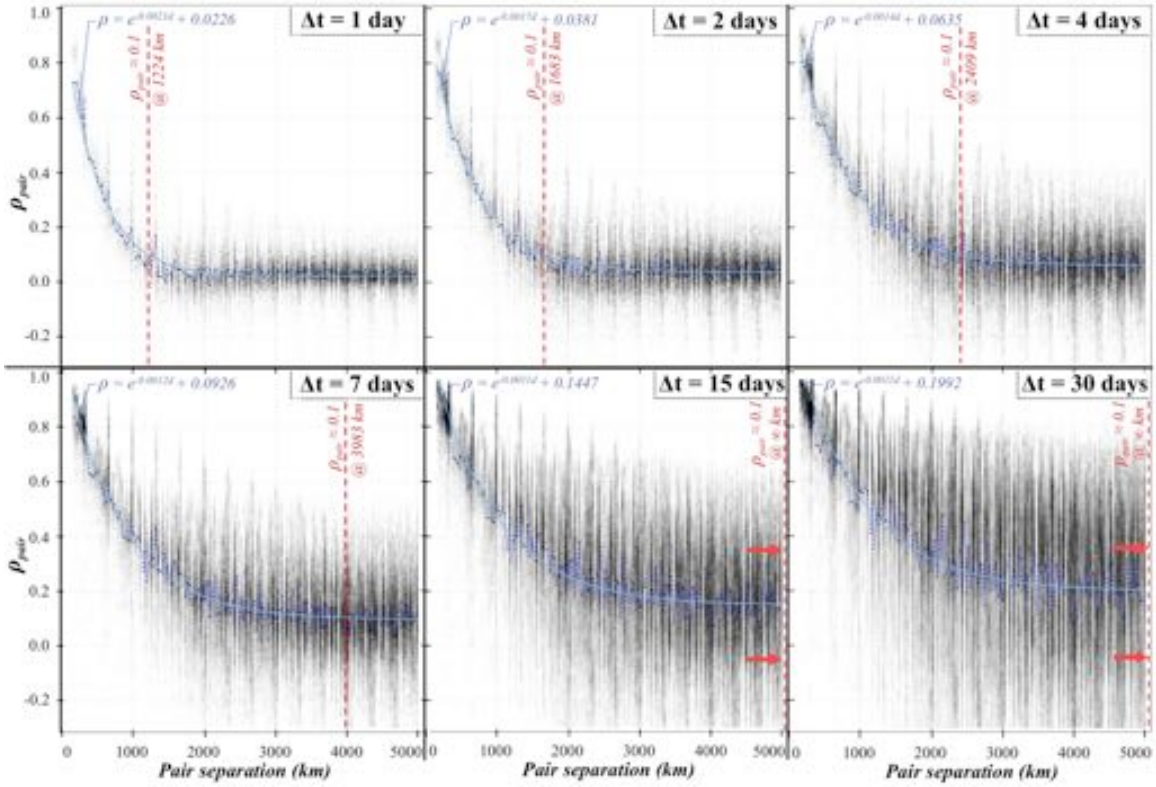
Δt = 1 day
Δt = 2 days
Δt = 4 days
Δt = 7 days
Δt = 15 days
Δt = 30 days
Pair separation (km)
Pair separation (km)
Pair separation (km)

Figure 4.

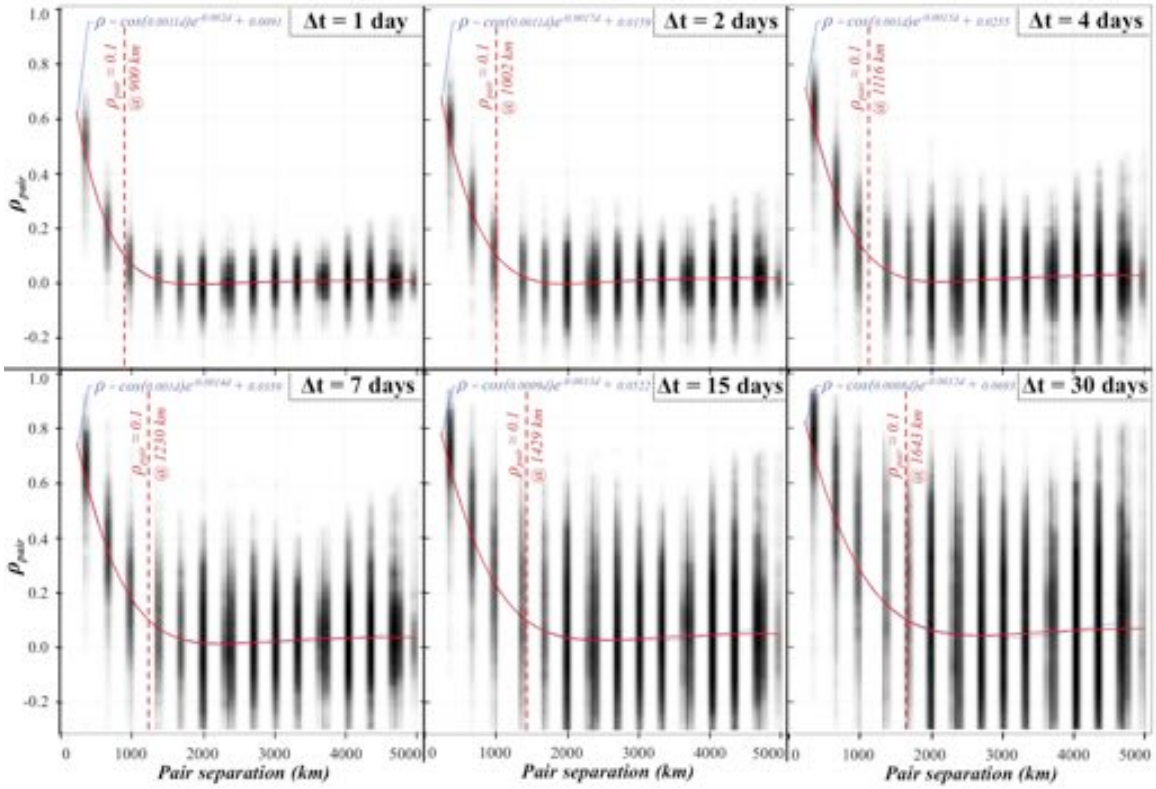
Δt = 1 day
Δt = 2 days
Δt = 4 days
Δt = 7 days
Δt = 15 days
Δt = 30 days
Pair separation (km)
Pair separation (km)
Pair separation (km)

Figure 5.

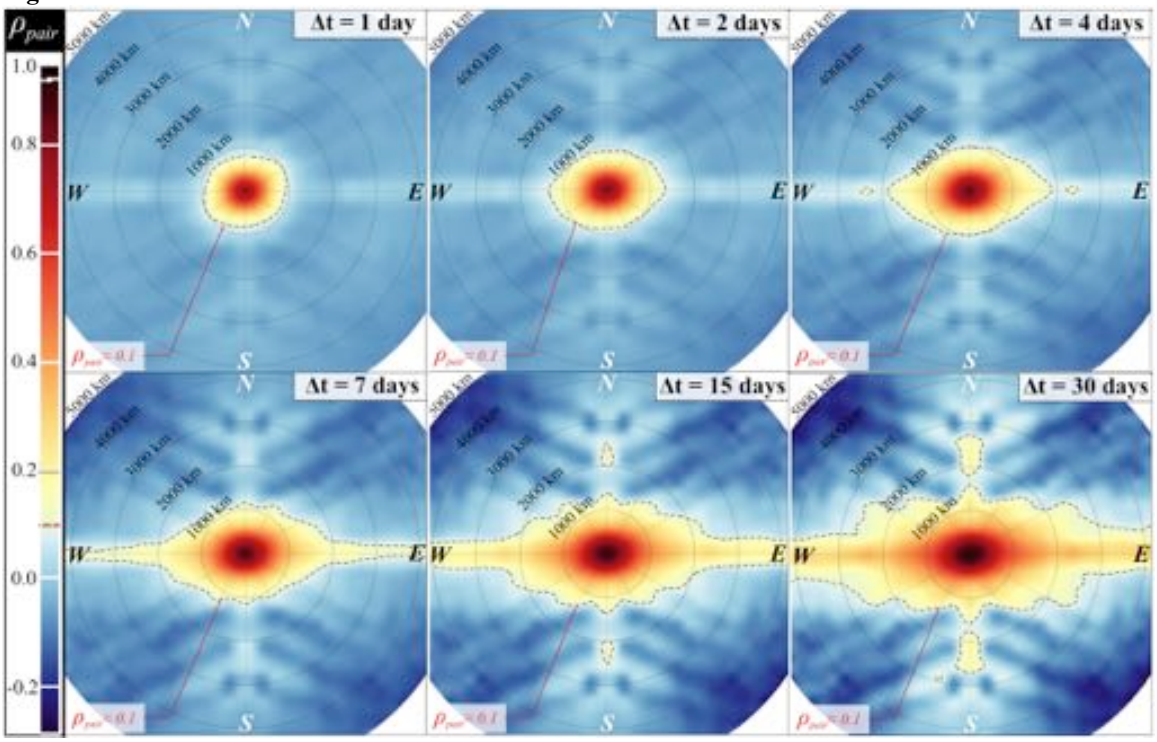

Table 1.

	All pairs			***E/W pairs***			***N/S pairs***			
	Decorrelation distance (km)	Trend parameters		*Decorrelation distance (km)*	Trend parameters		*Decorrelation distance (km)*	Trend parameters		
Δt *days*	ρ = 0.1 @	β x 10^{-3}	γ x 10^{-2}	*ρ = 0.1 @*	β x 10^{-3}	γ x 10^{-2}	*ρ = 0.1 @*	α x 10^{-3}	β x 10^{-3}	γ x 10^{-2}
1	*1123*	*2.11*	*0.62*	*1224*	*2.09*	*2.26*	*900*	*1.12*	*1.97*	*0.91*
2	*1380*	*1.76*	*1.15*	*1683*	*1.65*	*3.81*	*1002*	*1.09*	*1.70*	*1.59*
4	*1678*	*1.51*	*2.07*	*2409*	*1.37*	*6.35*	*1116*	*1.03*	*1.52*	*2.55*
7	*1975*	*1.35*	*3.07*	*3983*	*1.23*	*9.26*	*1230*	*0.98*	*1.41*	*3.59*
15	*2509*	*1.17*	*4.71*	*>5000*	*1.12*	*14.5*	*1429*	*0.89*	*1.28*	*5.22*
30	*3177*	*1.03*	*6.20*	*>5000*	*1.07*	*19.9*	*1643*	*0.83*	*1.16*	*6.93*