A Method for Clear-Sky Identification and Lone-Term Trends Assessment Using Daily Surface Solar Radiation Records

Journal Article

Author(s):
Ferreira Correa, Lucas; Folini, Doris; Chtirkova, Boriana; Wild, Martin

Publication date:
2022-08

Permanent link:
https://doi.org/10.3929/ethz-b-000566947

Originally published in:
Earth and Space Science 9(8), https://doi.org/10.1029/2021EA002197

Funding acknowledgement:
188601 - Towards an improved understanding of the Global Energy Balance: causes of decadal changes of solar radiation (SNF)

AGU ADVANCING EARTH AND SPACE SCIENCE

Earth and Space Science

RESEARCH ARTICLE
10.1029/2021EA002197

A Method for Clear-Sky Identification and Long-Term Trends Assessment Using Daily Surface Solar Radiation Records

Lucas Ferreira Correa[1], Doris Folini[1], Boriana Chtirkova[1], and Martin Wild[1]

[1]Institute for Atmospheric and Climate Sciences, ETH Zurich, Zurich, Switzerland

Key Points:

- A clear-sky detection method using only daily mean surface solar radiation is proposed and validated, with focus on decadal scale trends
- Method calibration requires in addition satellite cloud fraction data, to derive site and calendar month specific transmittance thresholds
- Method validation for 24 sites suggests its usefulness for deriving long term clear-sky trends worldwide, albeit with slight bias

Correspondence to:

L. F. Correa,
lucas.ferreira@env.ethz.ch

Citation:

Correa, L. F., Folini, D., Chtirkova, B., & Wild, M. (2022). A method for clear-sky identification and long-term trends assessment using daily surface solar radiation records. *Earth and Space Science*, *9*, e2021EA002197. https://doi.org/10.1029/2021EA002197

Received 22 DEC 2021
Accepted 25 JUL 2022

Author Contributions:

Conceptualization: Lucas Ferreira Correa
Formal analysis: Lucas Ferreira Correa, Doris Folini, Boriana Chtirkova, Martin Wild
Funding acquisition: Martin Wild
Investigation: Lucas Ferreira Correa, Doris Folini, Martin Wild
Methodology: Lucas Ferreira Correa
Project Administration: Martin Wild
Supervision: Doris Folini, Martin Wild
Validation: Lucas Ferreira Correa
Visualization: Lucas Ferreira Correa
Writing – original draft: Lucas Ferreira Correa
Writing – review & editing: Doris Folini, Boriana Chtirkova, Martin Wild

Abstract Time series of clear-sky irradiance are fundamental for the understanding of changes in the Earth Radiation budget, since they allow to examine radiative processes in the cloud-free atmosphere. Clear-sky data is usually derived from all-sky irradiances using one of several clear-sky methods proposed in the literature. However, most of the available clear-sky methods require additional in situ measurements and/or high temporal resolution (sub-daily), which restricts the derivation of clear-sky time series to a few well equipped stations. Here we propose a new clear-sky identification method that aims to overcome this problem, with the ultimate goal of deriving multidecadal clear-sky trends for many sites globally. The method uses site specific monthly transmittance thresholds to derive long term clear-sky time series for any station worldwide that has daily mean irradiance data. We exemplify the method for 24 stations. Transmittance thresholds are derived by combining 29 years (1990–2018) of satellite cloud cover data with in situ irradiance measurements. The thresholds are then applied to the whole time series (independent of satellite data availability) to screen out cloudy days. Comparison of our results with reference data derived using Long and Ackerman's (2000, https://doi.org/10.1029/2000jd900077) method shows good agreement after bias correction, especially for decadal trends. While limitations of the method, such as anomalies representation, are highlighted and discussed, validation results encourage its use to derive long term clear-sky time series and associated decadal-scale trends around the globe.

1. Introduction

The assessment of cloud-free (clear-sky) irradiance is of major importance for the study of radiative processes in the atmosphere, and it is yet a challenge. Such type of data is fundamental for screening out cloud effects in the radiative budget and visualizing the effects of gases and aerosols. This is key for understanding the role of cloud free radiative processes in climate. However, there is usually no procedure to store separately measurements taken under these conditions, and there is no universal method for deriving clear-sky irradiances from all-sky measurements.

Gueymard et al. (2019) have highlighted and evaluated 21 different methods proposed by several authors for clear-sky identification. These methods are usually based mostly on synoptic cloud data and on sky clarity indices (Younes & Muneer, 2007). Previous studies such as Ododo et al. (1996) and Younes and Muneer (2006) have evaluated the quality aspects involved in the usage of synoptic cloud cover as a clear-sky indicator. The usage of sunshine duration was also evaluated by previous studies (e.g., Barbaro et al., 1981). These methods are proven to be reliable and have been used for several applications in the last few decades (e.g., Manara et al., 2016; Sanchez-Lorenzo et al., 2009; Stjern et al., 2009; Yamasoe et al., 2021). Other methods proposed the use of transmittance, diffuse ratio and other indices for clear-sky identification (e.g., Perez et al., 1990; Renner et al., 2019; Thevenard & Brunger, 2001; Wild et al., 2021). These methods usually rely on the optical effects of clouds: having a high albedo, clouds play a major role in reflecting solar radiation. When this radiation is reflected back to space, the atmospheric transmittance is decreased, and when the radiation is scattered to the surface, the diffuse radiation increases. Thus, a cloud free scenario would be expected to have high transmittance and low diffuse ratio. However, there is not a clear agreement between authors on what would be the optimal thresholds for such indicators, nor on whether such universal optimal thresholds exist. In the context of sky clarity indices, the Long and Ackerman (2000) method may serve as a high quality reference. In this method the authors propose four tests to eliminate cloudy observations and extrapolate clear-sky data to fill these cloudy observations with clear-sky representative data. This method has been extensively used for clear-sky identification for its reliability, and has been used as a basis for the evaluation of different clear-sky methods (Renner et al., 2019; Reno & Hansen, 2016;

Younes & Muneer, 2007). The method relies on the availability of sub-daily observations, the typical time resolution being a few minutes.

Even though these methods are proven to be reliable within their limitations, and are extensively used for various applications in climate research, their applications still have two significant limitations: their temporal and spatial coverage. The existing clear-sky methods are usually dependent on additional and/or sub-daily station measurements. This implies that the clear-sky irradiances from any station can only be derived if the station has the additional measurements (diffuse radiation, synop cloud cover, etc.) necessary for such procedures or high temporal resolution (up to 30 min). The seasonal maximum transmittance method (Wild et al., 2021) is an exception to this dependence, and its use is compared to the use of the method proposed in this paper. In the case of the Long and Ackerman (2000) method the ideal temporal resolution is even higher, since this method requires minute data as input. This dependency is a limiting factor if one wants to achieve extensive global coverage and long term clear-sky data.

In this paper, we propose a new method for clear-sky identification, based on daily mean surface solar radiation observations. Additional data is only needed to calibrate the method once per site, when daily mean fractional cloud cover satellite data are used to define transmittance thresholds specific to calendar month and site. The method thus overcomes the necessity of additional and/or subdaily in situ observations for the derivation of a clear-sky time series. Consequently, it can be applied to any data base of existing long-term daily mean surface solar radiation measurement data from around the globe. Ultimately focusing on decadal clear-sky trends, some systematic bias of the method with respect to the absolute values of clear-sky irradiance is tolerable. In fact, based on validation results for 24 sites, we recommend to apply the method to study long-term variability of clear-sky irradiances rather than (daily) absolute values of clear-sky irradiances.

We define a clear-sky day as a day in which there is no significant influence of clouds on the total downward solar radiation at the surface. The identification of long term clear-sky trends, as aimed with this method, is fundamental for better understanding phenomena like Global Dimming and Brightening and its effects on the climate (Wild, 2009). The representation of inter annual variability is also discussed. A full description, validation and comparison to other methods, as well as the strengths and weaknesses of this method are presented in the following sections.

2. Data

In this work we have used daily surface global solar irradiance data from the World Radiation Data Centre (WRDC, Voeikov Main Geophysical Observatory, 2022) and from the Baseline Surface Radiation Network (BSRN, Driemel et al., 2018; Ohmura et al., 1998). There was a total of 24 stations, 23 from BSRN and one from WRDC. The BSRN stations were chosen due to availability of enough data to thoroughly validate the method. That is, long enough overlap with satellite cloud cover data (see below) and high time resolution (minutes) allowing the application of the Long and Ackerman (2000) clear-sky algorithm. The temporal resolution of the WRDC stations is not sufficient for clear-sky derivation using the Long and Ackerman (2000) method, which was used as a reference in the validation. Thus, the choice of the WRDC station in Madrid, Spain, was motivated by its climatic features and availability of synop cloud cover data. This station is much less cloudy than the other stations at which the synop cloud cover was analyzed (see Table 1), and this contrast between stations allows one to discuss the role of the cloudiness to the quality of the results (see Section 4.1). Data from BSRN undergoes the quality control tests as recommended by Long and Dutton (2002). WRDC data already have quality flags associated upon download. An additional consistency verification was performed in both datasets to remove any remaining doubtful value or inhomogeneous data. Any doubtful data have been removed (i.e., inhomogeneous data, physically impossible values). Global daily solar irradiance data was converted into transmittance via division by the incoming daily mean Top of Atmosphere (TOA) irradiance. TOA irradiance was calculated for each station with a simple model, in which we calculated the cosine of the solar zenith angle at every second from sunrise until sunset, combined it with the Sun-Earth distance (dependent on the day of the year), length of the day and the solar constant (1361 W/m^2) and finally integrated to obtain the daily incoming radiation at the TOA for a given day and location. Twenty nine years (1990–2018) of daily mean (daytime) satellite-derived cloud fraction data from CLARA-A2 (Cloud, Albedo and Radiation from AVHRR, Karlsson et al., 2017, 2020), version 2.0 (1990–2015) and version 2.1 (2016–2018), was also used. This product is gridded, has a global 0.25°

Table 1
Stations Used in the Validation Procedures

Station	Coordinates	Data set	T	Sat cc	CS days	RMSE (W/m2)	Bias (W/m2)
Lauder, New Zealand	45.04°S 169.69°E	BSRN	0.5202	56.7%	17.1%	6.4	−3.9
Alice Springs, Australia	23.80°S 133.89°E	BSRN	0.6650	31.5%	52.0%	4.9	−2.3
Darwin, Australia	12.42°S 130.89°E	BSRN	0.5842	42.1%	44.1%	15.9	−12.6
Momote, Papua New Guinea	2.06°S 147.42°E	BSRN	0.4843	72.7%	4.7%	11.4	−9.3
Kwajalein, Marshall Islands	8.72°N 167.73°E	BSRN	0.5583	65.1%	13.9%	11.3	−10.2
Tamanrasset, Algeria	22.79°N 5.52°E	BSRN	0.6876	32.2%	48.8%	11.6	−6.9
Goodwin Creek, Mississippi, USA	34.25°N 89.87°W	BSRN	0.5058	50.0%	34.0%	14.5	−9.2
Tateno, Japan	36.06°N 140.13°W	BSRN	0.4796	59.0%	26.0%	14.9	−10.2
Billings, Oklahoma, USA	36.60°N 97.52°W	BSRN	0.5464	48.1%	32.1%	6.8	−4.4
Southern Great Plains, Oklahoma, USA	36.60°N 97.48°W	BSRN	0.5654	48.4%	31.8%	6.9	−4.6
Desert Rock, Nevada, USA	36.63°N 116.02°W	BSRN	0.6751	31.7%	53.9%	4.3	−1.9
Chesapeake Light, North Atlantic Ocean, USA	36.90°N 75.71°W	BSRN	0.5456	53.1%	29.0%	9.8	−6.4
Bondville, Illinois, USA	40.06°N 88.36°W	BSRN	0.5159	60.4%	23.2%	11.1	−8.0
Boulder, Colorado, USA	40.12°N 105.24°W	BSRN	0.5791	51.8%	28.4%	12.5	−7.8
Madrid, Spain	40.45°N 3.73°W	WRDC	0.5743	47.6%	34.5%	[a]	[a]
Rock Springs, Pensylvannia, USA	40.72°N 77.93°W	BSRN	0.4559	63.1%	21.2%	15.0	−11.2
Carpentras, France	44.08°N 5.06°E	BSRN	0.5413	44.1%	39.9%	6.7	−4.4
Payerne, Switzerland	46.81°N 6.94°E	BSRN	0.4318	63.5%	20.4%	11.8	−9.0
Fort Peck, Montana, USA	48.32°N 105.10°W	BSRN	0.5429	57.4%	22.1%	7.3	−4.8
Regina, Canada	50.20°N 104.71°W	BSRN	0.5203	67.2%	24.9%	7.3	−5.2
Camborne, UK	50.22°N 5.32°W	BSRN	0.4049	68.3%	13.8%	14.3	−9.2
Lindenberg, Germany	52.21°N 14.12°E	BSRN	0.4093	66.0%	16.6%	12.0	−8.2
Toravere, Estonia	58.25°N 26.46°E	BSRN	0.3901	65.2%	14.5%	14.5	−10.2
Lerwick, UK	60.14°N 1.18°W	BSRN	0.3469	75.6%	6.1%	15.5	−11.1

Note. BSRN, Baseline Surface Radiation Network; WRDC, World Radiation Data Centre. Site average[b] transmittance (T); Site average[b] satellite cloud cover (Sat cc); Percentage of days flagged as clear-sky with our method (CS days); root-mean-square error when comparing daily mean clear-sky irradiance [in W/m^2] using Long and Ackerman's (2000) algorithm and daily mean clear-sky irradiance [in W/m^2] using our method[b] (RMSE); and mean bias, defined as the mean difference between daily mean clear-sky irradiance [in W/m^2] using the Long and Ackerman's (2000) algorithm and daily mean clear-sky irradiance [in W/m^2] using our method [method - reference] (Bias).

[a]not possible to apply Long and Ackerman's (2000) algorithm to the data due to its temporal resolution. [b]all values calculated from daily means.

resolution and is available since 1982. We did not use the first few years to avoid missing and/or problematic data reported in the documentation. The measurements are taken on board of polar orbiting satellites, meaning that the cloud product is derived from, usually, two samples per day (one in the morning, one in the afternoon) in lower and middle latitudes. Even though this might look as a limited sample, it should be enough for a good representation of mesoscale to large-scale meteorological events, due to their time scales of hours to a day (in the case of mesoscale) or longer (in the case of synoptic to global scales). Smaller scale phenomena, such as turbulence, could be misrepresented. However, due to their scales, these processes should, in general, not have major effects on the observed daily cloud cover. The CLARA-A2 product documentation reports a hit rate (overall global frequency of correct cloudy and cloud free estimations) of 79.7% when using CALIPSO-CALIOP (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation) measurements as a reference. Our comparison of the satellite cloud fraction with the daily mean synop cloud fraction in four sites (Madrid, Payerne, Lindenberg, and Saentis) led to correlation coefficients ranging from 0.796 (Saentis) to 0.853 (Lindenberg). Mean daily cloud cover derived from synop observations, also used in this study, was downloaded from the European Climate Assessment and Data set (Klein Tank et al., 2002).

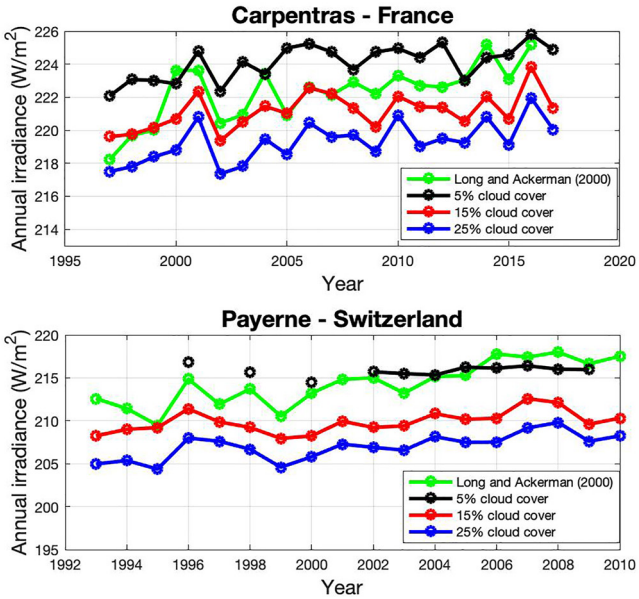

Figure 1. Annual clear-sky irradiances using different satellite cloud fraction thresholds compared to a reference clear-sky time series, based on Long and Ackerman's (2000) algorithm, for Carpentras—France, and Payerne—Switzerland.

3. Methods

The method developed in this study takes daily mean surface solar radiation measurement data as input. Its objective is to identify days with little to no effects of clouds on the irradiances, and then flag them as clear-sky. It relies on all-sky daily mean atmospheric transmittance (also referred to as clearness index by some authors). It is obtained by taking the ratio between daily mean (measured) surface solar radiation and daily mean (computed) top of atmosphere irradiance, which is then compared with (to be defined) clear-sky transmittance thresholds.

While the threshold definition requires the use of satellite-based cloud cover data, the method can subsequently be applied to the entire time series of daily surface solar radiation measurements of that station, even if it extends outside the period covered by satellite data. This then allows also the identification of clear-sky days in the long-term historic surface solar radiation records.

To avoid ambiguity, we start by introducing some important concepts used in the following sections, and associate them with letters for clarity:

- Individual month (*mi*): refers to a single month in the time series (i.e., January of 2003, October of 2007, etc.).
- Calendar month (*mc*): refers to every calendar month of the time series (i.e., all January months in a time series).
- Individual transmittance threshold (*thi*): refers to the transmittance threshold calculated for each of the individual months (*mi*).
- Transmittance threshold (th_{mc}): refers to the transmittance thresholds defined for each calendar month.

In the following sections, we discuss how to define a *thi*, and how they are used to define a th_{mc} for each calendar month at each station, to derive a clear-sky time series.

3.1. Definition of Individual Transmittance Thresholds (*thi*) and Transmittance Thresholds (th_{mc})

The method implementation starts by finding a *thi* for every individual month (*mi*) within the overlapping period of in situ and satellite data. At the end of the process, every station will have 12 th_{mc}, one for each calendar month. To start, we associate each station to its closest pixel from the grid of the CLARA-A2 daily cloud fraction product, making the station to be a point within the $0.25 \times 0.25°$ grid from the satellite. In situ data will be compared to satellite data, thus, it is necessary that the irradiance time series cover at least part of the satellite data period. Whereas one needs surface and satellite time series overlapping in order to apply this method, all stations were selected to accommodate a minimum overlapping between surface and satellite products of 8 years. This choice and its implications are discussed in Section 3.2.

Daily transmittance and daily cloud fraction data are sorted by individual months (*mi*) and then evaluated within each *mi*. The number of low cloud fraction days of each *mi* will be used as a proxy to indicate the probability of occurrence of clear-sky days at the station in that calendar month. However, defining "low cloud fraction days" based on satellite cloud cover data involves considering the uncertainties of representativity of the satellite grid for the station (Hakuba et al., 2013, 2014; Schwarz et al., 2017). One has to consider that the station is just a point within the satellite grid, thus, the satellite cloud fraction can be visualized as a probability of clear-sky occurrence at the station. For this reason, selecting just the days with satellite cloud fraction equal zero would result in a very low probability of cloud occurrence at the station, but with a high probability of missing many clear-sky days. Meanwhile, a 20% cloud cover at the scale of the satellite data grid may still be compatible with no clouds at the station itself. This creates a trade-off, where lower grid scale cloud fraction values imply missing many of clear-sky days at the station, while higher cloud fraction values mean including many cloudy days. Figure 1 illustrates the choice of different cloud cover percentages for the application of the proposed clear-sky method at two selected observation stations from BSRN.

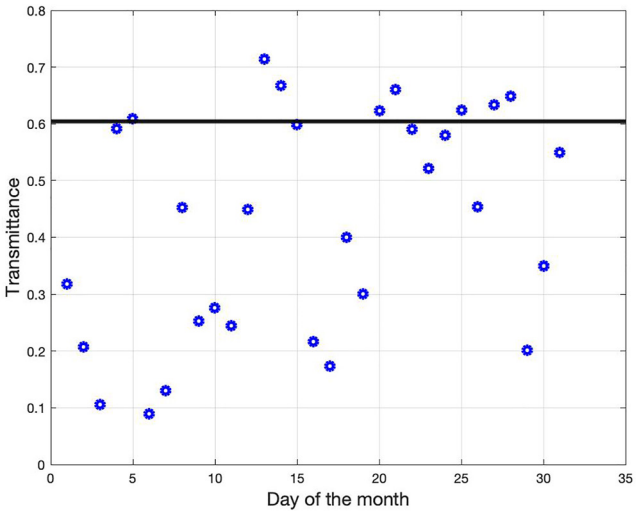

Figure 2. Daily transmittance in January of 1991 in Madrid, Spain. The black line marks the individual transmittance threshold set for this month based on the information of the percentage of cloud-free days from satellite information.

Even though 5% would reduce the bias (the absolute values would be closer to the reference, here generated using the Long and Ackerman (2000) algorithm), it would drastically reduce the availability of data, especially in a cloudier site, which is the case for the BSRN site Payerne (Switzerland), shown in this example. The use of 15% cloud fraction allows for more data, however, it still restricts significantly the representation of trends (potentially due to removal of too many clear-sky days). Finally, the use of 25% increases the absolute bias when compared to lower percentages, however it optimizes the representation of trends, still showing reasonable filtering of cloudy days, as discussed in Sections 4 to 4.2. For that reason we empirically chose 25% as the upper limit to define "low cloud fraction days."

Once we have the percentage of "low cloud fraction days" in a *mi* (i.e., percentage of days with cloud cover lower than 25%), we apply this percentage to the daily transmittance distribution of that *mi* to define a *thi*. For example, if 10% of a *mi* days are flagged as "low cloud fraction," then the 10% highest daily transmittances in that month are flagged as "very likely" to be clear-sky, and the lowest value of this group is set as the *thi* for that *mi*. Note that the days behind *thi* are not necessarily the same days identified via the cloud fraction criterium. The reason is again related to the representativity of the satellite data: days with low cloud fraction (less than 25%) in gridded satellite data do not necessarily coincide with the days with highest station specific transmittance. Figure 2 illustrates this process for January of 1991 in Madrid, Spain.

In January of 1991, 8 (25.8%) of the 31 days of the month had less than 25% of cloud cover in the pixel where the Madrid station is located. Consequently, the 74th percentile of the distribution of daily transmittances was set to be the transmittance threshold for that month, resulting in $thi = 0.6040$. With this approach, the cloudier the month is (according to satellite cloud fraction), the higher the transmittance threshold will be in terms of percentile (with respect to the distribution of values). The procedure is repeated for every month for which both in-situ and satellite data are available, leading to between 8 (minimum years we require) and 29 (maximum number of years of satellite data available) individual transmittance thresholds (*thi*) for each calendar month (*mc*) at each station, depending on the number of overlapping years between the station and the satellite data. In the case of no days with less than 25% of cloud cover, the value is kept as undefined (NaN).

Finally, we chose the 50th percentile of the *thi* distribution of each calendar month (*mc*) to define the Transmittance Threshold (th_{mc}) of that month, resulting in one th_{mc} per *mc*. Days whose transmittance is higher than th_{mc} for the respective calendar month will be flagged as clear-sky by our method. The choice of the 50th percentile is intuitive but somewhat arbitrary, and reveals another trade-off. A choice of lower percentiles to define the thresholds, being less restrictive, might allow more contamination by cloudy days being mistakenly flagged as clear-sky, while higher percentiles might be too restrictive, removing more clear-sky days from the clear-sky time series. A discussion about this trade-off can be found in Section 4.1.

3.2. Minimum Number of Overlapping Years

One important aspect when constructing the method is choosing a minimum number of overlapping years (period with both satellite and in situ data) required to calibrate the method for a specific station. Fewer overlapping years mean fewer individual transmittance thresholds (*thi*), which in turn implies that their distribution for each calendar month *mc* is constructed by fewer values, ultimately impacting the robustness of the transmittance thresholds th_{mc}. The issue is illustrated on Figure 3. We assume the *thi* to be normally distributed with a mean value of 0.65 and a standard deviation of 0.04 (or, alternatively, 0.02). The values are inspired by actual data. Around 60% of the total months at the stations used in this study have a standard deviation of the *thi* distribution within this range. From this normal distribution, we randomly draw a set of values and calculate their 50th percentile. We vary the size of the set drawn (between 1 and 100 values) and for each set size we repeat the drawing 1,000 times. We thus obtain 1,000 estimates of the 50th percentile for each set size, as illustrated in Figure 3. The illustrated

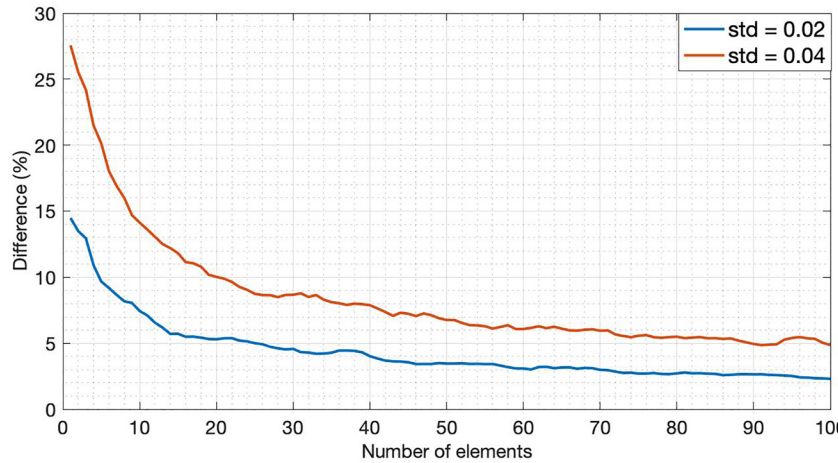

Figure 3. Difference between the minimum and the maximum 50th percentile when generating random normal distributions with 1–100 elements 1,000 times. Slopes show the moving mean of 5 values.

uncertainty represents the percentage difference between the highest and the lowest value of the 50th percentile (y-axis) for each data set size (x-axis). The curves are given as moving means (five values) for each number of elements in a set.

We observe that a higher standard deviation (0.04 instead of 0.02) results in larger uncertainties. The actual distribution of individual transmittance thresholds (*thi*) varies significantly by month and station, thus, the normal distribution of random numbers will not always mimic the behavior of this distribution. However, the distributions of *thi* are not random, they reflect the inter annual behavior of transmittance and satellite cloud cover at the station. To visualize what happens in real-world cases, we performed a similar analysis as shown in Figure 3, but in this case with real-world measurements. We calculated the uncertainty (difference between maximum and minimum 50th percentile) for 1,000 possible sets of 8 and 15 years of data (randomly generated) at four selected stations. The result is shown in Figure 4. The stations were chosen for having the longest records (up to 24 years of data).

The first obvious aspect, already highlighted in the previous Figure 4, is that the sets of 15 years of data (blue error bars) have lower uncertainties than the sets of 8 years of data (black error bars). This reinforces the importance of more overlapping years for more consistent results. The uncertainties with 8 years of data range mostly below 10%, consistent with the 0.02 standard deviation random normal distributions (shown in Figure 3). The highest uncertainty for these sites is for January in Payerne, when it becomes as high as 19% for the sets of 8 years of data. High uncertainties also occur in other months in Payerne, but never higher than 16%, also consistent with the 0.04 standard deviation random normal distributions. The higher uncertainty of less overlapping years affects three major aspects of the clear-sky detection and trend assessments which are discussed later in the text. With a higher uncertainty, the number of false positives (cloudy days mistakenly flagged as clear-sky) is increased, the correlation between the results of our method and the reference method is reduced and the confidence bounds of the bias correction are widened. Each of these aspects is discussed in more detail in Section 4.

Considering the uncertainties found both in the observations and in the simulations, the slopes on Figure 3 then are used as a reference to select the minimum number of overlapping years to allow use of our method for clear-sky derivation. We suggest at least between 8 and 10 years of overlapping data for the use of our method. Less overlapping years would result in a less accurate climatological (if the overlapping years are not representative for the station) and statistical (since any anomalous year would significantly affect the distribution) representation. More years and a lower *thi* standard deviation then obviously yield more robust results.

Within the stations used in this study, the one with the least overlapping years is Darwin, which has 13 years, but a *thi* standard deviation of ∼0.02 for all months, except February (∼0.05).

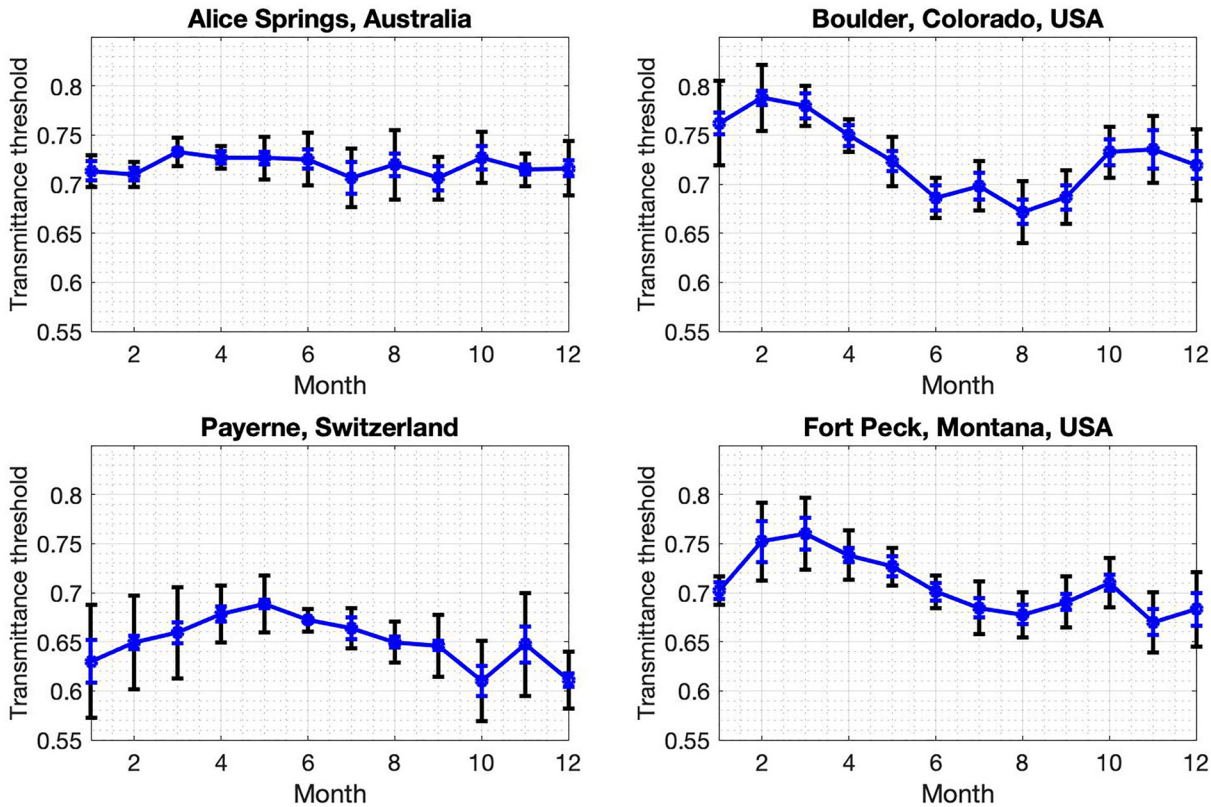

Figure 4. Transmittance thresholds th_{mc} (marker) and uncertainties (difference between maximum and minimum 50th percentile) for 1,000 possible sets of 15 years of overlapping data (blue error bars) and 8 years of data (black error bars) in selected stations. Transmittance thresholds were calculated with the whole time series. All stations have between 21 and 24 years of data and are from the Baseline Surface Radiation Network data set.

3.3. From Daily to Monthly Data

After the derivation of daily clear-sky time series, the amount of remaining data may vary from higher than 30% (i.e., Carpentras, France) to as low as 4.7% (i.e., Momote, Papua New Guinea). For several applications, one might want to convert those time series into monthly means. This was done in the current study in order to validate the method. Only months with more than one clear-sky day had the monthly values calculated. All daily values in a month were normalized by the monthly mean incoming TOA radiation. Monthly mean incoming TOA radiation was divided by incoming TOA radiation of the day flagged as clear-sky, and the resulting value was multiplied by the incoming surface solar radiation of the day flagged as clear-sky. With this procedure, we expect to avoid significant biases in the monthly values due to sun geometry. After this procedure, a simple mean of all normalized daily clear-sky values is calculated to generate monthly mean clear-sky values.

4. Validation

To validate this clear-sky method we compared it with the reference clear-sky time series, based on the Long and Ackerman (2000) algorithm, and to synop cloud cover data. Data from 24 stations, in five continents, 23 from BSRN and one from WRDC, were used in the validation procedures. The BSRN stations were chosen for their availability of data within the minimum overlapping period between satellite and in situ data. The stations from WRDC did not have enough temporal resolution to apply Long and Ackerman's (2000) algorithm. The individual WRDC station (Madrid, Spain) was selected as it matches with available synop cloud cover data used for the comparison, and because this station's cloudiness features contrasts with the other stations where synop cloud cover was analyzed, raising one of the aspects to be discussed. Table 1 shows the annual mean transmittance and satellite cloud cover at each of these sites, the percentage of clear-sky days as flagged by our method, the RSME obtained when comparing daily irradiances from our method to Long and Ackerman's (2000) and the bias in the clear-sky time series derived by our method. It should be noted that some stations have similar transmittance

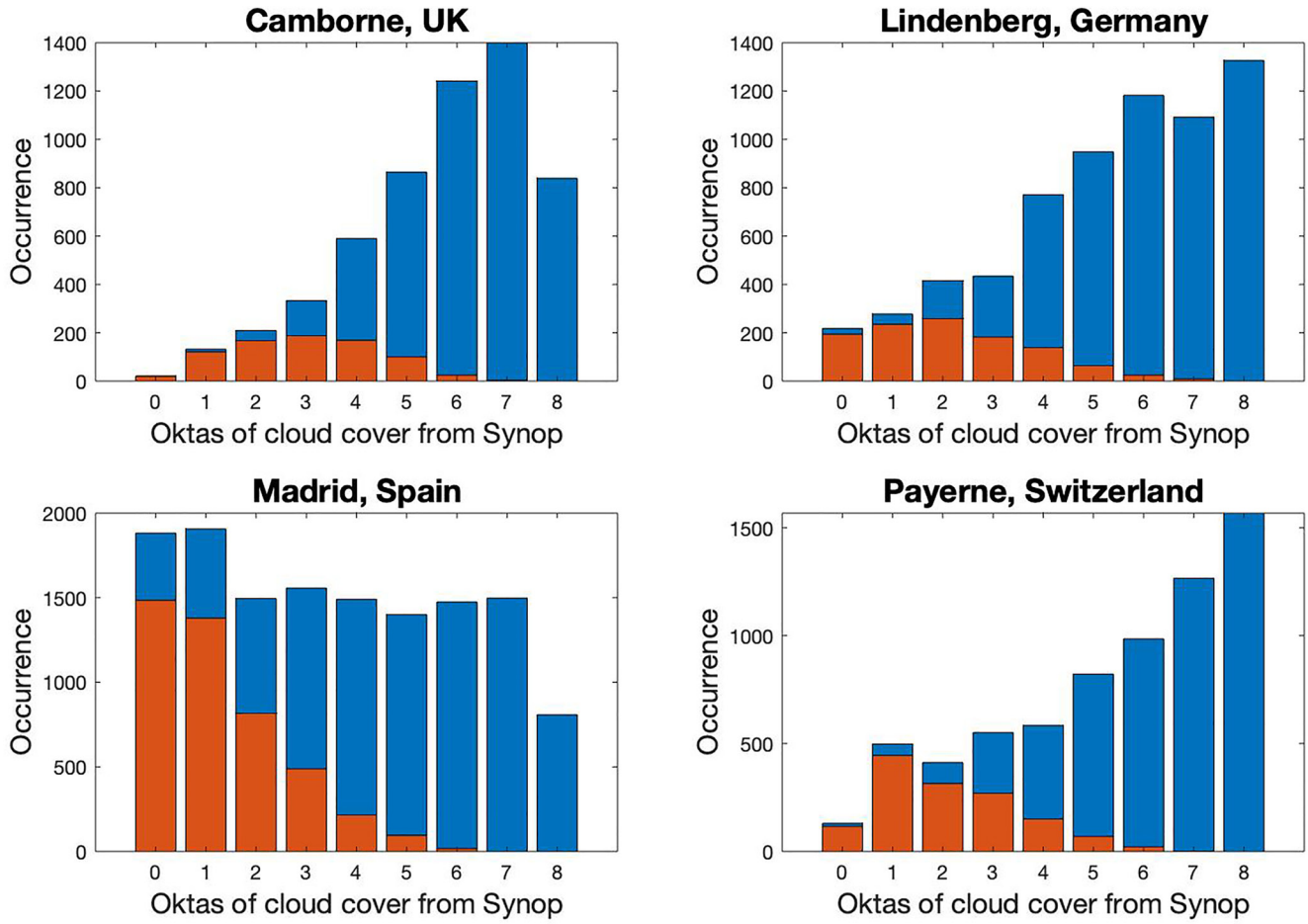

Figure 5. Occurrence of daily mean (rounded to the nearest integer) oktas of synop cloud cover (red + blue bars). Red part refers to the occurrence of days flagged as clear-sky by the method proposed in the present paper.

and fairly different satellite cloud cover (or vice versa), since cloud cover does not directly relate to cloud optical depth. Plus, clouds are not the only components affecting the transmittance (aerosols and water vapor also play a major role).

The number of clear-sky days and the RMSE vary significantly from site to site. The RMSE is very likely to be associated with the occurrence of partly cloudy days erroneously flagged as clear-sky by our method. This inserts something that we refer to as "cloud noise." Its implications are discussed in more details in the following sections.

4.1. Synop Cloud Cover on Days Flagged as Clear-Sky

The clear-sky days identified by our method were compared to synop cloud cover data from their respective sites. Synop cloud clover is the most traditional cloudiness measurement and has been used for more than a century for various applications. Several studies compared synop cloud cover with satellite cloud cover (e.g., Bojanowski et al., 2014; Kotarba, 2017), and even though they highlight inconsistencies when comparing the two methods, they also highlight their capability of addressing real-world cloudiness. Thus, even though both synop and satellite data have their limitations, a good agreement between clear-sky days identified by our method and clear-sky days identified using synop measurements would definitely be an indication on the ability of the method to filter cloudy days. Figure 5 shows the occurrence of each of the oktas from synop cloud cover and their respective classification according to our method for a few sites. The sites were chosen due to the availability of synop cloud cover data from the European Climate Assessment and Data Set (Klein Tank et al., 2002).

Table 2
Occurrence of True/False Positives/Negatives When Comparing Days Flagged as Clear-Sky by Our Method With Days With 2 or Less Oktas of Cloud Cover According to Synop Data

	True positive	False positive	True negative	False negative	% of accuracy on positives	% of true positives
Camborne	308	494	4,945	54	85.1	38.4
Lindenberg	687	419	5,335	223	75.5	62.1
Madrid	3,682	820	7,410	1,601	69.7	82.2
Payerne	876	515	5,261	164	84.2	63.0

Note. % of Accuracy on Positives Refer to the Number of True Positives Compared to the Total Clear-Sky Days According to Synop Cloud Cover (True Positive + False Negative); % of True Positives Refer to the Number of True Positives Compared to the Total Number of Clear-Sky Days by the Method (True Positives + False Positives).

We observe that at all stations our method was able to identify most of the low cloud cover days (according to synop data) as clear-sky. One should highlight, however, that a significant amount of the days with 0 or 1 oktas of cloud cover in Madrid where flagged as cloudy by our method. This is the result of a trade-off. Slightly lower thresholds (i.e., using a percentile lower than the 50th when getting the th_{mc} from the *thi* distributions) would flag most of these days as clear-sky, nevertheless, it would increase the number of days with high synop cloud cover, flagged as clear-sky. This trade-off does not have a strong impact on the results for a relatively sunny site such as Madrid, since the proportion of low to high synop cloud cover days flagged as clear-sky is much bigger than one. However, for relatively cloudy sites, such as Camborne, this proportion is close to one, thus the increase in high synop cloud cover days flagged as clear-sky would have a much more significant impact. Another way of visualizing this is by the analysis of true/false positives/negatives. For the means of this analysis, we define days with 2 or less oktas of cloud cover as clear-sky, and if our method also flags these days as clear-sky, we have a true positive. If our method flags these days as cloudy, we have a false negative. On the other hand, if the day has more than 2 oktas of cloud cover, and our method flags it as clear-sky, that is a false positive. Finally, a true negative occurs when both methods agree on a cloudy day. Table 2 shows the occurrence of true/false positives/negatives in these four stations. We observe that a slight increase in false positives would affect more the results in Camborne than in Madrid.

If we change the classification in the table to 4 or less oktas, we see an increase in the percentage of true positive values from 83.5% in Camborne, to 97.1% in Madrid. The remaining 2.9%–16.5% are very likely to be "false positives" (cloudy days flagged as clear-sky by the method), since they have more than 50% cloud cover as reported from synop data. This indicates a reasonable agreement between clear-sky identification by our method and by synop cloud cover, but it also highlights the potential contamination with false positives. Due to the daily temporal resolution, it is virtually impossible to completely avoid all the cloud signals, however, through reducing the occurrence of false positives, we then minimize the uncertainty it brings to the clear-sky time series we are constructing.

4.2. Comparison With Long and Ackerman's Algorithm

The comparison with clear-sky time series derived using the Long and Ackerman (2000) algorithm explores the ability of our method to identify clear-sky days, to reproduce clear-sky trends, and to reproduce month to annual variability. Figure 6 shows parts of two time series of daily irradiance before and after clear-sky filtering with our method, compared to Long and Ackerman's clear-sky time series.

Figure 6 illustrates that our method was able to screen out cloudy days, and keep only days that had a good fit with the clear-sky time series from Long and Ackerman's algorithm. In fact, we see that the difference in daily mean irradiances between the two clear-sky time series was lower than 5% of the absolute values in 68% of the days in Payerne, and in 85% of the days in Carpentras. This difference was within 10% in 90% of the days in Payerne and in 98% of the days in Carpentras. And the difference was within 15% in 99% of the days for both sites. This reinforces the ability of our method to identify clear-sky days and also highlights a slightly better agreement at the sunnier site (i.e., Carpentras) between the two.

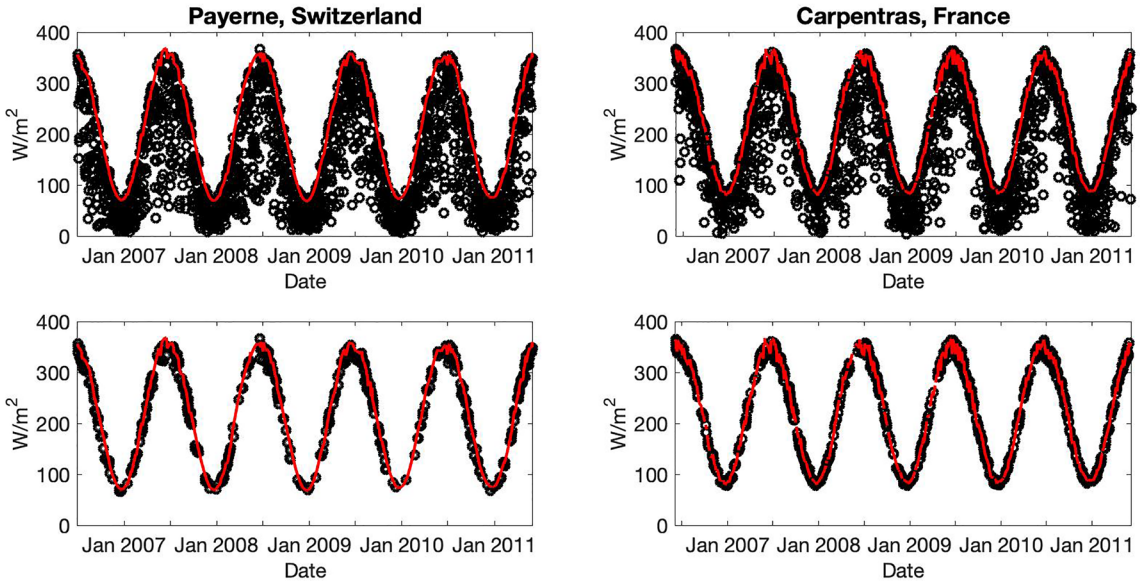

Figure 6. Time series of daily mean irradiance at the surface in Payerne, Switzerland (July 2006–May 2011), and in Carpentras, France (July 2006–June 2011). Red lines represent clear-sky identified by the Long and Ackerman (2000) algorithm. Black dots in upper graphs show daily irradiances for all days over the periods considered. Black dots in lower graphs show only irradiances from days flagged as clear-sky by our method.

To visualize the method's ability to represent clear-sky inter-annual variability as well as trends, we compared clear-sky irradiance anomalies time series from different sites. Figure 7 shows the time series of annual anomalies at eight of the BSRN sites used in this study. They were chosen to represent different conditions: both cloudier stations (such as Momote) and sunnier stations (such as Desert Rock).

With respect to the inter-annual variability of clear-sky irradiance anomalies, we observe a reasonable agreement in most cases, especially at sites with high mean transmittance (and less clouds), like in Desert Rock, for example, at cloudier sites, such as Lindenberg, Tateno and Toravere, even though the trends are still well represented, the quality of the inter-annual variability representation is low. In the extreme case of Momote, located in the equatorial region, under the influence of the Intertropical Convergence Zone, where few cloud-free days occur, thus most of the monthly, and therefore annual, values obtained by our method are missing. This indicates that the cloud-free processes have minor effects at that site, and it is reinforced by the flat line of the clear-sky time series obtained from the Long and Ackerman (2000) algorithm. The range of anomaly values is also similar between our method and the reference time series, however, it should be noted that the absolute values in our method are usually around 5%–10% lower than the reference clear-sky values. This happens because, using daily time steps, we cannot filter only purely clear-sky scenes, like it is done in the reference algorithm using minute screening of clouds. This allows some remaining background cloud noise with our method, which is evident in the RMSE and Bias values in Table 1, and which might be a limiting factor for an accurate representation of higher frequency components of the time series (i.e., anomalies).

The correlation coefficients between the daily clear-sky anomalies time series from our method and from Long and Ackerman's (2000) method seem to be inversely proportional to the average satellite cloud cover at the station. The two highest correlations coefficients are at Desert Rock and Alice Springs, which have the lowest average satellite cloud cover. The lowest correlation coefficients are at Momote and Lerwick, which are also the stations with highest average satellite cloud cover. This is in line with the obvious hypothesis that cloudier stations have higher "cloud noise," thus will have a worse representation of the anomalies in the time series.

While the analysis of the variability of the anomalies shows some potential limitations of the method, the trend analysis shows its major strength: the ability to reproduce decadal trends. The decadal trends were calculated for 21 of the 23 BSRN sites (Lerwick, UK, and Momote, Papua New Guinea, were not included because of too much missing data due to lack of clear-sky days, as seen from Figure 7 for Momote's case) using clear-sky days determined by our method and by Long and Ackerman's algorithm. Figure 8 shows the comparison between the two.

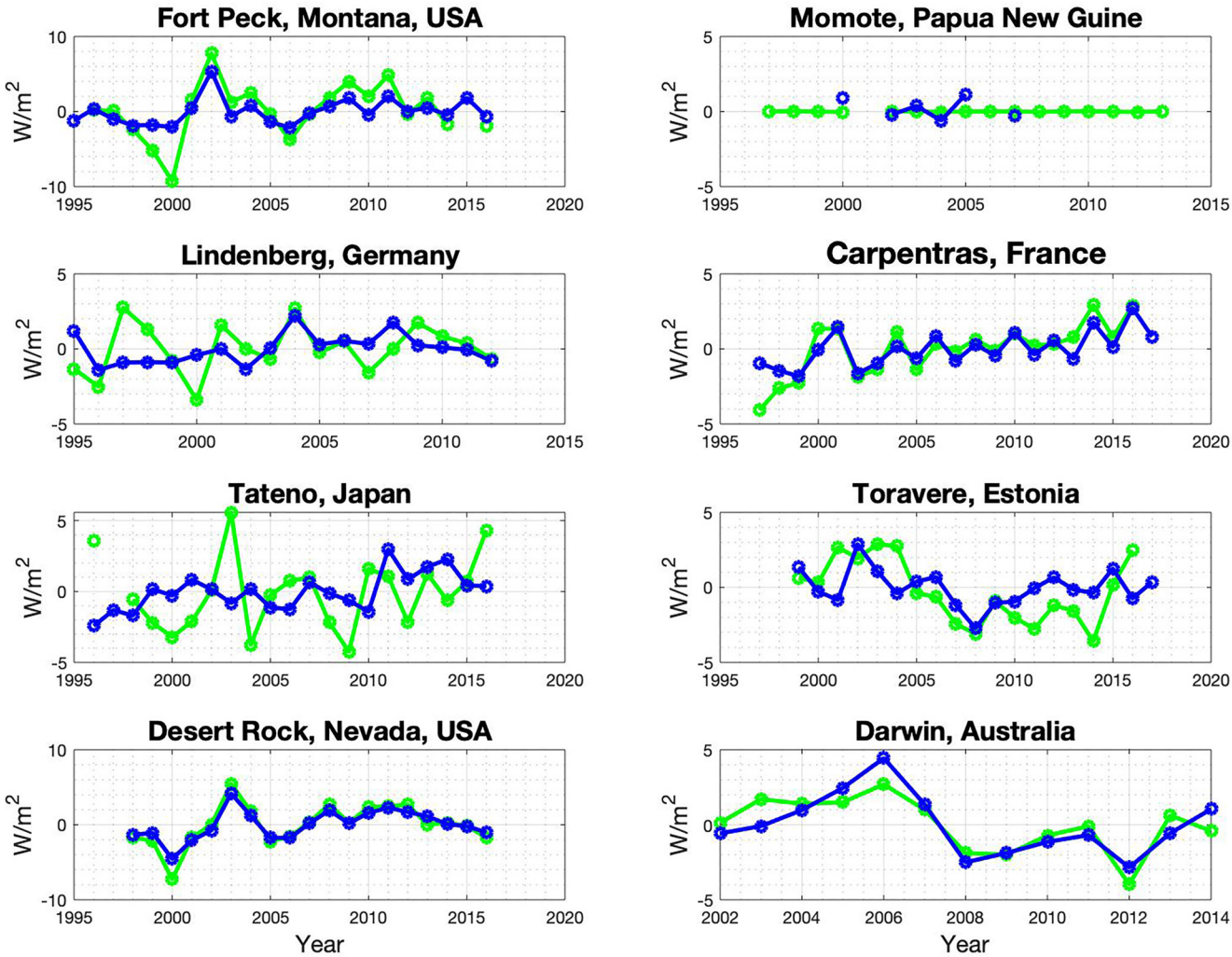

Figure 7. Time series of annual anomalies of clear-sky irradiance at several sites from Baseline Surface Radiation Network. Blue lines show the output from our method, green lines show the output from Long and Ackerman's (2000) algorithm. Trends (in W/m^2 per decade): Fort Peck: 1.75 reference, 0.69 method; Momote: not enough data to calculate*; Lindenberg: 0.62 reference, 0.51 method; Carpentras: 2.15 reference, 1.08 method; Tateno: 1.15 reference, 1.09 method; Toravere: −0.80 reference, −0.41 method; Desert Rock: 1.59 reference, 1.31 method; Darwin: −2.38 reference, −1.73 method. *Later discussed in this section.

We see a strong correlation ($R^2 = 0.92$) between decadal trends determined by both methods, and, most importantly, the methods agree in the sign of the trends at all of the 21 sites. That is, any brightening/dimming sign shown in Long and Ackerman's clear-sky time series, was also reproduced in the time series based on our method. The results also show a systematic bias in the trends obtained by our method (trends are usually lower, in absolute values), potentially associated to the previously mentioned cloud noise. The average absolute trend over these stations is 1.69 for the reference method, against 1.09 for our method, and the slope of the regression is 1.50. The observed linear relationship allows the use of the slope to correct the bias. We have an average correction factor of 1.50 (with 95% confidence bounds of 1.27 and 1.71, thus a $\pm$ 0.23 uncertainty), when considering all the stations. However, when considering the cloudiness of the stations, we observe slightly different values. When doing the same regression analysis only using the 10 stations with mean satellite cloud cover of 55% or more (see Table 1), we get a R^2 of 0.94, and a correction factor of 1.86 (with 95% confidence bounds of 1.48 and 2.24). On the other hand, when doing the same analysis using the eight stations with 50% or less mean satellite cloud cover (see Table 1), we again obtain a strong R^2 of 0.95, but a correction factor of 1.30 (with 95% confidence bounds of 0.99 and 1.613). This result indicates that cloudier stations (in this case, stations with mean satellite cloud cover equal to or higher than 55%) require a bias correction up to 25% higher than the average, while sunnier stations (in this case, stations with mean satellite cloud cover equal or lower than 50%) require a bias correction up to 15% lower

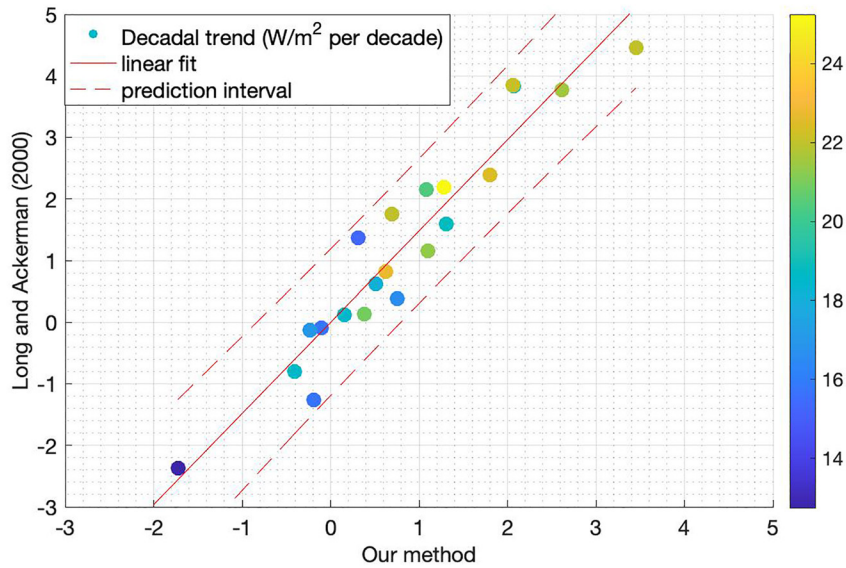

Figure 8. Comparison of decadal trends in clear-sky time series obtained using our method and using Long and Ackerman's algorithm for 21 Baseline Surface Radiation Network stations. Colors indicate the length (in years) of the time series. Prediction interval with a 95% confidence level.

than the average. This difference is very likely to be associated with the cloud noise, which imposes a higher bias to cloudier stations. The bias correction of 1.50 ± 0.23 should be applied to any clear-sky decadal trend calculated using our method in order to find the estimated real world clear-sky trend.

4.3. Comparison to the Seasonal Maximal Transmittance Method

The seasonal maximum transmittance method is another alternative for long-term clear-sky trend analysis without the necessity of additional measurements. It has been used by Wild et al. (2021) to analyze decadal SSR trends in Central Europe. This method relies on the maximum daily transmittance per season (thus four transmittances per year) to derive annual means and then long term clear-skies time series (Wild et al., 2021). We applied this method to the stations used in this study in order to compare the clear-sky trends those obtained by our method. We kept the time series in transmittance units in order to avoid any biases associated with the conversion to irradiances. A similar comparison as in the previous sections using the same 21 stations in shown in Figure 9.

We see a small decrease in the R^2 of the fit from our method using transmittance time series (0.85) when compared to the trends using irradiance values (0.92). However, it has a significantly better fit when compared to the clear-sky trends from seasonal maximum transmittance method (0.59), which has higher uncertainties. The results indicate that, even though both methods show a good qualitative representation of clear-sky trends, our method brings a significant improvement in the quantitative representation of such trends, when correcting the bias. The slope of the regression line is 1.42 for our method, indicating that we underestimate the absolute trend magnitude as compared to Long and Ackerman (2000). By contrast, the slope of the regression line is 0.52 if seasonal maximum transmittances are used. This seems plausible as using the maximum value per season is likely to result in a bias toward an overall too transparent atmosphere (too high transmittance).

5. Discussion

The whole validation process highlights most of the strengths and weaknesses of our method. A first aspect to be highlighted is its applicability. The combination of satellite and in situ data for calibration of the method allows it to be applied to any station around the globe, depending only on the overlap of satellite and in situ data necessary for the threshold definitions.

The results in Section 4.1 show that the method is able to identify most of the days that are more likely to be clear-sky, when comparing to synop cloud cover. At sites where clouds are more frequent, such as in Camborne

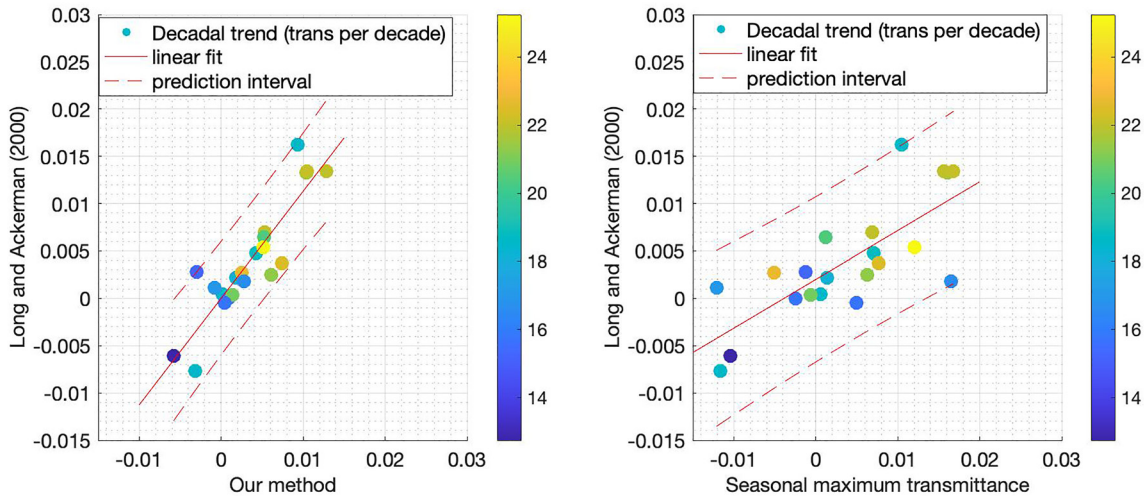

Figure 9. Comparison of decadal trends in clear-sky time series obtained using our method (left graph) and using the seasonal maximum transmittance method (right graph) with the clear-sky decadal trends from Long and Ackerman's algorithm for 21 Baseline Surface Radiation Network stations. Decadal trends in units of transmittance per decade. Colors indicate the length (in years) of the time series. Prediction interval with a 95% confidence level.

(UK), it is more likely for the transmittance thresholds to allow a day with the presence of clouds to be erroneously flagged as clear-sky (Figure 5). This should affect partly cloudy days, with presence of broken clouds. Overcast days, when transmittance is strongly affected by clouds, should not suffer from this problem for cloud identification.

The comparison with clear-sky time series derived from Long and Ackerman's algorithm allowed the verification of several aspects of the time series, especially the variability of the anomalies and trends. The representation of the anomalies in the clear-sky time series is probably the most important weakness to be highlighted. Even though some sites have shown a good representation of the anomalies (with a correlation coefficient >0.6), at other sites the ability of the method to reproduce short term variations was limited. The weak correlations between our method and the reference time series were found at cloudier sites. This might also indicate the presence of the background cloud noise caused by a few partly cloudy days mistakenly flagged as clear-sky. This background cloud noise represents an important limitation for the reproducibility of high frequency components. On the other hand, the ability of our method to reproduce long term trends is an important strength to be highlighted. The lower frequency components of the time series are less affected by the cloud noise caused by a few days mistakenly flagged as clear-sky, thus, in this context our method appears to be a useful alternative. The clear-sky time series derived by our method were able to reproduce the sign of the trend as determined in the reference time series in all of the cases analyzed, and the comparison with the reference data set showed a good linear relationship of the decadal trends in surface solar radiation. This analysis also allowed us to verify the improvement of the quantitative representation of clear-sky decadal trends by our method when compared to the seasonal transmittance method. However, there is a systematic bias that has to be taken into account. According to the slope of the linear fit of this comparison, the reference trends are usually around 50% larger than the trends derived by our method's clear-sky time series. This implies that an average 1.5 ± 0.23 correction factor should be applied to the result in order to get the expected real world trends. For more accurate estimates, the cloudiness of the station should be taken into account. Cloudier stations (in this study, stations with mean satellite cloud cover equal to or higher than 55%) have higher biases, due to their higher cloud noise, requiring then a higher correction factor, which was found to be up to 25% higher than the average. On the other hand, sunnier stations (in this study, stations with mean satellite cloud cover equal or lower than 50%) require a correction factor up to 15% lower than the average. One should take these aspects into consideration when using decadal trends derived using our method.

We already discussed our method's limitation to reproduce high frequency components of the clear-sky time series in cloudy regions, but we should also discuss its ability to generate a clear-sky time series at all. While at sites like Madrid, more than 30% of the days are flagged as clear-sky, at other sites this value could be much lower, such as 6.1% in the case of Lerwick or even 4.7% in the case of Momote, Papua New Guinea. Such low

availability of cloud-free days may represent a limitation for the analysis of cloud-free time series. However, it should also be seen as an indication of the radiative balance being controlled by cloud-mediated processes. Our method aims at visualizing what happens on days when clouds are not a significant aspect. If there are no cloud-free days at a specific site, then the cloud-free processes possibly have only minor direct effects on the radiative balance on time scales larger than a day.

Finally, as described in the methods section, we set thresholds for satellite cloud fraction and for *thi* distribution. These are the values proposed in this study, for allowing a reasonable trend representation as well as minimizing the occurrence of false positives. The results presented here highlight the strengths and weaknesses of the method using these values and indicate the trade-off that was considered when selecting them. These aspects should be considered when potentially using different values.

6. Conclusions

We presented a new method for the identification of clear-sky trends based on daily solar irradiance data at the surface. For the site specific method calibration only, daily satellite derived cloud cover data are also required. The method uses optimal transmittance thresholds to derive daily clear-sky time series. The use of satellite data upon method calibration allows the method to be applicable to any station on the globe, given that a part of the station irradiance time series overlaps with the satellite period. In our validation procedure, we compared our results to a reference clear-sky time series, derived using the algorithm by Long and Ackerman (2000). This verified our method's ability to reproduce high frequency (i.e., anomalies) and low frequency (i.e., decadal trends) components of the time series in comparison to the reference method. We also compared the clear-sky identification of our method with synop cloud cover data. The full validation procedure allowed us to identify both the strengths and the limitations of the method. The most important limitation is the remaining "cloud noise" in the clear-sky time series, resulting from partially cloudy days erroneously flagged as clear-sky. This issue is obviously more frequent in cloudier sites, affecting particularly the high frequency reproducibility of the clear-sky time series and producing higher biases in the decadal trends calculated for cloudier locations. On the other hand, the reproducibility of the low frequency components of the time series (i.e., decadal trends) has shown a strong correlation when compared to the reference data set. This leads to the conclusion that this method is adequate for long-term clear-sky time series (i.e., decadal trends), and not ideal for short term variability (i.e., anomalies). We found an average bias correction factor of 1.5 ± 0.23 for the long term trends, meaning that any trend obtained using our method should be multiplied by 1.5 in order to find the expected real world trend for the respective site. When considering the cloudiness of the station for the bias correction, we found that cloudier stations (here defined as stations with mean satellite cloud cover equal to or higher than 55%) might require a correction factor up to 25% higher than the average (1.86), while sunnier stations (here defined as stations with mean satellite cloud cover equal to or lower than 50%) might require a correction factor up to 15% lower than the average (1.30). We understand that, just like any other clear-sky identification method, the proposed one is not perfect, but by presenting it we expect to allow a broader assessment (in time and space) of clear-sky time series, to better understand changes in the radiative processes in the cloud-free atmosphere.

Data Availability Statement

The Baseline Surface Radiation Network (BSRN) data used to apply and validate the method presented in this study is available at the BSRN website (https://bsrn.awi.de/). For this study, it was retrieved via the ftp server ftp://ftp.bsrn.awi.de/. The World Radiation Data Centre station data used to apply and evaluate the method is available for registered users at http://wrdc.mgo.rssi.ru/. The satellite cloud fraction data from CLARA, used to apply the method described in the paper, can be found in the CM SAF website (https://www.cmsaf.eu/) and downloaded using the Web User Interface at https://wui.cmsaf.eu/.The synop cloud cover data, used to evaluate the method in this study, can be downloaded from the European Climate Assessment & Dataset website (http://www.ecad.eu).

Acknowledgments
This study was funded by Swiss National Science Foundation grant no. 200020_188601. Open access funding provided by ETH Zurich. We are grateful to the Baseline Surface Radiation Network community for its continuous efforts to maintain and distribute high quality radiation data. The authors also highly acknowledge the work of the CM-SAF team, which produces the CLARA data set, used in this study, available for every interested user. Access to the code used by the authors to apply the method can be provided upon request.

References

Barbaro, S., Cannata, G., Coppolino, S., Leone, C., & Sinagra, E. (1981). Correlation between relative sunshine and state of the sky. *Solar Energy*, *26*(6), 537–550. https://doi.org/10.1016/0038-092x(81)90166-3

Bojanowski, J. S., Stöckli, R., Tetzlaff, A., & Kunz, H. (2014). The impact of time difference between satellite overpass and ground observation on cloud cover performance statistics. *Remote Sensing*, *6*(12), 12866–12884. https://doi.org/10.3390/rs61212866

Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., et al. (2018). Baseline Surface Radiation Network (BSRN): Structure and data description (1992–2017). *Earth System Science Data*, *10*(3), 1491–1501. https://doi.org/10.5194/essd-10-1491-2018

Gueymard, C. A., Bright, J. M., Lingfors, D., Habte, A., & Sengupta, M. (2019). A posteriori clear-sky identification methods in solar irradiance time series: Review and preliminary validation using sky imagers. *Renewable and Sustainable Energy Reviews*, *109*, 412–427. https://doi.org/10.1016/j.rser.2019.04.027

Hakuba, M. Z., Folini, D., Sanchez-Lorenzo, A., & Wild, M. (2013). Spatial representativeness of ground-based solar radiation measurements. *Journal of Geophysical Research: Atmospheres*, *118*(15), 8585–8597. https://doi.org/10.1002/jgrd.50673

Hakuba, M. Z., Folini, D., Sanchez-Lorenzo, A., & Wild, M. (2014). Spatial representativeness of ground-based solar radiation measurements—Extension to the full Meteosat disk. *Journal of Geophysical Research: Atmospheres*, *119*(20), 11–760. https://doi.org/10.1002/2014jd021946

Karlsson, K. G., Anttila, K., Trentmann, J., Stengel, M., Fokke Meirink, J., Devasthale, A., et al. (2017). CLARA-A2: The second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data. *Atmospheric Chemistry and Physics*, *17*(9), 5809–5828. https://doi.org/10.5194/acp-17-5809-2017

Karlsson, K. G., Anttila, K., Trentmann, J., Stengel, M., Solodovnik, I., Meirink, J. F., et al. (2020). CLARA-A2.1: CM SAF cLoud, Albedo and surface RAdiation dataset from AVHRR data - Edition 2.1, Satellite Application Facility on Climate Monitoring. https://doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V002_01

Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., Böhm, R., Demarée, G., Gocheva, A., et al. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *22*(12), 1441–1453. https://doi.org/10.1002/joc.773

Kotarba, A. Z. (2017). Inconsistency of surface-based (SYNOP) and satellite-based (MODIS) cloud amount estimations due to the interpretation of cloud detection results. *International Journal of Climatology*, *37*(11), 4092–4104. https://doi.org/10.1002/joc.5011

Long, C. N., & Ackerman, T. P. (2000). Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *Journal of Geophysical Research*, *105*(D12), 15609–15626. https://doi.org/10.1029/2000jd900077

Long, C. N., & Dutton, E. G. (2002). *BSRN Global Network Recommended QC Tests,V 2.0*. BSRN Technical Report.

Manara, V., Brunetti, M., Celozzi, A., Maugeri, M., Sanchez-Lorenzo, A., & Wild, M. (2016). Detection of dimming/brightening in Italy from homogenized all-sky and clear-sky surface solar radiation records and underlying causes (1959–2013). *Atmospheric Chemistry and Physics*, *16*(17), 11145–11161. https://doi.org/10.5194/acp-16-11145-2016

Ododo, J. C., Agbakwuru, J. A., & Ogbu, F. A. (1996). Correlation of solar radiation with cloud cover and relative sunshine duration. *Energy Conversion and Management*, *37*(10), 1555–1559. https://doi.org/10.1016/0196-8904(96)86837-3

Ohmura, A., Dutton, E. G., Forgan, B., Fröhlich, C., Gilgen, H., Hegner, H., et al. (1998). Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research. *Bulletin of the American Meteorological Society*, *79*(10), 2115–2136. https://doi.org/10.1175/1520-0477(1998)079<2115:bsrnbw>2.0.co;2

Perez, R., Ineichen, P., Seals, R., Michalsky, J., & Stewart, R. (1990). Modeling daylight availability and irradiance components from direct and global irradiance. *Solar Energy*, *44*(5), 271–289. https://doi.org/10.1016/0038-092x(90)90055-h

Renner, M., Wild, M., Schwarz, M., & Kleidon, A. (2019). Estimating shortwave clear-sky fluxes from hourly global radiation records by quantile regression. *Earth and Space Science*, *6*(8), 1532–1546. https://doi.org/10.1029/2019ea000686

Reno, M. J., & Hansen, C. W. (2016). Identification of periods of clear sky irradiance in time series of GHI measurements. *Renewable Energy*, *90*, 520–531. https://doi.org/10.1016/j.renene.2015.12.031

Sanchez-Lorenzo, A., Calbo, J., Brunetti, M., & Deser, C. (2009). Dimming/brightening over the Iberian Peninsula: Trends in sunshine duration and cloud cover and their relations with atmospheric circulation. *Journal of Geophysical Research*, *114*(D10), D00D09. https://doi.org/10.1029/2008jd011394

Schwarz, M., Folini, D., Hakuba, M. Z., & Wild, M. (2017). Spatial representativeness of surface-measured variations of downward solar radiation. *Journal of Geophysical Research: Atmospheres*, *122*(24), 13–319. https://doi.org/10.1002/2017jd027261

Stjern, C. W., Kristjánsson, J. E., & Hansen, A. W. (2009). Global dimming and global brightening—An analysis of surface radiation and cloud cover data in northern Europe. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *29*(5), 643–653. https://doi.org/10.1002/joc.1735

Thevenard, D., & Brunger, A. (2001). *Typical weather years for international locations*. ASHRAE Research Project.

Voeikov Main Geophysical Observatory. (2022). World Radiation Data Centre website. http://wrdc.mgo.rssi.ru/

Wild, M. (2009). Global dimming and brightening: A review. *Journal of Geophysical Research*, *114*(D10), D00D16. https://doi.org/10.1029/2008jd011470

Wild, M., Wacker, S., Yang, S., & Sanchez-Lorenzo, A. (2021). Evidence for clear-sky dimming and brightening in central Europe. *Geophysical Research Letters*, *48*(6), e2020GL092216. https://doi.org/10.1029/2020gl092216

Yamasoe, M. A., Rosário, N. M. É., Almeida, S. N. S. M., & Wild, M. (2021). Fifty-six years of surface solar radiation and sunshine duration over São Paulo, Brazil: 1961–2016. *Atmospheric Chemistry and Physics*, *21*(9), 6593–6603. https://doi.org/10.5194/acp-21-6593-2021

Younes, S., & Muneer, T. (2006). Improvements in solar radiation models based on cloud data. *Building Services Engineering Research and Technology*, *27*(1), 41–54. https://doi.org/10.1191/0143624406bt143oa

Younes, S., & Muneer, T. (2007). Clear-sky classification procedures and models using a world-wide data-base. *Applied Energy*, *84*(6), 623–645. https://doi.org/10.1016/j.apenergy.2007.01.016