Norwegian
Meteorological
Institute

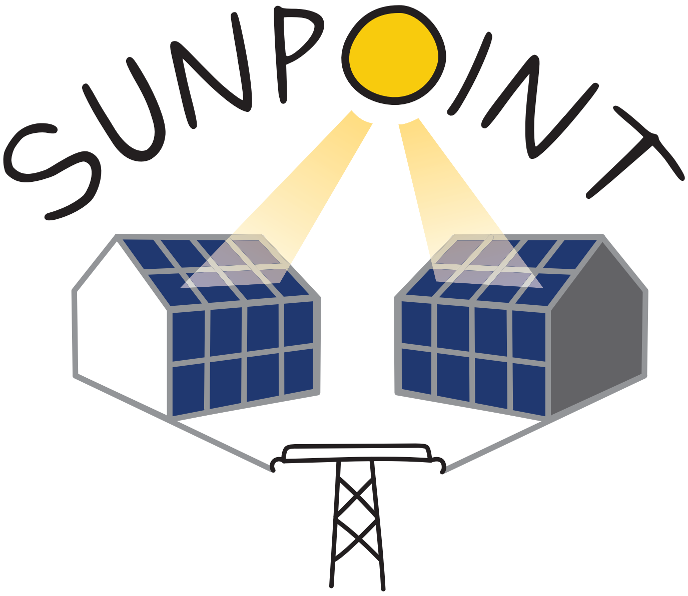
METreport

No. 8/2024
ISSN 2387-4201
Observations

Quality control of radiation data for solar resource mapping

SunPoint project report

Andreas Dobler[1], Erik Berge[1], Hans Olav Hygen[1],
Sigbjørn Grini[2], Heine Nygard Riise[3]
[1]Norwegian Meteorological Institute
[2]Norconsult AS
[3]Institute for Energy Technology (IFE)



Norwegian
Meteorological
Institute

METreport

Title Quality control of radiation data for solar resource mapping	**Date** 17.04.2024
Section Division for model and climate analysis, Division for climate modelling and air pollution and Division for climate services	**Report no.** 8/2024
Author(s) A. Dobler, E. Berge, H. O. Hygen, S. Grini, H. N. Riise	**Classification** ● Free ○ Restricted
Client Norwegian Research Council project *SunPoint*: SUn in Norway – POtential and INTegration of the solar energy resource	**Client's reference** NFR-320750

Abstract

This report describes a methodology for selecting and quality controlling hourly global solar irradiation measurements in Norway. Most of the considered data is openly available through the Frost API by the Norwegian Meteorological Institute `frost.met.no`. Initially, 106 stations have been taken into account. Requirements for data coverage and visual quality assessments resulted in a refined dataset of 47 stations. The data from these stations underwent an automated quality control (QC) routine, leading to the flagging of data points exceeding thresholds of upper and lower theoretical limits.

The resulting dataset and the QC flags have been made available in NetCDF format along with the visual quality plots. One of the main challenges in the QC procedure has been the estimation of clear-sky values for high altitudes. This has been addressed using data provided by the CAMS McClear Clear-Sky Irradiation service. Overall, the QC procedure flagged 28% of the hourly data. Removing the flagged data reduces the estimated average solar potential at the selected stations by about 1% from $872\ kWh/m^2$ to $863\ kWh/m^2$.

Keywords

Global solar irradiation, FROST, observations

Disciplinary signature

Responsible signature

Abstract

This report describes a methodology for selecting and quality controlling hourly global solar irradiation measurements in Norway. Most of the considered data is openly available through the Frost API by the Norwegian Meteorological Institute `frost.met.no`. Initially, 106 stations have been taken into account. Requirements for data coverage and visual quality assessments resulted in a refined dataset of 47 stations. The data from these stations underwent an automated quality control (QC) routine, leading to the flagging of data points exceeding thresholds of upper and lower theoretical limits.

The resulting dataset and the QC flags have been made available in NetCDF format along with the visual quality plots. One of the main challenges in the QC procedure has been the estimation of clear-sky values for high altitudes. This has been addressed using data provided by the CAMS McClear Clear-Sky Irradiation service. Overall, the QC procedure flagged 28% of the hourly data. Removing the flagged data reduces the estimated average solar potential at the selected stations by about 1% from 872 kWh/m^2 to 863 kWh/m^2.

Contents

1 Introduction

In this report, a quality control (QC) procedure for global solar irradiation measurements in Norway is described. The methodology has been developed and applied within the Norwegian Research Council project *SunPoint*: SUn in Norway – POtential and INTegration of the solar energy resource (NFR-320750). The main focus lies on using the data for mapping the solar resource potential in Norway. However, the quality and evaluation of radiation measurements is also of interest in other applications, e.g. hydrological modelling or the bias-adjustment of model data.

The measurements have been collected from the open database of the Norwegian meteorological institute `frost.met.no` and extended with a few additional stations directly collected within the SunPoint project. The evaluation has been carried out over the reference period 2016–2020.

This report provides information on the QC and automated flagging routines applied to the measurement data. An introduction to the theory of solar geometry and to solar radiation measurement instruments can for instance be found in *Grini* (2015). Chapter 2 gives an overview on the data used. Chapter 3 describes the methods used for the selection of stations and shows the selected stations. In chapter 4 an overview on the QC flagging procedure is provided. In chapter 5 the QC method and its implications are discussed and chapter 6 gives a short summary of the report and our findings.

The complete and the quality controlled measurements, this report, the output of the visualizations and the scripts used for the visualization and QC are openly available at `https://zenodo.org/records/8082726`.

2 Data

Data retrieval from `frost.met.no` is following the Climate and Forecast (CF) standard names (see `https://cfconventions.org/standard-names.html`). The CF standard name for global solar irradiation is *surface_downwelling_shortwave_flux_in_air* and the variable name is *rsds* (radiation;shortwave;downwelling;surface) with units W/m^2. It is the sum of direct and diffuse solar radiation incident on the surface per unit area.

Sometimes this is also referred to as *global radiation* (e.g. on `seklima.met.no`; *globalstråling* in Norwegian). The measurements from `frost.met.no` are radiation fluxes through a horizontal surface. The common term for this within the solar energy sector is *global horizontal irradiance* (GHI) which will be used in the following. The annual GHI, or solar potential, is then the sum of all hourly GHI values in a year (with unit kWh/m^2)

2.1 Station data

As of 02.05.2022, there have been 104 stations listed in `frost.met.no` providing hourly GHI data. For this project, additional station data has been provided by the Norwegian University of Life Sciences (NMBU) for Ås, the University in Oslo (UiO) for Finse, Jotun A/S for Sandefjord, and the University in Bergen (UiB) for Bergen. For Ås and Bergen, the data available from `frost.met.no` is equal to the data collected but less complete. Thus the data retrieved from `frost.met.no` has been replaced by the collected data. For Finse and Sandefjord, the additional data has been added, resulting in a base dataset of 106 stations. An overview of the station locations is shown in Fig. 1 and a list is provided in the Appendix (Table A1). The complete dataset for the period 2016–2020 has been converted into a NetCDF file and is available at `https://zenodo.org/records/8082726`.

The time stamps in the collected hourly data are at full hours and refer to the end of the measuring period, i.e. the values represent averages over the previous hour. This is also true for the GHI data available at `frost.met.no`. Contrary, data currently available at the download portal from the Norwegian Institute of Bioeconomy Research (NIBIO) (`https://lmt.nibio.no/agrometbase/getweatherdata_new.php`) represent averages over the hour following the measurement time stamp, i.e. they are shifted by one hour.

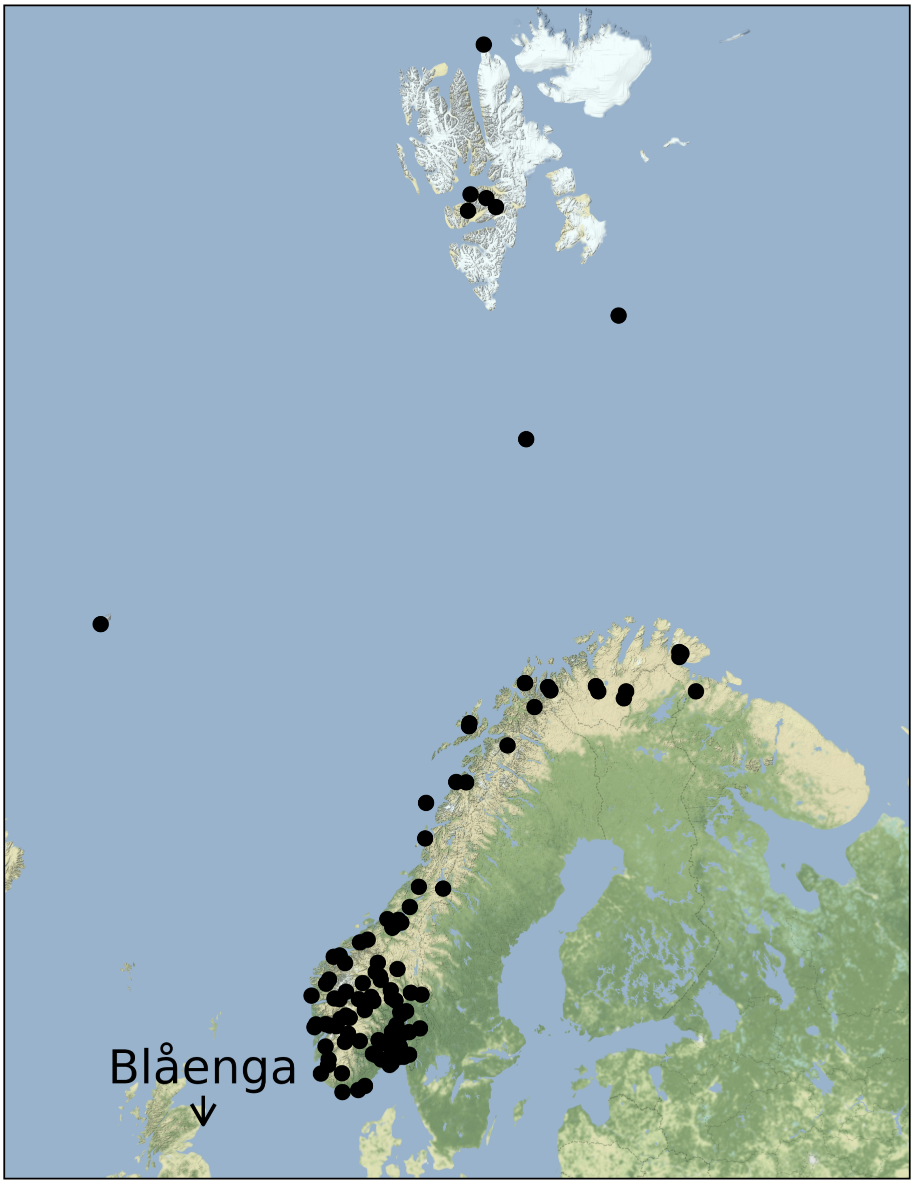


Figure 1: Map of the 106 Norwegian stations providing hourly GHI data. Black dots indicate the station locations, where Blåenga refers to the measurement station in Antarctica.

2.2 Clear-sky data

Modeled data for hourly GHI under clear-sky conditions has been downloaded via the Copernicus Atmosphere Monitoring Service (CAMS) McClear Clear-Sky Irradiation service. At the time of writing, the service is hosted at `https://www.soda-pro.com/web-services/radiation/cams-mcclear`. It provides clear-sky data interpolated to any specific location on the globe, based on aerosol, ozone and water vapour from the CAMS global forecasting system and incorporating properties like surface albedo and elevation (*Lefèvre et al.*, 2013; *Gschwind et al.*, 2019). The use of the service is currently free of charge and unlimited but needs registration. After registration, the data can also be downloaded using the *iotools.get_cams* routine from the *pvlib* tool for python (*Jensen et al.*, 2023). The downloaded data is based on version 3.5 of the McClear clear-sky model which reduces the bias with respect to ground measurements compared to earlier versions (*Lefèvre*, 2023).

3 Station selection

3.1 Data coverage

In a first step, stations with low data coverage in the reference period 2016–2020 have been excluded. The thresholds we have applied are

- at least 80% overall coverage
- at least 60% coverage for all months of the year, i.e. data for at least 60% of all dates in January, February, etc. in the reference period must be available.

Figure 2 shows daily mean GHI from the 104 stations available in `frost.met.no`. The data goes back to 1982, but most of the data has become available during the last decade. For the reference time period (2016–2020), about 60-70 stations have good coverage, while the rest of the stations shows only a limited or no coverage at all (Fig. 3).

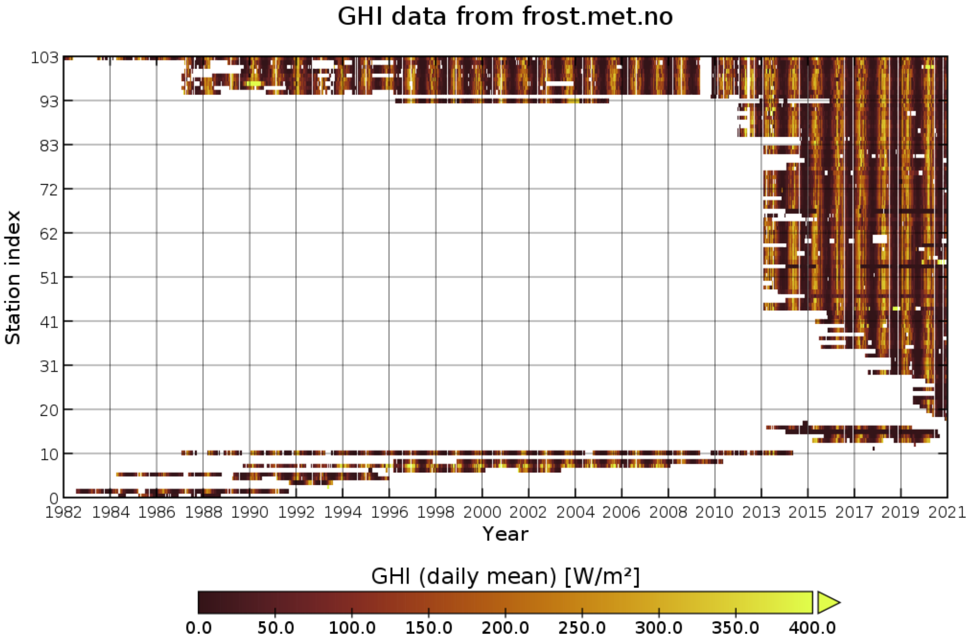


Figure 2: Daily means of GHI data retrieved from frost.met.no. White areas represent missing data. The station index refers to the order of appearance in the frost.met.no database.

The requirement of at least 80% overall coverage reduces the dataset from 106 to 72 stations with enough data. The criterion of at least 60% coverage for any month of the year removes three additional stations, resulting in 69 stations. The complete list of the

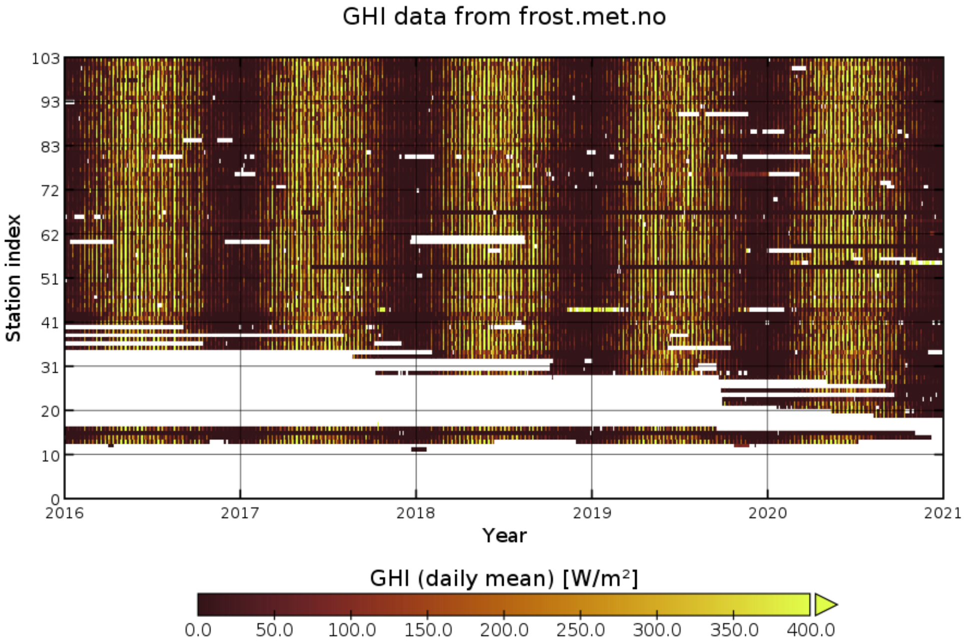


Figure 3: GHI retrieved from frost.met.no for the time period 2016–2020. White areas represent missing data. The station index refers to the order of appearance in the frost.met.no database.

106 stations in the Appendix (Table A1) includes a remark on which stations that were removed due to low overall or monthly coverage.

3.2 Data quality visualization

After the removal of stations with too low data coverage, the remaining stations have been evaluated more generally to remove stations with a low data quality. This has been done with the use of various visualization methods. On the one hand, a visualization of the data in various ways is crucial in identifying suspicious values in the station data. On the other hand, visually controlling a large set of stations quickly becomes time consuming and should not be too extensive. In this study we are inspecting about 70–100 stations, making the visualization extensive but doable and valuable.

To evaluate whether there are significant shortcoming in the station data, the following visualization techniques have been used:

- time-series of measured GHI
- diurnal and annual cycles of measured and clear-sky GHI (heat-maps)

- ratios of measured to clear-sky GHI (scatter-plots)
- frequencies of measured to clear-sky GHI ratios (histograms)

Examples of the visualizations and some typical evaluations are provided in the sections below. A complete set of the visualization plots for all stations is available at `https://zenodo.org/records/8082726`. A more extensive visualization tool is presented in *Blanc et al.* (2022) but only the four techniques listed above have been implemented here.

3.2.1 Time-series

Simple time-series of GHI can already reveal some data issues. In Fig. 4, a clearly negative trend in the measured GHI values at the Sjufellet station is visible.

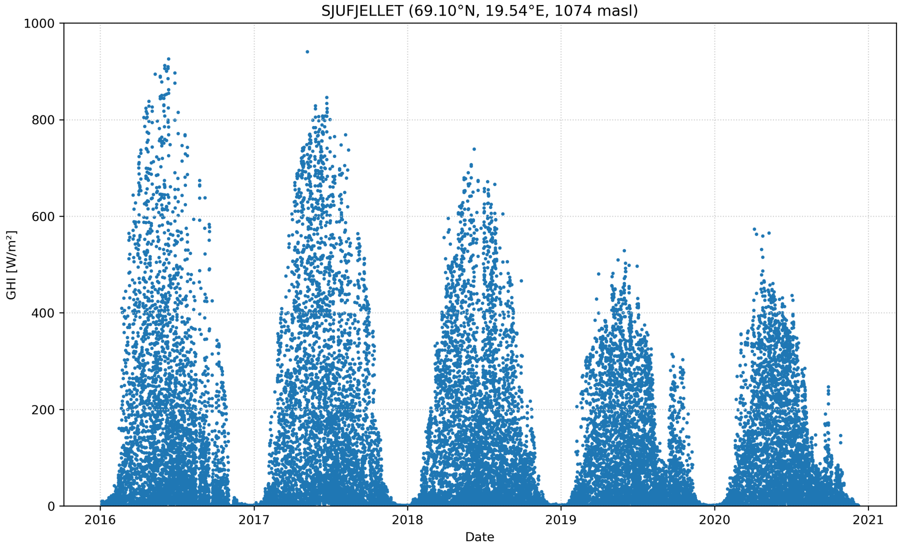


Figure 4: Hourly GHI values from the Sjufjellet station.

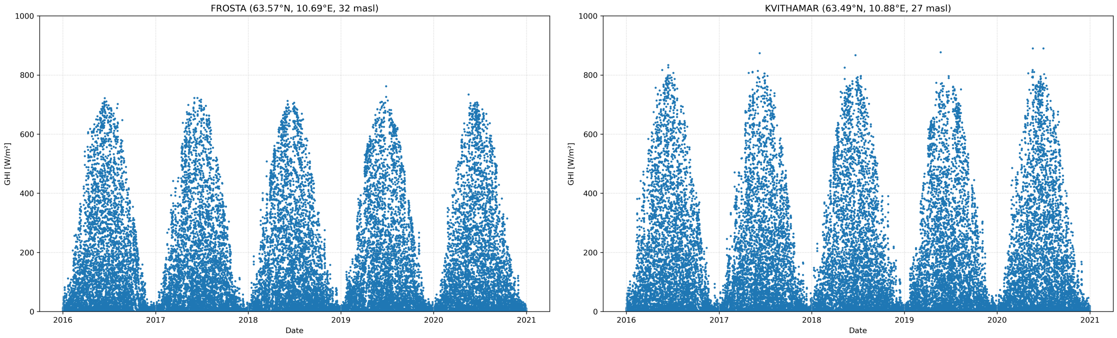


Figure 5: As Fig. 4 but for the two stations Frosta (left) and Kvithamar (right).

Another application of time-series is to compare close-by stations: Fig. 5 shows that the maximum GHI values measured at the Frosta station are too low when compared with the data from the closest station at Kvithamar (located only 22.4 km away).

3.2.2 Heat-maps

Heat-maps visualize the diurnal and the annual cycle of measured GHI at a station. They are often combined with a corresponding plot for clear-sky irradiance. Examples for Roverud, Skjetlein and Oslo are shown in Figs. 6 – 8. For the Roverud station (Fig. 6) a sharp horizontal gradient in measured GHI at about 08:00 UTC is visible throughout the whole period, indicating lacking radiation in the morning. Further investigations have shown that the station is affected by shadowing effects of nearby trees to the southeast. Roverud has therefore been excluded from the final, quality assured data set. Note that meanwhile the trees have been removed and the Roverud data from 2022 onwards is no longer affected by the tree shading.

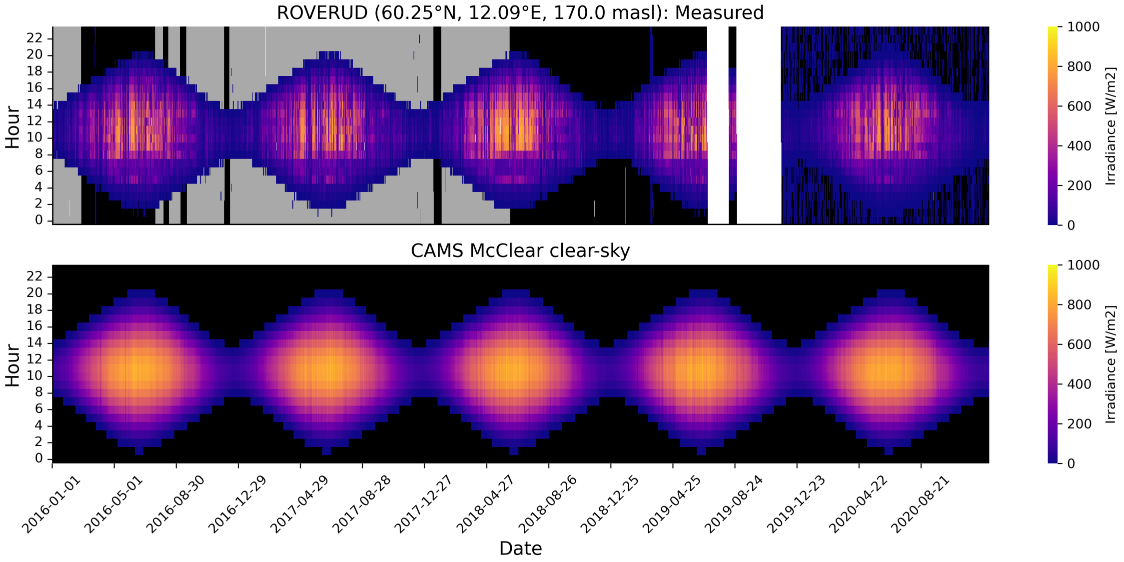


Figure 6: Annual (x-axis) and diurnal (y-axis) cycles of measured GHI (top) and estimated clear-sky irradiance (bottom) at the Roverud station for the period 2016–2020. Values equal to zero are coloured black, missing data are white and negative values shown in grey.

For the Skjetlein station (Fig. 7), the heat-map does not reveal any shadowing issues in the station data. However, there are small positive (dark blue) and negative (grey) values visible during night time in the station data, especially at the beginning of the period. Although this may indicate some issues with the station data, the overall impact on the solar resource potential is assumed to be small and the night time values can easily be filtered out, e.g. by using a night mask derived from clear-sky radiation. Thus, the station has not been removed from the final station selection.

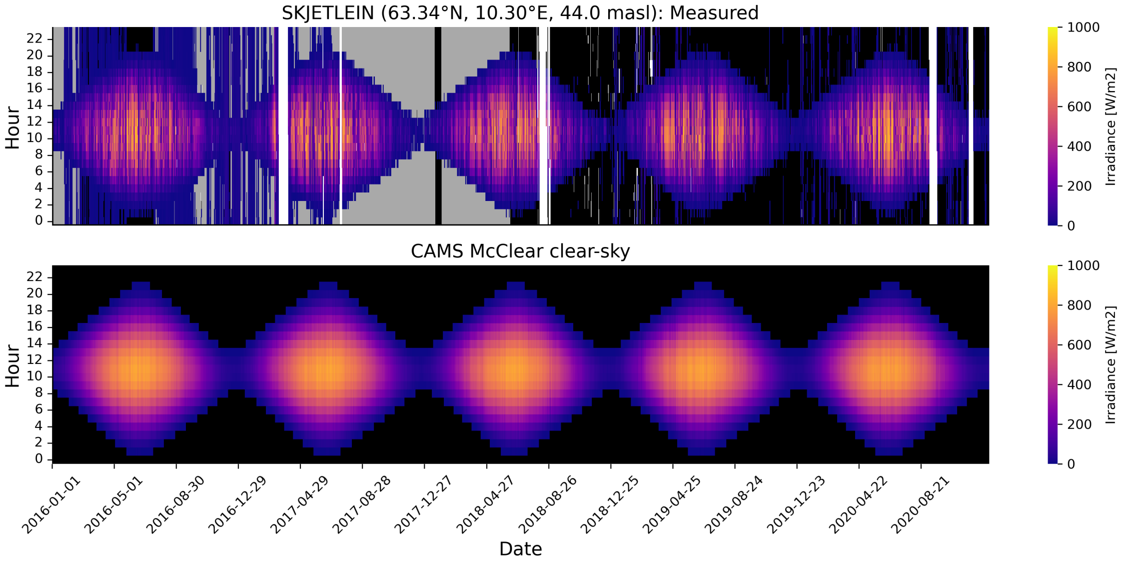


Figure 7: As Fig. 6 but for the station at Skjetlein.

Small negative values appear also in the station data form Oslo-Blindern (Fig. 8). Otherwise, the heat-map reveals a reasonable performance throughout the period, with a phase of maximum observed GHI during summer 2018.

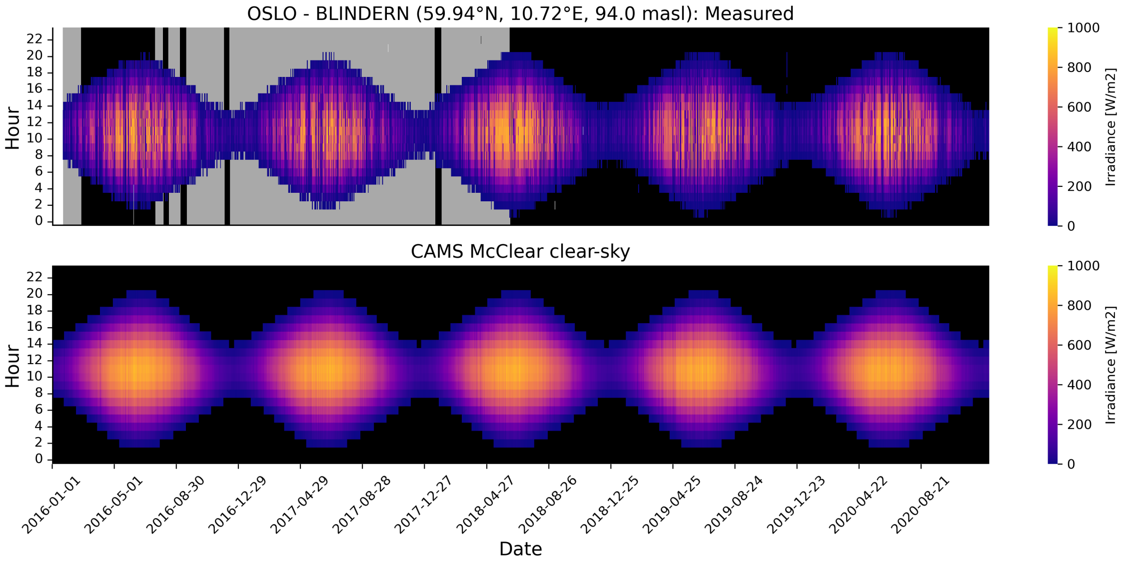


Figure 8: As Fig. 6 but for the station at Oslo-Blindern.

3.2.3 Scatter-plots and Histograms

Fig 9 shows measured against clear-sky irradiance for two stations in Bergen: Bergen-Florida located at the university and Flesland located at the airport. For Flesland, there is a large number of measurements exceeding the clear-sky values with a factor of 1.3 or more (yellow points). Comparing it to the nearby station at Bergen-Florida, where the exceeding of a factor of 1.3 is less common, the irradiance data at Flesland seem doubtful.

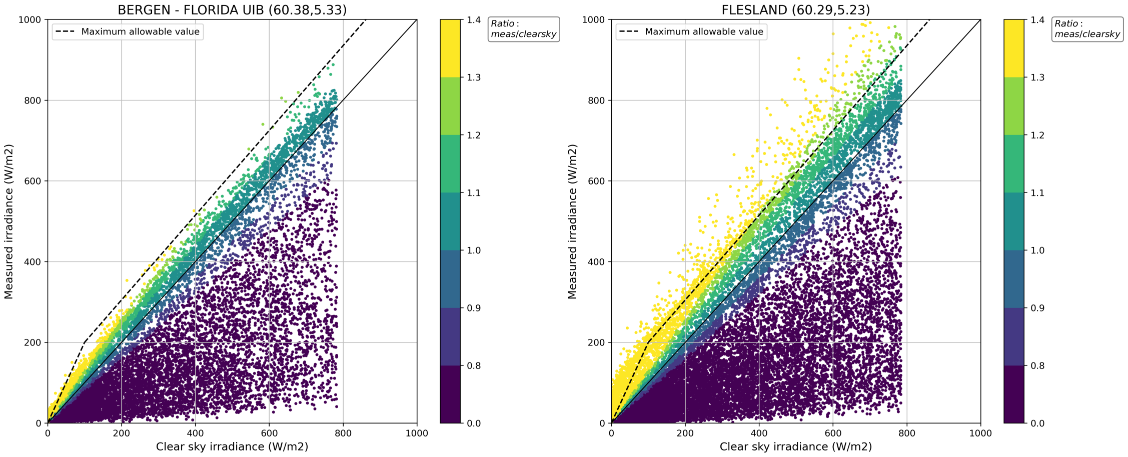


Figure 9: Measured GHI (y-axis) vs. estimated clear-sky irradiance (x-axis) for the stations at Bergen-Florida (left) and Flesland (right). The colouring shows the ratio of measured to clear-sky irradiance. A 1:1 line is shown in solid black, while the dashed line gives a maximum allowable ratio to the clear-sky estimate (see section 4 for details).

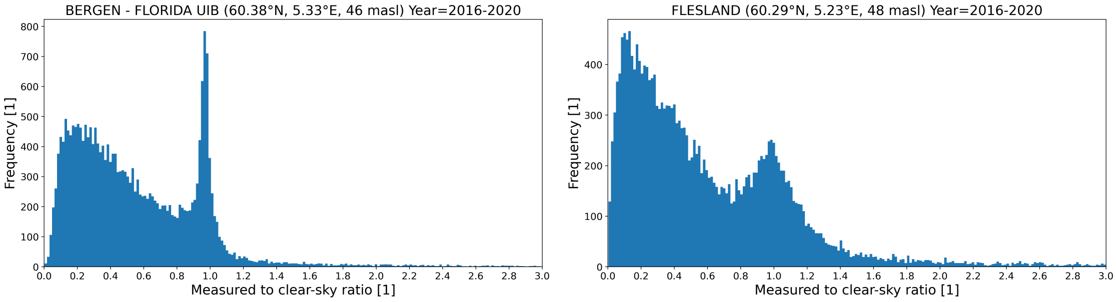


Figure 10: Frequencies of measured to clear-sky irradiance ratios at the Bergen-Florida (left) and Flesland (right) stations.

This is supported by the histogram of measured to clear-sky GHI ratios at the two stations (Fig. 10). While the Bergen-Florida data shows a clear peak and drop around one (clear-sky days), the peak is much smaller at the Flesland station and the values decrease more slowly with increasing ratios. Additionally, we have also found that the diurnal cycle on clear-sky days in the Flesland measurements is skewed (not shown), indicating a possible tilt in the measuring device. Thus, taking all theses findings and the fact that there is a well performing station close by into account, the Flesland station has been removed from the selection.

3.3 Selected stations

In total, 22 of the 69 stations with enough temporal coverage have been removed due to quality issues, resulting in the 47 remaining stations shown in Table 1. The 22 stations removed by visualization checks are listed in Table 2 including the reason for removal. A map of the Norwegian mainland showing the locations of the final and the rejected stations is provided in Fig. 11. While the southern part of Norway shows a good coverage for various conditions, the station density in the middle and north of Norway is rather low.

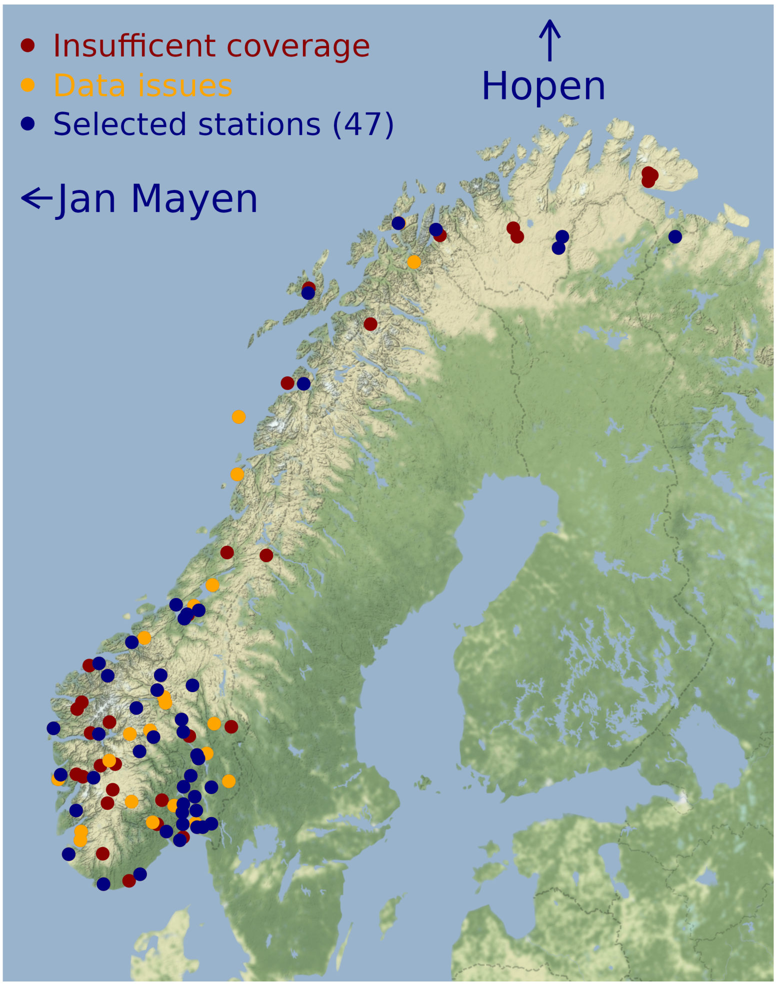


Figure 11: Map of the Norwegian mainland with the selected and rejected stations. Blue dots and arrows (Jan Mayen and Hopen) indicate the locations of selected stations, while yellow dots show stations with data issues and red dots mark stations with too low temporal coverage.

Table 1: Overview of the selected stations sorted from south to north. The column *GHI raw* gives the mean annual GHI in the raw station data (after removing negative values). *GHI cleaned* denotes the corresponding values after the data flagged by the automated quality control has been removed (see section 4).

Station name	**Latitude** [°N]	**Longitude** [°E]	**Altitude** [masl]	**GHI raw** $[kWh/m^2]$	**GHI cleaned** $[kWh/m^2]$
LYNGDAL	58.13	7.05	6	972	970
LANDVIK	58.34	8.52	6	956	953
SÆRHEIM	58.76	5.65	87	932	931
TJØLLING	59.05	10.12	19	1026	1023
GJERPEN - ÅRHUS	59.23	9.58	41	920	888
RÅDE - TOMB	59.32	10.81	12	1001	993
ØSAKER	59.32	11.04	45	993	985
RAMNES - KILE VESTRE	59.38	10.24	39	976	966
RAKKESTAD	59.39	11.39	100	1037	1021
SANDE - GALLEBERG	59.62	10.22	60	903	902
ÅS	59.66	10.78	92	1005	1002
ETNE II	59.66	5.95	8	971	881
LIER	59.79	10.26	39	947	944
OSLO - BLINDERN	59.94	10.72	94	953	952
ÅRNES	60.13	11.39	160	960	956
HØNEFOSS - HVERVEN	60.14	10.27	126	959	956
ULLENSVANG FORSØKSGARD	60.32	6.65	12	785	784
GRAN	60.36	10.56	245	912	908
BERGEN - FLORIDA UIB	60.38	5.33	46	790	789
ØSTRE TOTEN - APELSVOLL	60.70	10.87	264	915	913
KISE PA HEDMARK	60.77	10.81	128	948	942
HEMSEDAL SKISENTER	60.84	8.50	1344	799	794
LØKEN I VOLBU	61.12	9.06	521	947	938
NJØS	61.18	6.86	45	814	811
GAUSDAL - FOLLEBU	61.22	10.26	375	965	946
FURENESET	61.29	5.04	7	823	821
FÅVANG	61.46	10.19	200	960	940
JUVVASSHØE	61.68	8.37	1894	1115	1093
DOVRE-LANNEM	62.02	9.21	560	831	826
ALVDAL	62.11	10.63	478	872	864
LINGE	62.29	7.22	34	856	846
SNØHEIM	62.30	9.35	1475	1010	995
LEBERGSFJELLET	62.52	6.87	625	830	823
TINGVOLL	62.91	8.19	23	815	812
SKJETLEIN	63.34	10.30	44	838	816
TRONDHEIM - GLØSHAUGEN	63.42	10.41	60	838	833
KVITHAMAR	63.49	10.88	27	842	833
RISSA III	63.59	9.97	23	817	815
VALNESFJORD	67.28	15.10	20	727	723
SORTLAND - KLEIVA	68.65	15.28	14	712	702
ISKORAS II	69.30	25.35	591	805	783
PASVIK - SVANVIK	69.46	25.50	131	693	689
KARASJOK - MARKANNJARGA	69.46	30.04	27	705	704
NORDNESFJELLET	69.56	20.42	697	727	720
TROMSØ - HOLT	69.65	18.91	20	698	692
JAN MAYEN	70.94	-8.67	10	566	563
HOPEN	76.51	25.01	6	544	543
Mean	62.55	10.99	231	872	863

Table 2: Overview of stations with enough time coverage but with quality issues in the GHI data. These stations have been removed from the selection. The last column lists the reason for removal.

Station name	Reason for removal
LIARVATN	Wrong values
HJELMELAND	Too low values, esp. towards the end
RYGGE - HUGGENES	Shaded to the east
BØ	Poor correspondence with clear-sky values
HOKKSUND	Too low values, except outliers
MØSSTRAND II	Too low values, wrong values in 2020
ROVERUD	Shaded by trees south east
FLESLAND	Missing peak in the frequency of measured/clear-sky ratio around 1 on clear days. Too high maxima
MIDTSTOVA	Max values are too low (e.g. compared to Finse)
ILSENG	Many suspect values, wrong values in 2020
FILEFJELL - KYRKJESTØLA	Wrong values after 2017
BEITOSTØLEN II	Many suspect values, wrong values after 2017-05
RENA - ØRNHAUGEN	No clear obs/clear-sky peak, suspicious values, probably also tilted sensor
OTTA - SKANSEN	Wrong values
HØVRINGEN II	Too low values
SURNADAL - SYLTE	Shaded to the west
FROSTA	Too low values, esp. maxima
MÆRE III	Many suspect values around winter 2019
TJØTTA	Wrong values in 2020
MYKEN	Decreasing trend
SJUFJELLET	Decreasing trend
TROMSØ	Many suspect or missing values

4 Automated quality control

An automated quality control (QC) routine has been applied to all stations in the reference period 2016-2020. The QC incorporates all hourly GHI values and checks whether they are exceeding physically plausible maximum or minimum values and whether the temporal variability and distribution of the data is as expected. The routine is based on the tests described in *Grini* (2015) where more details and references can be found. However, while *Grini* (2015) is focusing on low elevated stations around the Oslofjord, our study covers the whole of Norway. This includes more stations and varying conditions, for instance higher altitudes and latitudes. Thus, some of the tests needed to be adjusted accordingly (see chapter 5). This resulted in a QC routine of eight tests, as listed in Table 3.

Table 3: Overview of quality control tests. A specific flag is assigned to a data point for each test it fails (last column). GHI is the hourly surface downwelling shortwave radiation, I_E the extra-terrestrial irradiance, θ the zenith angle and I_{CS} clear-sky radiation. μ, σ denote the daily mean and standard deviation, respectively.

Test	**Flagging criteria**	**Flag**
Offset	$GHI \leq -12\ W/m^2$ or $GHI > 6\ W/m^2$ if $\theta > 93\ °$	O
TOA	$GHI > I_E$	T
Clear Sky	$GHI > f * I_{CS} + a$, where $f = 2,\ a = 0$ if $I_{CS} \leq 100\ W/m^2$ $f = 1.05,\ a = 95$ if $I_{CS} > 100\ W/m^2$	CS
Low 2	$GHI < 10^{-4}\ (80 - \theta)\ I_E$ if $\theta \leq 80\ °$	2
Difference	$\frac{\lvert d(GHI/I_E)\rvert}{dt} \geq 0.75$ if $\theta < 80\ °$	D
Consistency	$\sigma(GHI/I_E) < 1/16 * \mu(GHI/I_E)$ or $\sigma(GHI/I_E) > 80$	C
Missing	GHI = NA	M
Negative	$GHI < 0$	N

Additionally to the station data, the test routine needs the following input: 1) extra-terrestrial irradiance, 2) zenith angles and 3) clear-sky radiation. 1) and 2) are calculated by the corresponding (default) methods of *pvlib*, 3) is taken from the McClear Clear-Sky Irradiation service (see section 2.2).

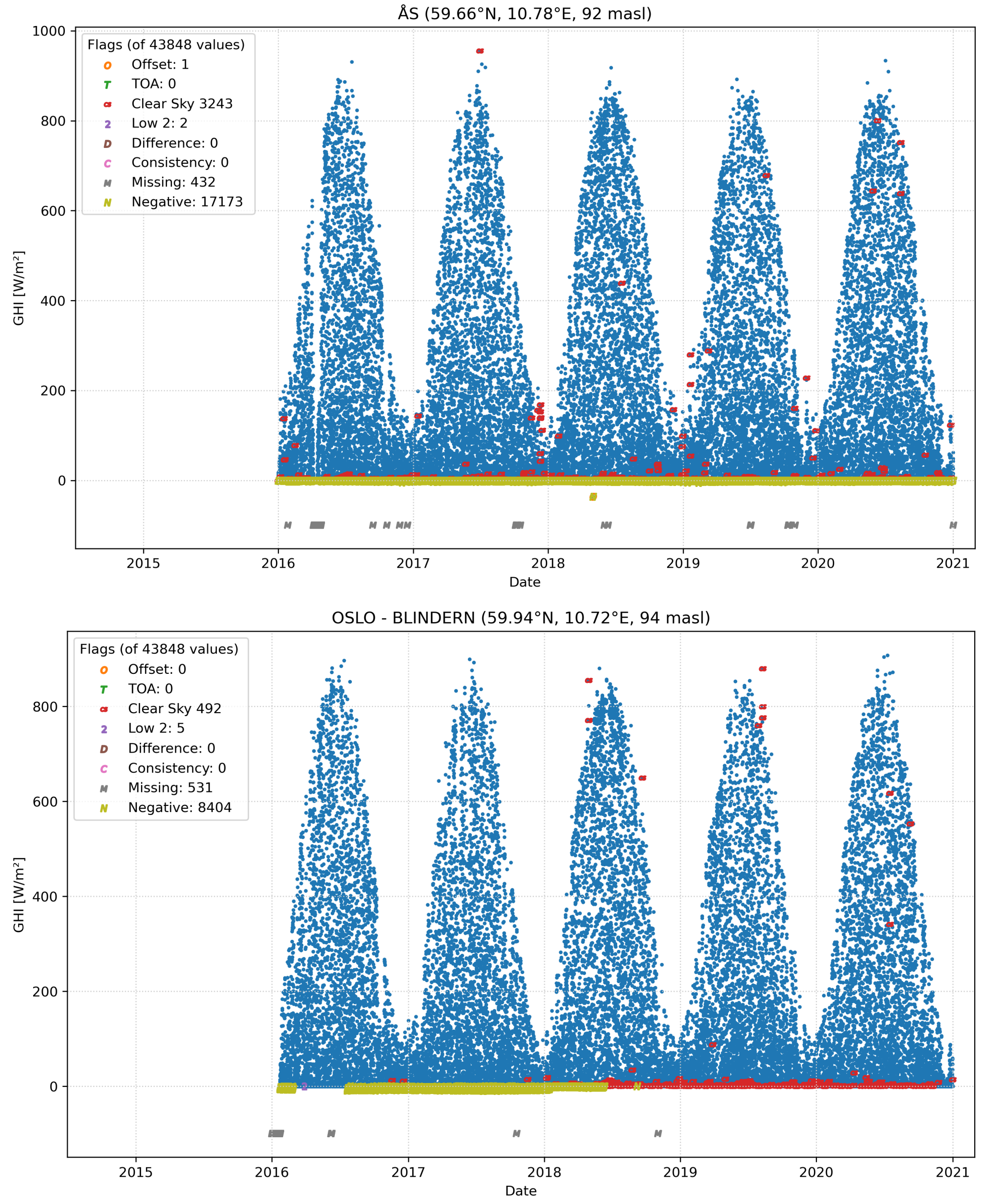


Figure 12: Hourly GHI values and assigned flags for the data from stations at Ås and Oslo - Blindern.

The automated QC routine puts flags on the data for each test the data does not pass. This is illustrated in Figures 12 and 13, showing the flags assigned to the station data at Ås, Oslo-Blindern, Juvvasshøe and Skjetlein. The Ås and Oslo-Blindern station show flags mostly related to small negative numbers and low values which are above the clear-sky limits (Fig. 12). The flags at Juvvasshøe and Skjetlein are more diverse with numerous clear-sky flags also assigned to high values.

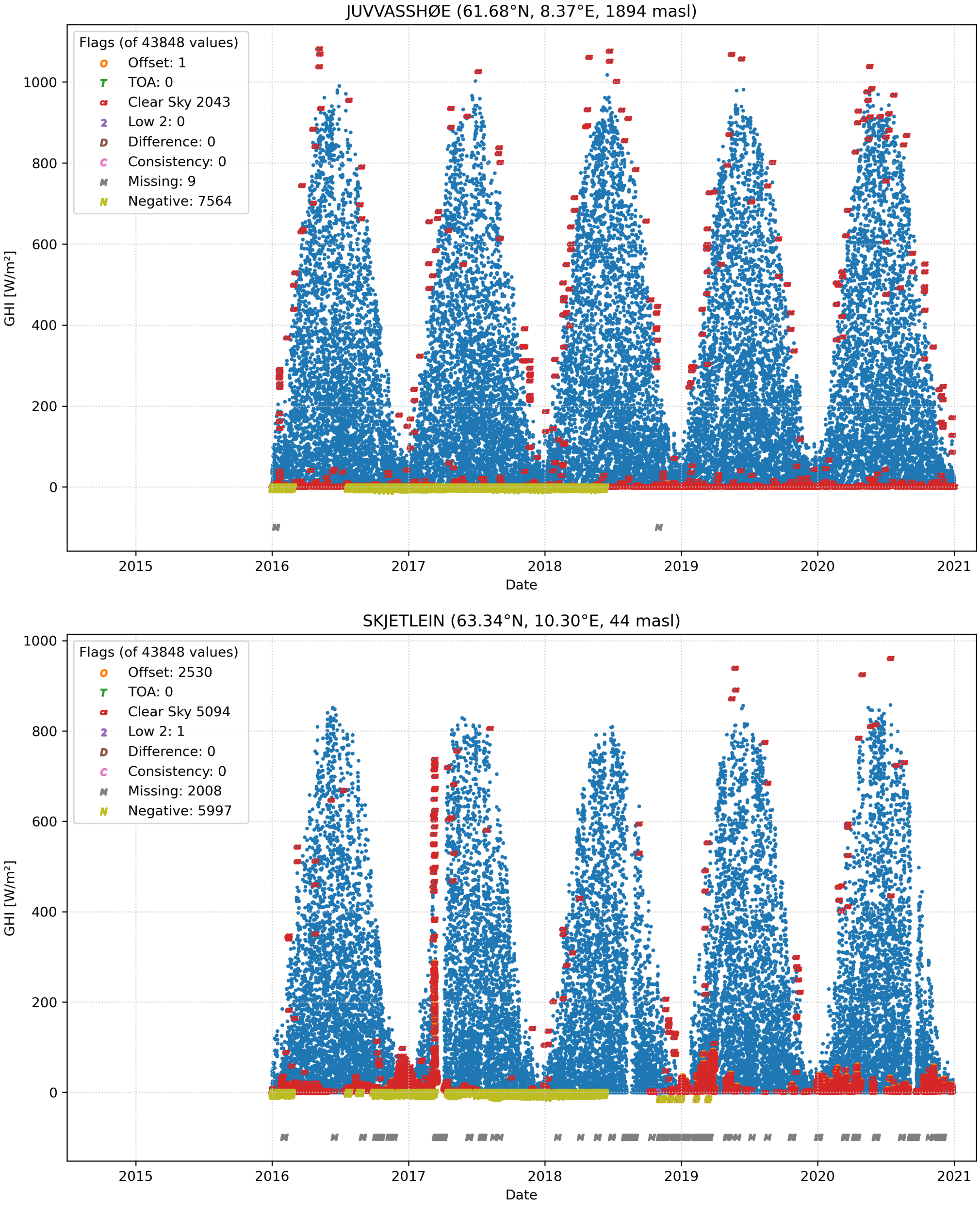


Figure 13: Hourly GHI values and assigned flags for the data from stations at Juvvasshøe and Skjetlein.

5 Discussion

Plots showing the flags assigned by the automated quality control (QC) procedure for all stations and a NetCDF file containing the flags for the selected 47 stations are available on `https://zenodo.org/records/8082726`. One of the main uncertainties in the QC lies in the references clear-sky data and the associated thresholds. For our application the clear-sky data needs to be provided at a wide range of locations, making a (semi-) automated retrieval necessary. The default *pvlib* package already provides clear-sky data at any location, e.g. using the Ineichen clear-sky method. However, *Stein et al.* (2012) have shown that complex clear-sky models accounting for atmospheric parameters are more accurate compared to simpler models, and their errors tend to be less variable in space and time. This is especially true for higher elevations, whereas at low elevations simpler models may work as well as more complex ones (*Stein et al.*, 2012). Thus, we concluded that for the application in this study, more complex models are better suited and the clear-sky data has been collected using the McClear Clear-Sky Irradiation Service.

Due to the usage of a different clear-sky model and the varying conditions at around 100 different stations throughout Norway, the threshold function for the measured to clear-sky GHI ratio (Clear Sky test) given in *Grini* (2015) has been adjusted. A subjective evaluation of the threshold exceedances, with focus on the selected 47 stations, resulted in the definition of a continuous function with different thresholds for low ($\leq 100\ W/m^2$) and high clear-sky values (see Table 3). Figure 14 provides scatter plots for the data from the Ås and Juvvasshøe station including the new threshold function. At Ås, only a few data points are exceeding the threshold, especially for values above $100\ W/m^2$. At Juvvasshøe, there are more data points with a higher ratio and over the maximum allowable value.

The McClear model provides larger clear-sky values than the Ineichen method for stations at higher altitudes. Using the McClear clear-sky data at Juvvasshøe reduces the number of flagged values from about 4500 (not shown) to 2043 (out of 43848, Fig. 13). The Juvvasshøe station is generally well maintained and its measurements reliable (personal communication), indicating a lower number of flagged values is more preferable. Results from similarly located stations (Finse-UiO, Snøheim) support that the new threshold function in combination with the McClear clear-sky data provides a sound basis for the Clear Sky test at high elevated locations. Using the Ineichen data results in too many flags assigned due to exceedance of the clear-sky limit.

In terms of the total irradiance potential, the effect of the adjusted threshold function and the new clear-sky model is small at low elevated stations such as Ås, in agreement with *Stein et al.* (2012). For higher elevated stations however, the clear-sky data provides a larger uncertainty in quantifying the actual irradiance potential. At Juvvasshøe for instance, removing all data flagged by the Ineichen model reduces the potential from 1115 kWh/m^2 to 1073 kWh/m^2 (i.e. -4%, not shown) while using the McClear model the potential is only reduced to 1093 kWh/m^2 (-2%, Table 1).

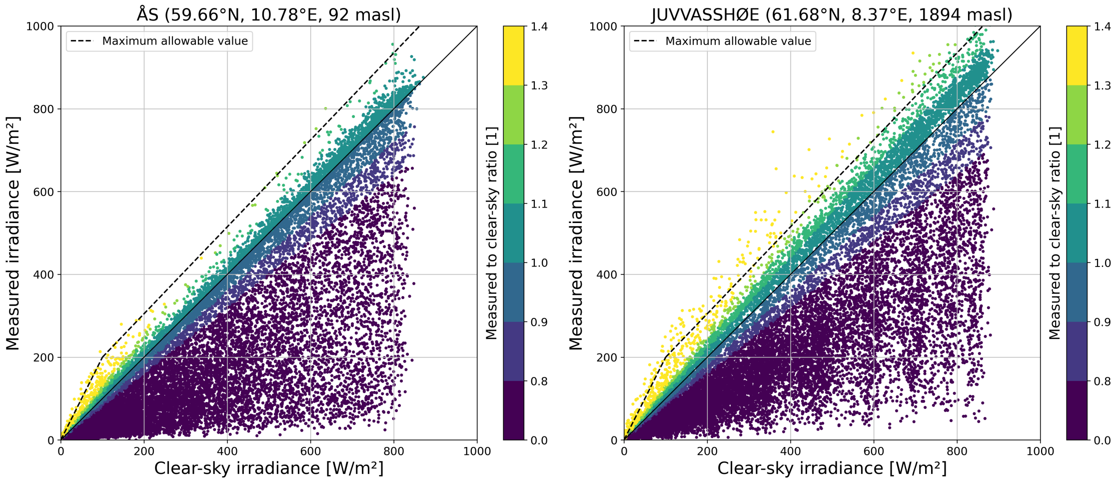


Figure 14: Measured GHI (y-axis) vs. estimated clear-sky irradiance (x-axis) for the stations at Ås (left) and Juvvasshøe (right). The colouring denotes the measured to clear-sky GHI ratio. A 1:1 line is shown in solid black. The dashed line shows the maximum allowable ratio to the clear-sky estimate, i.e. the (new) threshold function for the Clear Sky test.

To test for too low values in the data, *Grini* (2015) has applied two lower limit tests based on extraterrestrial irradiance. The first test checks whether the daily mean of the ratios between hourly measured GHI and extraterrestrial irradiance is below 3%, i.e. whether less then 3% of the extraterrestrial irradiance reaches the surface averaged over a day. However, we found that for low solar elevation conditions, typical to Norway from late autumn to early spring, this test fails too often. Although the impact on the total irradiance potential by rejecting this data would be small, we decided to not apply the test. We use the second lower limit test only which is less restrictive for conditions under low solar elevation (Low 2 in table 3).

Consistency and offset flags can indicate general problems with the measurement devices. The QC procedure found no consistency issues in the selected stations. Except for the Etne II station, only a few data points were flagged for offset. Generally, these offsets are of small absolute magnitude ($< 10\ W/m^2$) and have only a minor effect on the total irradiance potential. For the Etne II station, a large amount of small positive values during nighttime has been flagged. This offset is only apparent during nighttime and the remaining values seem not to be affected. Note that the QC method flags negative and offset values separately, i.e. values below $-12\ W/m^2$ are flagged as offset and negative.

For the estimation of the total irradiance potential (Table 1), all the values flagged by the QC procedure have been removed (i.e. set to NaN). The removal reduces the number of available hourly GHI values by 28% in total, ranging from 15% to 52% for the single stations. However, the majority of these values are small deviations from zero at nighttime, which are either flagged by the test for negative values or clear-sky exceedance at night. Thus, the overall impact of removing the flagged data on the total irradiance potential is below 2% at most of the stations (Table 1) and flagging did not lead to an exclusion of additional stations. For the Etne II station however, removal of numerous small positive values during nighttime, reduces the potential by 10% and this estimation should be considered with some caution.

As the strict removal of the flagged data may not be appropriate for other applications of the data, the unfiltered data and the flags are available together with the filtered data at `https://zenodo.org/records/8082726`.

6 Summary

We have collected surface measurements of hourly global horizontal irradiance (GHI) data for 106 locations in Norway for the time period 2016-2020. The data has been retrieved from the Norwegian meteorological institute's open database `frost.met.no` and extended by additional data from four locations collected within the SunPoint project. To provide a reliable basis for mapping the solar resource potential, stations with insufficient data coverage have been excluded as they can introduce biases and inaccuracies in the final results. We have set a requirement of at least 80% overall coverage and 60% coverage for any month of the year, which has reduced the number of stations to 69.

A quality control (QC) procedure based on various visualization techniques has been implemented to identify significant shortcomings in the station data. The station data has been inspected using time-series analysis, visualization of diurnal and annual cycles (heatmaps), and comparison to clear-sky GHI data to detect anomalies, spurious trends, and inconsistencies that could have compromised the data integrity. Based on this, another 22 stations have been rejected due to issues such as shadowing effects or data inconsistencies, resulting in a final number of 47 selected stations. Additionally, the data from all stations has been subjected to an automated flagging process testing for the exceedance of theoretical limits, temporal variability, and data distributions.

Finally, the data from the 47 selected stations has been collected in a NetCDF file to make it freely available and easily accessible. This quality controlled GHI data is provided at `zenodo.org/records/8082726`. Files including the flags from the flagging routine for the 47 stations and the cleaned data, i.e. a dataset where all flagged data has been removed, are available there as well, together with this report, QC and visualization scripts and the output from the visualization tests for all 106 stations. In total, 28% of the hourly data has been flagged, where the majority are small deviations from zero at nighttime. The removal of all flagged data at the 47 selected stations resulted in a reduction of the average solar potential from 872 kWh/m^2 to 863 kWh/m^2 (-1%).

We found that a main uncertainty in the QC procedure lies in the clear-sky values estimated for high altitudes and the associated limits that should not be exceeded by the measurements. We argue that using clear-sky data from the CAMS McClear Clear-Sky Irradiation service together with a relative simple threshold function provides a flexible and sound basis also for high elevated locations.

References

Blanc, P., R. Jolivet, L. Ménard, and Y.-M. Saint-Drenan (2022), Data sharing of in-situ measurements following GEO and FAIR principles in the solar energy sector, doi: 10.23646/AC2M-8504, working paper.

Grini, S. (2015), Quality control of global solar irradiation measured at four stations in eastern Norway, Master's thesis, Norwegian University of Life Science, Ås, `http://hdl.handle.net/11250/292854`.

Gschwind, B., L. Wald, P. Blanc, M. Lefèvre, M. Schroedter-Homscheidt, and A. Arola (2019), Improving the McClear model estimating the downwelling solar radiation at ground level in cloud-free conditions - McClear-v3, *Meteorologische Zeitschrift*, *28*(2), 147163, doi:10.1127/metz/2019/0946.

Jensen, A. R., K. S. Anderson, W. F. Holmgren, M. A. Mikofski, C. W. Hansen, L. J. Boeman, and R. Loonen (2023), pvlib iotools – Open-source Python functions for seamless access to solar irradiance data, *Solar Energy*, *266*, 112,092, doi: 10.1016/j.solener.2023.112092.

Lefèvre, M. (2023), Solar radiation evaluation and quality assurance report #40, September – November 2022, *Tech. rep.*, MINES Paris, PSL university, `https://atmosphere.copernicus.eu/sites/default/files/custom-uploads/EQC-solar/CAMS2_73_2021SC2_D1.3.1-2023Q2_RAD_validation_report_SON2022_v1.pdf`.

Lefèvre, M., A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola, A. Benedetti, J. W. Kaiser, and J.-J. Morcrette (2013), McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions, *Atmospheric Measurement Techniques*, *6*(9), 24032418, doi:10.5194/amt-6-2403-2013.

Stein, J. S., C. W. Hansen, and M. J. Reno (2012), Global horizontal irradiance clear sky models : implementation and analysis, doi:10.2172/1039404.

Appendix: Station list

Table A1: List of all 106 stations considered and their location. Stations included in the final selection are listed in bold. In case the station has not been included the reason is given in the last column (C: Not enough overall data coverage; CM: Not enough monthly coverage; Q: Quality issues). Quality issues are detailed in Table 2.

Station	Latitude [°N]	Longitude [°E]	Altitude [masl]	Exclusion
SOLHOM I KVINESDAL	58.77	7.02	650	C
MAZE - RUOGONJARGA	69.46	23.69	277	C
BIRKEBEINEREN SKISTADION	61.14	10.5	484	C
BLÅENGA	-77.51	-34.21	375	C
GJENGEDAL - DALHEIM	61.66	5.99	355	C
GAUPNE	61.41	7.29	6	C
KJEVIK	58.2	8.08	12	C
ÅS	59.66	10.78	92	
TRONDHEIM - VOLL	63.41	10.45	127	C
KVAMSKOGEN - JONSHØGDI	60.39	5.96	455	C
BODØ - VÅGØNES	67.29	14.45	33	C
SUOLOVUOPMI - LULIT	69.58	23.53	381	C
FLESBERG	59.87	9.4	183	C
HAUKELISETER TESTFELT	59.81	7.21	990	C
SJUFJELLET	69.1	19.54	1074	Q
LIARVATN	59.05	6.12	300	Q
KVAM - AKSNESET	60.34	6.22	13	C
ÅNSTADBLÅHEIA	68.72	15.31	500	C
HANSBU	60.08	7.42	1160	C
BALESTRAND - BALE	61.2	6.53	15	C
SANDANE	61.79	6.18	51	C
BRUSDALEN	62.48	6.48	69	C
NORDLI - SANDVIKA	64.46	13.6	420	C
REINDALSPASSET	78.06	17.04	181	C

Continued on next page

Table A1: List of all 106 stations considered and their location. Stations included in the final selection are listed in bold. In case the station has not been included the reason is given in the last column (C: Not enough overall data coverage; CM: Not enough monthly coverage; Q: Quality issues). Quality issues are detailed in Table 2. (Continued)

Station	**Latitude [°N]**	**Longitude [°E]**	**Altitude [masl]**	**Exclusion**
ISTJØRNDALEN	78.01	15.21	188	C
JANSSONHAUGEN VEST	78.18	16.41	250	C
VERLEGENHUKEN	80.06	16.24	8	C
ULVIK - HJELTNES	60.56	6.93	42	C
TRYSIL - NORDRE KANKEN	61.32	12.19	1020	C
GAMANJUNNI	69.48	20.58	1237	C
BERGEBYDALEN	70.23	28.96	150	C
TORVHAUGDALEN	70.31	29.1	268	C
REINHAUGEN	70.34	28.96	470	C
BJØRNØYA	74.5	19	16	C
PLATÅBERGET III	78.23	15.38	450	C
DOVRE-LANNEM	62.02	9.21	560	
HEMSEDAL SKISENTER	60.84	8.5	1344	
FLESLAND	60.29	5.23	48	Q
BERGEN - FLORIDA UIB	60.38	5.33	46	
TRONDHEIM - GLØSHAUGEN	63.42	10.41	60	
TROMSØ	69.65	18.94	100	Q
KARASJOK - MARKANNJARGA	69.46	25.5	131	
HOPEN	76.51	25.01	6	
JAN MAYEN	70.94	-8.67	10	
ILSENG	60.8	11.2	182	Q
GAUSDAL - FOLLEBU	61.22	10.26	375	
JUVVASSHØE	61.68	8.37	1894	
OTTA - SKANSEN	61.78	9.54	309	Q
HØVRINGEN II	61.89	9.48	940	Q
RÅDE - TOMB	59.32	10.81	12	
RYGGE - HUGGENES	59.4	10.75	35	Q
LIER	59.79	10.26	39	

Continued on next page

Table A1: List of all 106 stations considered and their location. Stations included in the final selection are listed in bold. In case the station has not been included the reason is given in the last column (C: Not enough overall data coverage; CM: Not enough monthly coverage; Q: Quality issues). Quality issues are detailed in Table 2. (Continued)

Station	**Latitude [°N]**	**Longitude [°E]**	**Altitude [masl]**	**Exclusion**
HØNEFOSS - HVERVEN	60.14	10.27	126	
GRAN	60.36	10.56	245	
BEITOSTØLEN II	61.25	8.92	965	Q
HOKKSUND	59.76	9.89	15	Q
RAMNES - KILE VESTRE	59.38	10.24	39	
TJØLLING	59.05	10.12	19	
GJERPEN - ÅRHUS	59.23	9.58	41	
MØSSTRAND II	59.84	8.18	977	Q
GVARV - NES BIOFORSK	59.38	9.21	94	C
BØ	59.42	9.03	105	Q
ØSAKER	59.32	11.04	45	
LYNGDAL	58.13	7.05	6	
HJELMELAND	59.23	6.15	43	Q
ETNE II	59.66	5.95	8	
MIDTSTOVA	60.66	7.28	1162	Q
FILEFJELL - KYRKJESTØLA	61.18	8.11	956	Q
NJØS	61.18	6.86	45	
LINGE	62.29	7.22	34	
LEBERGSFJELLET	62.52	6.87	625	
TINGVOLL	62.91	8.19	23	
SURNADAL - SYLTE	62.99	8.69	5	Q
SKJETLEIN	63.34	10.3	44	
MÆRE III	63.94	11.43	59	Q
RISSA III	63.59	9.97	23	
OVERHALLA - SKOGMO	64.51	12.02	32	CM
RENA - ØRNHAUGEN	61.38	11.5	872	Q
MYKEN	66.76	12.49	17	Q
VALNESFJORD	67.28	15.1	20	

Continued on next page

Table A1: List of all 106 stations considered and their location. Stations included in the final selection are listed in bold. In case the station has not been included the reason is given in the last column (C: Not enough overall data coverage; CM: Not enough monthly coverage; Q: Quality issues). Quality issues are detailed in Table 2. (Continued)

Station	Latitude [°N]	Longitude [°E]	Altitude [masl]	Exclusion
LOSISTUA	68.19	17.79	740	CM
SORTLAND - KLEIVA	68.65	15.28	14	
ALVDAL	62.11	10.63	478	
SNØHEIM	62.3	9.35	1475	
ISKORAS II	69.3	25.35	591	
FÅVANG	61.46	10.19	200	
SANDE - GALLEBERG	59.62	10.22	60	
RAKKESTAD	59.39	11.39	100	
ÅRNES	60.13	11.39	160	
ULLENSVANG FORSØKSGARD	60.32	6.65	12	
ROVERUD	60.25	12.09	170	Q
FROSTA	63.57	10.69	32	Q
NORDNESFJELLET	69.56	20.42	697	
OSLO - BLINDERN	59.94	10.72	94	
PASVIK - SVANVIK	69.46	30.04	27	
KVITHAMAR	63.49	10.88	27	
ØSTRE TOTEN - APELSVOLL	60.7	10.87	264	
LØKEN I VOLBU	61.12	9.06	521	
LANDVIK	58.34	8.52	6	
SÆRHEIM	58.76	5.65	87	
FURENESET	61.29	5.04	7	
TJØTTA	65.83	12.43	21	Q
TROMSØ - HOLT	69.65	18.91	20	
KISE PA HEDMARK	60.77	10.81	128	
FINSE - UIO	60.59	7.52	1206	C
SANDEFJORD - JOTUN	59.11	10.25	18	CM