Effects of clouds and aerosols on downwelling surface solar irradiance nowcasting and sort-term forecasting

Kyriakoula Papachristopoulou[1,2], Ilias Fountoulakis[3,2], Alkiviadis F. Bais[4], Basil E. Psiloglou[5], Nikolaos Papadimitriou[6], Ioannis-Panagiotis Raptis[1], Andreas Kazantzidis[6], Charalampos Kontoes[2], Maria Hatzaki[1], Stelios Kazadzis[7]

[1]Laboratory of Climatology and Atmospheric Environment, Section of Geography and Climatology, Department of Geology and Geoenvironment, National and Kapodistrian University of Athens, Athens, GR-15784, Greece
[2]Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens (IAASARS/NOA), Athens, GR-15236, Greece
[3]Research Centre for Atmospheric Physics and Climatology, Academy of Athens, Athens, Greece
[4]Laboratory of Atmospheric Physics, Aristotle University of Thessaloniki, Thessaloniki, Greece
[5]Institute for Environmental Research and Sustainable Development, National Observatory of Athens (IERSD/NOA), Athens, GR-15236, Greece
[6]Laboratory of Atmospheric Physics, Department of Physics, University of Patras, GR 26500, Patras, Greece
[7]Physikalisch-Meteorologisches Observatorium Davos, World Radiation Center (PMOD/WRC), Davos 7260, Switzerland

Correspondence to: Kyriakoula Papachristopoulou (kpapachr@noa.gr)

Abstract. Solar irradiance nowcasting and short-term forecasting are important tools for the integration of solar plants in the grid. Understanding the role of clouds and aerosols in those techniques is essential for improving their accuracy. In this study, we introduce the improvements in the existing nowcasting/short-term forecasting operational systems SENSE/NextSENSE, based on a new configuration and upgrading of cloud and aerosol inputs methods and also, we investigate the limitations of such model evaluation with surface-based sensors due to cloud effects. We assess the real-time estimates of surface global horizontal irradiance (GHI) produced by the improved SENSE2 operational system at high spatial and temporal resolution (~5km, 15min) for a domain including Europe and Middle East-North Africa (MENA) region and the short-term forecasts of GHI up to 3h ahead by the NextSENSE2 system, against ground-based measurements from 10 stations across the model domain, for a whole year (2017).

Results show that the GHI estimates are within +/-50 W/m^2 (or +/-10%) of the measured GHI for 61% of the cases, after the new model configuration and a proposed bias correction. The bias ranges between -12 W/m^2 to 23 W/m^2 (or 2% to 29%) with mean value 11.3 W/m^2 (2.3%). The correlation coefficient is between 0.83 to 0.96 with mean value 0.93. Statistics are improved a lot when integrating in daily and monthly scales (mean bias 6.6 W/m^2 and 5.7 W/m^2, respectively). We demonstrate that the main overestimation of the SENSE2 GHI is linked with the underestimation of cloud optical thickness from the Meteosat Second Generation (MSG) satellites, while the relatively low overestimation linked with aerosol optical depth (AOD) forecasts (derived from Copernicus Atmospheric Monitoring Service - CAMS) results in low overestimation of clear sky GHI. The highest deviations are associated with cloudy atmospheric conditions with clouds obscuring the sun over the ground-based station. Thus, they are much more linked with satellite/ground-based comparison limitations than the actual model

performance. The NextSENSE2 GHI forecasts based on the cloud motion vector (CMV) model, outperform the smart persistence forecasting method, which assumes the same cloud conditions for the future time steps. The forecasting skill (FS) of the CMV based model compared to the persistence approach increases with cloudiness (FS up to ~20%), linked mostly to periods with changes in cloudiness, that persistence by definition fails to predict. Our results can be useful for further studies on satellite-based solar model evaluations and, in general, for the operational implementation of solar energy nowcasting and short-term forecasting, supporting solar energy production and management.

1 Introduction

Climate change mitigation along with energy production in a sustainable manner could be addressed with the deployment of renewable energy technologies (Edenhofer et al., 2011; Pörtner et al., 2022). Diverse technologies of renewable energy are investigated worldwide, and their deployment has been increasing, with solar energy markets growing rapidly, with a prospect to be the major source of energy supply in next decades (Arvizu et al., 2011; IEA, 2022). Since solar energy resources are strongly affected by atmospheric conditions, they are highly variable in space and time. The availability of solar resources is primarily affected by clouds and aerosols (e.g., Fountoulakis et al., 2021; Papachristopoulou et al., 2022). Therefore, there is a need for operational nowcasting and short-term solar forecasting for real time energy production, to better integrate solar energy exploitation technologies with national and international power systems. To increase the accuracy of those nowcasting and forecasting tools, it is imperative to understand the role of clouds and aerosols in their implementation.

The continued development and improvement of earth observation (EO) techniques in the last two decades constitute a big source of new atmospheric information. Satellite-based information and atmospheric models provide high quality data of atmospheric composition and state at high spatial and temporal resolution, considered as big data. By exploiting that multiplatform information, accurate estimates of surface solar radiation can be produced in real time (nowcasting), with numerous applications in different fields like solar energy and health (e.g., Kosmopoulos et al., 2018, 2021). The operational solar energy forecasting methods are categorized into three base methods (Sengupta et al., 2021; Yang et al., 2022) with the time horizon (few seconds to few days) and the exogenous data, i.e., sky cameras, satellite data and numerical weather predictions (NWP). Additionally, there are many statistical and machine-learning methods, which are often combined with NWP data to improve their outputs (post-processing or blending). The use of cloud motion vector (CMV) technique on satellite data is commonly used for solar forecasting of few hours ahead (e.g., Garniwa et al., 2023; Kallio-Myers et al., 2020), giving better results than the persistence solar forecasting approach, a method that assumes constant cloudy conditions for the future time steps.

SENSE is a nowcasting system of surface solar radiation developed under the EU project Geo Cradle, as a collaboration of the Beyond centre of EO research and satellite remote sensing at the National Observatory of Athens, Greece, in collaboration with the Physics and Meteorological Observatory Davos World Radiation Center, Switzerland (Kosmopoulos et al., 2018). It is a combination of geophysical input parameters from satellite-based and model data sources and a neural network (NN)

technique, trained on precalculated surface solar radiation simulations (look up table – LUT) by radiative transfer modelling. It uses the cloud optical thickness (COT) product from the Meteosat Second Generation (MSG) satellite and aerosol optical

70   depth (AOD) forecasts form the Copernicus Atmospheric Monitoring Service (CAMS) as inputs to the NN to derive the solar radiation in real time. The validation of this method showed a good agreement on daily and monthly levels; however, various sources of uncertainties have been identified, concerning mainly the use of NN especially during high irradiance atmospheric conditions, the COT, and the structure/density of atmospheric parameters in the LUTs. In the new version of the model, that has been used in the present study, the model configuration changed, and these uncertainties were limited.

75   The NextSENSE system was developed as a continuation of SENSE, during the EU project E-shape and by the same research groups. It was first introduced in the study of Kosmopoulos et al. (2020), as a novel short-term solar energy forecasting system (3h ahead), based on forecasts of satellite derived COT using a CMV technique, with solar irradiance estimated by the SENSE model. The employed CMV technique is based on state-of-the-art image processing technologies (dense optical flow). In the same study, a first evaluation of the CMV forecasts accuracy was performed, against the real MSG COT and the deviations of

80   forecasted irradiances compared with nowcasting outputs ranged from 18% to 34% under changing cloudy conditions, outperforming the persistence method for certain conditions. However, this first evaluation was based on the satellite-derived COT, so the aim of the current study is to compare the irradiance forecasts with ground-based measurements. Additionally, as NextSENSE is based on the same hierarchy of SENSE with only addition the CMV analysis, all improvements of SENSE2 are inherited in the NextSENSE2 system too.

85   The present study aims to investigate the role of clouds and aerosols in nowcasting and short-term forecasting of global horizontal irradiance (GHI), using ground-based measurements, by:

- Introducing the SENSE2 and NextSENSE2 upgrades of SENSE and NextSENSE systems, respectively.
- Validating the improved nowcasted GHI using ground based pyranometer measurements for 1 year (2017).
- Investigating cloud and aerosol effects on GHI estimates.

90   - Proposing a possible correction for GHI estimation based on MSG COT real time information.
- Validating CMV forecasted GHI and benchmarking the results with those by the persistence method.

2 Data and methods

2.1 SENSE2

SENSE2 is an operational system that produces fast estimates of GHI in real time every 15min, for a wide area including

95   Europe and Middle East-North Africa (MENA) region at high spatial resolution (~5km), calculated from earth observation (EO) data and LUTs derived from radiative transfer model (RTM) simulations. The SENSE2 presented in this study (Fig. 1) is an improved system, compared to the previous SENSE version, in terms of the parameterizations for radiative transfer

calculations and, mainly, the improvement of aerosol and cloud representation in the model using a more detailed look-up-table (LUT) and multiparametric equations for different aerosol and cloud scenes, respectively.

100 The first improvements of the SENSE2 system are that:

- the computations of clear-sky GHI are performed in the previous day, for the whole domain (1.5M pixels), every 15 min and, for the current day, the real-time cloud information is applied to provide all skies GHI in real time (no NN is used).

- the computations of clear-sky GHI are based on a new, more detailed LUT of ~16M combinations of simulated GHI
105 at the earth's surface, that was generated using the GRNET High Performance Computing Services and the computational resources of ARIS/GRNET infrastructure. The RTM simulations were performed using the libRadtran package (Emde et al., 2016; Mayer & Kylling, 2005) and Table 1 summarizes the input variables and their resolution resulted to the ~16M runs.

RTM simulations were performed spectrally from 280 to 3000nm, with 1nm spectral resolution using the DISORT radiative
110 transfer solver in pseudo-spherical mode (Buras et al., 2011). The molecular absorption parameterization of representative wavelength approach – REPTRAN was used in the solar range (Gasteiger et al., 2014) to account for the absorption of atmospheric gases for the whole solar spectrum. The Kurudz 1.0nm (Kurucz, 1994) extraterrestrial solar spectrum and the U.S. Standard Atmosphere (Anderson et al., 1986) were used as inputs. The default aerosol model of Shettle (1989) was used as the basis with modified aerosol optical properties of AOD, single scattering albedo (SSA) and Angstrom exponent (AE) varied
115 according to Table 1. The spectral global irradiances were integrated over the spectral range of the simulations to derive the GHI.

The clear-sky GHI modelled values by SENSE2 (Fig. 1) are calculated in the previous day, by linear interpolation in the 7 dimensions (7D) of the precalculated GHI LUT using the corresponding inputs. Specifically, the solar zenith angle (SZA) values are precalculated for every grid cell of the domain (1.5M in total), for 15 min time step. The main input parameter for
120 the clear-sky computations is the forecasted AOD at 550nm from CAMS (CAMS AOD hereafter). The forecasts for the day of interest are values from the CAMS run of the previous day initialized at 00:00 UTC (e.g., the AOD used to simulate the GHI for the 24th of a month, has been derived from the CAMS run that started on the 23rd at 00:00 UTC). Climatological values are used for the interpolation in the 7D LUT for the additional aerosol optical properties of SSA and AE (MAcv2 climatology, Kinne, 2019), for the water vapor (WV) (CAMS reanalysis (Inness et al., 2019)), the total ozone column (TOC) (OMI TOC
125 data (Bhartia, 2012) based climatology) and surface albedo (GOME-2 database (Tilstra et al., 2017, 2021)). It should be mentioned that the interpolation procedure in the 7D LUT was added in the new SENSE2 to further improve the accuracy of the GHI estimations. Finally, since the results of the RTM runs are at sea level and for the mean earth-sun distance, a post correction of the clear-sky GHI values from the LUT is performed for the surface elevation following the methodology described in Fountoulakis et al., (2021) and the actual earth-sun distance for the particular day of the year (DOY).

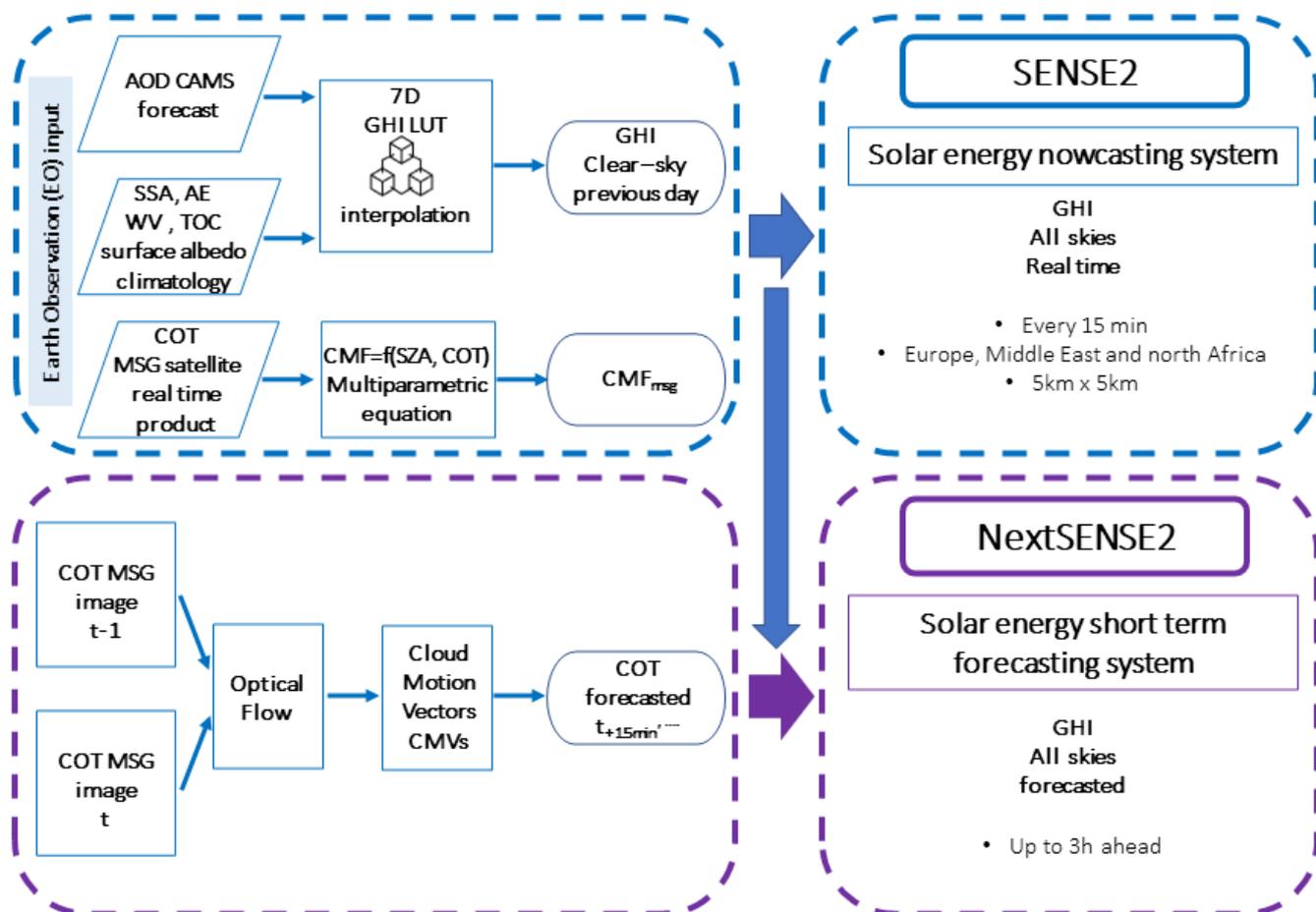



Figure 1 Schematic overview of solar energy nowcasting system (SENSE2) and short-term forecasting (NextSENSE2) up to 3h ahaid.

Table 1 Input parameters to radiative transfer simulations performed at the ARIS GRNET supercomputer resulted to the 7D GHI LUT.

Parameter	Range	Resolution
Solar zenith angle (SZA in deg)	1 to 89	1
Aerosol optical depth at 550nm (AOD)	0 to 2, 2.5, 3.0	0.05
Single Scattering Albedo (SSA)	0.6 to 1	0.1
Angstrom exponent (AE)	0 to 2	0.4
Total Ozone Column (TOC in DU)	200 to 500	100
Water Vapor (WV in cm)	0.5 to 3	0.5
Surface albedo	0.05 to 0.8	0.15

Another improvement is related to the cloud representation in real time, using multi-parametric equations for different cloud scenes, based on the Cloud Modification Factor (CMF) concept, instead of using the COT as an input parameter in the RTM. The computation of the all-skies GHI in real time, every 15min, is based on the COT product we extract operationally in real time using broadcasted MSG satellite data and the software package provided by EUMETSAT Satellite Application Facilities

140   of Nowcasting and Very Short Range Forecasting, NWC SAF. To provide timely the all skies GHI SENSE2 product for 1.5M pixels, neither the running of radiative transfer simulations nor the multi-dimension interpolations would be sufficiently fast. Instead, a multi-parametric equation was constructed, fitted on libRadtran simulations for a wide range of COT values for different SZAs (points in Fig.2a). In our simulations, spherical droplets were assumed having typical climatological mean heights (base at 2km, 3km height) and microphysical properties. Typical values for the effective radius (R_{eff} = 10 μm) and the

145   liquid water path (LWP = 1 g/m^3) were used, given the unavailability of those data and their small impact on GHI (Oumbe et al., 2014; Qu et al., 2017). The simulated GHI for each COT was divided by the GHI for COT=0 (clear sky) for the same SZA to derive the CMF (Eq. 1). The CMF ranges from zero (overcast conditions) to 1 (clear sky) and it is easy to use to provide all skies GHI by simply multiplying clear-sky GHI with CMF (Eq. 3). The libRadtran-derived CMF for each SZA were fitted against COT using the hyperbolic tangent function. The resulting fits are shown as solid lines in Fig.2a and are mathematically

150   expressed by the multi-parametric Eq. 2.

$$CMF = \frac{GHI}{GHI_{clr}} \qquad (1)$$

$$CMF = 1 - \tanh^b(COT^a) \qquad (2)$$

where a and b are polynomials of SZA

$$a = 2.24 \cdot 10^{-1} + 2.81 \cdot 10^{-4} \cdot SZA - 2.18 \cdot 10^{-5} \cdot SZA^2 + 3.71 \cdot 10^{-7} \cdot SZA^3 - 2.65 \cdot 10^{-9} \cdot SZA^4 \qquad (2a)$$

155   $$b = 12.2 + 5.27 \cdot 10^{-3} \cdot SZA - 2.24 \cdot 10^{-3} \cdot SZA^2 + 8.33 \cdot 10^{-6} \cdot SZA^3 + 3.94 \cdot 10^{-8} \cdot SZA^4 \qquad (2b)$$

The real time MSG COT is used as input in Eq. 2 every 15 min, along with SZA, for ~1.5M pixels, to calculate the CMF (CMFmsg hereafter). Apart from being very fast, the use of this formula to calculate CMFmsg is also accurate, as it can be seen by the comparison of the CMF values derived by Eq. 2 against those of libRadtran runs (Fig. 2b). CMF differences are less than 0.015 (or 1.5%) for SZA lower than 70 degrees, while they are up to 0.03 (3%) for SZAs between 80 and 90 degrees,

160   showing the very good representation of the CMF as a function of COT with Eq. 2. In terms of accuracy this means that using Eq. 2 is almost the same as running RTM simulations, but in terms of computational time is by far more efficient in the operational mode. Finally, by multiplying CMFmsg with clear-sky GHI, the all-skies GHI product is provided (Eq. 3), in less than 1 min for 1.5M pixels.

$$GHI = GHI_{clr} * CMFmsg \qquad (3)$$

(a)   (b)

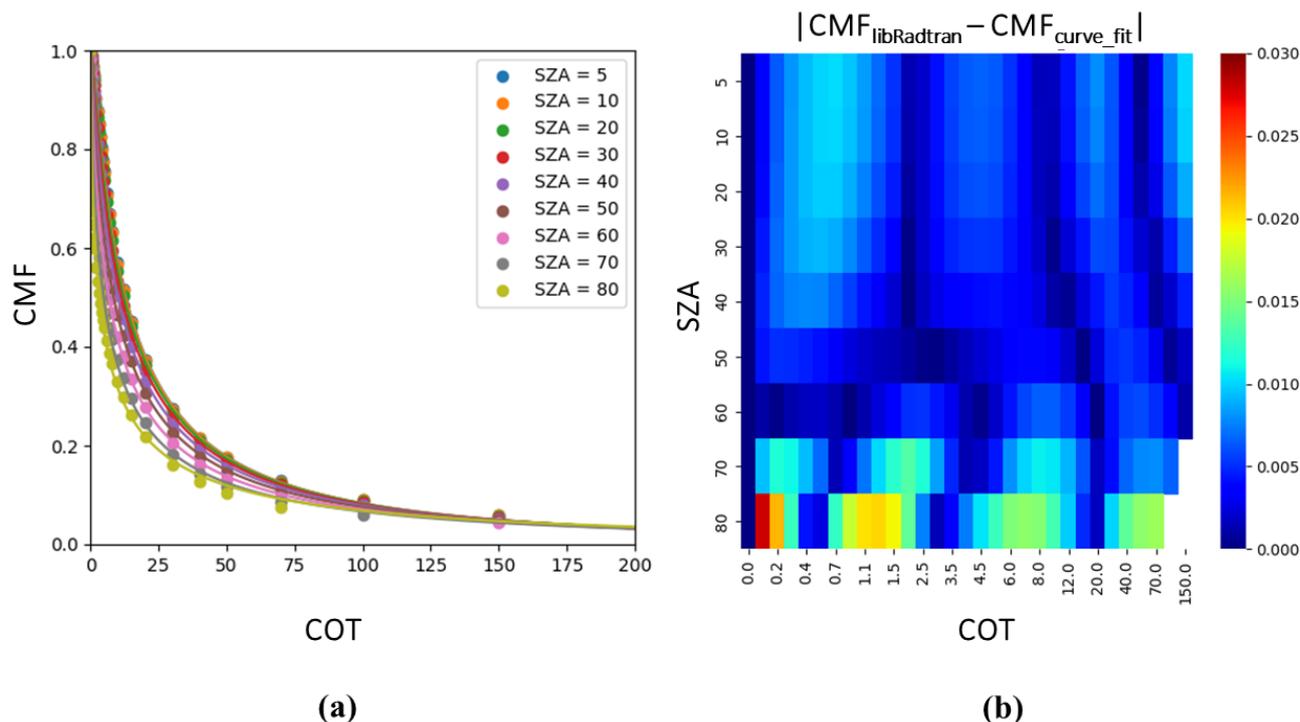



Figure 2 (a) Cloud modification factor (CMF) versus cloud optical thickness (COT) and solar zenith angle (SZA) based on radiative transfer simulations of global irradiances using the libRadtran package. CMF is the ratio of global horizontal irradiance (GHI) values to those under cloudless conditions (COT=0). (b) Differences between the CMF directly derived from libRadtran simulations against those derived from Eq. 2, as a function of COT and SZA.

The new SENSE2 configuration was built for the improvement GHI nowcasting on the 15 min time scale. Additionally, it allows a greater flexibility for the system to:

- include reanalysis or measured data of AOD and other optical properties e.g., CAMS reanalysis or AERONET measurements.

- extent to other output products. Apart from GHI, direct normal irradiance (DNI) or total irradiance on tilted surface could be also produced. By introducing spectral information and making the appropriate modifications, products related to specific spectral regions could be also derived (e.g., UV Index by using real time input for TOC, photosynthetically active radiation (PAR) etc).

- run a past time series of one or few locations, autonomously, using as input actual measurements. In this case, if there is no time constrain, model runs could be performed without the parametrizations (LUT and multiparametric functions).

2.2 NextSENSE2

NextSENSE2 is the operational system that provides forecasts of GHI up to 3-hours ahead with a 15-min time step by applying a CMV technique to the MSG COT product (Fig. 1). In this section, we describe the method employed to produce forecasted

COT, which is the main input to derive the operational forecasts of GHI. All the other EO inputs and the radiative transfer parameterizations for fast estimates of forecasted GHI are the same as those described in the previous section for the SENSE2 model.

We use CMVs to predict the motions of the clouds and project their future positions. The CMVs in NextSENSE2 are calculated by applying a state-of-the-art optical flow algorithm from the computer vision community. Optical flow is the apparent motion of objects between consecutive frames, caused by the relative movement between the object and a camera. We apply the Farnebäck (2003) two-frame motion estimation technique to images of COT product (Kosmopoulos et al., 2020). By applying the algorithm to two consecutive images of satellite derived COT the optical flow displacement vectors are calculated. This CMV field is applied to the later COT image (real) to get the next COT image (forecasted COT). This procedure is repeated for 12 times resulting to the 3h forecasting horizon. Main assumptions are the brightness constancy and that the cloud's displacements are only two dimensional (image plane). More details regarding the CMV model and forecasted COT can be found in Kosmopoulos et al. (2020).

2.3 Smart persistence

We benchmarked the CMV forecasted GHI of NextSENSE2 system with the so-called smart persistence approach, which assumes that the state of the clouds remains constant for future time steps, while all other variables like SZA etc. dynamically change. Hence, it uses the same COT values from the later satellite information as input to the next 12 time-steps in order to forecast GHI up to 3h ahead.

2.4 Ground based irradiance measurements

To validate the modelled GHI, ground-based measurements from pyranometers were utilized. The 1-min GHI ground based measurements were collected from stations of the Baseline Surface Radiation Network (BSRN; Driemel et al., 2018) that are within the study area and have data throughout 2017, and from two additional stations at Athens (ASNOA: NOA's Actinometric Station) and Thessaloniki. Table 2 summarizes the information of the 10 in total stations utilized and Fig. 3 depicts their geographical locations.

BSRN station-to-archive files were accessed and manipulated using the SolarData v1.1 R package (Yang, 2019). The function that reads the data from the station-to-archive files also computes several auxiliary variables such as solar zenith angle, clear sky irradiances using the Ineichen-Perez clear sky model (Ineichen & Perez, 2002) and extraterrestrial GHI. Using the same methodology, the Ineichen-Perez clear sky model values were also computed for the non BSRN station data.

The BSRN recommended Quality Check (QC) tests (Long & Dutton, 2010) where performed to the collected measurements, to ensure the best quality of the measurements. Measurements that were not respecting the above QC tests, were flagged, and set to as a missing value.

Table 2 Detailed information about the ground-based sations used in this study.

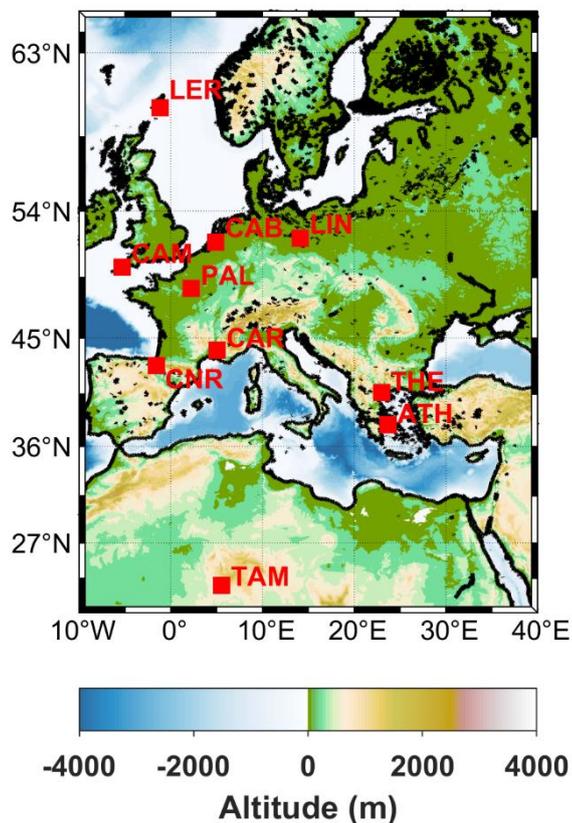
| Name | Ground based pyranometer | | | location | AERONET station |
	network	Lat. (°N)	Lon. (°E)		
ATH - Athens	-	37.9	23.7	Europe/Greece	Co – located
CAB – Cabauw	BSRN	51.9711	4.9267	Europe/Amsterdam	Co – located
CAM – Camborne	BSRN	50.2167	-5.3167	Europe/London	Co – located
CAR – Carpentras	BSRN	44.083	5.059	Europe/Paris	Co – located
CNR – Cener	BSRN	42.816	-1.601	Europe/Madrid	Co – located
LER – Lerwick	BSRN	60.1389	-1.1847	Europe/London	Co – located
LIN – Lindenberg	BSRN	52.21	14.122	Europe/Berlin	Co – located (metObs LIN)
PAL – Palaiseau, SIRTA Obser	BSRN	48.713	2.208	Europe/Paris	Co – located
TAM – Tamanrasset	BSRN	22.7903	5.5292	Africa/Algiers	Co – located
THE -Thessaloniki	-	40.63	22.96	Europe/Greece	Co – located

215

Figure 3 Locations of the ground-based stations measuring global horizontal irradiance (GHI) that are used in the current study. These are eight BSRN stations, plus Athens, and Thessaloniki, Greece.

2.5 Ground based aerosol information

To assess the CAMS AOD forecasts used as input to the model, ground-based measurements of AOD from the AERONET

220   network (Holben et al., 1998) were used. All the ground-based stations with pyranometer data (BSRN, Athens and Thessaloniki) have a collocated AERONET station (see Table 2). The level 2, version 3 direct sun (Giles et al., 2019) AOD data at 500nm were collected and using the Ångström exponent for 440-675nm, the AOD values at 550nm were derived. Only for Cabauw, measurements of AOD at 500 nm weren't available, therefore AOD at 440 nm was used instead and converted to 550 nm using the Ångström exponent for 440-675 nm.

225   **2.6 Evaluation metrics**

For the validation of the SENSE2/NextSENSE2 derived GHI values against ground-based measurements, common statistical metrics have been adopted. Given that the error is defined as the difference between modelled values (x_{m_i}) and observed values (x_{o_i}), we have the following common metrics:

Mean Bias error:

230
$$MBE = \frac{1}{N}\sum_{i=1}^{N}(x_{m_i} - x_{o_i}) \qquad (4)$$

Root mean square error:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{m_i} - x_{o_i})^2} \qquad (5)$$

And Pearson correlation coefficient R.

235   An additional metric the forecast skill (FS) was used to assess the performance of CMV forecasted GHI using persistence model as a benchmark model:

$$FS = 1 - \frac{rRMSE_{CMV}}{rRMSE_{pers.}} \qquad (6)$$

3 Results and Discussion

The results are discussed separately for the evaluation of nowcasted GHI (Section 3.1) related to SENSE2 outputs (modelled

240   GHI hereafter) and the evaluation of the short-term forecasted GHI (Section 3.2), namely the NextSENSE2 product (forecasted GHI, hereafter). The comparisons between ground based and estimated GHI where restricted to SZAs below 75° (i.e., for solar height above 25° from the local horizon) due to the limitations in the field of view of the satellite.

The CMF derived from the ground-based measurements of GHI was used in our analysis to categorize the cloudiness conditions. Specifically, the CMF was calculated as the ratio (Eq. 7) of measured GHI to the clear sky irradiance calculated by the Ineichen-Perez clear-sky model (Ineichen & Perez, 2002) (See Section 2.4).

$$CMF = \frac{GHI_{measured}}{GHI_{clr}} \qquad (7)$$

Three categories according to CMF are used in the following: CMF>0.9 for clear sky conditions, 0.4<CMF<0.9 for partially cloudy conditions and CMF<0.4 for overcast conditions.

3.1 Nowcasting

3.1.1 Overall performance

Figure 4 presents the overall performance of the SENSE2 system at the 15min time scale, by comparing the modelled GHI values against ground-based measurements, for all stations, for a whole year (2017). We can see that most of the points (Fig. 4a, number of cases N>600) fall on the 1:1 line (blue line) which indicates the overall good performance of the system, with a correlation coefficient of 0.93. For 58 % of the cases, the absolute differences between modelled and ground-based measurements of GHI are within +/-50 W/m^2 or +/-10 % (Fig. 4b). The SENSE2 system mostly overestimates the GHI, which corresponds to points above identity line (Fig. 4a), with MBE 23.8 W/m^2 (4.9%). This overestimation is attributed to the underestimation of cloud related information from satellite (MSG COT), when we compare point measurements with a pixel in satellite images corresponding to a wide area of almost 5 km x 5 km. This overestimation is more pronounced for low irradiances (low left corner of Fig. 4a with GHI<250 W/m^2), indicating the challenging task of accurately modelling GHI based on satellite data for stations with enhanced cloudiness. Especially at high latitudes, measurements with clouds and at large SZAs are more frequent. Lerwick is the most northern station and at the same time the station with the greatest MBE (Fig. 4b). There are also points below the identity line, which indicate that the measured GHI is greater than the modelled. This is attributed partly to the irradiance enhancement by clouds that occurs often when the sky above the ground station is partially cloudy, and the sun is visible. In this case, the reflection of solar radiation by clouds increases the diffuse component, and hence the GHI, at the ground, even though the satellite "sees" a cloudy pixel. The cloud effect on GHI estimates is investigated in more detail in Section 3.1.3.

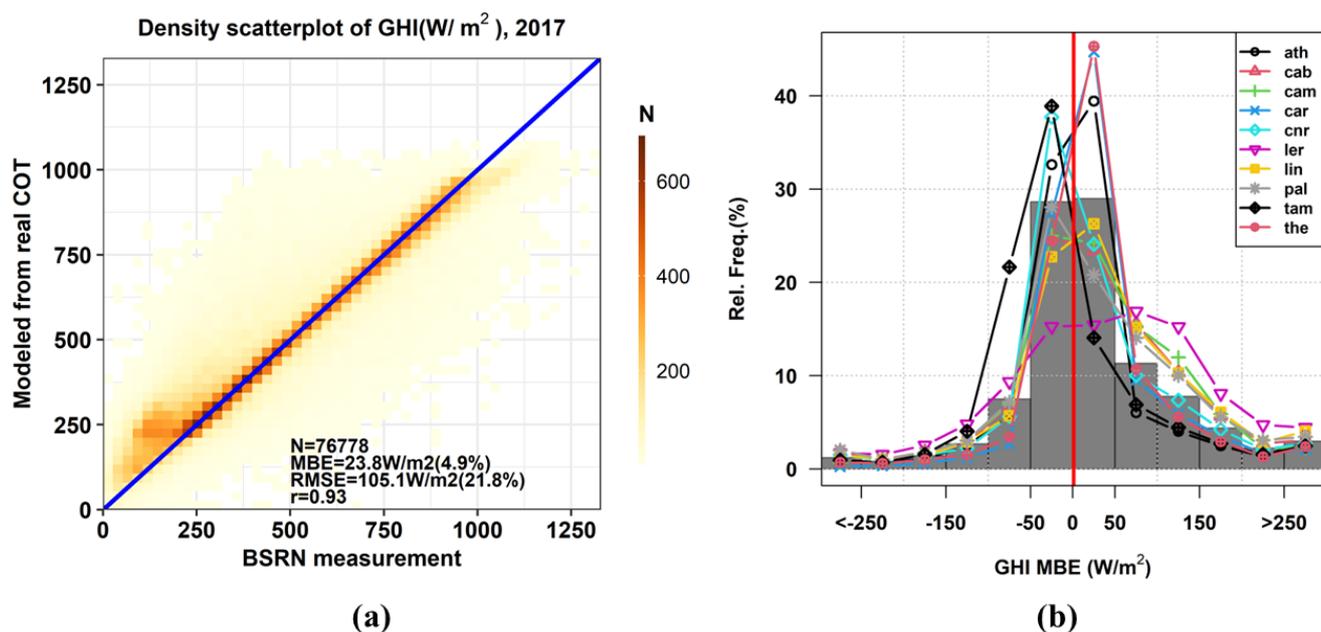



Figure 4 (a) Comparison of the global horizontal irradiance (GHI) modeled versus measured for all ground based stations, for 2017. (b) Relative frequency of GHI MBE for all stations (grey bars) and for each station (lines with different symbols and colors).

270   All the statistical metrics are drastically improved with increasing time scale for all stations (Fig. 5). Stations with similar results are the northern most stations CAB, CAM, LER, LIN, and PAL. At the 15 min time scale their MBE, RMSE and correlation coefficient range 29-43 W/m^2, 104-131 W/m^2 and 0.82-0.90, respectively. Those statistics are improved for the monthly means to 5-10 W/m^2 for MBE, to 7-13 W/m^2 for RMSE and to r~1. Similar results were found for the rest southernmost stations (MBE, RMSE and correlation coefficient range from -7 to 30 W/m^2, from 84 to 104 W/m^2 and from

275   0.93 to 0.95, respectively, for 15 min time scale and from -4 to 8 W/m^2, from 6 to 10 W/m^2 and r~1, respectively, for monthly means). The overall MBE and RMSE is reduced to 6.6 W/m^2 (3.3%) and 15.4 W/m^2 (7.7%) for the daily mean GHI and to 5.7 W/m^2 (3.2%) and 9.2 W/m^2 (5.2%) for the monthly means, while correlation coefficient reaches values up to almost 1, which was anticipated since the cloud effect is smoothed out for larger time scales.

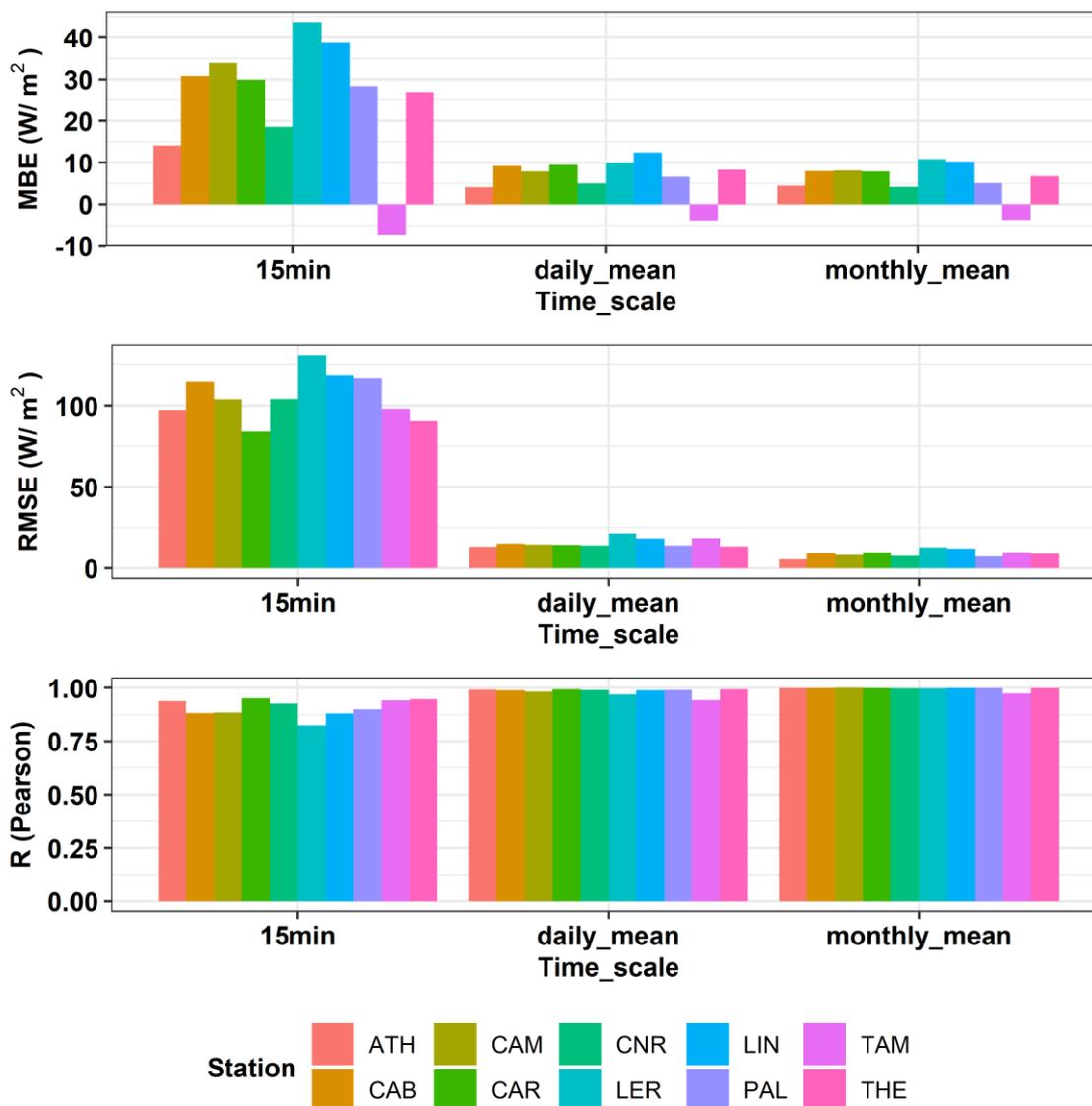



Figure 5 Comparison of the global horizontal irradiance (GHI) modeled versus measured per ground based station, for 2017, for different time scales (15 min, daily mean and montlhy mean).

3.1.2 Aerosol effect on retrieved solar irradiance

The CAMS AOD forecasts used as input to the operational model were assessed against ground-based measurements from the AERONET network, and the related uncertainty introduced to modelled GHI has been calculated. The AERONET AOD direct sun measurements were matched with CAMS AOD forecasts (1h time resolution) interpolated to the 15 min time steps of the model. The closest AERONET measurement +/- 10 minutes around the 15 min time steps were matched (or the mean value if more than one measurements were available). To estimate the model uncertainties due to forecasted AOD, the clear sky GHI

was calculated using as input first the forecasted CAMS AOD and second, the synchronized AERONET AOD measurements.
290   The comparison for AOD and modelled GHI is presented in Fig. 6 per station in terms of MBE.

CAMS forecasts mostly overestimate AOD with MBE 0.015 (10%) for all stations which results to an underestimation of modelled clear sky GHI -2.7 W/m^2 (-0.4%). The greatest overestimation 0.05 (~50%) was found for CAM and CNR which resulted to the greatest underestimation of clear sky modelled irradiances -8.5 W/m^2 (-1.4%). Underestimation of AOD was found for CAB and THE, with MBE<0.01 (<3%), resulting in negligible overestimation of modelled irradiances (MBE< 295   1W/m^2 or 1%).

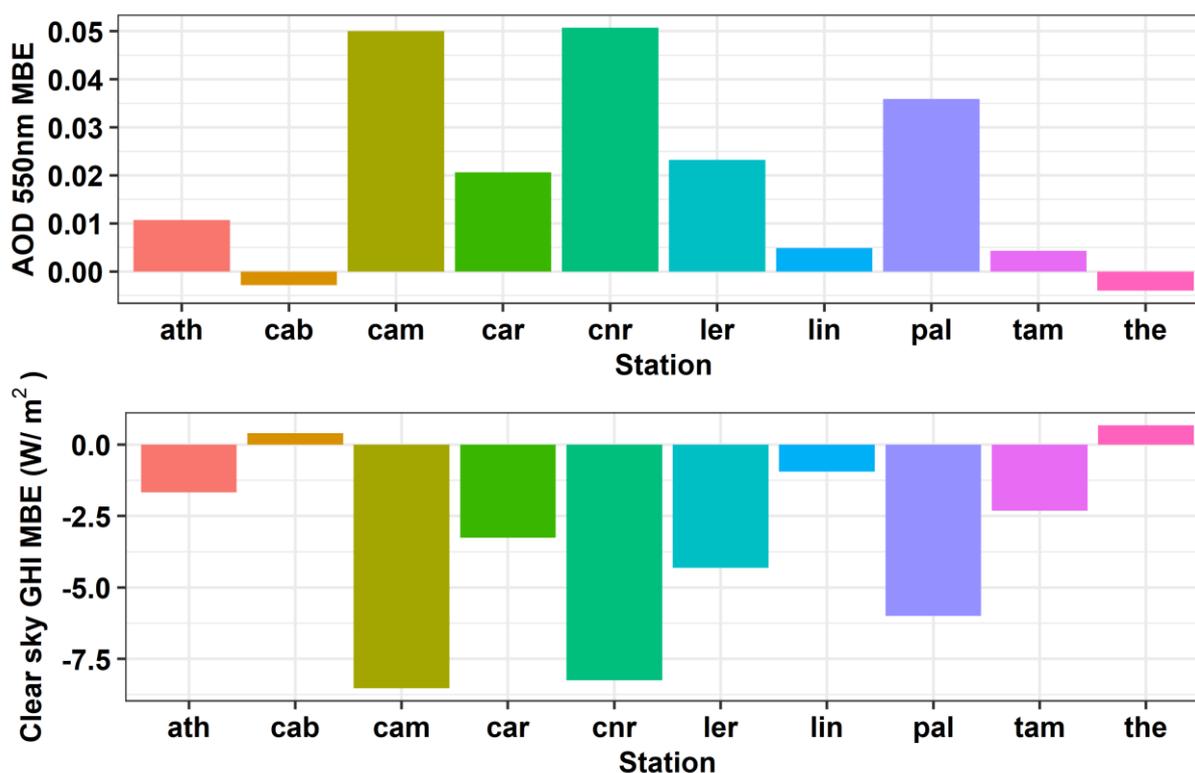



Figure 6 Upper panel: Mean bias error (MBE) of aerosol optical depth (AOD) at 550nm forecasted by CAMS (1 day ahead forecast) compared to AOD measured by ground based sun photometers of the AERONET network. Lower panel: MBE of global horizontal irradiance (GHI) modeled under clear sky conditions using as input CAMS forecasted AOD at 550nm versus measured values
300   **(AERONET).**

3.1.3 Cloud effects on retrieved solar irradiance

Overall, the model relatively overestimates GHI, as we saw in Section 3.1.1, which can be attributed to the underestimation of cloud information from satellite (MSG COT). To decompose the error induced due to satellite cloud information, we classify the cloudiness conditions using CMF (Fig 7a). We can see that the GHI is overestimated by the model under cloudy conditions 305   (CMF<0.9), while for clear sky conditions (CMF>0.9) the model closely resembles the measured GHI. The category of

CMF>0.9 contains apart from the conditions with actual clear sky, also the CMF >1 which exists due to the cloud enhancement of solar irradiance, responsible for the underestimation of modeled GHI in this category (see Section 3.1.1). For partially cloudy conditions (CMF <0.9 and >0.4), the MBE is 81.6 W/m^2 (22.8%) and the greatest error occurs for the low CMF values (CMF<0.4) (MBE=100.1 W/m^2 or 73.1%). In this category, the lowest values of measured GHI are found (<250 W/m^2). The

310   high modeled GHI values could be attributed to the underestimation of MSG COT product due to complex geometries related to high latitudes and large SZAs. Additionally, there are also situations when the low measured GHI values are related to partially cloudy sky, with clouds blocking the sun, a situation that cannot be resolved by the satellite COT retrieval. To demonstrate this last situation, we compared the modelled and measured GHI values for MSG COT=0 (Fig. 7b, clear sky from the satellite point of view). In this case, the model overestimates GHI with MBE 13.6 W/m^2 (2.3%). Most of the points are

315   above the 1:1 line, especially, for measured GHI<250 W/m^2. The interesting part is that the same case stands for the whole range of measured GHI, indicating that it is a general limitation of satellite that it cannot take into account clouds that obscure the sun.

To verify that the sun visibility is the main reason for the overestimation, we tried to separate those instances by using the pyrheliometer measurements of direct irradiance (DNI – direct normal irradiance) available by the BSRN network. The DNI

320   measurements (1min) were divided with the clear-sky DNI, again from the Ineichen-Perez clear sky model (Ineichen & Perez, 2002). According to this ratio values, we classified as sun visible the situations with ratio >0.8. This threshold was selected to account for the strong effect of aerosols in DNI, given that a monthly mean climatological value for aerosol attenuation factor is used by DNI clear-sky model (Ineichen & Perez, 2002). We classified as sun obscured situations with ratio <0.6, and we omitted as unclassified situations those with ratio values between 0.6 and 0.8, to be confident that direct irradiance is blocked

325   by clouds by more than 40%. The results of the GHI comparison between modeled and measured values were grouped based on the sun visibility classification and are presented in Fig. 7c. We can see that the sun visible situations give quite good results (points close to the 1:1 line). In addition, several cloud enhancement events (points below 1:1 line), cannot be captured by the model. In contrast, the model overestimates GHI (MBE 74.5 W/m^2 or 24.5%) when the sun is obscured, a situation that cannot be inferred from satellite information. Comparing station (point) data with model-derived, based on satellite pixel-based cloud

330   retrievals, introduces deviations linked with the cloud features within this pixel. Even if the satellite imager can resolve a cloud within the pixel, the COT product for this pixel is a constant value, namely a spatially homogeneous cloud optical property for the corresponding area. In this atmospheric scene, GHI measured at the ground level, with the sun obscured or not, will affect the comparison dramatically. Cases with partial cloudiness and the sun obscured as seen from the ground sensor (almost total attenuation of direct irradiance) will be associated with low measured irradiance that cannot be captured by the model. This is

335   the main reason of the overall model overestimation.

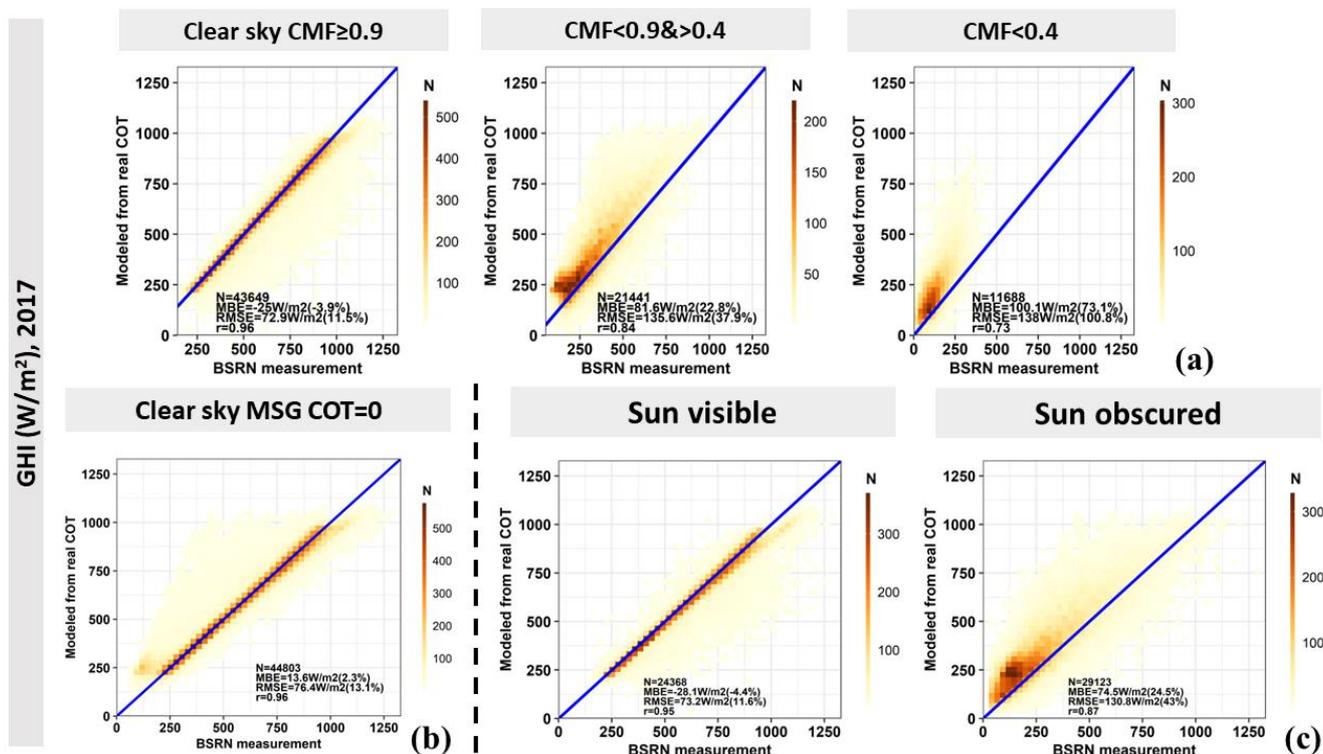



Figure 7 Comparison of the global horizontal irradiance (GHI) measured versus modeled for all ground based stations (a) for different cloudiness conditions based on the cloud modification factor (CMF) derived by the ratio of GHI ground based measurements devided by clear-sky GHI (clear-sky model) (b) for clear-sky conditions as determined by MSG satellite product by zero values of cloud optical thickness (COT=0) (c) for conditions characterized as sun visible or obscured.

Since the main source of errors in this analysis is associated with clouds, we further investigate the cloud input from satellite data. The MSG COT is transformed to CMFmsg using Eq. (2) and this is the cloud related input in the SENSE2 model. We compared the CMFmsg with the CMF derived from GHI measurements (Eq. 7) and the results are presented in Fig. 8, as relative frequency distributions of CMFmsg, CMF and their difference (CMFmsg – CMF), for all cases and different cloudiness conditions. In Fig. 8a and b, for all situations, we can see the overall overestimation of CMF (0.17) by satellite (underestimation of MSG COT), which is the reason for the overall overestimation of modelled GHI. This overestimation is linked to the partially cloudy (CMF <0.9 and >0.4, Fig. 8g and h) and the overcast (CMF<0.4, Fig. 8i and j) conditions. The large relative frequencies of CMFmsg=1 in Fig. 8i are situations when the satellite cannot resolve any clouds giving COT=0 or CMFmsg=1, but the measurements demonstrate the presence of clouds. There is also a large fraction of enhancement events (CMF > 1) which is not captured by the CMFmsg (Fig. 8a). During those events satellite cloud product gives mostly no clouds (CMFmsg=1, Fig. 8c and e).

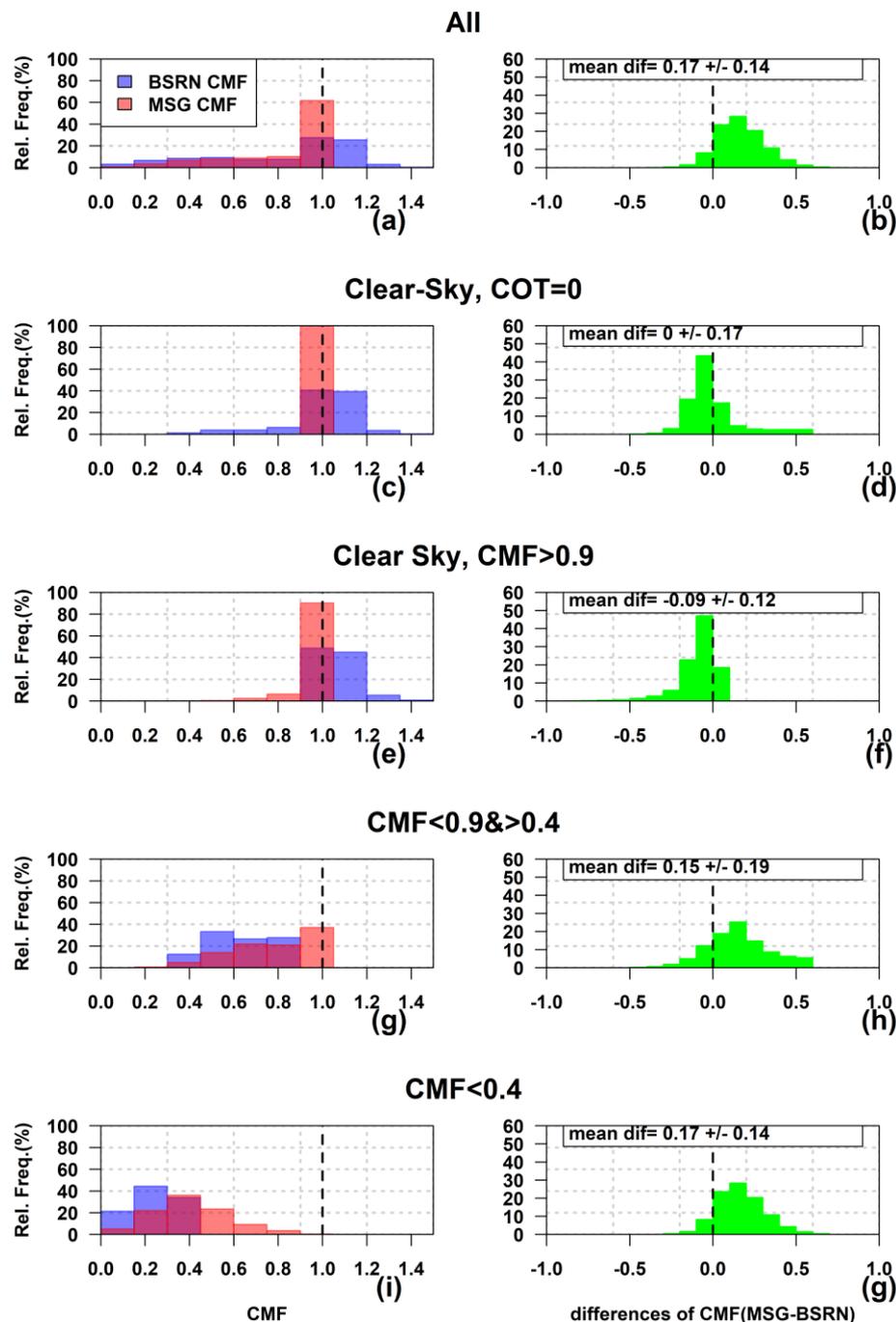



Figure 8 Left panels (a, c, e, g, i): Distribution of cloud modification factor (CMF) from measurements of global horizontal irradiance (GHI) (blue bars) and from MSG cloud optical thickness (COT) (red bars), under all cases and under different cloudiness conditions. Right panels (b, d, f, h, g): Distribution of the CMF differences between those derived from measurements of GHI and those derived from MSG satellite COT.

3.1.4 Bias correction based on cloud input

The model overall relatively overestimates GHI, which is attributed to the CMFmsg overestimation (COT underestimation) under cloudy conditions. Based on the main conclusions from CMF differences from the previous section, we tried to find if there is any common pattern for all stations, in those CMF differences (modelled against measured) as a function of CMFmsg, since the latest is the only operationally available input, every 15min. Additionally, we found that those differences hardly change with SZA (Fig. A1), so we will investigate their relation only with CMFmsg.

We calculated the mean CMF difference and their standard deviation per CMFmsg bins for every station and the results are presented in Fig.9. A pattern of mean CMF differences was found for almost all station (apart from TAM, ATH and THE) with CFMmsg overestimation up to almost 0.1 starting from CMFmsg bin 0.3 up to 0.8, related also with low standard deviations over those bins.

As we discussed in the previous section, this CFMmsg overestimation (up to ~0.1) is mostly related with the sun obscured conditions. Nevertheless, the sun's visibility is information that couldn't be provided by satellites. Consequently, we tried to correct CMFmsg (the operational input) with those CMF differences (modelled against measured). We used the mean of the of CMF differences per CMFmsg bin from seven out of ten stations (excluding TAM, ATH and THE) to derive the correction factor (the correction hereafter), which is depicted as the thick black dashed line in Fig. 9. The correction was applied to CMFmsg values falling in the bins 0.3-0.8 only. The correction was applied to all stations, including TAM, ATH and THE, which act as a test bed (low frequency of cloudy cases) for the general correction derived from the rest seven stations.

Table 3 summarizes the statistics of corrected modeled GHI against ground-based measurements. We can see that the correction was successful for all stations. LER and CAM are the two stations with the greatest improvement in their statistics, following CAB and LIN, which was anticipated since those are stations at higher latitudes associated with high cloudiness. Even ATH and THE stations, that were independent from the correction construction procedure, exhibit better results after applying the correction. TAM is the only station where statistics weren't improved. Due to its rare cloudiness, this station's statistics were already good, indicating that probably a hybrid approach of the correction according to the area's cloudiness would be better. Overall, after the correction for 61% of the cases the GHI differences (modeled against measured) were within +/-50 W/m^2 (or +/-10%). The MBE for all stations was also improved to 11.3 W/m^2 (2.3%), compared to the uncorrected values (23.8 W/m^2 or 4.9%).

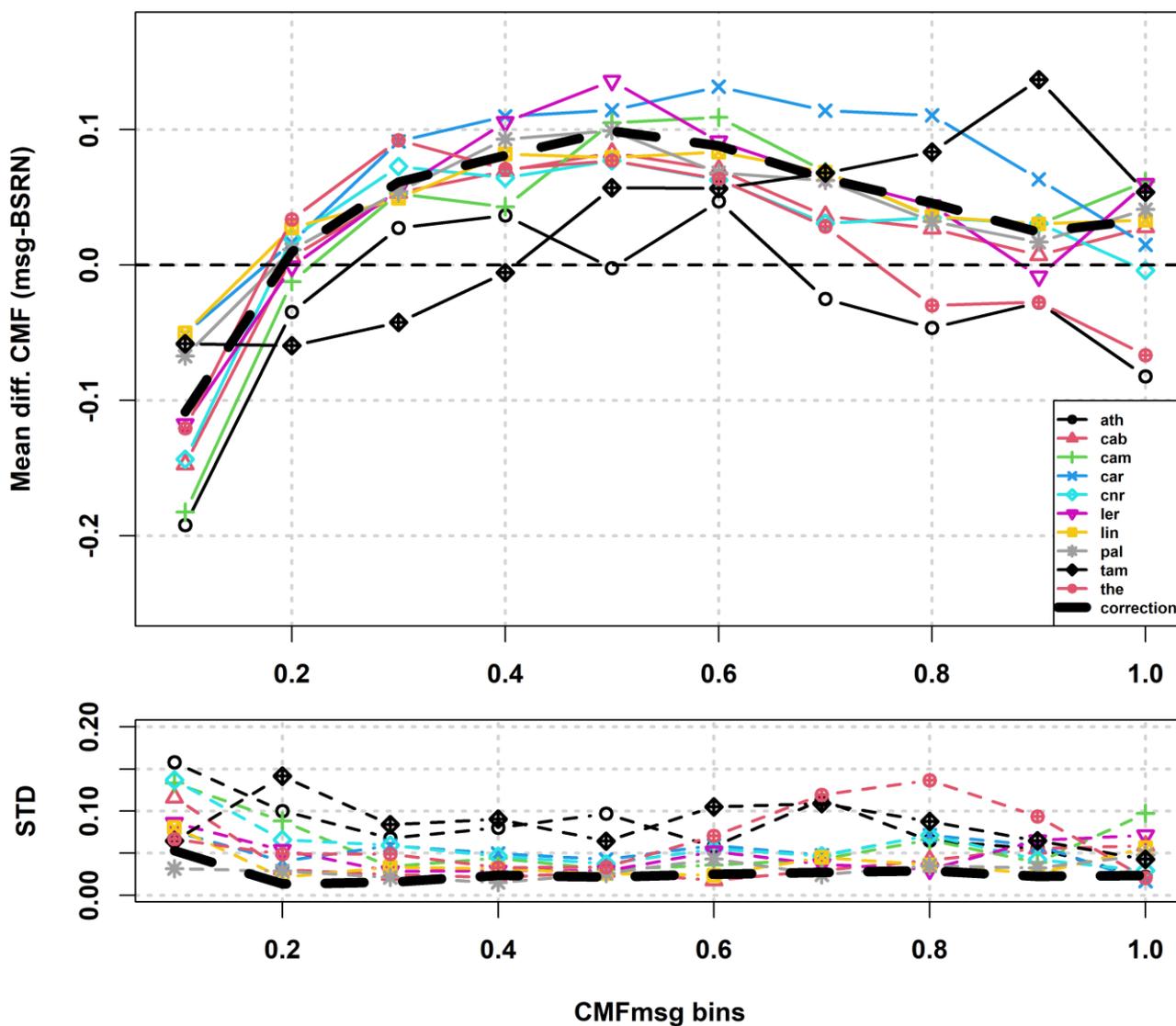



Figure 9 Upper panel: Mean cloud modification factor (CMF) difference between modeled (from MSG satellite cloud optical thickness) values CMFmsg and those derived from global horizontal irradiance (GHI) measuremnts per modeled CMFmsg bins. Lower panel: The coresponding standard deviation (STD) of CMF differences per modeled CMF bins. The different colors in lines and different symbols correspond to different stations.

390

Table 3 Performance of nowcasted irradiances before and after correction with CMFmsg.

BSRN station	N	MBE (W/m²)	MBE cor. (W/m²)	RSME	RSME cor.	r	r cor.
ATH	9472	14	8	97	97	0.94	0.94
CAB	7749	31	11	115	110	0.88	0.88
CAM	2376	34	13	104	98	0.89	0.89
CAR	8985	30	23	84	80	0.95	0.96
CNR	8806	19	6	104	102	0.93	0.93
LER	5191	44	21	131	124	0.82	0.83
LIN	7989	39	21	119	114	0.88	0.88
PAL	8011	28	11	117	113	0.90	0.91
TAM	9011	-8	-12	98	98	0.94	0.94
THE	9188	27	18	91	88	0.95	0.95

3.2 Short term forecasting

3.2.1 Overall performance - Benchmark with Persistence method

Figure 10 summarizes the performance of the CMV method (green points) in predicted GHI as function of forecasting horizon, by providing main statistics, after comparison with GHI ground based measurements, from all ten stations, for a whole year (2017). Detailed results per station, for representative statistics and selected time steps (+60, +120, +180 mins) can be found in Table 4. As a benchmark, the results of the commonly used persistence forecasting method are presented also in Fig. 10 (black points). We can see that the CMV model systematically outperforms persistence for all time steps. It is interesting that the first time step (+15 min) is not the one with the maximum difference between the CMV and persistence statistics (or the maximum of CMV FS%), indicating that for such short time interval the probability of changing cloudiness is low, which favours the persistence method. The second time step is the one with the maximum of CMV FS% (best performance) compared to persistence up to ~10%. As the forecasting horizon increases all metrics deteriorate for both methods, while, at the same time, persistence is systematically worse than CMV.

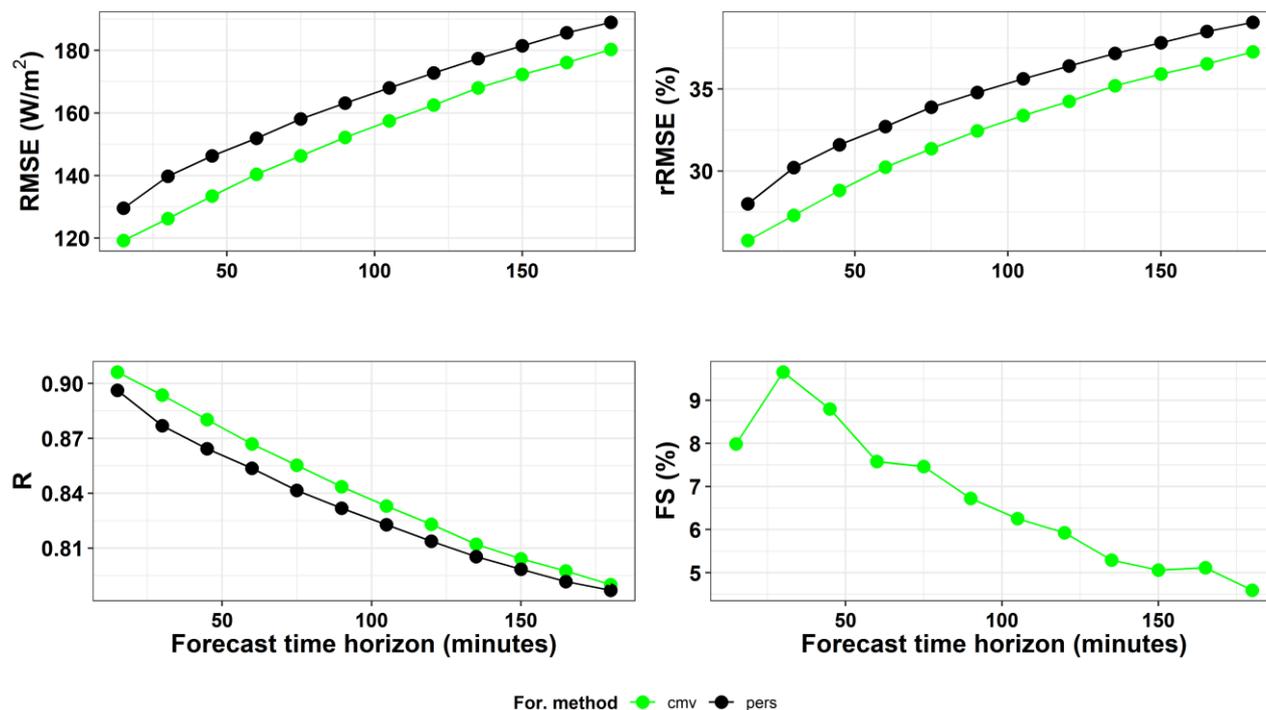



Figure 10 Performance statistics of CMV modeled (green points) and persistance method (black points) forecasted global horizontal irradince (GHI) for every 15min time step up to 3h ahead.

410 **Table 4 Performance statistics of CMV modeled forecasted global horizontal irradince (GHI) for the 60min, 120min and 180min time steps.**

station	Mean CMF	rRSME (%) time step (min)			R time step (min)			FS (%) time step (min)		
		+60	+120	+180	+60	+120	+180	+60	+120	+180
ATH	0.97	22.0	24.5	25.3	0.90	0.87	0.86	0	1.7	3.2
CAB	0.68	39.7	45.6	49.6	0.80	0.73	0.68	10.5	7.5	5.8
CAM	0.63	54.1	62.0	68.9	0.66	0.58	0.50	2.7	3.4	1.5
CAR	0.85	22.8	26.3	29.6	0.90	0.86	0.82	8.1	5.6	3.1
CNR	0.81	31.0	34.8	37.4	0.85	0.79	0.75	2.4	2.9	2.6
LER	0.61	50.8	56.0	59.6	0.73	0.67	0.62	10.4	9.5	9.4
LIN	0.68	39.7	45.9	51.2	0.81	0.74	0.69	13.9	10.9	8.8
PAL	0.68	38.9	44.9	48.8	0.82	0.75	0.71	11.6	7.6	6.0
TAM	0.87	21.4	23.1	24.7	0.89	0.86	0.82	2.2	1.0	-1.3
THE	0.91	23.9	27.2	29.2	0.89	0.86	0.84	6.8	5.1	2.4

An interesting grouping of stations resulted by comparing main statistics (rRMSE and FS) for both forecasting methods with stations' mean CMF, representing their mean cloudiness (Fig.11). Three time steps were selected (+60, +120, +180 min) and

415 are depicted with increasing transparency. Two groups of stations are evident. Those with high mean cloudiness (LER, CAM, PAL, LIN, and CAB), which show worse rRMSE than those of lower cloudiness (ATH, THE, TAM, CAR and CNR), independently of the method. Again, the CMV model (green symbols) outperforms the persistence method (black symbols) for all stations for these time steps (except of TAM for +240min). The interesting finding is that the FS (%) of the CMV method increases with decreasing CMF, namely the forecasting skill of CMV model is higher compared to persistence for

420 stations with higher cloudiness, demonstrating the applicability of the CMV forecasting method on GHI under cloudy conditions.

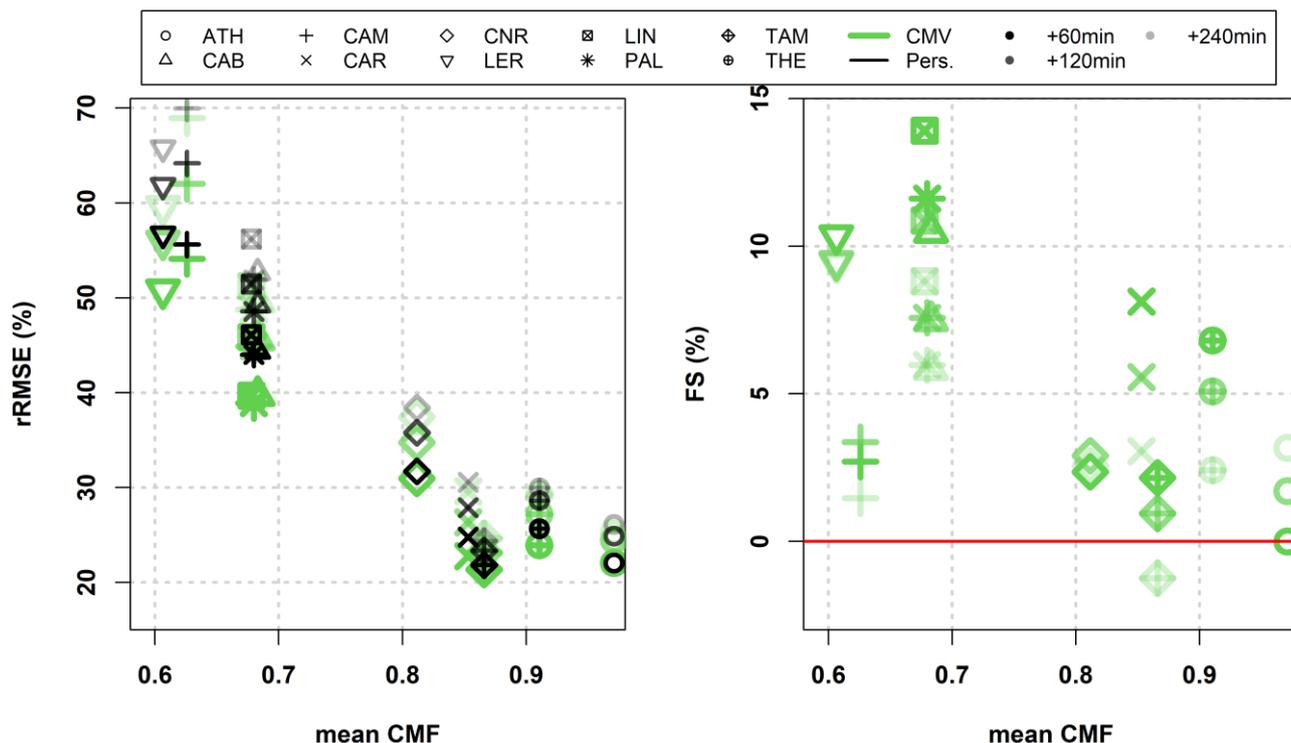



Figure 11 Mean bias error (MBE) and its relative values (rMBE%), relative root mean square error (rRMSE%) and forecasting skill expressed in percentage (FS%) of CMV model (green symbols) and persistence method (black symbols) forecasted global horizontal irradiance (GHI) versus the average cloudiness of stations (mean CMF) for the time steps +60, +120 and +240min (increasing transparency of symbols).

425

430

3.2.2 Performance for different cloudy conditions

To demonstrate the value of the CMV model against the persistence method that assumes the same cloudy conditions for all future time steps, we compare their performance under different cloudy conditions and transitions in cloudiness. Figure 12 presents the RMSE for both models (CMV green points and persistence black points) and the CMV model FS% as a function

435   of CMF, for three time-steps (+60, +120 and +180 min). Persistence performs better than CMV model under clear sky conditions, namely CMF=1, for all times steps (as expected, as there is no change in cloudiness). This is also true for the CMF bin 0.9 only for time step +180 min and for CMF bin >1 for all time steps, a bin which contains mainly clear sky cases. For cloudy conditions, namely CMF<0.9, the CMV model outperforms persistence for all time steps (apart from +180 min time step and CMF bin 0.9). The cloudier the conditions (the smaller the CMF), the better the performance of CMV model and the

440   greater the CMV FS% (up to ~ 20% for time step +60 min). The FS of CMV model decreases slightly with forecasting horizon, however, for the maximum of the forecasting horizon (+180 min) remains quite high ~+10% for CMF bins <0.7.

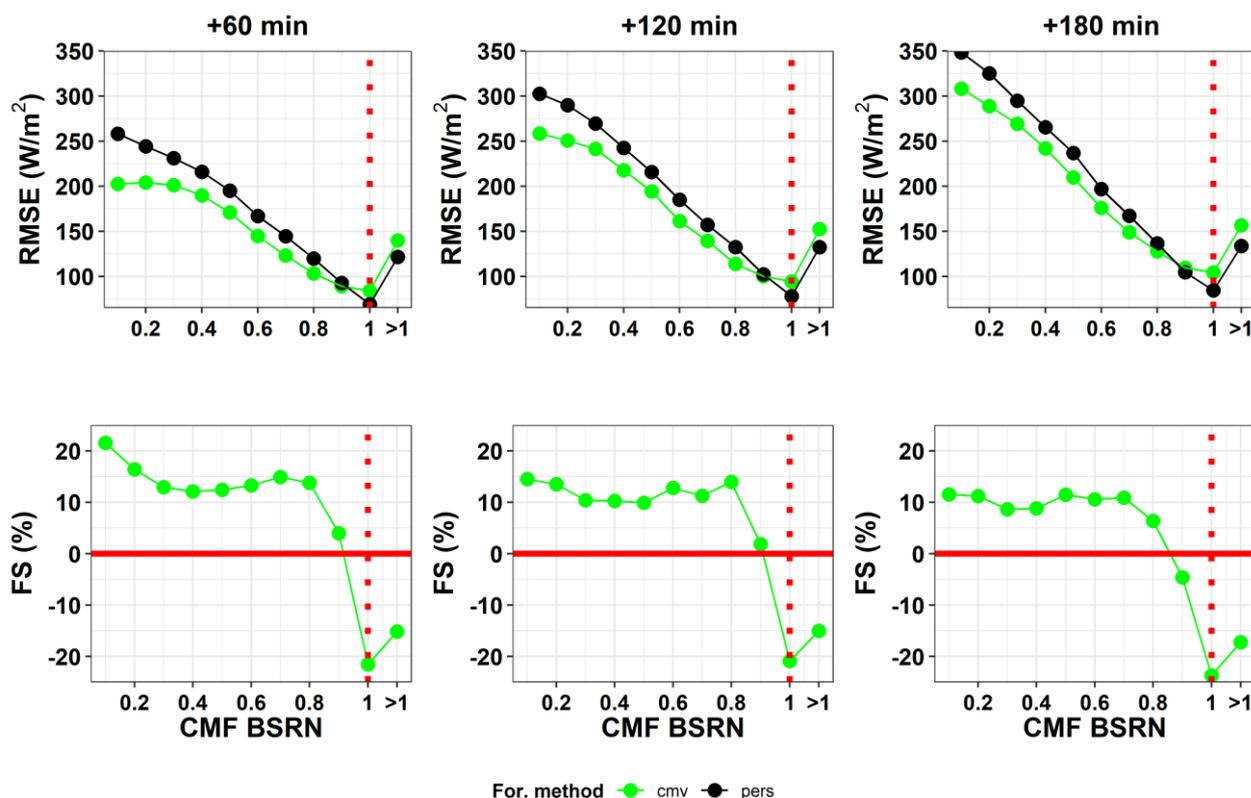



Figure 12 Upper panel: Root mean square error (RMSE) of forecasted global horizontal irradiances (GHI) for all stations with the CMV model (green symbols) and with the persistence method (black symbols) versus cloud modification factor (CMF) derived from
445   **GHI measurements, for 3 time steps (+60, +120 and +180min). Lower panel: The CMV model forecasting skill against CMF classes for the same time steps with upper panel.**

To demonstrate the better performance of CMV method compared to the persistence for all time steps under cloudy conditions, we calculated CMV FS% for partially cloudy conditions (CMF <0.9 and >0.4) and overcast conditions (CMF<0.4) and the results are presented in Fig. 13. We can see again that the FS of CMV model decreases with time, however, the minimum
450 value is ~10% for both categories. The maximum of FS is for both categories at +15 min time step at ~16% and ~22% for CMF<0.9&>0.4 (triangle symbols) and CMF<0.4 (cross symbols), respectively.

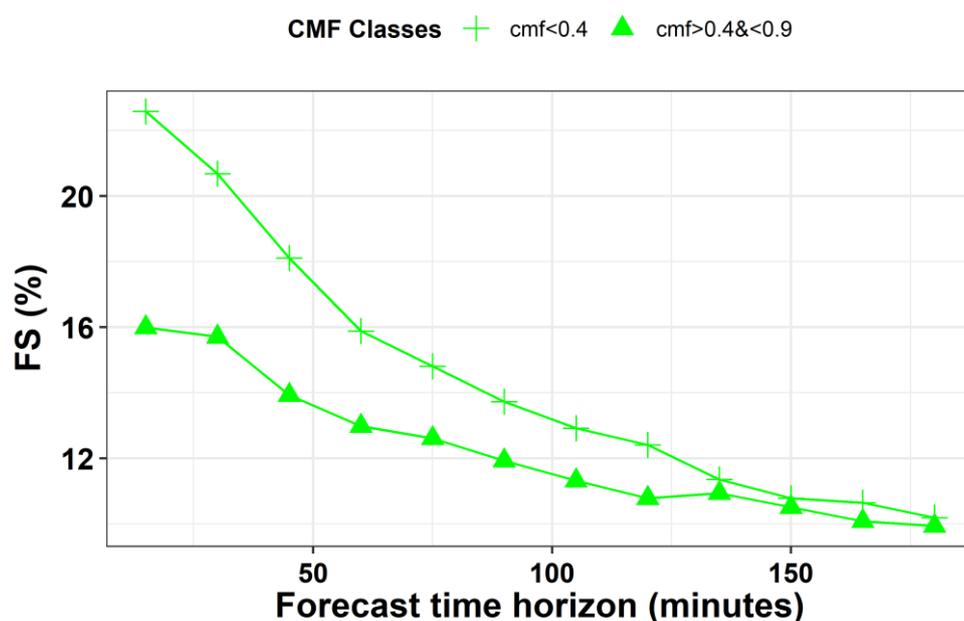



Figure 13 Forecasting skill (FS) expressed in percenateges of CMV model against the persistence method as a function of time horizon, from all stations, for two different cloudiness conditions: CMF >0.4 and <0.9 (triangles) and CMF <0.4 (crosses).

455 The performance of the CMV model against the persistence method was assessed also for changing cloudiness, in terms of CMF changes (from ground-based measurements). The CMF changes were calculated for a time interval of 60min (as $\Delta CMF=CMF_{t+60} - CMF_t$), and the results of the CMV model FS (%) are presented for the +60min time step, as a function of CMF changes in Fig. 14. The high negative value of FS for zero CMF changes bin indicates that the persistence method is better for that bin, which was anticipated, since we have zero or almost zero changes of CMF, which practically is the
460 persistence method definition. Persistence is still better than CMV for CMF changes from cloudy to clearer conditions, up to the +0.3 CMF change bin, but the FS is less negative than the zero bin. For CMF changes with higher magnitudes (bins > 0.4) from cloudy to clearer conditions, CMV is better that persistence, with FS values 15% for the CMF change bin +0.6. Consistent results were found for the opposite situation, namely from clearer to cloudy conditions, with CMV model always being better that persistence, with FS values up to ~20%.

465 Our analysis for different cloudiness conditions, highlights the limited ability of the persistence method compared to the CMV based nextSENSE2 to accurately forecast GHI under cloudy conditions (CMF values<0.9) and to follow the transitions in cloudiness (especially from clearer to cloudy conditions).

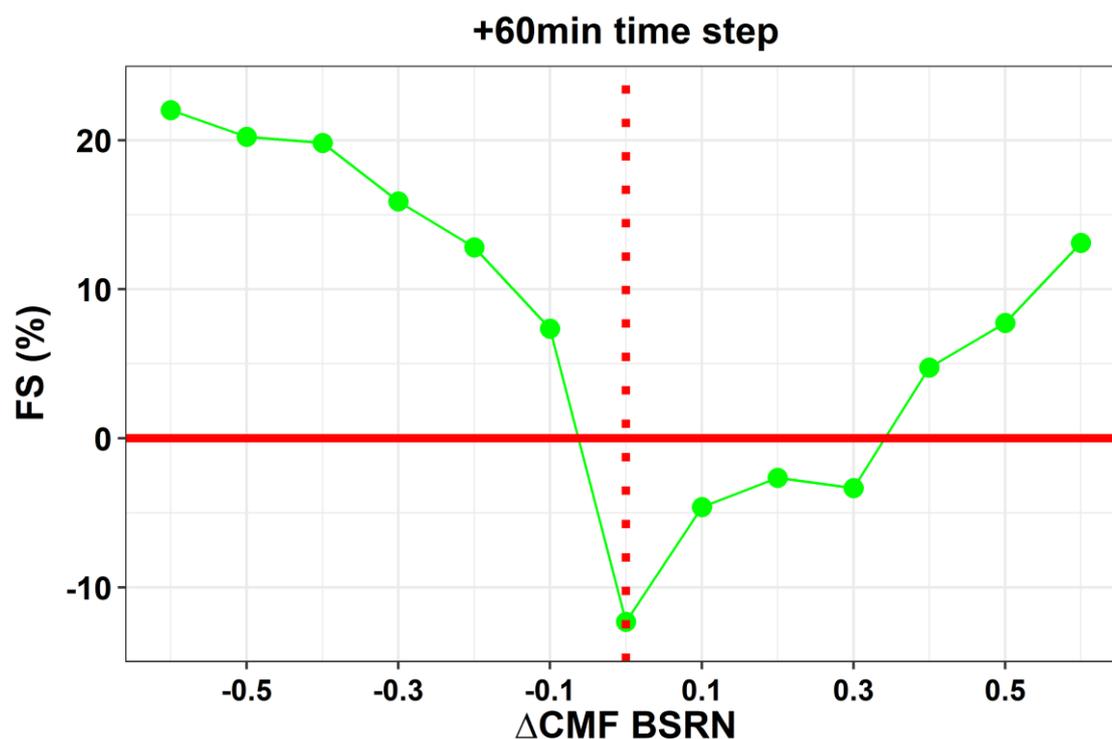



Figure 14 Forecasting skill (FS) expressed in percenateges of CMV model against the persistence method from all stations as a
470 **function of CMF changes within 60 min time interval (from time 0 to +60 min time step).**

4 Summary and conclusions

Our motivation is the continuous improvement of the EO based estimates and the accuracy of short-term forecasts of available solar resources to support solar energy exploitation systems on a regional scale (Europe and MENA region). In this study, we improve the SENSE/nextSENSE nowcasting/short-term forecasting operational systems, and analyze in detail the cloud related

475 uncertainties, discriminating also situations based on sun visibility, using ground-based measurements.

In terms of the aerosol related inputs, the slight overestimation of CAMS AOD that was found against the AERONET retrievals (<10%) resulted to SENSE2 clear-sky GHI underestimation lower than 1%, highlighting the applicability of CAMS forecasts as EO inputs for operational solar resources nowcasting. In terms of modeled all skies GHI, it was found that clouds are the main source of uncertainty. The SENSE2 mostly overestimates GHI with MBE 23.8 W/m² (4.9%), which was attributed to the

480 MSG cloud underestimation (overestimation of CMFmsg by ~0.17). We demonstrated that the most difficult situations to be modeled are related to high spatial variability of solar radiation within the satellite pixel due to clouds (e.g., the sun obscurity,

an information not possible to be derived from satellite data). Based on our cloud related analysis using ground-based data, a correction for the modelled GHI was used, resulting to an overall improvement of the SENSE2 modelled GHI with 61% of the cases within +/-50 W/m² (+/-10%) of measured GHI and a final MBE of SENSE2 11.3 W/m² (2.3%). Our main analysis was

485   based on the 15 min time scale, however based on the application hourly, daily or monthly data could be used. The daily and monthly SENSE2 GHI showed much better statistics (MBE 6.6 W/m² and 5.7 W/m², respectively). NextSENSE2 was also improved due to the SENSE improvements. We also show that compared to the persistence method, the model works much better (as expected) at locations with increased cloudiness and for frequent cloudiness changes.

  The data and methods involved for the estimation/prediction of the GHI in this study also reveal their limitations. The pixel-

490   based approach of the model inputs (satellite and models) could not always reflect the reality above a (point) ground-based station. However, the model inputs are the state of the art of EO data and can be readily available in regional or global scale, at high spatial and temporal resolution, hence the GHI product is representative of an area (~5km x 5km in this model), which is useful for PV parks covering a wide area. In general, evaluating the performance of such EO based GHI models with ground-based measurements must account for these comparison spatial representativity issues. The optical flow algorithm for

495   calculating CMVs is also based on assumptions like 2D clouds and brightness constancy. However, it is a method based on cloud inputs from satellite data in real time and the applicability of those methods is demonstrated here compared with the persistence approach.

  Since satellite cloud information is the only real time input, a new straightforward configuration for estimating GHI was applied (SENSE2). The advantage of calculating clear sky GHI from the previous day, is what increases the accuracy of this

500   product, since it is based on a detailed LUT of ~16M combinations of 7 different inputs, considering apart from AOD, additional aerosol optical properties and atmosphere/surface state inputs. Thus, the uncertainties in the estimated clear sky GHI practically result only from uncertainties in the inputs. The new scheme of calculating all skies GHI by multiplying the clear sky GHI with CMFmsg (derived in real time by multiparametric function of MSG COT and SZA) was improved by applying a suitable CMFmsg correction. The correction was successful and improved the model performance, especially for areas with

505   high cloudiness. Additionally, the new configuration of the SENSE2 is more flexible, and it is easy to adapt and provide more products like DNI, UV index, PAR, etc. which is one of the prospects for the new model or run in a retrospective way using reanalysis data or in situ observational data.

  According to the results, high resolution (every 15 min, at ~5km x 5km) and quite accurate GHI real time estimates/forecasts are produced from the upgraded SENSE2/NextSENSE2 operational systems that can contribute to solar energy systems

510   management and planning.

Appendix A

To see if the CMF differences (MSG modelled against measured) changing with SZA, the MBE of CMF was calculated for bins of SZA every 10 degrees. The observed CMF was considered the one derived from GHI measurements (Eq. 7) and the

515   modelled one derived by the Eq. 2. The results are presented in Fig. A1, for all cases and under different cloudiness conditions,

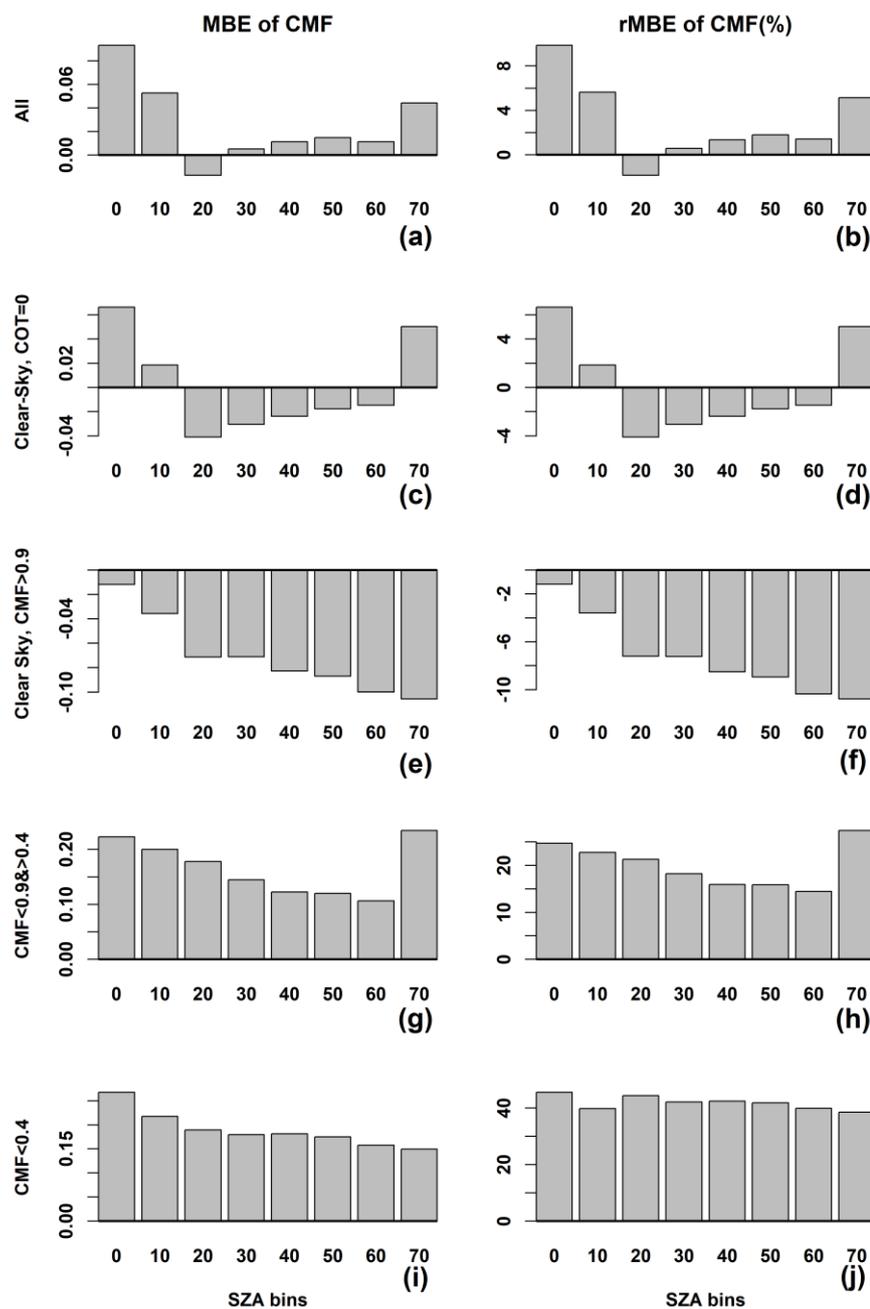
along with the relative values of CMF MBE expressed in percentages. We can see again the fact that the main overestimation of CMF values by MSG COT comes from the cloudy conditions (CMF <0.9). Specifically, for the partially cloudy conditions (CMF <0.9 and >0.4) the MBE reach values up to ~ 0.20 and for overcast skies (CMF <0.4) there are SZA bins (0 and 70 degrees) that the MBE reach values up to 0.4. However, for those two categories the MBE hardly changes with SZA.

520

Figure A 1 Cloud modification factor (CMF) mean bias error (MBE – left column) and relative MBE (% - right column) as a function of solar zenith angle (SZA) under all cases and under different cloudiness conditions.

Author contributions.

525 Idea and initialization, KP and SK; Model parameterization, IF, KP, IPR; resources, AFB, BEP, IPR, CK; data provision and curation, AFB and BEP; cloud function approach, NP and AK; overview and revision, CK, MH, SK; 1st draft writing, visualization, analysis and interpretation, KP; writing, review and editing, all authors. All authors gave final approval for publication.

530 **Acknowledgments.**

We would like to thank the 8 site instrument operators and technical staff of the BSRN network stations who made the ground-based measurements feasible. We acknowledge the principal investigators and co-investigators and their staff for establishing and maintaining the 10 AERONET sites used in this study. This study contains modified Copernicus Atmosphere Monitoring Service information [2023], and neither the European Commission nor ECMWF is responsible for any use that may be made

535 of the Copernicus information or data it contains. We acknowledge the Eumetsat SAFNWC services as well as the OMI team for providing all the necessary data used in this study. We acknowledge the free use of the GOME-2 surface LER database provided through the AC SAF of EUMETSAT. The GOME-2 surface LER database was created by the Royal Netherlands Meteorological Institute (KNMI). This work was supported by computational time granted from the National Infrastructures for Research and Technology S.A. (GRNET) in the National HPC facility - ARIS - under project ID pa210301- SO-LISIS.

540 CK, SK, IF and KP would like to acknowledge the European Commission project EuroGEO e-shape (GA no. 820852). AK would like to acknowledge the co-financing by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T1EDK - 00681). SK, IF and KP would like to acknowledge the COST Action HARMONIA, CA21119, supported by COST (European Cooperation in Science and Technology).

545

Financial support.

This research has been supported by the European Commission project 'EXCELSIOR': ERATOSTHENES: Excellence Research Centre for Earth Surveillance and Space-Based Monitoring of the Environment (grant no. 857510) and the European Commission project EuroGEO e-shape (GA no. 820852).

550 **References**

Anderson, G. P., Clough, S. A., Kneizys, F. X., Chetwynd, J. H., & Shettle, E. P. (1986). *AFGL atmospheric constituent profiles (0.120 km)*. AIR FORCE GEOPHYSICS LAB HANSCOM AFB MA.

Arvizu, D., Balaya, P., Cabeza, L., Hollands, T., Kondo, M., Konseibo, C., Meleshko, V., Stein, W. H., Tamaura, Y., & Xu, H. (2011). *Direct solar energy*.

555   Bhartia, P. K. (2012). OMI/Aura TOMS-Like Ozone, Aerosol Index, Cloud Radiance Fraction L3 1 day 1 degree x 1 degree V3, NASA Goddard Space Flight Center, Goddard Earth Sciences Data and Information Services Center (GES DISC). *Goddard Earth Sciences Data and Information Services Center (GES DISC)*, *10*.

Buras, R., Dowling, T., & Emde, C. (2011). New secondary-scattering correction in DISORT with increased efficiency for forward scattering. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *112*(12), 2028–2034.
560   https://doi.org/10.1016/j.jqsrt.2011.03.019

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., … König-Langlo, G. (2018). Baseline Surface Radiation Network (BSRN): Structure and data description (1992-2017). *Earth System Science Data*, *10*(3), 1491–1501. https://doi.org/10.5194/essd-10-1491-2018

565   Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Seyboth, K., Matschoss, P., Kadner, S., Zwickel, T., Eickemeier, P., Hansen, G., & Schlömer, S. (2011). IPCC special report on renewable energy sources and climate change mitigation. *Prepared By Working Group III of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK*.

Emde, C., Buras-Schnell, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., Kylling, J., Richter, B., Pause, C., Dowling,
570   T., & Bugliaro, L. (2016). The libRadtran software package for radiative transfer calculations (version 2.0.1). *Geoscientific Model Development*, *9*(5), 1647–1672. https://doi.org/10.5194/gmd-9-1647-2016

Farnebäck, G. (2003). Two-Frame Motion Estimation Based on Polynomial Expansion. In *Lecture Notes in Computer Science* (Vol. 2749, Issue 1, pp. 363–370). https://doi.org/10.1007/3-540-45103-X_50

Fountoulakis, I., Kosmopoulos, P., Papachristopoulou, K., Raptis, I. P., Mamouri, R. E., Nisantzi, A., Gkikas, A., Witthuhn,
575   J., Bley, S., Moustaka, A., Buehl, J., Seifert, P., Hadjimitsis, D. G., Kontoes, C., & Kazadzis, S. (2021). Effects of aerosols and clouds on the levels of surface solar radiation and solar energy in cyprus. *Remote Sensing*, *13*(12), 1–26. https://doi.org/10.3390/rs13122319

Garniwa, P. M. P., Rajagukguk, R. A., Kamil, R., & Lee, H. J. (2023). Intraday forecast of global horizontal irradiance using optical flow method and long short-term memory model. *Solar Energy*, *252*, 234–251.
580   https://doi.org/10.1016/j.solener.2023.01.037

Gasteiger, J., Emde, C., Mayer, B., Buras, R., Buehler, S. A., & Lemke, O. (2014). Representative wavelengths absorption parameterization applied to satellite channels and spectral bands. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *148*, 99–115. https://doi.org/10.1016/j.jqsrt.2014.06.024

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R.,
585   Campbell, J. R., Welton, E. J., Korkin, S. V., & Lyapustin, A. I. (2019). Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. *Atmospheric Measurement Techniques*, *12*(1), 169–209. https://doi.org/10.5194/amt-12-169-2019

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., & Smirnov, A. (1998). AERONET - A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, *66*(1), 1–16. https://doi.org/10.1016/S0034-4257(98)00031-5

IEA. (2022). *Renewables 2022*. https://www.iea.org/reports/renewables-2022, License: CC BY 4.0

Ineichen, P., & Perez, R. (2002). *A NEW AIRMASS INDEPENDENT FORMULATION FOR THE LINKE TURBIDITY COEFFICIENT* (Vol. 73, Issue 3). www.elsevier.com/locate/solener

Inness, A., Ades, M., Agustí-Panareda, A., Barr, J., Benedictow, A., Blechschmidt, A. M., Jose Dominguez, J., Engelen, R., Eskes, H., Flemming, J., Huijnen, V., Jones, L., Kipling, Z., Massart, S., Parrington, M., Peuch, V. H., Razinger, M., Remy, S., Schulz, M., & Suttie, M. (2019). The CAMS reanalysis of atmospheric composition. *Atmospheric Chemistry and Physics*, *19*(6), 3515–3556. https://doi.org/10.5194/acp-19-3515-2019

Kallio-Myers, V., Riihelä, A., Lahtinen, P., & Lindfors, A. (2020). Global horizontal irradiance forecast for Finland based on geostationary weather satellite data. *Solar Energy*, *198*, 68–80. https://doi.org/10.1016/j.solener.2020.01.008

Kinne, S. (2019). The MACv2 aerosol climatology. *Tellus, Series B: Chemical and Physical Meteorology*, *71*(1), 1–21. https://doi.org/10.1080/16000889.2019.1623639

Kosmopoulos, P. G., Kazadzis, S., Schmalwieser, A. W., Raptis, P. I., Papachristopoulou, K., Fountoulakis, I., Masoom, A., Bais, A. F., Bilbao, J., Blumthaler, M., Kreuter, A., Siani, A. M., Eleftheratos, K., Topaloglou, C., Gröbner, J., Johnsen, B., Svendby, T. M., Vilaplana, J. M., Doppler, L., … Kontoes, C. (2021). Real-time UV index retrieval in Europe using Earth observation-based techniques: System description and quality assessment. *Atmospheric Measurement Techniques*, *14*(8), 5657–5699. https://doi.org/10.5194/amt-14-5657-2021

Kosmopoulos, P. G., Kazadzis, S., Taylor, M., Raptis, P. I., Keramitsoglou, I., Kiranoudis, C., & Bais, A. F. (2018). Assessment of surface solar irradiance derived from real-time modelling techniques and verification with ground-based measurements. *Atmospheric Measurement Techniques*, *11*(2), 907–924. https://doi.org/10.5194/amt-11-907-2018

Kosmopoulos, P., Kouroutsidis, D., Papachristopoulou, K., Raptis, P. I., Masoom, A., Saint-Drenan, Y. M., Blanc, P., Kontoes, C., & Kazadzis, S. (2020). Short-term forecasting of large-scale clouds impact on downwelling surface solar irradiation. *Energies*, *13*(24). https://doi.org/10.3390/en13246555

Kurucz, R. L. (1994). Synthetic Infrared Spectra. *Symposium - International Astronomical Union*, *154*(13), 523–531. https://doi.org/10.1017/S0074180900124805

Long, C. N., & Dutton, E. G. (2010). *BSRN Global Network recommended QC tests, V2. x.*

Mayer, B., & Kylling, A. (2005). Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use. *Atmos. Chem. Phys.*, *5*(7), 1855–1877. https://doi.org/10.5194/acp-5-1855-2005

Oumbe, A., Qu, Z., Blanc, P., Lefèvre, M., Wald, L., & Cros, S. (2014). Decoupling the effects of clear atmosphere and clouds to simplify calculations of the broadband solar irradiance at ground level. *Geoscientific Model Development*, *7*(4), 1661–1669. https://doi.org/10.5194/gmd-7-1661-2014

Papachristopoulou, K., Fountoulakis, I., Gkikas, A., Kosmopoulos, P. G., Nastos, P. T., Hatzaki, M., & Kazadzis, S. (2022). 15-Year Analysis of Direct Effects of Total and Dust Aerosols in Solar Radiation/Energy over the Mediterranean Basin. *Remote Sensing*, *14*(7). https://doi.org/10.3390/rs14071535

625 Pörtner, H.-O., Roberts, D. C., Poloczanska, E. S., Mintenbeck, K., Tignor, M., Alegría, A., Craig, M., Langsdorf, S., Löschke, S., & Möller, V. (2022). *IPCC, 2022: Summary for policymakers*.

Qu, Z., Oumbe, A., Blanc, P., Espinar, B., Gesell, G., Gschwind, B., Klüser, L., Lefèvre, M., Saboret, L., Schroedter-Homscheidt, M., & Wald, L. (2017). Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat-4 method. *Meteorologische Zeitschrift*, *26*(1), 33–57. https://doi.org/10.1127/metz/2016/0781

630 Sengupta, M., Habte, A., Wilbert, S., Gueymard, C., & Remund, J. (2021). *Best practices handbook for the collection and use of solar resource data for solar energy applications*. National Renewable Energy Lab.(NREL), Golden, CO (United States).

Shettle, E. P. (1989). Models of aerosols, clouds and precipitation for atmospheric propagation studies, paper presented at Conference on Atmospheric Propagation in the UV, Visible, IR and MM-Region and Related System Aspects, NATO
635 Adv. Group for Aerosp. *Res. and Dev., Copenhagen*.

Tilstra, L. G., Tuinder, O. N. E., Wang, P., & Stammes, P. (2017). Surface reflectivity climatologies from UV to NIR determined fromEarth observations by GOME-2 and SCIAMACHY. *Journal of Geophysical Research*, *122*(7), 4084–4111. https://doi.org/10.1002/2016JD025940

Tilstra, L. G., Tuinder, O. N. E., Wang, P., & Stammes, P. (2021). Directionally dependent Lambertian-equivalent reflectivity
640 (DLER) of the Earth's surface measured by the GOME-2 satellite instruments. *Atmospheric Measurement Techniques*, *14*(6), 4219–4238. https://doi.org/10.5194/amt-14-4219-2021

Yang, D. (2019). SolarData package update v1.1: R functions for easy access of Baseline Surface Radiation Network (BSRN). *Solar Energy*, *188*, 970–975. https://doi.org/10.1016/j.solener.2019.05.068

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
645 Meer, D., & Peters, I. M. (2022). A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. In *Renewable and Sustainable Energy Reviews* (Vol. 161). Elsevier Ltd. https://doi.org/10.1016/j.rser.2022.112348

650