Ten years of 1 Hz solar irradiance observations at Cabauw, the Netherlands, with cloud observations, variability classifications, and statistics

Wouter B. Mol[1], Wouter H. Knap[2], and Chiel C. van Heerwaarden[1]

[1]Meteorology and Air Quality Group, Wageningen University, Wageningen, The Netherlands
[2]Royal Netherlands Meteorological Institute, De Bilt, The Netherlands

Correspondence: Wouter Mol (wbmol@wur.nl)

Abstract. Surface solar irradiance varies on scales down to seconds, of which detailed, long-term observational datasets are rare but in high demand. Here, we present an observational dataset of global, direct, and diffuse solar irradiance sampled at 1 Hz, and fully resolved variability until at least 0.1 Hz, over a period of 10 years, from the Baseline Surface Radiation Network (BSRN)

5   station at Cabauw, the Netherlands. The dataset is complemented with irradiance variability classifications, clear-sky irradiance and aerosol reanalysis, information about the solar position, observations of clouds and sky type, and wind measurements up to 200 meters above ground level. Statistics of variability derived from all time series include approximately 185,000 detected events of both cloud enhancement and cloud shadows. The Cabauw measurement site has additional observations freely available at the open data platform of the Royal Netherlands Meteorological Institute. This paper describes the observational site,

10   quality control, classification algorithm with validation, and the processing method of complementary products. Additionally, we discuss and showcase (potential) applications, including limitations due to sensor response time. These observations and derived statistics provide detailed information to aid research into how clouds and atmospheric composition influence solar irradiance variability, and to help validate models that are starting to resolve variability at higher fidelity. The main datasets are available at https://doi.org/10.5281/zenodo.7093164 (Knap and Mol, 2022) and https://doi.org/10.5281/zenodo.7092058 (Mol

15   et al., 2022b), see the data availability section for the complete list.

1   Introduction

Clouds generate large intra-day surface solar irradiance variability, the spatiotemporal scales of which reach down to seconds or less (Yordanov et al., 2013; Tabar et al., 2014; Gueymard, 2017; Kivalov and Fitzjarrald, 2018), or tens of meters (Lohmann et al., 2016; Mol et al., 2022a). Observing, understanding, and forecasting irradiance variability at these scales is important for

20   a range of applications. Solar energy production and electricity grid stability is negatively affected by fast and local irradiance variability (Liang, 2017; Yang et al., 2022). Numerical weather prediction models are incapable of forecasting variability at these short scales, but as their resolution keeps increasing and sub-grid scale irradiance variability parameterizations are developed, they require more detailed observations for validation. Cloud resolving models and the development of more accurate

3D radiative transfer calculations in academic setups (e.g. Jakub and Mayer, 2015; Gristey et al., 2020; Veerman et al., 2022), likewise require detailed and accurate observations of solar irradiance. Heterogeneity of solar irradiance and resulting surface fluxes is also an increasingly important topic in the field of land-atmosphere interaction (Helbig et al., 2021), with a non-linear response of vegetation's photosynthesis for varying light intensities (Pearcy and Way, 2012) or diffuse irradiance penetration into canopies (Mercado et al., 2009; Durand et al., 2021).

Existing observational datasets of surface solar irradiance at the sub-minute scale are rare, in particular for multiple years or longer and with separate direct and diffuse irradiance measurements. Notable examples of such datasets include those used in previously mentioned studies (Tabar et al., 2014; Gueymard, 2017; Kivalov and Fitzjarrald, 2018; Lohmann, 2018; Gristey et al., 2020). In this work, we present such a dataset, which consists of 10 years of 1 Hz resolution global, direct, and diffuse irradiance, supplemented with meteorological observations for interpretation and data analysis. To the best of our knowledge, this is a unique observational dataset given its time span, temporal resolution, and multi-component measurements. The separation of global horizontal irradiance (GHI) into direct and diffuse components is important for distinguishing and characterizing the different types of atmospheric conditions and specific conditions under which irradiance variability is generated. Most notably, the phenomenon of cloud enhancement, where clouds scatter additional sunlight to cloud-free spots on the surface to significantly increase total irradiance (Gueymard, 2017; Yordanov, 2015; Mol et al., 2022a), is by definition a combination of direct and diffuse irradiance and cannot be understood with only GHI observations.

In this work, we will describe the 10-year observational dataset of solar irradiance, all supplementary meteorological observations and related processing, the time series variability classification algorithm, statistical datasets derived from all time series and classifications, and examples of how the data can be used. Sections 2 and 3 are a complete and more elaborate version of the condensed dataset description published in Journal of Geophysical Research: Atmospheres (Mol et al., 2022a).

2 Observational data description

All in-situ observations in this dataset are taken at the Ruisdael Observatory in Cabauw (previously known as CESAR, https://ruisdael-observatory.nl/cabauw/) of the Royal Netherlands Meteorological Institute (KNMI). The observatory, hereafter referred to as Cabauw, is located in a rural area in the south west of the Netherlands at 51.97°N, 4.92°E (Figure 1a). The climate in the Netherlands is a typical ocean-influenced west coast climate, with relative mild and wet winters despite its high latitude, and milder summers than further inland. KNMI provides an overview of the current Dutch climate and trends on their website (https://www.knmi.nl/klimaat), including the long term increasing trend of incoming solar radiation and recent extremes (e.g. that of spring 2020, van Heerwaarden et al., 2021). The following section describes all the observational data we use from Cabauw, the supplementary modeled clear-sky irradiance, atmospheric composition re-analysis, calculated solar positions, satellite-derived cloud types, and ground based cloud cover.
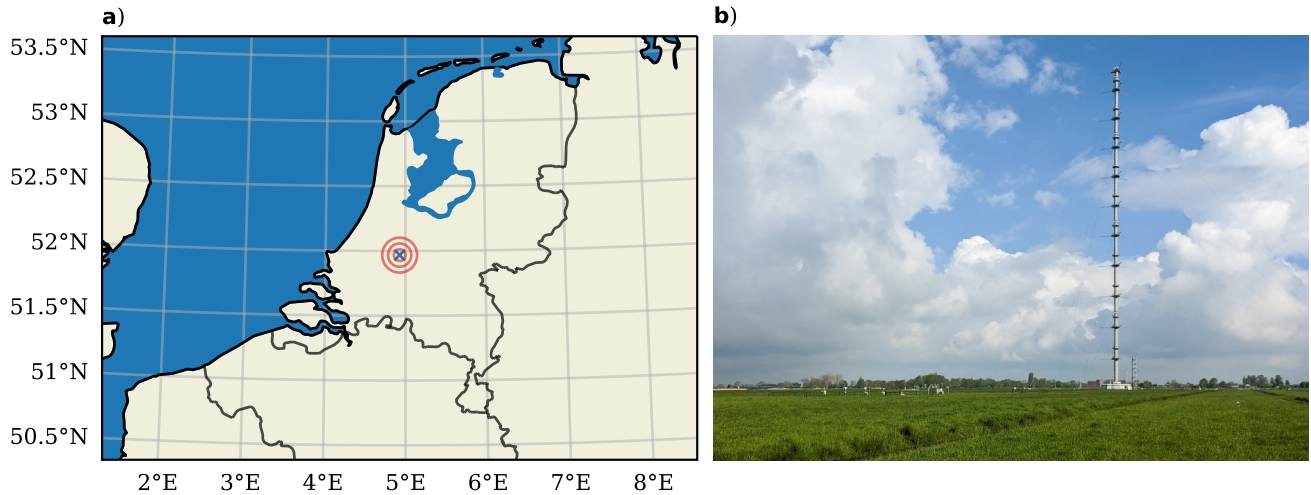
Figure 1. Ruisdael Observatory Cabauw, the Netherlands, where all observations in our dataset are located. The geographical location is marked in **(a)** with a cross, the circles are the 5 to 15 km radii for satellite cloud type extraction. A photograph of the 213 meter high tower at the Cabauw site, with the BSRN station among other instruments in the bottom left, is shown in **(b)**.

2.1 Surface solar irradiance observations

55 The surface solar irradiance station at Cabauw is part of the Baseline Surface Radiation Network (BSRN, Driemel et al., 2018), operational since 2005. The BSRN station measures all components of the surface radiation balance. Observations are logged at a 1 Hz frequency, reprocessed to 1 minute quality controlled and validated data, made available publicly at the PANGAEA data repository (Knap, 2022) with instrument metadata, maintained by station scientist Wouter Knap (KNMI). While 1 minute resolution is enough for many applications, in particular those concerned with the surface net radiation balance at longer time

60 scales, much of the cloud driven irradiance variability occurs at sub-minute scales. For the purpose of research into cloud-driven irradiance variability at these short time scales, a separate 10-year subset of solar irradiance at 1 Hz resolution has been released (Knap and Mol, 2022), and is described in here. This subset spans from 2011-02 until 2020-12. The three components are global horizontal irradiance (GHI) and diffuse irradiance (DIF), measured with Kipp & Zonen CM22 pyranometers, and direct normal irradiance (DNI), measured with a Kipp & Zonen CH1 pyrheliometer. ~~These instruments are thermopile based~~

65 ### 2.1.1 Sensor response time and resolved variability

The pyranometer and pyrheliometer instruments are thermopiles, meaning there is a non-zero response time to variations in incoming radiation, ~~and the~~ i.e. the time it takes for the thermopile to adjust to changes in irradiance signal. Thus, the true resolved resolution is not 1 Hz, though it depends on the magnitude and rate of change of irradiance variability whether this impact is noticeable. According to manufacturer's specifications, the CH1 pyrheliometer has a 7 second (95%) or 10 second

70 (99%) response time (Kipp & Zonen, 2001), and 1.66 seconds (66%) or 5 seconds (95%) for the CM22 pyranometers (Kipp & Zonen, 2004). This results in a likely underestimation of variability at 1 Hz, but the exact extent to which that happens is hard to quantify given the absence of solar irradiance measurements with fast responding sensors at a similar location, and for long enough to sample diverse weather conditions. Figure 2 illustrates the power spectral density (PSD) of a year of BSRN 1 Hz data. Tabar et al. (2014) present spectra (their Figure 2) for one year of data from a semiconductor pyranometer, thus with a >

75 1 Hz response time, which shows an order of magnitude higher PSD at 1 Hz compared to 0.1 Hz as in our Figure 2. However, these data were collected in Hawaii, which is a very different geographical location and climate than Cabauw. Alternatively, two weeks of summer time 10 Hz irradiance observations from fast radiometers (Mol and Heusinkveld, 2022) show a steeper decline between 0.1 and 1 Hz than the BSRN data. This supports the idea that the true cloud-driven irradiance variability does not always extend towards 1 Hz and higher frequencies, and that often the slow response time has no noticeable effect.
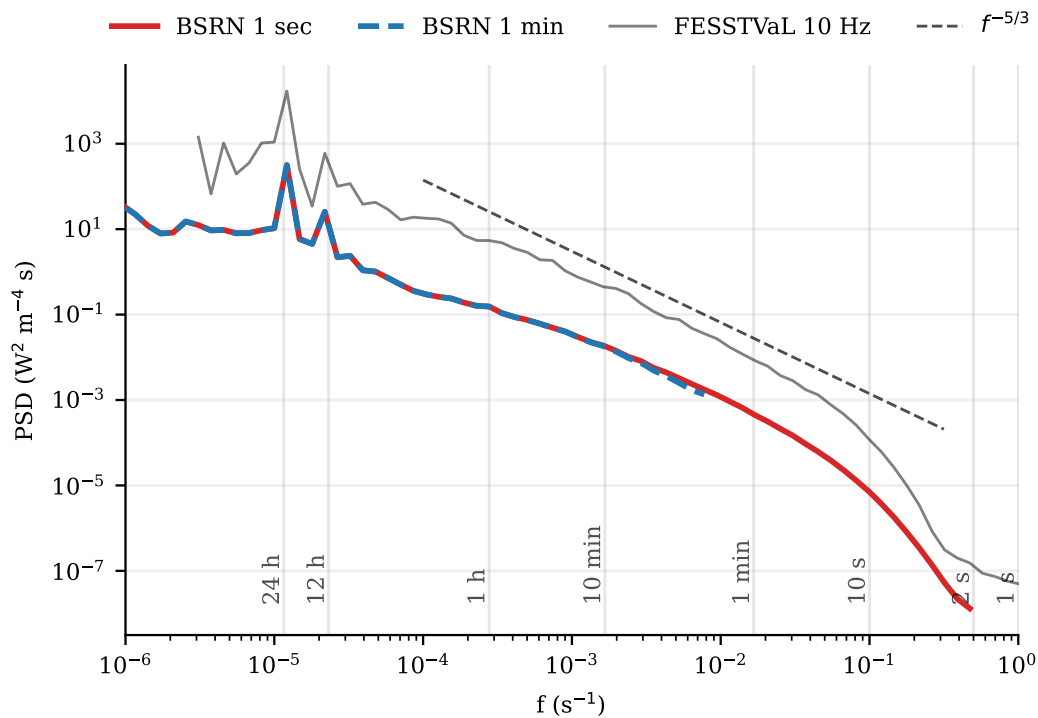
Figure 2. Power spectral density of one year (2016) of 1 Hz BSRN data (GHI component). The one minute spectrum is based on resampled 1 Hz data. As a reference, $f^{-5/3}$ scaling is added to emphasize the steep decline after 0.1 Hz towards 1 Hz. An additional comparison is the the spectrum from a semiconductor radiometer deployed during the FESSTVaL field campaign from June 14 to 30, 2021 near Berlin, Germany.

80 Based on the technical specifications of the pyranometer and pyrheliometer, we are at least confident that variability is resolved up to 0.1 Hz (10 seconds). This is also supported by Stratum et al. (2023), where they show agreement between the BSRN dataset presented in this paper and irradiance spectra from semi-realistic large-eddy simulation up to 0.1 Hz (their

Figure 6). Given the uncertainty of how much of the true variability is resolved between 0.1 and 1 Hz, we advice analyses at 1 Hz only in combination with additional constraints, such as knowledge of cloud properties and their velocity to estimate the fastest possible changes in irradiance. For example, in Mol et al. (2022a), knowledge of wind speed, cloud size distributions, and cloud edge transparency were combined to utilize the data down to 1 Hz.

One last implication is that the slow response time slightly reduces the contrast in data. Ehrlich and Wendisch (2015) demonstrate a reconstruction technique of the true 1 Hz signal through deconvolution, essentially a sharpening technique, which is not a trivial exercise. We have not applied this ~~technique~~here, as we cannot validate whether it works reliably for our dataset, but we mention it ~~here~~ as an option to anyone who might want to attempt to apply the method despite its challenges. The only pre-processing we ~~do~~ apply to the data, namely gap filling and quality control, is discussed in Section 3.1.

2.2 Supplementary irradiance data

2.2.1 Solar position and direct horizontal irradiance

Information about the sun's position is important for quality control, data analysis, and interpretation of results. Calculations of the sun's position (elevation and azimuth angle) are done using the python package PySolar (https://github.com/pingswept/pysolar) at a 1 minute resolution, linearly interpolated to 1 second. The calculations are for the purposes of this research area indistinguishable from highly accurate peer reviewed code such as the Solar Position Algorithm (SPA, https://midcdmz.nrel.gov/spa/). Using the solar elevation angle α (degrees above horizon, or as zenith angle $\theta = 90 - \alpha$), we calculate the horizontal component of direct irradiance: $DHI = DNI \cdot \sin(\alpha)$. An alternative calculation is $DHI = GHI - DIF$, which in the case of a good measurement setup should be equal to $DNI \cdot \sin(\alpha)$, and is the basis for one of the checks in data quality control (discussed in Section 3.1).

2.2.2 Clear-sky irradiance and atmospheric composition

Clear-sky global horizontal irradiance (GHI_{cs}) is the total downwelling horizontal solar irradiance in the absence of clouds. We use CAMS McClear version 3.5 (Gschwind et al., 2019) as the GHI_{cs} reference for our dataset, released in September 2022. CAMS McClear includes corrections based on atmospheric composition re-analysis such as aerosols and total column atmospheric water vapour. This allows us to define times when GHI exceeds GHI_{cs} as those purely driven by clouds as accurately as possible, as opposed to definitions based on simpler models, which is necessary for the irradiance classification algorithm described in Section 3.3. Atmospheric composition input for McClear is included in their publicly available dataset (https://www.soda-pro.com/web-services/radiation/cams-mcclear), which we add to our dataset for context. The only further processing applied to this data is linear interpolation from 1 minute to 1 second to match the irradiance observations.

2.3 Additional in-situ measurements

2.3.1 Wind profiles from Cabauw

A 213 meter high tower provides wind speed and direction measurements at 2, 10, 20, 40, 80, 140 and 200 meters above ground level at a 10 minute interval (Wauben et al., 2010). Figure 1b shows the tower with respect to the BSRN site, which is a few 115 hundred meters to the south. We apply no further processing to this data, apart from creating daily files from the monthly files. The original tower data (including temperature, visibility, and humidity) is shared publicly by the KNMI on their open data platform: https://dataplatform.knmi.nl/dataset/cesar-tower-meteo-lb1-t10-v1-2.

2.3.2 Nubiscope

Detailed cloud cover observations, for analysis, validation of satellite observations (Section 2.4), and irradiance based sky 120 type classification (Section 3.3.2), we use the observations from a nubiscope located within a few meters from the BSRN instrumentation (Wauben et al., 2010). In 10 minutes, this instruments makes a hemispherical scan of the sky using infrared sensors to determine cloud fraction and sky type of various categories. We subset the nubiscope to three years (2014 to 2016) for validation and analyses purposes. If necessary, additional data (2008-05 to 2017-04) is publicly available at https://dataplatform.knmi.nl/dataset/cesar-nubiscope-cldcov-la1-t10-v1-0. Again, no further processing is applied by us, apart 125 from turning monthly files into daily files.

2.4 Satellite observations

The satellite product CLAAS2 ~~Benas et al. (2017)~~ (Benas et al., 2017) provides cloud cover, cloud top pressure (CTP), and cloud optical thickness (COT), available every 15 minutes during day time at an approximate spatial resolution of 20 km^2 over Cabauw. We provide three years of satellite data (2014-2016) for both validation and cloud type analyses, and describe the post 130 processing steps in Section 3.2.1.

3 Processing and Methods

3.1 Quality control and completeness of irradiance data

One of the first steps in the processing is constructing daily files from the raw instrument data, which occasionally misses a few seconds of data. In such cases, we apply linear gap filling between measurement points, after which we apply quality control 135 and derive all other variables. Gap filled data points are not flagged. Data quality control for the irradiance measurements is necessary to mask out station maintenance, malfunctioning instruments, or other cases of bad data like those caused by precipitation. Maintenance happens on a ~~nearly daily basis~~ semi-daily basis (Monday, Wednesday, Friday) to ensure the high BSRN quality standards are met, and is ~~therefore~~ the most common source of anomalous measurements. It is typically brief and only involves sensor cleaning, though sometimes instrumentation is disabled or replaced due to quality issues such that

there are gaps of hours up to a few days. For the official 1-minute BSRN dataset (Knap, 2022), all measurements during such periods are filtered. The 1 Hz version includes quality flags ('good' or 'bad' data) derived from the official dataset, where 'good' means all three components (GHI, DNI, DIF) are valid. The 1 Hz version includes the original measurements, and quality flags have to be applied to filter bad data, such that the user can decide on the strictness of filtering themselves. We independently determined data quality at the 1 Hz level by performing the following checks, which is a result of a trial and error process through manual data inspection:

1. The absolute rate of change of the DIF and DNI components with respect to clear-sky between two seconds has to be below 5% and 20%, respectively.

2. The same for GHI, except the limit is 5% for cloudy conditions and 20% for sunny conditions. This leads to some false positives, which are reset if GHI and DHI changes are well-correlated.

3. Invalid measurements are padded by 180 seconds before and after to be on the safe side.

4. The residuals $\Delta Q_{abs} = |GHI - (DHI + DIF)|$ and $\Delta Q_{rel} = |GHI / (DHI + DIF) \cdot 100 - 100|$ have to be below 10% and 20 W m^{-2} for a 15 minute time frame, respectively. This time frame is necessary, because the instruments are a few meters apart, which leads to decorrelation of the individual components and larger residuals the shorter the time scale.

5. For 'good' quality, all three components have to pass the tests. If one or more components include missing or bad data, the data for that time if flagged as 'bad'.

The implementation of these rules are in the published processing code (see Section 6), and can be modified to adjust the strictness of quality control. There are only minor differences between these custom 1 Hz based quality flags and the official ones 1-min BSRN dataset flags. For all data during during day time (solar elevation angle above 0 degrees), 97.98% of the flags are similar, 1.26% are bad for custom flags and good in official, and 0.77% are good in custom but bad in official. Most of these mismatches originate from just a few days and resulting data is otherwise in very close agreement, with the vast majority of data being of good quality. Figure 3 illustrates for the whole 1 Hz dataset the data availability after (custom) quality control, and shows that most months and years have near 100% complete and good data for all three measured irradiance components. Figure 9 shows illustrates that after quality control, DIF + DHI = GHI for every month of the year, as an example data reliability. a proof data quality. The minor negative bias of DIF+DIR that is visible for some months is within 0.3 to 0.6%, which far exceeds constraints imposed by the official BSRN standard. These residuals are deemed insignificant and have no implications for the use cases of this dataset, as these concern irradiance variability.

3.2 Nubiscope and satellite processing

The nubiscope and satellite data are used both as a validation dataset for irradiance derived sky types (described in Section 3.3), and to provide observations of clouds and sky type for data analysis. Here, we describe the processing applied to the cloud observations, and how the validation dataset is created.
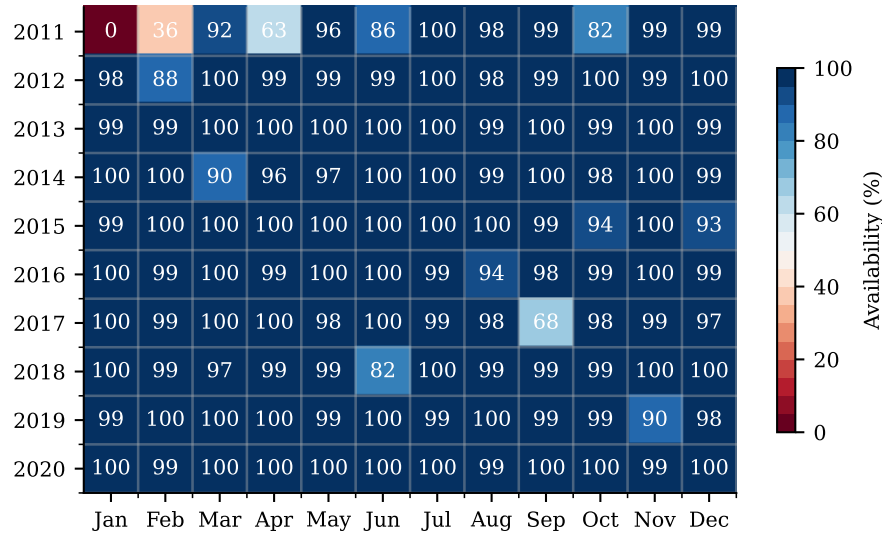
Figure 3. BSRN Cabauw 1 Hz data availability per month of available years, during day light (solar elevation angle above 0 degrees), after custom quality control. Numbers are rounded off percentages.

3.2.1 Satellite processing

We classify cloud types using a simple cloud top pressure (CTP) and cloud optical thickness (COT) categorization (see NOAA (2022) ISCCP algorithm description, their Figure 20). The 9 types are cumulus (Cu), stratocumulus (Sc), stratus (St) for low clouds; altocumulus (Ac), altostratus (As), and nimbostratus (Ns) for middle clouds; cirrus (Ci), cirrostratus (Cs), and cumulonimbus (Cb) for high clouds. Cu, Ac, and Ci are the optically thinnest clouds for each altitude, and St, Ns, and Cb the thickest, where Cb spans from low to high altitude as the only exception in this list. In this study we use it to group together cloud conditions of various altitudes and optical thicknesses in a more intuitive way, though analyses can be done on the input COT and CTP data rather than derived cloud types. The main limitations are that both the cloud fraction and actual cloud optical thickness contribute to a higher reported optical thickness in a satellite pixel, which is a result of limited spatial resolution, and higher clouds can obscure lower clouds. The spatial satellite product is converted to a time series representative for the BSRN station by determining the most common cloud class within a 5, 10, or 15 km radius around Cabauw, illustrated by the circles in Figure 1a. A smaller radius is not possible due to satellite resolution (pixel area ≈ 20 km^2) and larger radii become unrepresentative for Cabauw. Cloud cover is derived by calculating the fraction of pixels with clouds within a given radius, which is likely an overestimation due to sub-pixel cloud fractions not always being 1. Overall agreement with the nubiscope is not bad, however, as illustrated by the similar probability densities in Figure 4. Correlation coefficients between the two only show marginal improvement between 10 and 15 km. The satellite derived cloud cover overestimates the extremes at [0.0-0.1) and [0.9-1.0] compared to values bins [0.1-0.2) and [0.8-0.9), and does not have the nuances the nubiscope can resolve. For r = 5 km, there are only 4 pixels, so cloud cover from this is too coarse for most applications, but the dominant cloud type

derived from this narrow area around Cabauw is expected to be most representative. This might change for high altitude clouds at low solar elevation angles for example, requiring perhaps more sophisticated approaches, and therefore we have included the original spatial satellite fields in our dataset.
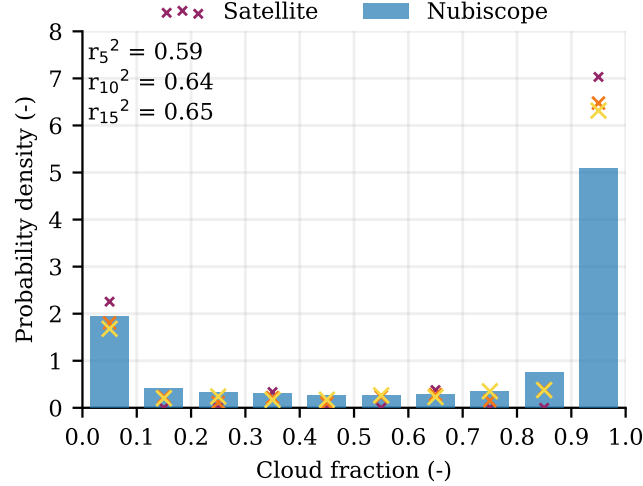
Figure 4. Comparison of cloud fraction derived from satellite to the ground-based nubiscope. The analysis is done for radii 5, 10, and 15 km, and show the probability density for 10 bins between 0 and 1 cloud fraction. Satellite radii go from 5 km (dark, small) to 15 km (light, large) cross markers. Correlation coefficients are shown in the top left for each radius. Data ranges from 2014-01 to 2016-12, and is interpolated (nearest-neighbour) to a common 5 minute time axis.

3.2.2 Validation dataset

We derive a validation dataset with clear-sky and overcast classifications based on a combination of the nubiscope and satellite data, to be used for statistical verification of sky types based on irradiance observations (Section 3.3.2). Because both instruments have their limitations, the idea is to only identify a situation as clear-sky or overcast if both datasets are in agreement. The rest of the preprocessing involves interpolation to a common 1-minute resolution grid, and masking out any data if either of the two products is missing for given a given time. The validation dataset is provided in Mol et al. (2022b) for all three radii, though we mostly use r = 10 km, for years 2014 until 2016. Disagreement between the two observational datasets is common, with the nubiscope being roughly 3 or 1.5 times more conservative with classifying a sky type as clear or overcast, respectively. This likely has to do not only with the difference in type of observation (ground-based versus remote sensing), but also in that the nubiscope is a more sensitive instrument, as we illustrated in Figure 4 and discussed in Section 3.2.1. In most cases, the nubiscope is more conservative, such that both clear-sky and overcast conditions are mostly controlled by what the nubiscope sees. Figure 8 illustrates this best, with the satellite and nubiscope differing in seasonal cycle for clear-sky conditions and in yearly averages for both clear-sky and overcast.

3.3 Irradiance classifications

The main addition to the core 1 Hz irradiance time series is the classification of measurements into various categories that describe the type of irradiance variability. We calculate two sets of classification types, one being an instantaneous classification to give a qualification to single measurement point, and the other a more indirect qualification of sky type based on longer time frames. First we describe what the classifications represent and how they are calculated, and then how they are further processed to derive a wide range of interesting statistics about irradiance variability. Examples are shown in Figures 5, 6, and 7, and the public dataset (Mol et al., 2022b) includes similar quicklooks for all 10 years of time series.
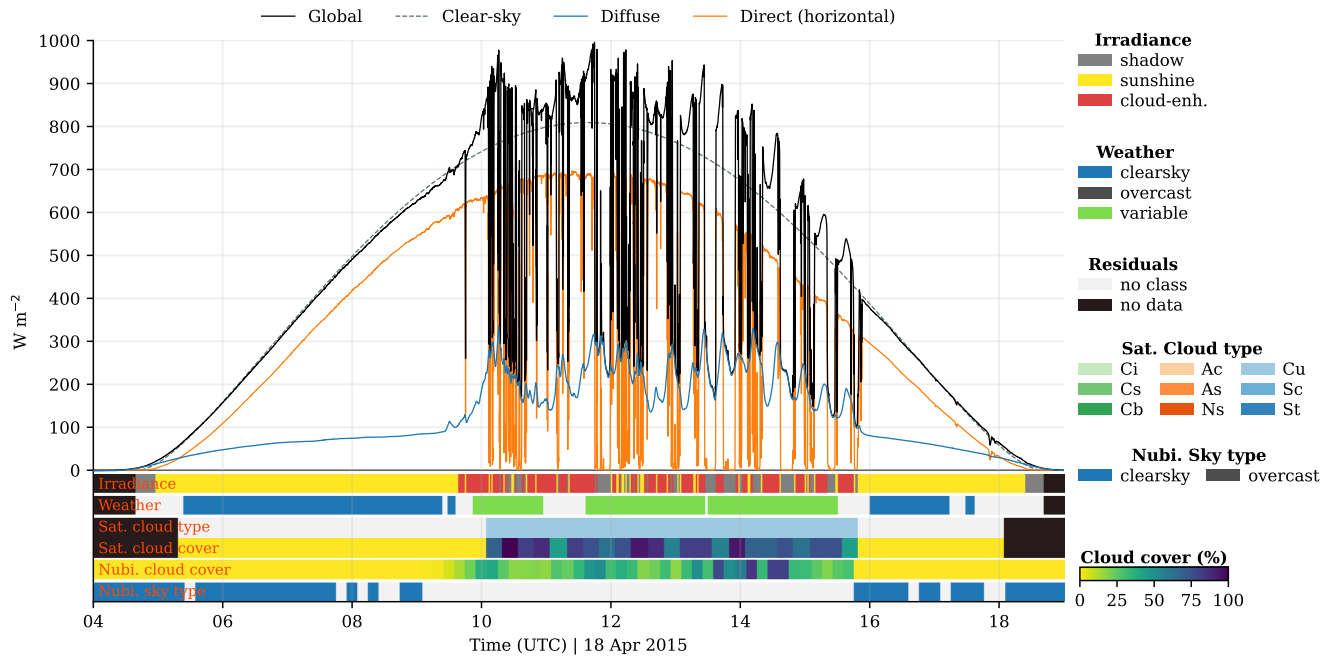
Figure 5. Surface solar irradiance time series at 1 Hz, irradiance classifications, and cloud observations for 18 April 2015. It starts with clear-sky conditions and turns into highly variable surface irradiance thanks to scattered boundary layer clouds. The three measured irradiance components, global, diffuse, and direct horizontal irradiance are shown together with modeled (CAMS McClear) clear-sky irradiance.

3.3.1 Cloud shadow and enhancement

One of the most noticeable drivers of intra-day irradiance variability are broken cloud fields making patterns of cloud shadows and cloud enhancements (e.g. Yordanov et al., 2015; Gueymard, 2017; Veerman et al., 2022). Cloud shadows are where (most) direct irradiance is blocked, and cloud enhancement where light scattered by clouds locally coincides with direct irradiance to increase global horizontal irradiance above clear-sky values. We define a shadow where direct normal irradiance (DNI) is below 120 W m^{-2}, which is the inverse of what the World Meteorological Organization defines as sunshine (DNI $\geq$ 120 W
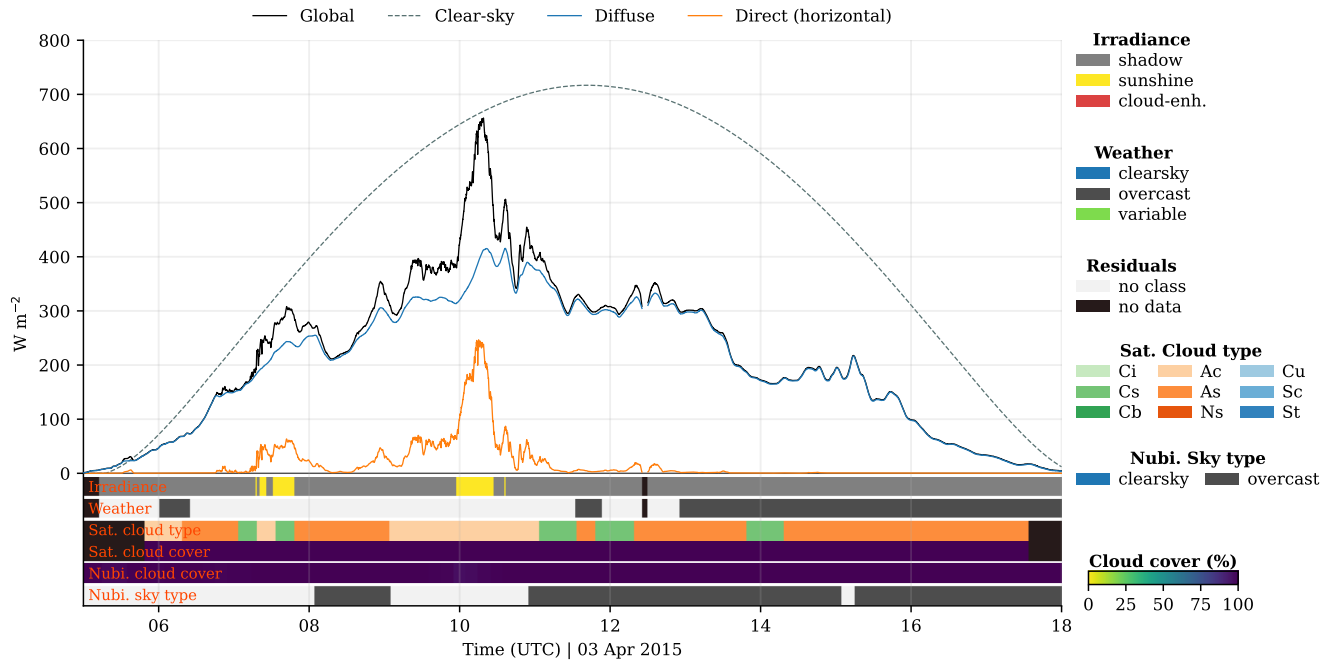
Figure 6. Surface solar irradiance observations of a mostly overcast day (3 April 2015), similar layout as Figure 5. The satellite and nubiscope observations indicate overcast conditions of mostly opaque mid to high level clouds (altostratus (As) or cirrostratus (Cs)), though there are brief periods of partially transparent clouds (altocumulus (Ac)).

m^{-2}, WMO (2014)), and a straightforward implementation. Cloud enhancement requires a more careful approach. We define cloud enhancement as a single measurement where global horizontal irradiance (GHI) exceeds the reference clear-sky (GHI$_{cs}$). In reality, observed GHI can still fluctuate noticeably in cloud-free conditions, and the GHI$_{cs}$ reference may not be perfect, such that there is some uncertainty in the detection for weak cases of cloud enhancement. To prevent false positives in the detection algorithm, we first apply an activation threshold of GHI exceeding GHI$_{cs}$ by 1% and 10 W m^{-2}. Both a relative and absolute threshold are used, as clear-sky irradiance ranges from 10^0 to 10^3 W m^{-2} as function of the solar elevation angle (i.e., time of day). 10 W m^{-2} is based on 1% of the typical order of magnitude for clear-sky irradiance around noon for Cabauw. When the threshold is reached, adjacent measurements are also marked as cloud enhancement so long as they exceed GHI$_{cs}$ by 0.1%. Edge cases at low solar elevation angles are removed by requiring DNI to be at least 10 W m^{-2}. All of these thresholds are chosen to enable us to capture all but the weakest of cloud enhancements, which arguably are not important. The residual third class is called 'sunshine', which is the WMO definition of sunshine minus cloud enhancement. Detection criteria can be adjusted in the code and recalculated, or another level of filtering can be applied after classification through the derived event statistics (discussed in Section 3.4). Examples of the classifications are shown in the top color-coded bar beneath the time series in Figures 5, 6, and 7.
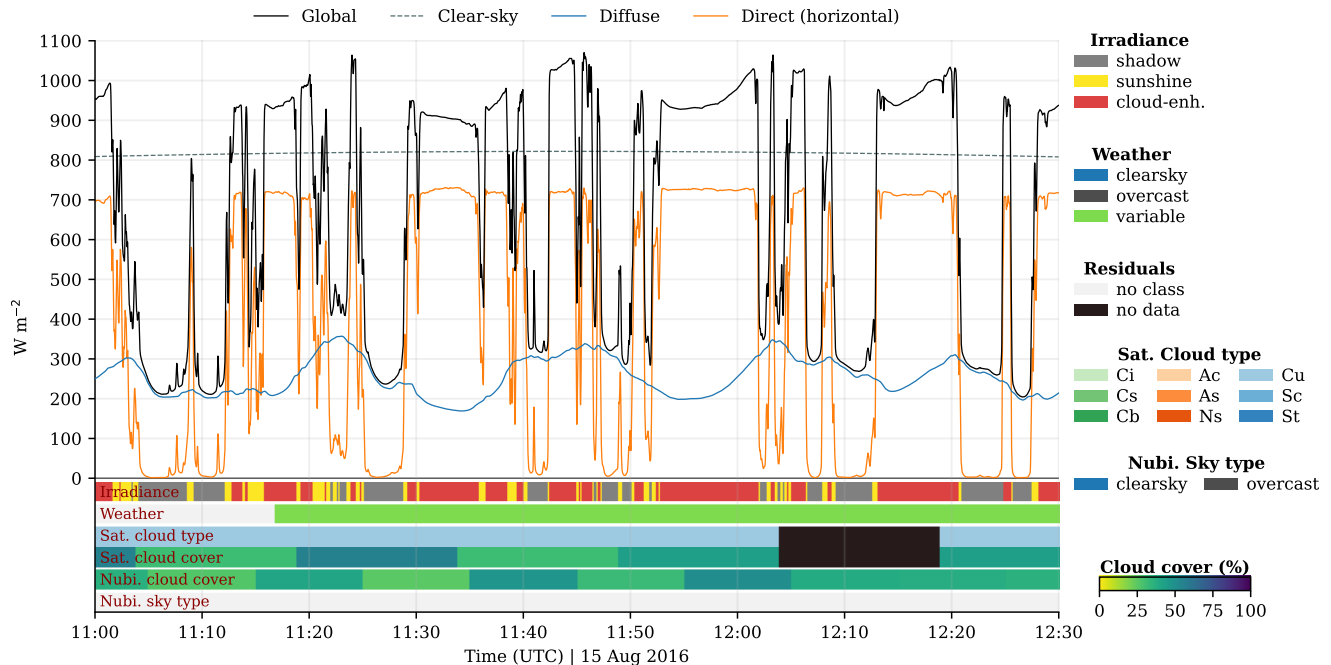
Figure 7. Detailed example of cloud-driven irradiance variability. Similar to Figure 5, but for 90 minutes of 15 August 2016.

3.3.2 Overcast, clear-sky, and variable irradiance

The second group of classifications represents irradiance 'weather' type, or sky type, based on irradiance data only. The weather types are clear-sky and overcast for smooth and predictable surface irradiance, and a third class of 'variable' irradiance representing pronounced and unpredictable 3D radiative effects to contrast the former two. The way these classifications are derived is partially based on subjective thresholds and assumes good quality clear-sky data, and thus we validate against satellite and ground-based cloud cover observations.

We classify as clear-sky those points in the time series for which, in a 15 minute centered moving window, GHI stays within 3% or 5 W m^{-2} of GHI$_{cs}$ with a maximum standard deviation of the ratio GHI/GHI$_{cs}$ = 0.01 within that window. This irradiance based algorithm emphasizes smoothness, and thus predictability, more than exactly matching GHI$_{cs}$, so as to not rely too much on CAMS McClear being perfectly accurate. Clear-sky conditions are uncommon, occurring between 5 to 15% of the time depending on observational method (Figure 8). Skill scores indicate that the irradiance based classification misses almost half the cases (probability of detection close to 50%), and is generally too conservative (bias < 1) with respect to the validation dataset (see Table 1). The negative bias against the validation dataset, which is not what Figure 8 shows, is due to the skill scores being calculated for cases only when there is agreement between the satellite and nubiscope, rather than for all available data in Figure 8. The order of magnitude of occurrence is similar to what the validation dataset suggests, but the seasonal cycle is not reproduced, though seasonality between the irradiance classification and satellite alone are similar. For

Table 1. Skill scores for irradiance based sky type classifications compared to the validation dataset (satellite + nubiscope). Scores are based on a contingency table approach. POD is Probability Of Detection, FAR means False Alarm Ratio.

Year	Accuracy	Bias	POD	FAR	Samples (hours)
Clear-sky					
2014	0.969	0.624	0.478	0.233	3346
2015	0.962	0.610	0.494	0.191	3311
2016	0.968	0.666	0.486	0.271	3244
Overcast					
2014	0.951	0.979	0.912	0.068	2914
2015	0.954	0.971	0.914	0.059	2946
2016	0.956	0.980	0.923	0.058	2872
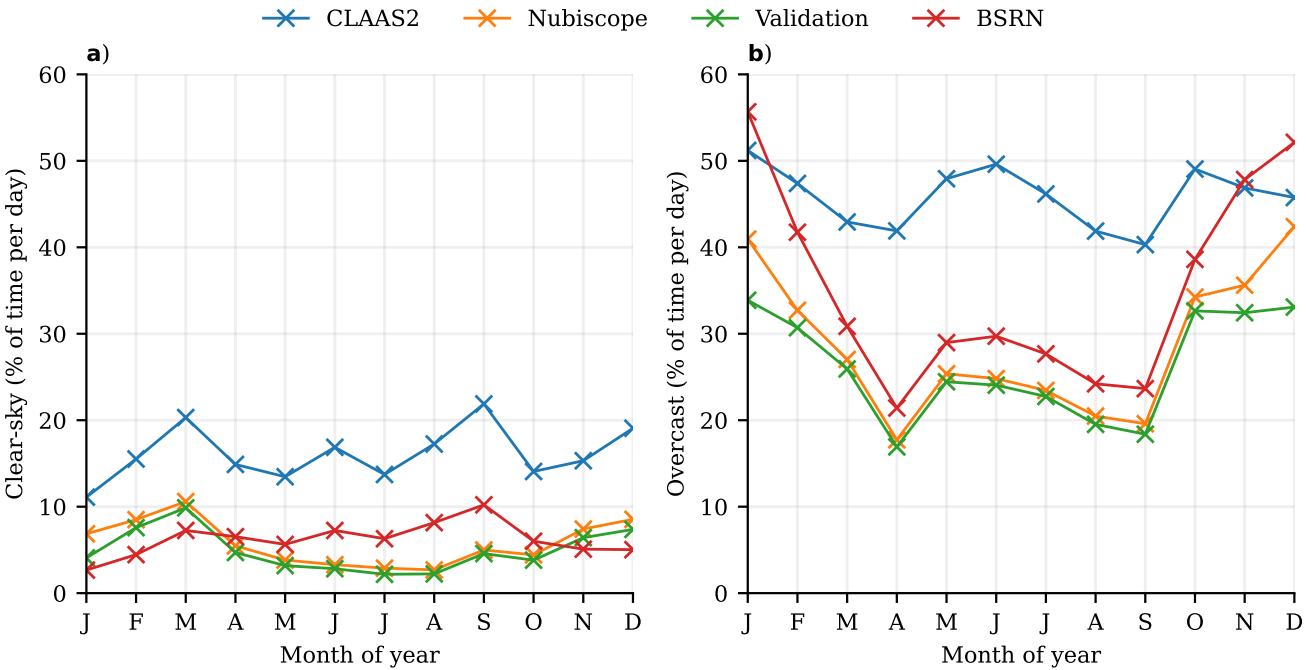
Figure 8. Comparison of sky type classifications based on satellite time series (r = 10 km), nubiscope, and 1 Hz irradiance observations. The validation period is from 2014 until 2016, in units relative to available data during daylight (solar elevation angle $\alpha > 0$ degrees).

2014 and 2015, the seasonal correlation to the nubiscope is also much better, but in 2016 there were many cases where the nubiscope saw thin cirrus (cloud cover $< 5\%$, August 17, 18, 23, 24, and 25) rather than clear-sky. In all these cases, GHI $<$ GHI$_{cs}$, consistent with thin cirrus attenuating incoming solar radiation slightly, and not part of the validation set because of disagreement between satellite and nubiscope. It appears (from Table 1 and Figure 10) that there is poor skill in the irradiance-based classification, though manual inspection of quicklooks gives a different impression and most of the bias shown in Figure 10 appears to stem from cases with thin cirrus. If one wants to filter out the thin cirrus cases, the classification threshold can be set more strict and thereby limit cases to more true clear-sky. Figure 6 shows a case with overall agreement between the nubiscope, satellite, and irradiance based clear-sky classification. We refer the reader to the public dataset (Mol et al., 2022b) with time series quicklooks for many more examples.

We define overcast weather as a period of 45 minutes for which the sum of DNI is below 1% of GHI$_{cs}$, and the average below 10 W m^{-2} to catch rare edge cases. This class is indicative of continuous, optically thick and persistent cloud cover, which is a common occurrence in the Netherlands (20 to 50% of the time depending on the season) and does well against the validation dataset. Probability of detection is high (92%), with 6% false alarms and only a slight negative bias of -1.7%. Scores slightly move toward positive (+2.6%) or negative (-5.1%) bias when shortening or lengthening the moving window to 15 minutes and 60 minutes respectively. Although a cloud cover of 100% as seen by the nubiscope and satellite is be classified as overcast in the validation dataset, cloud cover does not imply the optical thickness is high enough to block all direct irradiance. This distinction is illustrated in Figure 6, which show good agreement for overcast conditions overall, but there is still some irradiance variability with 100% cloud cover around 10 UTC. This example emphasizes our definition of overcast as smooth and predictable diffuse irradiance weather as opposed to a sky type with 100% cloud cover. The seasonal cycle between the irradiance based sky type and nubiscope correlate well (Figure 10), whereas the satellite yearly cycle is less pronounced.

Finally, the variable weather class is defined as any 60-minute window in which 10 transitions from a shadow to cloud enhancement or vice versa occur, built upon the instantaneous classifications defined in Section 3.3.1. It is indicative of weather associated with a characteristic bi-modal distribution of irradiance, cloud enhancements, and at least a handful of large fluctuations in a short time frame, all of which current numerical weather prediction models cannot reproduce. This classification ~~doesn~~ does well in locating highly variable irradiance conditions, of which examples are shown in Figure 5 and 7, and can for example be used to find case studies.

3.4   Event statistics

Within the classified irradiance time series, we call sections of cloud enhancements or shadows 'events'. The 10 years of irradiance time series contain 184,447 cloud shadows and 186,685 cloud enhancement events. For every event, the start and end time are used to select complementary radiation and meteorological data, such that every cloud enhancement and shadow event can be characterized. Notable examples are statistics of event duration, maximum cloud enhancement strength, minimum direct irradiance 'min(DNI)', mean 200 meter wind speed, dominant cloud type, maximum cloud top height, and mean solar elevation angle. Event statistics such as these allows for filtering of events according to additional criteria, e.g. for comparing

events of different magnitudes or finding the most extreme cases of cloud enhancement for a given cloud type. Event statistics for cloud shadows and enhancements for are included in the public dataset (Mol et al., 2022b).

3.5 Daily statistics

For case study selection or climatological overviews, we calculate daily statistics, which are mostly aggregates of irradiance and classification data. This statistic file is included in the public dataset. We use this to create figures such as Figures 3, 9, and 10, and to find specific case studies as detailed in Table 2.

4 Examples and use cases

The following section provides some examples and (potential) use cases of the dataset, including previously completed work. Veerman et al. (2022) and Tijhuis et al. (2022) research 3D radiative transfer modelling approaches for cumulus case studies, where 1 Hz irradiance time series and statistics are used as validation. In Mol et al. (2022a), we show how the spatiotemporal scales of cloud shadows and enhancements are described by power laws and driven by cloud size distributions, using the event statistics as described in Section 3.4.

Figures 9 and 10 give an overview of the seasonal and yearly variability of solar irradiance and its classifications that characterize the mid latitude climate of Cabauw. Figure 9 also partially serves ~~are~~ as validation of the BSRN instrumentation, with the direct and diffuse components adding up to GHI as should be the case. Figure 10 illustrates the typically overcast conditions during winter, and highly variable irradiance conditions during summer with significant year to year variability.

In order to find specific types of case studies for analysis, you can use either the event statistics or daily statistics to query and filter specific conditions. As an example, we use the daily statistics file and Python's Xarray to find case studies of the most variable irradiance throughout the day or specific cases where overcast conditions transition to clear-sky (or vice versa), which have a potential for brief periods of strong variability and cloud enhancement. These example are shown in Table 2, and the code is publicly available (Section 6).

5 Conclusions

In this paper, we describe a high resolution, 10-year long observational dataset of detailed surface solar irradiance, complemented with meteorological data. Using time series classification algorithms, we derive statistics about sky type and irradiance variability. We provide examples and use cases of this dataset to illustrate its potential, ranging from case study selection and model validation to fundamental insight into drivers of irradiance variability. With all data and processing code publicly available, the user is free to modify our classification algorithm to their liking and validate against independent observations, or expand upon the large set of statistics already provided. Quicklooks for all available days from 2011-02 until 2020-12 are provided to get familiar with the dataset contents and get an impression of the many different types of weather conditions at
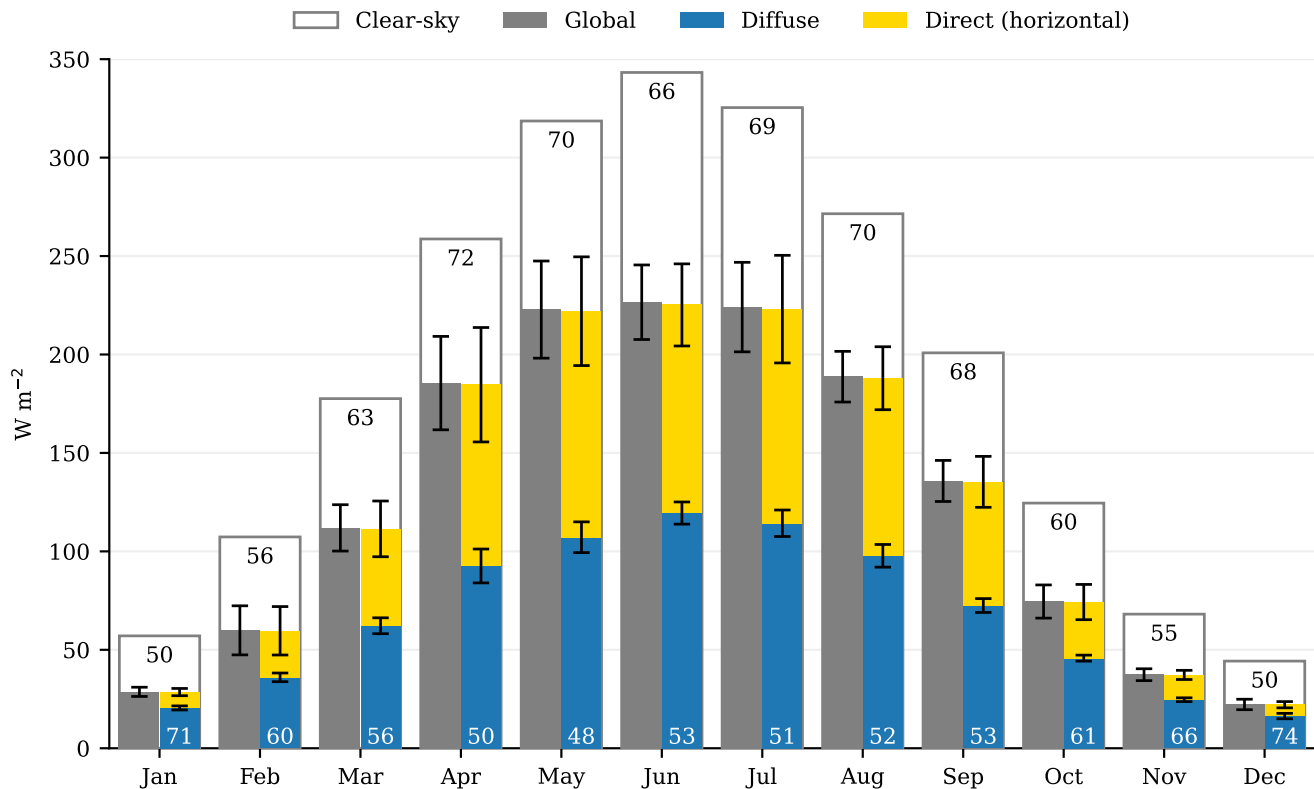
Figure 9. Surface irradiance climate for period 2011-02 until 2020-12 based on the 1 Hz dataset of all three components. The numbers in the bottom indicate the percentage of diffuse irradiance (DIF) to global horizontal irradiance (GHI). The error bars indicate, for each component, the year-to-year standard deviation. The white bars encompassing the three components is the clear-sky irradiance (GHI$_{cs}$) for each month, based CAMS McClear, where the numbers in the top are the percentage of GHI compared to GHI$_{cs}$. Only days with $> 95\%$ of data completeness are included.

Cabauw. We believe this dataset is of great use in research into cloud driven irradiance variability, and it provides a necessary validation reference for models that are starting to resolve the full spectrum of variability.

6 Code and data availability

1 Hz GHI, DIF and DNI observations of the BSRN station at Cabauw are published on Zenodo (Knap and Mol, 2022), Irradiance time series classifications, supplementary data, quality control, event and daily statistics, and satellite data are published as a separate, complementary dataset on Zenodo (Mol et al., 2022b). Satellite data for an area around Cabauw is taken from the CLAAS2 open access dataset described in Benas et al. (2017), and included in the previous dataset for 2014 to 2016. Also included there is the nubiscope data (Wauben et al., 2010), taken from the KNMI Data Platform (https:
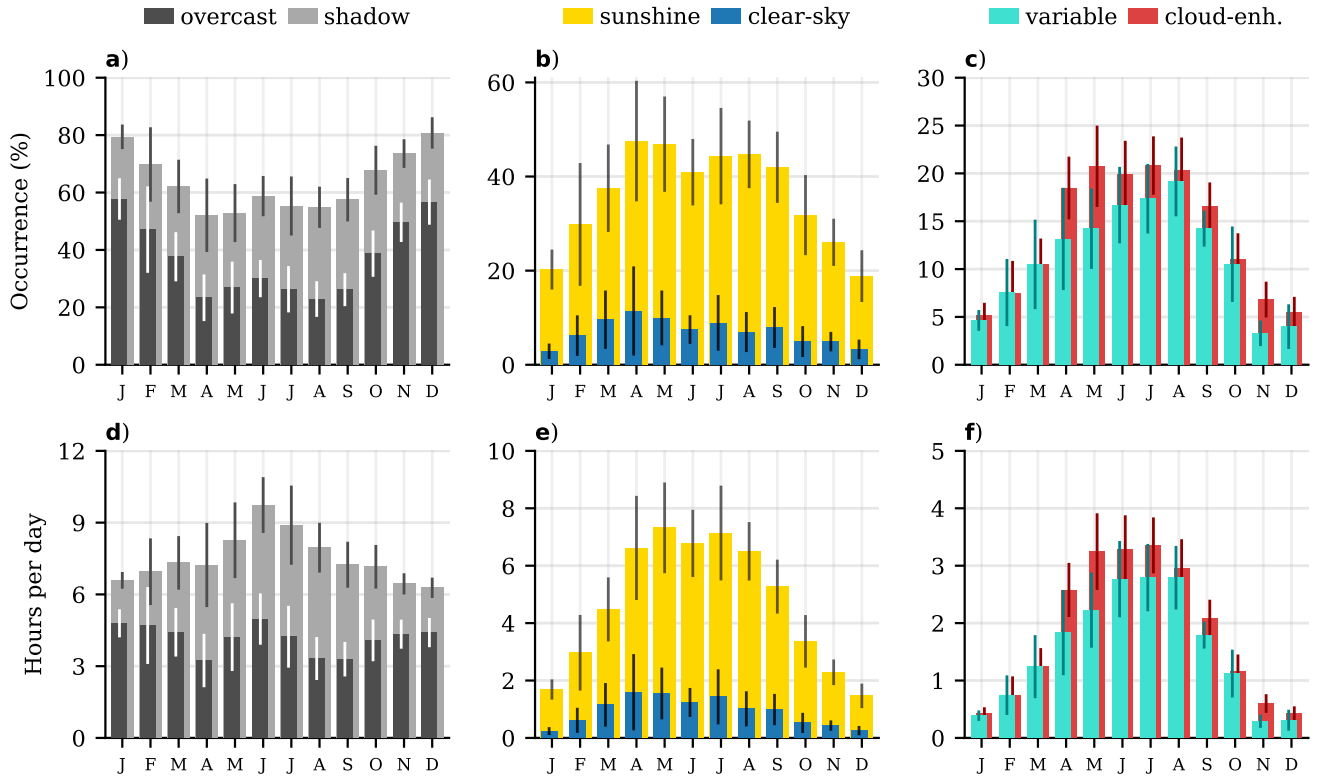
Figure 10. Instantaneous and weather classification ~~climatology~~ occurrence per month showing the relative (top row) and absolute (bottom row) ~~occurrence~~ climatology of each classification throughout the year for all available data (2011-02 to 2020-12). Here, sunshine, also includes the portion marked as cloud enhancement, such that shadow + sunshine = 100%. See Section 3.3 for the six classification definitions shown here. The relative occurrence is expressed as a percentage of daylight (solar elevation angle > 0°), absolute occurrence is expressed in average hours per day. Error bars indicate the year-to-year standard deviation.

//dataplatform.knmi.nl/dataset/cesar-nubiscope-cldcov-la1-t10-v1-0) for years 2014 to 2016. All code to reproduce the classifications from the irradiance observations, event and daily statistics, and figures presented in this paper is archived at https://zenodo.org/record/7851741.

Author contributions. WM performed the data analysis, produced the public datasets, and wrote the manuscript. WK is the BSRN Cabauw station scientist and has maintained and produced the original 1 Hz irradiance dataset. CvH has helped shape the manuscript and is project PI. All authors have contributed to the final manuscript.

Table 2. Examples of finding case studies using daily statistics. The table shows for a custom period (2014-2016) the top 5 cases for two example queries. The first is the absolute most variable irradiance weather. The second are cases with at least 5% overcast, variable and clear-sky, sorted by those with the most variability, which is a way to find case studies with overcast to clear-sky transitions.

Top cases	Most variable	Overcast <-> clear-sky
1	2015-07-22	2015-04-30
2	2015-07-17	2016-05-17
3	2016-07-02	2015-05-12
4	2014-08-09	2016-08-06
5	2015-06-18	2016-04-19

Competing interests. The authors declare that they have no conflict of interest.

325 *Acknowledgements.* WM and CvH acknowledge funding from the Dutch Research Council (NWO) (grant: VI.Vidi.192.068).

References

Benas, N., Finkensieper, S., Stengel, M., van Zadelhoff, G.-J., Hanschmann, T., Hollmann, R., and Meirink, J. F.: The MSG-SEVIRI-based cloud property data record CLAAS-2, Earth System Science Data, 9, 415–434, https://doi.org/https://doi.org/10.5194/essd-9-415-2017, publisher: Copernicus GmbH, 2017.

330 Driemel, A., Augustine, J., Behrens, K., Colle, S., Cox, C., Cuevas-Agulló, E., Denn, F. M., Duprat, T., Fukuda, M., Grobe, H., Haeffelin, M., Hodges, G., Hyett, N., Ijima, O., Kallis, A., Knap, W., Kustov, V., Long, C. N., Longenecker, D., Lupi, A., Maturilli, M., Mimouni, M., Ntsangwane, L., Ogihara, H., Olano, X., Olefs, M., Omori, M., Passamani, L., Pereira, E. B., Schmithüsen, H., Schumacher, S., Sieger, R., Tamlyn, J., Vogt, R., Vuilleumier, L., Xia, X., Ohmura, A., and König-Langlo, G.: Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017), Earth System Science Data, 10, 1491–1501, https://doi.org/10.5194/essd-10-1491-2018, publisher: Copernicus GmbH, 2018.

Durand, M., Murchie, E. H., Lindfors, A. V., Urban, O., Aphalo, P. J., and Robson, T. M.: Diffuse solar radiation and canopy photosynthesis in a changing environment, Agricultural and Forest Meteorology, 311, 108 684, https://doi.org/10.1016/j.agrformet.2021.108684, 2021.

Ehrlich, A. and Wendisch, M.: Reconstruction of high-resolution time series from slow-response broadband terrestrial irradiance measurements by deconvolution, Atmospheric Measurement Techniques, 8, 3671–3684, https://doi.org/10.5194/amt-8-3671-2015, publisher: Copernicus GmbH, 2015.

Gristey, J. J., Feingold, G., Glenn, I. B., Schmidt, K. S., and Chen, H.: Surface Solar Irradiance in Continental Shallow Cumulus Fields: Observations and Large-Eddy Simulation, J. Atmos. Sci., 77, 1065–1080, https://doi.org/10.1175/JAS-D-19-0261.1, publisher: American Meteorological Society, 2020.

Gschwind, B., Wald, L., Blanc, P., Lefèvre, M., Schroedter-Homscheidt, M., and Arola, A.: Improving the McClear model estimating the downwelling solar radiation at ground level in cloud-free conditions – McClear-v3, metz, 28, 147–163, https://doi.org/10.1127/metz/2019/0946, 2019.

Gueymard, C. A.: Cloud and albedo enhancement impacts on solar irradiance using high-frequency measurements from thermopile and photodiode radiometers. Part 1: Impacts on global horizontal irradiance, Solar Energy, 153, 755–765, https://doi.org/10.1016/j.solener.2017.05.004, 2017.

350 Helbig, M., Gerken, T., Beamesderfer, E. R., Baldocchi, D. D., Banerjee, T., Biraud, S. C., Brown, W. O. J., Brunsell, N. A., Burakowski, E. A., Burns, S. P., Butterworth, B. J., Chan, W. S., Davis, K. J., Desai, A. R., Fuentes, J. D., Hollinger, D. Y., Kljun, N., Mauder, M., Novick, K. A., Perkins, J. M., Rahn, D. A., Rey-Sanchez, C., Santanello, J. A., Scott, R. L., Seyednasrollah, B., Stoy, P. C., Sullivan, R. C., de Arellano, J. V.-G., Wharton, S., Yi, C., and Richardson, A. D.: Integrating continuous atmospheric boundary layer and tower-based flux measurements to advance understanding of land-atmosphere interactions, Agricultural and Forest Meteorology, 307, 108 509, https://doi.org/10.1016/j.agrformet.2021.108509, 2021.

Jakub, F. and Mayer, B.: A three-dimensional parallel radiative transfer model for atmospheric heating rates for use in cloud resolving models—The TenStream solver, Journal of Quantitative Spectroscopy and Radiative Transfer, 163, 63–71, https://doi.org/10.1016/j.jqsrt.2015.05.003, 2015.

Kipp & Zonen: CH1 Pyrheliometer Instruction Manual, retrieved 26-07-2022 from https://www.kippzonen.com/Download/42/CH-1-Pyrheliometer-Manual-English, 2001.

360 Kipp & Zonen: CM22 precision pyranometer instruction manual, retrieved 26-07-2022 from https://www.kippzonen.com/Download/55/CM-22-Pyranometer-Manual, 2004.

Kivalov, S. N. and Fitzjarrald, D. R.: Quantifying and Modelling the Effect of Cloud Shadows on the Surface Irradiance at Tropical and Midlatitude Forests, Boundary-Layer Meteorol, 166, 165–198, https://doi.org/10.1007/s10546-017-0301-y, 2018.

365 Knap, W.: Basic and other measurements of radiation at station Cabauw (2005-02 et seq), https://doi.org/10.1594/PANGAEA.940531, 2022.

Knap, W. H. and Mol, W. B.: High resolution solar irradiance variability climatology dataset part 1: direct, diffuse, and global irradiance, https://doi.org/10.5281/zenodo.7093164, 2022.

Liang, X.: Emerging Power Quality Challenges Due to Integration of Renewable Energy Sources, IEEE Transactions on Industry Applications, 53, 855–866, https://doi.org/10.1109/TIA.2016.2626253, 2017.

370 Lohmann, G. M.: Irradiance Variability Quantification and Small-Scale Averaging in Space and Time: A Short Review, Atmosphere, 9, 264, https://doi.org/10.3390/atmos9070264, number: 7 Publisher: Multidisciplinary Digital Publishing Institute, 2018.

Lohmann, G. M., Monahan, A. H., and Heinemann, D.: Local short-term variability in solar irradiance, Atmospheric Chemistry and Physics, 16, 6365–6379, https://doi.org/https://doi.org/10.5194/acp-16-6365-2016, publisher: Copernicus GmbH, 2016.

Mercado, L. M., Bellouin, N., Sitch, S., Boucher, O., Huntingford, C., Wild, M., and Cox, P. M.: Impact of changes in diffuse radiation on

375   the global land carbon sink, Nature, 458, 1014–1017, https://doi.org/10.1038/nature07949, number: 7241 Publisher: Nature Publishing Group, 2009.

Mol, W. and Heusinkveld, B.: Radiometer grid at Falkenberg and surroundings, downwelling shortwave radiation, FESSTVaL campaign, https://doi.org/10.25592/uhhfdm.10273, Project: FESSTVaL (Field Experiment on submesoscale spatio-temporal variability in Lindenberg), a measurement campaign initiated by the Hans-Ertel-Center for Weather Research., 2022.

380 Mol, W., van Stratum, B., Knap, W., and van Heerwaarden, C.: Reconciling observations of solar irradiance variability with cloud size distributions, https://doi.org/10.48550/arXiv.2209.10284, arXiv:2209.10284 [physics], 2022a.

Mol, W. B., Knap, W. H., and van Heerwaarden, C. C.: High resolution solar irradiance variability climatology dataset part 2: classifications, supplementary data, and statistics, https://doi.org/10.5281/zenodo.7462362, Part of the Shedding Light On Cloud Shadows (SLOCS) project, funded by the Dutch Research Council (NWO), grant number VI.Vidi.192.068, 2022b.

385 NOAA: Climate Algorithm Theoretical Basis Document, Tech. rep., https://ncei.noaa.gov/pub/data/sds/cdr/CDRs/Cloud_Properties-ISCCP/AlgorithmDescription_01B-29.pdf, 2022.

Pearcy, R. W. and Way, D. A.: Two decades of sunfleck research: looking back to move forward, Tree Physiology, 32, 1059–1061, https://doi.org/10.1093/treephys/tps084, 2012.

Stratum, B. v., Heerwaarden, C. C. v., and Arellano, J. V.-G. d.: The Benefits and Challenges of Downscaling a Global Reanalysis with

390   Doubly-Periodic Large-Eddy Simulations, preprint, Preprints, https://doi.org/10.22541/essoar.168167362.25141062/v1, 2023.

Tabar, M. R. R., Anvari, M., Lohmann, G., Heinemann, D., Wächter, M., Milan, P., Lorenz, E., and Peinke, J.: Kolmogorov spectrum of renewable wind and solar power fluctuations, Eur. Phys. J. Spec. Top., 223, 2637–2644, https://doi.org/10.1140/epjst/e2014-02217-8, 2014.

Tijhuis, M., van Stratum, B., van Heerwaarden, C., and Veerman, M.: An Efficient Parameterization for Surface Shortwave 3D Radiative

395   Effects in Large-Eddy Simulations of Shallow Cumulus Clouds, https://doi.org/10.1002/essoar.10511758.2, 2022.

van Heerwaarden, C. C., Mol, W. B., Veerman, M. A., Benedict, I., Heusinkveld, B. G., Knap, W. H., Kazadzis, S., Kouremeti, N., and Fiedler, S.: Record high solar irradiance in Western Europe during first COVID-19 lockdown largely due to unusual weather, Communications Earth & Environment, 2, 1–7, https://doi.org/10.1038/s43247-021-00110-0, number: 1 Publisher: Nature Publishing Group, 2021.

Veerman, M. A., van Stratum, B. J. H., and van Heerwaarden, C. C.: A case study of cumulus convection over land in cloud-resolving simulations with a coupled ray tracer, Geophysical Research Letters, n/a, e2022GL100 808, https://doi.org/10.1029/2022GL100808, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2022GL100808, 2022.

Wauben, W., Bosveld, F., and Baltink, H. K.: Laboratory and field evaluation of the Nubiscope, WMO TECO, Helsinki, 2010.

WMO: Chapter 8 - Measurement of sunshine duration, https://library.wmo.int/index.php?lvl=notice_display&id=19664, accessed 06-Jul-2022, 2014.

Yang, D., Wang, W., Gueymard, C. A., Hong, T., Kleissl, J., Huang, J., Perez, M. J., Perez, R., Bright, J. M., Xia, X., van der Meer, D., and Peters, I. M.: A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality, Renewable and Sustainable Energy Reviews, 161, 112 348, https://doi.org/10.1016/j.rser.2022.112348, 2022.

Yordanov, G. H.: A study of extreme overirradiance events for solar energy applications using NASA's I3RC Monte Carlo radiative transfer model, Solar Energy, 122, 954–965, https://doi.org/10.1016/j.solener.2015.10.014, 2015.

Yordanov, G. H., Saetre, T. O., and Midtgård, O.: 100-millisecond Resolution for Accurate Overirradiance Measurements, IEEE Journal of Photovoltaics, 3, 1354–1360, https://doi.org/10.1109/JPHOTOV.2013.2264621, conference Name: IEEE Journal of Photovoltaics, 2013.

Yordanov, G. H., Saetre, T. O., and Midtgård, O.-M.: Extreme overirradiance events in Norway: 1.6 suns measured close to 60°N, Solar Energy, 115, 68–73, https://doi.org/10.1016/j.solener.2015.02.020, 2015.