

Correlations In Spatial Variability When Accounting For Cloud Advection

Joseph Ranalli
Penn State Hazleton
Hazleton, PA
jar339@psu.edu

Abstract—Spatiotemporal variability of irradiance has been a topic of interest in the literature. This study attempts to separate advective and uncorrelated portions of the spatially distributed irradiance by comparing cross correlations between site pairs within an irradiance measurement network and accounting for time lag when calculating those cross correlations. Following techniques in the literature, cross correlations are computed as a function of site pair separation distance for multiple wavelet transform timescales. Results show that the well known form of the correlation’s decline can be maintained by considering the lagging correlation and the component of the site pair separation distance perpendicular to the overall cloud motion. This may open the door for additional fidelity in modeling of spatiotemporal irradiance that more formally represents how the loss of correlation with distance depends on the cloud motion vector.

Index Terms—variability, spatial aggregation, cloud motion vector, wavelet variability model

I. INTRODUCTION

Spatiotemporal variability of irradiance is a topic of interest for modelers of photovoltaic (PV) power generation. All solar generation is made up of spatially distributed modules that collect the solar resource, converting solar irradiance to electricity. In contrast, irradiance monitoring efforts often take the form of pyranometer stations that provide individual point-based measurements. Even the best measurements consist of a few point-based measurement locations. The output of spatially distributed PV generators is known to exhibit a lower degree of variability than these point measurements. This difference can be attributed to the spatial smoothing effect as a distributed plant aggregates (integrates) the irradiance over its geographic extent. Models of this effect exist, and represent a range of different phenomenological bases.

II. BACKGROUND

Models of spatial aggregation of irradiance attempt to relate the variability of a single point irradiance time series to that of an entire PV plant.

One of the most widely used models of this phenomenon is the Wavelet Variability Model (WVM) [1]. The WVM represents a distributed plant as a set of discrete points and infers a reduction in variability based on the reduction of correlation across this point population. It uses a functional representation of the correlation between spatially distant points i and j that depends on the separation distance, d_{ij} ,

the cloud motion speed, V_c , and the wavelet mode timescale, \bar{t} [2]:

$$\rho_{ij} = \exp\left(-\frac{d_{ij}}{\frac{V_c \bar{t}}{2}}\right) \quad (1)$$

In computing the plant’s output, the magnitude of each wavelet mode is scaled by a factor that depends on the correlation for all possible site pairs making up the plant. The shortest scales (highest frequencies) experience the greatest reduction, due to the inverse relationship between the correlation and the timescale. Here, the scale of the wavelet mode for the aggregate plant, w_P , is scaled down relative to the reference as the square root of the sum of all site pair correlations as found via Eq. 1.

$$w_P(t) = w_{ref}(t) \sqrt{\frac{\rho_{sum}}{N^2}} \quad (2)$$

The basis of the scaling, along with the derivation of the best fit to real data, is laid out in the two sources that were cited in the preceding text [1], [2]. The correlations described are instantaneous correlations, i.e. the site pairs are tested only for simultaneous variability, without allowing for any time lag. However, as noted by the original authors [2], “*along-wind sites can become negatively correlated ... if there is frozen cloud field advection.*” These negative correlations were also observed in data by other investigators [3], [4]. This occurrence actually implies that analyzing only instantaneous (i.e. zero-lag) correlation can be misleading and does not provide any discrimination of the relationship between the actual variability signal (i.e. cloud-advection-induced variability) and uncorrelated “noise” in the spatial variability of the plant.

The Cloud Advection Model (CAM) [5], [6] was developed as an attempt to describe the fully frozen condition, which is akin to assuming perfect correlation between all site pairs in the direction of the cloud motion, when accounting for time lag. The original reports of the CAM describe improvements in representing the frequency domain characteristics of the aggregate plant response. However, they also describe situations in which the assumption of perfect correlation is too strong. That is, high frequency oscillations are observed to be incoherent, implying a loss of lagging correlation with increasing frequency. As such, some modification is necessary to account for the decorrelation present in the observed data.

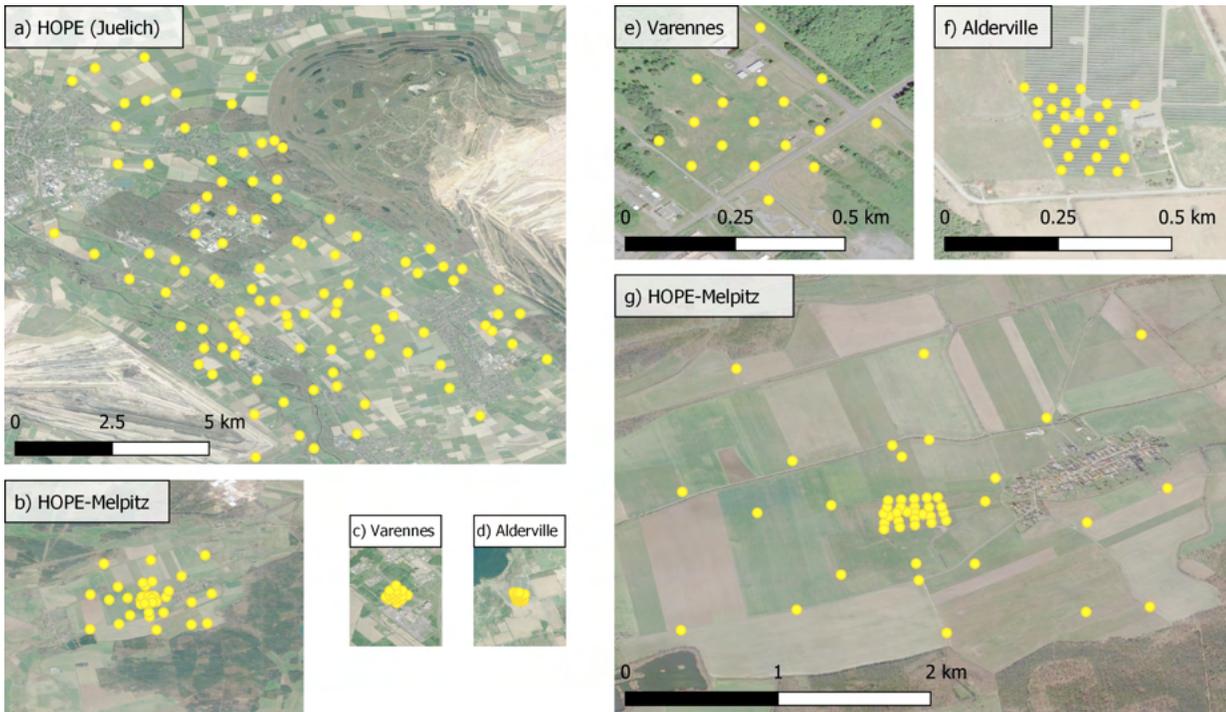


Fig. 1. Maps of the measurement campaign layouts for each of the datasets used. Note that subplots a-d share a scale, while e-g repeat previous figures at a higher scale to show additional detail for these configurations. a) HOPE (Jülich), b) HOPE-Melpitz, c) Varennes, d) Alderville, e) Varennes (detail view), f) Alderville (detail view), g) HOPE-Melpitz (detail view).

The analysis in this paper seeks to describe the middle ground between the WVM (no advective lag) and the CAM (frozen advection) by investigating the relationship between *lagging* correlations between multi-scale wavelet modes and the distance of site-pair separation.

III. DATA

Two different distributed irradiance data sources were used in this study. The HOPE campaign [7] made distributed measurements of irradiance in Germany using a network of point sensors. The original campaign was conducted with 100 sensors near Jülich, Germany, while a second campaign was conducted with 50 of those sensors installed near Melpitz, Germany. The data for both campaigns have a time interval of 1 second. An additional distributed irradiance dataset consists of two distributed measurement sites in Eastern Canada [8], one near Varennes, Quebec and one near Alderville, Ontario. Data from these sites was sampled at a maximum rate of 100 Hz, but data were only recorded for time intervals when the irradiance underwent a significant change. As a result, back-filling was necessary to complete the time series. In order to provide a consistent time step with the HOPE data, the Canadian site measurements have been resampled by averaging to a 1 second sample rate for this study. This has the added benefit of removing the somewhat unrealistic step changes from the irradiance timeseries.

Throughout the remainder of this paper, the four sites will be abbreviated with designations JUL (Jülich), MEL (Melpitz), VAR (Varennes) and ALD (Alderville). An image of the

distribution of the sensors for each campaign is given in Fig. 1.

These datasets represent various levels of geographic extent and variable cloud motion conditions. The largest site is JUL covering a total region of roughly 10km x 5km along its diagonal axis, while the two Canadian sites are smallest at approximately 0.25km x 0.25km. MEL spans roughly 3km x 2km, but also contains a relatively dense section of centrally located points. For this study, data used from VAR and ALD was limited to the publicly available, "variable" and "highly variable" days on the Natural Resources Canada (NRCAN) website ¹.

A. Calculation Methodology

Time series were obtained for each dataset and all were resampled to a sampling period of 1 second. Irradiance was converted to clearness index using the simplified solis clear sky model as implemented in PVLIB-Python [9]. A time window of interest was chosen for each dataset, during which the cloud motion vector was observed to be approximately constant. For ALD and VAR, this window was from 08:00-16:00 on the days indicated. For MEL, the window chosen was September 8, 2013 from 09:00-13:00 and for JUL, it was May 15, 2013 from 09:00-15:00. Cloud motion vectors were computed using the method described by Jamaly and Kleissl [10], which was

¹<https://www.nrcan.gc.ca/energy/renewable-electricity/solar-photovoltaic/18409>

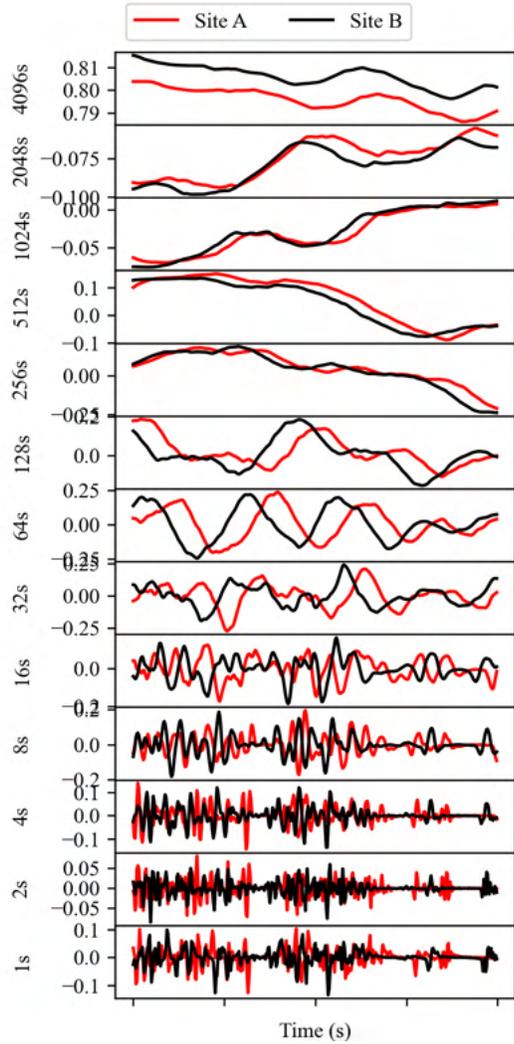


Fig. 2. Comparison of multiple timescales of wavelet modes for a 5 minute period for two separate sites in the ALD dataset.

compared to and yielded similar results to the method of Gagné et al [11], [12].

Wavelet modes were calculated using the implementation of the WVM in PVLIB-Python [9]. A total of 12 modes with timescales from 2s up to 4098s were computed for each time series. We excluded the shortest timescale as its parent wavelet shape was inconsistent with the top-hat wavelet used for all other modes. The cross correlation between every possible site pair, at every wavelet timescale, was computed using the signal processing library within scipy [13]. Scaling was applied to the cross correlation results, such that the value represented the correlation coefficient, ranging from zero to one. Results were computed at all possible lags without any biasing applied, which provides a slight favoritism to smaller lags. An example of the wavelet modes for two separate sites is shown in Fig. 2. The lagged nature of the correlation between these sites is clearly visible, especially for intermediate timescale modes.

IV. RESULTS

Plots similar to those of Lave et al [2] were generated to show the relationship between pairwise correlation, separation distance and wavelet timescale. To understand the results with respect to cloud advection, the site-pair separation distance was further differentiated into parallel and perpendicular directions, relative to the cloud motion vector. The absolute value of the perpendicular distance is used in the colorization of the plot. The results for the ALD very-variable day are shown in Fig 3.

As is evident from the leftmost plot, the proposed correlation of Lave et al. [2] is a good fit for the zero-lag correlation case. However, when computing the the maximum correlation allowing for a temporal offset (lag), we achieve the result in the central plot in Fig 3, which exhibits an upward shift for nearly every point, but especially those with small separation distances perpendicular to the cloud motion. Locations with a larger perpendicular separation are not as significantly affected, indicating that the correlation for these points is already well explained by the zero-lag case. We attempt to correct for this biasing of the correlation by replacing the absolute distance in Eq. 1 with the perpendicular separation distance in computing the x-axis variable (see Eq. 3). As seen in the right panel of Fig. 3, use of this distance helps return the dispersion of the data to the prediction line. The suitability of this match indicates that the uncorrelated variability tends to occur perpendicular to, rather than along, the cloud motion.

$$\rho_{ij,lag} = \exp\left(-\frac{d_{\perp ij}}{\frac{V_c}{2}t}\right) \quad (3)$$

While Fig. 3 maintains continuity with how this data is presented elsewhere in the literature, the x-axis scaling has the undesirable effect of compressing all the short timescale points almost onto the y-axis. As utilizing the peak lagging correlation mostly affects short- and moderate timescales, it would be helpful to utilize an axis scaling that better highlights those timescales. In Fig. 5a, we re-present the data from Fig. 3, but include the cloud speed coefficient in the exponent. This results in a plot that linearly represents the curve fit described by Eq. 3. The same effects described previously are clearly visible in this data.

One exception to the quality of fit is the moderate number of scattered points that appear in the lower right of the rightmost plot. Recoloring these points by their raw absolute distance, as in Fig. 4, demonstrates that though these are predominantly points that, despite having very small perpendicular separation distances, have large separation distances in the direction of cloud motion. The lagged correlation of these points remains much lower than would be expected for frozen advection. This implies that for large spatial distances, the cloud-advection is not truly frozen and highlights the fact that both cloud-advection and uncorrelated noise components play some role in the overall spatiotemporal variability.

Results for multiple test sites are shown in the other parts of Fig. 5. The best performing cases are ALD highly variable,

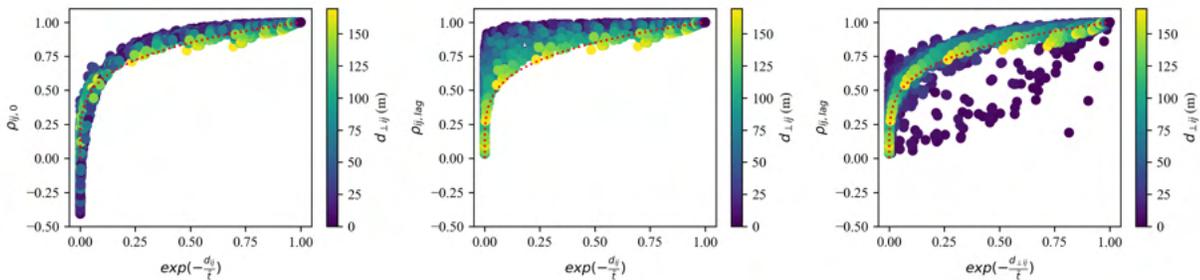


Fig. 3. Dependence of correlation on separation distance and timescale as in [2] for ALD, Aug. 12, 2015 from 08:00 through 16:00. left) correlation at zero lag and absolute distance, center) maximum correlation at any lag and absolute distance, right) maximum correlation at any lag and distance perpendicular to cloud motion. All plots are colored by the distance perpendicular to the cloud motion direction. Dotted line is correlation as proposed by Lave et al.

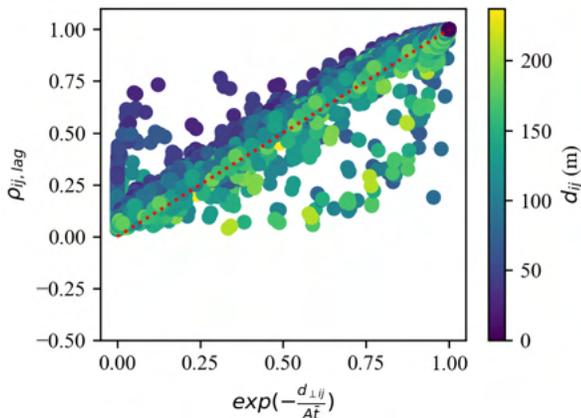


Fig. 4. Data from the rightmost panel of Fig. 5a, colored to show the absolute pair separation distance.

VAR variable and MEL (a, c and e). All three cases show a clear increase in the correlation between short timescale modes when allowing for lag (as in the center column). The improvement provided by utilizing the model based on Eq. 3 is also evident in the right column of these plots. Less benefit is seen when considering ALD variable, VAR highly variable and JUL cases (b, d and f). For all three of these cases with lesser performance, we do observe that tracking the peak lagging correlation does increase the overall level of correlation, pushing the points toward the upper left of the plot in the central column. The predominant difference seen for these cases is the degree to which points with a large perpendicular separation distance are also shifted to the upper left, implying that the correlation is stronger than expected at greater separation distances perpendicular to the cloud motion. These points experience less shifting when switching to the modified form of the equation, as they correspond to points whose absolute and perpendicular separation distances are similar. Points with short perpendicular separation distances are in fact shifted closer to alignment with the $x = y$ predicted line.

To actually quantify the suitability of the modified model,

TABLE I
QUALITY OF FIT FOR EACH CASE.

Site	Day	Cloud Spd. (m/s)	Base ρ	Peak Lag ρ	Modified Model ρ
ALD	HV	8.5	0.976	0.934	0.964
ALD	V	30.4	0.899	0.870	0.825
VAR	HV	40.7	0.884	0.725	0.750
VAR	V	10	0.981	0.958	0.985
MEL	Sept 8	19.5	0.967	0.928	0.955
JUL	May 15	17.8	0.932	0.887	0.889

the quality of the best-fit line between the model and the data points can be computed for each of the plots. These results are shown in Table I. For each case except ALD variable, switching to the modified distance measurement improves our ability to predict the correlation between site pairs when allowing for lag. As seen in the table, the cases with the best performance corresponded to those with the lowest cloud motion speeds.

V. CONCLUSION

This study demonstrates that when considering separated site-pair correlations, the use of a lagging cross correlation can improve discrimination of the variability induced by advection from that introduced by "noise" in the cloud field. Accounting for the lagging correlation increases the level of measured correlation between wavelet modes, particularly for sites whose time series are closely related by virtue of lying parallel to the cloud motion direction. The analysis demonstrated that the empirical model proposed previously by Lave et al. [2] better fits the lagging correlation when scaling by the perpendicular separation between sites relative to the cloud motion, rather than the absolute distance. This confirms that the uncorrelated "noise" in the wavelet modes is more significant in the direction perpendicular to the cloud motion, and frozen advection is important along the direction of cloud motion. In the test cases considered here, performance of this modified model was better for conditions with slower cloud motion speeds. Further studies are warranted to further investigate the use of this modified distance measurement and to incorporate discrimination of advection- and randomness-

induced variability into models of the smoothing of plant power generation time series.

“SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nature Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020.

VI. ACKNOWLEDGMENT

The author would like to acknowledge Sophie Pelland of Natural Resources Canada for sharing access to the NRCAN test site data and code for variability analysis, and Matthew Lave of Sandia National Laboratories for sharing WVM-related modeling code.

REFERENCES

- [1] M. Lave, J. Kleissl, and J. S. Stein, “A Wavelet-Based Variability Model (WVM) for Solar PV Power Plants,” *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 501–509, Apr. 2013.
- [2] M. Lave and J. Kleissl, “Cloud speed impact on solar variability scaling – Application to the wavelet variability model,” *Solar Energy*, vol. 91, pp. 11–21, May 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0038092X13000406>
- [3] V. P. A. Lonij, A. E. Brooks, A. D. Cronin, M. Leuthold, and K. Koch, “Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors,” *Solar Energy*, vol. 97, pp. 58–66, Nov. 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0038092X13003125>
- [4] B. Elsinga, “Chasing the Clouds: Irradiance Variability and Forecasting for Photovoltaics,” Dissertation, Utrecht University, Nov. 2017. [Online]. Available: <http://dspace.library.uu.nl/handle/1874/356774>
- [5] J. Ranalli, E. E. Peerlings, and T. Schmidt, “Cloud Advection and Spatial Variability of Solar Irradiance,” in *47th IEEE Photovoltaic Specialist Conference (PVSC)*. Virtual: IEEE, Jun. 2020, p. 8.
- [6] J. Ranalli and E. E. M. Peerlings, “Cloud advection model of solar irradiance smoothing by spatial aggregation,” *Journal of Renewable and Sustainable Energy*, vol. 13, no. 3, p. 033704, May 2021, publisher: American Institute of Physics. [Online]. Available: <https://aip.scitation.org/doi/abs/10.1063/5.0050428>
- [7] A. Macke, P. Seifert, H. Baars, C. Barthlott, C. Beekmans, A. Behrendt, B. Bohn, M. Brueck, J. Bühl, S. Crewell, T. Damian, H. Deneke, S. Düsing, A. Foth, P. D. Girolamo, E. Hammann, R. Heinze, A. Hirsikko, J. Kalisch, N. Kalthoff, S. Kinne, M. Kohler, U. Löhnert, B. L. Madhavan, V. Maurer, S. K. Muppa, J. Schween, I. Serikov, H. Siebert, C. Simmer, F. Späth, S. Steinke, K. Träumner, S. Trömel, B. Wehner, A. Wieser, V. Wulfmeyer, and X. Xie, “The HD(CP)² Observational Prototype Experiment (HOPE) – an overview,” *Atmospheric Chemistry and Physics*, vol. 17, no. 7, pp. 4887–4914, Apr. 2017. [Online]. Available: <https://www.atmos-chem-phys.net/17/4887/2017/acp-17-4887-2017-discussion.html>
- [8] A. Gagné, D. Turcotte, N. Goswamy, and Y. Poissant, “High resolution characterisation of solar variability for two sites in Eastern Canada,” *Solar Energy*, vol. 137, pp. 46–54, Nov. 2016. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0038092X16303000>
- [9] W. Holmgren, C. Hansen, and M. Mikofski, “pvlb python: a python package for modeling solar energy systems,” *Journal of Open Source Software*, vol. 3, no. 29, p. 884, Sep. 2018. [Online]. Available: <https://joss.theoj.org/papers/10.21105/joss.00884>
- [10] M. Jamaly and J. Kleissl, “Robust cloud motion estimation by spatio-temporal correlation analysis of irradiance data,” *Solar Energy*, vol. 159, pp. 306–317, Jan. 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0038092X17309556>
- [11] A. Gagné, N. Ninad, J. Adeyemo, D. Turcotte, and S. Wong, “Directional Solar Variability Analysis,” in *2018 IEEE Electrical Power and Energy Conference (EPEC)*, Oct. 2018, pp. 1–6, iSSN: 2381-2842.
- [12] S. Pelland, A. Gagné, M. A. Allam, D. Turcotte, and N. Ninad, “Spatiotemporal Interpolation of High Frequency Irradiance Data for Inverter Testing,” in *2021 IEEE 48th Photovoltaic Specialists Conference (PVSC)*, Jun. 2021, pp. 0211–0218, iSSN: 0160-8371.
- [13] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors,

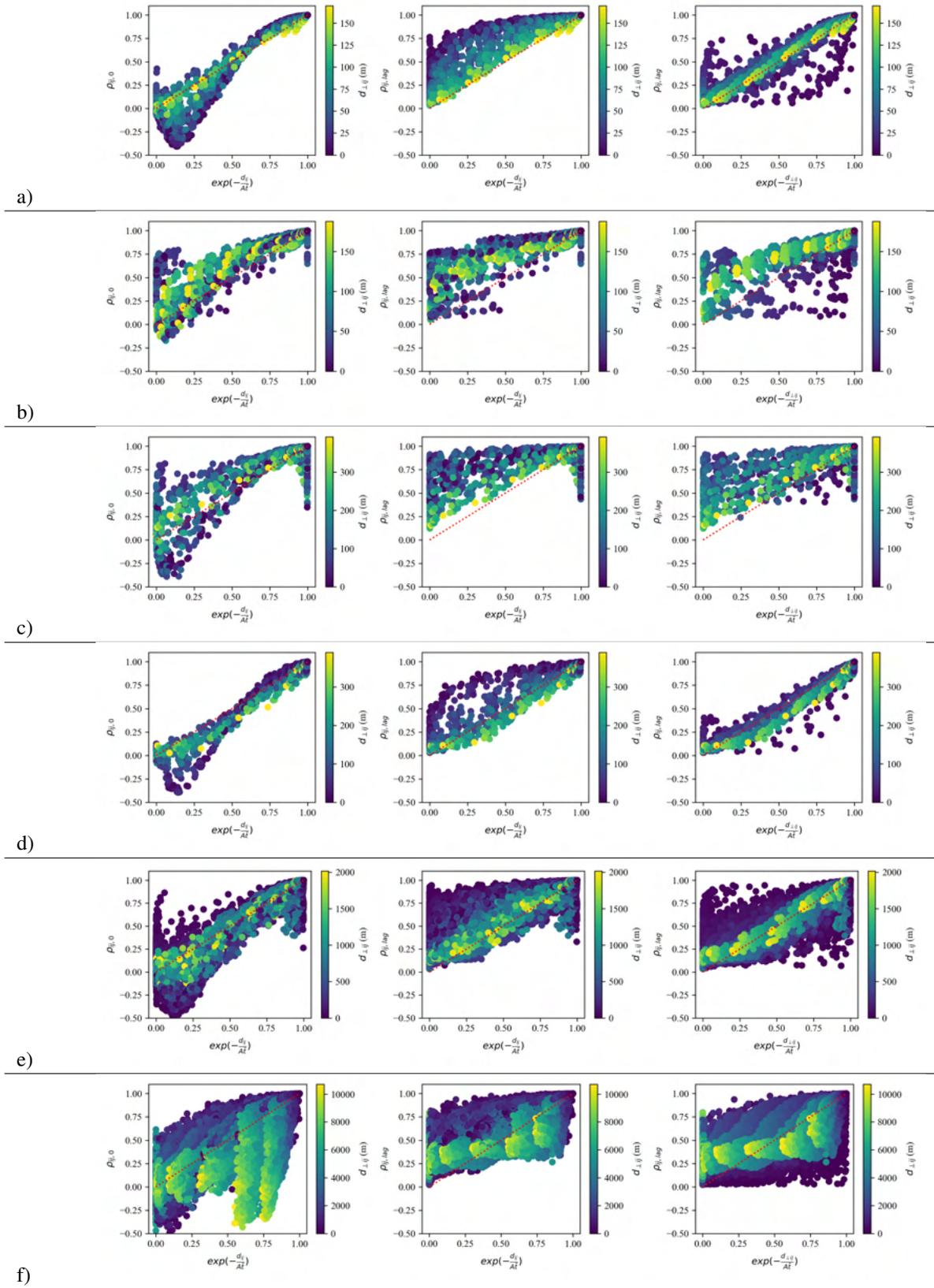


Fig. 5. Similar to Fig. 3, but with linearized scaling on the x-axis. left col) correlation at zero lag and absolute distance, center col) maximum correlation at any lag and absolute distance, right col) maximum correlation at any lag and distance perpendicular to cloud motion. Rows show the different test sites: a) ALD, Highly Variable (same as Fig. 3). b) ALD, Variable. c) VAR, Highly Variable. d) VAR, Variable. e) MEL, Sept 8, 2013 09:00-13:00. f) JUL, May 15, 2013 09:00-15:00