sciendo

ANNALS OF WEST UNIVERSITY OF TIMISOARA PHYSICS
Vol. LXIV, 2022

DOI : 10.2478/awutp-2022-0007

BIAS-CORRECTION OF DIFFUSE SOLAR IRRADIANCE MODELED THROUGH PARAMETRIC MODELS

Robert Blaga

Faculty of Physics, West University of Timisoara, V. Parvan 4, 300223 Timisoara, Romania

Article Info

Received: 15.08.2022
Accepted: 09.09.2022

Keywords:
clear-sky models; diffuse solar irradiance; high turbidity; aerosol multiple scattering; model output statistics

Abstract

In this paper, a bias-correction of clear-sky models in prediction of the diffuse solar irradiance is presented. High quality ground date from two networks is used for model input and validation: aerosol optical depth from AERONET and diffuse solar irradiance from BSRN. Four established clear-sky models are tested: REST2, Yang, PS and MAC2. Two types of bias-corrections are developed: i) a correction for multiple aerosol scattering (MSC), accounting for extinction at high turbidity, and ii) a general regression on the model errors as a function of Sun elevation angle (h) and turbidity (β) correction (MOS). While the impact of MSC proves marginal, the MOS correction shows good results. Two versions of MOS are developed, one on the global dataset and a site-specific adaptation. Both versions bring significant improvements over the original implementation of the models, while the site-specific version reduces the errors with a few additional percentage points. The aggregate nRMSE of the REST2 model is reduced from 17.8% (no correction) to 11.3% (global MOS) and 9.4% (site-specific MOS).

1. Introduction

As solar energy becomes increasingly popular and accessible, the implementation of solar projects in ever newer geographies and micro-climates is becoming a reality. The planning and operation of PV plants requires accurate predictions for short- and long-term energy production. Lacking high spatio-temporal resolution and accuracy solar irradiance data with worldwide coverage, the employment of estimation models remains an essential tool in the planning stage of such projects.

The solar energy reaching ground level can be obtained by solving the equations for radiative transfer through the atmosphere, given that the atmospheric composition and the extraterrestrial irradiance are known. As the extinction processes due to the various atmospheric species are independent from each other, a general solution for the radiative

transfer of a spectral solar flux from the top of the atmosphere to the ground, can be written as:

$$G_b(m, l_i, \lambda) = G_0(m, \lambda) \prod_i [\tau_i(m, l_i, \lambda)] \ \tau_R(m, \lambda) \tau_a(m, \alpha, \beta, \lambda) \quad (1)$$

where G_0 is the extraterrestrial irradiance, m and λ are the optical air mass and the wavelength of the radiation, τ_R and τ_a are the atmospheric transmittances for Rayleigh scattering and aerosols, and the index *i* covers absorption by all remaining atmospheric species. The quantities l_i represent columnar values of the various atmospheric components. In the absence of clouds, after Rayleigh and aerosol scattering, the most relevant atmospheric species that contribute to the attenuation of a solar beam are water vapor and ozone. Sometimes NO_2 and the collective impact of trace gases are also considered.

Approximate transmittance functions for Rayleigh scattering and aerosol extinction can be obtained through analytical reasoning. For all other species, the transmittances are not analytical functions, taking non-zero values where absorption lines/bands exist. In practice, spectral averaging is often employed in order to make the transmittances easier to work with. The most common approach is to replace the transmittance functions in Eq. (1) with their spectrally averaged (*broadband*) counterparts, as follows:

$$\bar{\tau}_i(m, l_i) = \int \tau_i(m, l_i, \lambda) G_0(m, \lambda) d\lambda \, / \int G_0(m, \lambda) d\lambda, \quad (2)$$

and using a fitting function to approximate the integrals. The broadband beam component of a solar beam is then written the same as Eq. (1), but without the spectral dependence. It must be noted that in passing from Eq. (1) to a broadband irradiance through Eq. (2), physicality is lost. Nonetheless, broadband or *parametric* models have shown good accuracy in modeling the solar irradiance under various scenarios [1].

As the solar beam propagates through the atmosphere, scattering and absorption by the atmospheric components occur, collectively called *extinction*. Eq. (1) describes the direct or *beam* solar irradiance that arrives at the ground after undergoing all extinction processes. The second component of solar irradiance that reaches the ground is the *diffuse* irradiance that arrives at a given point indirectly, through various scattering processes in the atmosphere. This component is much more difficult to model, because it comprises many different phenomena which do not stack multiplicatively. Furthermore, the solution of the radiative transfer equation needs to be integrated over all solid angles (while the beam component propagates only in the direction of the Sun). Thus, for broadband modelling an ad hoc

expression for the diffuse sector is used. Generally, the diffuse component can be modelled as:

$$G_b = G_0 \prod_i(\tau_i) \times D \tag{3}$$

where the function D captures the effects of aerosol and Rayleigh scattering. In the simplest approaches to modeling the scattering sector, it is simply modeled as the scattered component of the initial beam, multiplied by the downward scattering factor γ:

$$D = \gamma\,(1 - \tau_R \tau_a) \tag{4}$$

More complex models consider separately the Rayleigh and aerosol scattering, and can include other aerosol optical parameters, like the single scattering albedo (ϖ) and the aerosol asymmetry factor (g).

$$D = \gamma_a\, D_a + \gamma_R\, D_R \tag{5}$$

The extinction by aerosols is usually modelled through the Ångström equation [2]:

$$\tau_a = \exp(-m\beta\lambda^{-\alpha}) \tag{6}$$

where α and β are called the Ångström exponent and turbidity, respectively—collectively, Ångström parameters—while λ is the wavelength of the radiation expressed in μm.

Under a clear sky, aerosols are the main attenuating factor. As such, inaccuracy in modeling the aerosol sector translates into the largest errors in the modeled solar irradiance. Blaga et. al. (2021) have explored the way in which the modeling errors naturally split into intrinsic model errors and aerosol modeling errors. At high aerosol loading, especially, the accuracy of parametric models breaks down [3], potentially due to spectral effects not captured by the and other effects [4].

In this section we study the use of different bias-correction methods to improve the accuracy of clear-sky models in estimating the diffuse solar irradiance. In Sec. 2 the working database and the used model set is presented. Two bias-correction procedure are tested, one accounting for multiple aerosol scattering at high turbidity, and the second a general regression-based approach that eliminates the mean bias of the models. The results are presented in Sec. 3, by comparing the original models with the improved models using a

global and a site-specific bias-correction, respectively. The results are summarized and contextualized in the concluding section.

2. Data and methods

In this section we present the working database, which is built by merging data from two ground-based data repositories. The merging algorithm is tuned to select cvasi-cloudless days, which permit the isolation of the extinction through the effects of aerosols. Four clear-sky models are chosen from the literature to test the impact of bias-correction. Two corrections are implemented: one developed on the whole dataset and one specific for each location.

2. 1. Data

The dataset used in this study is comprised of measured data from two ground-based networks, the Aerosol Robotic Network (AERONET) [5] and the Baseline Surface Radiation Network (BSRN) [6]. Data from the two networks is of the highest presently available quality and often considered ground truth, used as reference for validating other models [7, 8]. Aerosol optical depth (AOD) and other columnar data from AERONET are used here as input for parametric models, and diffuse solar irradiance (DIF) from BSRN is used for validation purposes. Data from stations that are included in both networks are used to build the working dataset.

The BSRN data has a 1-min resolution, with relatively few gaps. Additional quality controls are applied on the raw data as described in Refs. [9, 10]. A screening for cvasi-clear days is performed using the indicator I introduced in Ref. [11]. The resulting dataset is merged with the AERONET data. The latter has a 15-min resolution and measurements are present only when clouds do not obscure the Sun. The lvl.2 version of the data, which is used here, benefits from an additional, manual screening for clouds. The full set of operations for merging the two data sources is described in detail in Ref. [12]. The combined dataset contains 289335 entries from 12 locations spread out on 5 continents. The working dataset is summarized in Table 1 and is available on request from the author. The diffuse solar irradiance as a function of aerosol optical depth is shown in Fig. 1a.

Table 1. Summary of data included in this study. N is the number of data entries for each station and the average β is reported across the whole (multi-year) dataset.

No	Location (Country)	BSRN index	BSRN station			Climate	*N*	β_{mean}	β_{max}
			Lat. [°]	Long. [°]	Alt. [m]				
1	Boulder (USA)	BOS	40.13	-105.24	1689	BSk	15506	0.023	0.310
2	Cener (Spain)	CNR	42.82	-1.60	471	Cfb	8404	0.036	0.499
3	Darwin (Australia)	DAR	-12.43	130.89	30	Aw	8141	0.054	0.214
4	Fukuoka (Japan)	FUA	33.58	130.38	3	Cfa	11649	0.133	0.887
5	Gobabeb (Namibia)	GOB	-23.56	15.04	407	Bwh	77952	0.040	0.436
6	Izana (Spain)	IZA	28.31	-16.50	2373	Csb	54441	0.043	1.065
7	Langley Research Center (USA)	LRC	37.10	-76.39	3	Cfa	9314	0.030	0.462
8	Palaiseau (France)	PAL	48.713	2.21	156	Cfb	9476	0.054	0.514
9	Sioux Falls (USA)	SFX	43.730	-96.62	473	Dfa	6153	0.036	0.223
10	Tamanrasset (Algeria)	TAM	22.790	5.53	1385	Bwh	16300	0.109	1.382
11	Toravere (Estonia)	TOR	58.254	26.46	70	Dfb	6606	0.055	0.839
12	Xianghe (China)	XIA	39.754	116.96	32	Dwa	15674	0.278	1.744

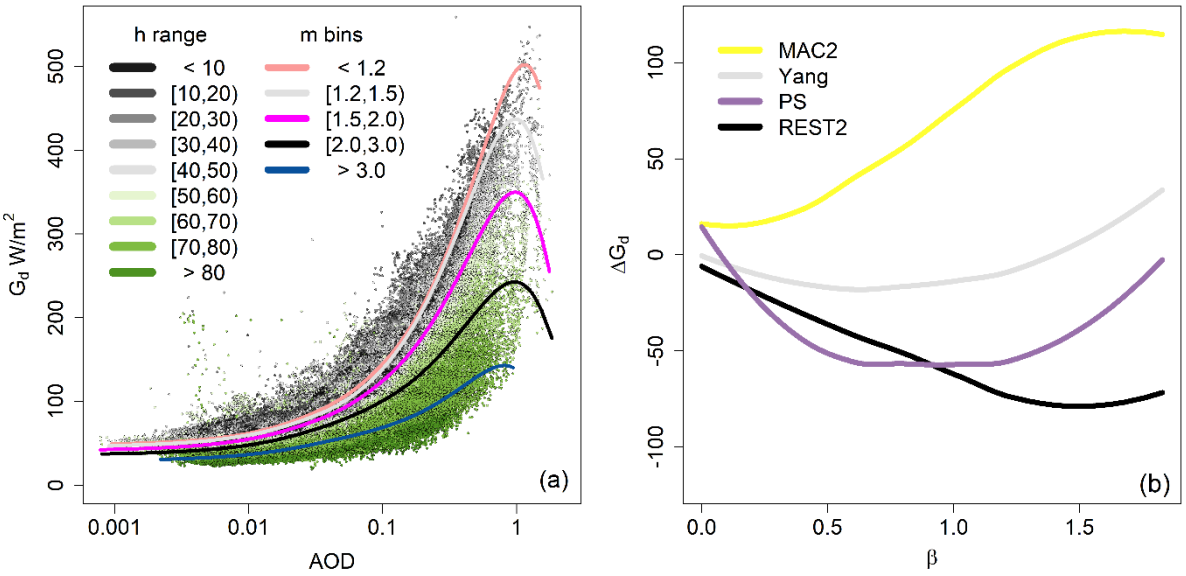


Figure 1. (a) DIF as a function of AOD. The range that the Sun elevation angle (h) takes is color coded in shades ranging from black to green. Average DIF for bins of air mass (m) are also represented. (b) DIF bias of four clear-sky models as a function of Ångström turbidity (β).

2. 2. Clear-sky models

The REST2 model is universally recognized as one of the most accurate models available in the literature [13], and a model which maintains a good performance under all atmospheric conditions [14]. The v5 version of REST2 is used here as a reference model

[15]. Two test models, the PS model developed by Paulescu and Schlett [16] and the one by Yang [17] (designated henceforth as Yang), are employed for this study. The final model which is included is the MAC2 model [18]. MAC2, with the implementation described in [13], has been found by Sun et al. [13] to be the best model in the literature (marginally surpassing REST2) for estimating global horizontal irradiance (GHI). However, MAC2 is subpar in predicting the diffuse component [19, 20].

2. 3. Bias-correction

Given that parametric models are not physical models, they will always have some degree of inaccuracy, which can be separated into random and systemic errors. The bias as a function of β for the four clear-sky models is represented in Fig. 1b. Besides MAC2, which strongly overestimates the diffuse irradiation, the other models all underestimate it. The mean bias error over the whole dataset for the four models is -12.9% (REST2), 5.6% (PS), -4.1% (Yang), and 24.1% (MAC2). The positive mean bias for PS is due to the majority of measurements clustering at small β, where the model has a small positive bias. We observe that bias for all models has one behavior up until β = 0.5 or 1, and then transition to a different behavior at high turbidity. Thus, we attempt a two-pronged bias-correction for the models.

Note that the DIF, from physical considerations, presents a local maximum as a function of β. The reasoning is as follows. For no aerosols, the DIF is small, being determined purely by Rayleigh scattering. As the turbidity increases, the diffuse component also increases. On the other hand, at very high aerosol loading both the diffuse and global irradiance are small, vanishing as the turbidity goes to infinity. Thus, a maximum point must exist as a function of β. Physically, this behavior is due to multiple scattering of aerosols in the presence of absorption. The aerosol asymmetry factor and forward scattering fraction also play a role. Without the presence of these parameters, the diffuse irradiance would keep increasing asymptotically towards a value determined by the limit $\mathbf{D}(\tau_{\mathbf{a}} \rightarrow \mathbf{0})$ in Eqs.(4)-(5) (see also Fig. 2a and the discussion below). Thus, only models that capture these effects at least to some degree have a chance of correctly modeling the DIF at high turbidity.

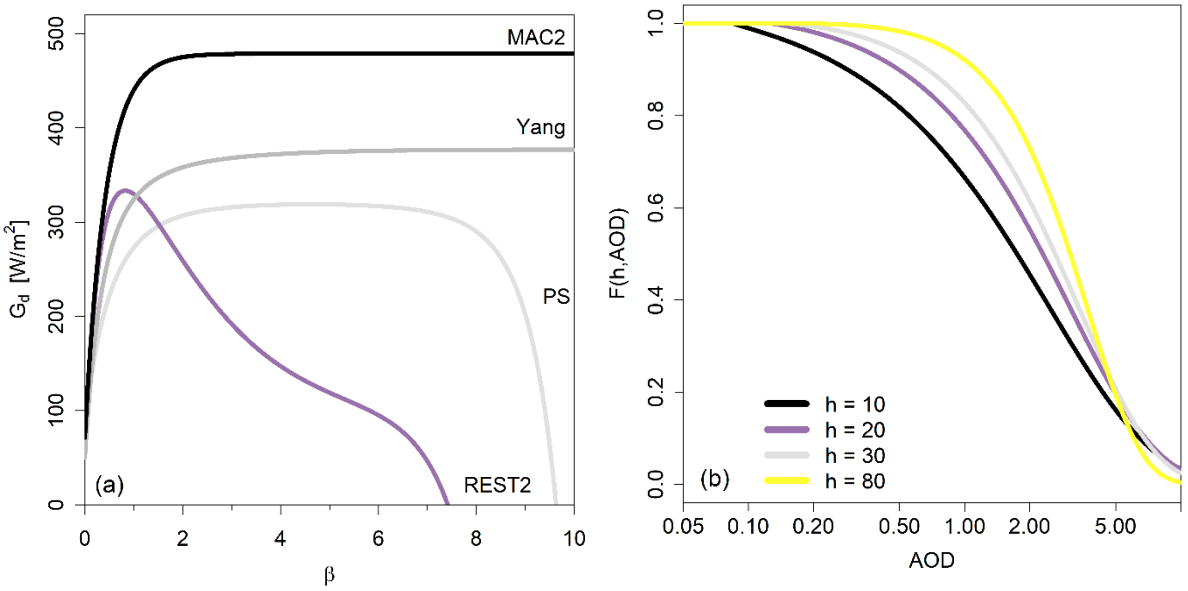


Figure 2. (a) Behavior of four clear-sky models as a function of β, at extreme aerosol loading. A local maxim of the REST2 model around β = 1 is clearly visible. (b) The multiple scattering correction as a function of AOD, at various Sun elevation angles.

As seen in Fig. 2a, REST2 is the only model from our set that captures the maximum point, with all others plateauing at a value where the direct component vanishes, and the diffuse equals the global irradiance, as estimated by each model. Multiple scattering can be taken into consideration by adding a correction factor to the downward scattering fraction. The REST2 model is based on the spectral code SMARTS2 [21, 22]. In SMARTS2, the author introduces a correction factor for aerosol multiple scattering by fitting a correction factor on measured data. The fitting function is structurally similar to the one introduced by Skartveit and Olseth (1988) for Rayleigh scattering [23]. Here, in a very coarse attempt, we test the usage of the Skartveit and Olseth (1988) correction factor, implemented *as is* for the aerosol optical depth instead of the Rayleigh one. The multiple scattering correction factor F (MSC), is expressed as follows:

$$F = \exp\left[-\left(\frac{t_a - \varepsilon}{\sigma_a}\right)^{0.72+\sin h}\right]$$
$$\sigma_a = 3.65 - 2.3\exp(-4\sin \mathrm{h}) \quad (7)$$
$$\varepsilon = 0.17[1 - \exp(-8\sin h)]$$

and multiplies the aerosol or total downward scattering fraction from Eqs. (4)-(5). In Eq.(7), t_a is the AOD. The MSC is illustrated in Fig. 2b.

Next, in order to correct for the overall bias in the models, we regression model on the modelling errors, as a function of the two most important parameters: the (sine of) the Sun elevation angle (h), and the Ångström turbidity (β). When applied to Numerical Weather Prediction outputs and other types of forecasts, this is usually called Model Output Statistics (MOS) [24, 25]. We apply here a 4th order linear regression, as implemented by Mathiesen and Kleissl (2011) [26], of the form:

$$\delta G = a_0 + a_1\beta + a_2\beta^2 + a_3\beta^3 + a_4\beta^4 + \\ + a_5 \sin h + a_6 \sin h^2 + a_7 \sin h^3 + a_8 \sin h^4 \quad (8)$$

and the bias-corrected irradiance is obtained as:

$$G_{d,c} = G_d + \delta G(h, \beta) \quad (9)$$

Due to the simple nature of the correction, the models sometimes return negative values for the solar irradiance. On these occasions the irradiance is set to zero.

3. Results

The bias correction algorithms are tested on the dataset described in Sec 2.1, containing measured AOD and DIF data from 12 collocated stations of the AERONET and BSRN networks. The dataset is split into a 25% fitting dataset and a 75% one for validation. The partition is performed for each location separately. The algorithms are fitted on the former dataset, while the latter is used to analyze the increase in performance due to the corrections. Both a general correction, and a site-adapted version of MOS are implemented. The performance of the original clear-sky models is compared to the models with MOS, and the models with MOS plus the MSC.

3. 1. The impact of bias-correction

The results with globally fitted bias correction, over the whole dataset, are shown in Fig. 3. We immediately observe from that the MOS correction brings significant improvements. While the original biases of the models vary between -12% and 24%, after the MOS correction all nMBEs are below 1%, as shown (see Fig. 3a). For all models, the MOS correction also lowers the nRMSE (see Fig. 3b). Notably, while REST2 is inferior to Yang in

the original implementation (18.5% vs. 15.4%), after applying MOS, REST2 becomes considerably better (10.8% vs. 13.9%). Even for MAC2, which performs the worst of all models, application of MOS brings down the nRMSE from 33.6% to 22.1%.

The impacts of the multiple scattering correction, on the other hand, seem to be marginal, bringing a less than 1% improvement for MAC2, and a loss in accuracy for all other models. This may be a sign of the inadequacy of the ad hoc correction described in Sec. 2.3, and the overall small number of high turbidity measurements as a fraction of the dataset. More testing should be performed under conditions of high turbidity. In the remainder of the manuscript, we drop the MSC from the analysis. Henceforth, the subscript *c* notation is used to denote the bias-corrected model, using exclusively the MOS correction (i.e. $REST2_c$).

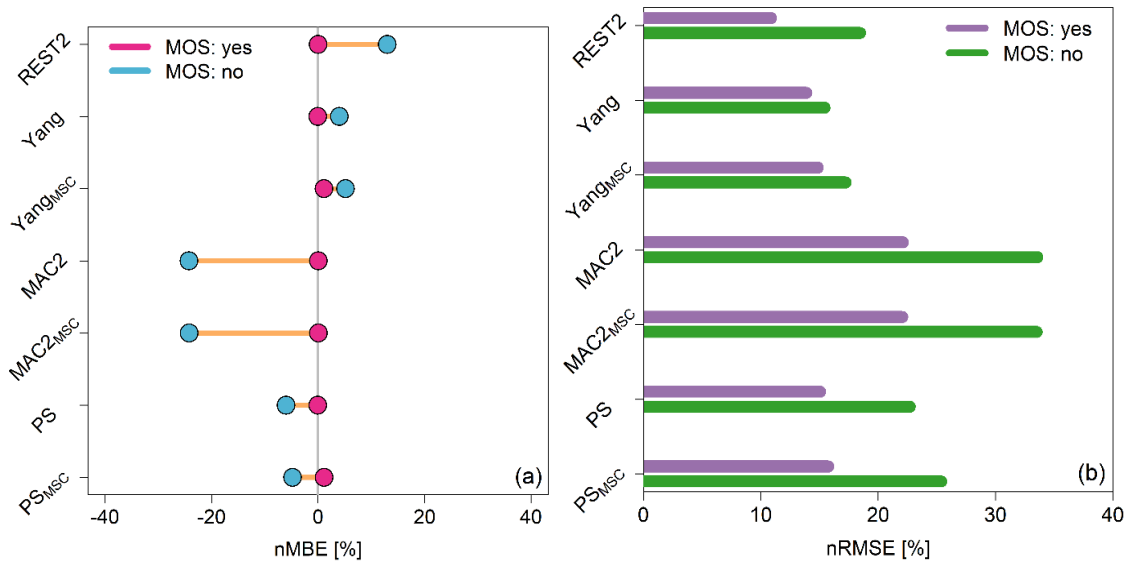


Figure 3. Performance of the bias-corrected model over the whole dataset. The (a) nMBE and (b) nRMSE of the four models, with and without the MOS and the MSC correction, are shown.

3. 2. Site-specific fitting

The question that naturally follows is whether the global bias correction leads to improvements at all stations. The nRMSE for the REST2 model with and without the MOS correction is shown in Fig. 4a for each station. We observe that, indeed, for most stations the correction leads to a significant improvement. The only stations where the correction leads to higher nRMSE are GOB and TAM. We note that these stations have an arid climate (B climate in Table 1). It is likely that the large number of temperate and continental climate locations (climate C and D in Table 1) strongly influence the global correction, leading to

weaker results at locations experiencing heavy desert dust. Thus, a site-specific MOS is also implemented.

The aggregate nRMSE for the whole dataset is shown in Fig. 4b, for the four models. We observe that the global corrections cut the errors heavily compared to the original model, and the site-specific corrections lower them by a few extra percentage points. The results broken down by location are listed in Table 2. Note that, by large, the relative differences between the three model versions (original, global correction and site-based correction), are similar for all stations. For TAM and GOB, for which the global correction was counterproductive, the site-based correction shows a notable improvement.

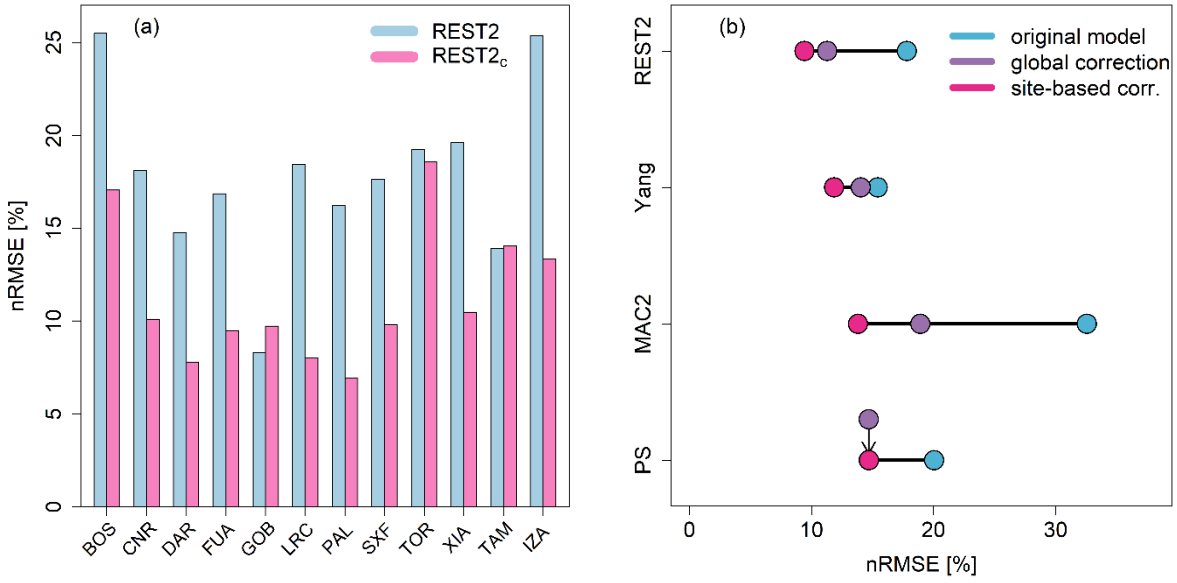


Figure 4. (a) nRMSE of the REST2 model for each station from our dataset, both with and without MOS correction. (b) Difference in the aggregate nRMSE of the four models, between three different implementations: no, global and site-based MOS correction.

Overall, based on our analysis, when high precision is required, a site-specific correction is recommended. If data is not available and moderate precision is sufficient, the global correction can be used. A midway approach, worth pursuing as a compromise, between a site-based and a global correction would be a general correction for each climate type.

Table 2. nRMSE of the four clear-sky models, in the three implementations, for the 12 stations included in the dataset.

BSRN code	REST2			Yang			MAC2			PS		
	Original	Global MOS	Site-based MOS	Original	Global MOS	Site-based MOS	Original	Global MOS	Site-based MOS	Original	Global MOS	Site-based MOS
BOS	25.49	17.08	15.56	21.58	19.14	16.63	31.74	14.27	13.19	19.28	19.51	19.81
CNR	18.10	10.11	9.02	15.95	13.33	11.97	37.01	17.43	13.83	16.63	13.31	13.25
DAR	14.75	7.79	6.91	11.85	12.94	11.00	38.65	21.22	15.07	19.48	14.76	14.68
FUA	16.85	9.47	8.32	17.06	13.32	11.26	22.50	15.12	11.81	18.50	13.39	13.33
GOB	8.30	9.74	5.96	12.99	10.99	10.88	34.35	17.75	14.27	19.33	11.69	11.74
LRC	18.44	8.03	7.89	13.11	11.37	11.16	27.69	11.50	46.51	17.98	11.38	11.53
PAL	16.22	6.93	5.50	13.09	9.87	8.36	32.17	14.06	11.97	14.04	9.62	9.63
SXF	17.63	9.80	9.00	20.34	18.18	17.53	26.87	13.24	13.53	19.39	17.96	17.76
TOR	19.23	18.57	13.79	15.78	15.3	14.7	46.91	27.92	16.90	20.98	16.42	16.36
XIA	19.60	10.48	8.18	17.30	14.62	11.47	33.12	23.40	17.09	23.47	17.18	17.24
TAM	13.91	14.04	9.08	15.79	18.46	10.80	28.17	23.11	15.47	25.57	18.81	18.71
IZA	25.36	13.34	12.94	10.32	10.91	9.10	31.67	27.93	15.57	25.92	12.23	12.31

4. Conclusion

In this study, we have tested the impact of bias-correction algorithms on the accuracy of clear-sky models in predicting the diffuse solar irradiance. Four clear-sky models from the literature were chosen, and ground-based data from two public repositories was used for model input and validation. Two types of bias-correction were tested: a general regression-type bias correction (MOS) and a simplistic correction for aerosol multiple scattering (MSC). While the effect of the implemented MSC proved marginal, the MOS correction had a large impact on model performances. Two versions of MOS were tested: one developed on the global dataset and a site-specific implementation. The global MOS was shown to greatly improve at performance of the models for most stations, except the arid ones. The site-specific MOS reduced the model errors with a few additional percentage points, including the arid stations. Overall, the conclusion from our study is that site-specific bias correction should always be used when multi-year locally measured data is available. If such data is not available, a global bias correction is a good compromise in most cases.

References

[1] C.A. Gueymard, Sol Energy 86 (2012) 2145-2169

[2] A Ångström, Geogr Ann A 11 (1929) 156-166

[3] D. Calinoiu, N. Stefu, R. Boata, R. Blaga, N. Pop, E. Paulescu, A. Sabadus, M. Paulescu, Energ Convers Manage 164 (2018) 175-187

[4] C.A. Gueymard, J. A. Ruiz-Arias, Renew Sust Energ Rev 45 (2015) 379-396

[5] Aerosol Robotic Network (AERONET).https://aeronet.gsfc.nasa.gov/ (accessed Sep 2022)

[6] Baseline Surface Radiation Network (BSRN). https://bsrn.awi.de/ (accessed Sep 2022)

[7] C.A. Gueymard, D. Yang, Atmos Environ 225 (2020) 117216.

[8] D. Yang, J. M. Bright, Solar Energy 210 (2020) 3-19.

[9] C. A. Gueymard, J. A. Ruiz-Arias, Solar Energy 128 (2016) 1-30.

[10] E. Paulescu, R. Blaga, Solar Energy 180 (2019) 75-84.

[11] R. Blaga, M. Paulescu, Sol Energy 174 (2018) 606-616.

[12] R. Blaga, D. Calinoiu, M. Paulescu, *in preparation*.

[13] X. Sun, J. M. Bright, C. A. Gueymard, B. Acord, P. Wang, N. A. Engerer, Renew Sust Energ Rev 111 (2019) 550-570.

[14] R. Blaga, D. Calinoiu, N. Stefu, R. Boata, A. Sabadus, E. Paulescu, N. Pop, O. Mares, S. Bojin, M. Paulescu, Meteorol Atmos Phys 133 (2021) 1395-1407

[15] C. A. Gueymard, Sol Energy 82 (2008) 272-285

[16] M. Paulescu, Z. Schlett, Theor Appl Climatol 75 (2003) 203-212
[17] K. Yang, G. W. Huang, N. Tamai, Sol Energy 70 (2001) 13-22
[18] J. A. Davies, D. C. McKay, Sol. Energy 29 (1982) 55-64
[19] X. Sun, J. M. Bright, C. A. Gueymard, X. Bai, B. Acord, P. Wang, Renew Sust Energ Rev 135 (2021) 110087
[20] R. Blaga, D. Calinoiu, M. Paulescu, J Renew Sustain Ener 13 (2021) 023701
[21] C. A. Gueymard, Sol Energy 71 (2001) 325-346
[22] C. A. Gueymard, (Vol. 1, p. 85). Cocoa, FL: Florida Solar Energy Center (1995)
[23] A. Skartveit, J. A. Olseth, Sol Energy 41 (1988) 19-20
[24] E. Lorenz, J. Hurka, D. Heinemann, H. G. Beyer, IEEE J Sel Top Appl 2 (2009) 2-10.
[25] D. Yang, D. van der Meer, Renew Sust Energ Rev 140 (2021) 110735.
[26] P. Mathiesen, J. Kleissl, Sol Energy 85 (2011) 967-977.