Importance of Time Averaging Conventions

Daniel Perez-Astudillo[1] and Dunia Bachour[1]

[1] Qatar Environment and Energy Research Institute, HBKU, Qatar Foundation, Doha (Qatar)

Abstract

Original or raw solar radiation data is preferably created at relatively high temporal resolutions, such as one-minute values for direct measurements. However, and especially when combining different sources or with different kinds of data, the data actually seen by end users is frequently already averaged at some other temporal resolution, commonly 15-, 30- or 60-minute averages. One also commonly missing piece of information is how those averages were obtained; in particular, what the timestamps actually mean; for example, does the hourly value marked as 10 am include 9:01 to 10:00, or 10:00 to 10:50, or something else? This study presents an estimation of errors, using real measurements, resulting from a user assigning the 'wrong' interval to the timestamps, at different averaging scales, from 5 to 60 minutes.

Keywords: solar resources, temporal resolution, solar data aggregation

1. Introduction

Ground-level data of solar radiation is the main input for solar resource assessment, modelling and solar power related studies and applications; these data, called 'solar data' for short from here on, provide the basis on which the results of those activities are built, and can have hefty financial impacts, for instance when those results are used at the level of large-scale, commercial solar power plants. Usually, solar data has two origins: from direct measurements taken by monitoring stations, ideally collecting measurements as 1-minute averages, or from satellite-derived observations, for which currently the widest coverages are done with a resolution of 15 minutes (instantaneous observations every 15 minutes for any particular site). Since monitoring stations only provide accurate data during their time of operation and only for a relatively limited spatial radius (of the order of 20-30 km in the best cases) around their location, a common practice includes a combination with other sources of data. In addition, the generator of the data sets is not always also the end user of the data, so end users generally receive data sets from a third party, usually after some filtering and temporal averaging has been made.

In ideal cases, end users receive not only the time series of irradiances or irradiations, but also some additional information on how the time series was created. However, this is not always the case, and even when some details are provided, a couple of items are frequently missing:

- Whether the values represent the 'instant' given by the timestamp or an average between timestamps.
- In the case of averages, which time interval is represented by the average, e.g. the interval since the previous timestamp, or from the current to the next one?

In the case of hourly averages, for instance, some models provide values using solar time, so each 'hour' can fall in the middle of the hours in local time.

Clearly, the user should ensure that enough information is provided. But when the information is unavailable, or when the user simply assumes that the timestamps of the received data are consistent with the user's own timestamp convention, the effects of this misuse can be important, and a quantification is presented here.

2. Methodology

Ground-level measurements collected at one-minute resolution during two years at one site were used here. For a more complete analysis, the three 'components' are studied, that is, direct normal irradiance G_b, global horizontal

doi:10.18086/swc.2019.42.07 Available at http://proceedings.ises.org

irradiance G, and diffuse horizontal irradiance G_d. The station is located in Doha, Qatar (25.33 N, 51.43 E) (Perez-Astudillo and Bachour, 2014). Qatar has a desert climate, with very hot and humid summers, and mild winters with some cloudiness and sporadic rain. Although clouds do not play an important role all year in Qatar, aerosols are a varying factor through the year, being reflected mainly in variations in the direct irradiance, which does not show a clear seasonal dependency in Doha.

The one-minute measurements are then quality-flagged following BSRN tests (Long and Dutton, 2002), and then averaged at different 'Nmin' timescales, namely Nmin = 5, 10, 15, 30 and 60 (hourly) minutes. The averages for any given interval are done with the entries that did not fail the quality checks (QC); if more than 50% of the entries in an interval did not pass QC (missing entries are counted as no pass), the average for the interval is reported as missing.

To quantify the errors induced by wrong timestamp assignment, two sets of averages were obtained:

- 'PRE': the averaged value represents the average of the previous Nmin minutes, including the minute of the timestamp. For instance, the 10-minute averages at 10:50 are obtained from the minutes 10:41 to 10:50, both inclusive. Hourly averages at hour = 10 include the minutes from 09:01 to 10:00.
- 'POST': the averaged value represents the average of the next Nmin minutes, including the minute of the timestamp. For instance, the 10-minute averages at 10:50 are obtained from the minutes 10:50 to 10:59, both inclusive. Hourly averages at hour = 10 include the minutes from 10:00 to 10:59.

Then, using the PRE dataset as reference, the differences between both are calculated, by pairing entries with the same timestamp, in terms of mean bias difference (MBD), mean absolute difference (MAD), root-mean-square difference (RMSD):

$$MBD = \frac{1}{n} \sum_i \left(x_{post,i} - x_{pre,i}\right) \qquad \text{(eq.1)}$$

$$MAD = \frac{1}{n} \sum_i \left|x_{post,i} - x_{pre,i}\right| \qquad \text{(eq.2)}$$

$$RMSD = \sqrt{\sum_i \frac{\left(x_{post,i} - x_{pre,i}\right)^2}{n}} \qquad \text{(eq.3)}$$

where i runs from 1 to n, the total number of included data pairs (with the same timestamp) $x_{pre,i}$ and $x_{post,i}$ which are the irradiance values from the PRE and POST datasets, respectively. The relative values rMBD, rMAD and rRMSD are obtained by dividing each of the above by the average of the PRE values.

3. Results

3.1. Irradiance profiles

The main effect that one can expect to see in this comparison is a shift in the irradiance profiles. Figure 1 shows a sample profile of the beam irradiance, G_b, during a cloudy day, at various time averaging intervals, from 5 minutes to 1 hour. The shift is more clearly visible at large time averagings; indeed, a time-shifted profile of hourly values is relatively easy to spot because sunrise and sunset, as well as the peak in irradiance, can be clearly seen away from their expected times. As the averaging periods get smaller, however, visual inspections become less effective in identifying time shifts; even small asymmetries between morning and afternoon irradiance profiles don't necessarily signal a time shift, since they may be attributed to different atmospheric conditions, depending on the location.

While the shift is visible, it is not the only effect on these profiles since, as defined in the previous section, the averages (when comparing an entry in PRE with the 'previous' entry in POST) differ in two minutes, namely, the first and the last. This has a larger effect at small intervals, due to the smaller number of included values, but as the averaging interval sizes increase up to an hour, the effect of these two minutes is further reduced, and the shift is more important, as previously discussed.

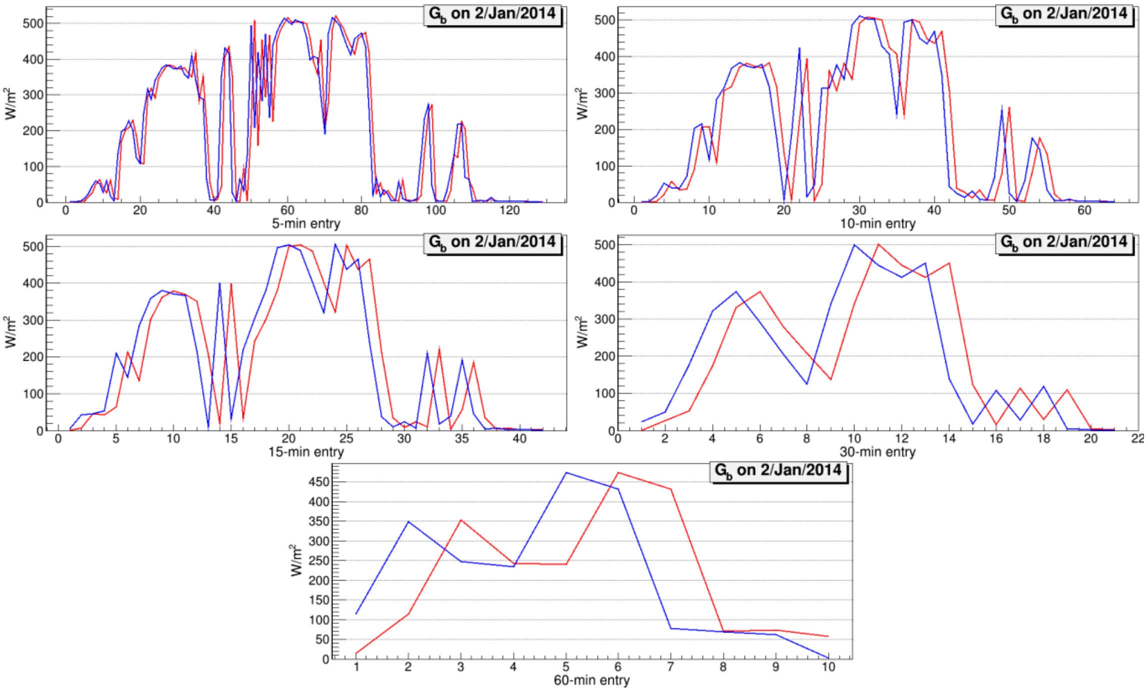

Fig. 1: Comparisons at different time averaging intervals of direct normal irradiance during one day, at daytime, from the PRE (red) and POST (blue) datasets. The averages are not only time-shifted but also slightly changed, especially at small time intervals and under unstable conditions.

3.2. Statistical comparisons

Table 1 shows the mean bias, mean absolute bias, and root-mean-square differences (all in W/m^2), together with their relative values (in %) between the POST and PRE datasets for G_b, G and G_d, for the two years under study. Note that only daytime values are included in these comparisons. In addition, since the QC formulas use the solar zenith angle (SZA) at the considered times, the SZA for each dataset was averaged in the same way as the irradiances, and therefore, sunrise and sunset happen at slightly different timestamps between the datasets; a pair of entries was included here only if SZA in both datasets was under 90 degrees at that timestamp.

Tab. 1: Results of statistical differences in POST averages with respect to PRE averages, over two years (only daytime entries considered). MBD, MAD and RMSD are in W/m^2; rMBD, rMAD and rRMSD are in %.

	MBD	rMBD	MAD	rMAD	RMSD	rRMSD
5 min						
G_b	-0.12	-0.03	18.32	4.16	41.46	9.42
G	-0.10	-0.02	16.97	3.43	34.31	6.94
G_d	-0.04	-0.02	6.34	3.37	12.55	6.66
10 min						
G_b	-0.31	-0.07	28.84	6.49	52.98	11.92
G	-0.25	-0.05	28.94	5.80	43.91	8.80
G_d	-0.12	-0.06	11.11	5.86	18.80	9.91
15 min						
G_b	-0.50	-0.11	38.25	8.55	62.95	14.07
G	-0.40	-0.08	40.18	8.00	53.19	10.59
G_d	-0.20	-0.10	15.33	8.03	23.86	12.50
30 min						
G_b	-1.00	-0.22	64.57	14.09	92.70	20.22
G	-0.96	-0.19	73.95	14.37	85.69	16.65

G_d	-0.55	-0.28	26.48	13.55	36.19	18.52
1 hour						
G_b	0.74	0.16	111.18	23.62	148.56	31.56
G	-0.45	-0.09	141.03	26.52	157.40	29.59
G_d	-0.21	-0.10	46.43	23.14	59.66	29.73

The relative values from Table 1 are plotted in Fig. 2, for all the different time averaging sizes. Since the mean bias only accounts for the difference in the irradiance averages obtained with both datasets, the time shift effects are mostly cancelled out, resulting in small bias values only due to the different averaged values; the bias increases at the larger intervals due to the different averages (between datasets) obtained near sunrise and sunset. The time shift effects can be seen more clearly on MAD and RMSD, where even at 5 minutes the differences are already above 3% and reach around 30% at hourly level.

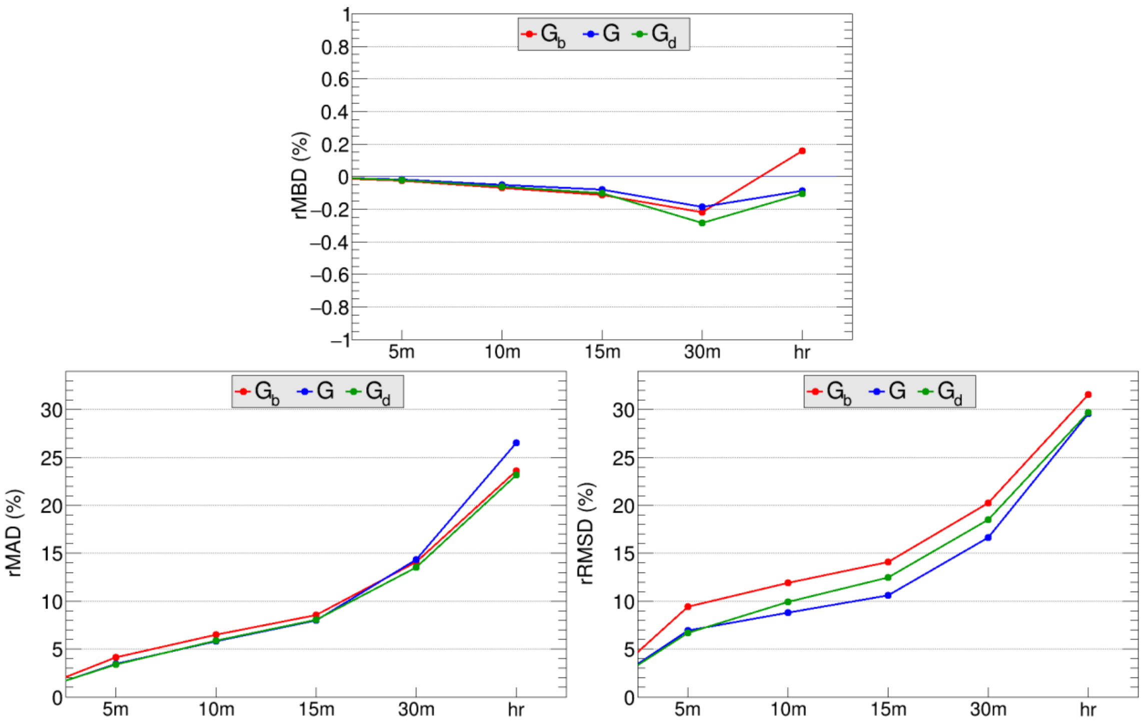

Fig.2: Relative MBD, MAD, and RMSD of the POST averages with respect to the PRE averages at different time scales, for two years of measurements. Only daytime entries are considered.

4. Conclusions

Time series of solar radiation are the foundation of not only solar resource studies but also different applications, from modelling to 'real-world' systems, such as commercial solar power plants. It is thus critical, when dealing with solar radiation measurements, to eliminate or reduce as much as possible all sources of errors that may negatively affect the data along the way from sensors to final products. One aspect that does not often receives a high importance, is that when a user receives a dataset that includes time-averaged values, the definition of the averaged interval is not always provided, or clear, and the user may either have to figure out the convention, or without much thought just assume that the timestamp convention is the same used by his/her own data, or by other datasets at his/her disposal. The study presented here shows a quantification of the effect that an incorrect assumption of the averaged periods can have on a time series, by using two years of data collected at one minute resolution, and averaged at different time scales up to hourly, with two different averaging conventions. The results obtained here show that the errors are already noticeable at the short averaging intervals: at 5 minutes, the mean absolute bias and RMSD are already over 3% and 6%, respectively, while at hourly averages these errors have increased to over 26% and 31%, respectively. These results were obtained for Doha, where short-term irradiance is relatively stable during about half of the year due to low cloudiness. Mean bias results, as discussed in the text, should not be greatly affected by higher radiation variability. However, MAD and RMSD, which contain contributions from each averaged interval, will likely show larger values in highly variable conditions.

5. References

Perez-Astudillo, D., Bachour, D. DNI, GHI and DHI Ground Measurements in Doha, Qatar. Energy Procedia 49, 2398-2404. doi: 10.1016/j.egypro.2014.03.254

Long, C.N., Dutton, E.G., 2002. BSRN Global Network recommended QC tests, V2.0 Available online at https://epic.awi.de/id/eprint/30083/1/BSRN_recommended_QC_tests_V2.pdf Last accessed 30/Jul/2019