Geographic smoothing of solar photovoltaic electric power production in the western USA

Kelly Klima[1], Jay Apt[1, 2], Mahesh Bandi[3], Paul Happy[4], Clyde Loutan[5] and Russell Young[4]

[1] Department of Engineering and Public Policy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
[2] Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
[3] Collective Interactions Unit, OIST Graduate University, Onna, Okinawa, 9040495 Japan
[4] SoCore Energy, 225 W. Hubbard, Suite 200, Chicago, IL 60654
[5] California Independent Systems Operator (CAISO), 250 Outcropping Way, Folsom, CA 95630

E-mail: apt@cmu.edu

Keywords: solar photovoltaic, geographic smoothing, grid-connected PV systems, spectral analysis, intermittency, variability

Supplementary material for this article is available online

Abstract
We examine the geographic smoothing of solar photovoltaic (PV) generation from 15 large utility-scale plants in California, Nevada, and Arizona and from 19 installations on the roofs of commercial buildings in California. Plant sizes for the utility-scale generators were 125-315 MW and the plants cover an area ~300 x 800 km. The commercial rooftop PV generators were 80-520 kW and cover ~175 x 900 km. Examining the power output of these generators in the frequency domain, we quantify the smoothing obtained by combining the output of geographically separated plants. Utility-scale and commercial rooftop plants exhibit similar geographic smoothing, with 10 combined plants reducing the amplitude of fluctuations at 1 hour to 18-28% of those seen for a single plant. The smoothing observed for these western USA PV generators is greater than that seen in the Indian state of Gujarat. PV does not exhibit as much geographic smoothing as is seen for combining wind plants.

1. Introduction

Solar photovoltaics (PV) provided 7.5% of Germany's 2015 net electricity consumption [1] and PV's market share is increasing in other countries. Solar power production has significant short-term variability caused by clouds, diurnal intermittency, and seasonal variability. Analysis of the effects of this variability and intermittency on power systems began well before the recent increase in PV's market penetration [2,3]. It has long been recognized that a portion of the short-term variability in PV plant generation might be mitigated by summing the outputs from geographically dispersed plants.

Several authors have examined observed geographic smoothing of the output of PV plants[1]. Wiemken et al [4] examined the standard deviation and generation duration curves of the average daily power generation from 100 rooftop PV systems in Germany using 5 minute time resolution data. Murata, Yamaguchi, and Otani [5] used a metric of the largest PV output fluctuation in a given time interval to examine geographic smoothing of PV generators in Japan, finding modest decreases in their output fluctuation coefficient as up to 20 small (0.1-5.6 kW) generators were summed. Mills and coauthors [6] used the 3σ power output change rate (ramp rate) of utility-scale plants in Arizona, finding a reduction in the 1-min and 10-min extreme ramp rates when geographically separated plants are added. Marcos et al [7] used data from 7 PV plants with a total installed capacity of 20 MW and separated by 6-360 km to examine the reduction in the largest observed power fluctuation at discrete periods of 1, 4, 20, 60 and 600 seconds by adding plants together. They found that adding the first few plants together quickly reduced fluctuations, but that adding additional plants was of limited value. Lave, Stein and Ellis [8] examined the sum of residential rooftop PV in Ota City, Japan (2.1 MW total) and a 19 MW utility-scale plant in Colorado. The plants were similar in geographic extent (the utility-scale plant covered roughly twice the area as the rooftop installations). Using the maximum output change in 1 sec over a single day at each site, they find that (1) the ramp rates are reduced by roughly an order of magnitude for the entire plant as compared to those at a single point, and (2) the relative variability decays exponentially as additional houses, or (in the case of the utility-scale plant) inverters are added. They find that the smoothing decreases quickly as the first few houses or inverters are added, then much more slowly as additional units are added.

A few groups have studied the decrease in variability as sub-arrays within a larger PV plant are added together. Van Haaren, Morjaria and Fthenakis [9] examined sub-arrays at one 48MW plant, plotting a cumulative distribution function of the 1-min ramp rate for increasing sub-array size. They make an estimate of the highest ramp rates for a plant that is covered by a cloud. If the cloud is larger than the areal extent of the plant and is moving at mean velocity $\bar{v}$ parallel to the shortest side of a rectangular array with short dimension L, they argue that the maximum ramp rate in time interval Δt will be $\bar{v}\,\Delta t / 2L$ for the case where $L / \bar{v} > \Delta t$; they examine other cases numerically.

[1] Observations of output data from plants is preferred to work using measured or simulated solar irradiance, in part because real solar plants have power electronics that respond to changes differently than do irradiance monitors. For example, the inverters have a maximum power output that limits responses to cloud focusing; they also have finite response rates.

Methods that use the step change over a particular interval or the standard deviation of output fluctuations in a period of time give information on fluctuations at selected time scales. However, there is a method that can allow the determination of how much smoothing is obtained at any desired time scale. The power output of a PV plant contains fluctuations at many frequencies. Curtright and Apt [10] obtained 10 sec time resolution data over two years from a 4.6 MW utility-scale array in Arizona, Fourier transformed the time domain data into the frequency domain, and found that the power spectrum shows a falloff toward higher frequencies that goes roughly as frequency $f^{-1.3}$ (in addition to the expected peak at a frequency corresponding to 24 hours and its harmonics).

Klima and Apt [11] used information from the frequency domain to examine geographic smoothing of 20 utility-scale (3-221 MW) plants in the Indian state of Gujarat at 1-2 min time resolution. They found that "Interconnecting approximately 20 plants yields a 25%-45% reduction in variability depending on [the] frequency examined." This technique allows the decrease in variability by adding plants together to be examined at any desired time scale. Similar to wind [12], there is more smoothing at time scales of 10 minutes than at time scales of 6 hours. Also similar to wind, there is considerably more amplitude in the variations at long time scales. In agreement with Marcos et al and with Lave, Stein and Ellis, they found quickly diminishing returns as more plants are summed. However, the Gujarat data showed quite modest geographic smoothing for these plants even over 400 km. At time scales of one hour, for example, the Gujarat data showed that only half the amplitude of the variability at this time scale was eliminated; for wind plants with a similar geographic spread 95% of the variability is extinguished.

Here we use both utility-scale and rooftop PV data from locations in the US states of California, Nevada and Arizona to study smoothing using the frequency domain method. We compare geographic smoothing of commercial building rooftop PV to that for utility scale PV, finding that geographic separation provides a similar degree of smoothing for rooftop PV and for larger PV plants. We conclude by proposing a theory that may provide some insight into the difference between geographic smoothing for wind and that for solar photovoltaic power.

2. Data

For utility-scale PV plants in the USA, we used data provided by the California Independent Systems Operator (CAISO) for one year at 1-minute resolution for 15 plants in California, Nevada, and Arizona at one-minute time resolution. Plant sizes were 125-315 MW. Southern California Edison (SCE) and SoCore Energy provided one year of 15-minute resolution data on 19 rooftop PV arrays of 80-520 kW placed on commercial buildings such as drugstores. Plant locations for the CAISO and SCE data are shown in Figure 1. Additional characteristics of the data sets are provided in the online supplementary material.

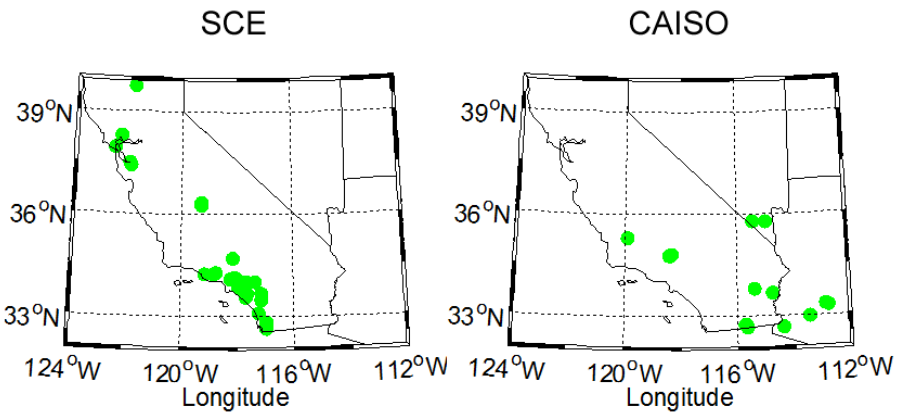

Figure 1. *Locations of solar PV systems. The green dots represent sites with excellent data that were used. The CAISO data cover ~300 x 800 km, and the Southern California Edison (SCE) data ~175 x 900 km. The CAISO data are utility-scale PV arrays, while the SCE data are rooftop arrays on commercial buildings.*

Sample generation characteristics for the CAISO data and SCE rooftop data are shown in Figure 2. Many of the SCE commercial rooftop inverters were sized so that the output reached the rated capacity on sunny days and thus has a characteristic flat profile near local noon during clear days during spring, summer, and fall months. Comparing geographic smoothing between the CAISO and SCE data only for January 2016 avoided any distortions that may arise from the saturated output.

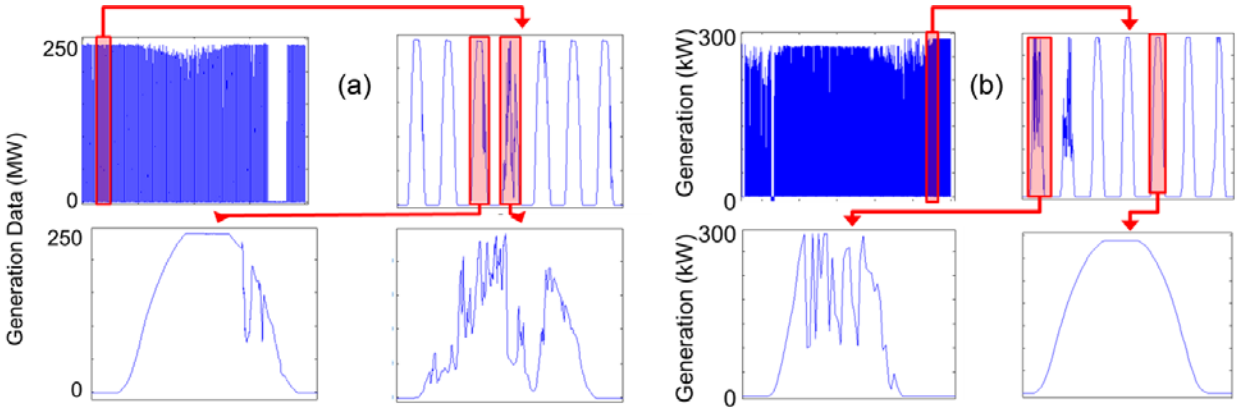

Figure 2. *Generation data for one representative CAISO large plant (a) and for an SCE commercial building rooftop plant (b) for one year beginning July 1, 2015, one week, a partly cloudy day, and a clear day. The power electronics for both are sized so that the maximum output saturates on sunny summer days.*

3. Methods

Following [11] and [12], we examined the generation data in the frequency domain. To handle the observed uneven time steps in the Gujarat data, we used the Lomb periodogram [13] as coded in Press et al [14]. Although we find no difference between the Lomb periodogram

procedure and a standard Fourier transform procedure, to ensure our procedure was similar between datasets, we also used the Lomb periodogram for all data sets.

An attribute of the Fourier or Lomb methods of estimating the power spectral density (PSD) is that increasing the temporal length of the dataset does not reduce the standard deviation of the PSD at any frequency. To increase the signal-to-noise ratio, we used the standard technique of partitioning the dataset into time segments. Figure 3 shows representative PSDs for CAISO and SCE.

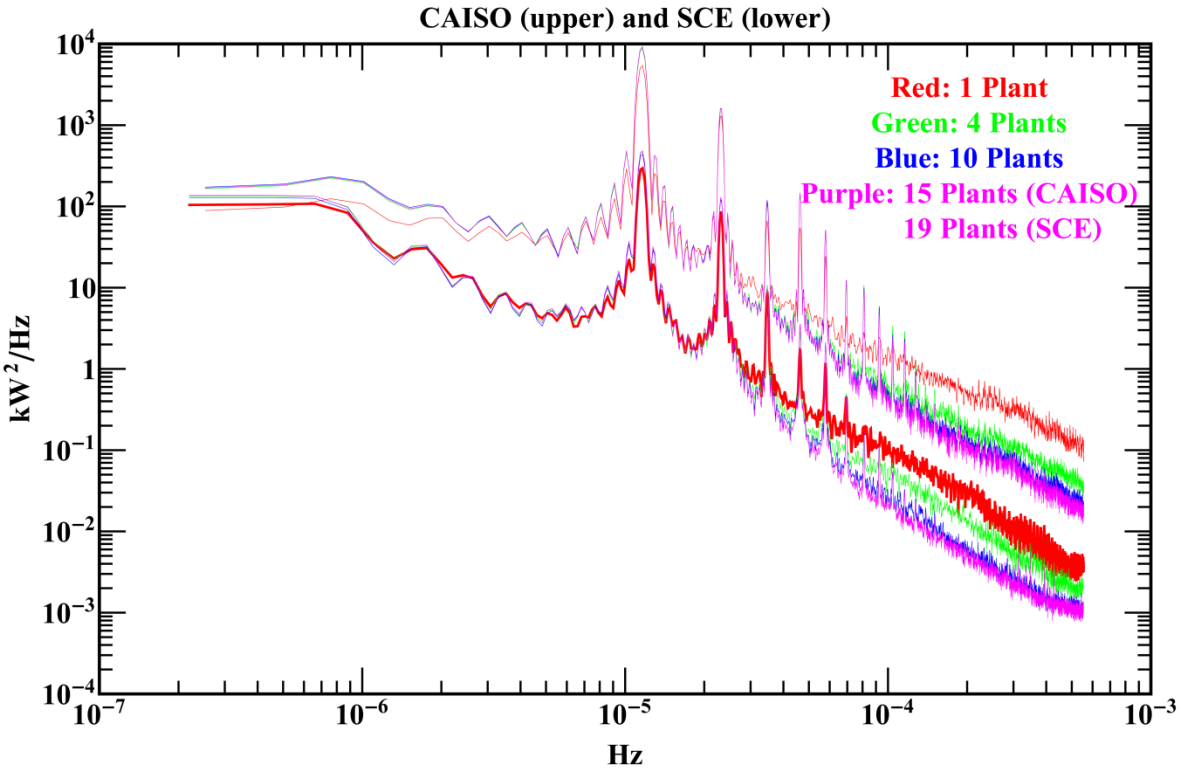

Figure 3. *Power spectral densities for a representative plant in CAISO and SCE (red), the sum of 4 plants (green), 10 plants (blue), and all plants (purple). The SCE data are snapshots taken at 15-minute intervals, so the Nyquist frequency corresponds to 30 minutes. The PSDs for the 1-minute CAISO data have been truncated at the same frequency. 32-segment averaging has been used.*

Next, to understand the potential for smoothing variability in plants, we summed plants, calculated the PSD, and compared the slopes.

We investigated how interconnecting combinations of plants in each data set can potentially provide smoothing (Figure 4) . The procedure (as in [11] and [12]) was to compare the line of best fit for the two PSDs at particular frequencies by taking the ratio of the single plant value to that of the interconnected plants value. If no smoothing occurs when solar plants are interconnected, the result should be close to one at all frequencies. If there is a reduction in variability then there will be frequencies for which the fraction is less than one. This procedure

allows the determination of the fraction of the amplitude fluctuations that are smoothed by adding a given number of plants at any given frequency (for example, the frequency corresponding to 6 hours or to 1 hour, Figure 5).

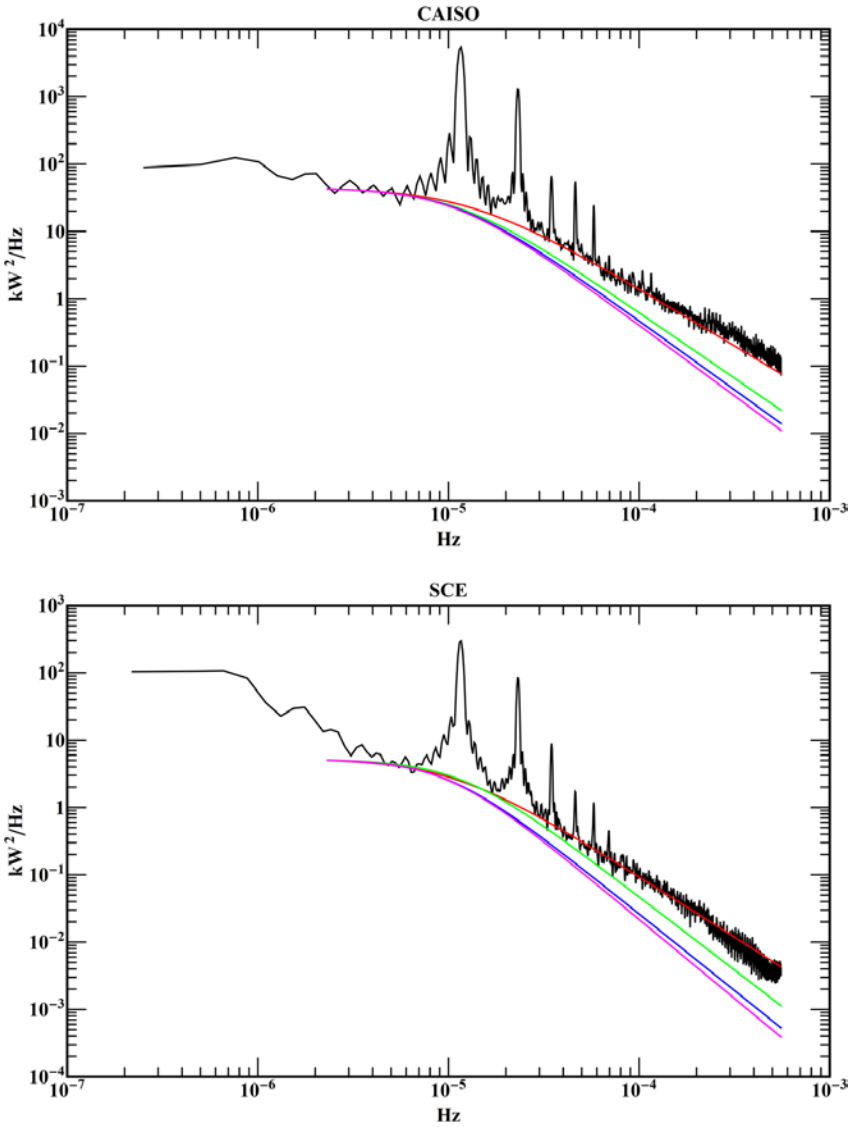

Figure 4. *Power spectral density (PSD) plots for one CAISO utility-scale PV plant (top) and one SCE commercial rooftop plant (bottom).Colored lines are the fits of the form* $A/(1+Bf^{\alpha})$ *to the PSDs for one plant (red), 4 plants (green), 10 plants (blue), and 15 plants for CAISO or 19 plants for SCE (purple).PSDs for the combination of plants are not shown for clarity.*

4. Results

We compare the smoothing observed for the utility-scale CAISO plants in California, Nevada, and Arizona (~300 x 800 km) to that for SCE rooftop PV in California (~175 x 900 km) and to the observed smoothing for wind plants over a similar geographic extent in Texas. Utility-scale

and commercial rooftop PV plants exhibit similar geographic smoothing, considerably less than is observed for wind.

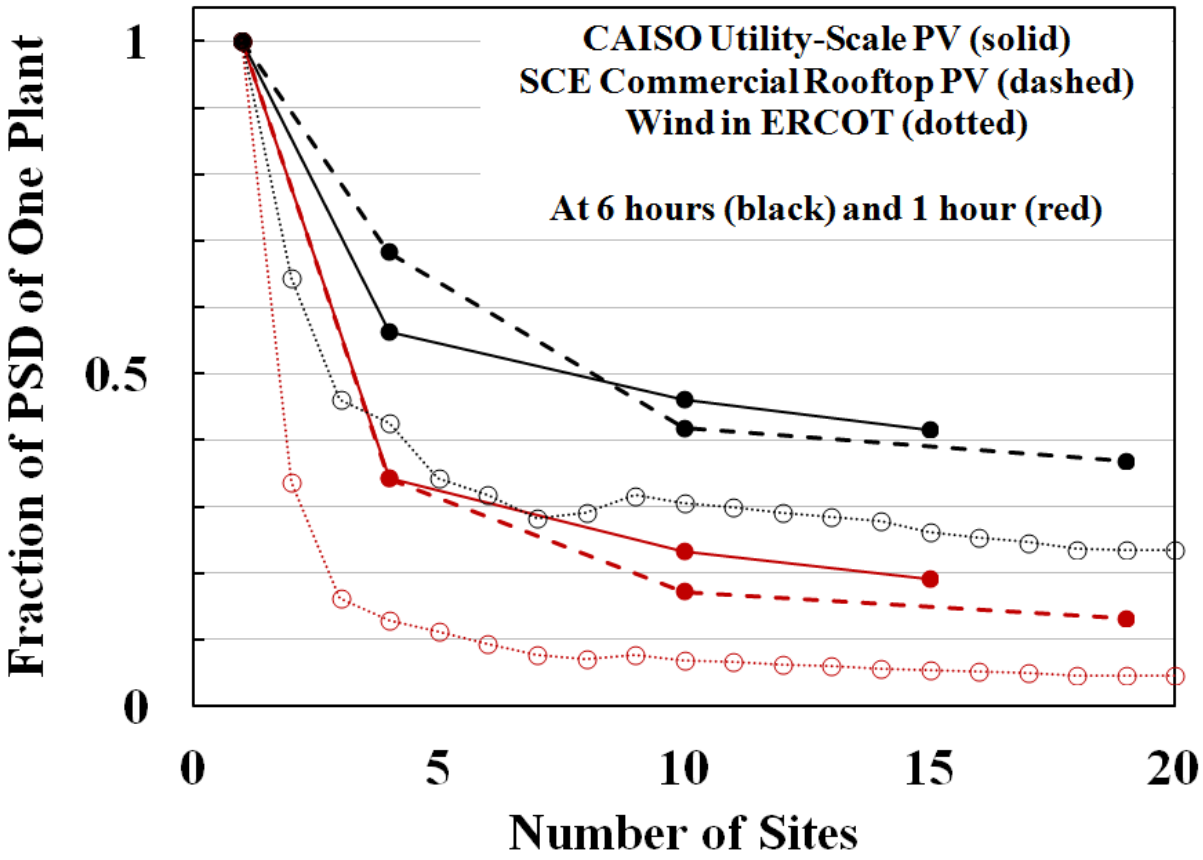

Figure 5. *Comparison of smoothing at 6 hours in black and 1 hour in red among the 120-315 MW utility-scale plants in CAISO (solid lines) with the 5-400 kW SCE commercial rooftop plants in California (dashed lines) and with the smoothing found by adding geographically separated ERCOT wind plants over 500 km [12] (dotted lines with open circles). The CAISO data cover ~300 x 800 km, and the Southern California Edison (SCE) data ~175 x 900 km. The ERCOT wind data cover ~200 x 500 km. A full year of data is used for all PV plants shown here, beginning July 1, 2015; the wind data of [12] shown here was for the full year 2008.*

5. Discussion

When geographically separated PV plants are added together, the slope of the summed power spectrum steepens; we observe nearly identical steepening when rooftop plants are summed as when utility-scale plants are summed (Figure 4).

We observe no significant difference in geographic smoothing between physically large utility-scale PV arrays and small commercial rooftop arrays. Geographic smoothing of PV variability is modest compared to geographic smoothing of wind power variability over comparable geographic areas (Figure 5).

Power spectral density plots are given in the supplementary data for the 19 rooftop and 15 utility-scale plants examined here. As for utility-scale PV generation examined in earlier research [10, 11], the spectra are relatively flat at frequencies lower than ~10^{-5} Hz, and fall off at higher frequencies.

We find that PV plants experience less geographic smoothing than do wind plants. Wind and solar photovoltaic power fluctuations arise from quite different physical considerations. Wind speed varies around a usually non-zero mean value, causing wind power to fluctuate, albeit with large magnitude, above and below a normally steady power. The largest length scales of atmospheric flows spanning hundreds of km and representing the lowest frequencies (inverse daylight time ~ 1.2 X 10^{-5} Hz) influence all smaller length scales, hence higher wind speed fluctuation frequencies. Consequently, the largest atmospheric flow scales of hundreds of km represent the correlation length for wind power fluctuations [15], and control the geographic smoothing of wind plants situated distant from each other.

In solar PV however, since half the earth faces the sun at any point in time, all solar photovoltaic plants are globally correlated through time lags controlled by earth's rotation about its axis. As a result, solar PV has much longer (global) correlation length than the relatively local correlation length for wind power (~ 100s of km). Cloud passage over solar PV plants locally disrupts this global correlation by causing a negative ramp in power output. In sharp contrast with wind, solar PV fluctuations are strongly asymmetric about the (mean) clear sky index. Negative power output ramps due to cloud passage contribute large magnitude fluctuations below the clear sky index (leading order effect), while cloud focusing contributes very small magnitude fluctuations above the clear sky index (higher order effect).

The geographic smoothing observed for PV in both CAISO and SCE is more substantial than that that seen in Gujarat [11]. One hypothesis to explain the difference is that if the weather in the Gujarat desert region is consistently clear that there may be fewer fluctuations to smooth than in the southwest USA. We explored this hypothesis by noting that the 10-year average monthly cloud fraction in Gujarat is indeed considerably lower than for the regions in the USA we examined, but only in 8 months of the year. In June and September, the two areas have similar cloud fractions, and in July and August Gujarat is cloudier (supplementary data, section 5). When we performed our analysis dividing the data into one set comprising June through September and another comprising the remaining months, we saw no significant difference in the geographic smoothing. Thus we are forced to reject this hypothesis and simply note that there is considerably less geographic smoothing for utility-scale plants in Gujarat than for the plants in the USA.

Acknowledgements

This research is based on work supported by the Solar Energy Research Institute for India and the U.S. (SERIIUS) funded jointly by the U.S. Department of Energy subcontract DE AC36-08G028308 (Office of Science, Office of Basic Energy Sciences, and Energy Efficiency and Renewable Energy, Solar Energy Technology Program, with support from the Office of International Affairs) and the Government of India subcontract IUSSTF/JCERDC-SERIIUS/2012 dated 22nd November 2012 by National Science Foundation grant no. ACI-1053575. J.A. received partial support from the Carnegie Mellon Climate and Energy Decision

Making Center (CEDM), formed through a cooperative agreement between the National Science Foundation and CMU (SES-0949710). M.M.B. was supported by the Collective Interactions Unit at the Okinawa Institute of Science and Technology Graduate University.

References

[1] Wirth S 2016 Recent Facts about Photovoltaics in Germany, Fraunhofer ISE
[2] Woyte A, Van Thong V, Belmans R, Nijs J 2006 Voltage fluctuations on distribution level introduced by photovoltaic systems *IEEE Transactions on Energy Conversion* 21(1) 202–209
[3] Paatero JV, Lund PD 2007 Effects of large-scale photovoltaic power integration on electricity distribution networks *Renewable Energy* 32(2) 216–234
[4] Wiemken E,Beyer HG, Heydenreich W and Kiefer K 2001 Power characteristics of PV ensembles: experiences from the combined power production of 100 grid connected PV systems distributed over the area of Germany *Solar Energy* 70 513-518
[5] Murata A, Yamaguchi H and Otani K 2009 A method of estimating the output fluctuation of many photovoltaic power generation systems dispersed in a wide area *Electrical Engineering in Japan* 166(4) 645-652
[6] Mills A, Ahlstrom M, Brower M, Ellis A, George R, Hoff T, Kroposki B, Lenox C, Miller N, Milligan M, Stein J, and Wan Y 2011 Dark shadows *IEEE Power and Energy Magazine* 9(3) 33–41.
[7] Marcos J, Marroyo L, Lorenzo E and Garcia M 2012 Smoothing of PV power fluctuations by geographical dispersion *Progress in Photovoltaics: Research and Applications* 20 226-237
[8] Lave M, Stein JS and Ellis A 2012 Analyzing and simulating the reduction in PV powerplant variability due to geographic smoothing in Ota City, Japan and Alamosa, CO *IEEE Photovoltaic Specialists Conference (PVSC)*, 2
[9] van Haaren R, Markaria M and Fthenakis V 2012 Empirical assessment of short-term variability from utility-scale solar PV plants *Progress in Photovoltaics: Research and Applications* 22(5) 548-559
[10] Curtright A and Apt J 2007 The character of power output from utility-scale photovoltaic systems *Progress in Photovoltaics: Research and Applications* 16 241-247
[11] Klima K and Apt J 2015 Geographic smoothing of solar PV: Results from Gujarat *Environmental Res. Lett.* 10 104001
[12] Katzenstein W, Fertig E and Apt J 2010 The Variability of Interconnected Wind Plants *Energy Policy* 38(8) 4400-4410
[13] Lomb NR 1976 Least-squares frequency analysis of unevenly spaced data *Astrophys. Space Sci.* 39 447–62
[14] Press WH, Teukolsky SA, Vetterling WT and Flannery BP 1992 Numerical Recipes in FORTRAN: The Art of Scientific Computing 2nd ed (Cambridge: Cambridge University Press)
[15] Bandi MM 2017 Spectrum of Wind Power Fluctuations *Phys. Rev. Lett.* 118, 028301

Geographic smoothing of solar photovoltaic electric power production in the western USA

Supplementary information

Kelly Klima[1], Jay Apt[1, 2], Mahesh Bandi[3], Paul Happy[4], Clyde Loutan[5], and Russell Young[4]

[1] Department of Engineering and Public Policy, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
[2] Tepper School of Business, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
[3] Collective Interactions Unit, OIST Graduate University, Onna, Okinawa, 9040495 Japan
[4] SoCore Energy, 225 W. Hubbard, Suite 200, Chicago, IL 60654
[5] California Independent Systems Operator (CAISO), 250 Outcropping Way, Folsom, CA 95630

E-mail: apt@cmu.edu

Table of Contents:

1. Commercial rooftop SCE data: Location and size of the plants 2
2. Commercial rooftop SCE Data: Power spectral densities calculated using the Lomb periodiogram 3
3. Utility-scale CAISO data: Location and size of the plants 23
4. Utility-scale CAISO data: Power spectral densities calculated using the Lomb periodiogram 24
5. Gujarat and California cloud regimes 40

1. Commercial rooftop SCE data: Location and size of the plants

Plant name, location, and generation data (15 minute intervals) for 19 small-scale photovoltaics were provided by SoCore Energy for April 1, 2015 to March 31, 2016. Table S1 shows the location and installed capacity of the plants examined in the Main Text. Plant name and location were provided by SoCore Energy. Locations have been rounded to the nearest tenth of a degree. In addition, we provide the maximum generation observed in the provided dataset.

Table S1: Location and maximum generation of the plants for the commercial rooftop plants. Please see Figure 1 in the main text for a map of these locations.

Site	Latitude	Longitude	Maximum generation (kW)
Plant 1	38.3	-122.3	96.76
Plant 2	39.9	-121.8	240.87
Plant 3	34.1	-118.3	86.81
Plant 4	33.8	-117.8	338.00
Plant 5	34.0	-117.7	337.50
Plant 6	33.6	-117.7	287.72
Plant 7	33.5	-117.1	363.10
Plant 8	33.9	-117.7	313.21
Plant 9	33.7	-117.7	436.64
Plant 10	34.1	-118.1	409.75
Plant 11	33.9	-118.0	410.44
Plant 12	32.8	-117.0	385.11
Plant 13	34.3	-118.9	388.16
Plant 14	36.3	-119.3	411.83
Plant 15	32.6	-117.0	382.64
Plant 16	33.1	-117.2	363.19
Plant 17	36.4	-119.3	458.15
Plant 18	33.7	-117.2	514.99
Plant 19	38.0	-122.5	257.74

2. Commercial rooftop SCE Data: Power spectral densities calculated using the Lomb periodiogram

This section contains the power spectral densities (PSDs) for the 19 plants for the full year via the Lomb periodiogram and 32 time segments. Since the time resolution of the data is 15 minutes, the Nyquist frequency is 5.6×10^{-4} Hz, and the plots have that as the highest frequency. The lowest frequency is determined by the 1-year length of the data and the 8 segment averaging, yielding 2.53×10^{-7} Hz.

Table S2: Summary table of slopes in the log domain at high frequency of the commercial rooftop plants.

Site	Slope
Plant 1	-1.81
Plant 2	-1.78
Plant 3	-1.79
Plant 4	-1.79
Plant 5	-1.78
Plant 6	-1.78
Plant 7	-1.77
Plant 8	-1.78
Plant 9	-1.79
Plant 10	-1.91
Plant 11	-1.78
Plant 12	-1.78
Plant 13	-1.78
Plant 14	-1.80
Plant 15	-1.79
Plant 16	-1.81
Plant 17	-1.79
Plant 18	-1.78
Plant 19	-1.81

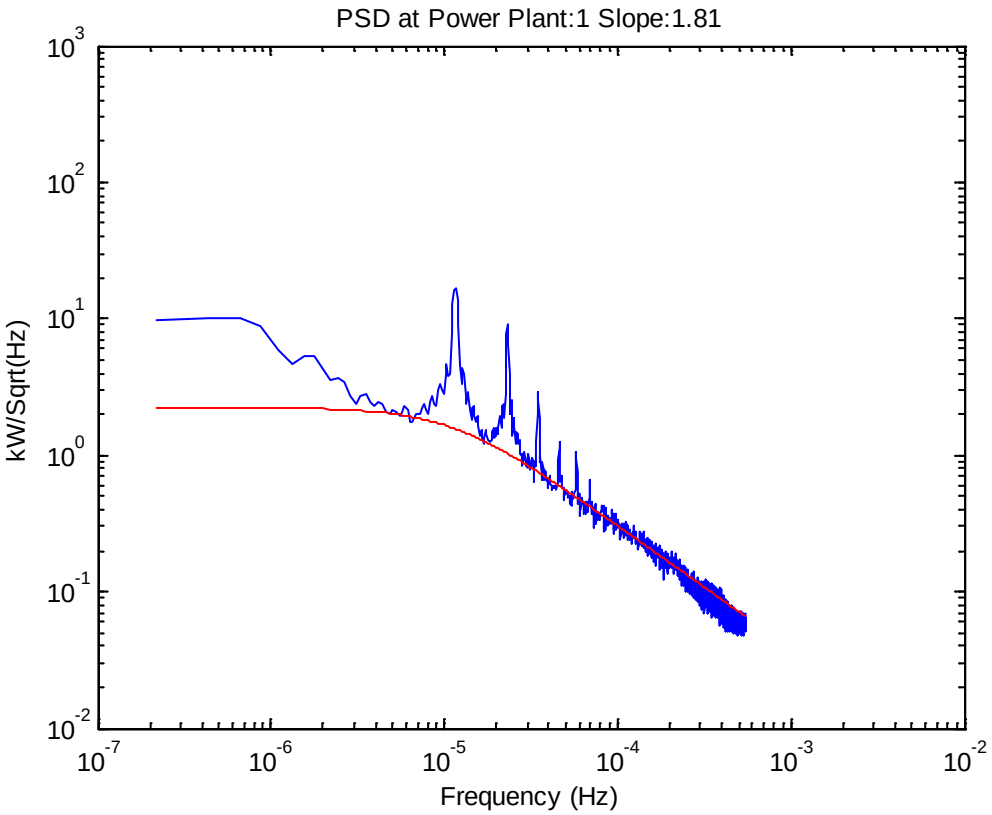

Figure S1: PSD (blue) with line of best fit (red) for Plant 1 (A0001). The slope at high frequencies in the loglog domain (*m*) is -1.81.

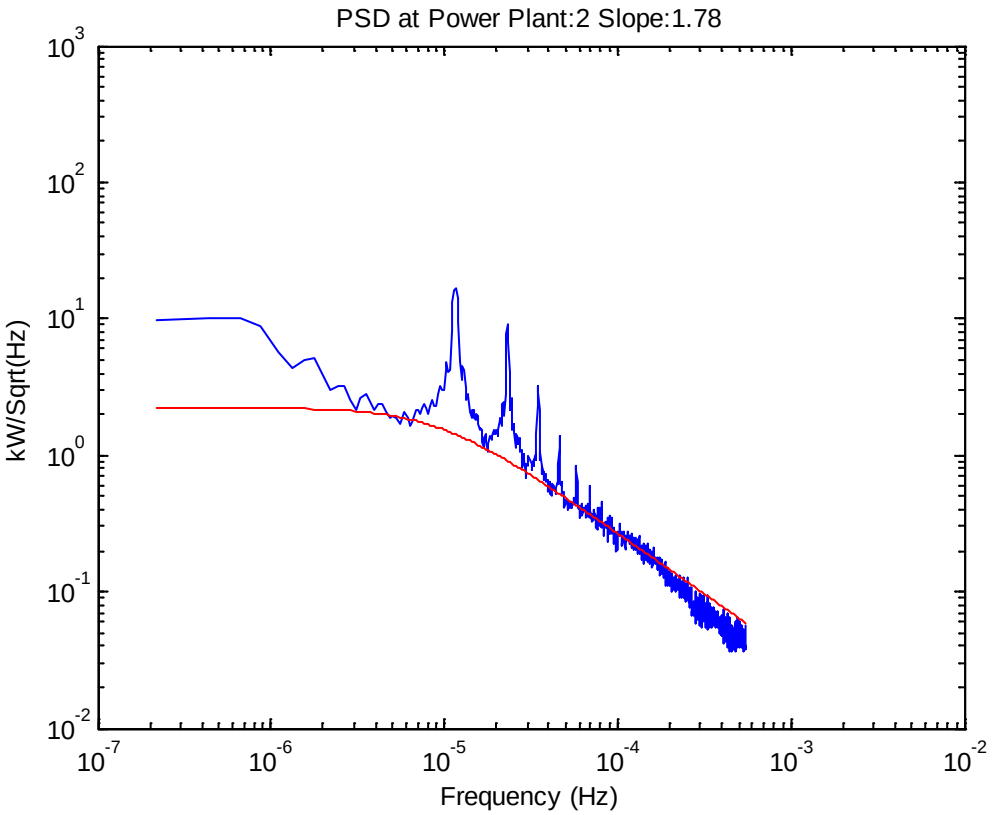

Figure S2: PSD (blue) with line of best fit (red) for Plant 2 (A0011). The slope at high frequencies in the loglog domain (*m*) is -1.78.

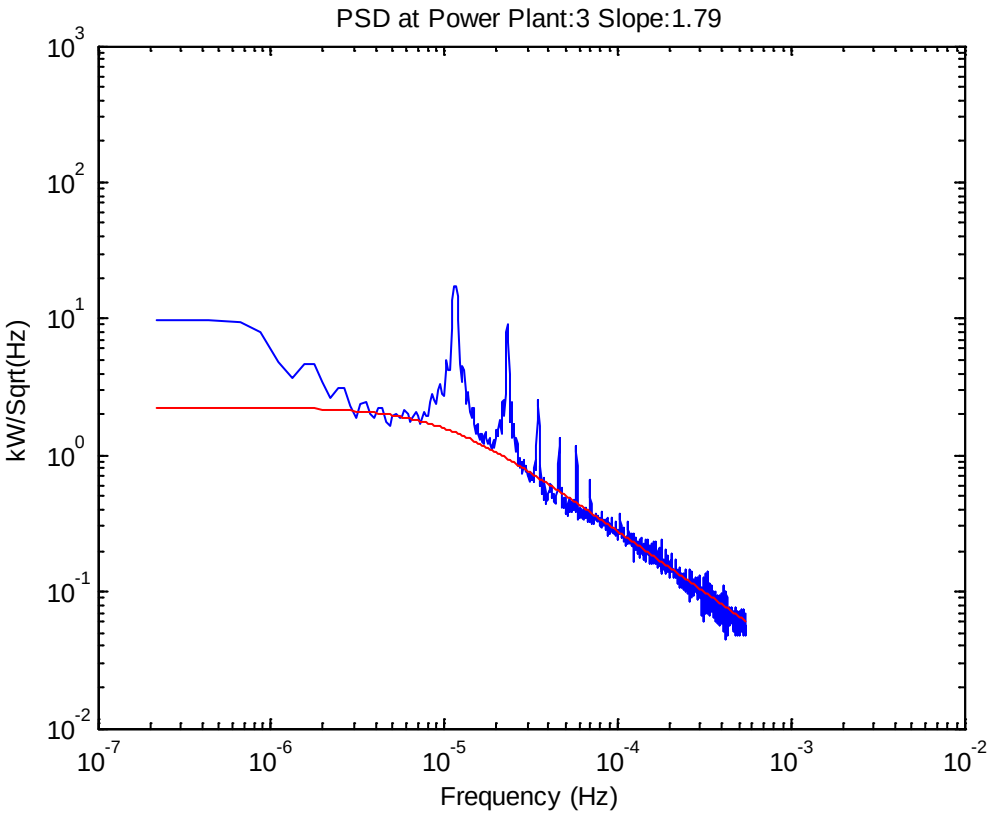

Figure S3: PSD (blue) with line of best fit (red) for Plant 3 (A0024). The slope at high frequencies in the loglog domain (*m*) is -1.79.

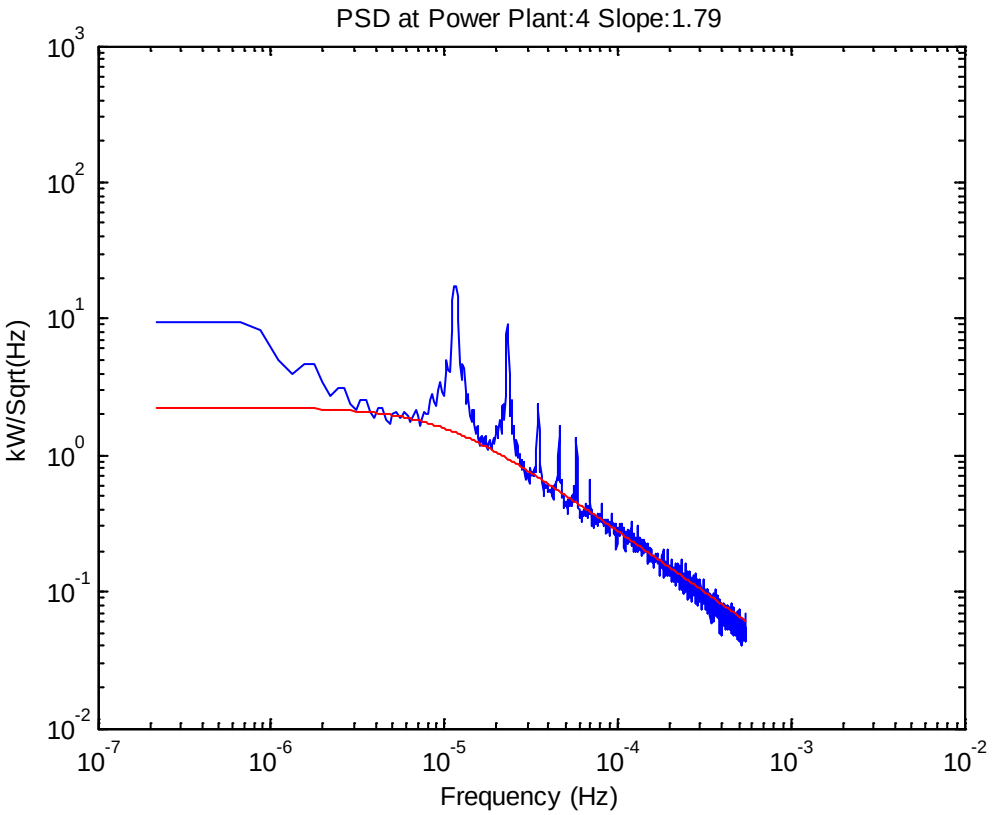

Figure S4: PSD (blue) with line of best fit (red) for Plant 4 (A0056). The slope at high frequencies in the loglog domain (*m*) is -1.79.

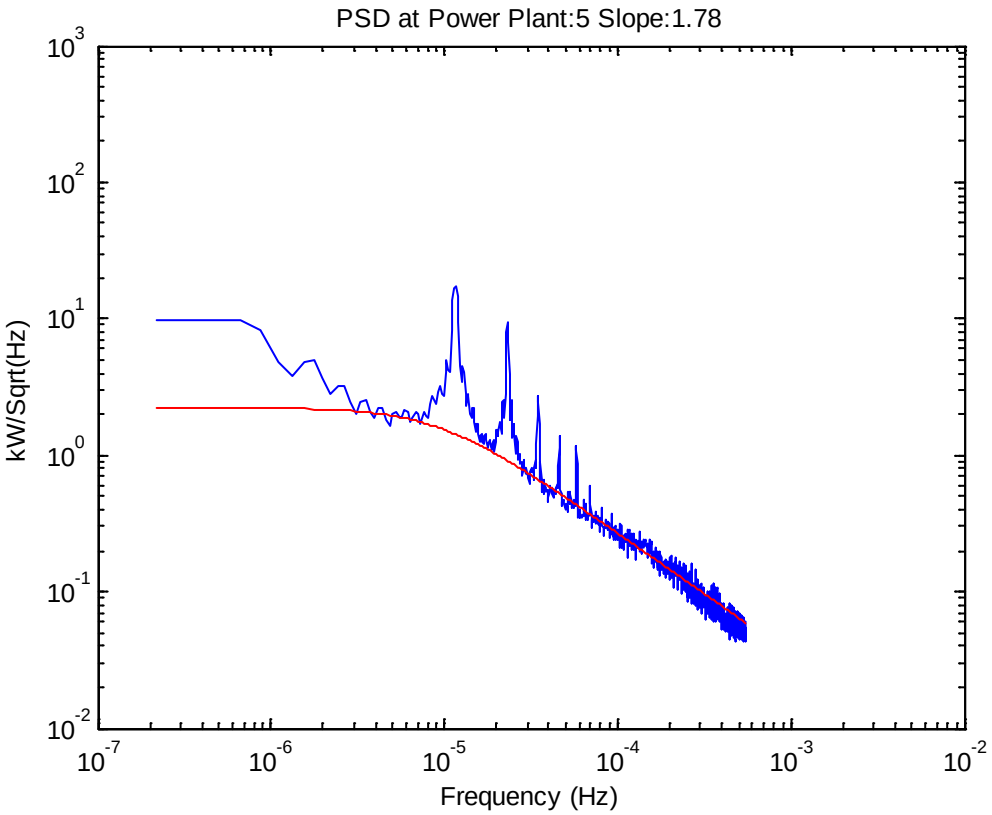

Figure S5: PSD (blue) with line of best fit (red) for Plant 5 (A0058). The slope at high frequencies in the loglog domain (*m*) is -1.78.

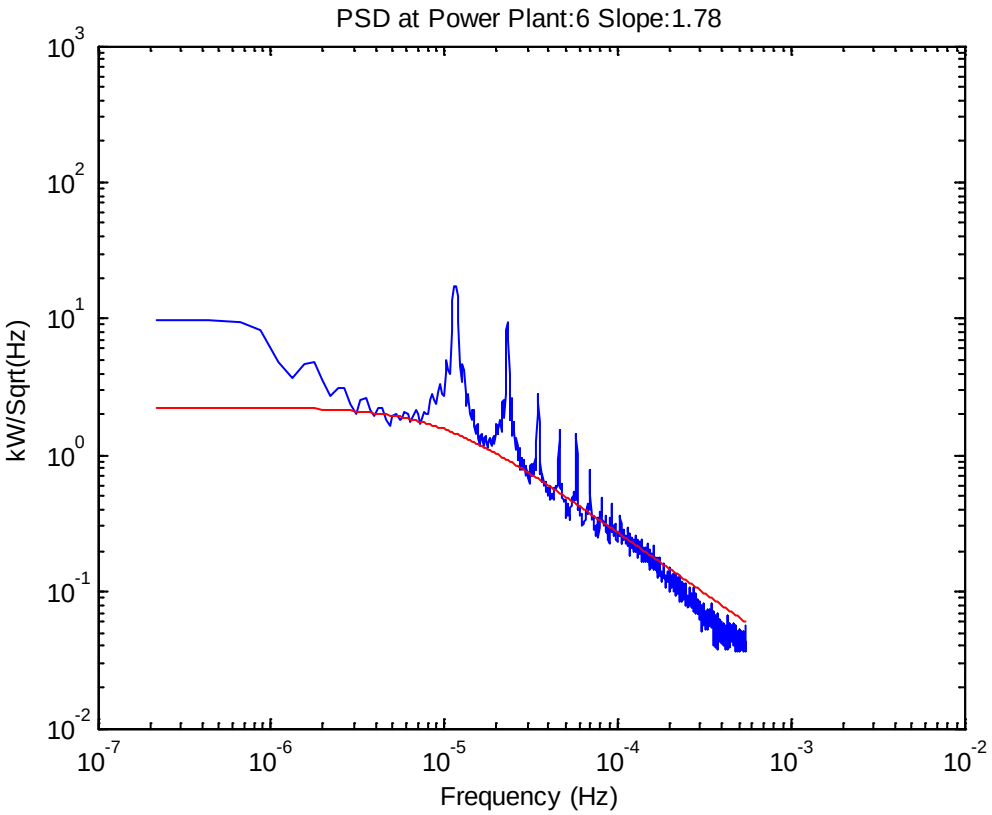

Figure S6: PSD (blue) with line of best fit (red) for Plant 6 (A0059). The slope at high frequencies in the loglog domain (*m*) is -1.78.

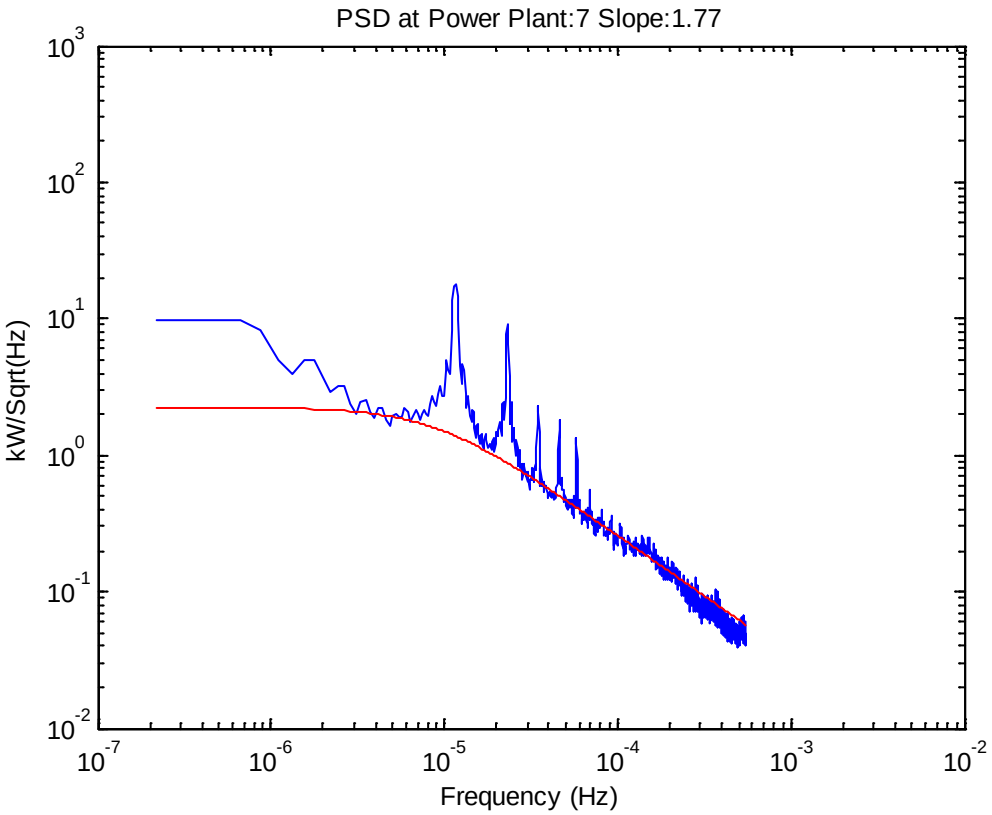

Figure S7: PSD (blue) with line of best fit (red) for Plant 7 (A0061). The slope at high frequencies in the loglog domain (*m*) is -1.77.

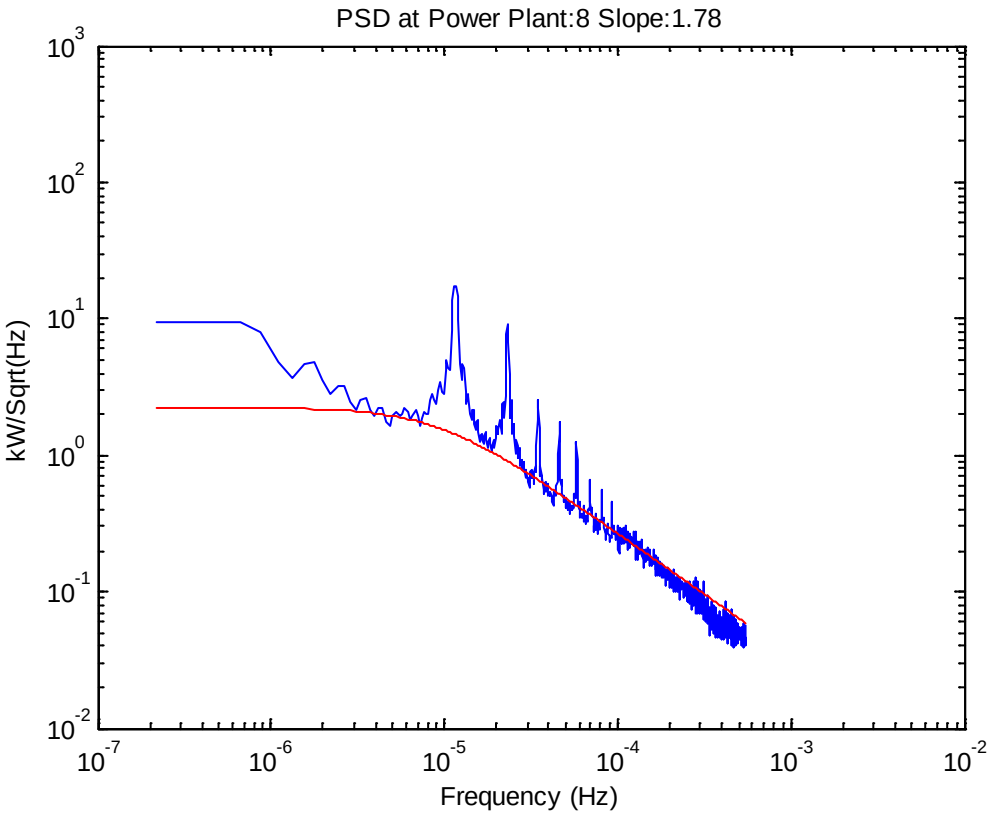

Figure S8: PSD (blue) with line of best fit (red) for Plant 8 (A0062). The slope at high frequencies in the loglog domain (*m*) is -1.78.

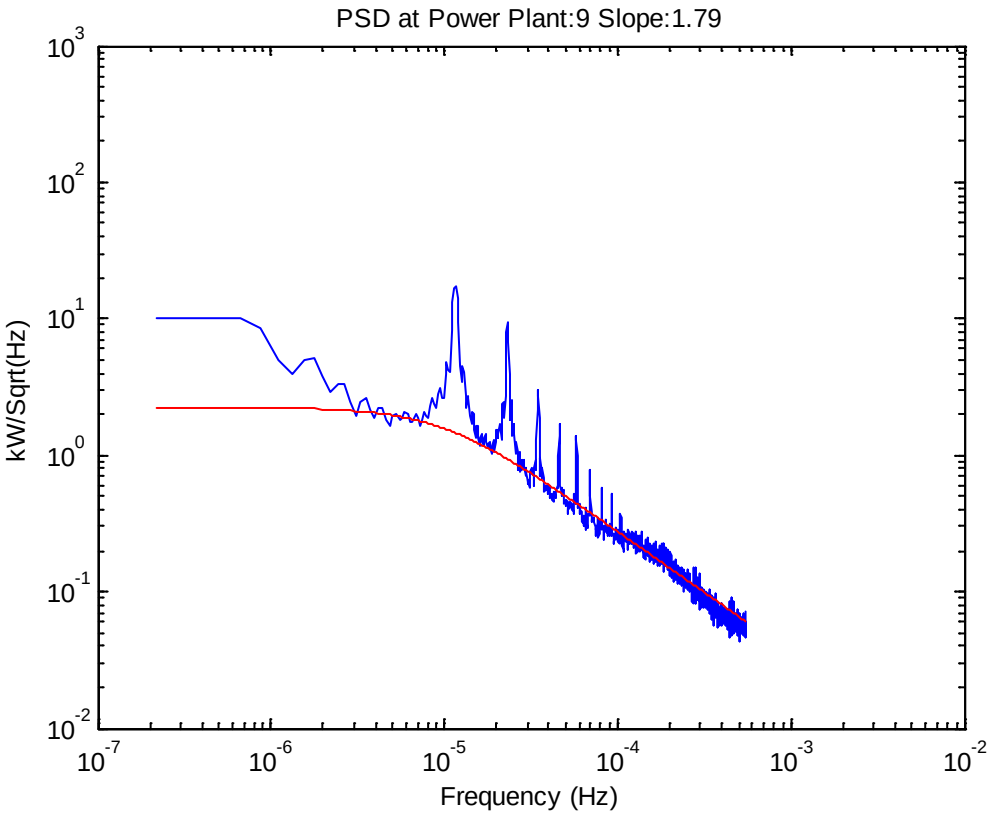

Figure S9: PSD (blue) with line of best fit (red) for Plant 9 (A0067). The slope at high frequencies in the loglog domain (*m*) is -1.79.

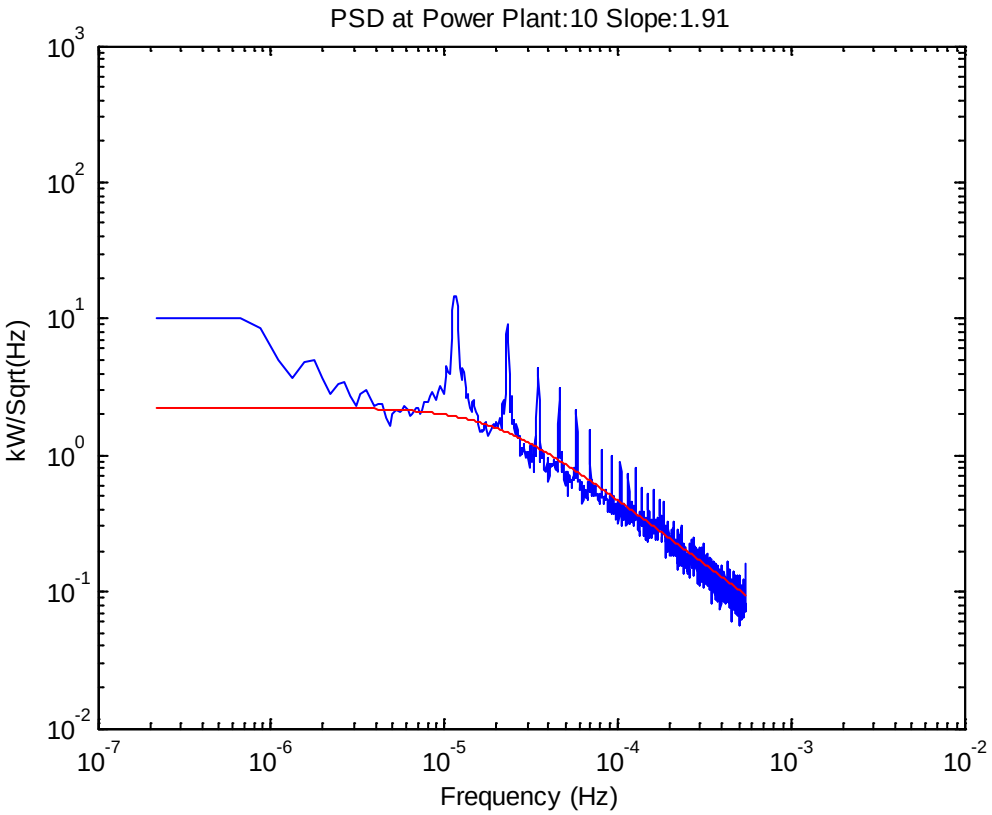

Figure S10: PSD (blue) with line of best fit (red) for Plant 10 (A0089). The slope at high frequencies in the loglog domain (*m*) is -1.91.

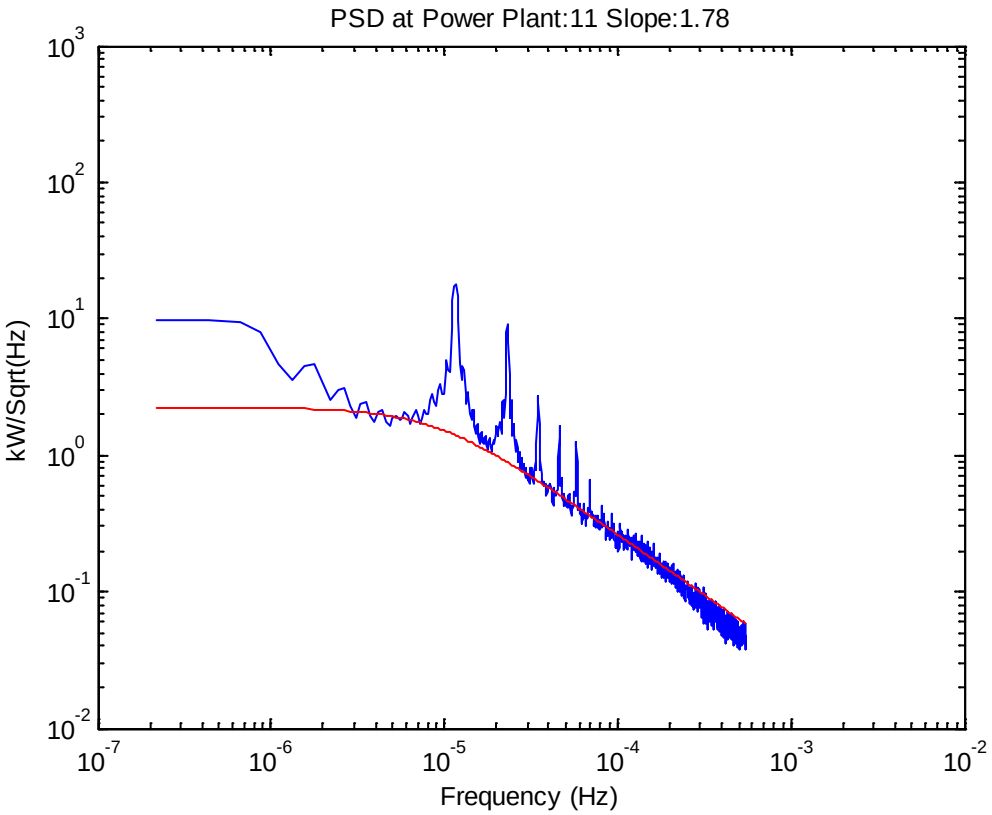

Figure S11: PSD (blue) with line of best fit (red) for Plant 11 (A0090). The slope at high frequencies in the loglog domain (*m*) is -1.78.

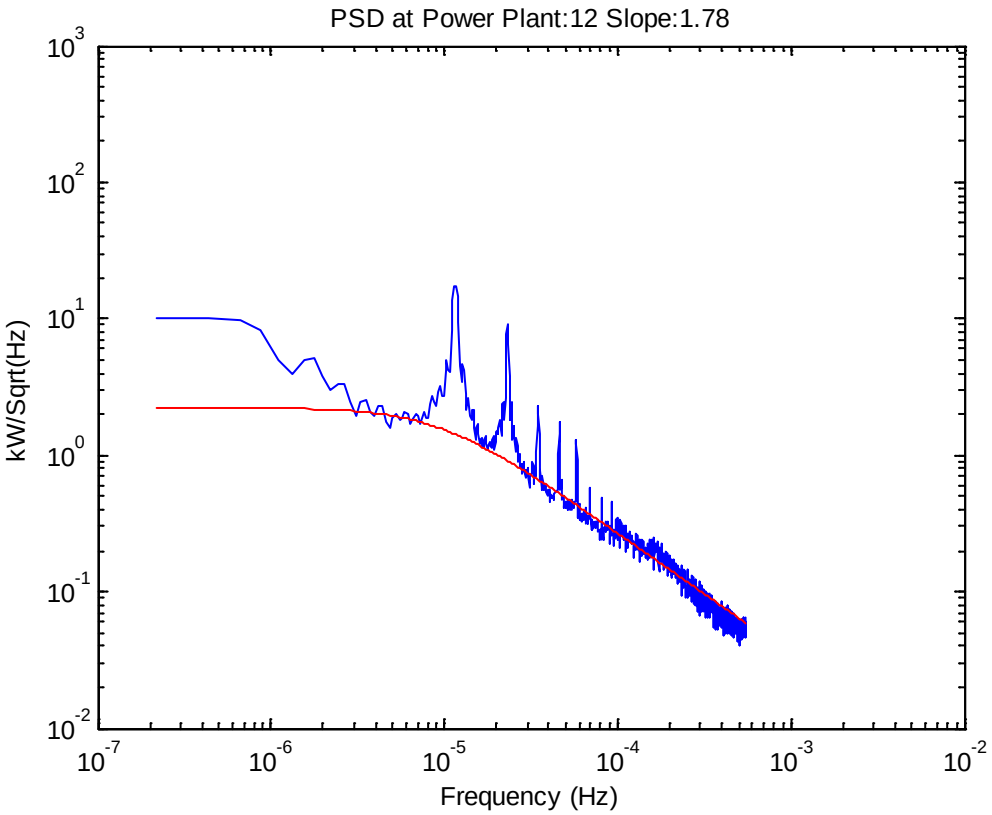

Figure S12: PSD (blue) with line of best fit (red) for Plant 12 (A0092). The slope at high frequencies in the loglog domain (*m*) is -1.78.

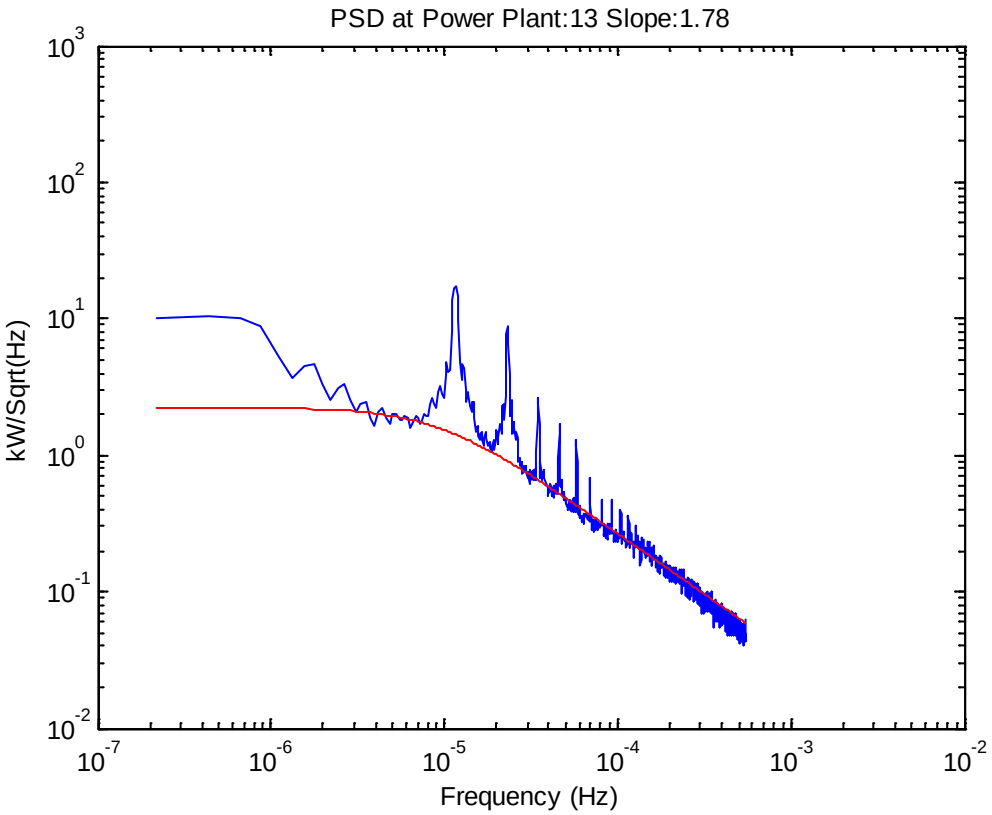

Figure S13: PSD (blue) with line of best fit (red) for Plant 13 (A0095). The slope at high frequencies in the loglog domain (*m*) is -1.78.

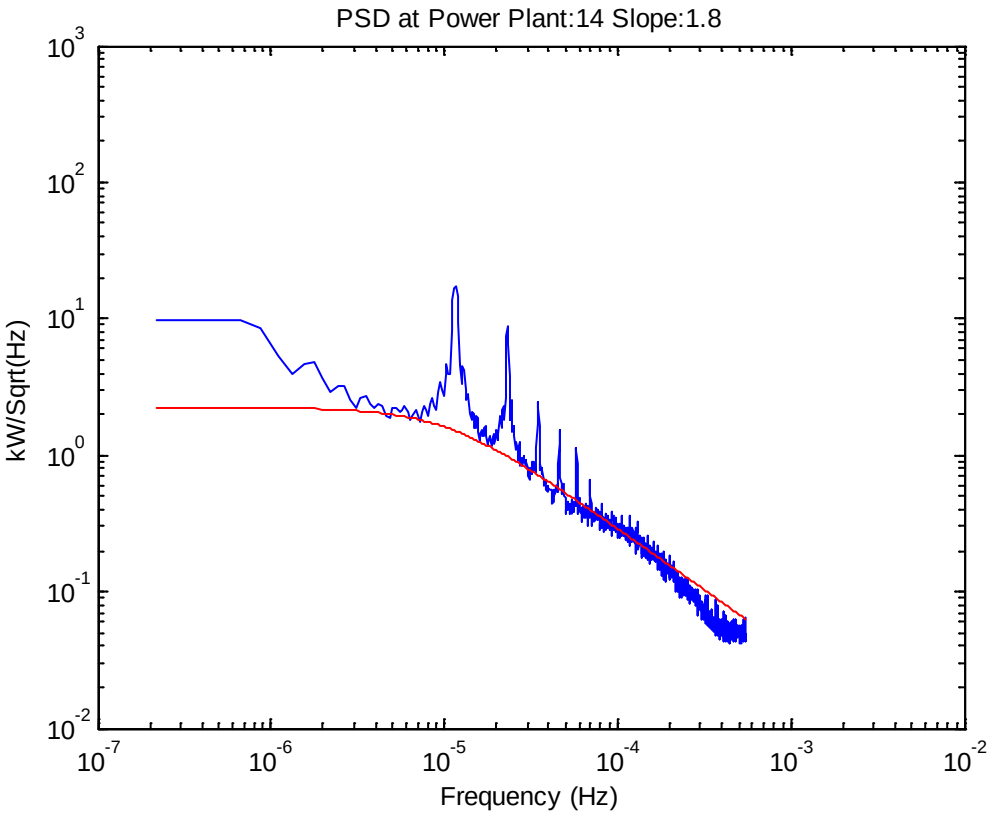

Figure S14: PSD (blue) with line of best fit (red) for Plant 14 (A0098). The slope at high frequencies in the loglog domain (*m*) is -1.8.

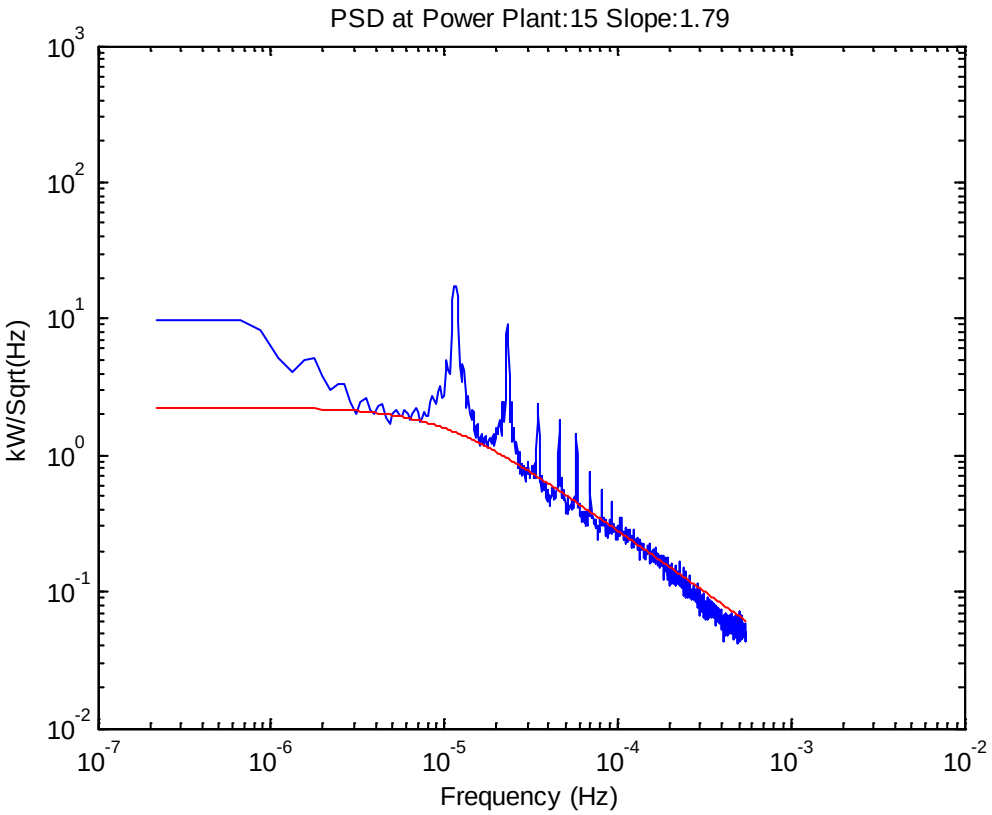

Figure S15: PSD (blue) with line of best fit (red) for Plant 15 (A0099). The slope at high frequencies in the loglog domain (*m*) is -1.79.

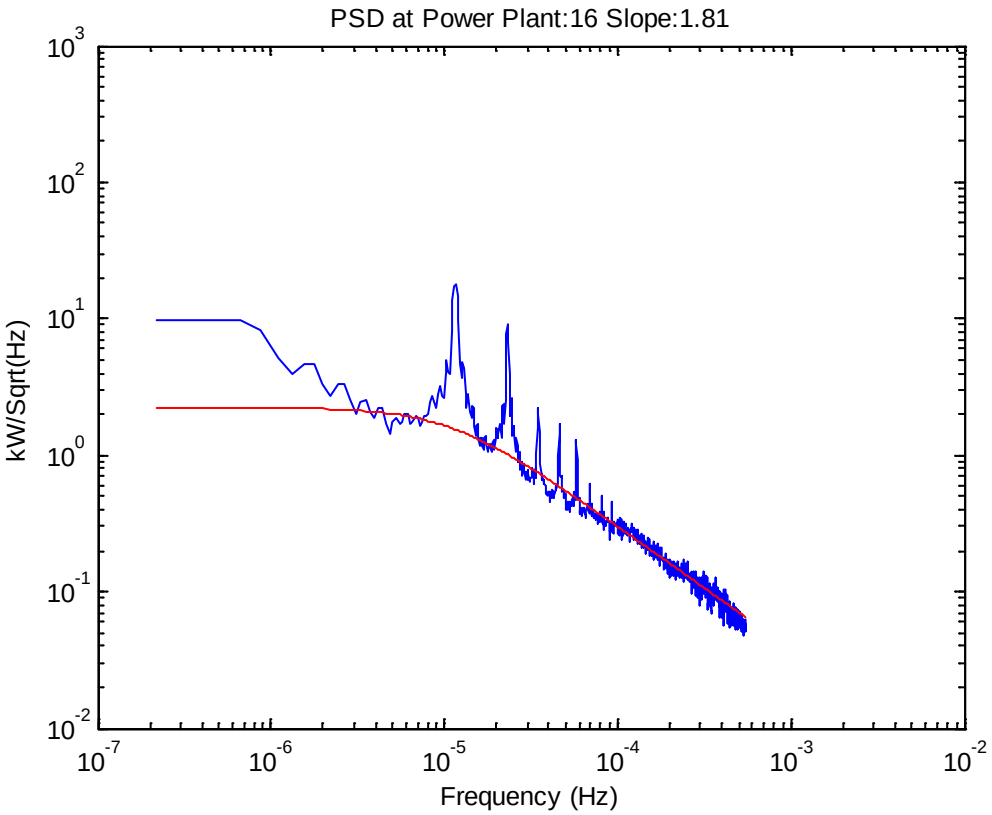

Figure S16: PSD (blue) with line of best fit (red) for Plant 16 (A0104). The slope at high frequencies in the loglog domain (*m*) is -1.81.

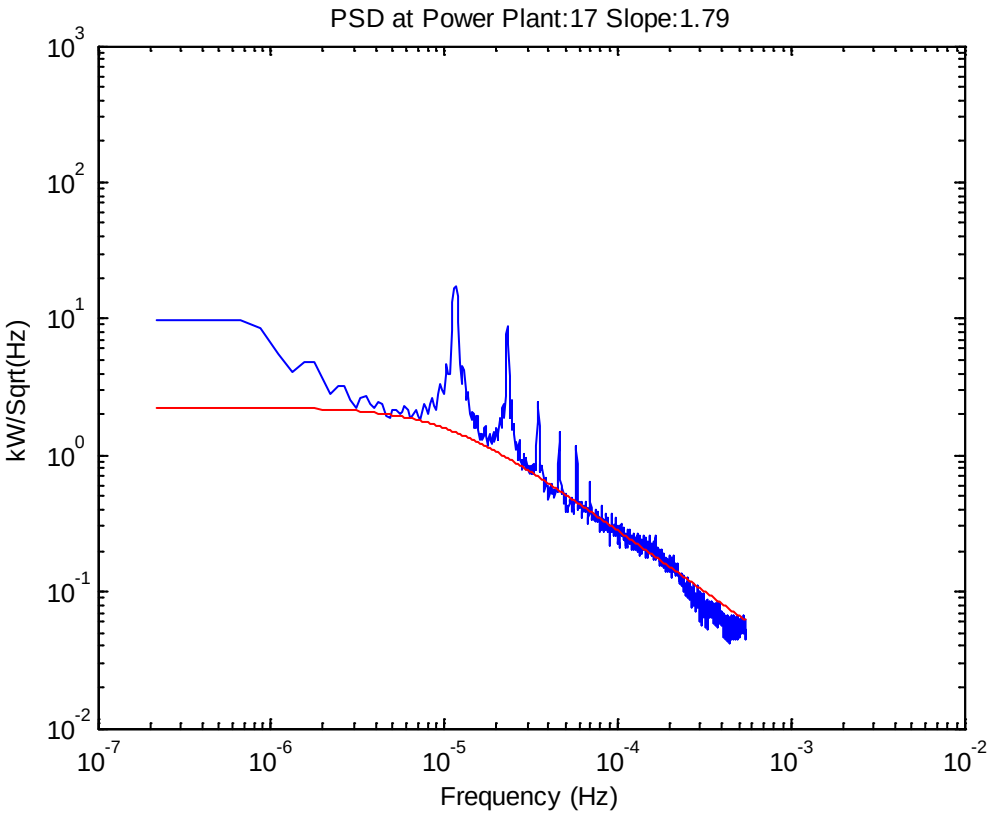

Figure S17: PSD (blue) with line of best fit (red) for Plant 17 (A0113). The slope at high frequencies in the loglog domain (*m*) is -1.79.

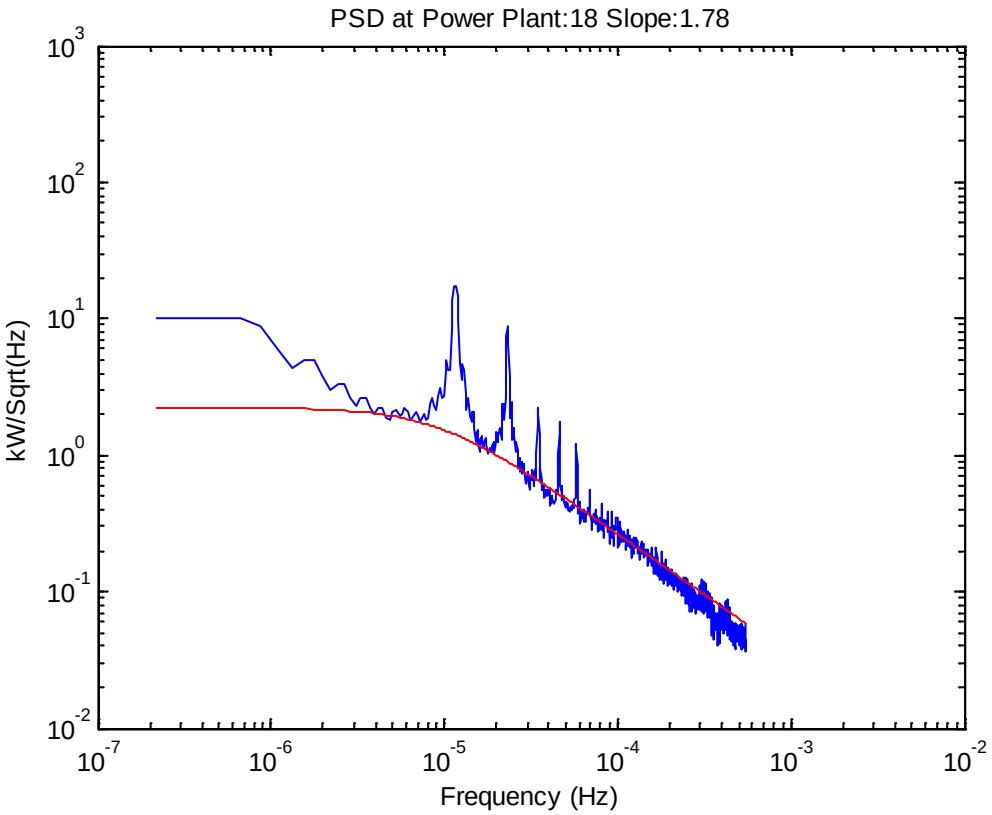

Figure S18: PSD (blue) with line of best fit (red) for Plant 18 (A0114). The slope at high frequencies in the loglog domain (*m*) is -1.78.

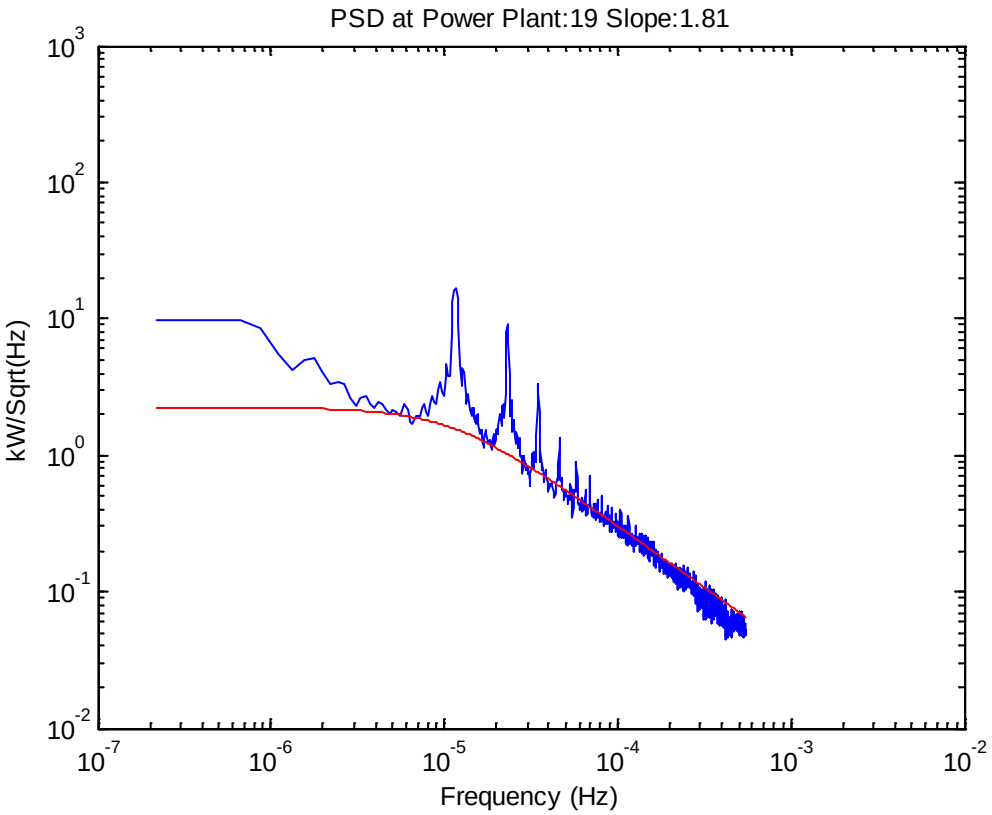

Figure S19: PSD (blue) with line of best fit (red) for Plant 19 (A0122). The slope at high frequencies in the loglog domain (*m*) is -1.81.

3. Utility-scale data: Location and size of the plants

Plant name, location, and generation data (1 minute intervals) for 15 commercial scale photovoltaic plants were provided by CAISO for July 1, 2015 to June 30, 2016. Per agreement with the data provider, the latitude and longitude of these plants has been rounded to the nearest degree, and the capacity of these plants has been rounded to the nearest 10 MW.

Table S3. Location and maximum generation of the utility-scale plants. Please see Figure 1 in the main text for a map of these locations.

Plant	Latitude	Longitude	Size (MW)
Plant 1	34	-115	300
Plant 2	33	-113	300
Plant 3	35	-118	280
Plant 4	34	-115	260
Plant 5	34	-115	250
Plant 6	35	-118	240
Plant 7	35	-120	210
Plant 8	33	-114	200
Plant 9	36	-115	250
Plant 10	36	-115	160
Plant 11	33	-113	170
Plant 12	33	-116	150
Plant 13	33	-116	140
Plant 14	33	-116	130
Plant 15	33	-113	130

4. CAISO Utility-scale data: Power spectral densities calculated using the Lomb periodiogram

This section contains the power spectral densities (PSDs) for the 15 plants for the full year via the Lomb periodiogram as for the commercial rooftop data in section 2. Although the data are at 1 minute, these plots have been cut off at the same high frequency limit as those for the rooftop plants in section 2.

Table S4: Summary table of slopes in the log domain at high frequency of the CAISO utility scale plants.

Plant	Slope
Plant 1	-1.67
Plant 2	-1.60
Plant 3	-1.69
Plant 4	-1.70
Plant 5	-1.60
Plant 6	-1.62
Plant 7	-1.68
Plant 8	-1.73
Plant 9	-1.69
Plant 10	-1.63
Plant 11	-1.63
Plant 12	-1.74
Plant 13	-1.59
Plant 14	-1.67
Plant 15	-1.74

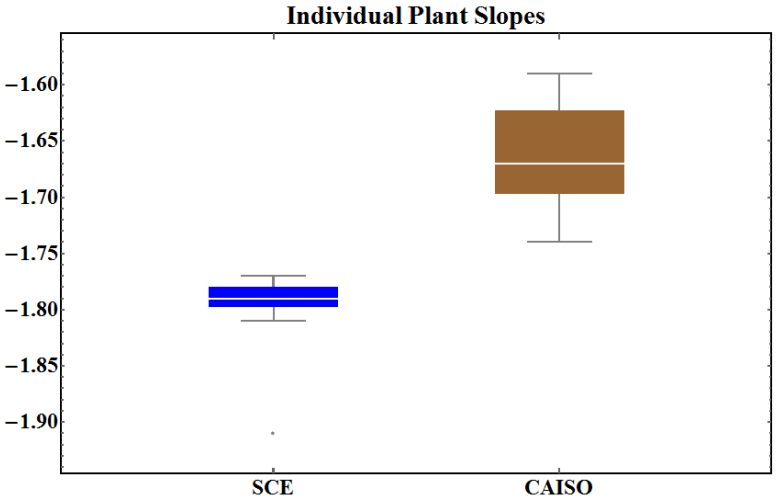

Figure S20: Box-and-whisker plot comparison of PSD slopes of the 19 commercial rooftop SCE plants to those of the 15 utility-scale CAISO plants.

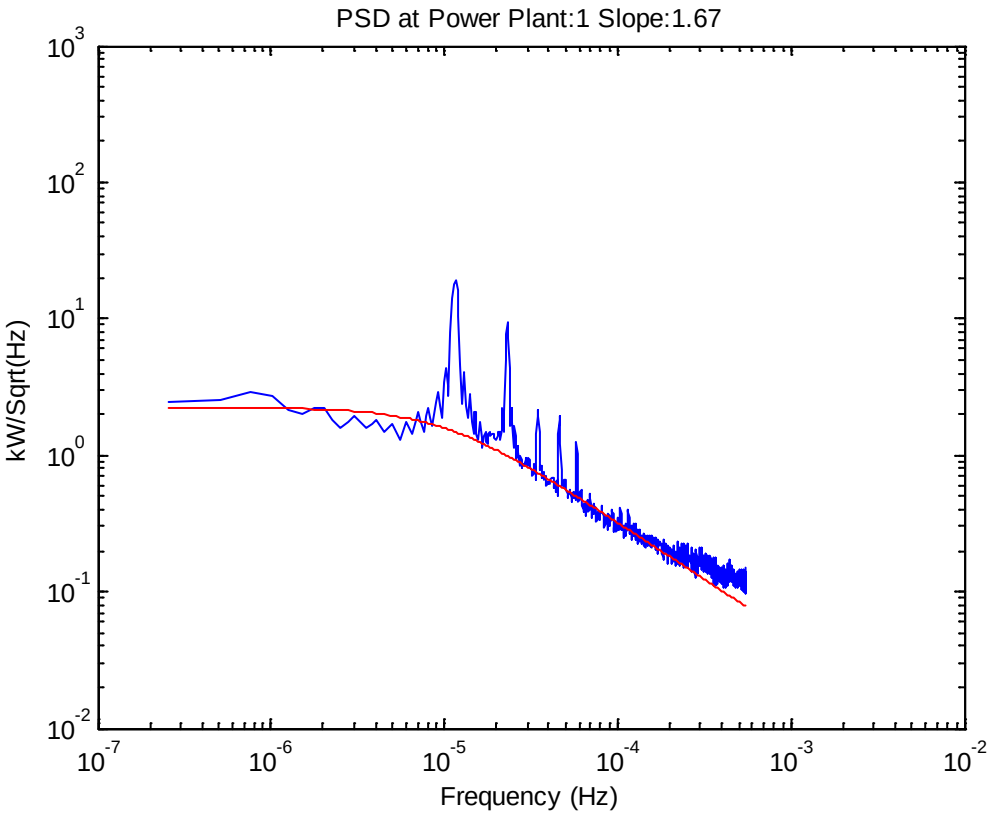

Figure S21: PSD (blue) with line of best fit (red) for Plant 1. The slope at high frequencies in the loglog domain (*m*) is -1.67.

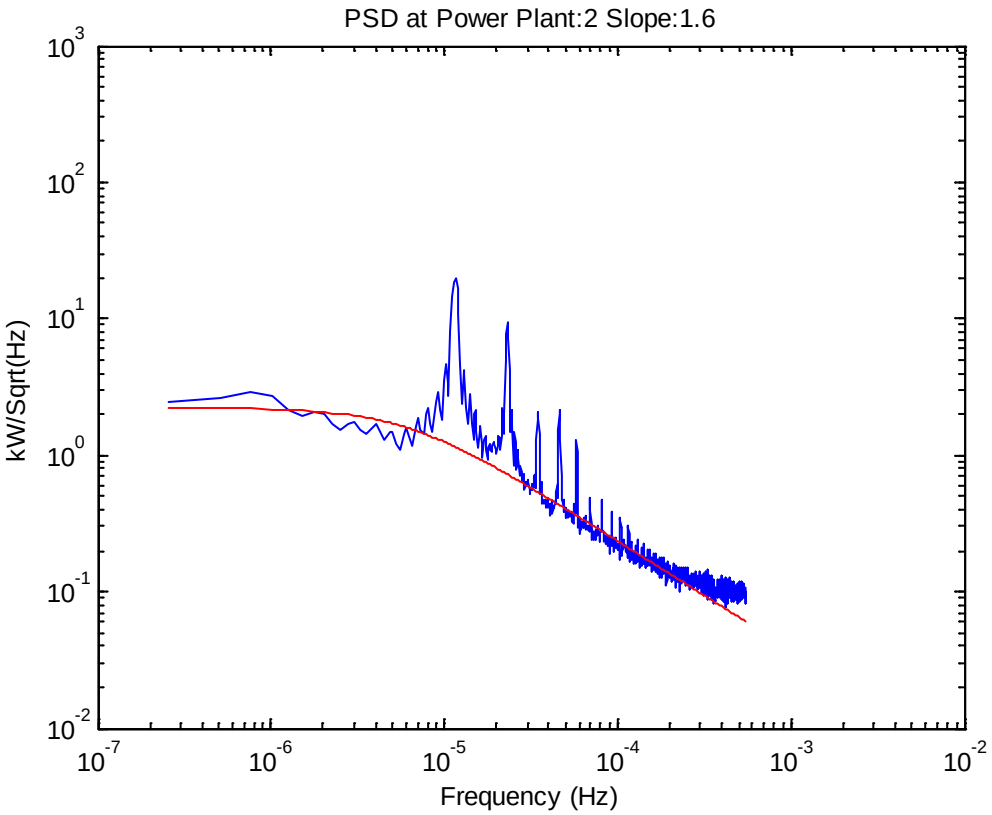

Figure S22: PSD (blue) with line of best fit (red) for Plant 2. The slope at high frequencies in the loglog domain (*m*) is -1.6.

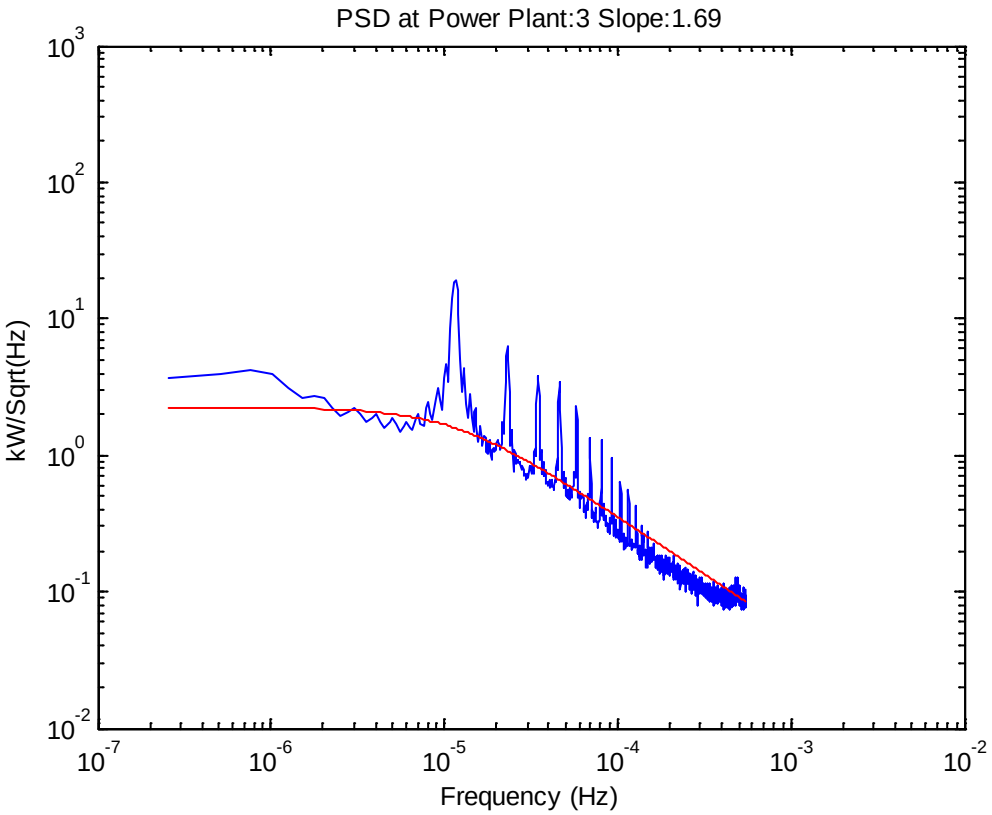

Figure S23: PSD (blue) with line of best fit (red) for Plant 3. The slope at high frequencies in the loglog domain (*m*) is -1.69.

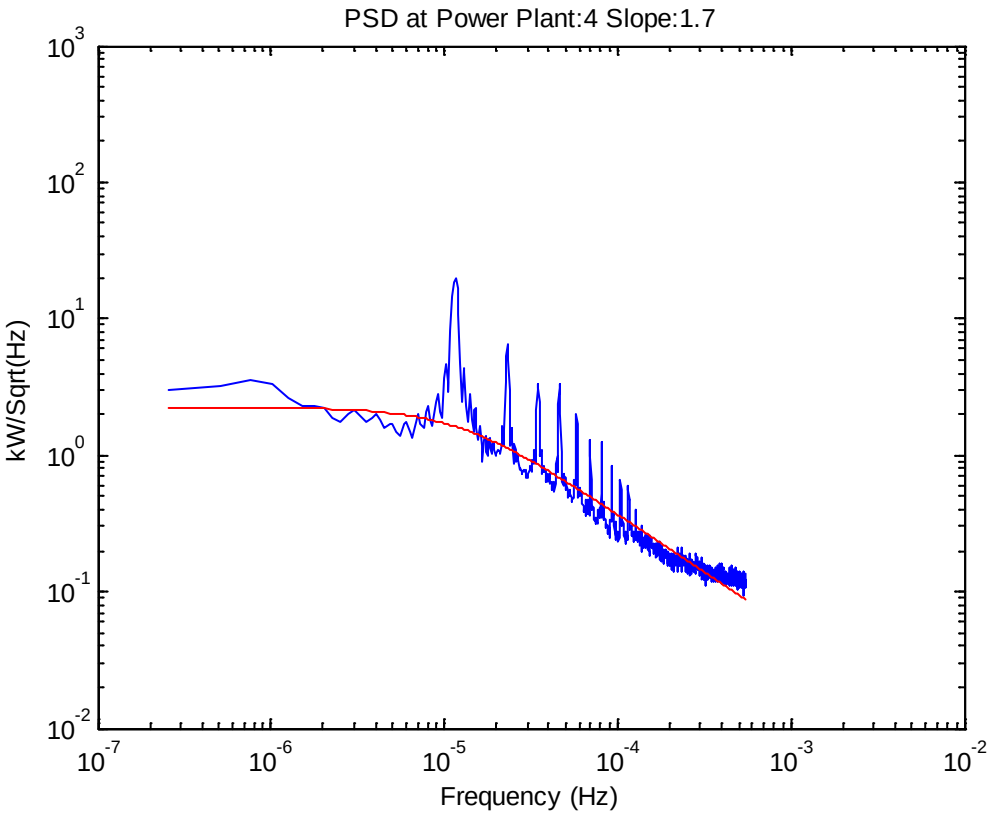

Figure S24: PSD (blue) with line of best fit (red) for Plant 4. The slope at high frequencies in the loglog domain (*m*) is -1.7.

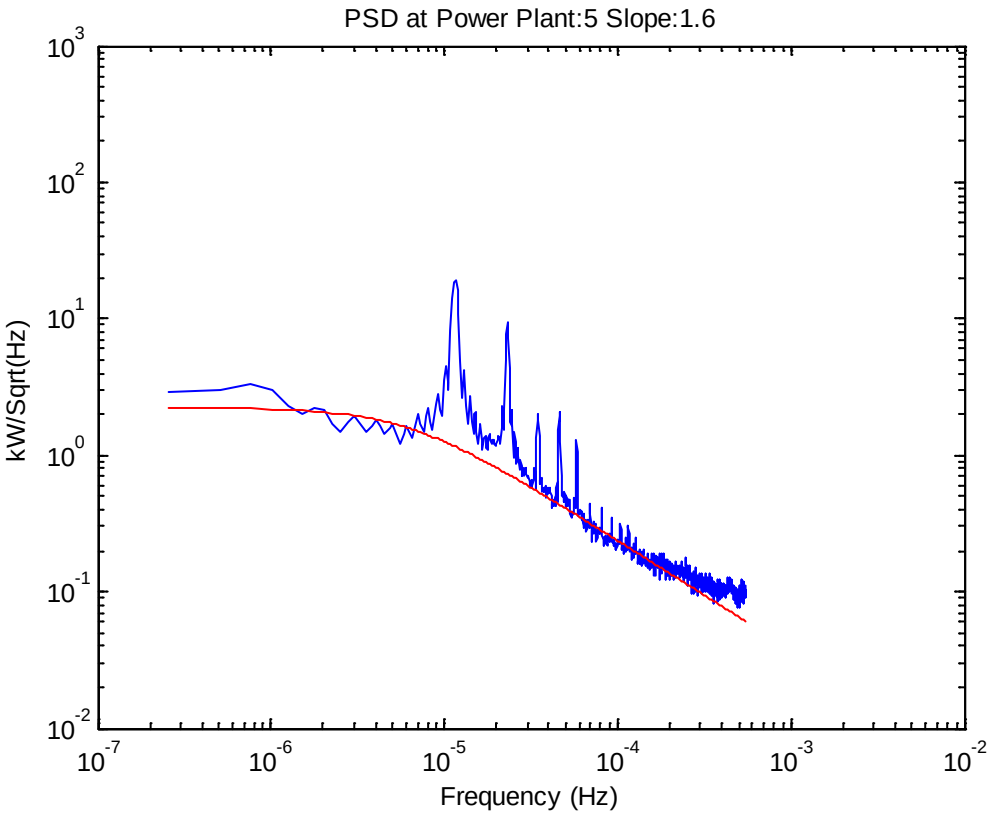

Figure S25: PSD (blue) with line of best fit (red) for Plant 5. The slope at high frequencies in the loglog domain (*m*) is -1.6.

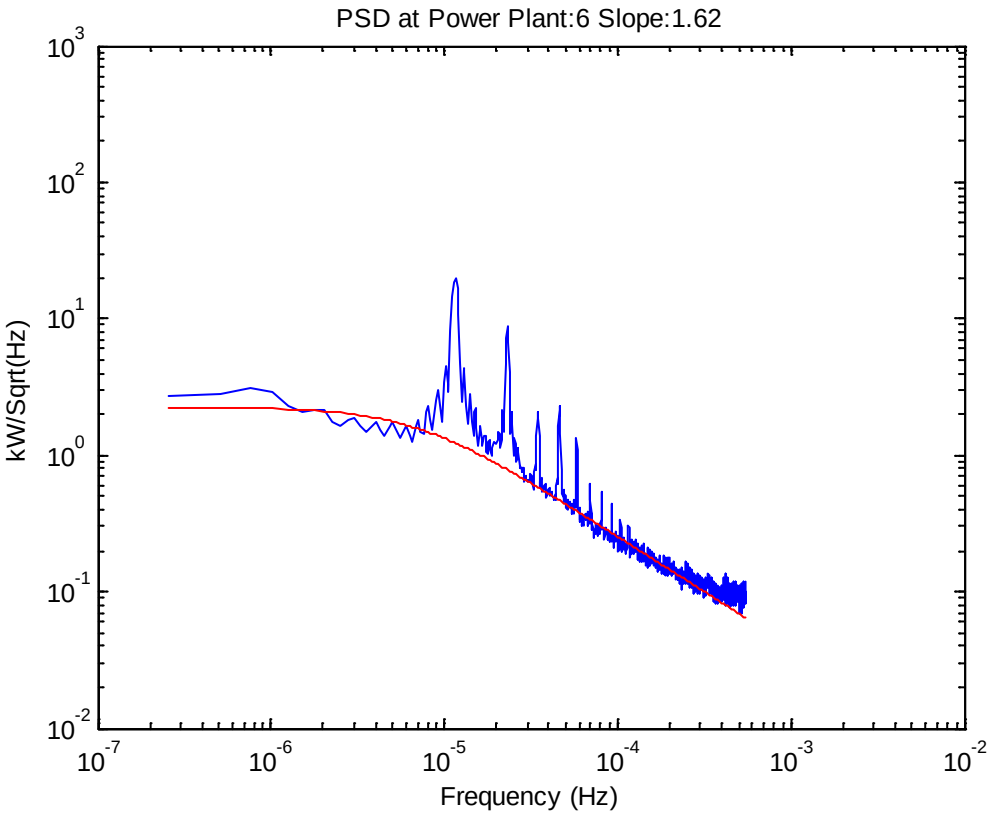

Figure S26: PSD (blue) with line of best fit (red) for Plant 6. The slope at high frequencies in the loglog domain (*m*) is -1.62.

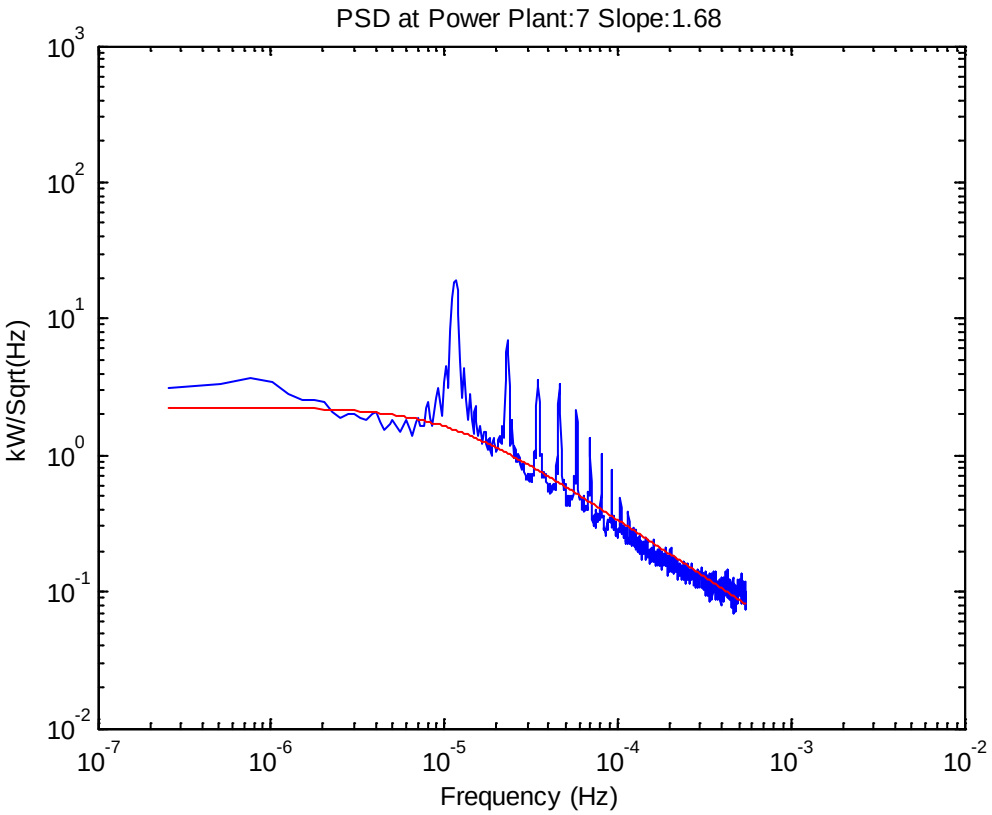

Figure S27: PSD (blue) with line of best fit (red) for Plant 7. The slope at high frequencies in the loglog domain (*m*) is -1.68.

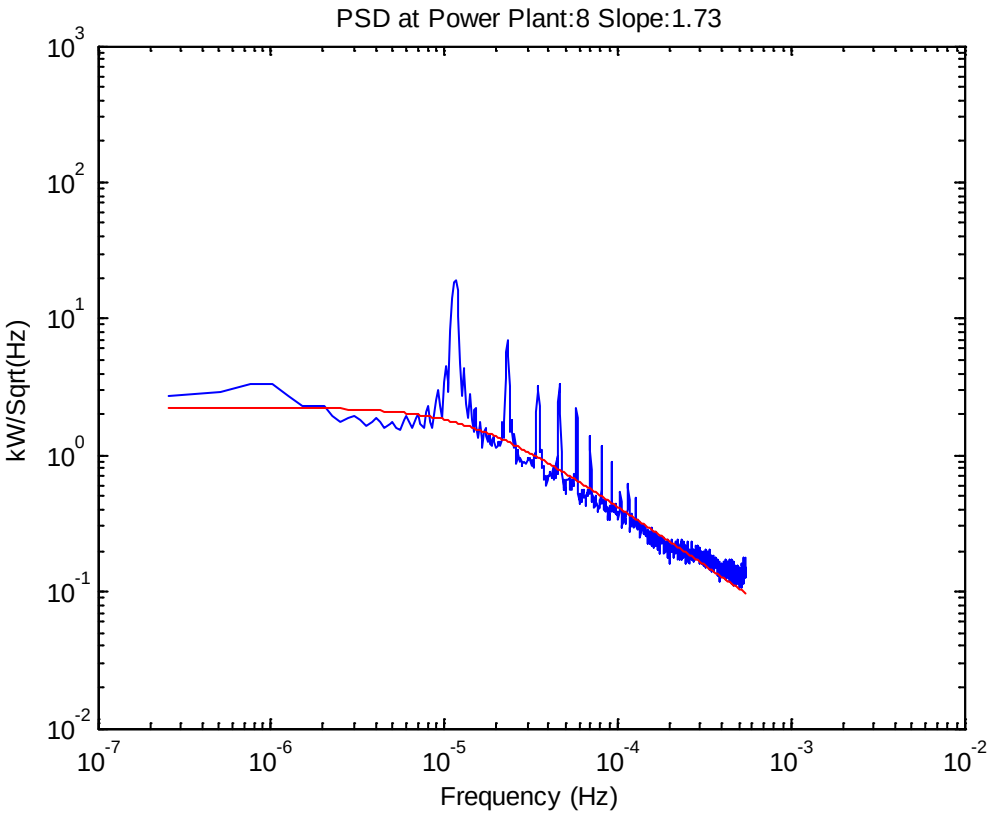

Figure S28: PSD (blue) with line of best fit (red) for Plant 8. The slope at high frequencies in the loglog domain (*m*) is -1.73.

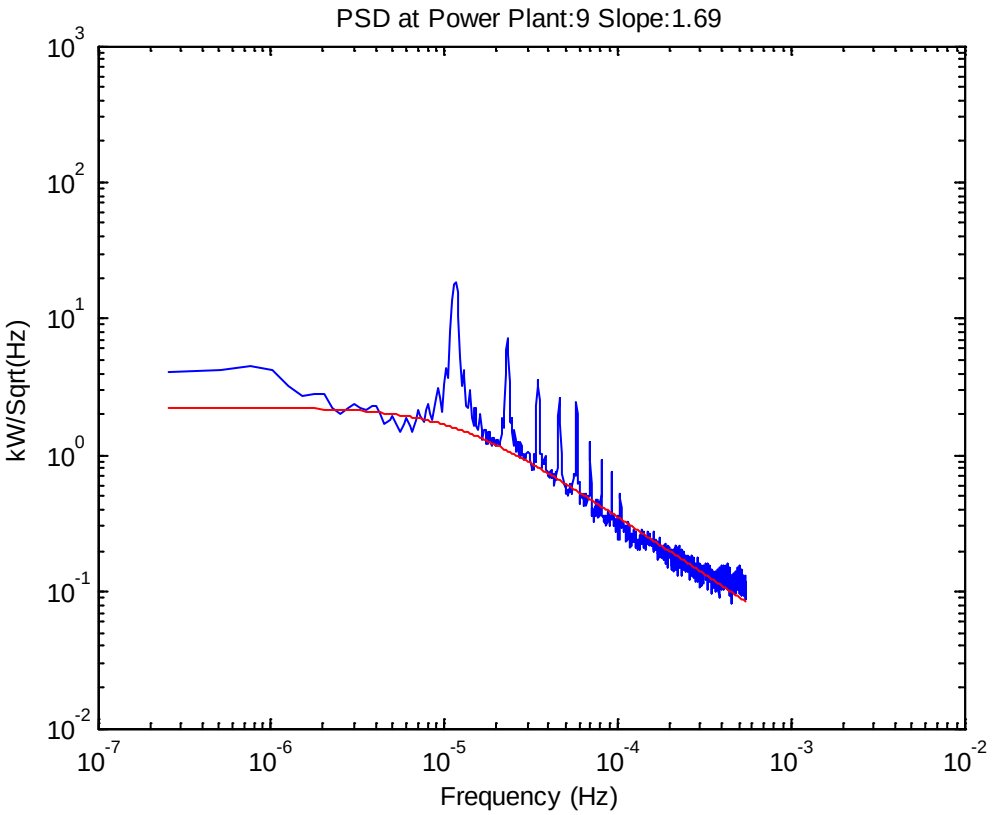

Figure S29: PSD (blue) with line of best fit (red) for Plant 9. The slope at high frequencies in the loglog domain (*m*) is -1.69.

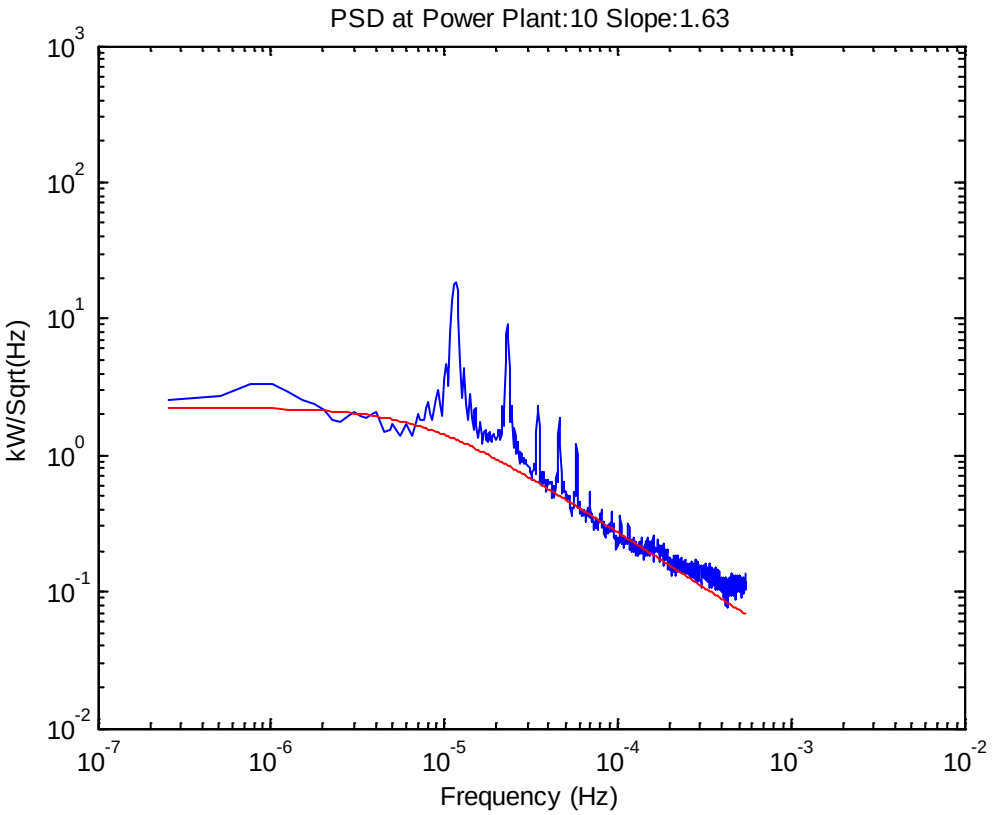

Figure S30: PSD (blue) with line of best fit (red) for Plant 10. The slope at high frequencies in the loglog domain (*m*) is -1.63.

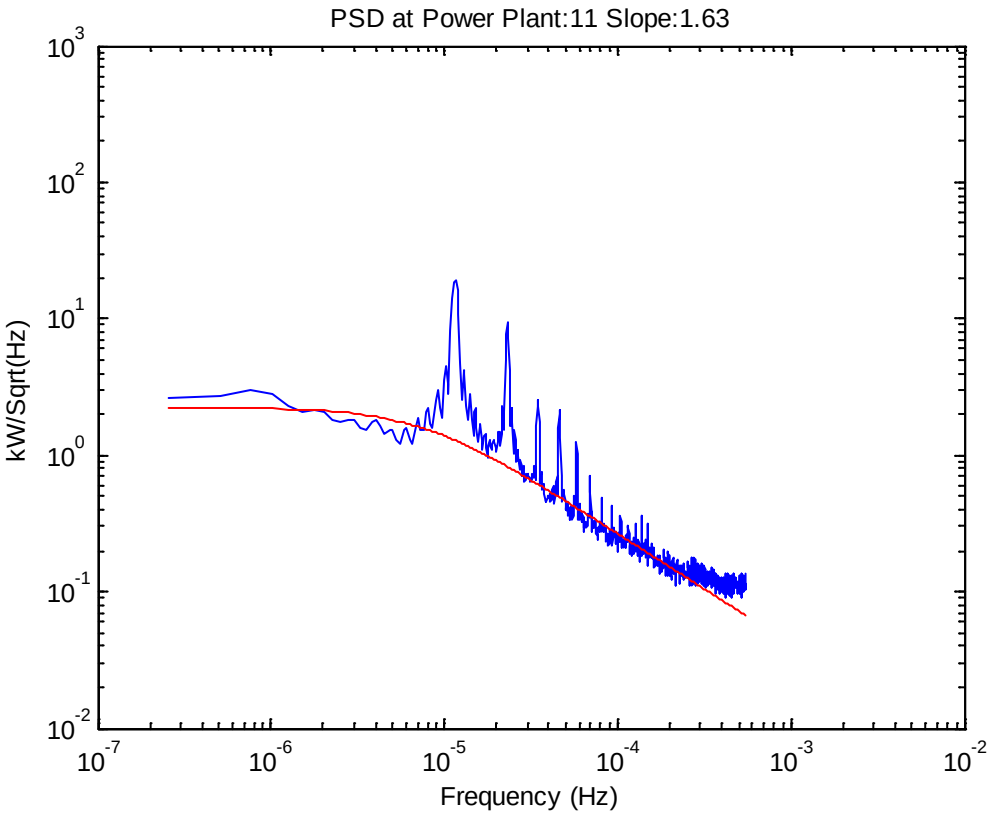

Figure S31: PSD (blue) with line of best fit (red) for Plant 11. The slope at high frequencies in the loglog domain (*m*) is -1.63.

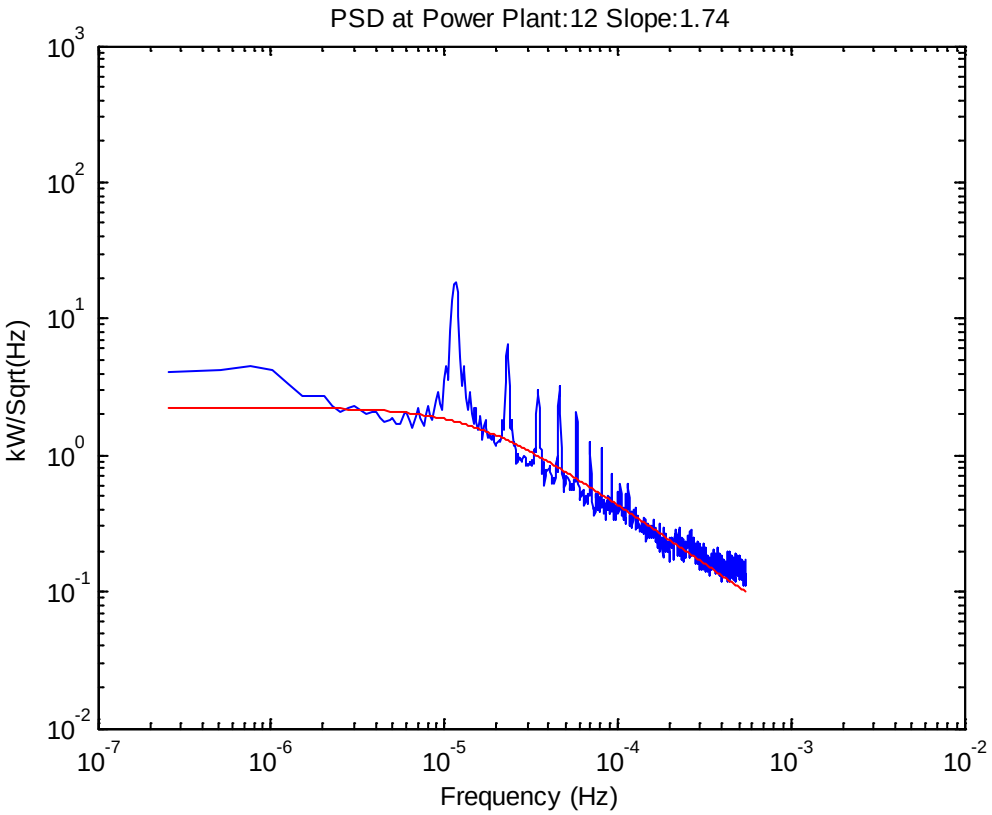

Figure S32: PSD (blue) with line of best fit (red) for Plant 12. The slope at high frequencies in the loglog domain (*m*) is -1.74.

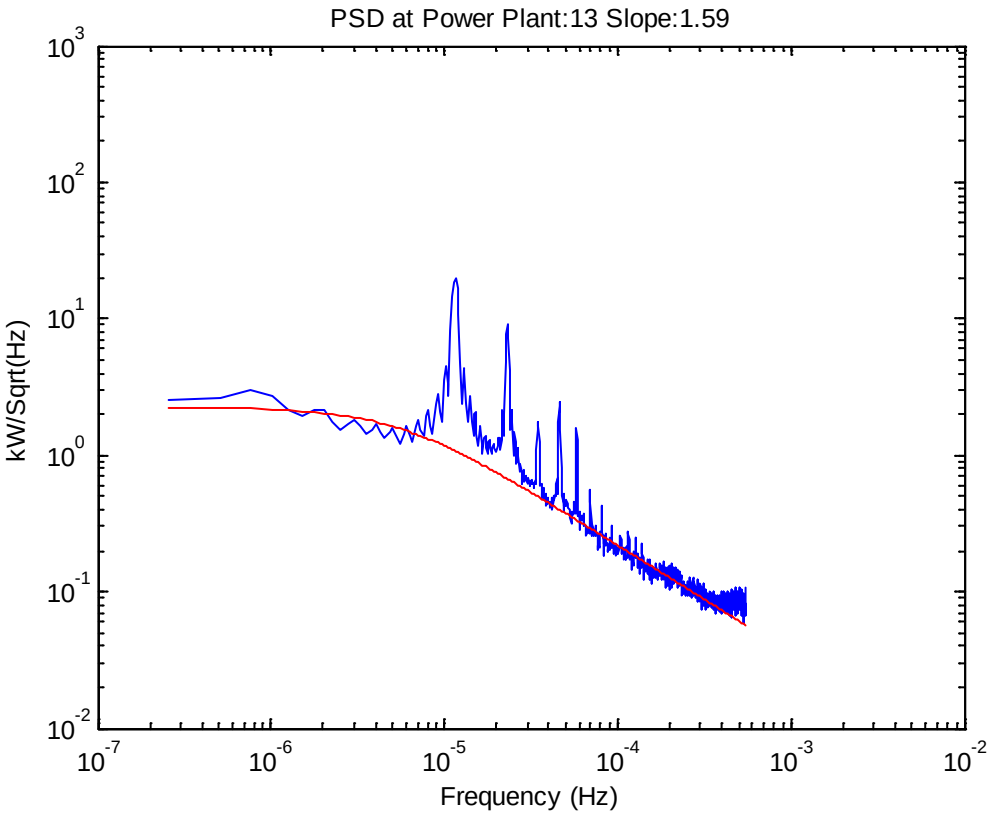

Figure S33: PSD (blue) with line of best fit (red) for Plant 13. The slope at high frequencies in the loglog domain (*m*) is -1.59.

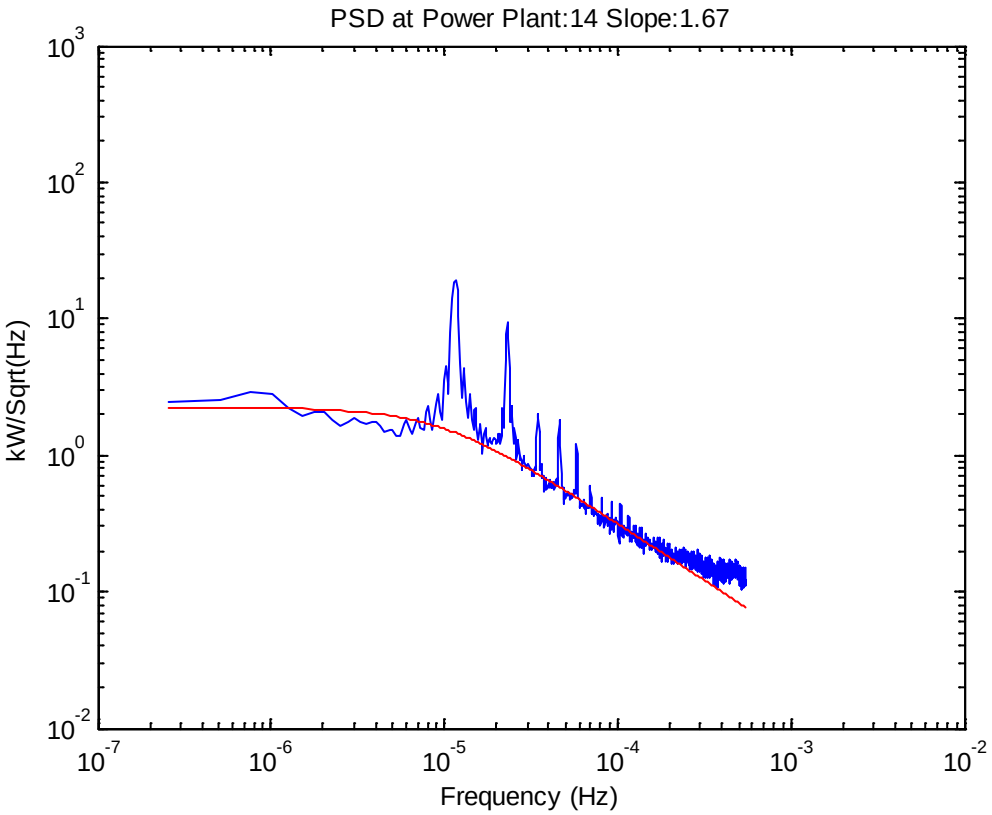

Figure S34: PSD (blue) with line of best fit (red) for Plant 14. The slope at high frequencies in the loglog domain (*m*) is -1.67.

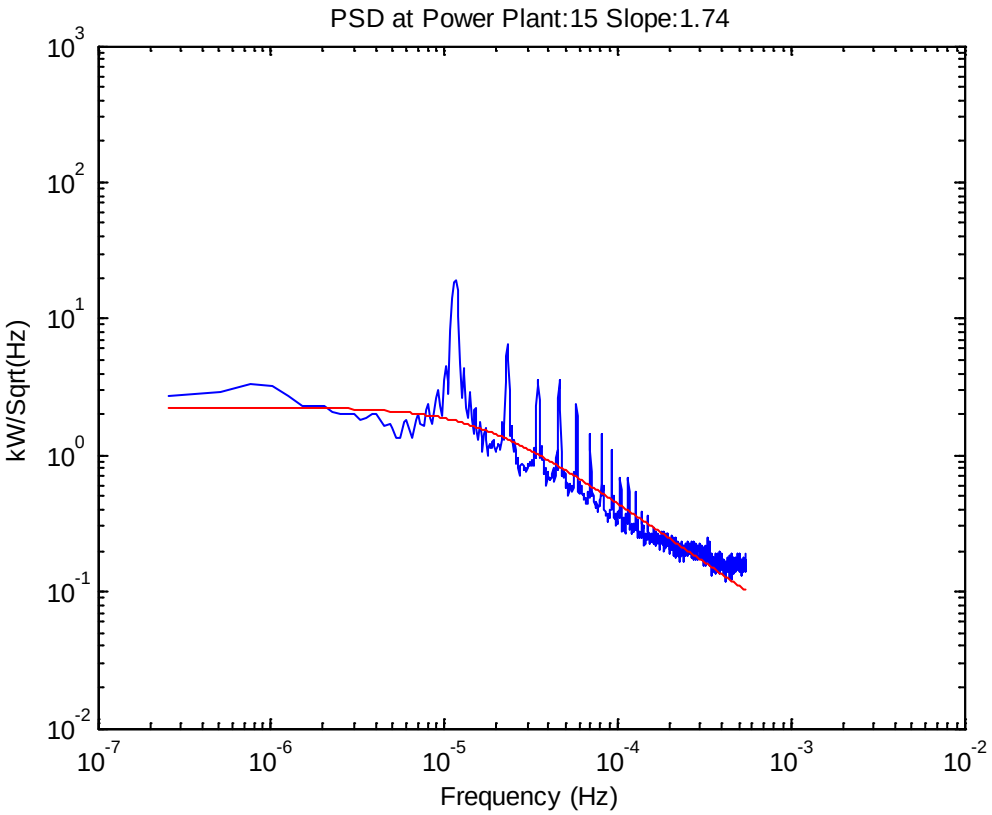

Figure S35: PSD (blue) with line of best fit (red) for Plant 15. The slope at high frequencies in the loglog domain (*m*) is -1.74.

5. Gujarat and California cloud regimes

In the main text we consider whether the difference in geographic smoothing between the utility-scale plants in the Indian state of Gujarat and the US state of California arises from the absence of persistent cloud activity in Gujarat during the 8 non-monsoon months.

The Gujarat PV plants are at 21-24 degrees north latitude, and per the Köppen classification system has an arid hot desert (BWh) climate over the spatial area examined[1]. The California plants are at 32.7 – 36.6 degrees north latitude, and have Köppen climate classifications of cold desert (BWk), hot and cold steppe (BSh and BSk), and temperate with hot and dry summers (Csa).

A qualitative understanding of the differences in cloud activity in the two regions can be aided by examining satellite images of the regions over a long period. The Moderate Resolution Imaging Spectroradiometer (MODIS) on the Terra satellite provides daily near-global images that approximate true color[2]. While the images have some areas for each day with no coverage (due to the orbit and sun parameters), they are useful for forming a picture of the differences between the two regions (Figures S35 and S36). A 200x200 pixel region encompassing Gujarat of the 0.1 degree resolution MODIS image for each day in January 2016, with each row being 1 week. The black no-data regions in some frames are due to unfavorable orbital and sun geometries.

[1] Peel, M.C.; Finlayson, B.L.; McMahon, T. A. (2007). Updated world map of the Köppen–Geiger climate classification. Hydrology and Earth System Sciences. 11 1633–1644.

[2] http://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD_143D_RR&date=2016-08-23

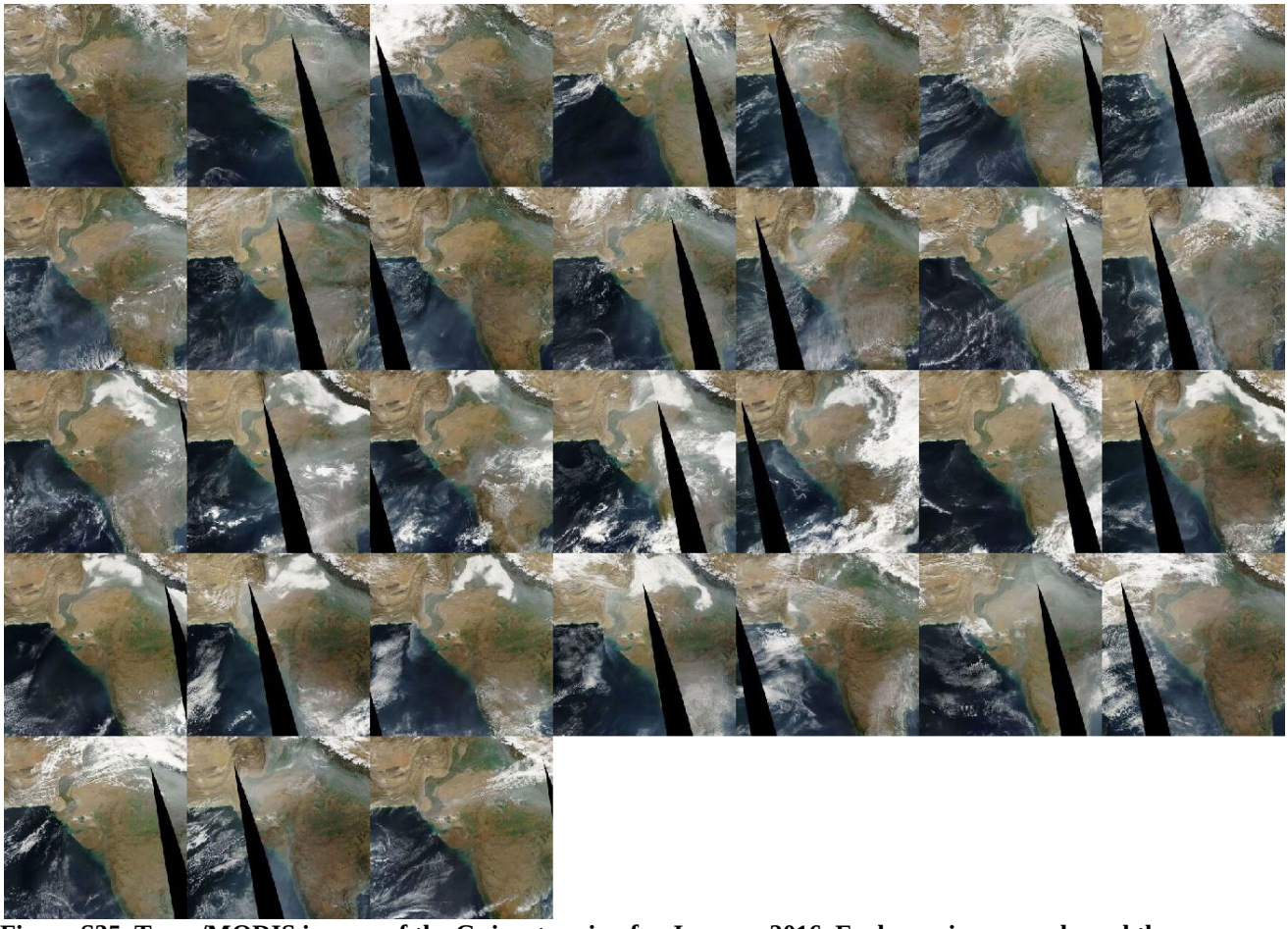

Figure S35. Terra/MODIS images of the Gujarat region for January 2016. Each row is one week, and the frames are sequential beginning January 1 and are 200x200 pixel portions of the 0.1 degree resolution MODIS image.

The generally clear Gujarat weather is in contrast to the California ISO plant geographic area weather for the same month, January 2016, shown below.

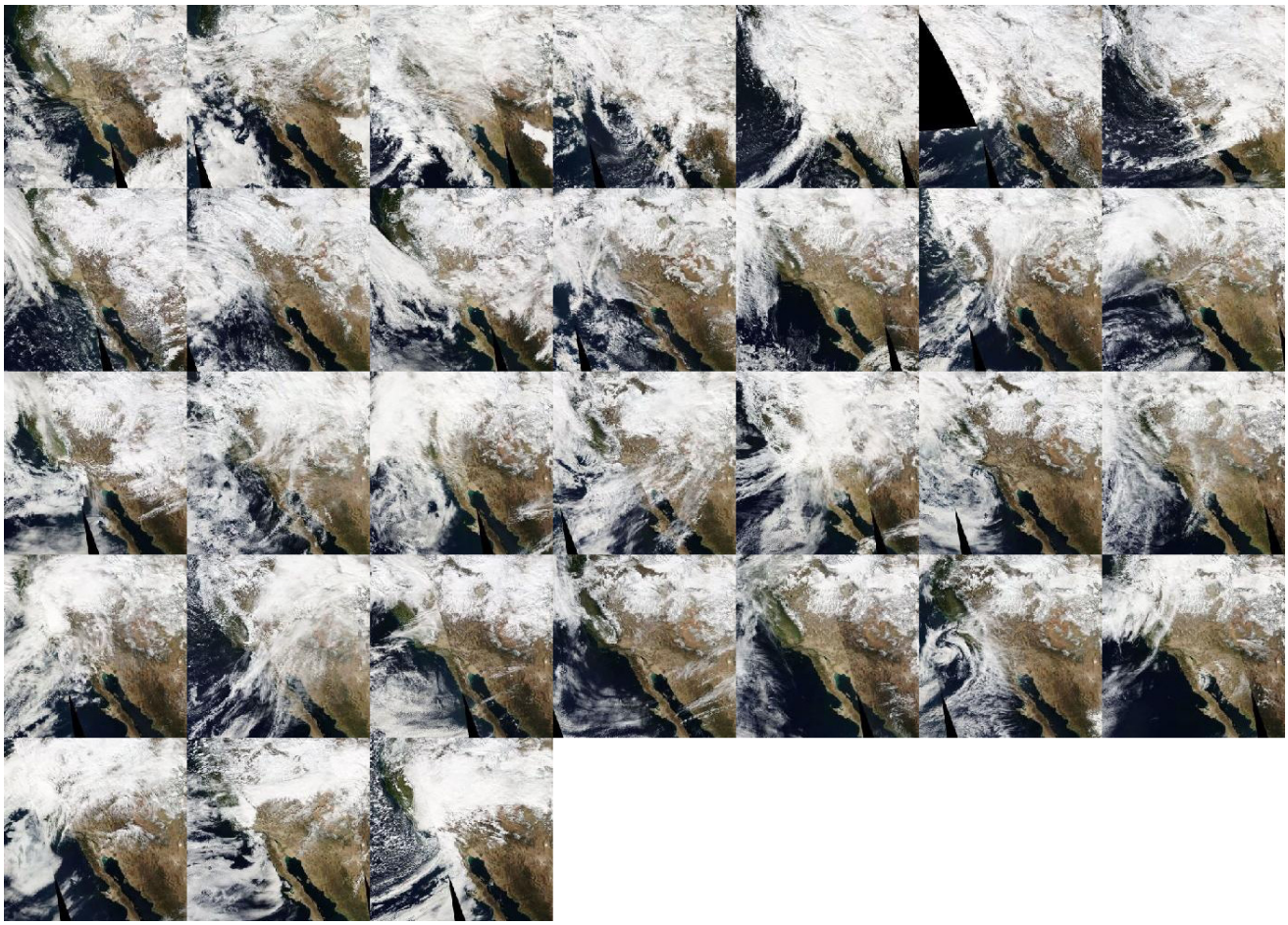

Figure S36. Terra/MODIS images of the region containing the CA-ISO PV plants for January 2016. Each row is one week, and the frames are sequential beginning January 1 and are 200x200 pixel portions of the 0.1 degree resolution MODIS image.

One measure of the cloud fraction can be obtained from the European Space Agency's Tropospheric Emission Monitoring Internet Service (TEMIS) that derives the effective cloud fraction from spectra at 758-777 nm near the oxygen A-band[3]. Monthly data averaged over a 1x1 degree box are available from August 2002-March 2012, and can be accessed through the Royal Netherlands Meteorological Institute Climate Explorer[4]. The 116 months of data for an area encompassing the Gujarat and California ISO (CA-ISO) sites produced by the Climate Explorer tool have been plotted in Figure S37, with the mean and colored bands representing 17% and 83% confidence intervals.

The plot for the CA-ISO PV plant region shows that the monthly average cloud fraction never falls to the low values (~5%) seen in Gujarat during the 8 non-monsoon months, even during the clearest months in CA-ISO.

[3] A description is available at http://www.temis.nl/fresco/fresco.html
[4] http://climexp.knmi.nl/select.cgi?id=someone@somewhere&field=fresco6_cloud

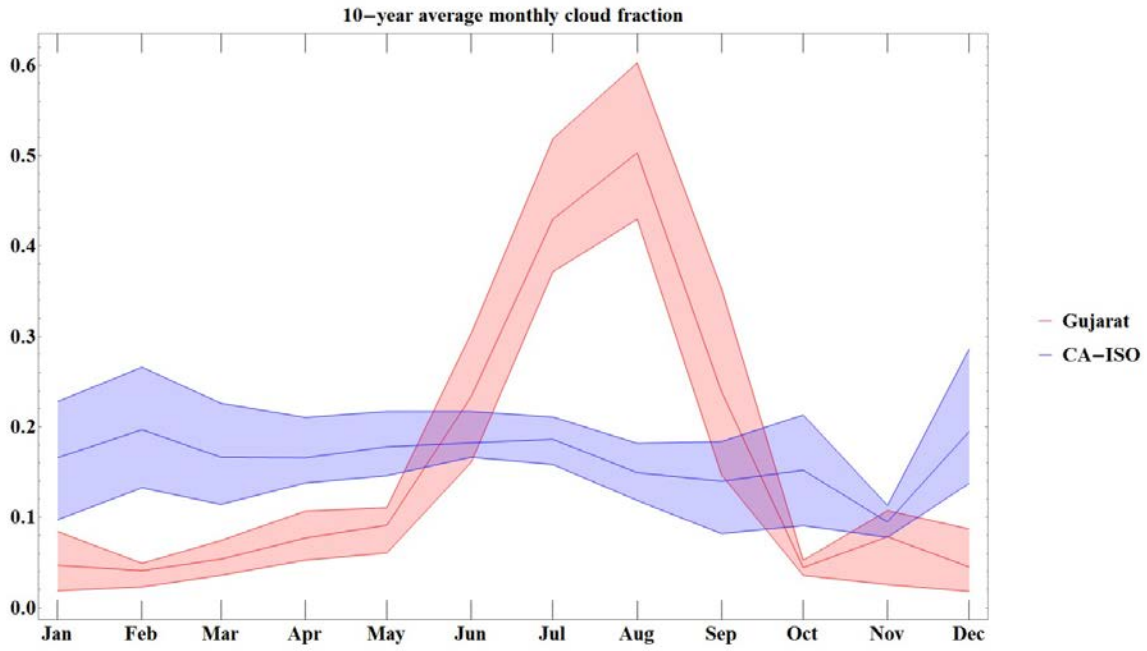

Figure S37. Average monthly cloud fraction for Gujarat (red) and the CA-ISO PV plant region (blue), with 17% and 83% confidence interval bands for the 116 months from August 2002-March 2012. Geographic boundaries used were east longitude 69 to 72, north latitude 21 to 24 (Gujarat) and longitude -112.8 to -120.4, latitude 32.6 to 36.6 (CA-ISO).

The single grid point data for the Arizona PV array used for the subarray analysis is, of course, more variable than the larger areas over which the Gujarat and CA-ISO plants are spread. Figure S38 compares the three regions.

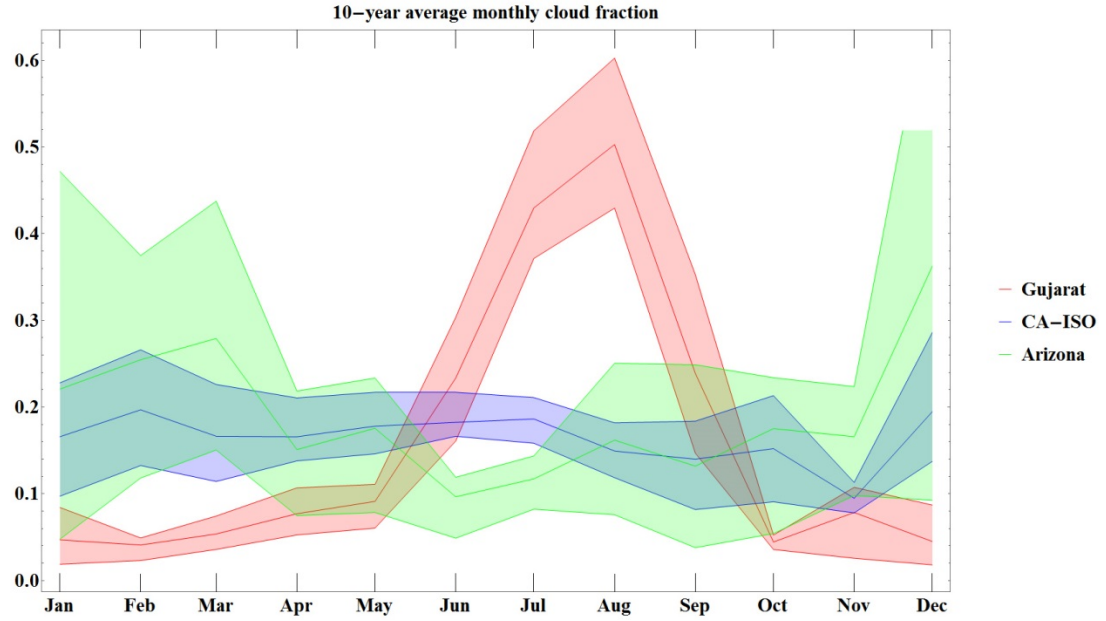

Figure S38. Average monthly cloud fraction for Gujarat (red), the CA-ISO PV plant region (blue), and the single Arizona plant, with 17% and 83% confidence interval bands.