Dependence of sub-hourly solar variability statistics on time interval and cloud vertical position

Mónica Zamora Zapata[1] and Jan Kleissl[2]

[1)] *Departamento de Ingeniería Mecánica, Universidad de Chile*

[2)] *Department of Mechanical and Aerospace Engineering, University of California San Diego*

(*mzamora@uchile.cl.)

(Dated: 13 June 2022)

Solar variability corresponds to strong variations of the solar irradiance, caused mainly by the presence of clouds. Practical uses of solar resource data, such as the design of photovoltaic solar plants, usually employs several years of hourly data, neglecting subhourly features. The effect of clouds on short-time variability can differ by cloud type, suggesting that some cloud effects could be ignored when working with hourly data. In this work, we study compare statistics of solar variability calculated at different time intervals, and separate the analysis by cloud categories. We use 1 minute solar data and cloud radar products from the ARM CACTI campaign in Córdoba, Argentina, where a wide variety of clouds exist. We classify the clouds based on their vertical position and observe solar variability using the mean and standard deviation of the clear sky index for varying time intervals of 5, 15, 30, and 60 minutes. Time intervals affect the mean and standard deviation of the clear sky index differently for each cloud type: longer time intervals neglect small variability and overestimate the mean clear sky index of low and mid clouds, while high clouds do not change as much. The effect is also palpable when measuring ramps: the percentile 95 of the ramps obtained for 1 minute is 21 times greater compared to 1 hour. This ratio varies per cloud type, with the strongest differences occurring for mid clouds, having ramps that are 73 times stronger.

I. INTRODUCTION

The inherent variability of the solar resource is a big challenge for increasing renewable energy penetration in the electric grid[1]. Variability can occur at different timescales: seasonal changes, diurnal changes, or at very short timescales of minutes or seconds[2]. In the PV industry, historical hourly data is typically used to design a plant, meaning that the variability at the longer timescales is captured and that the behavior at shorter timescales is usually neglected. Short term variability matters both at a local scale, affecting the performance of electrical equipment, and at a global scale, affecting the electric grid's balance and economic dispatch when using hourly schedules[3], which makes variability a challenge for increasing PV penetration.

Quick changes in the solar resource are mainly caused by passing clouds. 'A cloud can diminish the solar irradiance that reaches the surface, or it can also augment in a process known as cloud enhancement, which occurs by forward scattering through thin clouds or on the sides of clouds[4,5]. The resulting variability is a compound effect of the optical properties of the cloud field, its spatial organization, its motion in space, as well as its own dynamics. Different types of clouds have distinct ways of evolving: some move with the wind without changing much, while others can either rise and grow, or dissipate in the span of an hour. Each location in the world has meteorological conditions that favor the existence of some clouds during the year, resulting in unique climatological records of cloudiness, thus of solar variability. Learning how each cloud type affects solar variability can facilitate a systematic analysis of cloud effects and expand it to other locations.

Previous works have studied the link between solar variability and cloud type. Hinkelman et al.[6] characterized solar ramps –the change of solar irradiance over a time interval– per cloud type. They used 1 min solar irradiance data from the SURFRAD network (continental US), and GOES satellite images with 30 min and 4 km spatial resolution to distinguish 12 cloud categories. They found that the features of the ramps are characteristic for each cloud type, leading to overall differences in each site due to the different frequency that they can display. Reno et al.[7] used a GOES satellite product (GSIP) with 6 cloud types and hourly resolution, and 1-minute solar irradiance for 2 sites in the US, creating hourly statistics of average and standard deviation of the clear sky index. They found that different cloud types correspond to distinct variability features and ramp rates. Lohmann et al.[8] characterized solar variability not only in time but also in space using a network of sensors in Germany, and classified sky conditions as clear, overcast or mixed using the average and standard deviation of the clear sky index. They found that mixed conditions were linked to more variability and stronger ramps.

The aforementioned studies have similar conclusions but since the cloud classes differ, it is hard to compare the results in a quantitative way. Satellite products have improved but their weakest feature is resolution both in time and space; therefore, they prevent us from having more detailed information on local cloud features. Ground-based methods for observing cloud properties also exist, including derived products from sky imagers, ceilometers, radars, and lidar. Currently, one of the most complete products can be obtained from radars, as they have great time resolution and can distinguish different layers of clouds. Very recent work has used these type of products to demonstrate improvements in solar variability forecasts[9], with not much attention given to time resolution issue. Thus, exploring both satellite and ground products is important to complement our understanding of the link between cloud types and solar variability.

The present work explores the dependence of the calculation of statistics of solar variability to time intervals and cloud types, using 1 minute resolution data. The high temporal resolution will allow us to determine the impact of neglecting sub-hourly features. The data is from the ARM CACTI campaign in Argentina, where a wide variety of clouds exist. The paper is structured as follows: Section II describes the data and methods to calculate solar variability and the cloud classification, Section III presents the main findings and analyzes the effect of time resolution, and Section IV contains the conclusions.

II. DATA AND METHODS

A. Data

We use data from the mobile ARM CACTI (Cloud, Aerosol, and Complex Terrain Interactions) campaign in Córdoba, Argentina (32,12 ºS, 64.73º W) which was deployed during 2018-2019. This location was chosen by ARM due to its unique features, which display a large variety of cloud types: “orographic boundary layer clouds, deep convection, and some mesoscale systems uniquely observable from a single fixed site”[10]. The unique variety of cloud conditions makes it interesting not only for atmospheric research but also for studies on solar variability.

For the solar resource, we retrieve the global horizontal irradiance (GHI) from the surface radiation product QCRAD1LONG[11], which is available at 1 min resolution from 2018/9/23 to

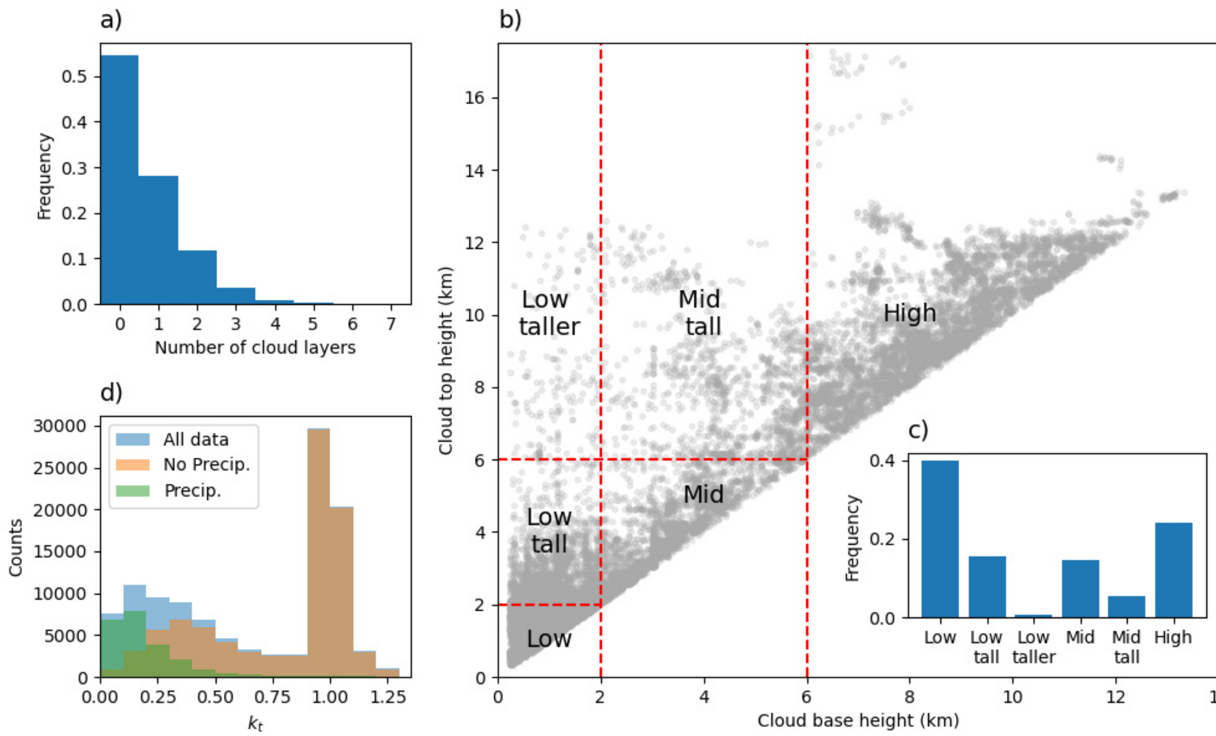

FIG. 1. Statistics of the dataset and cloud classification scheme: a) histogram of the number of cloud layers, with a maximum of 7, b) cloud base and top heights of the observations of single cloud layers (gray points), with a visualization of the cloud classification (dashed red lines), c) histogram of cloud types, and d) the histogram of the clear sky index k_t (ratio between measured and clear sky global horizontal irradiance), emphasizing the effect of the cases labeled as precipitating events.

2019/5/1. For the cloud properties, we use the ARSCLKAZRBND1KOLLIAS product, derived from radar and micropulse lidar results[12], which can recognize up to 10 layers of clouds, detecting base and top heights with a resolution of 4 s. Due to the mismatch in time resolution, we downsample the cloud data to 1 minute using the nearest reading available.

B. Cloud classification

For each time, there can be be a number of cloud layers. Fig. 1a shows the histogram of the number of cloud layers present for the times considered in the study, where clear sky conditions and single cloud layers dominate.

The cases with a single cloud layer are classified into 6 categories based on its cloud vertical position (Fig. 1b). Both cloud base and vertical extent are considered: low, low tall, and low taller all have cloud base heights lower than 2 km but the latter 2 have top heights greater than 2 or 6 km, respectively. Similarly, mid and mid tall clouds have cloud base heights lower than 6 km but the latter has top heights greater than 6 km. Finally, high clouds correspond to cloud base heights greater than 6 km. Fig. 1c shows that the least common type are low taller clouds, which could correspond to extremely tall convective clouds like cumulunimbus. In this work, the classification is restricted to single cloud layers as they are the most common in this dataset and because multiple layers pose a greater challenge for a systematic analysis.

Precipitation events can lead to incorrect readings of cloud base heights. Measured precipitation rates are available for this site but at an hourly rate which is not useful for our sub-hourly analysis. Therefore, we discard all events with readings of cloud base heights at the surface level. Even though this selection may also leave fog events out, we do not expect their exclusion to impact our results greatly. Fig. 1d shows the histogram for k_t, where the events labeled as precipitation have a very low k_t, indicating that they may indeed correspond to thick precipitating clouds.

C. Clear sky index and rolling statistics

We use the implementation of the Perez clear sky model in pvlib[13,14] to obtain a clear sky irradiance GHI_{cs}. We then compute the clear sky index $k_t = \mathrm{GHI}/\mathrm{GHI}_{cs}$. Daily values of Linke turbidity are found iteratively by finding a value that sets the clear sky index closer to 1 for the clear portions of the day.

Due to errors in the clear sky model near sunrise and sunset, when the irradiance is low, k_t can reach unrealistically high values. Therefore, we exclude the times with solar elevation lower than 20°, which is a common practice[15]. In our case, while this results in an estimated loss of 20% of the daytime data, there are still more than 6 hours available for a day in May, and since the removed hours are associated with lower energy generation their impact is not significant.

Since the dataset has 1 minute resolution, we are able to compute rolling statistics of ${k_t}^2$ for longer time intervals (Δt) of 5, 15, 30, and 60 min. We note that there is no resampling, meaning that all calculations use the 1 minute data. The average clear sky index, $\overline{k_t}$, and its standard deviation, σ_{k_t}, are calculated at the center of each time interval Δt with a trapezoidal method, as

$$\overline{k}_{t,i}(\Delta t) = \frac{1}{\Delta t} \sum_{j=i-n/2+1}^{i+n/2} \frac{k_{t,j-1} + k_{t,j}}{2} \delta t \text{ , and} \tag{1}$$

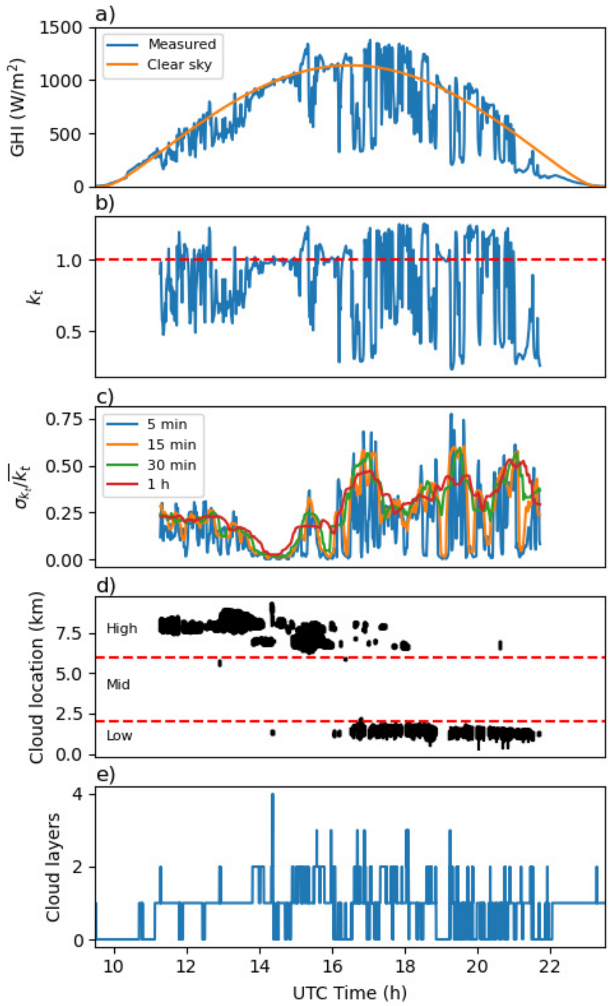

FIG. 2. Time series for January 21, 2019: a) measured and clear sky GHI at 1 min resolution, b) instantaneous clear sky index, c) normalized standard deviation of the clear sky index for different time intervals, d) the vertical position of passing clouds (shaded areas cover cloud base to cloud top heights), and e) number of cloud layers. Note that data with elevation greater than 20° has been omitted in b,c, and d).

$$\sigma^2_{k_t,i}(\Delta t) = \frac{1}{\Delta t^2} \sum_{j=i-n/2+1}^{i+n/2} \frac{1}{2}\left((k_{t,j-1} - \overline{k}_{t,j-1}(\Delta t))^2 + (k_{t,j} - \overline{k}_{t,j}(\Delta t))^2\right)\delta t\ , \qquad (2)$$

where i is the discrete time index, δt is the data time resolution, and $n = \Delta t/\delta t$.

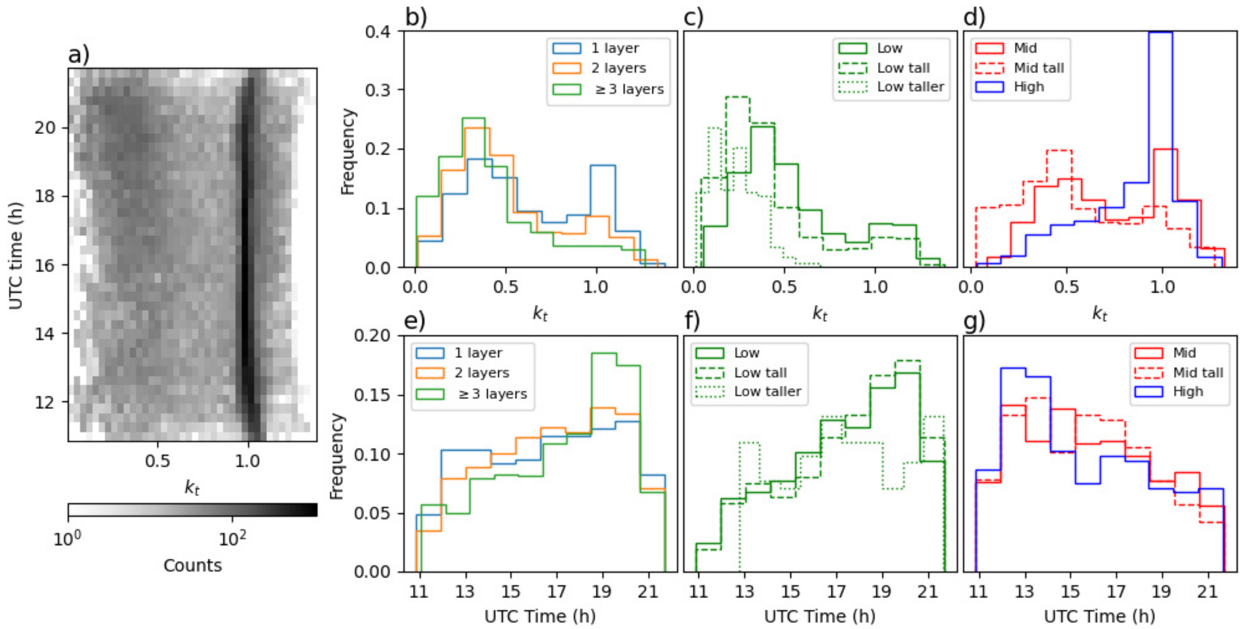

FIG. 3. Overall clear sky index statistics: a) shows the distribution of k_t over time in the day. The distribution of k_t and time in the day is then separated per number of cloud layers (b,e), and by cloud categories (c,d,f,g).

III. RESULTS

A. Sample day readings

Fig. 2 shows the solar and cloud readings for January 21, 2019. This day had high clouds passing in the morning and low clouds in the afternoon, with up to 4 layers of clouds, and no precipitation (Fig. 2d,e). In this case, there is a noticeable difference: low clouds induce a stronger variability than high clouds. This effect is evident in Fig. 2a,b,c, with the normalized standard deviation $\sigma_{k_t}/\overline{k}_t$ being higher in the afternoon for all the time intervals considered. There is also a time window of nearly clear sky situations ($k_t \approx 1$) around 14:00 UTC, which corresponds to a period of scattered and presumably optically thinner high clouds (Fig. 2d).

B. Overview of site conditions

The observed distribution of k_t at the site is bimodal (Fig. 1d), with a predominance of near clear conditions ($k_t \approx 1$). This is also evidenced in Fig. 1a, where a 55% of the times correspond to cloudless sky conditions. Regarding the number of cloud layers, a single layer is the most common (28.1%), followed by two (11.8%); three or more layers are less frequent (5.5%). Based on the cloud classification performed for the cases with a single cloud layer (Fig. 1b,c), low clouds had the highest frequency (39.8%), followed by high (23.9%), low tall (15.5%), mid (14.6%), mid tall (5.3%), and, lastly, low taller clouds (<1%).

Fig. 3a shows the distribution of k_t as a function of UTC time, with the counts in a logarithmic scale due to the frequent clear or near-clear sky conditions found in this dataset. For early mornings and late afternoons we see a slight increase of k_t, related to the errors that occur near sunrise and sunset, which were the reason to exclude lower elevation angles. There is a darker patch of lower k_t in the evening hours meaning that there are more clouds impacting solar radiation in the afternoon than in the morning. We can further explore these features by looking at the statistics of only cloudy events.

Fig. 3b,e shows the histograms of k_t and UTC time, respectively, for cloudy events and by the number of cloud layers. Only single cloud layers show a bimodal distribution for k_t, 2 layers or more display lower values of k_t. In terms of time, 3 or more layers are much more frequent in the evening, while 1 and 2 layers show a slightly flatter distribution along the day but still an increasing frequency with time in the day.

Finally, we can look the effect that each cloud type has on k_t and their frequency throughout the day (Fig. 3c,d,f,g). The highest values of k_t are linked mostly to high clouds and to some mid clouds. Mid tall and all low cloud types generally display lower values of k_t, with low taller clouds exhibiting the lowest k_t. The strong difference between low and high clouds is expected since the latter are often optically thinner[16]. High clouds are more frequent in the morning, while low and low tall clouds are more frequent in the evening, which could be a sign of them being surface-driven convective clouds such as Cumulus, which tend to develop in the afternoon when the surface heating is stronger. Low taller clouds seem to have a uniform distribution (but note the low number of samples), while mid and mid tall clouds are more frequent in the late morning and early afternoon.

C. Solar variability and time intervals

The variability measure σ_{k_t} depends on the time interval considered. We will analyze this dependence first by looking at the overall relationship between σ_{k_t} and $\bar{k}_t$, and then separating the analysis by cloud category.

Fig. 4 shows the overall effect of the time interval on the joint distribution of $\bar{k}_t$ and σ_{k_t}.

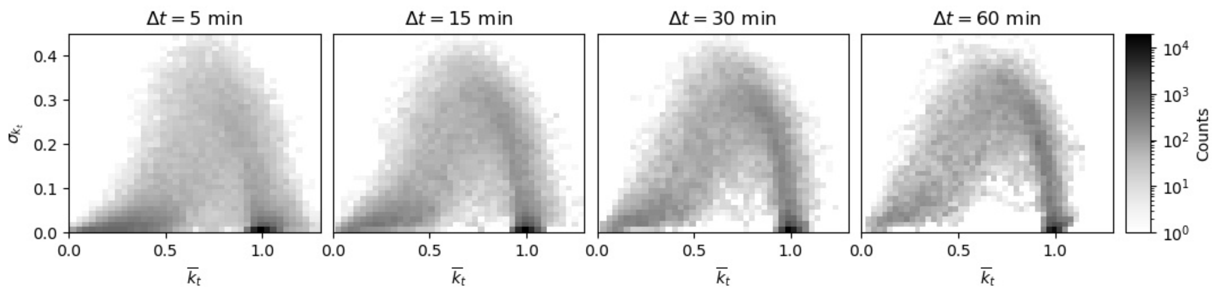

FIG. 4. Joint distribution of the clear sky index mean, $\overline{k}_t$, and standard deviation, σ_{k_t}, for different time intervals, Δt.

The time interval has a profound effect on both the mean and standard deviation of k_t: longer time windows diminish the maximum values, and by smoothing out the temporal scales, a more defined pattern is seen in the parameter space. In other words, some information is lost: by comparing the 5 and 60 minute statistics, the latter misses the concentration of events linked to small $\overline{k}_t$ and σ_{k_t}, overestimating solar variability, finding more moments close to clear sky instances ($\overline{k}_t \approx 1$), and completely missing events with middle values of $\overline{k}_t$ and low σ_{k_t}. While the effect is progressive with time interval, the differences between distributions are minor for 30 and 60 min. This solar variability conclusions are valid only for the site considered since the frequency of certain cloud types and clear days may affect different ranges of $\overline{k}_t$ and σ_{k_t}.

We now separate the analysis, observing the time interval effect for each cloud category. Fig. 5 shows the joint distributions of $\overline{k}_t$ and σ_{k_t} per cloud type and time window. First of all, statistics for the low taller clouds are included for completeness but the low number of samples does not give statistically significant results.

The segregated analysis allows confirming that different cloud types occupy different regions of the parameter space. When looking at 5 min statistics, low and low tall clouds have a stronger presence in the left bottom region with lower clear sky index and lower variability. The distribution progressively changes with greater time interval, reducing the frequency of cases with short term variability. While the maximum frequency stays within that region for 15 and 30 min statistics, it is lost for the 60 min case, resulting in a more uniform distribution of $\overline{k}_t$ and σ_{k_t}. For high clouds, the 5 min statistics shows a stronger presence at the right bottom corner, with low variability and $\overline{k}_t \approx 1$. Longer time windows increase the cases with more variability, likely through including the effect of nearby minutes. Overall, the distribution of high clouds is less affected by time interval, which confirms the tendency to be more spatially uniform. Lastly, mid and mid tall clouds occupy

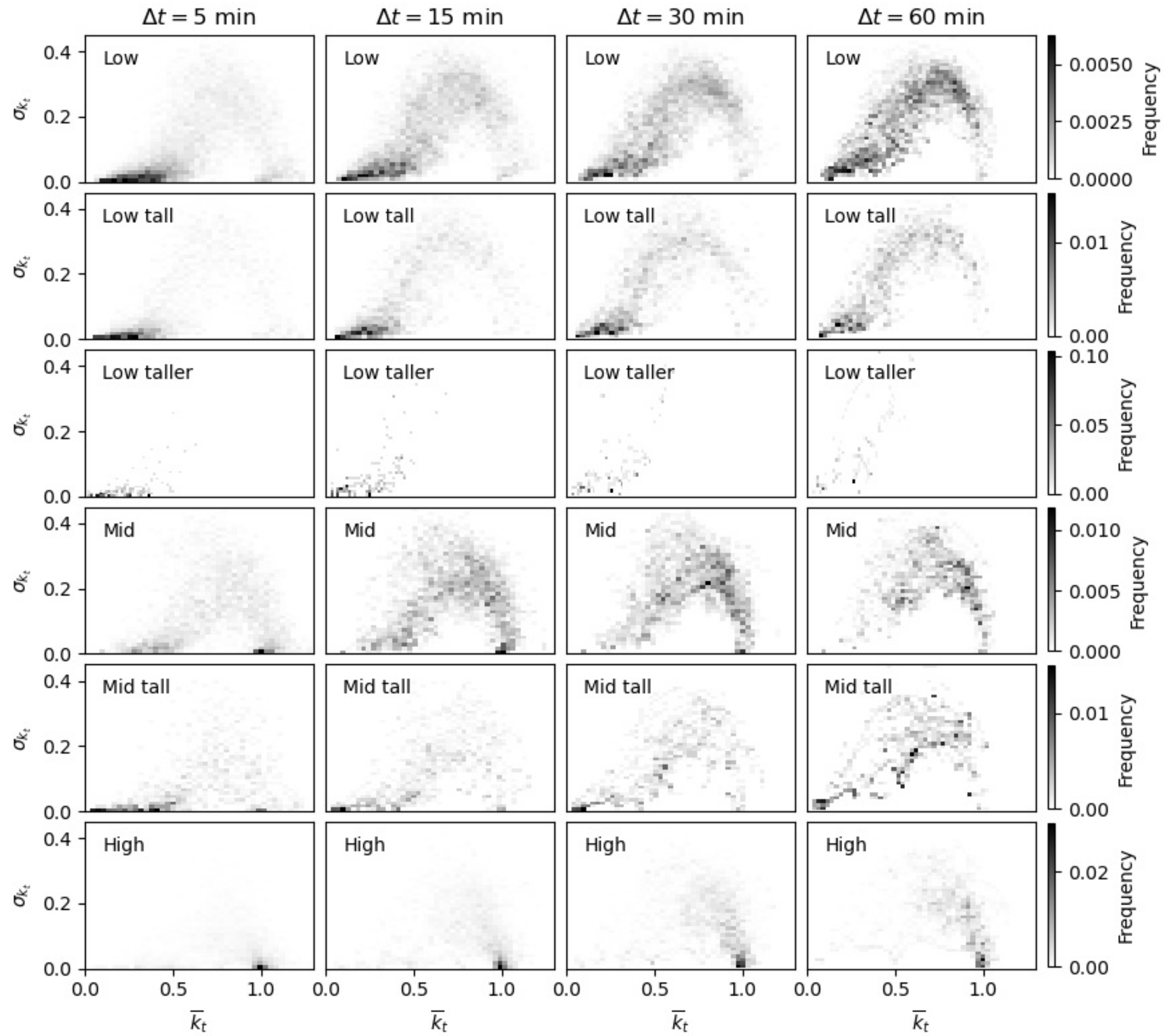

FIG. 5. Joint distributions of $\overline{k}_t$ and σ_{k_t} by cloud type (rows) and time interval (columns).

a broader range in the parameter space, with 5 minute statistics peaking in the left and right bottom corners. Similarly to low clouds, the peak regions are quickly lost with longer time intervals.

Summarizing, hourly intervals tend to overestimate variability because quick changes can be underrepresented. This effect is observed for all types of clouds but it affects low and mid clouds more strongly. Not only variability is misrepresented but also $\overline{k}_t$: it can be overestimated for low and mid cloud types. From a statistical point of view, the mean overestimation of clear sky index, comparing the hourly and 1 minute values, is not found to vary greatly across cloud types (no change for high clouds and 2-18% for the rest). However, a comparison based on means only is too simplistic as it neglects the fact that the distributions are not normal, with some displaying

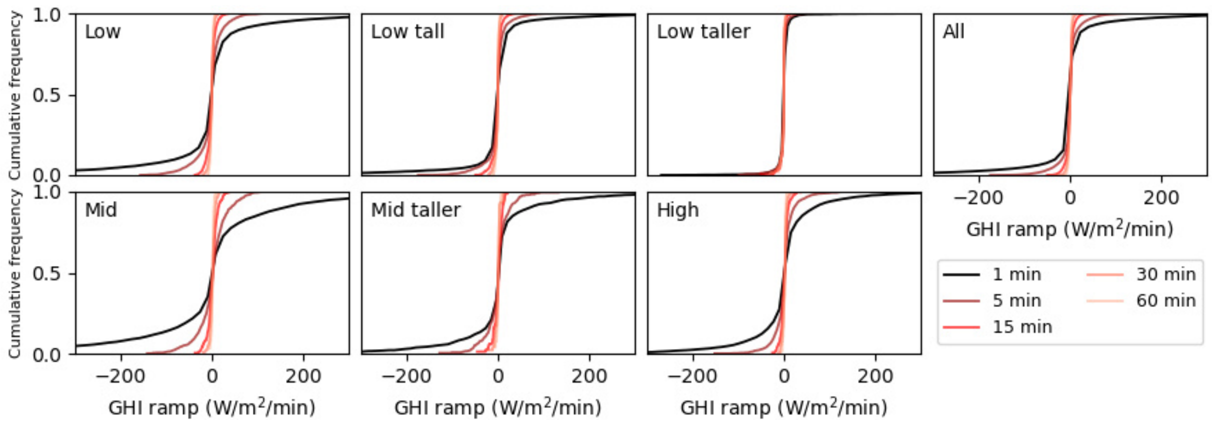

FIG. 6. Cumulative distribution of GHI ramps per cloud type and time resolution

a bimodal behavior (for full details of the distributions and tabulated means and medians see Appendix A).

D. Quantification of ramps

Fig. 6 shows the cumulative distribution of GHI ramps, which has been typically reported in previous studies[6]. At a first glance, both time intervals and cloud type matter. On the one hand, the effect of time is the same in all cases, longer time intervals underestimate the strength of the ramps because it smooths out the quick sub-hourly features. On the other hand, we see that the strongest ramps are linked to mid clouds, followed by mid taller, then high and low clouds.

We can further quantify the effect by comparing the percentiles 5 and 95, representing extreme values, shown in Table I. Longer intervals can greatly underestimate the strength of GHI ramps, with 1 minute ramps being up to 73 times stronger for mid clouds (p5), 37 times for low clouds (p95), and 31 times for high clouds (p5). The frequency of clouds and clear days at each particular site will determine the overall ramp distribution. For this case, the overall extreme ramps are 21 times stronger for 1 minute intervals when compared to hourly ones.

IV. CONCLUSIONS AND FUTURE WORK

We have analyzed the sub-hourly features of solar variability and its relationship to cloud type and time intervals, using 1 minute solar data and cloud radar products from the mobile ARM campaign CACTI in Argentina. Single layer clouds were classified by their vertical position, and

Time	All		Low		Low tall		Low taller		Mid		Mid tall		High	
window	p5	p95	p5	p95	p5	p95	p5	p95	p5	p95	p5	p95	p5	p95
1 min	-95.5	93.5	-192.6	164.6	-64.8	62.4	-13.7	11.6	-286.9	261.4	-133.2	143.6	-117.0	117.6
5 min	-35.8	33.7	-57.6	40.3	-43.8	29.1	-12.9	8.3	-60.2	59.8	-50.0	42.4	-34.7	38.2
15 min	-14.5	12.5	-18.1	13.5	-18.0	11.2	-8.1	3.7	-19.8	21.1	-18.7	13.4	-13.0	14.2
30 min	-7.8	6.4	-9.4	6.3	-10.2	4.7	-12.0	2.7	-7.1	10.0	-8.0	9.9	-6.7	7.7
60 min	-4.8	4.3	-5.7	4.4	-7.2	3.3	-7.9	1.5	-3.9	5.4	-4.4	5.5	-3.7	5.2

TABLE I. Statistics of the GHI ramps in W m^{-2} min^{-1}, per cloud type and per time interval, where p5 and p95 correspond to percentiles 5 and 95 of the observed data.

time intervals of 5, 15, 30 and 60 minutes were used to compute rolling statistics of the clear sky index (the mean $\overline{k}_t$ and the standard deviation σ_{k_t}).

This site shows a majority of clear conditions, followed by single clear layers. Each cloud type is associated with a different distribution of clear sky index and time of the day.

Solar variability was studied through the joint distribution of $\overline{k}_t$ and σ_{k_t}, finding that the choice of time interval profoundly affects the distribution for each cloud type. Longer time windows overestimate σ_{k_t} because they underrepresent the smaller scale dynamics, and they also overestimate $\overline{k}_t$ for the low and mid clouds, while high clouds properties are not as affected by temporal aggregation. Secondly, we quantified the change in GHI ramps with time interval and for each cloud type. The extreme values, quantified by percentiles 5 and 95, decrease with longer intervals, as expected but the effect varies per cloud type. Mid clouds generate the strongest ramps, with extreme ramps that are 71 times stronger when comparing 1 min data to hourly data. For comparison, the same ratio for all sky conditions is 19 times.

While this work did not resample the data to reproduce variability statistics based on hourly data, we expect those statistics to also overestimate both $\overline{k}_t$ and σ_{k_t}, probably at higher rates since the time intervals will not be shorter than 1 h and quick changes are filtered out. As more cloud products with high resolution become available, future work should aim to improve cloud classifications and other variables such as cloud optical thickness at finer time resolutions. The effects of multi layered clouds have also been left for future work, as more data would be preferred for statistical approaches. Finally, as the work by Riihimaki et al.[9] has shown, solar variability forecasting should be pursued with different techniques and cloud characterization methods.

ACKNOWLEDGMENTS

MZZ thanks the Faculty of Physical and Mathematical Sciences at Universidad de Chile, for a faculty incorporation grant.

DATA AVAILABILITY STATEMENT

Cloud and solar data are available at the ARM website: `https://adc.arm.gov/discovery/#/results/id::corarsclkazrbnd1kolliasM1.c1_cloud_base_best_estimate_macro_kazrarscl_cloud?dataLevel=c1&showDetails=true` and `https://adc.arm.gov/discovery/#/results/id::corqcrad1longM1.c2_BestEstimate_down_short_hemisp_lwbroad_qcrad_radio?dataLevel=c2&showDetails=true`, respectively. The code used in this study is available at `https://github.com/mzamora/SolarVarCACTI`.

Appendix A: Quantifying the overestimation of irradiance and variability

We have seen that for some cloud types, coarser time windows can lead to greater $\overline{k}_t$ and σ_{k_t}. In order to report the mean overestimation, we show in Fig. 7 the marginal distributions of k_t and σ_{k_t} per cloud type and time window, and in Table II the corresponding mean and median values. In a broad sense, we find that the mean k_t overestimation varies 2%-18% in all cloud types but the high clouds, for which there is no change. Nevertheless, many distributions are not normal and even bimodal and consequently, the mean values can misrepresent the changes in their distributions. Complementing with the median values, these are found to vary in a broader range (3%-29%), with a decrease occurring only for for high clouds, of 4%. For σ_{k_t}, the changes between 5 min and hourly windows are greater, with mean / median values being at least 2 / 2.5 times greater.

REFERENCES

[1]P. D. Lund, J. Byrne, R. Haas, and D. Flynn, *Advances in energy systems: The large-scale renewable energy integration challenge* (John Wiley & Sons, 2019).

[2]R. Perez, M. David, T. E. Hoff, M. Jamaly, S. Kivalov, J. Kleissl, P. Lauret, M. Perez, *et al.*, *Spatial and temporal variability of solar energy* (Now Publishers Incorporated, 2016).

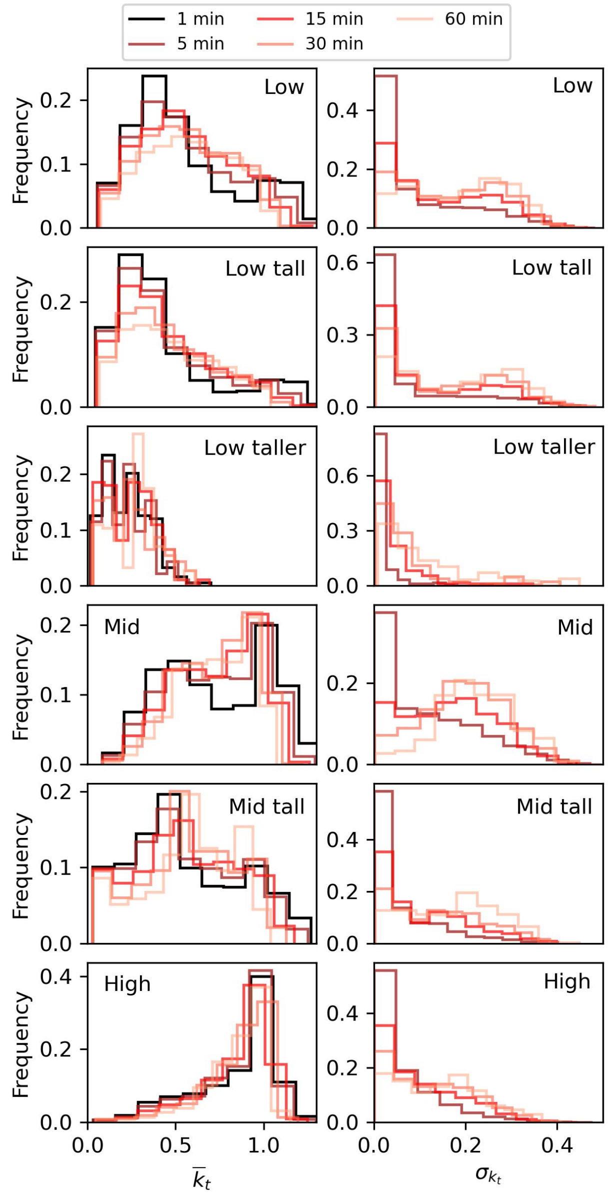

FIG. 7. Distributions of $\overline{k}_t$ and σ_{k_t} by cloud type and time window

[3]A. Mills and R. Wiser, "Implications of wide-area geographic diversity for short-term variability of solar power," Tech. Rep. (2010).

[4]Z. K. Pecenak, F. A. Mejia, B. Kurtz, A. Evan, and J. Kleissl, "Simulating irradiance enhancement dependence on cloud optical depth and solar zenith angle," Solar Energy **136**, 675–681 (2016).

[5]G. H. Yordanov, "A study of extreme overirradiance events for solar energy applications using nasa's i3rc monte carlo radiative transfer model," Solar Energy **122**, 954–965 (2015).

	Mean $\bar{k}_t$					Mean σ_{k_t}			
Cloud type	1 min	5 min	15 min	30 min	1 h	5 min	15 min	30 min	1 h
Low	0.55	0.55	0.56	0.56	0.57	0.10	0.14	0.17	0.19
Low tall	0.43	0.43	0.44	0.45	0.47	0.08	0.12	0.15	0.18
Low taller	0.22	0.23	0.24	0.25	0.27	0.02	0.05	0.08	0.12
Mid	0.73	0.73	0.73	0.73	0.72	0.12	0.17	0.20	0.22
Mid tall	0.56	0.55	0.56	0.56	0.57	0.06	0.11	0.14	0.17
High	0.84	0.84	0.84	0.84	0.84	0.07	0.10	0.12	0.14

	Median $\bar{k}_t$					Median σ_{k_t}			
Cloud type	1 min	5 min	15 min	30 min	1 h	5 min	15 min	30 min	1 h
Low	0.46	0.50	0.53	0.55	0.56	0.04	0.12	0.16	0.20
Low tall	0.34	0.35	0.38	0.41	0.44	0.02	0.07	0.13	0.19
Low taller	0.22	0.23	0.25	0.25	0.27	0.01	0.03	0.05	0.08
Mid	0.72	0.76	0.77	0.76	0.74	0.09	0.17	0.20	0.22
Mid tall	0.49	0.51	0.54	0.56	0.60	0.03	0.07	0.14	0.19
High	0.94	0.93	0.91	0.91	0.90	0.04	0.08	0.12	0.15

TABLE II. Mean and median values of $\bar{k}_t$ and σ_{k_t} by cloud type and time window

[6] L. M. Hinkelman, A. Heidinger, M. Sengupta, and A. Habte, "Relating solar resource variability to cloud type," (2013).

[7] M. J. Reno and J. S. Stein, "Using Cloud Classification to Model Solar Variability," , 20 (2013).

[8] G. M. Lohmann, A. H. Monahan, and D. Heinemann, "Local short-term variability in solar irradiance," Atmospheric Chemistry and Physics **16**, 6365–6379 (2016).

[9] L. D. Riihimaki, X. Li, Z. Hou, and L. K. Berg, "Improving prediction of surface solar irradiance variability by integrating observed cloud characteristics and machine learning," Solar Energy **225**, 275–285 (2021).

[10] A. Varble, S. Nesbitt, P. Salio, E. Avila, P. Borque, P. DeMott, G. McFarquhar, S. van den Heever, E. Zipser, D. Gochis, R. Houze, M. Jensen, P. Kollias, S. Kreidenweis, R. Leung, K. Rasmussen, D. Romps, C. Williams, and PNNL, BNL, ANL, ORNL, "Cloud, Aerosol, and Complex Terrain Interactions (CACTI) Field Campaign Report," Tech. Rep. DOE/SC-ARM-19-028, 1574024 (2019).

[11]L. Riihimaki, Y. Shi, and D. Zhang, "Data quality assessment for arm radiation data (qcrad1long),".

[12]K. Johnson and M. Jensen, "Active remote sensing of clouds (arscl) product using ka-band arm zenith radars (arsclkazrbnd1kollias),".

[13]P. Ineichen and R. Perez, "A new airmass independent formulation for the Linke turbidity coefficient," Solar Energy **73**, 151–157 (2002).

[14]W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlib python: a python package for modeling solar energy systems," Journal of Open Source Software **3**, 884 (2018).

[15]N. Engerer and F. Mills, "Validating nine clear sky radiation models in australia," Solar Energy **120**, 9–24 (2015).

[16]R. A. Houze, "Chapter 1 - types of clouds in earth's atmosphere," in *Cloud Dynamics*, International Geophysics, Vol. 104, edited by R. A. Houze (Academic Press, 2014) pp. 3–23.

[17]M. Schroedter-Homscheidt, M. Kosmale, S. Jung, and J. Kleissl, "Classifying ground-measured 1 minute temporal variability within hourly intervals for direct normal irradiances," Meteorologische Zeitschrift **27**, 161–179 (2018).