ESTIMATING DAY-AHEAD SOLAR FORECAST ERRORS FROM SITE-SPECIFIC VARIABILITY

Philippe Lauret[1], Josselin Le Gal La Salle[1] and Mathieu David[1]

[1] PIMENT Laboratory/University of La Réunion, Saint-Pierre (Réunion)

Abstract

This work introduces a straightforward statistical model to estimate the Root Mean Square Error (RMSE) of day-ahead solar irradiance forecasts using only a site-specific solar variability metric. The proposed method leverages a global dataset of 60 quality-controlled ground stations and employs a linear regression between RMSE and the standard deviation of hourly changes in the clear-sky index. This results in an efficient pre-deployment tool that helps practitioners assess forecast errors at a site before implementing a forecasting method. The model includes also a 95% uncertainty interval, providing valuable insight into the bounds within which future forecast errors are likely to fall. Validation on seven independent sites from the SURFRAD network confirms that the proposed model can reasonably predict RMSE forecast errors based on a site's nominal variability.

Keywords: Solar forecast error, Solar variability, Root Mean Square Error (RMSE), Forecast verification

1. Introduction

Accurate forecasting of solar irradiance is essential for the reliable integration of solar power into the grid. Forecast accuracy is known to vary with local atmospheric and climatic conditions, and among these, solar variability—the short-term fluctuation in irradiance due to clouds and weather patterns—plays a central role in determining the intrinsic difficulty of forecasting at a given site. Several recent studies have attempted to quantify this notion of forecast difficulty or predictability.

Yang (2022) proposed an advanced framework to estimate the upper and lower bounds of the forecast error based on a theoretical definition of predictability. His method combines a statistical reference model (the optimal convex combination of climatology and persistence) to infer the upper bound, and a perturbed numerical weather prediction (NWP) experiment to estimate the lower bound through a predictability error growth analysis. Although elegant, this physics-based approach is complex to implement, requires tuning several parameters, and depends on assumptions about NWP perturbations. Moreover, the concept of predictability—while theoretically appealing—remains somewhat abstract and not always straightforward to interpret or apply in practice for operational forecasters.

In parallel, Karimi-Arpanahi et al. (2023) introduced a complementary data-driven approach by defining a predictability index based on Weighted Permutation Entropy (WPE), a statistical metric that captures the temporal complexity of renewable generation time series. Their analysis of more than 300 PV systems across Australia showed that WPE correlates strongly with forecast errors and that considering predictability in decision-making can lead to better operational policies, higher profits for investors, and lower electricity costs for consumers. However, entropy-based metrics such as WPE require parameter tuning and lack direct physical interpretation, which limits their adoption by the broader solar forecasting community.

Building on these previous efforts, the present study proposes a simpler and more transparent alternative to estimate forecast errors without explicitly invoking the concept of predictability. Our approach directly relates the Root Mean Square Error (RMSE) of day-ahead irradiance forecasts to a solar variability metric, namely the standard deviation of hourly changes in the clear-sky index. This quantity is well established in the solar energy literature, easily derived from standard irradiance or satellite data, and intuitively represents the inherent variability of the site.

10.18086/swc.2025.04.21 Available at http://proceedings.ises.org

Similarly to the work of Karimi-Arpanahi et al. (2023), we assume that lower variability implies higher predictability, which in turn supports better-informed decisions and economic outcomes. Yet, the metric used here—solar variability—is physically meaningful, globally available, and readily interpretable by both researchers and practitioners.

The objective of this paper is therefore to develop and validate a straightforward, globally applicable statistical model that predicts the day-ahead RMSE of solar irradiance forecasts using only a site-specific variability indicator. The model provides not only an expected RMSE but also its 95% uncertainty interval, thereby defining practical bounds within which future forecast errors are likely to fall. The goal is to equip system planners, and decision-makers with a lightweight method to estimate expected forecast performance using minimal data inputs, specifically a single year's worth of hourly irradiance data. Furthermore, the predicted RMSE bounds will enable solar forecasters to efficiently assess how their forecasts perform in relation to these expected limits.

2. Model formulation

Our model hypothesis relates the RMSE of day-ahead forecasts to the variability of the clear-sky index. The clear-sky index $k_c = {}^{I}/_{I_c}$ is calculated using measured Global Horizontal Irradiance GHI (I) and a clear-sky model output (I_c) provided by Lefèvre et al. (2013). For the variability metric, we use the definition proposed by Perez et al. (2016) which is defined as the standard deviation of a one-year time series of hourly differences in the clear-sky index and read as $\sigma(\Delta k_c)$. The RMSE is computed as $RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\widehat{I_n} - I_n\right)^2}$ for N pairs of GHI forecast $\hat{I}$/observations I.

3. Model derivation

Sixty globally distributed locations (see Figure 1 and annex), with quality-controlled GHI data, provided by Subtask 1 of the IEA PVPS Task 16 program (Forstinger et al., 2021a, Forstinger et al., 2021b) were used to build the model. More specifically, we constructed a dataset containing site-specific variability along with RMSE of day-ahead operational high resolution (HRES) GHI forecasts from European Centre for Medium-Range Weather Forecasts (ECMWF). The operational HRES GHI forecasts (00Z run) were retrieved from the ECMWF's Meteorological Archival and Retrieval System (MARS). The RMSE of these forecasts and the site variability were computed for one year. Figure 2 highlights a clear dependence between RMSE and nominal variability, with a significant correlation of 0.72.

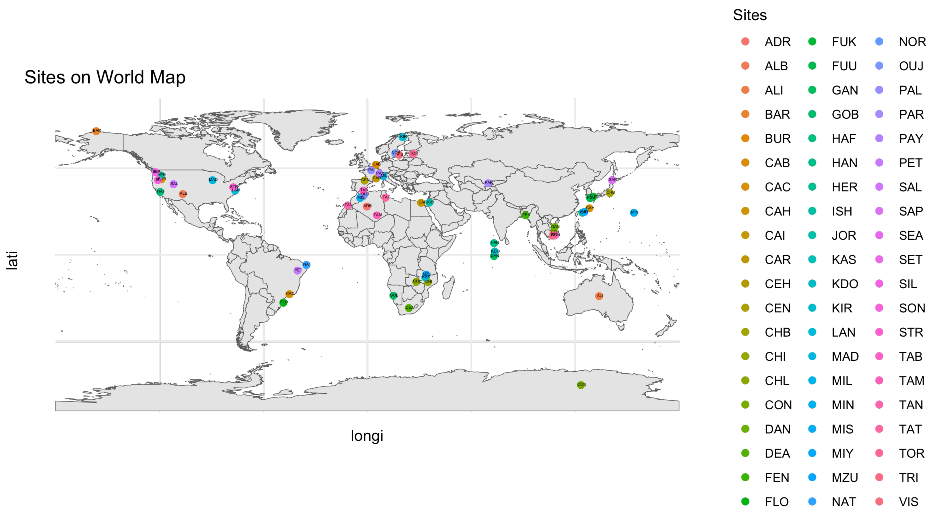

Figure 1: Geographical distribution of the sites used for model development.

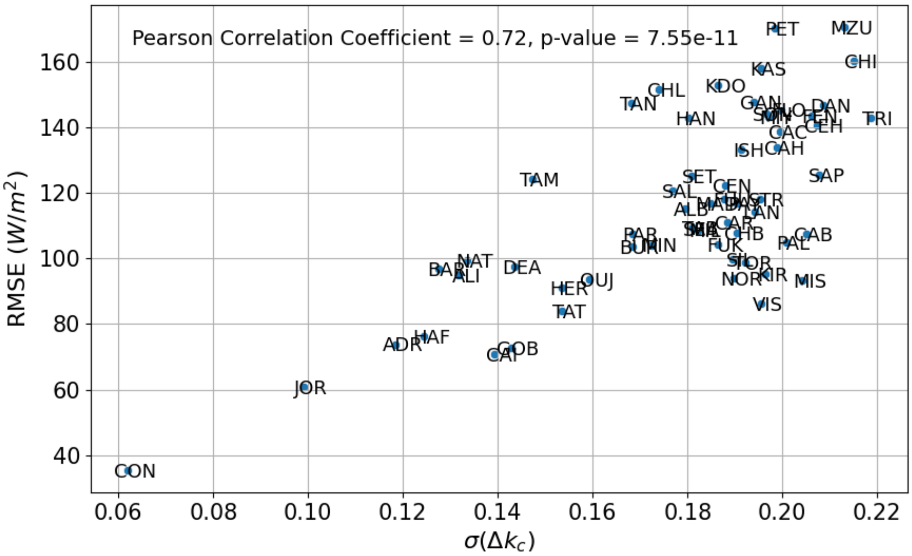

Figure 2: Empirical evidence of a relationship between nominal variability and RMSE.

Building on this experimental evidence, we derive a linear relationship between forecast error (RMSE) and nominal variability $\sigma(\Delta k_c)$. More precisely, a linear regression is fitted, and the 2.5% and 97.5% conditional quantiles are computed to define a 95% uncertainty interval, representing the RMSE bounds of the prediction model. The corresponding regression and quantile equations (denoted here D24 model equations) are presented in Table 1 and Figure 3.

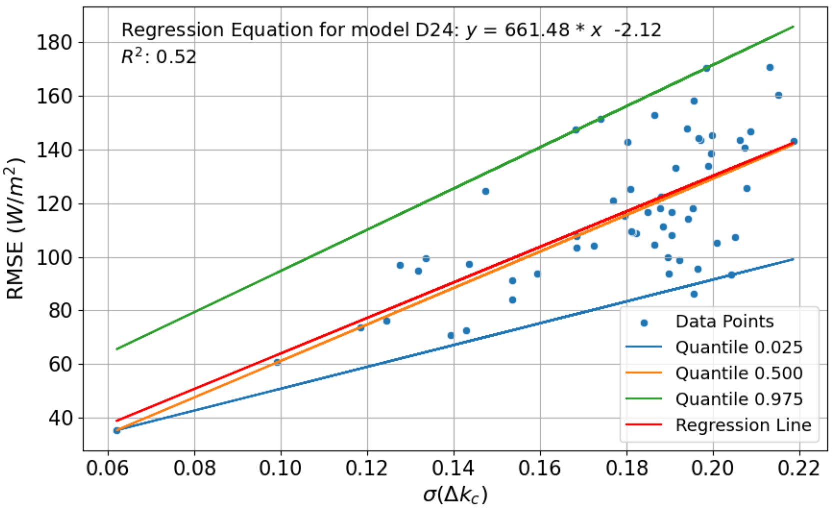

Figure 3: Day-ahead model D24 regression fit and 95% quantile bounds.

Table 1: D24 model equations: x = $\sigma(\Delta k_c)$ y = RMSE.

Quantile	Equation
2.5%	y = 406.53 x + 10.21
Mean (regression)	y = 661.48 x − 2.12
97.5%	y = 767.26 x + 18.02

Figure 4 presents the predictions given by the linear regression model for the 60 sites, ordered by increasing variability. The reference RMSE from the ECMWF day-ahead forecast is indicated by a black cross for each site. For most sites, the model provides a good agreement with the observed RMSE values. As expected, the uncertainty interval—defining the RMSE bounds—widens with increasing site variability. Although a few deviations are observed, the ECMWF reference day-ahead RMSEs generally fall within these bounds. The results also highlight the wide range of forecast errors captured by the proposed model.

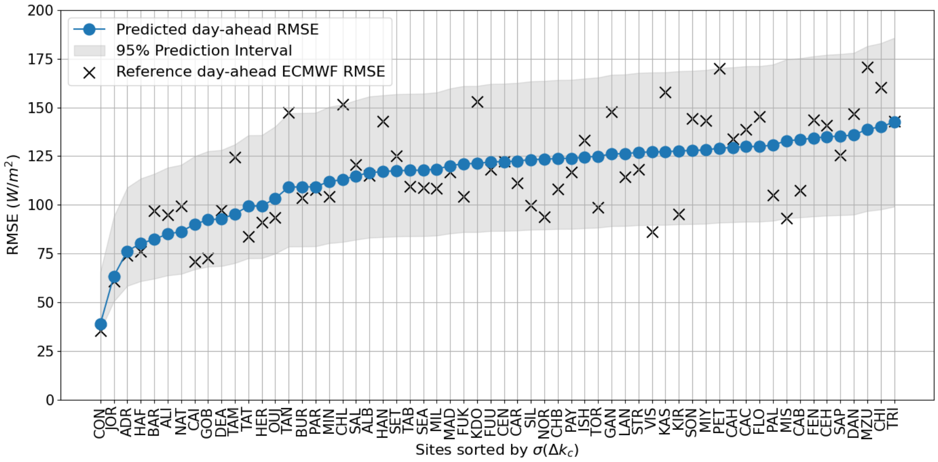

Figure 4: Deployment of the day-ahead prediction model across 60 sites. Reference RMSE values correspond to those produced by the IFS ECMWF forecast model.

4. Model verification

The proposed model is further evaluated on seven sites from the SURFRAD network, widely used for solar forecast development and validation. Several studies, including those by Perez et al. (2010), Perez et al. (2018), Yang (2022) and Zhang et al. (2022) have relied on this dataset to produce intra-day and day-ahead forecasts. The main characteristics of the SURFRAD stations are summarized in Table 2.

Table 2: SURFRAD stations. The reference IFS ECMWF RMSE is computed for the year listed in the Table. $\sigma(\Delta k_c)$, is the nominal variability which is used as input of the RMSE prediction model.

Site	Code	Year	Latitude	Longitude	Altitude (m)	$\sigma(\Delta k_c)$
Bondville	BON	2016	40.067	-88.367	213	0.190
Desert Rock	DRA	2016	36.626	-116.018	1007	0.137
Fort Peck	FPK	2016	48.317	-105.100	634	0.178
Goodwin creek	GWN	2016	34.255	-89.873	98	0.175
Penn State Uni.	PSU	2016	40.720	-77.930	376	0.189
Sioux Falls	SXF	2016	43.730	-96.620	473	0.183
Table Mountain	TBL	2016	40.125	-105.237	1689	0.210

RMSE model predictions (blue dots in Figure 5) are compared to the reference RMSE values derived from ECMWF operational forecasts (which serve here as test forecasts). In addition, we include published RMSE results from: (1) the North American Mesoscale (NAM) model forecasts (Yang and Perez, 2019), (2) post-processed NAM forecasts (Yang, 2019), (3) SolarAnywhere v4 forecasts (Perez et al., 2018) and (4) (Perez et al., 2013).

As shown in Figure 5, nearly all RMSE values lie within the predicted uncertainty interval. In addition, it must be noted that the RMSE bounds for the day-ahead case are comparable in magnitude (≈ 75 W/m²) to those reported by Yang (2022).

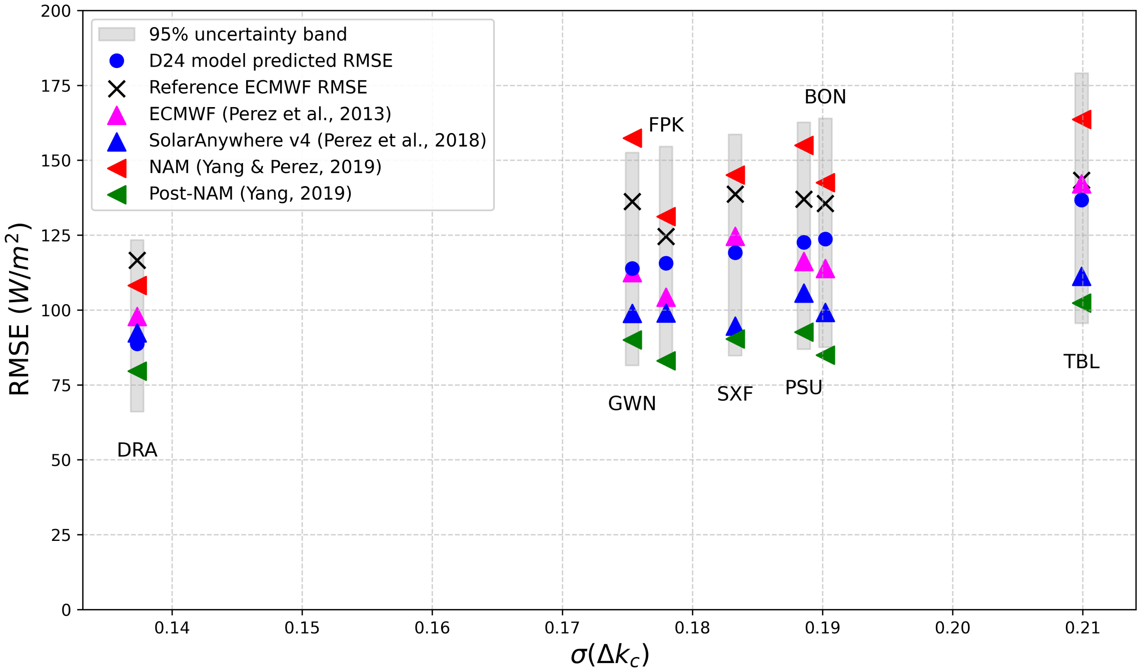

Figure 5: Verification of the proposed models on the SURFRAD stations.

5. Discussion

The previous verification stage confirms that the proposed model can rapidly provide a first estimate of forecast performance at a given site when its variability is known. This offers a practical tool for both forecasters and project developers to assess site-specific forecast difficulty and potential economic value. Unlike the physics-based method of Yang (2022), our approach is purely statistical, defining RMSE bounds through a 95% uncertainty interval. Although simplified, the resulting bounds show comparable magnitude to those of Yang (2022). Overall, the model constitutes an efficient first-step verification method that can be complemented, when needed, by more detailed physics-based analyses.

The prediction model relies on the knowledge of solar variability at the site of interest. Although ground-based GHI data may not always be available, this limitation can be overcome using satellite-derived GHI products with global coverage. Combined with freely accessible clear-sky time series, such as those provided by CAMS McClear (2024), these datasets enable the computation of the nominal variability metric for virtually any location worldwide. As a result, users can readily estimate both the expected RMSE forecast error and its associated bounds without requiring local ground measurements.

6. Conclusion

In this work, we developed a simple linear regression model to predict day-ahead RMSE values from site-specific solar variability. A preliminary verification shows that the model can reasonably reproduce forecast errors and their uncertainty ranges. These initial results suggest that the proposed model could serve as a valuable tool for early-stage project planning, benchmark evaluation, and rapid site assessment. The inclusion of uncertainty bounds further enables solar forecasters to contextualize forecast accuracy and identify outlier performance.

Future work may explore adding supplementary predictors to the model or testing other types of solar variability metrics.

7. Acknowledgments

This research received funding from the TwInSolar project under the European Union's Horizon Europe research and innovation program (Grant Agreement No. 101076447).

8. References

CAMS McClear. Copernicus Atmosphere Monitoring Service – McClear clear-sky irradiance model. Available at: https://www.soda-pro.com/web-services/radiation/cams-mcclear (accessed 22 Oct 2024).

Forstinger, A., Saint-Drenan, Y.-M., Wilbert, S., Jensen, A., Kraas, B., Peruchena, C.F., Gueymard, C., Ronzio, D., Yang, D., Collino, E., Martinez, J.P., Ruiz-Arias, J., Hanrieder, N., Blanc, P., 2021. IEA-PVPS Task 16 Reference Solar Measurements. MINES ParisTech.

Forstinger, A., Wilbert, S., Jensen, A., Kraas, B., Peruchena, C., Gueymard, C., Ronzio, D., Yang, D., Collino, E., Martinez, J., Ruiz-Arias, J., Hanrieder, N., Blanc, P., Saint-Drenan, Y., 2021. Expert quality control of solar radiation ground data sets. In: Proc. SWC 2021 – ISES Solar World Congress. International Solar Energy Society, pp. 1037–1048.

Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., Wald, L., Schroedter-Homscheidt, M., Hoyer-Klick, C., Arola, A., Benedetti, A., Kaiser, J.W., Morcrette, J.J., 2013. McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmospheric Measurement Techniques, 6, 2403–2418. https://doi.org/10.5194/amt-6-2403-2013

Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Renne, D., Hoff, T.E., 2010. Validation of short- and medium-term operational solar radiation forecasts in the US. Solar Energy, 84 (12), 2161–2172. https://doi.org/10.1016/j.solener.2010.08.014

Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Van Knowe, G., Hemker, K., Heinemann, D., Remund, J., Muller, S.C., Traunmüller, W., Steinmauer, G., Pozo, D., Ruiz-Arias, J.A., Lara-Fanego, V., Ramirez-Santigosa, L., Gaston-Romero, M., Pomares, L.M., 2013. Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. Solar Energy 94, 305–326.

Perez, R., David, M., Hoff, T.E., Jamaly, M., Kivalov, S., Kleissl, J., Lauret, P., Perez, M., 2016. Spatial and temporal variability of solar energy. Foundations and Trends in Renewable Energy, 1 (1), 1–44. https://doi.org/10.1561/2700000002

Perez, R., Schlemmer, J., Kivalov, S., Dise, J., Keelin, P., Grammatico, M., et al., 2018. A new version of the SUNY solar forecast model: A scalable approach to site-specific model training. In: Proc. IEEE 45th Photovoltaic Specialists Conference, pp. 1–6. https://doi.org/10.1109/PVSC.2018.8547650

Karimi-Arpanahi, S., Pourmousavi, S.A., Mahdavi, N., 2023. Quantifying predictability and its impact on decision-making in renewable generation forecasting. Applied Energy, 347, 121467. https://doi.org/10.1016/j.apenergy.2023.121467

Yang, D., 2022. Correlogram, predictability error growth, and bounds of mean square error of solar irradiance forecasts. Renewable and Sustainable Energy Reviews, 167, 112736. https://doi.org/10.1016/j.rser.2022.112736

Zhang, G., Yang, D., Galanis, G., Androulakis, E., 2022. Solar forecasting with hourly updated numerical weather prediction. Renewable and Sustainable Energy Reviews, 154, 111891. https://doi.org/10.1016/j.rser.2021.111891

Appendix: List of sites used to build the model

Table 3: List of sites used to build the RMSE prediction model. RMSE of HRES IFS forecasts were computed for year listed in the Table. σ(Δkc) is the nominal variability of the site.

Site	Code	Source	Year	Latitude	Longitude	Altitude (m)	$\sigma(\Delta k_c)$
Adrar	ADR	Enermena	2016	27.880	-0.274	262.0	0.118
Albuquerque	ALB	NOAA	2019	35.038	-106.622	1617.0	0.179
Alice Springs	ALI	BSRN	2018	-23.798	133.888	547.0	0.132
Barrow	BAR	NOAA	2017	71.323	-156.611	11.0	0.128
Burns	BUR	SRML	2018	43.519	-119.022	1265.0	0.168
Cabauw	CAB	BSRN	2019	51.971	4.927	0.0	0.205
Cachoeira Paulista	CAC	Sonda	2018	-22.690	-45.006	574.0	0.200
Cape Hedo	CAH	SKYNET	2018	26.867	128.248	65.0	0.199
Cairo University	CAI	Enermena	2018	30.036	31.009	104.0	0.139
Carpentras	CAR	BSRN	2018	44.083	5.059	100.0	0.188
Central Highlands	CEH	ESMAP	2019	12.753	107.876	277.0	0.207
Cener	CEN	BSRN	2018	42.816	-1.601	471.0	0.188
Chiba	CHB	SKYNET	2018	35.625	140.104	21.0	0.190
Chileka	CHI	ESMAP	2017	-15.680	34.972	769.0	0.215
Chilanga	CHL	ESMAP	2017	-15.548	28.248	1226.0	0.174
Concordia	CON	BSRN	2018	-75.100	123.383	3233.0	0.062
Da Nang	DAN	ESMAP	2019	16.013	108.186	11.0	0.209
Dar Es Salaam	DAR	ESMAP	2017	6.781	39.204	122.0	0.200
De Aar	DEA	BSRN	2018	-30.667	23.993	1287.0	0.144
Feni	FEN	ESMAP	2019	22.800	91.358	7.0	0.206
Florianopolis	FLO	BSRN	2018	-27.605	-48.523	11.0	0.200
Fukuoka	FUK	BSRN	2018	33.582	130.376	3.0	0.186
Fukue	FUU	SKYNET	2018	32.752	128.682	80.0	0.188
Gan	GAN	ESMAP	2017	-0.691	73.150	5.0	0.194
Gobabeb	GOB	BSRN	2018	-23.561	15.042	407.0	0.143
Hanford	HAF	NOAA	2018	36.314	-119.632	73.0	0.124
Hanimaadhoo	HAN	ESMAP	2017	6.748	73.170	4.0	0.180
Hermiston	HER	SRML	2018	45.818	-119.285	188.0	0.153
Ishigakijima	ISH	BSRN	2018	24.337	124.164	5.7	0.191
Izana	IZA	BSRN	2016	28.309	-16.499	2372.9	0.113
Univ. of Jordan	JOR	Enermena	2017	30.172	35.818	1012.0	0.099
Kasungu	KAS	ESMAP	2017	-13.015	33.468	1067.0	0.196
Kadhdhoo	KDO	ESMAP	2017	1.860	73.520	2.0	0.186
Langley Res. Center	LAN	BSRN	2018	37.104	-76.387	3.0	0.194
Madison	MAD	NOAA	2018	43.072	-89.411	271.0	0.185
Milan RSE	MIL	RSE	2018	45.476	9.255	150.0	0.182
Minamitorishima	MIN	BSRN	2018	24.288	153.983	7.1	0.172
Missour	MIS	Enermena	2017	32.860	-4.107	1107.0	0.204
Miyako	MIY	SKYNET	2017	24.737	125.327	50.0	0.197
Mzuzu	MZU	ESMAP	2017	-11.420	33.995	1287.0	0.213
Natal	NAT	Sonda	2018	-5.837	-35.206	58.0	0.134
Univ. of Oujda	OUJ	Enermena	2019	34.650	-1.900	617.0	0.159
Palaiseau	PAL	BSRN	2018	48.713	2.208	156.0	0.201
Payerne	PAY	BSRN	2018	46.815	6.944	491.0	0.190
Petrolina	PET	BSRN	2018	-9.068	-40.319	387.0	0.198
Salt Lake City	SAL	NOAA	2016	40.772	-111.955	1288.0	0.177
Sapporo	SAP	BSRN	2018	43.060	141.329	17.2	0.208
Seattle	SEA	NOAA	2017	47.687	-122.257	20.0	0.181
Seattle	SET	SRML	2017	47.654	-122.309	70.0	0.181

Silver Lake	SIL	SRML	2018	43.119	-121.059	1324.0	0.189
Song Binh	SON	ESMAP	2017	11.264	108.345	61.0	0.197
Sterling	STR	NOAA	2018	38.972	-77.487	85.0	0.195
Plataforma Solar	TAB	CIEMAT	2018	37.091	-2.358	500.0	0.181
Tamanrasset	TAM	BSRN	2017	22.790	5.529	1385.0	0.147
Tan-Tan	TAN	Enermena	2016	28.498	-11.322	75.0	0.168
Tataouin	TAT	Enermena	2016	32.974	10.485	210.0	0.153
Toravere	TOR	BSRN	2018	58.254	26.462	70.0	0.192
Tri An	TRI	ESMAP	2018	11.102	107.038	59.0	0.219
Kiruna	KIR	SMHI	2018	67.842	20.410	424.0	0.196
Norrköping	NOR	SMHI	2018	58.582	16.148	53.0	0.190
Parkent	PAR	p.C.	2016	41.320	69.740	1081.0	0.168
Visby	VIS	SMHI	2018	57.673	18.345	49.0	0.195