Performance of empirical models for diffuse fraction in Uruguay

G. Abal[a,b,*], D. Aicardi[a], R. Alonso Suárez[a], A. Laguarda[a,b]

[a] *Laboratorio de Energía Solar, UDELAR*
Av. L. Batlle Berres, km 508, CP 50000, Salto, Uruguay
[b] *Instituto de Física, Facultad de Ingeniería, UDELAR*
Herrera y Reissig 565, CP 11300, Montevideo, Uruguay

Abstract

Knowledge of diffuse solar radiation is required for the estimation of global irradiation on inclined surfaces or for estimating DNI for CSP applications. Since diffuse irradiance data is comparatively scarce relative to global horizontal irradiance (GHI) data, several methods are used to estimate the diffuse component of GHI. These methods have a local component and most of them have been developed using data recorded in the northern hemisphere, where long-term reliable measurements of diffuse irradiance are available. This work considers ten models for hourly diffuse irradiation and evaluates their performance, both in their original and locally adjusted versions, against data recorded at five sites from a subtropical-temperate zone in the southern part of South America (latitudes between 30° S and 35° S). The raw data has been quality-assessed by using a set of seven sequential filters which preserve the natural spread of the data while removing unphysical data points. The local adjustment and performance evaluation are done using random-sampling cross-validation techniques on an ensemble. The best estimates result from locally adjusted multiple-predictor models, some of which can estimate hourly diffuse fraction with uncertainty of 18% of the mean.

Keywords: diffuse radiation, solar resource assessment, DNI

1. Introduction

The diffuse component (DHI) of the solar radiation reaching the ground is the result of several interactions between the incident solar (beam) radiation and the atmosphere. These processes can be described by physical models provided enough information on the current composition of the local atmosphere (i.e. aerosol type and density, water vapor column, Ozone column, among others) are available [1]. This detailed information is recorded at a few specialized ground measuring sites, such as those from Aeronet (http://aeronet.gsfc.nasa.gov/).

The separation of the beam and diffuse components of GHI is required before estimating direct normal irradiance (DNI) or global irradiance on inclined surfaces. Recent efforts in solar resource assessment in Uruguay have emphasized the characterization and modeling of GHI on several time scales [2–5], but there is little information available on diffuse radiation for this region. DHI is comparatively hard to measure accurately over long periods of time, so most available data sets include only GHI. A simple way to do this separation is to use phenomenological approaches, based on estimating DHI from a small set of easily measured or calculated predictor variables. These models refer to a definite time scale (typically an hour, a day or a month) and usually relate the diffuse fraction (the fraction of global horizontal irradiance (GHI) which is diffuse) to the clearness index and eventually other variables. They are not universal and several comparisons of their performance at different locations have been reported [6–11].

Since the final uncertainties in solar resource estimation correlate with financial risks in utility-scale projects, a reasonable knowledge of the uncertainties in each step of the calculations is important for the assessment of the performance of solar en-

*Corresponding author: G. Abal, abal@fing.edu.uy

ergy conversion technologies [12]. The uncertainty of a diffuse-fraction model will depend on the degree of climatic similarity between the data sets used to develop the model and the climate in which it is being evaluated. Localized assessments are necessary both to select the best model and to characterize its uncertainty.

The diffuse fraction is not a function of clearness index alone. Proposals with additional variables [8, 13–16] may have lower uncertainties in diffuse fraction estimates at the expense of higher complexity. Gueymard and Ruiz-Arias have recently compared the performance of 140 diffuse fraction models published in the literature [6]. They used minute-based data from 54 research-class stations distributed over four climatic regions of the globe (only one of them is located less than 1000 km from the area of interest in this paper) and characterized the regional performance of each model. An important conclusion is that no current separation model is truly "universal", in the sense to have consistent accuracy over large climatic zones. In fact, the diffuse fraction reflects the typical composition of the local atmosphere, which may be influenced by (natural or man-made) phenomena affecting the water content or aerosol type and density at a specific region. Thus, diffuse fraction estimation is a problem with an important local component.

Phenomenological separation models should be adjusted to local data to remove most of their bias and significantly reduce related uncertainties. However, these models are frequently used as universal due to the absence of reliable local information on their performance. Many models for diffuse fraction have been derived from DHI data taken at locations in the northern hemisphere, some of them at locations near densely populated areas, where these kind of measurements first became available. These models may not perform as well in locations with different characteristics, as previously noted for Australia by Boland [17].

In this work, controlled-quality local diffuse irradiation data from five low-altitude sites with southern latitudes (between 30° S and 35° S) is used to evaluate the performance of ten well-known hourly diffuse-fraction models. A strong filtering procedure is applied to the hourly data. For each model, both the original version and a locally adjusted version are evaluated against independent data using a standard cross-validation technique. Two frequently used models for daily and monthly average diffuse fraction are also evaluated and locally adjusted. Information is provided on the best way to estimate diffuse fraction for this and similar geographical regions on an hourly, daily and monthly basis. More importantly, the uncertainty associated to each estimation procedure is characterized, so that it may be accounted for in engineering calculations for solar energy projects.

The paper is organized as follows. In Section 2, the solar radiation database, the typical uncertainty for each site and the filters applied on the raw data are discussed. In Section 3, hourly diffuse fraction models are briefly described and evaluated against local data. In Section 4, all hourly models are adjusted to local data and re-evaluated on a per-site basis using several common statistical indicators. A global adjusted version of each model is defined and evaluated. In Section 5, the data is reduced to daily totals and two daily and monthly average models for diffuse fraction are implemented, locally adjusted and evaluated. Finally, In Section 6 our conclusions are summarized.

2. Ground data

The data used in this work consists of simultaneous data sets for hourly global and diffuse horizontal irradiation from five sites located in a sub-tropical temperate zone of the south-eastern part of South America with homogeneous climatic characteristics shown in Figure 1. The area has a marked seasonality, no significant volcanic activity, low population density (except for the Buenos Aires metropolitan area) and it is not heavily industrialized.

2.1. Description of data sets

The location, instruments and number of hourly records (simultaneous global and diffuse irradiance) for each site are listed in Table 1. All sites are at low altitudes, with the highest (AR)

Location						Time period	Instruments & method for DHI			
Site	Code	LAT (°)	LON (°)	ALT (m)	owner	start - end	GHI	DHI	hours	uncertainty
Montevideo	AZ	-34.92	-56.17	58	LES	03/2014 - 08/2013	Delta-T SPN1 [4%]	Delta-T SPN1 [7%]	7961	9%
Salto	SM	-31.27	-57.89	41	INUMET	06/1998 - 12/2003	KZ CM11 [3%]	KZ CM11 + s-ring [4%]	20594	6%
Luján	LU	-34.58	-59.05	20	GERSolar	01/2011 - 06/2012	KZ CMP11 [3%]	Eppley 8-48 + s-ball [5%]	5934	7%
Artigas	AR	-30.40	-56.51	136	LES	02/2014 - 12/2015	Delta-T SPN1 [4%]	Delta-T SPN1 [7%]	7613	9%
Treinta y Tres	TT	-33.27	-54.17	20	LES	02/2014 - 12/2015	Delta-T SPN1 [4%]	Delta-T SPN1 [7%]	6634	9%

Table 1: Location of the measurement sites considered in this work (see Fig. 1 for the geographical distribution of the sites). Time period (month/year) and pyranometer manufacturer, model and estimated uncertainty for hourly averages. The method used for DHI measurement at SM was a Kipp & Zonen (KZ) CM-121 shadow-ring (s-ring). At LU it was a shading ball assembly (s-ball) based on a SOLYS2 tracking system. The last column indicates the valid daytime hours (F0 level, see Table 2) with simultaneous GHI and DHI measurements. And the last column is the overall estimated uncertainty for the normalized data from each site.

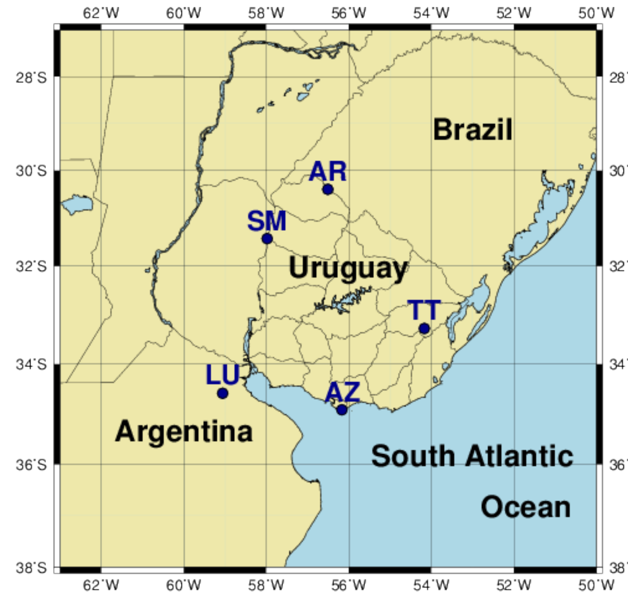

Figure 1: Location of the measuring stations considered in this work. Other details are provided in Table 1.

at 136 m above sea level. Except for AZ, all sites had a daily cleaning routine by local staff. For AZ, the cleaning routine was performed on a weekly basis. No ventilation devices where used.

The AZ site is located at the roof-top of the School of Engineering at Montevideo, an urban coastal location. GHI and DHI were measured and recorded at one-minute intervals between 2011 and 2013 using a new Delta-T SPN1 pyranometer. The data was recorded at 1-minute intervals using a Fisher-Scientific DT80 datalogger connected by cable to the internal network.

The SM site was at a supervised meteorological station run by the National Meteorological Service (INUMET), located close to the Salto air field in a semi-rural location. GHI and DHI were measured and recorded at 15 minute intervals, during six years using two CM11 (Secondary Standard) Kipp & Zonen (KZ) pyranometers. The raw data was recorded with a Campbell Scientific CR1000 datalogger and it was provided to us without any processing. DHI was measured with a manually adjusted shadow-band, also from KZ. The DHI data was corrected using the isotropic correction factor [18] as provided by the band manufacturer, $f = (1 - S)^{-1}$, with

$$S = \frac{2\theta_0}{\pi} \cos\delta \left(\omega_s \sin\phi \sin\delta + \sin\omega_s \cos\phi \cos\delta\right) \qquad (1)$$

where $\theta_0 = 0.185\,\text{rad}$ is the view angle of the shadow ring, δ the solar declination angle, ω_s the sunset hour angle (in radians) and ϕ the site latitude. This factor, which at the relevant latitudes varies yearly between 1.05 and 1.14, accounts for the portion of hemispherical sky radiation blocked by the shadow band under the assumption of an isotropic distribution of diffuse irradiance. According to the manufacturer, the correction from Eq. (1) is accurate to ±0.5%. A comparison of several correction methods for diffuse irradiance measurements based on shadow rings [19], suggests a typical uncertainty of 4% with respect to a shading-ball assembly measurement, provided secondary-standard class pyranometers are used in both cases. Based on these considerations and on our own verifications, we estimate a typical uncertainty of 3% for GHI and 4 % for DHI hourly data from this site, relative to the overall mean values.

The LU site is at a specialized research laboratory (GERSO-

LAR) of the National University at Luján (Argentina) located in a semi-rural area 50 km from the city of Buenos Aires. Three independent measurements (GHI, DHI and DNI) were recorded at 1-minute intervals between 2011 and 2012. Global irradiance was measured with a KZ CMP11 pyranometer, diffuse irradiance with a Black and White Eppley 8-48 pyranometer using a shade-ball assembly and the beam component was measured with an Eppley NIP pyrheliometer. These instruments were mounted on a new SOLYS2 tracking system from KZ . The calibration of pyrheliometers and pyranometers was done by periodic comparisons against a Kendall Absolute Cavity radiometer used as secondary standard. Further details on this data set can be found in Ref. [20]. The estimated uncertainty for data from this site is 3 % and 5 % for hourly measurements of GHI and DHI respectively, relative to the overall mean values. Although the shading-ball assembly method for diffuse measurements is potentially more accurate than a shadow-ring measurement, the Eppley 8-48 pyranometer used for diffuse measurements has a typical uncertainty of 5%, as declared by the manufacturer.

The TT site is part of an experimental station managed by the National Institute of Agronomical Studies (INIA) located in a rural area, about 50 km from the nearest populated areas. The AR site is at a meteorological station run by the National Meteorological Service (INUMET) located in a semi-rural area, 15 km from the town of Artigas. Two new Delta-T SPN1 pyranometers were installed by our laboratory at these sites in february 2014, and data was recorded at 1-minute intervals and sent on a daily basis to a dedicated server via the cellular (GSM) network. Data from these sites recorded between 2014 and 2015 has been used in this work. At both sites, redundant GHI measurement using KZ CMP11 and CMP6 pyranometers were installed and all the instruments received daily maintenance from the local staff. The sites AZ, TT and AR are part of a continuous solar radiation measurement network maintained by our laboratory since 2010. The pyranometers in this network are calibrated at our laboratory at two-year intervals, following ISO 9847:1992(E) norm procedures [21]. The Secondary Standard used as a reference is a KZ CMP22 pyranometer calibrated against the World Radiometric Reference (WRR) at the World Radiation Center (WRC) at Davos in August 2014.

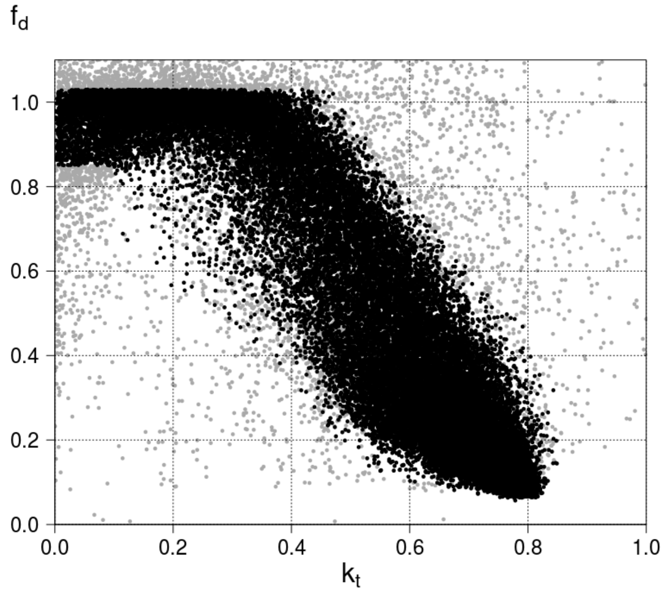

Figure 2: Hourly diffuse fraction, f_d, vs. clearness index, k_t, for all sites filtered to F7-level are shown in black. Unfiltered (F0-level) data for all sites is shown in the background (gray).

The SPN1 pyranometer has no moving parts and can operate over long periods of time without human intervention other than the cleaning procedures required by all hemispherical instruments. These radiometers are robust and allow continuous measurement of global and diffuse irradiance at remote locations in a cost-effective way as compared to other alternatives, such as rotating shadowband radiometers or tracker-based measurements. DHI data from SPN1 instruments account for about 46% of the data used in this work (Table 1) so we shall briefly discuss the accuracy of these instruments.

This pyranometer uses an array of seven thermopile sensors, each of them calibrated to consistently measure solar irradiance. A special mask under its dome shades at least one of the sensors from direct sunlight while leaving at least one of them exposed to direct sunlight, at all times and locations. This mask blocks approximately half of the hemisphere and the instrument computes individual values for GHI and DHI using a simple algorithm based on the maximum and minimum irradiance it measures in its seven sensors at a given time. The uncertainty stated by its manufacturer for individual measurements is ±8% (±10 W/m^2), both for GHI and DHI at 95% confidence level [22]. Several studies [23–25] have shown that these instruments easily comply with the stated uncertainty for GHI, but their DHI uncertainty can be higher. In a comparison made at NREL in 2009, Myers and Wilcox [23] reported for this instrument uncertainties between 4 to 7% for GHI and 7 to 11% for DHI. Another study [24] compared SPN1 measurements against KZ CM11 pyranometers (one of them equipped with a shadow band) and reported uncertainties of approximately 3% for GHI and 14% for DHI at the 1-minute time scale. More recently, a detailed study by Badosa et al. [25] has compared SPN1 measurements against high quality (sun tracker based) data and found similar uncertainties of approximately 5% for GHI and 12% for DHI. A negative mean bias of approximately 5% was also reported for DHI measurements when compared to measurements from a shade-ball assembly. Based on these results, the application of a 1.05 correction factor to the DHI output of this instrument has been recommended [22, 25]. In this work, this correction has been applied to the DHI data from the AR, AZ and TT sites before filtering.

Furthermore, we have recently recalled the three SPN1 instruments used in this work and calibrated them at our laboratory against two CMP-22 pyranometers (Secondary standards), one of them equipped with a shadowring. Additionally, a simultaneous measurement of GHI, DHI and DNI based on two new CMP10 pyranometers and two CHP1 pirheliometers mounted on a SOLYS2 tracking system where available for consistency checks. As a result, we have determined that the GHI uncertainty of the three SPN1 instruments is between 3 and 4 % and their DHI uncertainty, between 9 and 10 %, without correction factor. When this factor is included, we have verified that the DHI measurement is essentially unbiased and the uncertainty in DHI from the SPN1 instruments is between 6 and 7 %, in agreement with [25]. On this basis, we estimate the SPN1 uncertainty for GHI at 3 % and for (corrected) DHI at 7 %.

Based on the uncertainty estimate for each measurement, one can assign combined uncertainties to the diffuse fraction data for each site, as indicated in the last column of Table 1.

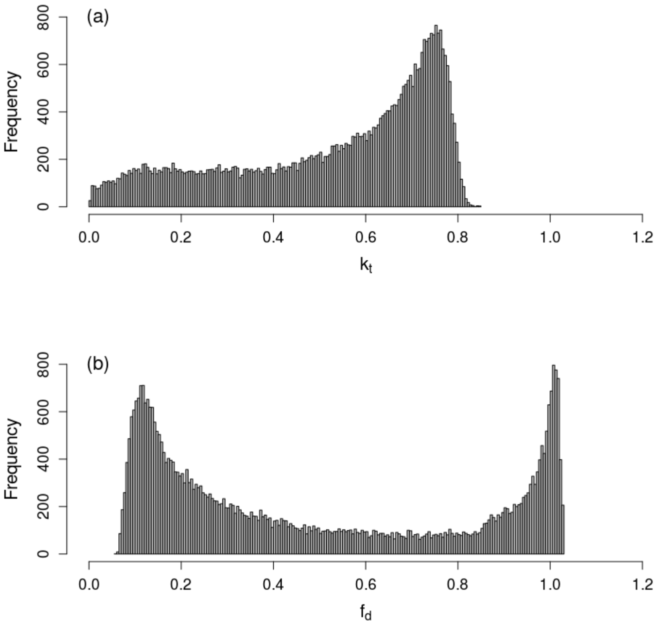

Figure 3: Distributions of (a) hourly k_t values and (b) hourly f_d values, both filtered to F7-level. Data from all sites is shown, since similar distributions are found on a per-site basis.

2.2. *Filtering criteria*

GHI data separates into its beam (I_{bh}) and diffuse (I_{dh}) components, $I_h = I_{bh} + I_{dh}$. As a first step, the data is normalized in order to remove most trends due to the apparent solar motion. The hourly clearness index, k_t, is defined as $k_t = I_h/I_{0h}$, where I_{0h} is the hourly solar irradiation on a horizontal surface at the top of the atmosphere. The hourly diffuse fraction, f_d, is the ratio $f_d = I_{dh}/I_h$. For cloudy conditions, $k_t \to 0$ and $f_d \to 1$. For clear-sky conditions, $k_t \approx 0.80$ and f_d takes low values (≈ 0.10) which depend on the composition of the local atmosphere, as shown in Figs. 2 and 3.

When working with diffuse (or beam) solar irradiation, quality assessment of the data is specially relevant [26, 27]. A filtering procedure is implemented, based on the sequential application of seven filters to the normalized hourly data records, as summarized in Table 2. The process starts with the set (F0)

Filter	Conditions	AZ		SM		LU		AR		TT		all sites	
		hours	%	hours	%	hours	%	hours	%	hours	%	hours	%
F0	$\cos\theta_z \geq 0$ & $I_h > 0$ & $I_{dh} > 0$	7961		20594		5934		7613		6634		48736	
F1	$\cos\theta_z \geq 0.1219$ ($\alpha_s \geq 7°$)	7062	11.3	18483	10.3	5372	9.5	6974	8.4	5987	9.8	43878	10.0
F2	$I_h \leq I_{hc}$ ($T_L = 2$)	6863	2.8	18348	0.7	5315	1.1	6909	0.9	5890	1.6	43325	1.3
F3	$k_t > 0.1$ & $I_d \geq I_{dc}$ ($T_L = 1.5$)	6796	1.0	18327	0.1	5300	0.3	6909	0.0	5890	0.0	43222	0.2
F4	$I_d \leq (600\,\mathrm{W/m^2})\,\alpha_s$	6773	0.3	18091	1.3	5142	3.0	6903	0.1	5875	0.3	42784	1.0
F5	$k_t \leq 0.10$ & $f_d \geq 0.85$	6601	2.5	18045	0.3	5132	0.2	6786	1.7	5853	0.4	42417	0.9
F6	$k_t \leq 0.85$ & $0.05 \leq f_d \leq 1.03$	6559	0.6	17895	1.0	4920	4.1	6558	3.4	5576	4.7	41472	2.2
F7	$\lvert t\rvert = \left\lvert \hat{f}_d - f_d \right\rvert / \sigma < 3$	6491	1.0	17616	1.4	4868	1.1	6486	1.1	5534	0.8	40995	1.2
all	% discarded		18.5		14.5		18.0		14.8		16.6		15.9

Table 2: Sequence of filters applied to the hourly irradiation data from each station. For each site, the number of hours that pass each filter and the percentage of records discarded are indicated. A total of 40995 valid daytime hourly records were used and 15.9 % of the daytime hours were discarded.

of daytime hours with positive global and diffuse hourly irradiation records for each site. F1 eliminates hours with low solar altitude ($\alpha_s < 7°$), for which the measurements become unreliable. F2 uses the ESRA clear-sky model [28, 29] with Linke turbidity parameter $T_L = 2$, to provide an upper bound I_{hc} for GHI. For the hours that pass this filter the hourly clearness index, $k_t = I_h/I_{0h}$, is calculated. Unless for very dark conditions, diffuse irradiation should be larger than the clear-sky estimate I_{dc}. Filter F3 uses the same clear-sky model with $T_L = 1.5$ to apply a lower bound on diffuse irradiation when $k_t > 0.1$. F4 places an upper bound on diffuse irradiation, $I_d \leq (600\,\mathrm{W/m^2})\,\alpha_s$, (solar altitude expressed in radians), based in Page's estimate for overcast irradiance [28]. For instance, for $\alpha_s \approx 80°$ this limit is $843\,\mathrm{W/m^2}$. For the hours that pass this filter, the diffuse fraction $f_d = I_d/I_h$ is calculated. At overcast sky conditions the diffuse fraction should be high. F5 removes low diffuse fractions found at overcast conditions with the requirement that if $k_t \leq 0.10$ then $f_d \geq 0.85$. On physical grounds, one would expect $0 \leq f_d \leq 1$, but these limits are relaxed to account for measurement error. F6 places boundaries on the normalized data by requiring $0.05 \leq f_d \leq 1.03$ and $k_t \leq 0.85$. The last filter, F7, is of a statistical nature and aimed to remove the few remaining outliers. A simple polynomial fit (P5) to the F6-level data is used to compute normalized residuals $t = (\hat{f}_d - f_d)/\sigma$, where σ is the sample RMSD. Since the residuals are (almost) normally distributed, an hour is considered an outlier (and discarded with 99.7% confidence level) if $|t| > 3$. Approximately 41000 hourly (k_t, f_d) records result from this procedure, as summarized in Table 2.

The thresholds for all filters have been selected on physical grounds after visual inspection of their effects on the data cloud. It is important to emphasize that the quality of the raw data and the specific choices made in the filtering procedure affect quantitatively the results. The hourly data set filtered to F7-level is shown in Fig. 2 against the background of F0-level data. The distributions for (filtered) k_t and f_d are shown in Fig. 3 (a) and (b) respectively. Note that in (a) the right peak in k_t is associated to clear days while in (b) the right peak in f_d is associated to overcast conditions and the left peak to clear-sky conditions. These distributions are similar to those reported in the literature [6, 30, 31]. Both variables have a bi-modal distribution, although bimodality is weakened at the hourly timescale in k_t.

3. Phenomenological models for hourly diffuse fraction

Phenomenological models attempt to capture the general trend of the diffuse fraction in terms of a set of readily available variables together with basic time and site information. These models are usually adjusted from a limited amount of data for a few locations. Even though such models are not universal [6], they are often used, at least for engineering purposes [32], over a wide range of locations and atmospheric conditions without information about the associated uncertainties. In this Section, ten well-known hourly diffuse fraction models are introduced. They have been selected with special attention to simplicity and usability and are locally adjusted and evaluated in Section 4. In order to easily refer to them, a short code is assigned to each model as indicated in Table 3. For each model, there is an original version and two locally adjusted versions (per site and global) as described in Section 4.

Model	Year	Ref.	Sites	Length	Param.	Pred.
P5	2015	–	5	40995 h	3	1
OH	1977	[33]	1	4 y	2	1
EKD	1982	[34]	5	5 y	4	1
RBD	1990	[13]	5	22000 h	9	2
SO2	1987	[35]	1	44687 h	6	2
BSL	2001	[17, 36]	7	NA	2	1
RBL	2010	[14]	7	NA	6	5
RA1	2010	[15]	21	~ 23 y	4	1
RA2s	2010	[15]	21	~ 23 y	5	2
RA2	2010	[15]	21	~ 23 y	7	2

Table 3: Models for hourly diffuse fraction considered in this work. NA indicates unknown metadata. "Length" indicates the size of the data set used to train the original model (years or hours); "Param." is the number of adjustable parameters in our implementation and "Pred." is the number of predictor variables.

3.1. Simple polynomial model (P5)

A simple polynomial model (P5) for diffuse fraction results from a polynomial function of k_t,

$$f_d = \begin{cases} 1 & k_t < 0.20 \\ a_0 + a_1\, k_t + a_2\, k_t^2 + a_3\, k_t^3 + a_4\, k_t^4 + a_5\, k_t^5 & 0.20 \le k_t \le 0.85 \\ c_0 & k_t > 0.85. \end{cases} \tag{2}$$

subject to continuity constrains for f_d and its derivative f_d' at the endpoints of the central interval. The resulting model has three independent parameters and it serves as a benchmark to evaluate more sophisticated approaches obtained from the literature. This model, adjusted to F6-level data, has been used to compute the residuals for discarding outliers in filter F7, as explained in Section 2.2. The values of the coefficients adjusted to F7-level data, for each site and globally as detailed in Section 4, are listed in Table 4.

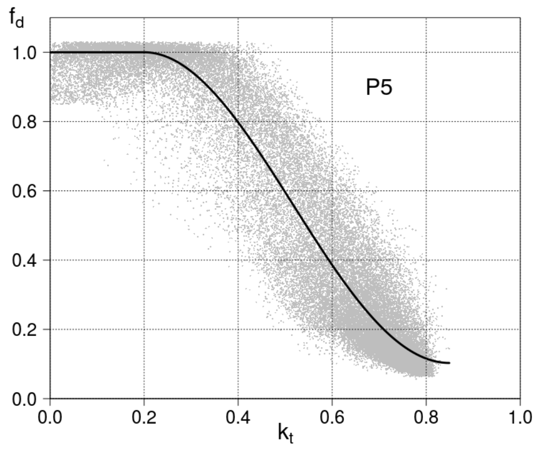

Figure 4: Global polynomial model, Eq. (2) against F7-level data. The coefficients are listed in the rightmost column of Table 4.

P5	AZ	SM	LU	AR	TT	Global
a_0	0.50	0.72	0.80	0.86	1.04	0.77
a_1	5.92	2.80	1.97	0.87	−1.45	2.16
a_2	−22.22	−6.62	−3.93	3.53	13.21	−3.91
a_3	29.51	−4.66	−5.97	−28.43	−43.80	−9.02
a_4	−19.54	14.13	10.96	39.51	48.79	17.00
a_5	6.09	−6.20	−3.56	−16.21	−17.60	−6.79
c_0	0.10	0.09	0.11	0.11	0.12	0.10

Table 4: Parameter sets for model P5, Eq. (2). See Section 4 for details.

3.2. Models OH and EKD

Among the most well-known models are those by Orgill and Hollands [33] (OH) and Erbs et al. [34] (EKD). Both models have been evaluated by several authors previously [6, 7, 37] and can be expressed as

$$f_d = \begin{cases} 1 + a_1\, k_t & k_t < k_a \\ b_0 + b_1\, k_t + b_2\, k_t^2 + b_3\, k_t^3 + b_4\, k_t^4 & k_a \le k_t \le k_b \\ c_0 & k_t > k_b. \end{cases} \quad (3)$$

These models differ mainly in their functional form in the central interval, where OH uses a linear expression ($b_2 = b_3 = b_4 = 0$). Two continuity constrains at $k_a = 0.35$ and $k_b = 0.75$ reduce the number of free parameters in this model to just two. Orgill and Hollands used four years of hourly data from a single site (Toronto, Canada) to obtain the coefficients shown in the first column of Table 5. Emphasizing the local nature of the model, they recommended the use of these parameters for latitudes between 43 °N to 54 °N [33].

OH	Original	AZ	SM	LU	AR	TT	Global
a_1	−0.25	−0.40	−0.29	−0.24	−0.33	−0.19	−0.28
b_0	1.56	1.51	1.60	1.60	1.56	1.63	1.59
b_1	−1.84	−1.86	−2.00	−1.95	−1.93	−1.99	−1.96
c_0	0.18	0.12	0.10	0.14	0.11	0.14	0.12

Table 5: Parameters for model OH, Eq. (3) with $b_2 = b_3 = b_4 = 0$.

For the EKD model uses $k_a = 0.22$ and $k_b = 0.80$ in Eq. (3). The original coefficients for this model, listed in Table 6, were obtained using data from five U.S. sites with latitudes between 31 °N and 42 °N with altitudes from 62 m to 1620 m above sea level [34]. Continuity constrains at k_a and k_b and a continuous derivative at k_b are assumed, so there are four independent parameters. The authors compared this correlation to 3 years of data from Highett, Australia (latitude 38 °S) to evaluate its usefulness in a different climate at a similar latitude. Since then, the EKD model has been used and evaluated world-wide [7, 32, 37, 38] and it has been recommended for universal use in engineering textbooks [32]. Both models (OH and EKD) are shown in the top panels of Fig. 5, in their original and locally adjusted versions, against F7 data.

3.3. Model RBD

The model by Reindl et al. [13] is an example of a simple, piecewise, multi-predictor model. These authors used 22000 hours of data from five sites in the U.S. and Europe, with latitudes ranging from 28 °N to 56 °N. An additional set of 3000 hours measured at Oslo, Norway (latitude 60 °N) were used for evaluation purposes. Reindl et al. considered a large set of 28 candidate predictor variables, including those commonly measured at meteorological stations, and used a piecewise linear model to fit the data. They concluded that, on an hourly basis, the best predictor variables were k_t and $\sin \alpha_s$, the sine of the solar altitude. Other relevant predictors might be ambient temperature and relative humidity, but they found this four-predictor model to perform only marginally better than the two-predictor

EKD	Original	AZ	SM	LU	AR	TT	Global
a_1	−0.09	−0.24	0.00	−0.06	−0.15	−0.10	−0.09
b_0	0.95	0.70	0.38	0.62	0.68	0.85	0.60
b_1	−0.16	2.63	6.54	3.70	2.91	0.98	3.97
b_2	4.39	−7.38	−21.25	−10.83	−7.75	−0.06	−11.74
b_3	−16.64	1.86	21.37	7.00	1.47	−9.75	7.76
b_4	12.34	2.67	−6.99	−0.30	3.24	8.62	−0.28
c_0	0.17	0.13	0.09	0.12	0.13	0.13	0.11

Table 6: Parameter sets for model EKD, Eq. (3).

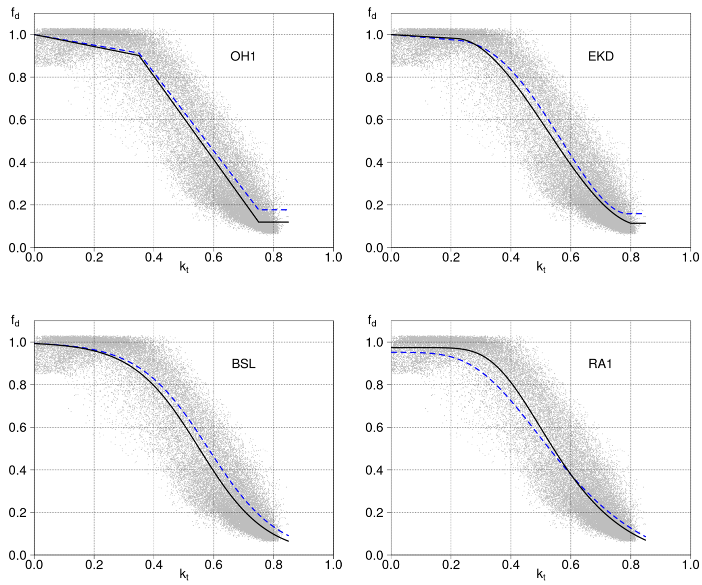

Figure 5: Four single-predictor models. For each case, the original model (dashed line) and the locally adjusted global model (full line) are shown with the data filtered at level F7 in the background.

one. Keeping in mind simplicity and usability, we shall consider only the two-predictor version (RBD) defined by,

$$f_d = \begin{cases} a_0 + a_1\, k_t + a_2\, \sin\alpha_s & f_d \leq 1 \quad\quad k_t < k_a \\ b_0 + b_1\, k_t + b_2\, \sin\alpha_s & 0.1 \leq f_d \leq 0.97 \quad k_a \leq k_t \leq k_b \\ c_0 + c_1\, k_t + c_2\, \sin\alpha_s & f_d \geq 0.1 \quad\quad k_t > k_b, \end{cases} \tag{4}$$

where $k_a = 0.30$ and $k_b = 0.78$ are fixed. The constrains within each interval are required to avoid unphysical values for possible combinations of the predictors. No continuity constrains are applied. The values of the parameters, as given in Ref. [13], are listed in the first column of Table 7. An evaluation of this model against modern data can be found in Refs. [6, 7, 37].

RBD	Original	AZ	SM	LU	AR	TT	Global
a_0	1.02	0.88	1.00	0.97	0.96	0.96	0.96
a_1	−0.25	−0.04	−0.09	−0.07	0.12	0.20	−0.01
a_2	0.01	0.14	0.01	0.02	−0.06	−0.07	0.01
b_0	1.40	1.39	1.44	1.49	1.42	1.50	1.45
b_1	−1.75	−1.94	−1.97	−1.96	−2.01	−2.01	−1.98
b_2	0.18	0.31	0.24	0.22	0.32	0.27	0.26
c_0	0	0.56	−0.34	−0.15	0.02	−0.38	−0.12
c_1	0.49	−0.51	0.69	0.43	0.13	0.73	0.38
c_2	−0.18	−0.03	−0.12	−0.09	−0.01	−0.10	−0.08

Table 7: Parameters for RBD model, Eq. (4).

3.4. Model SO2

Skartveit and Olseth [35] proposed a piecewise non-linear diffuse fraction model [SO2] whith the solar altitude α_s as an additional predictor. In particular, one of the interval boundaries depends on α_s. The diffuse fraction is parametrized as

$$f_d(k_t, \alpha_s) = \begin{cases} 1 & k_t \leq k_a \\ f(k_t, \alpha_s) & k_a \leq k_t \leq \alpha k_b(\alpha_s) \\ f(\alpha k_b, \alpha_s) & k_t \geq \alpha k_b(\alpha_s) \end{cases} \tag{5}$$

where $\alpha = 1.09$, $k_b(\alpha_s) = r + s\, \exp(-\alpha_s/\alpha_0)$ and $\alpha_0 = 0.291$ rad. The non-linear functions are

$$\begin{aligned} f(k_t, \alpha_s) &= 1 - (1 - d_1)\left[a\sqrt{K} + (1 - a)\, K^2\right], \\ K(k_t, \alpha_s) &= \frac{1}{2}\left\{1 + \sin\left[\pi\left(\frac{k_t - k_a}{k_b - k_a} - \frac{1}{2}\right)\right]\right\}, \\ d_1(\alpha_s) &= r' + s'\, \exp(-\alpha_s/\alpha_0). \end{aligned} \tag{6}$$

SO2	Original	AZ	SM	LU	AR	TT	Global
a	0.27	0.06	0.44	0.11	0.10	0.12	0.21
r	0.15	0.01	0.05	0.03	0.04	0.06	0.04
s	0.43	0.64	0.43	0.58	0.42	0.27	0.47
r'	0.87	0.92	0.87	0.91	0.89	0.90	0.90
s'	−0.56	−0.68	−0.53	−0.56	−0.55	−0.43	−0.55
k_a	0.20	0.05	0.24	0.13	0.11	0.14	0.15

Table 8: Parameters for model SO2, Eqs. (5) and (6). The second column is from Ref. [35] and the rest of the columns correspond are fits to the ground data considered in this work.

In Ref. [35], six parameters (a, r, s, r', s', k_a) are obtained from the data. These values, reproduced in the first column of Table 8, are valid for altitudes close to sea level at average snow-free conditions in Norway. This model is continuous at both interval boundaries where it also has an (approximately) continuous partial derivative $\partial f_d/\partial k_t$ at $k_t = \alpha_s k_b$. In spite of its apparent complexity, it has only six adjustable parameters and two predictors. The same authors have proposed a more complex model [16], which includes persistence and ground albedo among other effects, and shall not be considered here.

3.5. Logistic models (BSL, RBL)

In 2001 Boland et al. [36] proposed and later evaluated [17] a single predictor model (BSL) using a simple logistic function, with just two parameters derived from data from 8 sites over four continents. More recently [14], Ridley, Boland and Lauret have proposed a multiple-predictor version,

$$f_d = \left[1 + \exp\left(a_0 + a_1\, k_t + a_2\, t_s + a_3\, \alpha_s + a_4\, K_t + a_5\, \psi\right)\right]^{-1}. \tag{7}$$

The two-parameter, single-predictor logistic model (BSL) can be obtained from Eq. (7) by setting $a_2 = a_3 = a_4 = a_5 = 0$ and its original parameters from Ref. [17] are reproduced in Table 9.

The extended model (RBL) considers four predictors aside from k_t: t_s is the apparent solar time (in hours) at the mid-hour point, α_s is the solar altitude angle in degrees, K_t is the daily clearness index and ψ is a persistence parameter defined as the average of the lag and lead hourly clearness index, i.e. for the

BSL	Original	AZ	SM	LU	AR	TT	Global
a_0	−5.00	−4.50	−5.03	−4.98	−4.78	−5.23	−4.94
a_1	8.60	8.37	9.26	8.83	8.81	9.18	8.95
RBL							
a_0	−5.38	−5.07	−5.50	−6.02	−5.31	−5.85	−5.60
a_1	6.63	7.29	7.75	7.34	7.94	7.94	7.63
a_2	0.01	−0.03	0.02	0.03	0.00	0.00	0.01
a_3	−0.01	−0.02	−0.01	−0.01	−0.02	−0.02	−0.01
a_4	1.75	1.12	0.88	1.75	0.68	0.93	1.12
a_5	1.31	1.92	2.12	1.92	2.08	2.29	2.06

Table 9: Parameters for the logistic models BSL and RBL, Eq. (7). For BSL all parameters a_j with $j > 1$ are zero. The original values are from Refs. [14, 17].

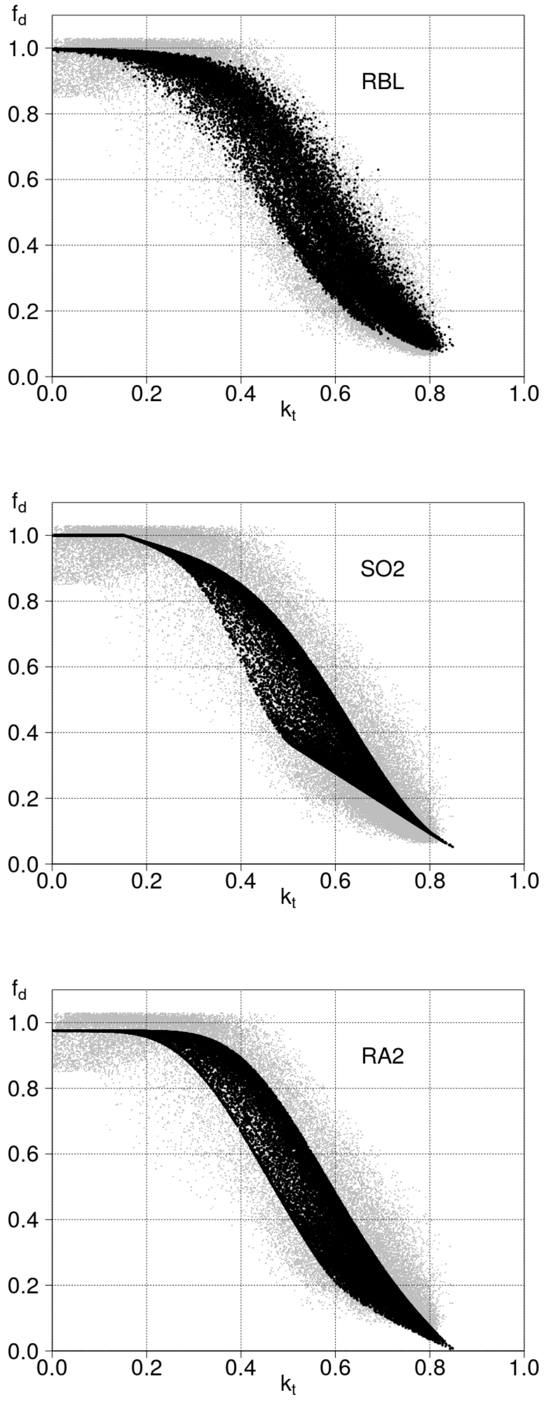

Figure 6: Best models (locally adjusted) with multiple predictors. Upper panel: model RBL, Eq. (7); center panel: model SO2, Eq. (5); Bottom panel: model RA2, Eq. (8). In the background, the data filtered to F7-level is shown. See Section 4 for details on the local adjustment and evaluation of these models.

j^{th} hour, $\psi(j) = \frac{1}{2}\left[(k_t(j-1) + k_t(j+1)\right]$, unless for sunrise or sunset hours where $\psi(j) = k_t(j \pm 1)$, respectively. Using a persistence and a daily variable in an hourly model implies that it can not be used in real time or for days with incomplete hourly information. However, it can capture daily trends in the data and offer an improved performance. The original values for these parameters have been obtained in Ref. [14], from data for seven locations worldwide. The coefficients obtained from the data considered in this paper are also shown in Table 9.

3.6. Double exponential models (RA1, RA2s, RA2)

Ruiz-Arias et al. [15], have proposed the use of a Gompertz (or double exponential) function for diffuse fraction estimation. These functions have previously been used in this context to discard outliers in the (k_t, f_d) plane [20, 27] as they can represent the general trend shown in Fig 2. In this work, we consider one (RA1) and two-predictor (RA2s, RA2) models with the general form

$$f_d = a_0 + a_1\, e^{-\exp\left(a_2 + a_3\, k_t + a_4\, k_t^2 + a_5\, m + a_6\, m^2\right)} \tag{8}$$

where m stands for the height-corrected relative air mass [39] evaluated at the midpoint of the hour. The single-predictor model (RA1) is obtained by setting $a_4 = a_5 = a_6 = 0$ in this expression and a simplified two-predictor model (RA2s) is obtained by setting $a_4 = a_6 = 0$. In Ref. [15] the parameters for each of these models are determined from a large set of high quality data from 21 sites worldwide. These sites range in latitude from 30 °N to 65 °N and in altitude from sea level to almost 2000 m. The dataset covers a range of climates but all sites are located in the northern hemisphere.

RA1	Original	AZ	SM	LU	AR	TT	Global
a_0	0.95	0.95	1.00	0.98	0.95	0.96	0.97
a_1	−1.04	−0.97	−1.07	−1.05	−0.92	−0.97	−1.01
a_2	2.30	2.96	2.82	2.96	3.57	3.46	3.07
a_3	−4.70	−6.07	−5.82	−5.75	−7.32	−6.68	−6.17
RA2s							
a_0	0.98	0.96	1.00	0.97	0.94	0.96	0.97
a_1	−1.02	−1.06	−1.17	−1.15	−1.01	−1.08	−1.11
a_2	2.88	3.36	3.05	3.32	3.94	3.75	3.38
a_3	−5.59	−5.87	−5.44	−5.56	−6.79	−6.25	−5.84
a_5	−0.11	−0.15	−0.11	−0.12	−0.17	−0.14	−0.13
RA2							
a_0	0.94	0.97	1.00	0.97	0.94	0.96	0.98
a_1	−1.54	−0.94	−1.40	−1.18	−1.11	−1.43	−1.24
a_2	2.81	2.72	3.25	3.36	4.30	4.10	3.47
a_3	−5.76	−1.01	−6.19	−5.49	−7.58	−7.85	−5.71
a_4	2.28	−5.78	1.62	0.10	1.37	2.80	0.32
a_5	−0.13	−0.45	−0.19	−0.19	−0.36	−0.22	−0.25
a_6	0.01	0.04	0.01	0.01	0.03	0.02	0.02

Table 10: Parameters for the double exponential single-predictor model RA1, based on Eq. (8) with $a_4 = a_5 = a_6 = 0$. The first column is from Ref. [15].

The site-independent set of coefficients recommended for each model in Ref. [15] are shown in the first column of Table 10 under "Original", together with the corresponding coefficients determined from the data considered in this work.

3.7. *Evaluation of original models*

In Table 3 the ten hourly diffuse fraction models and the length and scope of the data used to train the original version of each model are listed. These original models are evaluated against the F7-level set of ground measurements (described in subsection 2.2) and three performance indicators are calculated. For n data points (x_i, y_i) and corresponding estimates $\hat{y}_i$, the Mean Bias Deviation (MBD) is defined as

$$\mathrm{MBD} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i), \qquad \mathrm{rMBD} = 100 \times \frac{\mathrm{MBD}}{< y >} \quad (9)$$

where $< y >$ is the mean of the observations. It should be noted that this indicator has been defined in both ways (estimate - measurement or viceversa) in the literature. Eq. (9) implies that a positive bias is associated to overestimation by a given model, in accordance with standard usage [40].

The Root Mean Square Deviation (RMSD) quantifies the dispersion of the residuals,

$$\mathrm{RMSD} = \left[\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2\right]^{\frac{1}{2}}, \ \mathrm{rRMSD} = 100 \times \frac{\mathrm{RMSD}}{< y >}. \quad (10)$$

For hourly models, the relative forms are scaled using the average of F7-level hourly data, $< f_d >= 0.47$. In Section 5, $< F_d >= 0.46$ and $< \bar{F}_d >= 0.36$ are used to scale the daily and monthly average indicators, respectively. The rRMSD indicator characterizes the uncertainty introduced by the use of a given model to estimate the diffuse component of GHI. Thus, it is relevant to rank the adjusted models according to their capacity to describe the data. The rMBD gives information about systematic tendencies in the models to overestimate or underestimate the data. Combinations of these two indicators, such as Student's t parameter [41] or the $\mu_{1-\alpha}$ indicator [42] may be used to rank these original models, with some emphasis on the mean bias indicator. However, since the locally adjusted (essentially unbiased) versions of the models perform significantly better, we make no attempt to rank the original versions.

A Kolmogorov-Smirnov Index (KSI) is defined [4, 40, 43] using the cumulative distribution functions $F(Y)$ and $\hat{F}(Y)$ estimated from the f_d measurements and the corresponding model estimates respectively. KSI quantifies the distance between these distributions,

$$KSI = \int_0^1 D(Y)\,dY, \quad \text{with} \quad D(Y) = \left|\hat{F}(Y) - F(Y)\right|. \quad (11)$$

The function D helps visualize for which range of f_d the model estimates differ significantly from the data. As an example, in Fig. 8 we show this function for the RBL model. Other models have a similar form for $D(f_d)$ suggesting that their performance under clear sky conditions (low f_d) might be improved, for instance by using an accurate clear-sky model. Thus, KSI gives information about the similarity between the distributions of the measured and modeled diffuse fractions and discriminates well between different models.

These three basic indicators, MBD, RMSD and KSI, are combined into a single one which takes into account dispersion, absolute bias and likeness of distributions

$$\mathrm{CPI} = \frac{1}{3}(|\mathrm{rMBD}| + \mathrm{rRMSD} + 100 \times \mathrm{KSI}). \quad (12)$$

This overall indicator is similar to the combined index proposed by Gueymard [40].

Meaningful comparisons based on relative indicators calculated by different authors using different data sets are not always straightforward. Even when the same data set is used, differences in the filtering procedure may affect the performance indicators obtained for the same model. With this in mind, we note that some of the models considered here have been previously evaluated elsewhere. Jacovides et al. [10] evaluated the single variable models OH, EKD, RBD and BSL (among others) using 5-years of hourly data for a semi-rural site in Cyprus. They reported positive biases between 3-7% and rRMSD in a narrow range about 30% for all of them. Tapakis [44] used data from the same site in Cyprus (but for a 13-year period) to evaluate models OH, EKD, RBD and RA1 and obtained rMBD in a narrow range around 5% and rRMSD around 24%.

These relative indicators are calculated for each original model, on a per-site basis, in columns 3 to 8 of Table 13. Note that most models show positive biases (over-estimation of diffuse fraction) in the range 3-12%. As mentioned, similar results have been reported previously for Cyprus [10] and also in a comparison between several separation models with data from sites closer than 1000 km from our region of interest [45]. The exceptions to positive biases are the models based on Gompertz or double exponential functions (RA1, RA2s, RA2), which show negative biases (under 10%) for all our sites. In Ref. [15] these models are also reported with mostly negative biases (between -5% and -12%) when compared to independent data from 14 sites in the northern hemisphere. Many of the models considered here (including RA1 and RA2) have been recently compared (at the 1-minute time scale) against data from a BSRN site (SMS-São Martino da Serra), located about 500 km to the north of the area of interest of this work [6]. In this comparison, rRMSD between 24-29% and rMBD between 3-7% were found, with the biases from RA1 and RA2 having a different sign than those of the other models. These results are consistent with the left part of Table 13, with rRMSD's in the range 19-28%, depending on model and site. The AZ site, located at the most southern latitude and being the only coastal site in our analysis, has a higher dispersion (rRMSD's) than the rest. This site is representative of a special climatic regime with higher variability, humidity and cloudiness than inland sites.

In terms of global rRMSD, the best original models are RBL and RA2s with 20.7 % and 21.0% respectively, with the last having a lower bias. They are followed by RA2 and SO2 with rRMSD of 21.8% and 22.9%, respectively. Independent work performed in Argentina [11] used 4320 hours of data from one site in a similar climatic region and also found RBL and SO2 (in their original forms) as the two best models in terms of rRMSD. However, this work did not consider Gompertz-based models such as RA2s and RA2. Models RBL and SO2 have also been ranked among the best for Temperate Zones in Ref. [6].

The tendency for overestimation the diffuse fraction found in most models may be due to a clearer atmosphere with fewer aerosols on average, since most sites considered in this work are at semi-rural grasslands with low population density. However, this may also result from imperfect measurements. For instance, no ventilated pyranometers where used, so water droplets on the domes may affect measurements even at high solar altitudes. The use of the isotropic correction for the shadowring diffuse measurements (which affects more than 40% of the data) may produce some underestimation (of the order of 1%) in the diffuse fraction [19]. The exact choices made in the filtering procedure (Table 2) may also affect the mean bias results of the original models. Specifically, if a higher lower limit (i.e. 0.95 instead of 0.85) is chosen for diffuse fraction under cloudy conditions (F5) a slightly smaller bias would be obtained.

The models considered here have been originally adjusted to data mostly from the northern hemisphere and, in some cases, using small data sets. Since they have significant biases, it is worth deriving locally adjusted versions of these models, specifically adapted for this region of the world.

4. Locally adjusted models

Each model has been trained and evaluated on a per-site basis, using a standard cross-validation procedure. At each site, the F7-dataset is randomly divided into a training set (80%) and a testing set (20%) and the optimal set of coefficients for each model is obtained using standard regression techniques. The final coefficients and indicators result from averaging over an ensemble of 1000 elements, each sampling random 80-20 portions of the datasets. The size of the ensemble has been chosen as to warrant the repeatability of the procedure.

	AZ	SM	LU	AR	TT
w_i	0.14	0.32	0.26	0.14	0.14
f_i	0.16	0.43	0.12	0.16	0.13

Table 11: Weights w_i are based on the estimated uncertainty for the data from each site, indicated in Table 1. The second row shows the fraction f_i of data points from each site at F7-level (Table 2).

In addition to the locally adjusted version of each model for each site, global models are constructed from the locally adjusted models using the weighted average of the adjusted coefficients from each site as the global model coefficients. The weights w_i are determined from the estimated uncertainty for each site, σ_i, as indicated in the last column of Table 1. Thus, $w_i = C/\sigma_i^2$ with $C = \sum_i 1/\sigma_i^2$ (the sum runs over all sites). The resulting weight factors, listed in Table 11, give priority to higher quality data from the SM and LU sites. The coefficients for the globally fitted versions of each model are listed in the last column (under Global) of Tables 4 to 10.

Performance indicators for these globally adjusted versions are shown in the rightmost columns of Table 13. The per-site indicators are averaged (weighted using the fractions f_i of data points from each site) to obtain the global indicators for each adjusted model, listed under the "All" header in Table 13.

		Adjusted local models				
Model	Indicator	AZ	SM	LU	AR	TT
P5	$rMBD$ (%)	1.9	0.0	0.6	2.0	1.5
	$rRMSD$ (%)	25.8	19.2	22.5	22.1	22.9
	KSI (×100)	2.3	1.4	1.9	2.2	2.0
OH	$rMBD$ (%)	0.0	−0.8	−0.1	0.4	0.6
	$rRMSD$ (%)	25.8	20.1	22.6	22.2	22.9
	KSI (×100)	2.6	2.3	2.1	2.3	1.9
EKD	$rMBD$ (%)	0.4	0.1	0.3	0.8	0.7
	$rRMSD$ (%)	25.5	19.2	22.5	22.0	22.9
	KSI (×100)	2.2	1.4	1.9	1.9	1.8
RBD	$rMBD$ (%)	0.0	0.0	0.0	0.0	0.0
	$rRMSD$ (%)	23.3	17.3	20.8	18.4	20.5
	KSI (×100)	2.3	1.3	2.0	1.8	1.9
SO2	$rMBD$ (%)	0.9	0.0	0.4	1.5	1.3
	$rRMSD$ (%)	22.9	16.6	20.1	18.2	20.2
	KSI (×100)	1.9	1.2	1.6	1.4	1.6
BSL	$rMBD$ (%)	−0.2	−1.5	−0.6	−0.2	−0.1
	$rRMSD$ (%)	25.6	19.6	22.6	22.1	23.0
	KSI (×100)	2.2	1.6	2.0	1.6	1.6
RBL	$rMBD$ (%)	0.6	−0.8	−0.2	0.7	0.6
	$rRMSD$ (%)	21.2	15.7	17.7	17.4	19.0
	KSI (×100)	1.5	1.2	1.5	1.0	1.3
RA1	$rMBD$ (%)	0.0	0.0	0.0	0.0	0.0
	$rRMSD$ (%)	25.5	19.2	22.4	21.5	22.6
	KSI (×100)	2.5	1.3	2.0	2.1	2.1
RA2s	$rMBD$ (%)	−0.1	0.0	0.0	0.0	0.0
	$rRMSD$ (%)	23.2	16.9	20.3	18.3	20.3
	KSI (×100)	2.3	1.1	1.8	1.8	1.9
RA2	$rMBD$ (%)	0.0	0.0	0.0	0.0	0.0
	$rRMSD$ (%)	22.8	16.6	20.2	17.7	19.9
	KSI (×100)	2.2	1.0	1.8	1.6	1.9

Table 12: Per-site performance indicators for the locally adjusted models as compared to the F7-dataset on a per-site basis. Note that the local Gompertz models (RA1, RA2s and RA2) and the local RBD model are essentially unbiased at all sites. The best local model performance is from RBL at SM, with rRMSD below 16%.

		Original models						Adjusted global models					
Model	Indicator	AZ	SM	LU	AR	TT	All	AZ	SM	LU	AR	TT	All
P5	$rMBD$ (%)							4.5	2.8	−2.9	3.6	−3.1	1.7
	$rRMSD$ (%)							26.1	19.5	22.9	22.2	23.6	21.9
	KSI (×100)							2.3	2.1	2.3	2.5	2.8	2.3
OH	$rMBD$ (%)	10.7	8.7	4.2	9.6	3.0	7.9	3.8	1.6	−3.7	2.5	−4.0	0.7
	$rRMSD$ (%)	28.1	22.9	23.2	24.4	23.4	24.1	26.2	20.3	22.9	22.4	23.4	22.3
	KSI (×100)	5.9	6.0	3.8	5.2	3.6	5.3	2.7	2.9	2.2	2.3	2.2	2.6
EKD	$rMBD$ (%)	10.8	8.8	3.8	9.8	3.3	7.9	4.0	2.2	−3.4	3.0	−3.7	1.2
	$rRMSD$ (%)	28.0	22.2	22.9	24.1	23.1	23.6	25.9	19.6	22.8	22.1	23.5	21.9
	KSI (×100)	5.4	5.0	2.5	4.9	2.3	4.4	2.1	2.4	2.3	2.4	2.7	2.4
RBD	$rMBD$ (%)	11.2	10.6	5.2	11.0	3.8	9.2	3.3	2.3	−3.8	2.6	−4.4	0.9
	$rRMSD$ (%)	26.6	21.9	22.5	22.9	22.7	23.0	23.8	17.9	21.3	18.9	21.4	19.8
	KSI (×100)	5.6	6.0	4.5	6.0	4.2	5.5	2.4	2.7	2.2	2.5	2.5	2.5
SO2	$rMBD$ (%)	13.4	12.2	7.0	12.6	6.2	11.0	3.8	2.3	−4.0	2.5	−3.8	1.0
	$rRMSD$ (%)	26.7	22.1	22.0	22.4	21.9	22.9	23.4	16.9	20.7	18.3	21.1	19.2
	KSI (×100)	6.6	6.6	4.3	6.2	3.7	5.8	1.9	2.2	2.5	1.5	2.3	2.1
BSL	$rMBD$ (%)	11.5	9.3	4.7	10.2	3.6	8.5	3.0	1.1	−4.3	1.9	−4.7	0.1
	$rRMSD$ (%)	28.6	23.3	23.3	24.9	23.5	24.4	25.9	19.9	23.0	22.2	23.6	22.1
	KSI (×100)	6.0	6.2	3.7	5.2	3.3	5.3	2.2	2.7	2.4	1.8	2.6	2.4
RBL	$rMBD$ (%)	10.3	8.7	2.5	8.9	3.0	7.6	3.8	2.7	−4.7	2.4	−3.4	1.3
	$rRMSD$ (%)	25.0	19.8	18.7	20.9	20.4	20.7	22.2	16.3	18.8	17.8	19.5	18.1
	KSI (×100)	5.3	5.6	3.2	4.6	3.2	4.9	2.0	2.4	2.6	1.4	1.9	2.1
RA1	$rMBD$ (%)	0.0	−1.1	−5.8	−0.4	−7.6	−2.2	3.4	1.5	−3.8	2.4	−4.3	0.5
	$rRMSD$ (%)	26.0	21.3	24.3	22.8	25.6	23.2	25.8	19.6	22.8	21.8	23.3	21.8
	KSI (×100)	4.4	5.2	5.0	4.4	5.6	5.0	2.1	2.3	2.5	2.1	2.8	2.3
RA2s	$rMBD$ (%)	−1.2	−1.8	−7.5	−1.5	−8.5	−3.2	3.8	1.7	−4.0	2.2	−4.1	0.7
	$rRMSD$ (%)	23.9	18.5	23.3	20.2	24.5	21.0	23.7	17.4	20.8	18.5	21.0	19.5
	KSI (×100)	3.7	4.3	5.3	3.6	5.6	4.4	2.3	2.2	2.3	1.7	2.4	2.2
RA2	$rMBD$ (%)	−2.5	−3.1	−8.4	−2.7	−10.1	−4.5	0.9	−0.9	−7.4	−0.6	−7.0	−2.2
	$rRMSD$ (%)	24.6	19.7	23.8	20.9	24.8	21.8	23.5	17.0	21.9	18.3	21.8	19.5
	KSI (×100)	4.0	4.5	5.3	3.7	5.9	4.6	2.3	2.1	2.3	2.5	2.8	2.3

Table 13: Performance indicators for the original and globally adjusted versions of the hourly models. The indicators listed under the columns labelled "All", are the average of the per-site indicadors, weighted by the fraction of data points at each site (see Table 2).

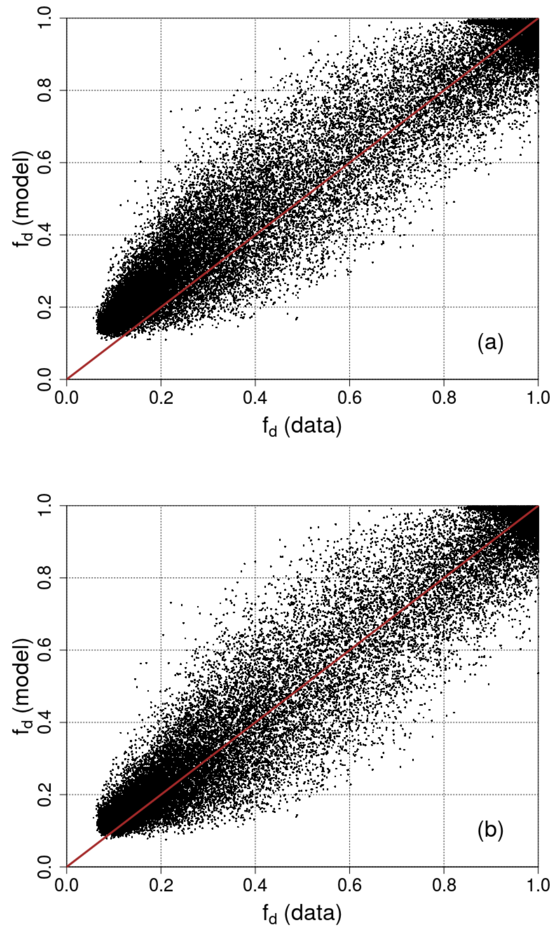

Figure 7: Scatter plot for the diffuse fraction from the RBL model: (a) model with original coefficients and (b) global model with adjusted coefficients. A line with slope 1 is drawn to guide the eye.

As expected, the adjusted versions have significantly lower biases than the original ones and the global adjusted versions have biases within ±2%. Original models have rRMSD indicators in the range 18-28% and locally fitted models in the range 16-26%, so the improvement in dispersion is not as significant as with bias. In Table 14, all models are listed as ordered by increasing CPI, defined in eq. (12). Ties in CPI are resolved by KSI or rRMSD (both yield the same result). The procedure is fairly robust: the first two models have the lowest KSI and rRMSD and any single variable model performs worst than any multi-variable model also in terms of KSI or rRMSD.

The best performing global model for this region is clearly RBL with CPI of 7.2 and rRMSD of 18.1%. It is followed by SO2 (19.2%) and RA2s tied with CPI=7.4 (RA2s has higher rRMSD 19.5% and lower bias than SO2), however RA2s is simpler to implement. RA2 and RBD follow tied with CPI=7.8, the first has lower rRMSD and KSI.

as noted, all multiple-predictor models perform better than any single-predictor model based on k_t alone, so it is worth including air mass m and other additional variables as predictors for describing diffuse fraction in this region. All local single-predictor models have rRMSD indicators around 22%, implying that the details of these models are not very relevant, as long as they are limited to k_t as the only predictor. In order to emphasize this point, we have introduced a simple fifth degree polynomial model with natural constrains (P5) which is ranked worst in terms of CPI but in terms of rRMSD and rMBD actually performs almost as well as the best adjusted single-predictor model (RA1).

Scatter plots for the original and globally adjusted versions of the RBL model are shown in Fig. 7 and the difference function D used to calculate the distance between the distributions from the data and the RBL model estimates (for the original and globally adjusted versions) is shown in Fig. 8. The effect of the local adjustment is seen in the reduction of the area under the difference function. Similar results are obtained for other models. The peak at low f_d values (clear sky conditions) suggests that some improvement in the model's performance may result if the low-end f_d estimation is done by using a locally tuned clear-sky model for diffuse radiation, such as [29]. Fig. 8 also shows a peak in D when $f_d \to 1$, so there is room for improvement at overcast conditions too. This potential for improvement should be considered when addressing the subject of improved physical models for diffuse fraction.

The RBL model, Eq. (7), stands out in its use of the daily clearness index (K_T) and a persistence parameter (which depends on the previous and the next hour) as predictors. This particular parametrization is probably related to its good performance. However, the use of a daily variable makes it inadequate for real-time (on-the-fly) estimation of hourly diffuse irradiation or for short-term forecasting applications. The second-best

	rMBD (%)		rRMSD (%)		KSI (x100)		CPI	Rank
Model	Original	Adjusted	Original	Adjusted	Original	Adjusted	Adjusted	
RBL	7.6	1.3	20.7	18.1	4.9	2.1	7.2	1
SO2	11.0	1.0	22.9	19.2	5.8	2.1	7.4	2
RA2s	-3.2	0.7	21.0	19.5	4.4	2.2	7.4	2
RA2	-4.5	-2.2	21.8	19.5	4.6	2.3	7.8	3
RBD	9.2	0.9	23.0	19.8	5.5	2.5	7.8	3
RA1	-2.2	0.5	23.2	21.8	5.0	2.3	8.2	4
BSL	8.5	0.1	24.4	22.1	5.3	2.4	8.2	4
EKD	7.9	1.2	23.6	21.9	4.4	2.4	8.5	5
OH	7.9	0.7	24.1	22.3	5.3	2.6	8.5	5
P5	-	1.7	-	21.9	-	2.3	8.7	6

Table 14: Overall performance indicators for the ten hourly models considered in this work in their original and adjusted versions. The "Rank" column orders the adjusted models using a combined performance indicator defined in Eq. (12). The horizontal line separates multiple-predictor models from single-predictor models.

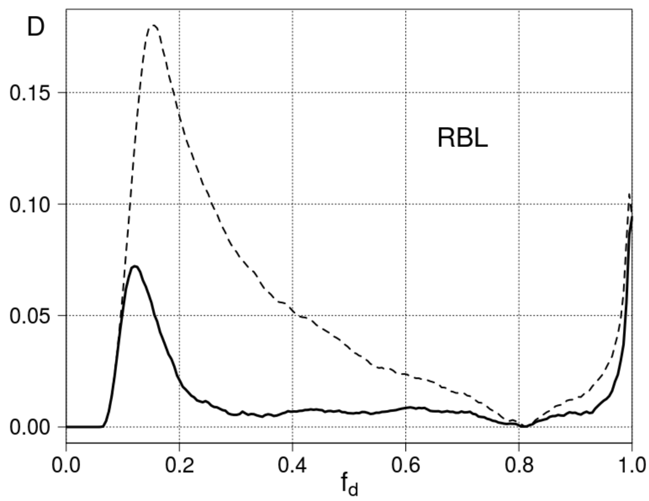

Figure 8: Difference between cumulative distribution functions, D, from Eq. (11), for the original (dashed line) and the adjusted global (full line) versions of the RBL model.

models are SO2 and RA2s with similar indicators. The adjusted RA2s model is almost unbiased and has a simpler parametrization than SO2. Thus, for the average user, the locally fitted RA2s model, Eq. (8), which can estimate hourly f_d with an uncertainty under 20%, may represent the best compromise between performance and simplicity.

5. Daily and monthly-mean diffuse fraction

In real applications, daily data for GHI may be the only information available near the required location. In order to estimate the daily solar resource on an inclined surface, a separation into daily diffuse and direct irradiation is previously required. In this Section, two models which have been widely used to obtain this separation are evaluated, in their original and locally adjusted versions. The case of monthly means of daily irradiation is also discussed.

In terms of the daily global and diffuse horizontal irradiation (H_h, H_{dh}) and the daily horizontal irradiation at the top of the atmosphere (H_{0h}), the daily clearness index and the daily diffuse fraction are defined as,

$$K_T = H_h/H_{0h} \qquad \text{and} \qquad F_d = H_{dh}/H_h. \tag{13}$$

The monthly-averaged versions of these quantities are

$$\overline{K}_T \equiv \overline{H}_h/\overline{H}_{0h} \qquad \text{and} \qquad \bar{F}_d \equiv \overline{H}_{dh}/\overline{H}_h, \tag{14}$$

where the averages are over daily data within each month.

For daily and monthly averaged daily data, the specific site-dependence is weaker than for hourly data. For simplicity

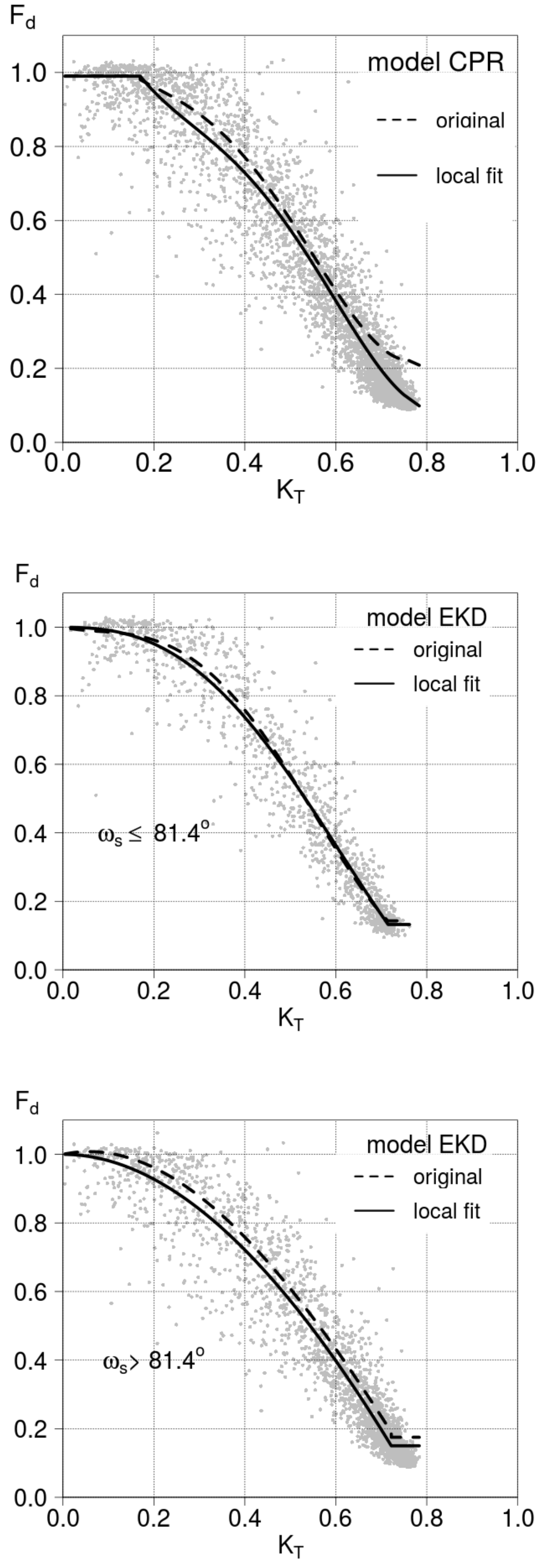

Figure 9: Top: CPR-d model for daily data, Eq. (15), (original and adjusted) against the bachground of the daily data. Middle: EKD-d model for daily data (original and adjusted), Eq. (16), with $\omega_s \leq 81.4°$ (winter). Bottom: EKD-d model for daily data (original and adjusted) with $\omega_s > 81.4°$.

and brevity, daily data from all sites is aggregated and a single locally adjusted version of each model is considered. Our objective is to assess the typical uncertainty associated to the diffuse-direct separation procedure at the daily and monthly time scales.

5.1. Data base

Daily data is obtained from hourly data as follows. For a given day, the subset of hours that pass F0, F1, F2 and F6 are considered. For days with complete hours, hourly data is accumulated to generate daily irradiation in the usual form, i.e. $H_h = \sum_j I_h(j)$ and $H_{dh} = \sum_j I_{dh}(j)$, where j is an hour index and the sums run over all daylight hours. Days with one or more missing hours are discarded. If a given month has at least 20 days with daily data, the monthly averages, $\bar{K}_T$ and $\bar{F}_d$, are computed otherwise the month is discarded. The results of this selection process are summarized in Table 15.

Site Code	Valid days	Valid months
AZ	549	16
SM	1538	53
LU	375	12
AR	506	18
TT	417	13
All sites	3385	112

Table 15: Summary of the filtered daily data for each station and valid months for calculating monthly-averages.

5.2. Models

Two frequently used models are considered in their daily and monthly averaged versions. The daily model by Collares-Pereira et al. [46] (CPR-d), is defined in four intervals in K_T,

$$F_d = \begin{cases} 1 & K_T \leq 0.17 \\ A_0 + A_1\,K_T + A_2\,K_T^2 + A_3\,K_T^3 + A_4\,K_T^4 & 0.17 < K_t \leq 0.75 \\ B_0 + B_1\,K_T & 0.75 < K_T < 0.80 \\ C_0 & K_T \geq 0.80. \end{cases} \tag{15}$$

where the original and locally adjusted values of the coefficients are listed in Table 16. The daily model from Erbs et al. [34] (EKD-d) is also considered,

$$F_d = \begin{cases} 1 + A_1\,K_T + A_2\,K_T^2 + A_3\,K_T^3 + A_4\,K_T^4 & K_T < 0.715 \\ B_0 & K_T \geq 0.715. \end{cases} \tag{16}$$

with coefficients listed in Table 16. Note that it includes a seasonal dependence through ω_s, the sunset hour angle: its coefficients have different values for ω_s below (i.e. winter) or above (rest of the year) a threshold of 81.4 °.

The monthly-average model proposed by Collares-Pereira [46] (CPR-m) uses the monthly averaged hour angle $\bar{\omega}_s$ (in rads) to introduce seasonal dependence,

$$\overline{F}_d = A - B\,\cos(A_2 + A_3\,\overline{K}_T) \tag{17}$$

where $A = A_0 + A_1\left(\bar{\omega}_s - \frac{\pi}{2}\right)$ and $B = B_0 + B_1\left(\bar{\omega}_s - \frac{\pi}{2}\right)$. The monthy mean sunset angle, $\bar{\omega}_s$, can be approximated by its value for the typical day of each month [47] with negligible error. The original and adjusted values for these coefficients are listed in Table 16.

Finally, the model for monthly-averaged diffuse fraction (EKD-m) by Erbs et al. [34] is also considered. It is defined by

$$\overline{F}_d = A_0 + A_1\,\overline{K_T} + A_2\,\overline{K_T}^2 + A_3\,\overline{K_T}^3, \tag{18}$$

with two sets of coefficients according to the value of $\bar{\omega}_s$. The monthly averaged clearness index is restricted to the interval $0.3 \leq \overline{K_T} \leq 0.8$ and the coefficients are listed in Table 16. The locally adjusted EKD-m model for $\omega_s \leq 81.4$ has some instabilities. On the other hand, inspection of Fig. 10 shows that the F_d data considered in this work has only weak seasonal dependence. Ignoring this dependence results in a stable locally adjusted model with similar performance indicators as those obtained by preserving the ω_s dependence. Thus, the locally adjusted EKD-m version does not include the ω_s dependence and a single set of local coefficients are listed in the last row of Table 16.

5.3. Evaluation

After a cross-validation procedure similar to the one used for the hourly models, performance indicators are obtained for the

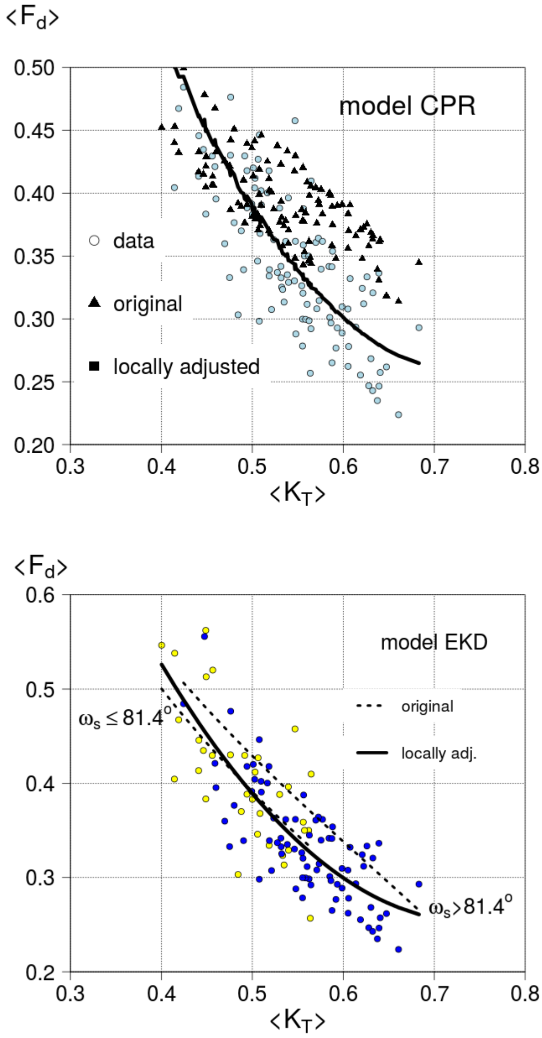

Figure 10: CPR model for monthly-averaged data, Eq. (17). Monthly averaged values are indicated within brackets < · > . EKD model for monthly-averaged data, Eq. (18). Data for $\bar{\omega}_s \leq 81.4\,°$ is shown with blue circles and with $\bar{\omega}_s > 81.4\,°$ with yellow circles. For each case, the original EKD monthly model, Eq. (16), is shown with dotted lines. The locally fitted model (with no dependence with ω_s) is shown with a full line. Monthly averaged values are indicated within brackets.

		Original models							
Model	restriction	A_0	A_1	A_2	A_3	A_4	B_0	B_1	C_0
CPR-d		1.19	−2.27	9.47	−21.87	14.65	0.63	−0.54	0.20
EKD-d	$\omega_s \leq 81.4°$	1	−0.27	2.45	−11.95	9.39	0.14		
	$\omega_s > 81.4°$	1	0.28	−2.56	0.85	0	0.18		
CPR-m		0.78	0.35	2.00	−1.80	*	0.51	0.26	*
EKD-m	$\omega_s \leq 81.4°$	1.39	−3.56	4.19	−2.14				
	$\omega_s > 81.4°$	1.31	−3.02	3.43	−1.82				
		Adjusted models							
CPR-d		1.49	−5.05	16.89	−29.15	16.52	0.79	−0.89	0.08
EKD-d	$\omega_s \leq 81.4°$	1	0	−0.46	−4.50	3.89	0.13		
EKD-d	$\omega_s > 81.4°$	1	0	−1.88	0.34	0	0.15		
CPR-m		6.21	0.56	0.93	−0.67	*	5.95	0.57	*
EKD-m		1.58	−3.67	2.68	−0.19				

Table 16: Original and locally adjusted parameters for daily and monthly-average models for diffuse fraction.

adjusted models. The performance indicators for the original and adjusted daily models are shown in Table 17. As expected, large bias indicators (between 5% and 10%) are obtained and at the daily timescale the original models also overestimate daily diffuse fraction in the region of interest. For both models, the local fit reduces the mean bias (below 1%) and leads to lower F_d estimates, as shown in Fig. 9. The locally fitted versions of both daily models perform similarly, with a small edge for EKD-d, estimating daily diffuse irradiance with rRMSD under 20% and the same KSI.

The monthly-averaged data, together with the estimates from the two models considered, are shown in Fig. 10. Due to the small size of the monthly dataset, all of it was used to fit and evaluate the monthly mean models. Even though this leads to artificially lower performance indicators, they are useful to compare the adjusted model to its original version and to compare monthly models between themselves.

Again, both original models tend to overestimate monthly-mean diffuse irradiation, as they do at the daily and hourly time scales and have large rRMSD (for averaged quantities) between 16% and 20%, with EKD-m outperforming CPR-m. The locally adjusted versions of both models are essentially unbiased and their rRMSD are significantly lower (under 13 %) as indicated in Table 17. As in the daily case, both adjusted models perform equally well.

It is worth using local adjusted models at the daily and monthly-average scales in order to reduce bias. The minimum uncertainties introduced when using the unbiased versions are 20% (daily) and 13% (monthly), respectively.

6. Conclusions

The uncertainty introduced by phenomenological models for diffuse fraction separation has been well characterized for a temperate region located in the southern part of South America. The daytime hourly data was quality assessed and almost 41000 hours of valid data from five sites (most of them in semi-rural areas) are the basis for this work. Ten models for hourly diffuse fraction have been implemented and evaluated in their original and locally adjusted forms. Half of the models considered use the clearness index as their single variable and the

	rMBD (%)		rRMSD (%)		KSI (x100)		CPI	
Model	Orig.	Adj.	Orig.	Adj.	Orig.	Adj.	Orig.	Adj.
EKD (d)	5.5	0.4	20.6	19.7	2.7	1.0	9.6	7.0
CPR (d)	10.2	0.7	22.8	19.9	4.9	1.0	12.6	7.2
EKD (m)	7.0	0.0	16.1	12.8	3.4	1.1	8.8	4.6
CPR (m)	9.8	0.0	20.2	12.8	4.8	1.1	11.6	4.6

Table 17: Statistical indicators for the daily and monthly mean diffuse fraction models considered in this work. For each model, the indicators with the original and the locally fitted coefficients are shown. The average daily diffuse fraction is $F_d = 0.46$ and the monthly average diffuse fraction is $\overline{F}_d = 0.36$.

other half includes other variables as predictors.

The five multi-variable models outperform, in terms of dispersion, any of the single variable models considered, so the best original models are the multi-variable models with uncertainties of at least 21%. Most original models over-estimate the diffuse fraction with biases in the range 3-12%, depending on the site. Gompertz based (double exponential) models are the exception and have small negative biases. These results are dependent on the quality of the experimental data. This may be a true effect due to a clearer atmosphere, which is plausible given the geographical characteristics and the relatively low industrialization and human density of the area under consideration. But the possibility that it is due to some residual bias present in the data after the filtering process cannot be ruled out at present. Further work is required, based on higher quality data for the area, before this overestimation can be confirmed.

A locally adjusted and a global version of each model where obtained and evaluated per-site using cross validation procedures. Our results clearly show that multiple-predictor models perform consistently better that any single-predictor ones, both in their original and local versions. The adjusted models do not show the overestimation tendency present in the original models. Mean biases are within ±5% and within ±2% for the global versions. The adjusted versions span a range of rRMSD between 16% and 26%, depending on site and model.

Using a combined performance indicator which takes into account bias, dispersion and similarity between the data and the modeled distributions, the ten adjusted hourly models have been ranked according to their overall performance in the region under consideration. The best of them, RBL [14], can estimate hourly diffuse fraction in the region of interest with a typical uncertainty of 18% and 1% bias. However, this model uses daily irradiation as an input and can't be used for real-time (on demand) separation or for predictive purposes. On a second level, are the SO2 and RA2s models [15, 16] with typical uncertainty under 20% and negligible bias. These models do not share the limitation of the RBL in regard to real-time use. The adjusted RA2s (double exponential) model has a simpler parametrization than the SO2 and, for the average user, RA2s may represent the best compromise between performance and simplicity.

At the daily and monthly mean timescale, two models (CP and EKD) were evaluated before and after adjusting them to to local data. In their original forms both tend to overestimate diffuse irradiation. In their adjusted versions, both daily models are essentially unbiased and perform similarly with typical uncertainty under 20%. In the monthly average case, both adjusted models are indistinguishable, with typical dispersion of about 13%.

In engineering applications, the overall uncertainty introduced by the diffuse radiation estimation should be carefully included in the calculations. The rather high biases found in some original models imply that caution is required before using phenomenological diffuse fraction models outside the regions for which their coefficients where estimated, even at similar latitudes or at a-priori similar climates, since average atmospheric composition may be different due to natural or human-related causes. Ideally, a local assessment of a proposed model against good quality local diffuse irradiation data should be considered.

Acknowledgements

The authors aknowledge the reviewers for their valuable comments and thank C. Raichijk (GERSOLAR-INEDES, Universidad Nacional de Luján, Argentina) for helpful comments on early versions of this manuscript. We thank R. Righini (GERSOLAR-INEDES, Universidad Nacional de Luján, Argentina) for providing us with the data for the Luján site and S. Arizcorreta, from INUMET (Uruguay), for recording and providing us the data for the Salto site. The authors acknowledge financial support from CSIC (UDELAR), from Agencia Nacional de Investigación e Innovación (ANII-FSE-2011-5976 and FSE-2013-10919) and from PEDECIBA (Uruguay).

References

[1] C. Gueymard, Solar Energy 82 (2007) 272.

[2] G. Abal et al., Anales de la IV Conferencia Latinoamericana de Energía Solar (IV ISES-CLA), pp. 1–12, Universidad Nacional del Cusco, Perú, 2010.

[3] R. Alonso Suárez et al., Annals of the Solar World Congress (SWC 2011), Kassel, Germany, 2011.

[4] R. Alonso Suárez et al., Solar Energy 86 (2012) 3205.

[5] R. Alonso Suárez, M. D'Angelo and G. Abal, Proceedings IV Congresso Brasileiro de Energia Solar e V Conferencia Latino-Americana da ISES, Sao Paulo, Brazil, 2012.

[6] C. Gueymard and J. Ruiz-Arias, Solar Energy 128 (2016) 1.

[7] S. Dervishi and A. Mahdavi, Solar Energy 86 (2012) 1796.

[8] H. Li et al., Renewable Energy 36 (2011) 1944.

[9] R. Tapakis, S. Michaelides and A. Charalambides, Solar Energy (2014) .

[10] C. Jacovides et al., Renewable Energy 31 (2006) 2492.

[11] C. Raichijk and F. Taddei, Avances en Energías Renovables y Medio Ambiente - ASADES 16 (2012) 11.23.

[12] C. Gueymard, Solar Energy 83 (2009) 432.

[13] D.T. Reindl, W.A. Beckman and J.A. Duffle, Solar Energy 45 (1990) 1.

[14] B. Ridley, J. Boland and P. Lauret, Renewable Energy 35 (2010) 478.

[15] J. Ruiz-Arias et al., Energy Conversion and Management 51 (2010) 881.

[16] A. Skartveit, J.A. Olseth and M.E. Tuft, Solar Energy 63 (1998) 173.

[17] J. Boland, B. Ridley and B. Brown, Renewable Energy 33 (2008) 575.

[18] A. Drummond, Arch. für Meteorologie, Geophysik und Bioklimatologie, Ser. B 7 (1956) 413.

[19] G. Sánchez et al., JOURNAL OF GEOPHYSICAL RESEARCH 117 (2012) D09206.

[20] C. Raichijk, Avances en Energías Renovables y Medio Ambiente 16 (2012) 11.17.

[21] . ISO Technical Committee ISO/TC 180, Solar Energy Sub-committee SC1: Climate, Measurement and Data, International Organization for Standarization (ISO) preprint ISO 9847:1992(E) (1992).

[22] J. Wood, Delta-T Devices preprint Rev. 3 (2015), also: private communication by Stephen Williams from Delta-T Devices.

[23] D. Myers and S. Wilcox, National Renewable Energy Laboratory preprint NREL/CP-550-45374 (2009).

[24] B. Psiloglou, S. Lykoudis and D. Kouvas, Advances in Meteorology, Climatology and Atmospheric Physics (Springer-Verlag, 2012) chap. ; Performance Assesment of an integrated sensor for simultaneous measurements of global and diffuse radiation components at Athens area, pp. 259–264.

[25] J. Badosa et al., Atmos. Meas. Tech. 7 (2014) 4267.

[26] M. Journée and C. Bertrand, Solar Energy 85 (2011) 72.

[27] S. Younes, R. Claywell and T. Muneer, Energy 30 (2005) 1533.

[28] T. Muneer, Solar radiation and daylight models, 2nd ed. ed. (Elsevier-Butterworth-Heineman, Oxford, 2004).

[29] C. Rigollier, O. Bauer and L. Wald, Solar Energy 68 (2000) 33.

[30] J. Tovar-Pescador, Modelling the Statistical Properties of Solar Radiation and Proposal of a Technique Based on Boltzmann Statistics, Modeling Solar Radiation at the Earth's Surface: Recent Advances, edited by V. Badescu, chap. 3, pp. 55–91, Springer, 2008.

[31] A. Ianetz and A. Kudish, A Method for Determining the Solar Global and Defining the Diffuse and Beam Irradiation on a Clear Day, Modeling Solar Radiation at the Earth's Surface: Recent Advances, edited by V. Badescu, chap. 4, pp. 93–112, Springer, 2008.

[32] J. Duffie and W. Beckman, Solar Engineering of Thermal Processes, Third ed. (Wiley and Sons, Hoboken, New Jersey, 2006).

[33] J. Orgill and G. Hollands, Solar Energy 19 (1977) 357.

[34] D. Erbs, S. Klein and J. Duffie, Solar Energy 28 (1982) 293.

[35] A. Skartveit and J. Olseth, Solar Energy 38 (1987) 271.

[36] J. Boland, L. Scott and M. Luther, Environmetrics 12 (2001) 575.

[37] C. Jacovides et al., Renewable Energy 35 (2010) 1820.

[38] R. Perez et al., ASHRAE Transactions 98 (1992) 354.

[39] F. Kasten and A. Young, Applied Optics 28 (1989) 4735.

[40] C. Gueymard, Renewable and Sustainable Energy Reviews 39 (2014) 1024.

[41] R. Stone, Solar Energy 51 (1993) 289.

[42] M. de Simón-Martín, M. Diez-Mediavilla and C. Alonso-Tristán, Solar Energy (2016), doi 10.1016/j.solener.2016.09.026.

[43] B. Espinar et al., Solar Energy 83 (2009) 118.

[44] R. Tapakis, A. Michaelides and A. Charalambides, Solar Energy (2015).

[45] C. Raichijk and A. Fasulo, Avances en Energías Renovables y Medio Ambiente - ASADES 13 (2009) 11.17.

[46] M. Collares-Pereira and A. Rabl, Solar Energy 22 (1979) 155.

[47] S. Klein, Solar Energy 19 (1977) 325.

Appendix A. Glossary

Symbol	Name	Unit
GHI	global horizontal irradiance	Wm^{-2}
DHI	diffuse horizontal irradiance	Wm^{-2}
DNI	beam or direct normal irradiance	Wm^{-2}
I_h	global horizontal hourly irradiation	Whm^{-2}
I_{dh}	diffuse horizontal hourly irradiation	Whm^{-2}
I_{bh}	beam horizontal hourly irradiation	Whm^{-2}
f_d	hourly diffuse fraction = I_{dh}/I_h	
I_{0h}	extraterrestrial hourly horizontal irradiation	Whm^{-2}
k_t	hourly clearness index = I_h/I_{0h}	
θ_z	solar zenith angle	rad
α_s	solar altitude angle	rad
δ	solar declination angle	rad
ϕ	latitude	rad
I_{dc}	clear-sky diffuse hourly horizontal irradiation	Whm^{-2}
I_{bc}	clear-sky beam hourly irradiation	Whm^{-2}
I_c	clear-sky global horizontal irradiation	Whm^{-2}
I_{sc}	hourly solar constant = 1367	Whm^{-2}
ϵ	eccentricity of the earth's orbit	
m	air mass	
T_L	Linke Turbidity at $m = 2$	
δ_R	Rayleigh optical thickness	
T_{rd}	diffuse transmittance function	
F_{da}	diffuse angular function	
H_{0h}	extraterrestrial daily irradiation = $\sum_{\text{day}} I_{0h}$	MJm^{-2}
H_h	global daily horizontal irradiation = $\sum_{\text{day}} I_h$	MJm^{-2}
K_T	daily clearness index = H_h/H_{0h}	
ω_s	sunset hour angle	rad
H_{dh}	diffuse daily horizontal irradiation = $\sum_{\text{day}} I_{dh}$	MJm^{-2}
F_d	daily diffuse fraction	
$\overline{H}_{0h}$	monthly mean extraterrestrial daily irradiation	MJm^{-2}
$\overline{H}_h$	monthly mean global daily horizontal irradiation	MJm^{-2}
$\overline{K}_T$	monthly mean clearness index = $\overline{H}_h/\overline{H}_{0h}$	
$\overline{H}_{dh}$	monthly mean diffuse daily horizontal irradiation	MJm^{-2}
$\overline{F}_d$	monthly mean diffuse fraction = $\overline{H}_{dh}/\overline{H}_h$	