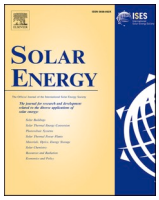
Solar Energy 269 (2024) 112363

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener



GISPLIT: High-performance global solar irradiance component-separation model dynamically constrained by 1-min sky conditions

José A. Ruiz-Arias [a,*], Christian A. Gueymard [b]

[a] Universidad de Málaga, Facultad de Ciencias, Física Aplicada I, Campus Teatinos s/n, 29071, Málaga, Spain
[b] Solar Consulting Services, Colebrook, NH 03576, USA

ARTICLE INFO

Keywords:
Solar irradiance
Components separation
Sky conditions
Direct irradiance
Diffuse irradiance

ABSTRACT

The separation of global horizontal irradiance (GHI) into its direct and diffuse components is necessary in a variety of applications, most specially in solar energy utilization, where knowledge of direct normal irradiance (DNI) is of paramount importance. Here a novel and efficient model, referred to as GISPLIT, is presented to perform this task accurately, using time series of measured data at 1-min resolution. To better describe the radiative effects of different cloud situations, the model takes advantage of a preliminary classification of the sky conditions into six sky types. An empirical submodel is assigned to each sky class to split GHI into its components, using a limited number of predictors that are related to GHI's magnitude and variability, and to coincident estimates of the clear-sky irradiance components. Those submodels are trained and validated using rigorously quality-assessed measurements from 120 radiometric stations over all continents and all five major Köppen-Geiger (KG) climate classes, totaling ≈64 million valid data points. Four model versions are evaluated using training data for either all KG climate regions combined or conditioned by KG climate, and either with or without additional support from machine learning. The validation of the four versions suggests that the conditioning by KG climate does not add any significant benefit over the "all-climates" training approach and that, overall, the model version trained with data from all KG climates combined and supported by machine learning generally predicts DNI with the best RMSE results at unseen sites, although with little difference over the other versions.

1. Introduction

The use of "separation" or "decomposition" models to derive the direct and diffuse components from global horizontal irradiance (GHI) is ubiquitous in solar energy applications because various solar engineering calculation segments require the solar irradiance components separately. However, the observations of GHI alone are more common than those of its components, and the most widespread satellite-derived datasets can only obtain GHI estimates from satellite imagery [1]. Thus, most frequently, the direct and diffuse components must be derived using one of the many empirical decomposition models that have been proposed and thoroughly reviewed in the literature [2,3]. Such models typically attempt to characterize the relationship between the diffuse/global ratio, K (hereafter "diffuse fraction") and the clearness index, K_T (i.e., the global/extraterrestrial horizontal irradiance ratio).

The large uncertainty in this decomposition process makes the estimates of the direct normal irradiance (DNI) component much less reliable than those of the GHI input, conversely to what happens in solar radiation measurements [4,5]. However, precise knowledge of DNI is of the utmost importance for focusing technologies (concentrating solar power, CSP, or concentrating photovoltaics, CPV) [6]. Hence, the accuracy of the specific decomposition model used to obtain historical DNI data series becomes an issue to guarantee a good system design, and ultimately the project's bankability and viability.

The difficulty of characterizing the aerosol load over arid areas [7–10], amplified by the large sensitivity of DNI to aerosol optical properties [11,12], adds to the dominant impact of mispredicted cloud properties as the main causes for biased DNI predictions when using either sophisticated radiative transfer models or even simpler separation models. Finally, the calculation of the global irradiance that is incident on a tilted plane (GTI), such as the plane of array of solar collectors, requires the separate knowledge of the direct and diffuse components. These calculations are made by solar engineers on a daily basis, who must therefore rely on an efficient method for the component separation process. This essentially rules out the use of sophisticated radiative transfer models because of their typical reliance on supercomputers and advanced input data. In addition, various studies, e.g., [4,13–15] have

* Corresponding author.
E-mail address: jararias@uma.es (J.A. Ruiz-Arias).

https://doi.org/10.1016/j.solener.2024.112363
Received 26 October 2023; Received in revised form 18 December 2023; Accepted 18 January 2024
Available online 31 January 2024

Nomenclature

Abbreviations

BSRN	Baseline Surface Radiation Network
CAELUS	Classification Algorithm for the Evaluation of cLoUdiness Situations
DIF	All-sky diffuse horizontal irradiance
DNI	All-sky direct normal irradiance
GHI	All-sky global horizontal irradiance
GISPLIT	Global Irradiance Splitting
GTI	All-sky global tilted irradiance
KG	Köppen-Geiger
PVPS	Photovoltaic Power Systems
SZA	Solar zenith angle

Symbols

D_h	All-sky diffuse horizontal irradiance
D_{hc}	Clear-sky diffuse horizontal irradiance
E_{0h}	Extraterrestrial horizontal irradiance
G_h	All-sky global horizontal irradiance
G_{hc}	Clear-sky global horizontal irradiance
K	Diffuse fraction
K_{cs}	Clear-sky index
K_{de}	Cloud-enhancement index
K_{ds}	Clear-sky diffuse fraction
K_T	Clearness index
m_r	Relative air mass
Z	Solar zenith angle

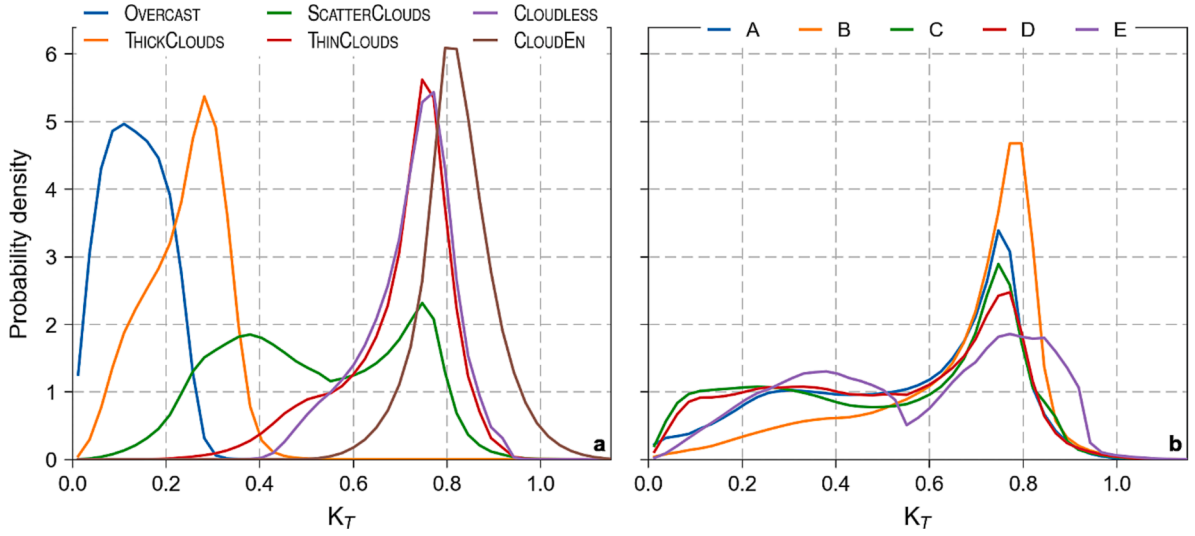


Fig. 1. Probability density distribution of observed K_T according to CAELUS sky types (a) and KG major climates (b). The observational dataset is made up of ≈50 million samples from 54 radiometric stations over the five primary KG climate zones. A detailed description of the dataset is provided in [30].

shown that the uncertainty in GTI, and thus in the power produced by tilted solar collectors, is a strong function of the errors introduced in the separation process, which is why innovative methods are needed to make it more accurate.

Considering the direct impact of cloudiness on the magnitude and variability of the diffuse fraction, it is logical to include as much cloud information as possible to improve that modeling. In this sense, it has been recently proposed to take each site's climatological characteristics into consideration with the help of the Köppen-Geiger (KG) climatic classification [16–19]. However, that classification is based on *seasonal* values of temperature and precipitation, which are only loosely related to the *real-time* cloud field optical characteristics, to the point that the marginal improvements hypothetically associated with this climatic stratification have been recently questioned [3]. Considering the present focus on modeling DNI at 1-min resolution, a main hypothesis in the present study is that a pure climate-specific approach is not sufficient, especially if the classification scheme is only indirectly related to cloud properties, such as the KG classification. At such high temporal resolution, it is obvious that the diffuse fraction varies considerably when the sky conditions change from cloudless to overcast, for instance. Even under purely cloudless conditions, the diffuse fraction—just like the direct/diffuse ratio—is a strong function of atmospheric turbidity [20–22], which thus needs to be accounted for. At the other extreme, the cloud enhancement phenomenon is now known to be relatively frequent [23–25] and to have a substantial impact on the determination of the diffuse fraction [26–28].

To overcome these inherent difficulties, the present approach exploits the strengths of CAELUS [29], a recent algorithm that can classify the 1-min sky conditions into six variability classes just from GHI observations, solar position, and clear-sky solar irradiance model estimates. One major finding of that study was that the sorting of K_T into sky classes is more discriminant than if rather conditioned by climatic classes. That conclusion was based on the fact that the K_T distributions based on sky classes show less overlap than those based on climate classes (Fig. 1).

In parallel—and alternatively to the conventional modeling approach for GHI separation, in which an empirical form for the K-K_T relationship is prescribed—machine-learning techniques have become an efficient modeling method that deserves careful consideration [31,32].

The innovative approach that is presented here to decompose GHI into its two components, hereinafter referred to as Global Irradiance Splitting (GISPLIT), conditions the GHI separation by CAELUS sky type, leaving the climate-conditioned separation described above as a supplementary option. The formulation of GISPLIT is done such that it can further combine conventional and machine-learning techniques to perform the separation. Hence, various versions of GISPLIT are developed to elucidate the improvement potential of 1) the climate-based

Table 1
Information on the 54 BSRN stations used in this work from the CAELUS database, sorted by Köppen-Geiger climate class. The "Acron." column indicates the site acronym according to BSRN. The Lat. and Lon. columns provide the station's latitude and longitude, respectively (in degrees); the Elev. column provides the site's elevation in meters above mean sea level; finally, the data period used for each station appears in the column labelled "Period".

Acron.	Station	Lat.	Lon.	Elev.	Period	Acron.	Station	Lat.	Lon.	Elev.	Period
Köppen-Geiger climate A stations:											
BER	Bermuda	32.3008	–64.766	8	2008–2012	KWA	Kwajalein	8.72	167.731	10	2012–2016
COC	Cocos Island	–12.193	96.835	6	2014–2018	MAN	Momote	–2.058	147.425	6	2008–2012
DAR	Darwin	–12.425	130.891	30	2010–2014	MNM	Minamitorishi.	24.2883	153.9833	7	2015–2019
DWN	Darwin Met Off.	–12.424	130.8925	32	2014–2018	NAU	Nauru Island	–0.521	166.9167	7	2008–2012
ISH	Ishigakijima	24.3367	124.1644	6	2014–2018	TIR	Tiruvallur	13.0923	79.9738	36	2014–2018
Köppen-Geiger climate B stations:											
ASP	Alice Springs	–23.798	133.888	547	2014–2018	GOB	Gobabeb	–23.5614	15.042	407	2014–2018
BOS	Boulder	40.125	–105.237	1689	2014–2018	PTR	Petrolina	–9.068	–40.319	387	2014–2017
DAA	De Aar	–30.6667	23.993	1287	2003–2004 2016–2018	SBO	Sede Boqer	30.8597	34.7794	500	2007–2011
DRA	Desert Rock	36.626	–116.018	1007	2014–2018	SOV	Solar Village	24.91	46.41	650	1999–2002
FPE	Fort Peck	48.3167	–105.1	634	2014–2018	TAM	Tamanrasset	22.7903	5.5292	1385	2014–2018
Köppen-Geiger climate C stations:											
BIL	Billings	36.605	–97.516	317	2013–2017	GCR	Goodwin Creek	34.2547	–89.8729	98	2014–2018
CAB	Cabauw	51.9711	4.9267	0	2014–2018	IZA	Izaña	28.3093	–16.4993	2373	2014–2018
CAM	Camborne	50.2167	–5.3167	88	2012–2016	LAU	Lauder	–45.045	169.689	350	2014–2018
CAR	Carpentras	44.083	5.059	100	2014–2018	LER	Lerwick	60.1389	–1.1847	80	2012–2016
CLH	Chesapeake L.	36.905	–75.713	37	2011–2015	LRC	Langley Res. C.	37.1038	–76.3872	3	2015–2019
CNR	CENER	42.816	–1.601	471	2014–2018	PAL	Palaiseau	48.713	2.208	156	2014–2018
E13	South. Gr. Plains	36.605	–97.485	318	2013–2017	PAY	Payerne	46.815	6.944	491	2014–2018
FLO	Florianopolis	–27.6047	–48.5227	11	2014–2015 2017–2019	SMS	São Martinho da Serra	–29.4428	–53.8231	489	2012–2016
FUA	Fukuoka	33.5822	130.3764	3	2014–2018	TAT	Tateno	36.0581	140.1258	25	2014–2018
Köppen-Geiger climate D stations:											
BON	Bondville	40.0667	–88.3667	213	2014–2018	SAP	Sapporo	43.06	141.3286	17	2015–2019
LIN	Lindenberg	52.21	14.122	125	2013–2017	SXF	Sioux Falls	43.73	–96.62	473	2014–2018
PSU	Rock Springs	40.72	–77.9333	376	2014–2018	TOR	Toravere	58.254	26.462	70	2014–2018
REG	Regina	50.205	–104.713	578	2007–2011	XIA	Xianghe	39.754	116.962	32	2007–2011
Köppen-Geiger climate E stations:											
ALE	Alert	82.49	–62.42	127	2009–2013	NYA	Ny-Ålesund	78.925	11.93	11	2014–2018
DOM	Concordia St.	–75.1	123.383	3233	2014–2018	SON	Sonnblick	47.054	12.9577	3109	2014–2018
EUR	Eureka	79.989	–85.9404	85	2007–2011	SYO	Syowa	–69.005	39.589	18	2014–2018
GVN	Georg von Neu.	–70.65	–8.25	42	2014–2018	TIK	Tiksi	71.5862	128.9188	48	2013–2017

conditioning, in addition to the built-in sky-type conditioning in GISPLIT, and 2) enhanced modeling based on machine-learning techniques, exemplified here by the partial use of extreme gradient boosting [33] to improve the separation process in the case of challenging sky situations (see Section 4). These two developments constitute secondary research questions of this work.

The whole study is based on a remarkably large solar irradiance database that extends worldwide and includes a record number of 120 ground stations and ≈64 million quality-assured data points (Section 2). The methodology is described in Sections 3 and 4. It differs substantially from all previous separation models and results in landmark progress in terms of both accuracy and universality, while maintaining a desirable 1-min temporal resolution. Validation results against ground observations are detailed in Section 5 before conclusions are reached in Section 6.

2. Solar irradiance data

Two different databases of 1-min solar irradiance ground observations are considered here. One of them is the quality-assured CAELUS database [30], consisting of 54 Baseline Surface Radiation Network (BSRN) site stations [34] with 5 years of data per site, except for Petrolina (Brazil) and Solar Village (Saudi Arabia), where only 4 years of data were available (Table 1). The second database, which includes 66 radiometric stations with 1 year of data per station (Table 2), is constructed from multiple public and private networks. It was put together through a collaborative effort that included these authors and other colleagues of the Solar Resource for High Penetration and Large-Scale Applications[1] [35] working group of the International Energy Agency's Photovoltaic Power Systems (PVPS) Programme Task 16. In what follows, this dataset is referred to as the PVPS database, part of which is published [35]. This database was subjected to exactly the same data treatment protocol as the CAELUS database for maximum quality. After stringent quality control and elimination of low-sun data points (solar zenith angle $> 85°$), a total of ≈50 million valid data points remain for analysis with the CAELUS database, and ≈14 million valid data points with the PVPS database. Fig. 2 shows the spatial distribution of the two databases' stations.

In terms of instrumentation, some differences exist between the two databases. For CAELUS, all stations monitor the three irradiance components with independent thermopile radiometers. In the case of the PVPS database, however, there are some stations that either just measure GHI and one of its components, or use a rotating shadowband irradiometer. Wherever the diffuse horizontal irradiance (DIF) is measured, the radiometer is properly blocked from the direct sun by a tracking shade.

Overall, the BSRN observations are expected to have somewhat lower uncertainty, which is why they have been used to train the separation model (Section 3).

The primary KG climate at the location of each radiometric station is used here to group the stations by climate conditions, which is required by two training versions of the separation model (Section 4) and in the

[1] https://iea-pvps.org/research-tasks/solar-resource-for-high-penetration-and-large-scale-applications/.

Table 2
Information on the 66 stations included in the PVPS database. Classification and columns are as in Table 1.

Acron.	Station (Network)	Lat.	Lon.	Elev.	Period	Acron.	Station (Network)	Lat.	Lon.	Elev.	Period
Köppen-Geiger climate A stations:											
CHI	Chileka, MW (1)	–15.6798	34.9723	767	2017	KDO	Kadhdhoo, MV (1)	1.8599	73.5203	0	2016
DAR	Dar Es Salaam, TZ (1)	–6.7811	39.2039	122	2016	MIY	Miyako, JP (2)	24.737	125.327	50	2017
FEN	Feni, BD (1)	22.8003	91.3582	5	2018	MLE	Male, MV (1)	4.1927	73.5281	–8	2016
GAN	Gan, MV (1)	–0.6906	73.15	3	2016	TOW	Townsville, AU (3)	–19.2483	146.7661	4	2018
HAN	Hanimaadhoo, MV (1)	6.7482	73.1696	2	2016						
Köppen-Geiger climate B stations:											
ADE	Adelaide Airp., AU (3)	–34.9524	138.5196	2	2015	MIS	Missour, MA (4)	32.8603	–4.1072	1107	2015
ADR	Adrar, DZ (4)	27.8783	–0.27	262	2015	OUJ	Univ. Oujda, MA (4)	34.6497	–1.9003	617	2017
ALB	Albuquerque, US (5)	35.038	–106.6221	1617	2018	PAN	Univ. Texas, US (7)	26.3059	–98.1716	45	2017
ALE	GIZ Richtersv., ZA (6)	–28.5608	16.7615	141	2017	POR	N. Mand. Univ., ZA (6)	–34.0086	25.6653	35	2016
ARI	Univ Arizona, US (7)	32.2297	–110.9553	786	2016	SAL	Salt Lake City, US (5)	40.7722	–111.9549	1288	2018
BRO	Broome Airp., AU (3)	–17.9475	122.2353	7	2015	SHA	Shagaya, KW (8)	29.2282	47.0624	0	2016
GAB	Gaborone, BW (6)	–24.661	25.934	1014	2015	SRR	NREL BMS, US (7)	39.742	–105.18	1829	2016
HAF	Hanford, US (5)	36.31	–119.63	73	2016	TAB	Plataf. Solar, ES (9)	37.0909	–2.3581	500	2016
JOR	Univ. Jordan, JO (4)	30.172	35.8183	1012	2015	TAT	Tataouin, TN (4)	32.9741	10.4851	210	2019
LAS	Univ. Nev.-LV, US (7)	36.107	–115.1425	615	2018	WAG	Wagga Wagga, AU (3)	–35.1583	147.4575	212	2018
LOY	Loy. Mar. Univ, US (7)	33.9667	–118.4226	27	2015						
Köppen-Geiger climate C stations:											
CAH	Cape Hedo, JP (2)	26.867	128.248	65	2018	LOC	Locarno, CH (12)	46.18	8.783	366	2015
CAP	Cape Grim, AU (3)	–40.6817	144.6892	95	2018	LUS	Lusaka, ZM (1)	–15.3946	28.3372	1262	2017
CAS	Casaccia-Rome, IT (10)	42.0417	12.3067	150	2018	MEL	Melbourne, AU (3)	–37.6655	144.8321	113	2018
CHB	Chiba, JP (2)	35.625	140.104	21	2018	MIL	Milan, IT (13)	45.4762	9.2546	150	2016
CHL	Chilanga, ZM (1)	–15.5483	28.2482	1226	2016	MUT	Mutanda, ZM (1)	–12.423	26.215	1318	2016
CHO	Choma, ZM (1)	–16.8383	27.0705	1284	2016	MZU	Mzuzu, MW (1)	–11.4199	33.9953	1285	2017
DUR	Durban, ZA (6)	–29.817	30.945	200	2016	ORE	Univ. Oregon, US (7)	44.0467	–123.0743	134	2017
FUU	Fukue, JP (2)	32.752	128.682	80	2015	PRE	Uni. Pretoria, ZA (6)	–25.7531	28.2286	1410	2015
HSU	C. P. Humboldt, US (7)	40.876	–124.08	36	2016	ROC	Rockhampton, AU (3)	–23.3753	150.4775	10	2015
JAE	Jaén, ES (11)	37.7878	–3.7781	459	2015	SEA	Seattle, US (5)	47.6868	–122.2567	20	2017
KAA	Kasama, ZM (1)	–10.1717	31.2256	1381	2017	STE	Stellenb. Univ., ZA (6)	–33.9281	18.8654	119	2016
KAO	Kaoma, ZM (1)	–14.839	24.931	1172	2017	STR	Sterling, US (5)	38.972	–77.4869	85	2019
KAS	Kasungu, MW (1)	–13.0153	33.4685	1065	2017	VUW	Vuwani, ZA (6)	–23.131	30.424	628	2018
KWA	KwaZulu, ZA (6)	–29.871	30.977	150	2016						
Köppen-Geiger climate D stations:											
BIS	Bismarck, US (5)	46.7718	–100.7596	503	2015	MAD	Madison, US (5)	43.0725	–89.4113	271	2016
DAO	Davos, CH (14)	46.813	9.844	1610	2019	NOR	Norrköping, SE (16)	58.583	16.152	43	2016
KAZ	Kanzelhöhe, AT (15)	46.6776	13.9024	1526	2016	ODE	Odeillo, FR (18)	42.494	2.029	1600	2019
KIR	Kiruna, SE (16)	67.842	20.41	424	2015	VIS	Visby, SE (16)	57.673	18.345	49	2015
LYN	Lyngby, DK (17)	55.7906	12.5251	40	2019						

(1) ESMAP: Energy Sector Management Assistance Program (https://www.esmap.org/).
(2) SKYNET: International Skynet Data Center (https://www.skynet–isdc.org/).
(3) BOM: Bureau of Meteorology, (https://www.bom.gov.au/).
(4) enerMENA: Solar radiation measurements in the MENA region [36].
(5) NOAA: National Oceanic and Atmospheric Administration, Global Monitoring Laboratory (https://gml.noaa.gov/grad).
(6) SAURAN: Southern African Universities Radiometric Network (https://sauran.ac.za/).
(7) NREL: National Renewable Energy Laboratory, Measurement and Instrumentation Data Center (https://midcdmz.nrel.gov/).
(8) KISR: Kuwait Institute for Scientific Research (https://www.kisr.edu.kw/en/).
(9) CIEMAT/PSA: Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (https://www.psa.es/en/index.php).
(10) ENEA: Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo económico sostenible (https://www.enea.it/en).
(11) UJAEN: University of Jaén, Spain (https://matras.ujaen.es/).
(12) MeteoSwiss: Federal Office of Meteorology and Climatology (https://www.meteoswiss.admin.ch/).
(13) RSE: Ricerca Sistema Energetico (https://www.rse-web.it/).
(14) PMOD: Physikalisch-Meteorologisches Observatorium Davos / World Radiation Data Center (https://www.pmodwrc.ch/en/home/).
(15) ZAMG: Zentralanstalt für Meteorologie und Geodynamik (https://www.zamg.ac.at/cms/en/).
(16) SMHI: Swedish Meteorological and Hydrological Institute (https://www.smhi.se/en).
(17) DTU: Technical University of Denmark (https://www.dtu.dk/english).
(18) CNRS: Centre National de la Recherche Scientifique (https://www.cnrs.fr/en).

validation analysis (Section 5). That classification has five primary levels (classes A, B, C, D, and E) and is determined here from a 30 arc-min resolution database [37].

The calculation of clear-sky solar irradiance, which is required by all the separation models included here, is as in Ruiz-Arias and Gueymard [29] using the SPARTA clear-sky solar irradiance model [38].

3. GHI separation modelling: New conceptual approach

The GHI separation with GISPLIT is conditioned by the CAELUS *sky types*, which divide the sky situations into six different variability classes. Broadly, these classes consider situations dominated mostly by overcast conditions (Overcast), thick, scattered, or thin clouds (ThickClouds, ScatterClouds or ThinClouds, respectively), cloudless skies (Cloudless), and cloud enhancements (CloudEn). This sky stratification is convenient from the standpoint of GHI separation models because the sky classes represent different situations for both K_T and K. Specifically, in the Overcast and ThickClouds classes, K_T is small and $K \approx 1$. Conversely, in the ThinClouds and Cloudless classes, K_T is large while K is small. Cloud enhancements, as represented in the CloudEn class, feature situations in which K_T is even greater than for cloudless skies, while K is not necessarily small. The remaining ScatterClouds class has the largest diversity in terms of K_T and K values, and expectedly is the most difficult to model.

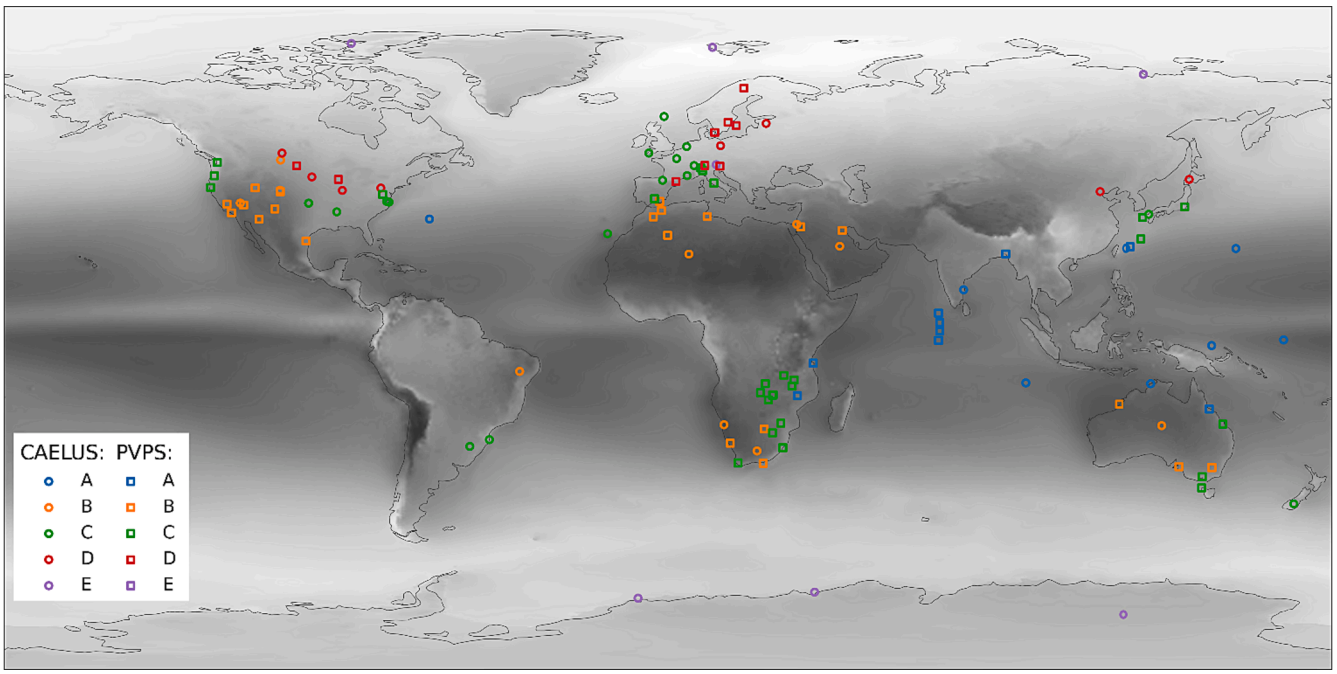


Fig. 2. Location of the radiometric stations for both the CAELUS database (circles) and the PVPS database (squares). The color of the markers indicates the primary KG climate at the station location. The background grayscale shows the mean GHI in the period 2006–2020 from the ECMWF's ERA5 reanalysis.

The predicted variable is K for all sky classes, except for Overcast and ThickClouds situations, for which DIF (hereinafter also referred to as D_h) is directly predicted from GHI (hereinafter also referred to as G_h) because these sky classes indicate situations with zero, or virtually zero, DNI, and thus $K \approx 1$.

In addition to the relative air mass—a readily available quantity—there are only three sequential irradiance inputs to the model: G_h; the modelled GHI under an ideal cloudless sky, G_{hc}; and the modelled DIF under the same ideal cloudless sky, D_{hc}. This is much less than for all separation models of the recent literature. Nevertheless, the necessary detection of variability patterns also requires that some of these inputs be also determined at a relaxed temporal resolution. Ultimately, the explanatory variables used in the sky-conditioned models are (i) the clearness index, $K_T = G_h/E_{0h}$, where E_{0h} is the extraterrestrial horizontal solar irradiance; (ii) a smoothed version of the clearness index, $\overline{K}_T = \overline{G}_h/E_{0h}$, where $\overline{G}_h$ is a centered 30-min moving average of G_h; (iii) the clear-sky index, $K_{cs} = G_h/G_{hc}$; (iv) the clear-sky diffuse fraction, $K_{ds} = D_{hc}/G_{hc}$; and (v) the cloud-enhancement index, $K_{de} = (G_h - G_{hc})/G_h$. All these explanatory variables are calculated with the same 1-min time grid as the observations. Of them, K_T and K_{cs} are mostly intended to account for atmospheric and cloud transmittances, respectively. These two indices describe such transmittances at 1-min time steps, when cloud enhancement can overwhelm solar irradiance variability. In contrast, $\overline{K}_T$ operates at the coarser time scale of 30-min because cloud enhancements are then smoothed out and solar irradiance variability is explained by other factors. Cloud enhancements are accounted for with the K_{de} index. Finally, K_{ds} provides information related to solar irradiance scattering in the cloudless atmosphere, and is thus helpful mostly for Cloudless and ScatterClouds sky types.

G_{hc} and D_{hc} must be obtained with a high-performance clear-sky solar radiation model, such as SPARTA [38], REST2 [39], or McClear [40], albeit only the first one has been tested so far in the present context. The inputs to the clear-sky solar radiation model are derived here from the NASA's MERRA-2 atmospheric reanalysis [41]. Since its predictions are provided with hourly temporal resolution and 0.5 × 0.625° spatial resolution, they are linearly interpolated, first to the location of each station, and then to the 1-min observational time grid.

MERRA-2 provides modeled atmospheric information since 1980, with a lag time of about two months, but other reanalysis products might be used as well, such as ECMWF's CAMS global reanalysis dataset[2] [42]. Furthermore, if the separation model needs to be operated for the recent past or near future, there are alternative data sources available, such as ECMWF's CAMS global atmospheric composition forecasts,[3] which are issued twice a day and provide hourly forecasts. Optionally, McClear[4] might also be appropriate for near-real-time data (but not for forecasts) because it provides data since 2004 with a lag time of two days.

Within each sky class, the separation of GHI with GISPLIT is modelled using various functional forms. In all cases, their coefficients are obtained by least-squares fitting to the 1-min observations, as follows:

- Overcast sky. Here, DNI is zero, or nearly zero, and thus D_h is expected to be $\approx G_h$. Hence, D_h is simply evaluated as a linear function of G_h:

$$D_h = a_0 + a_1 G_h \tag{1}$$

where the expected values of a_0 and a_1 are ≈ 0 and ≈ 1, respectively.

- ThickClouds sky. The situation is similar to the Overcast sky type. Therefore,

$$D_h = b_0 + b_1 G_h \tag{2}$$

- ScatterClouds sky. This sky type requires an elaborate treatment because it is associated with highly variable K_T and K values. The selected functional form is similar to that proposed by Engerer [43], but uses a different set of predictors:

[2] https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-reanalysis-eac4?tab=overview.

[3] https://ads.atmosphere.copernicus.eu/cdsapp#!/dataset/cams-global-atmospheric-composition-forecasts?tab=overview.

[4] https://www.soda-pro.com/web-services/radiation/cams-mcclear.

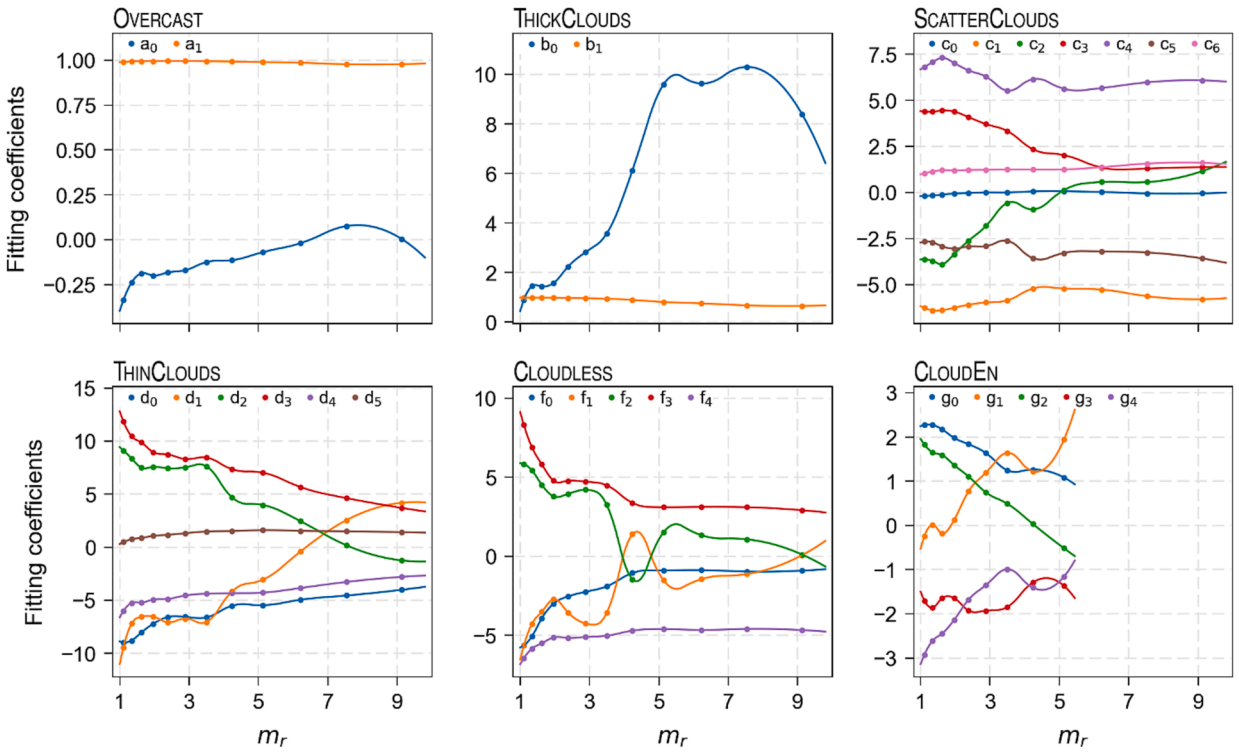


Fig. 3. Coefficients of the separation models fitted for each sky type for all climate-C sites of the training dataset combined. The markers are the fitting coefficients for each relative air mass interval. The lines are second-order interpolating splines of the coefficients.

$$K = c_0 + \frac{1 - c_0}{1 + \exp(c_1 + c_2 K_T + c_3 \overline{K}_T + c_4 K_{cs} + c_5 K_{ds})} + c_6 K_{de} \tag{3}$$

- ThinClouds sky. The functional form is similar to Eq. (3), with the same set of predictors, but now using $c_0 = 0$:

$$K = \frac{1}{1 + \exp(d_0 + d_1 K_T + d_2 \overline{K}_T + d_3 K_{cs} + d_4 K_{ds})} + d_5 K_{de} \tag{4}$$

- Cloudless sky. The functional form is similar to Eq. (4), but without the cloud-enhancement index term:

$$K = \frac{1}{1 + \exp(f_0 + f_1 K_T + f_2 \overline{K}_T + f_3 K_{cs} + f_4 K_{ds})} \tag{5}$$

- CloudEn sky. The function here is the same as Eq. (5):

$$K = \frac{1}{1 + \exp(g_0 + g_1 K_T + g_2 \overline{K}_T + g_3 K_{cs} + g_4 K_{ds})} \tag{6}$$

The numerical values of all coefficients are not fixed, but depend on sun position, as discussed below. The submodels above [Eqs. (1)–(6)] have been developed as the result of a systematic trial-and-error process. It involved multiple combinations of various functional forms and explanatory variables, including multilinear models and additional variability indices found in the literature. This long and arduous process is not described here for brevity. It is important to mention, though, that the potential role of surface albedo has been thoroughly tested because previous studies (e.g., [2]) determined that it has a significant impact on K. It was found, however, that adding that variable in the list of predictors did not improve things appreciably, even for sites under KG climate E—which are characterized by the presence of frequent or permanent snow, and are sometimes in mountainous areas. Two reasons for this apparent insensitivity to albedo might be that (i) the backscattering effects are already taken care of by K_{cs} and/or K_{ds}, and (ii) the uncertainty in the ground albedo data is large, while other modelling issues exist in mountainous areas. In short, the selection of the best submodel and list of predictors for each sky class followed a data-driven approach, to a large extent. In particular, the essential criterion was to select the submodel that had the lowest (or near lowest) root mean-square error (RMSE), along with the least possible number of variables to avoid overfitting. Once K is determined, the diffuse horizontal and direct normal components are readily obtained from $D_h = KG_h$ and $\mathrm{DNI} = (G_h - D_h)/\cos Z$, respectively, where Z is the solar zenith angle (SZA).

When the model parameterizations above are scrutinized carefully, some of the functional forms might appear counterintuitive. That happens, for instance, in the ScatterClouds and ThinClouds classes, which make use of the cloud-enhancement index as a predictor, conversely to the CloudEn class, which does not use it. This occurs because CAELUS already provides cloud enhancements as a specific sky class. Hence, the tangible relevance of the cloud–enhancement index vanishes within the CloudEn class. In parallel, however, there might remain situations in the ScatterClouds and ThinClouds classes that somehow *resemble* cloud enhancements. Within these classes, the cloud-enhancement index is then very informative to distinguish such particular situations from others. Moreover, for the Cloudless sky type, D_h could have logically been evaluated directly from a clear-sky solar irradiance model. When tested, however, this option yielded poorer results than Eq. (5). Most likely, this is due to errors in the prediction of clear-sky DIF because both DNI and DIF are much more sensitive than GHI to aerosol optical depth (AOD) in particular, and thus to the associated propagation of errors from the uncertain AOD input to the modeled clear-sky solar radiation [9,10,12].

In order to account for the 1-min solar position in the GHI separation, some authors have used SZA, or even the solar apparent time, as an additional explanatory variable (e.g., [43,44]). In GISPLIT, this role is indirectly accounted for through a conditional air-mass-dependent fitting. Specifically, the training dataset is divided into 12 log-spaced relative air mass (m_r) intervals from 1 to 10, and the models are independently fitted for each interval. The upper limit of 10 for m_r (i.e.,

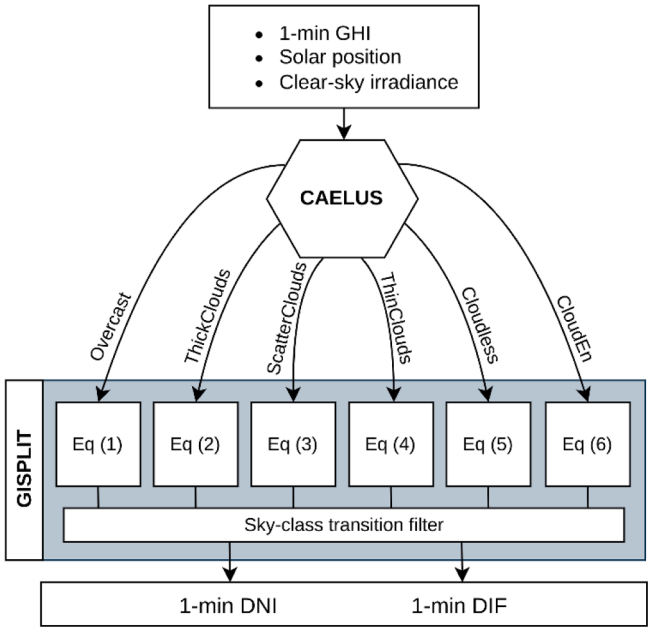


Fig. 4. Flowchart diagram of the GHI separation working principles in GISPLIT. In the climate-conditioned versions (i.e., G2 and G4, see Section 4.1), this schema is replicated for each climate class independently. In the versions augmented with extreme gradient boosting (i.e., G3 and G4, see Section 4.1), Eq. (3) and Eq. (6) are replaced with extreme gradient boosting models.

SZA≈84°) roughly corresponds with the maximum SZA for which CAELUS can classify the sky (i.e., 85°). (The relative air mass is evaluated as a function of SZA from Eq. (B.8) according to Gueymard [45] without pressure correction, while SZA is obtained for the central time of each 1-min period according to Blanco *et al.* [46].) Hence, the fitting results for each sky-aware submodel are 12 sets of coefficients, one for each m_r interval. For each prediction moment, the specific value of each fitting coefficient is interpolated to the actual 1-min air mass value using second-order regular splines. When the GHI separation needs to be done for SZA $> 84°$, the sky type can be prescribed equal to the last detected class, and K can be computed prescribing $m_r = 10$. Alternatively, any other separation model suited for such high SZAs might be used. The numerical values needed to operate all functions of the model are tabulated in Appendix A.

For illustration purposes, Fig. 3 shows the fitting coefficients when the model is trained using all climate-C site data combined (cf. Section 4). For the Overcast and ThickClouds classes, the coefficients a_1 and b_1 are ≈1 as expected, meaning that the calculation of DIF is mostly an air-mass-dependent bias correction—somewhat larger for ThickClouds, also as expected. For ScatterClouds, the coefficient c_0 is ≈0, which means that it could be removed from Eq. (3) without significant penalty. In parallel, the coefficients d_1 and d_2 for ThinClouds, as well as f_1 and f_2 for Cloudless, are similar in magnitude but with opposite signs. As they are associated with the variables K_T and $\overline{K}_T$, respectively, and these variables are foreseeably similar under the ThinClouds and Cloudless sky classes, both K_T and $\overline{K}_T$ could have been neglected in Eqs. (4) and (5). Indeed, for these two sky classes, K_{cs}, which is similar in magnitude to K_T and $\overline{K}_T$, has higher prediction power because its associated fitting coefficient (f_3) is greater than f_1 and f_2. However, in general, this is not so for all climate types or for all other sky types. For instance, for the ScatterClouds sky type, the coefficients c_2 and c_3 (associated with the variables K_T and $\overline{K}_T$, respectively), are not equal in magnitude and have opposite signs, hence they must be considered in Eq. (3). Under these circumstances, and because all sky types are interleaved in a GHI time series, it would not be sound to drop K_T and $\overline{K}_T$ out from the ThinClouds and Cloudless submodels, while keeping them in others. Therefore, in order to keep a good degree of homogeneity among all submodels, both K_T and $\overline{K}_T$ are used as predictors in all sky type models—except for Overcast and ThickClouds, as explained above. Another exception here is K_{de}, which is not considered in the Cloudless and CloudEn submodels because its use was found detrimental.

Fig. 3 also shows that the coefficients are only evaluated up to m_r≈5 (i.e., SZA≈78°) for the CloudEn variability class. This is because these situations very rarely occur for higher relative optical air masses. If that happens at prediction time, the value of the coefficients is linearly extrapolated to the corresponding m_r value.

Overall, despite the massive effort conducted to select the best-performing functional form for each sky class, there might be additional tweaks or alternative approaches that could still improve results. For instance, one possibility would be to explore more physics-based ways to construct the prediction submodels, rather than the multilinear logistic curve adopted here. The latter was considered preferable in the present context because that approach was already tested and found efficient in the literature. By doing so, the advantage is that any improvement provided by the present model can be directly traced back to its sky-type-based foundation, which is the main novelty introduced in this study.

Despite the benefits expected from breaking down the GHI separation process into various sky classes, this approach might also create artifacts around the transition edges between different sky classes, when considering the resulting 1-min time series of DNI and/or DIF. This has been thoroughly checked, and has been rarely observed in the transition from ScatterClouds to ThinClouds, or vice versa. However, to prevent and mitigate such issue at unseen sites, a 15-min moving average is applied to the transition edges between these two sky classes.

Fig. 4 shows a flowchart diagram of the general GISPLIT model architecture, which underlines the central role of CAELUS for the proper operation of the separation model. The first modeling stage involves the sky classification based on CAELUS sky classes [29], after which each input time step is bounded with a CAELUS sky class, so that a specialized separation function is eventually used for each sky type. Finally, a filter is used to smooth out the possible, although very rare, misalignments in the transition between sky classes, as highlighted above.

4. Detailed methodology

4.1. Model training

The CAELUS database is broken down into two separated datasets, one for model training (which involves all the data from each station, except the latest year), hereinafter referred to as the Train dataset (≈40 million quality-assured data points), and one for model validation (involving the latest year of data from each station), hereinafter referred to as the Valid dataset (≈10 million quality-assured data points). (There are exceptions for Petrolina and Solar Village, where only 4 years of data are available; they are all included in the Train dataset, whereas none is in the Valid dataset.).

Two versions of the separation model are actually fitted to the Train dataset, namely: (i) a single fit with all Train site data combined, referred to as GISPLIT version 1 (hereafter GISPLITv1 or G1), and (ii) a conditioned fit by major KG climate class (i.e., GISPLIT version 2, hereafter GISPLITv2 or G2). The latter implies 5 independent model fits (one per climate class), alternatively using all the training site data segregated by climate class. Therefore, G2 requires knowledge of the climate class prior to performing any GHI separation, i.e., at prediction time, because this information is needed to select the proper model parameters. (Note that the exact climate classification of any location depends on the spatial resolution of the selected KG database, hence the latter ultimately has an impact on the splitting model performance; this issue had not been reported yet in the literature.)

Another attempt at improving the overall model is prompted by the

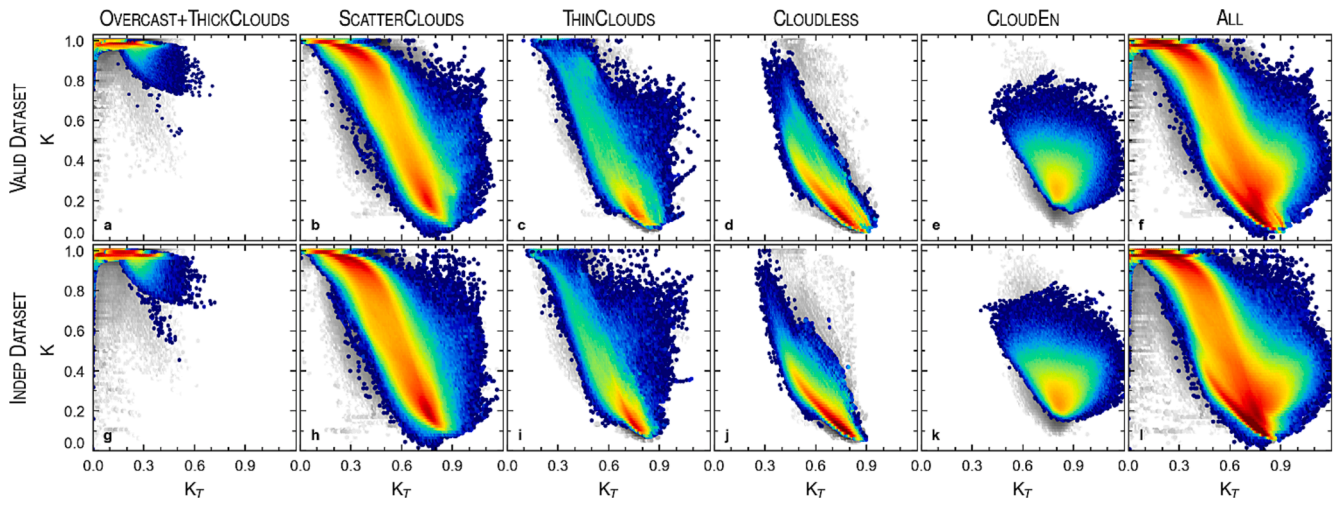


Fig. 5. Predicted K as a function of observed K_T for each sky class determined by CAELUS for the VALID dataset (panels a–e) and the INDEP dataset (panels g–k), and for all sky classes combined (panel f, for the VALID dataset, and panel l for the INDEP dataset) using GISPLITv1. The data points in gray color indicate the corresponding measurements in each sky class. Redder colors indicate a larger density of data points.

fact that the GHI separation for the SCATTERCLOUDS and CLOUDEN sky types is particularly challenging. That is because the former case involves a large variety of sky conditions (cf. Fig. 1a), whereas, in the latter case, a high temporal variability of the irradiance components exists during cloud enhancement events. Based on these concerns, an alternative modelling approach was devised in which Eq. (3) and Eq. (6) are replaced by extreme gradient boosting fits [32,33] supplied with the same explanatory variables as in these equations (Section 4.2). As described just above, two versions are trained, namely: (i) a single fit with all training data combined (GISPLIT version 3, hereafter GISPLITv3 or G3), and (ii) a conditioned fit by major KG climate class (GISPLIT version 4, hereafter GISPLITv4 or G4).

Overall, the inter-comparison of the four GISPLIT versions gives an opportunity to evaluate the potential advantage (in terms of improved accuracy) of (i) the climate-conditioned fit, and (ii) machine-learning techniques, at least for the most difficult sky types (i.e., SCATTERCLOUDS and CLOUDEN, as described in Section 4.2). In particular, if G2 appeared to be not clearly better than G1, the climate-conditioned fit would be counterproductive because it involves additional burden to determine the climate type (depending on the source of the KG gridded data and the associated spatial resolution). Most importantly, this issue could be especially troublesome for the separation of spatially-gridded GHI datasets, which have become commonplace, because the two grids can be expected to have differing spatial resolution. Similarly, if the performance of the separation model with machine-learning fits is not clearly better than without, Eqs. (3) and (6) could be considered sufficient for most applications.

4.2. Extreme gradient boosting model training

The extreme gradient boosting model is implemented using the Python interface of the XGBoost library, version 1.4.2, (https://xgboost.readthedocs.io/en/release_1.4.0) using the *gbtree* booster, with 100 regression trees, each with a maximum depth of 10, and a learning rate of 0.1. The remaining parameters are set to their defaults, including the loss function that minimizes the square error—as recommended in that library for regression problems. The number of regression trees is prescribed to 100. This is because a wide variety of preliminary tests using early stopping showed that if the number of trees was allowed to grow beyond that value, only very marginal performance improvements were obtained, while making the serialization size of the model grow disproportionately. Overall, it was found that 100 regression trees with a maximum depth of 10 was a reasonable configuration in terms of model performance and model size, while diminishing the risk of overfitting.

The training was performed using exactly the same irradiance data as for the G1 and G2 models, with the only difference that the relative optical air mass was included as one more prediction feature, and thereby the training was not conditioned by air mass.

4.3. Implementation

The huge observational datasets considered here, combined with the convoluted processes involved across the CAELUS sky classification and the GHI separation (e.g., sky-dependent functional forms, air-mass-dependent fits, or extreme gradient boosting models) induce practical difficulties that can only be overcome through special care with programming code and proper data treatment. A particularly critical issue concerns the management for the model's fitting coefficients. In all existing models so far, these coefficients are normally provided directly in the text (or in tables if they are too numerous), and can thus be easily translated to programming code. In contrast, the number of model coefficients is so large here that any attempt to tabulate them in a conventional way could make the coding task overwhelming and prone to errors. For instance, the G1 fit alone requires 324 coefficients (i.e., 27 coefficients for Eqs. (1) to (6), times 12 air mass intervals), and the G2 fit involves 5 times more (since there are 5 major KG classes). Because of their gradient boosting approach, versions G3 and G4 rely on even more coefficients. Hence, instead of having hard-coded coefficients embedded in the programming code, the solution adopted here is to dump them directly to separate *json* (JavaScript Object Notation) files after the fitting process. These files are shared as supplementary data in Appendix B.

5. Results and validation

5.1. Error metrics and validation datasets

The performance of the model is evaluated here separately with two different validation datasets: (i) the VALID dataset introduced in Section 4.1, and (ii) the PVPS database (≈14 million quality-assured data points), whose locations and radiometers are different than those of the GISPLIT training dataset (contrarily to the VALID dataset), and thus constitutes the backbone of a completely *independent* validation; for that reason, it is referred to as INDEP in what follows.

The performance of GISPLIT is evaluated as usual by the mean bias error (MBE) and RMSE metrics. Since the accuracy of any separation

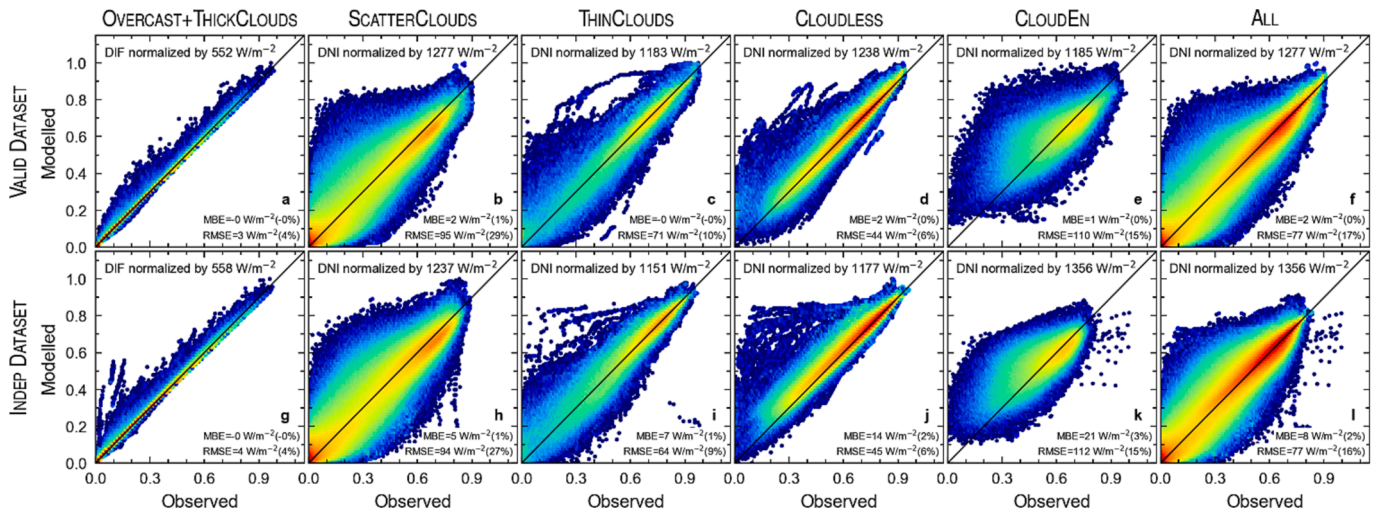


Fig. 6. Predicted vs. observed DIF or DNI for each sky class determined by CAELUS, and for all sky classes combined, for the VALID dataset (panels a–f) and for the INDEP dataset (panels g–l) using GISPLITv1. Panels a and g show the predicted DIF vs. observed DIF because DNI≈0 for OVERCAST + THICKCLOUDS. The rest of panels show predicted DNI vs. observed DNI. The DIF and DNI values are both normalized by the maximum value in each sky class to have the same scale in all panels and ease the visualization. The normalization value is shown in the upper part of each panel. The panels also show the MBE and RMSE statistics in each case.

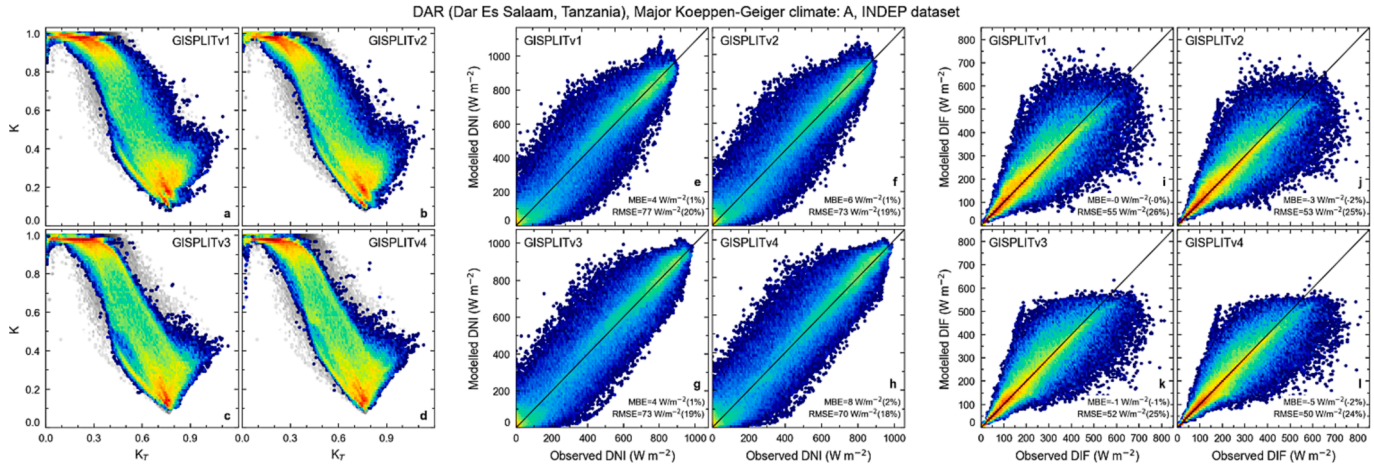


Fig. 7. Results of the four GISPLIT versions at Dar Es Salaam, Tanzania (KG class A, INDEP dataset). Panels a–d: Diffuse fraction in K-K_T space (the gray points in the background are the measured values). Panels e–h: Scatterplots of modelled vs. measured DNI. Panels i–l: Scatterplots of modelled vs. measured DIF. Panels e–l also show the MBE and RMSE statistics in each case.

model is way poorer than the actual observational uncertainty of ground radiometers at high-quality stations such as those under scrutiny here, both MBE and RMSE can be estimated from the direct comparison of each model's predictions and their observational counterparts [47].

5.2. Validation by sky type

The pertinence of using sky types, in terms of their specific impact on the results, needs to be evaluated since this is a key feature of the novel methodology proposed here. The impact of sky-type separation on GISPLITv1's performance in K-K_T space is shown in Fig. 5, when using both the VALID and the INDEP datasets along with all climate classes combined. The distribution of predicted values (color foreground overlay) agrees generally well with the pattern of observed values (gray background), except for the OVERCAST + THICKCLOUDS panels, which combine those two normally separate sky classes into a single one (Fig. 5a and 5 g), simply because they present common characteristics: (1) they both correspond to very low GHI magnitudes, thus making K so sensitive to small DIF variations that they are hardly seen by the model; and (2) the small GHI magnitude makes these situations almost irrelevant anyway. Note, in addition, that the observed patterns for these two sky situations are noticeably different in the VALID dataset and in the INDEP dataset (contrarily to what happens for the rest of sky situations). In particular, the INDEP dataset's observations are prone to somewhat lower K values. This can be explained, at least in part, by the different instruments and methods of observation (some with increased uncertainty) used in the radiometric sites of the INDEP dataset, as described in Section 2.

Similarly, Fig. 6 shows scatterplots of the predicted vs. measured GHI components with GISPLITv1, again segregated by sky class, and for both the VALID and INDEP datasets. For the OVERCAST + THICKCLOUDS combined situations (i.e., Fig. 6a and 6 g), the corresponding panels show the scatterplots of predicted DIF vs. measured DIF because DNI≈0; the rest of panels rather show similar scatterplots, but for DNI. For further reference, Fig. 6 also provides the MBE and RMSE statistics for each sky class. Considering all sky classes combined, the overall MBE and RMSE statistics for the predicted DNI are 0% and 17% for the VALID dataset, and 2% and 16% for the INDEP dataset, respectively. These results are important when comparing the present model to others from the literature.

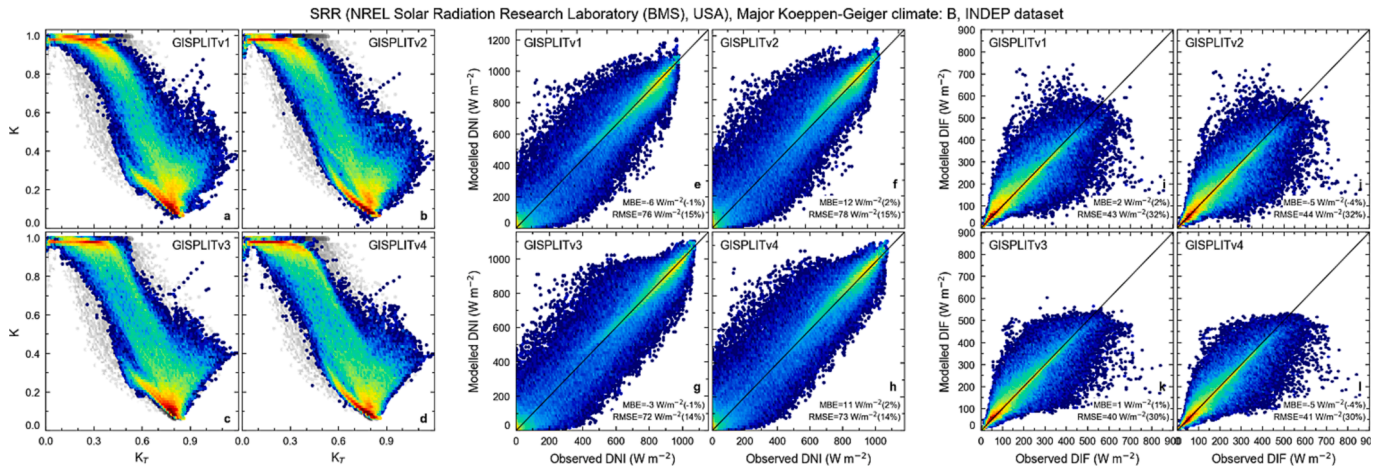


Fig. 8. Same as Fig. 6, but at the NREL Solar Radiation Research Laboratory (Baseline Measurement System), in Golden, Colorado, USA (KG class B, INDEP dataset).

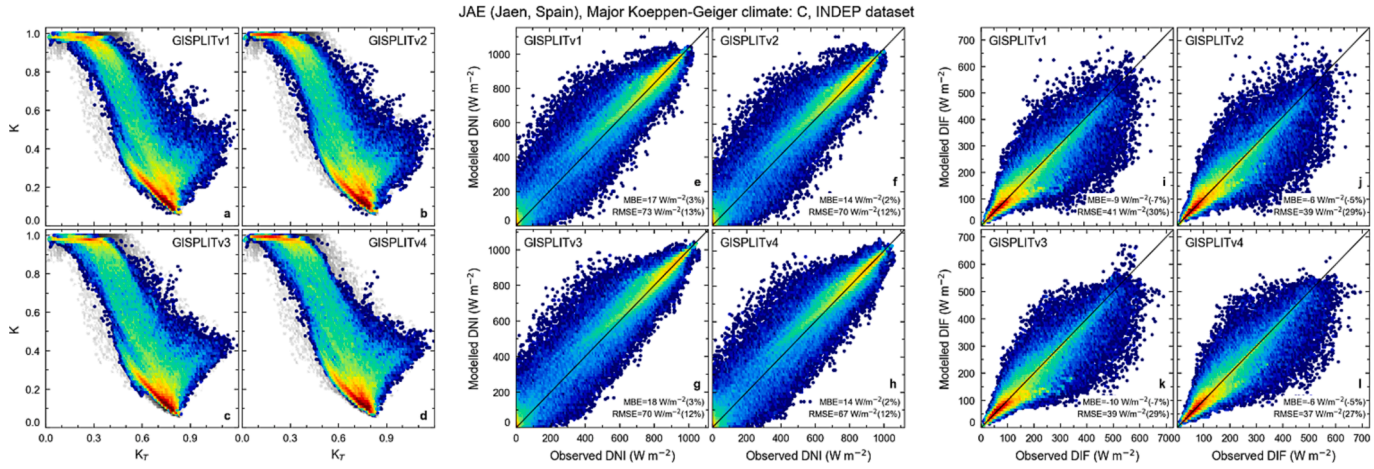


Fig. 9. Same as Fig. 6, but at Jaén, Spain (KG class C, INDEP dataset).

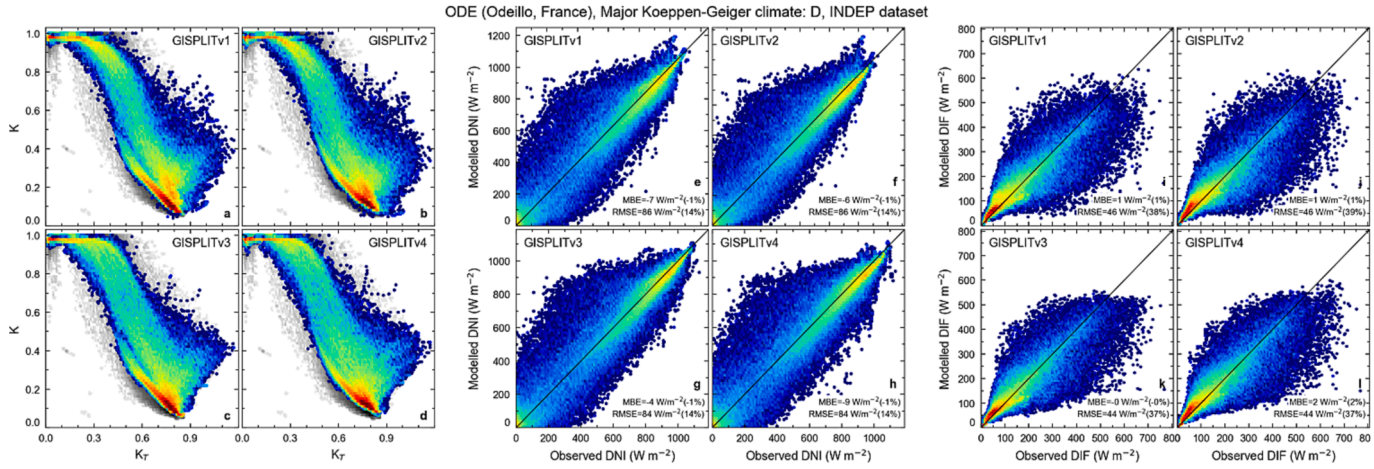


Fig. 10. Same as Fig. 6, but at Odeillo, France (KG class D, INDEP dataset).

5.3. Validation at representative stations

Because of space limitations, only a succinct validation is presented here, using a few representative sites. More complete results, involving the two complete validation datasets and various levels of disaggregation of the results, as well as other separation models of the literature, will be discussed in a forthcoming contribution.

The results of the four versions of GISPLIT are analyzed here visually in Figs. 7–11, using three types of plots: (i) the diffuse fraction in K-K_T space; (ii) the scatterplot of predicted vs. measured DNI; and (iii) the scatterplot of predicted vs. measured DIF. One representative station for each major KG climate class is selected. They are all extracted from the

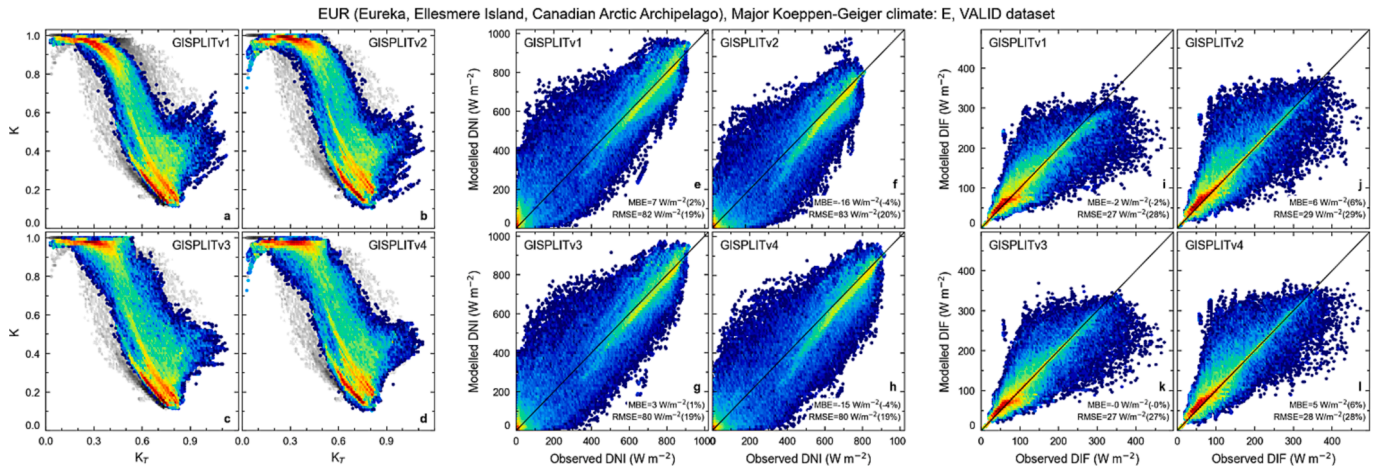


Fig. 11. Same as Fig. 6, but for Eureka, Canada (KG class E, VALID dataset).

Table 3
DNI's RMSE (in W/m^2, and in percent relative to the mean observed DNI, between brackets) for all VALID or INDEP site data combined (All climates), and by primary Köppen–Geiger climate. The lowest RMSE value is bold-faced in each case.

Dataset and Model	All climates		Climate A		Climate B		Climate C		Climate D		Climate E	
VALID												
GISPLITv1	76.8	(17.0)	79.0	(18.8)	72.3	(11.0)	69.6	(16.6)	74.6	(20.1)	97.6	(25.3)
GISPLITv2	73.1	(16.2)	77.6	(18.5)	66.7	(10.2)	68.4	(16.3)	73.2	(19.7)	86.2	(22.3)
GISPLITv3	72.0	(15.9)	74.6	(17.7)	67.5	(10.3)	66.3	(15.8)	70.8	(19.1)	88.4	(22.9)
GISPLITv4	**69.6**	**(15.4)**	**73.7**	**(17.5)**	**62.9**	**(9.6)**	**65.3**	**(15.6)**	**69.6**	**(18.7)**	**82.2**	**(21.3)**
INDEP												
GISPLITv1	76.6	(15.9)	79.6	(19.6)	73.1	(12.3)	77.9	(17.9)	78.2	(18.8)	–	–
GISPLITv2	77.0	(16.0)	75.1	(18.5)	75.0	(12.7)	78.7	(18.1)	78.6	(18.8)	–	–
GISPLITv3	**73.5**	**(15.3)**	73.5	(18.1)	**70.7**	**(11.9)**	**75.2**	**(17.3)**	**75.3**	**(18.1)**	–	–
GISPLITv4	74.8	(15.5)	**71.5**	**(17.6)**	73.0	(12.3)	77.0	(17.7)	75.9	(18.2)	**–**	**–**

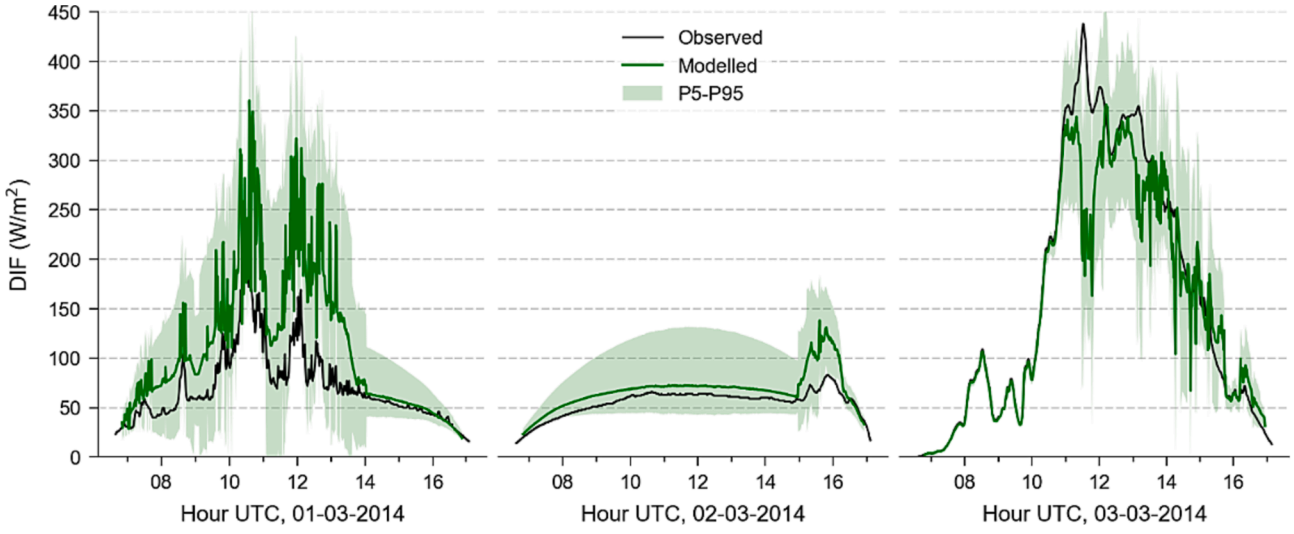


Fig. 12. Sample results of the separated DIF using GISPLITv1 and the P5–P95 separation range, compared to the observed DIF for the 1st to 3rd March 2014, at the Carpentras site of the BSRN network, as in Fig. 2 of Ruiz-Arias and Gueymard [29]. Percentiles P5 and P95 are obtained by linear interpolation of the data in Table 4.

INDEP dataset, except for KG class E, for which only the VALID dataset has stations. The results appear consistent both between the four GISPLIT versions and between the five climate classes. In general, the gradient boosting versions (G3 and G4) provide improvements over the conventional versions (G1 and G2), which are particularly visible (less scatter and/or outliers) in the case of the DIF scatterplots.

5.4. Comparative results for the four versions of the model

As indicated above, two objectives of this study are to evaluate whether more accurate results can be obtained either with (i) the stratification into KG classes; or (ii) the use of a machine-learning booster. To that effect, Table 3 displays the RMSE statistics for the 1-min predictions from all four GISPLIT versions against the measured

Table 4
Error percentiles of the diffuse fraction predicted by GISPLITv1 for all data from the Valid and Indep dataset combined, and evaluated as the observed K minus the GISPLITv1-*predicted* K.

	Overcast	ThickClouds	ScatterClouds	ThinClouds	Cloudless	CloudEn
P2	–0.147	–0.150	–0.275	–0.169	–0.063	–0.146
P10	–0.026	–0.026	–0.134	–0.077	–0.030	–0.092
P20	–0.009	–0.009	–0.074	–0.043	–0.017	–0.064
P30	–0.001	–0.001	–0.041	–0.026	–0.009	–0.044
P40	0.005	0.003	–0.012	–0.014	–0.002	–0.026
P50	0.007	0.007	0.009	–0.003	0.004	–0.007
P60	0.009	0.010	0.028	0.009	0.011	0.015
P70	0.012	0.013	0.051	0.024	0.020	0.042
P80	0.015	0.018	0.084	0.044	0.033	0.080
P90	0.020	0.027	0.145	0.083	0.059	0.146
P98	0.037	0.076	0.287	0.211	0.134	0.278

Table A1
Boundary limits for the relative air mass (AM) fitting intervals.

AM range	AM lower bound	AM upper bound
1	1.000000	1.211528
2	1.211528	1.467799
3	1.467799	1.778279
4	1.778279	2.154435
5	2.154435	2.610157
6	2.610157	3.162278
7	3.162278	3.831187
8	3.831187	4.641589
9	4.641589	5.623413
10	5.623413	6.812921
11	6.812921	8.254042
12	8.254042	10.000000

Table A2
Fitting coefficients for Overcast sky type [Eq. (1)] and all KG climates combined.

AM range	a_0	a_1
1	−0.270430	0.985281
2	−0.184382	0.989643
3	−0.070525	0.988872
4	−0.100241	0.989707
5	−0.160691	0.990882
6	−0.201713	0.991856
7	−0.202749	0.991523
8	−0.099830	0.985926
9	−0.078389	0.982755
10	0.032860	0.974661
11	0.049992	0.969138
12	−0.022059	0.966869

Table A3
Fitting coefficients for ThickClouds sky type [Eq. (2)] and all KG climates combined.

AM range	b_0	b_1
1	2.836804	0.969702
2	2.313497	0.974426
3	2.519388	0.970887
4	2.575589	0.968383
5	3.138318	0.960052
6	3.675868	0.948999
7	4.656980	0.926000
8	7.772397	0.863611
9	10.896965	0.776540
10	11.409539	0.706734
11	10.899992	0.649094
12	8.866133	0.620023

data from the two complete validation databases, with or without climate stratification. Using the Valid dataset, the best results are obtained with G4 for any climate, whereas G3 obtains the lowest RMSE with the Indep database. Moreover, the RMSEs vary somewhat with the KG climate class. Unsurprisingly, the largest deviations between G1 and G2, and between G3 and G4, are found for climate class E, but interestingly, they are also noticeably large for climate B. For the former, the explanation is that the measurements are carried out under challenging or extreme conditions [48], SZA is often very high, and the surface albedo remains typically large. For climate B, the likely reason is the higher prevalence of cloudless situations combined with the AOD-induced prediction uncertainty at these locations. However, because the mean DNI there is typically large, the deviations remain small in relative terms. The results for the other climate classes are similar to those that are obtained for the whole database, i.e., all climate classes combined.

Based on these results, two major conclusions are obtained: (i) the machine-learning augmentation generally helps increase the accuracy of the model, although not always significantly, depending on climate class; and (ii) the climate stratification also helps improve the model's performance, but this is only significant for climate E. All this suggests that a larger performance boost is gained from the machine-learning algorithm than from the climate stratification. Although a small improvement can be gained with the climate conditioning, that requires prior knowledge of the KG major climate at every location, which might depend on the selected KG source database. That might furthermore induce spatial prediction artifacts at the seams between climate zones. Overall, the added accuracy appears too small to justify the much-increased complexity, in general.

5.5. *Confidence intervals*

A side benefit of the sky-conditioned fit used in GISPLIT is that it can generate confidence intervals for the separation results, as exemplified in Fig. 12. To that aim, various percentiles of the diffuse fraction error are calculated for each sky type using all data from the Valid and Indep datasets combined (Table 4), which can then be converted to DIF percentile errors by multiplying them by the observed GHI in each case. This simple approach is only dependent on the sky condition. A more elaborate approach could refine these percentile errors by adding another stratification level based on clearness index, and possibly even on SZA.

6. Discussion and conclusions

The separation of GHI into its direct and diffuse components is important because it affects processes at the core of the whole solar energy processing chain. For instance, the solar industry depends on solar irradiance databases that typically require a GHI separation model to also provide the DNI and DIF components. In parallel, the number of

Table A4
Fitting coefficients for ScatterClouds sky type [Eq. (3)] and all KG climates combined.

AM range	c_0	c_1	c_2	c_3	c_4	c_5	c_6
1	−0.168923	−6.062615	−1.508863	4.193491	5.181752	−2.481742	1.024458
2	−0.147246	−6.204689	−3.160420	4.338477	6.550011	−2.663039	1.114486
3	−0.139633	−6.066288	−5.787197	4.192482	8.532064	−3.208247	1.291603
4	−0.078665	−6.045641	−6.643890	4.251738	9.303360	−3.760370	1.319456
5	−0.023811	−5.853729	−7.036026	4.024696	9.734440	−4.309999	1.285265
6	0.005094	−5.807881	−6.409352	3.775761	9.411570	−4.235079	1.343090
7	−0.020502	−5.783494	−4.924317	3.126877	8.523758	−3.708996	1.484430
8	−0.027818	−5.338959	−2.937058	1.910810	7.418074	−3.385720	1.603199
9	−0.119469	−5.397146	−1.468749	1.354740	6.420221	−2.673630	1.816240
10	−0.147949	−5.575542	−0.269800	0.724627	6.038634	−2.339497	1.869280
11	−0.118696	−5.887946	0.262464	0.565829	6.168884	−2.380477	1.813577
12	−0.099416	−6.037718	0.864560	0.465383	6.248804	−2.650968	1.789796

Table A5
Fitting coefficients for ThinClouds sky type [Eq. (4)] and all KG climates combined.

AM range	d_0	d_1	d_2	d_3	d_4	d_5
1	−4.285330	1.255239	3.829909	2.896615	−3.978833	0.554818
2	−5.294446	0.448661	3.979787	4.377741	−3.831376	0.673535
3	−6.491469	−1.270476	4.656898	6.320168	−3.551567	0.994920
4	−7.200966	−4.499367	5.851932	8.485546	−3.624856	1.346607
5	−6.991850	−8.584282	6.985822	10.375043	−4.286350	1.395247
6	−6.790056	−8.687491	6.737371	10.370912	−4.549546	1.510918
7	−6.541336	−9.835234	7.530466	10.306913	−4.833142	1.609632
8	−5.568492	−7.228173	4.503725	9.569243	−5.062407	1.679447
9	−5.129560	−5.912712	4.101678	8.349739	−4.787419	1.660875
10	−4.394846	−3.724484	2.117670	7.335274	−4.785777	1.603161
11	−3.913954	−0.632028	−0.118467	6.016344	−4.164976	1.544461
12	−3.607733	1.811084	−1.855264	4.922344	−3.437181	1.473780

Table A6
Fitting coefficients for Cloudless sky type [Eq. (5)] and all KG climates combined.

AM range	f_0	f_1	f_2	f_3	f_4
1	−1.192711	1.062605	3.192159	0.580028	−4.287379
2	−1.463937	2.584490	1.610016	0.809802	−3.800818
3	−1.036839	4.340858	−1.825881	1.616159	−3.752842
4	−0.623279	1.468524	0.093551	1.849545	−3.709492
5	−0.463827	1.432064	−0.423031	2.006101	−3.625812
6	−0.378256	−0.112933	0.848734	2.014385	−3.462603
7	−0.337838	−0.701105	1.236530	2.000446	−3.320745
8	−0.469989	0.728489	−0.551854	2.328537	−3.445325
9	−0.832497	−3.323310	3.086878	2.927484	−3.721025
10	−1.121246	−3.768125	3.050348	3.542450	−4.179149
11	−1.497161	−4.554619	3.487383	4.164519	−4.666296
12	−1.503478	−3.579747	1.879193	4.692566	−5.448330

Table A7
Fitting coefficients for CloudEn sky type [Eq. (6)] and all KG climates combined.

AM range	g_0	g_1	g_2	g_3	g_4
1	2.528170	0.475880	1.711019	−2.377725	−2.882007
2	2.427064	0.112267	1.639709	−2.035464	−2.621002
3	2.476252	−1.514925	1.384062	−0.766033	−2.787198
4	2.333035	−1.561375	1.127792	−0.535032	−2.707401
5	2.262218	−1.475785	0.788599	−0.416444	−2.641725
6	2.065023	−0.883507	0.488298	−0.603592	−2.316119
7	1.772686	−0.175208	0.117446	−0.751448	−1.946888
8	1.605940	0.091902	−0.309805	−0.603919	−1.964280
9	1.399476	0.978482	−0.920736	−0.790776	−1.602551
10	–	–	–	–	–
11	–	–	–	–	–
12	–	–	–	–	–

Table A8
Fitting coefficients for Overcast sky type [Eq. (1)] and KG climate A.

AM range	a_0	a_1
1	−0.656394	0.983670
2	−0.286845	0.988408
3	−0.181854	0.989689
4	−0.086805	0.989268
5	−0.106513	0.989177
6	−0.033340	0.986435
7	0.043915	0.982658
8	0.159806	0.976280
9	0.095269	0.974302
10	0.245972	0.964241
11	0.148298	0.964236
12	0.165868	0.957912

Table A9
Fitting coefficients for ThickClouds sky type [Eq. (2)] and KG climate A.

AM range	b_0	b_1
1	4.304892	0.960670
2	4.918151	0.958515
3	4.170482	0.959888
4	5.599840	0.942924
5	8.040371	0.911454
6	8.681748	0.886229
7	9.200023	0.854108
8	14.928363	0.748117
9	18.107636	0.635036
10	16.892266	0.576719
11	15.133573	0.532388
12	11.318045	0.554695

radiometric stations that only observe GHI worldwide is much greater than those that also monitor DNI and/or DIF. Hence, a full exploitation of the data in the former type of stations is only possible when one can count on a reliable separation model.

This work has tackled the separation of GHI into DNI and DIF with a new approach, referred to as GISPLIT, which uses an innovative method

Table A10
Fitting coefficients for ScatterClouds sky type [Eq. (3)] and KG climate A.

AM range	c_0	c_1	c_2	c_3	c_4	c_5	c_6
1	−0.167651	−5.805640	2.727195	3.909189	1.879837	−1.666776	0.822420
2	−0.115901	−5.960300	0.600489	4.257214	3.710370	−2.289587	1.065038
3	−0.084636	−5.857396	−2.497947	4.214328	6.073018	−2.840817	1.237982
4	−0.022826	−5.513450	−1.654621	3.760743	5.690122	−2.967539	1.216380
5	0.048042	−5.247286	0.680162	3.212790	4.494390	−2.900523	1.097639
6	0.074478	−4.990815	1.167509	2.498790	4.471235	−2.853132	1.158667
7	0.101374	−4.816338	2.280277	1.731573	4.133559	−2.693485	1.196584
8	0.203751	−4.401034	3.372784	0.619994	4.262508	−3.247373	1.036023
9	0.273697	−4.401246	4.080042	0.019061	4.564732	−3.530466	0.951185
10	0.341769	−4.443364	4.671100	−0.729797	5.198332	−4.073044	0.840270
11	0.347907	−4.639148	4.655462	−1.104242	5.603277	−4.164980	0.876214
12	0.368746	−4.966124	5.026218	−0.558449	5.765127	−4.416797	0.887853

Table A11
Fitting coefficients for ThinClouds sky type [Eq. (4)] and KG climate A.

AM range	d_0	d_1	d_2	d_3	d_4	d_5
1	−5.354574	−1.979047	4.480399	6.111432	−4.852934	0.791256
2	−5.605110	−0.858177	4.751803	5.217311	−3.992279	0.924208
3	−5.279324	−0.707743	4.800658	4.635633	−3.401802	1.004094
4	−4.829310	−2.565501	5.484564	5.037441	−3.548136	1.273912
5	−4.109059	−1.105312	4.530529	3.849139	−3.032887	1.244200
6	−4.424225	−0.600690	4.734261	3.643031	−2.662115	1.410721
7	−4.316503	1.025278	5.156552	2.002806	−1.639507	1.401672
8	−4.009664	2.139396	2.362278	2.801229	−2.010701	1.486278
9	−4.588721	3.586592	1.337246	3.204615	−2.127981	1.587570
10	−4.181167	3.986754	−0.732347	3.787669	−2.613944	1.598031
11	−4.056383	4.854996	−2.746832	4.055529	−2.514632	1.593231
12	−4.146663	3.564124	−3.208271	4.830139	−2.706207	1.612702

Table A12
Fitting coefficients for Cloudless sky type [Eq. (5)] and KG climate A.

AM range	f_0	f_1	f_2	f_3	f_4
1	−1.033214	−2.916872	4.620997	2.405105	−4.709904
2	−2.421706	−1.118823	3.998609	2.871034	−4.387345
3	−1.620163	−1.510108	4.024596	2.236200	−3.948132
4	−1.440001	−1.488602	4.170860	1.853427	−3.593230
5	−1.544383	0.811375	2.277661	1.589058	−3.212043
6	−1.561622	0.647620	2.553657	1.436798	−2.860246
7	−1.311328	0.992797	2.417315	0.974482	−2.551956
8	−0.798268	3.066336	0.649209	0.195074	−2.241835
9	−0.480334	2.222393	0.900908	0.160402	−2.164078
10	−0.432533	2.103613	0.261153	0.472750	−2.196910
11	−0.475899	3.463509	−1.144067	0.566575	−2.391671
12	−1.494487	6.003847	−2.001774	0.665326	−2.284539

Table A13
Fitting coefficients for CloudEn sky type [Eq. (6)] and KG climate A.

AM range	g_0	g_1	g_2	g_3	g_4
1	2.740233	3.009553	1.622893	−4.431401	−2.359735
2	2.781218	0.794773	1.605058	−2.823763	−2.503507
3	2.788144	−0.506508	1.390527	−1.815844	−2.450628
4	2.704475	−0.307199	1.104223	−1.781941	−2.236023
5	2.598873	0.219565	0.729514	−1.870360	−2.129149
6	2.450126	−0.154024	0.401198	−1.380783	−2.099692
7	2.160488	0.165348	0.089004	−1.270988	−1.938530
8	1.961863	−0.140515	−0.578202	−0.624059	−2.158320
9	2.018225	−1.882605	−1.123974	0.521693	−2.553936
10	–	–	–	–	–
11	–	–	–	–	–
12	–	–	–	–	–

Table A14
Fitting coefficients for Overcast sky type [Eq. (1)] and KG climate B.

AM range	a_0	a_1
1	−0.375108	0.984306
2	−0.260984	0.981449
3	−0.290487	0.984694
4	−0.594492	0.988305
5	−0.626525	0.988220
6	−0.304630	0.983805
7	−0.348521	0.981993
8	−0.016885	0.970643
9	0.353153	0.953446
10	1.099899	0.914144
11	1.089478	0.898912
12	0.916663	0.886234

Table A15
Fitting coefficients for ThickClouds sky type [Eq. (2)] and KG climate B.

AM range	b_0	b_1
1	3.182840	0.967380
2	2.348034	0.968235
3	2.620072	0.965367
4	4.771158	0.949122
5	4.720870	0.942572
6	8.113565	0.900551
7	12.169037	0.835388
8	13.053667	0.769291
9	16.304357	0.646093
10	18.309082	0.517170
11	14.854195	0.478911
12	11.543434	0.430205

to account for the role of GHI's variability in that process. Conversely to current models, which employ one or multiple GHI variability indices as predictors for the GHI separation, GISPLIT exploits such variability information in a preliminary classification step to separate the GHI data into 6 sky classes, each corresponding to specific spans of solar irradiance magnitude and variability, which are in turn related to different

Table A16
Fitting coefficients for ScatterClouds sky type [Eq. (3)] and KG climate B.

AM range	c_0	c_1	c_2	c_3	c_4	c_5	c_6
1	−0.192548	−6.109826	−1.838445	4.132585	5.541511	−2.397925	1.495795
2	−0.174677	−6.027566	−2.469644	4.136362	5.942677	−2.230535	1.330105
3	−0.117032	−6.019608	−1.204095	4.176261	5.111548	−1.985136	1.317841
4	−0.089883	−5.901868	−1.739732	3.992869	5.609738	−2.013985	1.367814
5	−0.047368	−5.633497	−1.944880	3.715390	5.854377	−2.243501	1.384282
6	0.018126	−5.377110	−1.048755	3.271397	5.502075	−2.292815	1.284309
7	0.002591	−5.100759	0.295159	2.245750	4.866129	−1.930779	1.419841
8	0.231692	−4.967224	7.480633	0.675815	2.205985	−1.444541	0.850911
9	0.317070	−5.186978	12.129436	0.315697	0.555784	−1.005000	0.650699
10	0.370544	−4.844744	16.661080	−2.169621	−0.266208	−0.992957	0.520458
11	0.392691	−4.349061	20.084617	−4.968154	−0.618589	−1.098614	0.443030
12	0.397430	−3.699776	22.689757	−5.522285	−1.983471	−0.848289	0.309141

Table A17
Fitting coefficients for ThinClouds sky type [Eq. (4)] and KG climate B.

AM range	d_0	d_1	d_2	d_3	d_4	d_5
1	−2.006790	1.689166	1.459408	2.303752	−4.438724	0.548643
2	−3.774653	1.761642	1.562456	3.855387	−4.158532	0.664096
3	−4.218747	0.335730	3.110162	4.226747	−3.832499	0.847225
4	−4.072261	0.591705	2.854155	4.059626	−3.448239	1.125370
5	−4.199395	0.045765	3.464978	4.077963	−3.096470	1.219341
6	−4.538839	1.222128	3.266189	3.580027	−2.437074	1.356670
7	−5.778576	−1.213263	5.301311	5.126204	−2.450863	1.590069
8	−4.434456	0.972127	2.279604	4.296954	−2.437820	1.563610
9	−3.675408	1.322247	1.851053	3.432598	−2.225957	1.466804
10	−3.780620	2.790071	0.789446	3.100125	−1.881443	1.465793
11	−3.476502	3.919474	−0.714234	2.807522	−1.724195	1.439414
12	−2.787684	4.448514	−2.339946	2.512432	−1.694376	1.293049

Table A18
Fitting coefficients for Cloudless sky type [Eq. (5)] and KG climate B.

AM range	f_0	f_1	f_2	f_3	f_4
1	−0.146201	−0.782346	3.820938	0.593319	−4.582175
2	−0.511763	2.121686	1.634827	0.268268	−3.913337
3	−0.694135	7.482701	−3.227272	0.032376	−3.498189
4	−0.652457	5.073989	−1.081865	0.158504	−3.205169
5	−0.571960	5.234959	−1.614106	0.314989	−2.983494
6	−0.419357	4.068465	−0.207652	−0.050117	−2.646037
7	−0.438601	6.338447	−2.161859	−0.294561	−2.293797
8	−0.609621	4.615089	−0.364357	−0.242355	−1.991841
9	−0.925328	4.570299	0.155901	−0.285632	−1.661737
10	−1.130139	5.357460	−0.142268	−0.439247	−1.322468
11	−1.183872	6.452329	−0.852517	−0.644400	−1.070727
12	−1.423409	7.983397	−2.258640	−0.383516	−1.226627

Table A19
Fitting coefficients for CloudEn sky type [Eq. (6)] and KG climate B.

AM range	g_0	g_1	g_2	g_3	g_4
1	2.543415	0.630910	1.641963	−2.466943	−2.792043
2	2.472645	0.222811	1.588923	−2.111225	−2.429652
3	2.580810	−0.310393	1.336039	−1.698453	−2.348222
4	2.478160	−1.616330	1.071488	−0.556824	−2.476743
5	2.427333	−1.499614	0.694312	−0.452976	−2.464936
6	2.328586	−1.451911	0.201360	−0.205716	−2.490762
7	2.078186	−1.086802	−0.166802	−0.138086	−2.347669
8	1.883310	−0.924069	−0.817180	0.152049	−2.292513
9	1.715260	−0.204570	−1.210243	−0.151133	−1.766749
10	–	–	–	–	–
11	–	–	–	–	–
12	–	–	–	–	–

Table A20
Fitting coefficients for Overcast sky type [Eq. (1)] and KG climate C.

AM range	a_0	a_1
1	−0.382027	0.990778
2	−0.292155	0.993732
3	−0.226153	0.994733
4	−0.219339	0.994861
5	−0.185650	0.994852
6	−0.165449	0.994937
7	−0.126367	0.994213
8	−0.106550	0.992457
9	−0.065219	0.989746
10	−0.017213	0.986111
11	0.098288	0.976810
12	0.013601	0.977989

Table A21
Fitting coefficients for ThickClouds sky type [Eq. (2)] and KG climate C.

AM range	b_0	b_1
1	0.718332	0.983012
2	1.361109	0.982482
3	1.436308	0.981767
4	1.504245	0.979601
5	2.195502	0.970876
6	2.723195	0.961218
7	3.570930	0.942386
8	5.985098	0.894769
9	9.054842	0.817903
10	9.267812	0.760182
11	10.224168	0.676447
12	8.312546	0.655540

cloudiness situations. Afterwards, it performs the GHI separation independently for each sky class. This approach dynamically accounts for the different sky situations that evolve, more or less rapidly, throughout a GHI time series. For instance, cloud enhancements and cloudless situations, which have a totally different nature, are not treated using the

Table A22
Fitting coefficients for ScatterClouds sky type [Eq. (3)] and KG climate C.

AM range	c_0	c_1	c_2	c_3	c_4	c_5	c_6
1	−0.183810	−6.318467	−3.469432	4.450085	6.681010	−2.580976	0.982123
2	−0.167922	−6.457180	−3.686034	4.448929	7.037291	−2.694323	1.134084
3	−0.124558	−6.433074	−3.795865	4.483544	7.224471	−2.877249	1.214243
4	−0.059026	−6.300078	−3.166601	4.437226	6.889208	−2.990942	1.177418
5	−0.029916	−6.152776	−2.300472	4.148794	6.389577	−2.815011	1.202197
6	−0.001347	−6.019307	−1.486724	3.736849	6.062896	−2.795808	1.203347
7	0.005309	−5.878466	−0.177008	3.322618	5.297706	−2.521483	1.196586
8	0.055192	−5.224913	−0.487820	2.334835	5.863958	−3.414029	1.230046
9	0.063757	−5.221876	0.311851	2.130418	5.447296	−3.182296	1.235534
10	0.008697	−5.260075	0.497899	1.436589	5.619257	−3.175531	1.393677
11	−0.065391	−5.625567	0.449525	1.400793	5.957761	−3.245656	1.591936
12	−0.062442	−5.785646	1.056819	1.454728	6.048573	−3.555806	1.622879

Table A23
Fitting coefficients for ThinClouds sky type [Eq. (4)] and KG climate C.

AM range	d_0	d_1	d_2	d_3	d_4	d_5
1	−8.909707	−9.540986	9.272279	11.732172	−6.007737	0.435942
2	−8.655596	−7.101935	8.312957	10.278250	−5.295024	0.703330
3	−7.745197	−6.393678	7.264786	9.647388	−5.335674	0.829393
4	−7.235813	−6.733584	7.622455	9.107393	−5.094648	1.056345
5	−6.584462	−7.294277	7.523184	8.866505	−5.057819	1.140807
6	−6.411283	−6.723026	7.568320	8.145559	−4.607634	1.259208
7	−6.517169	−6.952832	7.495120	8.388743	−4.473873	1.458297
8	−5.466865	−3.845580	4.423454	7.288545	−4.361456	1.518813
9	−5.474819	−2.791255	3.991684	6.876778	−4.247339	1.602665
10	−4.968422	−0.082246	2.504583	5.498966	−3.837467	1.530854
11	−4.591661	2.784493	0.346059	4.460082	−3.237158	1.484023
12	−3.967764	4.202462	−1.064470	3.580540	−2.814602	1.399587

Table A24
Fitting coefficients for Cloudless sky type [Eq. (5)] and KG climate C.

AM range	f_0	f_1	f_2	f_3	f_4
1	−5.664656	−5.634211	5.843156	8.334534	−6.577344
2	−5.155775	−4.511774	5.664640	7.015908	−5.980974
3	−3.913995	−3.511826	4.492100	5.851350	−5.594354
4	−2.936223	−2.745234	3.815159	4.743414	−5.172434
5	−2.485807	−3.536350	3.890050	4.772649	−5.219444
6	−2.174078	−4.164931	4.063065	4.710258	−5.173090
7	−1.808954	−3.520444	3.097597	4.486501	−5.113188
8	−0.999692	1.346027	−1.504644	3.414915	−4.787739
9	−0.812466	−1.451124	1.362084	3.118059	−4.679740
10	−0.798979	−1.315536	1.137551	3.112952	−4.699503
11	−0.848440	−0.999496	0.910434	3.031609	−4.609970
12	−0.835409	0.183006	−0.002891	2.853173	−4.654729

Table A25
Fitting coefficients for CloudEn sky type [Eq. (6)] and KG climate C.

AM range	g_0	g_1	g_2	g_3	g_4
1	2.296907	−0.342526	1.840453	−1.658682	−2.915569
2	2.303188	−0.224886	1.707224	−1.727791	−2.752042
3	2.155211	−0.250505	1.641246	−1.611859	−2.493844
4	1.963669	0.286064	1.382557	−1.766241	−2.102162
5	1.794227	0.977899	1.132686	−2.068521	−1.583884
6	1.591371	1.471618	0.746664	−2.104991	−1.222510
7	1.202041	1.819877	0.477509	−1.943121	−0.876332
8	1.287780	1.265281	0.030720	−1.342333	−1.363603
9	1.084658	2.018479	−0.476419	−1.430241	−1.143881
10	–	–	–	–	–
11	–	–	–	–	–
12	–	–	–	–	–

Table A26
Fitting coefficients for Overcast sky type [Eq. (1)] and KG climate D.

AM range	a_0	a_1
1	0.040413	0.986395
2	−0.168590	0.988266
3	−0.107487	0.988558
4	−0.109416	0.989505
5	−0.123451	0.989853
6	−0.220189	0.991907
7	−0.196506	0.991752
8	−0.210624	0.991477
9	−0.142188	0.988812
10	−0.096408	0.984805
11	−0.128588	0.984593
12	−0.172096	0.984640

Table A27
Fitting coefficients for ThickClouds sky type [Eq. (2)] and KG climate D.

AM range	b_0	b_1
1	2.928270	0.969052
2	1.551312	0.977291
3	1.869361	0.974968
4	2.256889	0.970348
5	2.475854	0.966362
6	2.957896	0.958774
7	4.039172	0.938645
8	6.674773	0.890103
9	9.734804	0.809020
10	9.768907	0.762013
11	8.052915	0.745592
12	9.003585	0.630937

same monolithic approach, as all existing separation models do.

In addition to the aforementioned innovation, the present study intended to shed light on the claim made in a few previous studies that the separation modeling could be improved if it is conditioned by primary KG climate class. To that aim, two parallel versions of GISPLIT have been devised, one that trains the model with all the training dataset

Table A28
Fitting coefficients for ScatterClouds sky type [Eq. (3)] and KG climate D.

AM range	c_0	c_1	c_2	c_3	c_4	c_5	c_6
1	−0.127986	−6.288903	0.947594	4.429818	3.377391	−2.418113	0.659432
2	−0.114155	−6.114018	−1.622179	4.262777	5.471530	−2.999146	0.834753
3	−0.120688	−5.839854	−4.462392	4.058722	7.623522	−3.920468	1.161365
4	−0.128138	−5.677780	−5.311812	3.948922	8.212628	−4.271338	1.396446
5	−0.045556	−5.503278	−5.343263	3.734825	8.495700	−4.759689	1.299284
6	−0.004081	−5.410887	−5.359281	3.683508	8.567112	−4.955557	1.363882
7	0.022480	−5.530671	−5.533885	3.619361	8.909204	−5.234465	1.362800
8	0.128673	−4.844708	−4.092068	2.524164	8.506729	−5.877084	1.221500
9	0.167657	−4.748124	−2.651324	2.277804	7.595546	−5.472440	1.133606
10	0.223695	−4.883140	−0.264139	1.147574	6.954241	−4.970569	1.003486
11	0.188877	−5.126976	1.722286	−0.167161	6.562104	−4.492904	1.145750
12	0.158483	−5.245658	4.203833	−1.756774	6.306062	−4.410647	1.264555

Table A29
Fitting coefficients for ThinClouds sky type [Eq. (4)] and KG climate D.

AM range	d_0	d_1	d_2	d_3	d_4	d_5
1	−6.218078	−4.827325	7.610705	6.322404	−4.024633	0.207216
2	−6.039574	−6.178465	7.940703	7.036517	−4.584531	0.351214
3	−6.410604	−5.331730	7.382290	7.193545	−4.400056	0.700683
4	−5.761065	−5.656046	6.512402	7.485817	−4.858513	0.846876
5	−6.086194	−8.210882	8.106639	8.515616	−5.044023	1.081227
6	−6.509036	−8.338712	7.860601	9.293600	−5.303562	1.387318
7	−6.865833	−9.170205	7.889388	10.303843	−5.824571	1.558770
8	−5.292107	−7.290715	5.020210	9.346069	−5.857776	1.676263
9	−5.050206	−6.797729	5.210597	8.320753	−5.168793	1.674146
10	−4.483523	−3.981779	2.577139	7.562246	−5.285585	1.663249
11	−3.645315	−1.291184	0.637713	5.934156	−4.681639	1.541377
12	−3.697387	3.070747	−2.429523	4.548576	−3.060924	1.486504

Table A30
Fitting coefficients for Cloudless sky type [Eq. (5)] and KG climate D.

AM range	f_0	f_1	f_2	f_3	f_4
1	−5.749120	−0.711563	3.127666	6.337195	−4.292969
2	−5.133027	−3.525413	6.113634	5.611610	−4.450581
3	−4.416700	−4.820633	5.578485	6.356645	−5.070370
4	−3.491407	−4.810835	5.284668	5.691578	−5.358944
5	−2.706180	−5.155929	5.052605	5.333276	−5.525264
6	−2.314904	−4.953869	5.320800	4.500941	−5.077293
7	−1.921049	−4.822490	4.581314	4.455506	−5.086411
8	−0.801643	1.863966	−1.822077	3.023794	−4.557968
9	−0.962562	−3.376612	3.499321	3.028858	−4.385917
10	−1.245521	−2.590770	2.885509	3.138688	−4.341684
11	−1.577581	−1.270911	1.733716	3.246349	−4.201945
12	−1.719818	1.188344	−0.850214	3.490872	−4.541389

Table A31
Fitting coefficients for CloudEn sky type [Eq. (6)] and KG climate D.

AM range	g_0	g_1	g_2	g_3	g_4
1	1.529366	2.223611	2.022348	−3.015128	−2.821832
2	2.058481	0.307109	1.707324	−1.803391	−3.079369
3	2.248664	−1.617210	1.437200	−0.372957	−3.492665
4	2.088567	−1.958899	1.260945	0.023365	−3.365999
5	2.015395	−1.857157	0.876420	0.091931	−3.083857
6	1.828915	−1.003302	0.661193	−0.384269	−2.527114
7	1.864879	−1.006226	0.199693	−0.199133	−2.573054
8	1.578219	−1.262806	−0.027981	0.185791	−2.382402
9	1.205854	1.043044	−1.190113	−0.705823	−1.082501
10	–	–	–	–	–
11	–	–	–	–	–
12	–	–	–	–	–

Table A32
Fitting coefficients for Overcast sky type [Eq. (1)] and KG climate E.

AM range	a_0	a_1
1	−0.757616	0.997394
2	−1.341361	0.998606
3	−0.590629	0.956902
4	−0.799572	0.986427
5	−0.962698	0.992962
6	−1.171474	0.998864
7	−1.261014	1.003373
8	−0.657147	0.987218
9	−0.794853	0.989207
10	−0.805722	0.984761
11	−0.991472	0.991209
12	−1.047819	0.993380

Table A33
Fitting coefficients for ThickClouds sky type [Eq. (2)] and KG climate E.

AM range	b_0	b_1
1	−0.071458	0.991764
2	−1.326152	0.995313
3	3.420097	0.938381
4	0.900369	0.974119
5	0.525748	0.977258
6	1.278616	0.971584
7	1.083440	0.965918
8	4.936016	0.895050
9	4.411426	0.879212
10	5.260956	0.833842
11	2.963451	0.859745
12	4.030433	0.787141

combined (GISPLITv1, or simply, G1), and another one for which the training dataset is split by primary KG climate class, with separate training for each class (GISPLITv2, or G2). Additionally, this study also explored to what extent machine-learning-based modeling could lead to perceptible performance improvements. To that aim, the GHI separation for the ScatterClouds and CloudEn sky types was also modelled with

Table A34
Fitting coefficients for ScatterClouds sky type [Eq. (3)] and KG climate E.

AM range	c_0	c_1	c_2	c_3	c_4	c_5	c_6
1	−0.268041	−7.325412	−0.091782	2.749135	5.371846	−0.872566	−0.207321
2	−0.444444	−7.519714	−2.201808	3.187409	7.167350	−4.200452	1.104740
3	−0.175905	−8.432679	−7.571224	4.631128	12.026985	−6.653413	0.904714
4	−0.146528	−7.620146	−7.672747	4.600271	11.324955	−5.631221	1.284436
5	−0.030430	−7.717333	−8.476633	4.839165	12.093732	−5.464887	1.223900
6	0.030323	−7.691183	−8.646968	5.293058	12.034121	−5.393757	1.234268
7	0.118417	−7.568815	−7.755847	4.806152	11.852932	−5.573987	1.039632
8	0.161240	−6.692281	−8.055880	4.397235	11.875243	−6.800156	0.961055
9	0.117201	−6.205669	−5.529691	2.933386	10.268320	−6.451235	1.142749
10	0.174165	−5.814552	−4.676779	2.536060	9.791834	−6.955263	0.985212
11	0.226881	−5.623031	−5.977946	2.475494	10.443147	−6.954748	0.871286
12	0.111223	−5.636834	−4.545690	1.612903	9.246084	−5.553912	1.188826

Table A35
Fitting coefficients for ThinClouds sky type [Eq. (4)] and KG climate E.

AM range	d_0	d_1	d_2	d_3	d_4	d_5
1	−41.182625	7.368646	33.580493	10.525168	−13.243563	1.141325
2	−29.504738	−22.090856	19.893241	34.520487	−12.018847	1.499117
3	−21.376182	−34.738389	25.058939	32.933280	−14.619420	1.528826
4	−13.778874	−18.564047	13.489309	20.660059	−9.412399	1.357672
5	−12.039132	−16.983537	12.801617	17.598133	−6.347883	1.416617
6	−9.288444	−13.470698	9.063047	15.008905	−6.498662	1.494311
7	−7.486830	−14.220383	8.473732	14.262746	−7.125291	1.617409
8	−6.858541	−11.740913	6.870406	12.717736	−6.693453	1.619429
9	−5.935080	−8.278886	5.248102	10.164204	−5.800602	1.558646
10	−4.941777	−7.024413	3.966606	9.142866	−6.172536	1.526082
11	−4.356211	−3.796456	1.398837	7.884246	−5.745212	1.516870
12	−3.844812	−1.600629	−0.464688	6.955156	−5.526584	1.492556

Table A36
Fitting coefficients for Cloudless sky type [Eq. (5)] and KG climate E.

AM range	f_0	f_1	f_2	f_3	f_4
1	1.109878	12.364196	−7.976018	−1.829031	−4.922972
2	−0.884304	−5.180875	3.911256	5.052965	−9.009687
3	−2.559209	3.094445	−4.724859	6.635537	−7.002750
4	−2.101670	−5.013375	3.397050	6.105448	−6.944501
5	−1.138660	−1.741118	0.435424	4.651126	−5.680600
6	−2.769980	−5.687657	4.148993	6.377312	−5.890869
7	−1.895863	−5.642230	4.082633	5.419560	−5.629021
8	−3.182559	−8.234161	6.586199	6.610879	−5.517753
9	−3.675885	−9.760173	8.377863	6.748899	−5.292858
10	−3.479839	−8.053492	6.398817	6.735294	−5.598072
11	−3.705306	−9.675267	8.037446	6.901438	−5.748937
12	−3.290543	−8.066958	6.070343	6.770991	−6.227049

Table A37
Fitting coefficients for CloudEn sky type [Eq. (6)] and KG climate E.

AM range	g_0	g_1	g_2	g_3	g_4
1	2.415248	−5.073501	2.094280	1.570622	−1.995205
2	2.874481	−4.422721	1.892312	1.075928	−5.688895
3	2.698923	−5.843026	1.411397	2.659266	−6.028276
4	2.380156	−3.571714	1.110499	1.163898	−5.009853
5	2.097730	−2.827965	0.467203	1.067480	−4.116028
6	2.119989	−2.621913	0.384222	0.852266	−3.806488
7	2.052495	−1.230673	−0.311302	0.263087	−3.689039
8	1.815141	−0.813173	−0.642812	0.167811	−2.796681
9	1.669848	0.275319	−1.259463	−0.174442	−2.667829
10	–	–	–	–	–
11	–	–	–	–	–
12	–	–	–	–	–

gradient-boosting algorithms, using alternately the entire training dataset combined (GISPLITv3, or G3), or the climate class separation avenue (GISPLITv4, or G4).

The huge observational database that has been compiled maximizes the robustness and significance of the results. It is made up of 120 ground stations with ≈64 million quality-assured data points, spanning different geographical regions and climate conditions worldwide. This huge database has multi-year time extension, and includes different high-performance radiometers and acquisition methods.

The thorough validation that has been conducted, including large datasets not used for model development, has shown a remarkable consistency of results for each of the four GISPLIT versions. The climate conditioning of the model provides only marginal performance improvement, thus discouraging such usage given the higher complexity. Generally, the machine-learning modeling approach provides more important accuracy benefits. The validation results presented here suggest that GISPLITv3 generally provides the best RMSE results at unseen sites. Additional analyses have already been made to evaluate the performance of the four GISPLIT versions in various other ways, and to compare them to other prominent models of the literature. These results confirm the superior performance of GISPLIT and will be detailed in a forthcoming publication.

Overall, this work has proposed a novel approach to perform the separation of GHI into its components. It is based on sky-type conditioning and only requires 1-min GHI observations, solar position, and estimates from a clear-sky solar irradiance model. Moreover, the sky-type classification method enables the calculation of confidence intervals, which might be useful in the process of evaluating the uncertainty in derived irradiance products. Nevertheless, the methodology adopted here is still data driven, and the functional forms selected for the driving mechanisms tend to amalgamate different atmospheric phenomena that are not necessarily related. Hence, future contributions should explore the use of more physically-based approaches.

CRediT authorship contribution statement

José A. Ruiz-Arias: Conceptualization, Data curation, Model development, Data analysis, Writing. **Christian A. Gueymard:** Conceptualization, Writing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the project PID2019-107455RB-C21 funded by MCIN/AEI/ 10.13039/501100011033, the project UMA20-FEDERJA-134 jointly funded by the FEDER 2014-2020 Operative Program and the Consejería de Economía, Conocimiento, Empresas y Universidad of the Junta de Andalucía, and by Solargis s.r.o. through the collaboration agreement 2021-124 with the University of Málaga. The authors would like to thank the scientists and personnel in charge of the BSRN stations for acquiring, processing and kindly sharing their datasets, which have been central to this study. Moreover, the authors acknowledge the scientists and personnel of the Global Modelling and Assimilation Office at NASA Goddard Space Flight Center who provided the MERRA-2 atmospheric data that were advantageously used to calculate the clear-sky solar irradiance at all sites. This work has been stimulated in great part by the authors' participation to Task 16 of the International Energy Agency's Photovoltaic Power Systems Programme. The other Task participants were instrumental in developing the robust quality-control algorithm prominently used here to improve the measured irradiance databases. The University of Málaga/CBUA provided the funding for open access.

Appendix A

GISPLIT model parameters

The following tables present the model parameters corresponding to Eqs. (1–6) for all climates combined (Tables A2-A7), as well as for each KG climate separately, according to classes A (Tables A8-A13), B (Tables A14-A19), C (Tables A20-A25), D (Tables A26-A31), and E (Tables A32-A37). Table A1 shows the relative air mass fitting ranges.

GISPLITv1 is fully described by the parameters in Tables A2-A7, whereas GISPLITv2 requires Tables A8-A37 to cover all possible KG primary climates. GISPLITv3 uses the same parameters (i.e., tables) as GISPLITv1 except those in Tables A4 and A7, which are replaced by extreme gradient-boosting model fits. Similarly, GISPLITv4 uses the same parameters as GISPLITv3, with the exception of those corresponding to the ScatterClouds and CloudEn sky types for all KG climates, which are substituted by extreme gradient-boosting model fits. The extreme gradient-boosting files are too large to be listed here, but they can be made available upon request.

Appendix B. Supplementary data

The GISPLIT parameters shown in Appendix A are also provided here for convenience in "json" (JavaScript Object Notation) files. Supplementary data to this article can be found online at https://doi.org/10.1016/j.solener.2024.112363.

References

[1] R. Perez, T. Cebecauer, M. Suri, Semi-empirical satellite models, in: J. Kleissl (Ed.), Sol. Energy Forecast. Resour, Assess., Elsevier, 2013.
[2] C.A. Gueymard, J.A. Ruiz-Arias, Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance, Sol. Energy 128 (2016) 1–30, https://doi.org/10.1016/j.solener.2015.10.010.
[3] D. Yang, Estimating 1-min beam and diffuse irradiance from the global irradiance: a review and an extensive worldwide comparison of latest separation models at 126 stations, Renew. Sustain. Energy Rev. 159 (2022) 112195, https://doi.org/10.1016/j.rser.2022.112195.
[4] C.A. Gueymard. Solar radiation resource: measurement, modeling, and methods. In: Letcher TM, editor. Compr. Renew. Energy 2nd Ed., vol. 1, Oxford: Elsevier; 2022.
[5] M. Sengupta, A. Habte, S. Wilbert, C. Gueymard, J. Remund, Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Third Edition. (2021), https://doi.org/10.2172/1778700.
[6] J.A. Ruiz-Arias, C.A. Gueymard. Solar resource for high-concentrator photovoltaic applications. In: Pérez-Higueras P, Fernández EF, editors. High Conc. Photovolt. Fundam. Eng. Power Plants, 2015, p. 261–302. https://doi.org/doi: 10.1007/978-3-319-15039-0_10.
[7] C.A. Gueymard, Uncertainties in modeled direct irradiance around the sahara as affected by aerosols: are current datasets of bankable quality? J. Sol. Energ-T ASME 133 (2011) https://doi.org/10.1115/1.4004386.
[8] C.A. Gueymard, J.A. Ruiz-Arias, Validation of direct normal irradiance predictions under arid conditions: a review of radiative models and their turbidity-dependent performance, Renew. Sust. Energy Rev. 45 (2015) 379–396.
[9] J.A. Ruiz-Arias, C.A. Gueymard, F.J. Santos-Alamillos, D. Pozo-Vázquez, Worldwide impact of aerosol's time scale on the predicted long-term concentrating solar power potential, Sci. Rep. 6 (2016) 30546, https://doi.org/10.1038/srep30546.
[10] J.A. Ruiz-Arias, Aerosol transmittance for clear-sky solar irradiance models: Review and validation of an accurate universal parameterization, Renew. Sustain. Energy Rev. 145 (2021) 111061, https://doi.org/10.1016/j.rser.2021.111061.
[11] C.A. Gueymard, Temporal variability in direct and global irradiance at various time scales as affected by aerosols, Sol. Energy 86 (12) (2012) 3544–3553.
[12] J.A. Ruiz-Arias, C.A. Gueymard, F.J. Santos-Alamillos, D. Pozo-Vázquez, Do spaceborne aerosol observations limit the accuracy of modeled surface solar irradiance?: Aerosol limits modeled solar radiation, Geophys. Res. Lett. 42 (2015) 605–612, https://doi.org/10.1002/2014GL062309.
[13] C.A. Gueymard, Direct and indirect uncertainties in the prediction of tilted irradiance for solar engineering applications, Sol. Energy 83 (3) (2009) 432–444.
[14] M. Hofman, G. Seckmeyer, Influence of various irradiance models and their combination on simulation results of photovoltaic systems, Energies 10 (2017) 1495, https://doi.org/10.3390/en10101495.
[15] M.J. Mayer, G. Gróf, Extensive comparison of physical models for photovoltaic power forecasting, Appl. Energy 283 (2021) 116239, https://doi.org/10.1016/j.apenergy.2020.116239.
[16] E.F.M. Abreu, P. Canhoto, M.J. Costa, Prediction of diffuse horizontal irradiance using a new climate zone model, Renew. Sustain. Energy Rev. 110 (2019) 28–42, https://doi.org/10.1016/j.rser.2019.04.055.
[17] R. Blaga, The impact of temporal smoothing on the accuracy of separation models, Sol. Energy 191 (2019) 371–381, https://doi.org/10.1016/j.solener.2019.08.078.
[18] J.P. Every, L. Li, D.G. Dorrell, Köppen-Geiger climate classification adjustment of the BRL diffuse irradiation model for Australian locations, Renew. Energy 147 (2020) 2453–2469, https://doi.org/10.1016/j.renene.2019.09.114.
[19] A.R. Starke, L.F.L. Lemos, C.M. Barni, R.D. Machado, J.M. Cardemil, J. Boland, S. Colle, Assessing one-minute diffuse fraction models based on worldwide climate features, Renew. Energy 177 (2021) 700–714.
[20] R. Blaga, D. Calinoiu, N. Stefu, R. Boata, A. Sabadus, E. Paulescu, N. Pop, O. Mares, S. Bojin, M. Paulescu, Quantification of the aerosol-induced errors in solar irradiance modeling, Meteorol. Atmospheric Phys. 133 (4) (2021) 1395–1407.
[21] M. Boraiy, M. Korany, Y. Aoun, S.C. Alfaro, M. El-Metwally, M.M. Abdel Wahab, P. Blanc, Y. Eissa, H. Ghedira, G. Siour, K. Hungershoefer, L. Wald, Improving direct normal irradiance retrieval in cloud-free, but high aerosol load conditions by using aerosol optical depth, Meteorol. Z. 26 (5) (2017) 475–483.
[22] C. Gueymard, F. Vignola, Determination of atmospheric turbidity from the diffuse-beam broadband irradiance ratio, Sol. Energy 63 (3) (1998) 135–146.
[23] C.A. Gueymard, Cloud and albedo enhancement impacts on solar irradiance using high-frequency measurements from thermopile and photodiode radiometers. Part 1: impacts on global horizontal irradiance, Sol. Energy 153 (2017) 755–765, https://doi.org/10.1016/j.solener.2017.05.004.
[24] M. Braga, A.K.V. de Oliveira, L. Burnham, S. Dittmann, R. Gottschalg, T. Betts, et al., Solar over-Irradiance Events: Preliminary Results from a Global Study (2020) 2764–2770, https://doi.org/10.1109/PVSC45281.2020.9300868.
[25] W.B. Mol, W.H. Knap, C.C. van Heerwaarden, Ten years of 1 Hz solar irradiance observations at Cabauw, the Netherlands, with cloud observations, variability classifications, and statistics, Earth Syst. Sci. Data Discuss. 2023 (2023) 1–19, https://doi.org/10.5194/essd-2022-456.
[26] A. Castillejo-Cuberos, R. Escobar, Detection and characterization of cloud enhancement events for solar irradiance using a model-independent, statistically-driven approach, Sol. Energy 209 (2020) 547–567, https://doi.org/10.1016/j.solener.2020.09.046.
[27] C.A. Gueymard, Cloud and albedo enhancement impacts on solar irradiance using high-frequency measurements from thermopile and photodiode radiometers. Part 2: performance of separation and transposition models for global tilted irradiance, Sol. Energy 153 (2017) 766–779, https://doi.org/10.1016/j.solener.2017.04.068.
[28] A.R. Starke, L.F. Lemos, J. Boland, J.M. Cardemil, S. Colle, Resolution of the cloud enhancement problem for one-minute diffuse radiation prediction, Ren. Energy 125 (2018) 472–484.

[29] J.A. Ruiz-Arias, C.A. Gueymard, CAELUS: Classification of sky conditions from 1-min time series of global solar irradiance using variability indices and dynamic thresholds, Sol. Energy 263 (2023) 111895, https://doi.org/10.1016/j.solener.2023.111895.
[30] A.J. Ruiz-Arias, C.A. Gueymard. CAELUS: Classification of sky conditions from 1-min time series of global solar irradiance using variability indices and dynamic thresholds 2023. https://doi.org/10.5281/ZENODO.7897639.
[31] M. Oh, C.K. Kim, B. Kim, C. Yun, J.-Y. Kim, Y. Kang, H.-G. Kim, Analysis of minute-scale variability for enhanced separation of direct and diffuse solar irradiance components using machine learning algorithms, Energy 241 (2022) 122921.
[32] R. Aler, I.M. Galván, J.A. Ruiz-Arias, C.A. Gueymard, Improving the separation of direct and diffuse solar radiation components using machine learning by gradient boosting, Sol. Energy 150 (2017) 558–569, https://doi.org/10.1016/j.solener.2017.05.018.
[33] T. Chen, C. Guestrin. XGBoost: A Scalable Tree Boosting System. Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., San Francisco California USA: ACM; 2016, p. 785–94. https://doi.org/10.1145/2939672.2939785.
[34] A. Driemel, J. Augustine, K. Behrens, S. Colle, C. Cox, E. Cuevas-Agulló, F.M. Denn, T. Duprat, M. Fukuda, H. Grobe, M. Haeffelin, G. Hodges, N. Hyett, O. Ijima, A. Kallis, W. Knap, V. Kustov, C.N. Long, D. Longenecker, A. Lupi, M. Maturilli, M. Mimouni, L. Ntsangwane, H. Ogihara, X. Olano, M. Olefs, M. Omori, L. Passamani, E.B. Pereira, H. Schmithüsen, S. Schumacher, R. Sieger, J. Tamlyn, R. Vogt, L. Vuilleumier, X. Xia, A. Ohmura, G. König-Langlo, Baseline surface radiation network (BSRN): structure and data description (1992–2017), Earth Syst. Sci. Data 10 (3) (2018) 1491–1501.
[35] A. Forstinger, Y-M. Saint-Drenan, S. Wilbert, A. Jensen, B. Krass, C. Fernández Peruchena, et al. IEA-PVPS Task-16 Reference Solar Measurements 2021. https://doi.org/10.23646/3491B1A6-E32D-4B34-9DBB-EE0AFFE49E36.
[36] D. Schüler, S. Wilbert, N. Geuder, R. Affolter, F. Wolfertstetter, C. Prahl, et al., The enerMENA meteorological network – solar radiation measurements in the MENA region, Cape Town, South, Africa (2016) 150008, https://doi.org/10.1063/1.4949240.
[37] H.E. Beck, N.E. Zimmermann, T.R. McVicar, N. Vergopolan, A. Berg, E.F. Wood, Present and future Köppen-Geiger climate classification maps at 1-km resolution, Sci. Data 5 (2018) 180214, https://doi.org/10.1038/sdata.2018.214.
[38] J.A. Ruiz-Arias, SPARTA: solar parameterization for the radiative transfer of the cloudless atmosphere, Renew. Sustain. Energy Rev. 188 (2023) 113833, https://doi.org/10.1016/j.rser.2023.113833.
[39] C.A. Gueymard, REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation – validation with a benchmark dataset, Sol. Energy 82 (2008) 272–285, https://doi.org/10.1016/j.solener.2007.04.008.
[40] M. Lefèvre, 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, J.-J. Morcrette, McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions, Atmospheric Meas Tech 6 (9) (2013) 2403–2418.
[41] R. Gelaro, W. McCarty, M.J. Suárez, R. Todling, A. Molod, L. Takacs, C.A. Randles, A. Darmenov, M.G. Bosilovich, R. Reichle, K. Wargan, L. Coy, R. Cullather, C. Draper, S. Akella, V. Buchard, A. Conaty, A.M. da Silva, W. Gu, G.-K. Kim, R. Koster, R. Lucchesi, D. Merkova, J.E. Nielsen, G. Partyka, S. Pawson, W. Putman, M. Rienecker, S.D. Schubert, M. Sienkiewicz, B. Zhao, The modern-era retrospective analysis for research and applications, version 2 (MERRA-2), J. Clim. 30 (14) (2017) 5419–5454.
[42] A. Inness, M. Ades, A. Agustí-Panareda, J. Barré, A. Benedictow, A.-M. Blechschmidt, J.J. Dominguez, R. Engelen, H. Eskes, J. Flemming, V. Huijnen, L. Jones, Z. Kipling, S. Massart, M. Parrington, V.-H. Peuch, M. Razinger, S. Remy, M. Schulz, M. Suttie, The CAMS reanalysis of atmospheric composition, Atmospheric Chem. Phys. 19 (6) (2019) 3515–3556.
[43] N.A. Engerer, Minute resolution estimates of the diffuse fraction of global irradiance for southeastern Australia, Sol. Energy 116 (2015) 215–237, https://doi.org/10.1016/j.solener.2015.04.012.
[44] D. Yang. Temporal-resolution cascade model for separation of 1-min beam and diffuse irradiance. J Renew Sustain Energy 2021;13:056101. https://doi.org/10.1063/5.0067997.
[45] C.A. Gueymard, Direct solar transmittance and irradiance predictions with broadband models. Part I: detailed theoretical performance assessment, Sol. Energy 74 (2003) 355–379, https://doi.org/10.1016/S0038-092X(03)00195-6.
[46] M.J. Blanco, K. Milidonis, A.M. Bonanos, Updating the PSA sun position algorithm, Sol. Energy 212 (2020) 339–341, https://doi.org/10.1016/j.solener.2020.10.084.
[47] C.A. Gueymard, A review of validation methodologies and statistical performance indicators for modeled solar radiation data: towards a better bankability of solar projects, Renew. Sustain. Energy Rev. 39 (2014) 1024–1034, https://doi.org/10.1016/j.rser.2014.07.117.
[48] X. Sun, J.M. Bright, C.A. Gueymard, X. Bai, B. Acord, P. Wang, Worldwide performance assessment of 95 direct and diffuse clear-sky irradiance models using principal component analysis, Renew. Sustain. Energy Rev. 135 (2021) 110087, https://doi.org/10.1016/j.rser.2020.110087.