Response to Referee #1

We thank the referee for all the valuable comments that have improved this manuscript. As a result, more information and analyses have been included. Additionally, following the referee's suggestion, the writing has been improved and the grammar mistakes have been corrected. Please see below our point-by-point replies to the specific comments, with the referee's comments in black colour and our corresponding replies colored in orange.

Comment 1 (Page 1, line 7): The title is rather general while the topic of the manuscript is rather limited and does not at all cover what is promised in the title. The words "surface" and "irradiance" should be included in the title and also preferable the location for which the study was made. A suggestion for the title is "Modeling the erythemal surface diffuse irradiance fraction for Badajoz, Spain".

Response 1: Following the referee's suggestion the title has been rewritten as: "Modeling the erythemal surface diffuse irradiance fraction for Badajoz, Spain"

C2 (Page 1, line 7): "inspired on" → "inspired from".

R2: The text has been corrected (line 9)

C3 (Page 1, line 12): The RAU3 acronym has no meaning unless the full manuscript is read. Maybe rather write "the best performing model (RAU3) is based on a model proposed by Ruiz-Arias et al. [2010] and shows values . . . ".

R3: The text has been rewritten following this suggestion (line 13-14).

C4 (Page 1, lines 13-14): Maybe write this sentence as: "The performance achieved by this entirely empirical model is better than those obtained by previous semi-empirical approaches, and therefore needs no additional information from other physically-based models."

R4: The text has been rewritten following the referee's suggestion (lines 14-16).

C5 (Page 2, line 40): Craig et al. reference missing in References.

R5: The correct reference is: Williamson et al. [2014] (line 40).

This reference was included in the reference list of the original manuscript as:

Williamson, C., R. Zepp, R. Lucas, S. Madronich, A. Austin, C. Ballare, M. Norval, B. Sulzberger, A. Bais, R. McKenzie, S. Robinson, D. Häder, N. Paul, and J. Bornman. Solar ultraviolet radiation in a changing climate, *Nature Climate Change*, 4, 434-441, 2014.

C6 (Page 2, line 41): It is the variations in the clouds and aerosols that affect the diffuse/direct partitioning and not the variations in the ultraviolet irradiance. Please clarify this sentence.

R6: This sentence has been rewritten as follows (line 40):

"*These variations in clouds and aerosols may affect not only the amount but also the diffuse/direct partitioning due to the stronger effectiveness of scattering at shorter wavelengths.*"

C7 (Page 2, line 55): Please include a reference and/or an example of a physically-based model.

R7: Models libRadtran [Mayer and Kylling, 2005], SBDART [Ricchiazzi et al., 1998] and TUV [Madronich and Flocke, 1997] have been mentioned in line 53-54. Their corresponding references have been also added to the reference list.

C8 (Page 2, line 58): You state that empirical model have a "wide use by the scientific community". May you please provide some references reflecting the wide use? In the next sentence, line 60, it is stated that "few studies have applied empirical models . . . " contradicting your claim on line 58.

R8: Lines 53-70 have been rewritten in order to have a clearer text and some references have been included. The final text is as follows (lines 56-62) :

"Hence, in this paper the empirical approach was preferred because of its simplicity and modest requirements in terms of ancillary data. The empirical approach has been widely use by the scientific community to estimate the diffuse component in the total solar spectrum [Orgill and Hollands, 1977; Iqbal, 1983; Reindl et al., 1990; Gonzalez and Calbo, 1999; De Miguel et al., 2001; Boland et al., 2008; Ridley et al., 2010; Ruiz-Arias, 2010; Engerer, 2015].

In the particular case of the UV range, the complexity in modeling the diffuse component increases due to the higher effectiveness of the Rayleigh scattering. As far as we are aware, ..."

C9 (Page 2, line 69): "contribution in" → "contribution to".

R9: The text has been corrected.

C10 (Page 3, line 78): The acronym UVI is introduced without explaining what it is and how it is related to UVER. It is not used later in the manuscript. Please explain UVI.

R10: Since UVER is explained in the next paragraph and we are here describing the climate general conditions, it has been prefered not to mention the UVI, and rewrite the text as follows (line 82):

" *... leading to noon irradiance values among the highest in Europe."*

C11 (Page 3, lines 84-86): It may be argued that at least two important processes that largely affect UVER are not included in the dataset used here. One is the effect of snow on the surface which significantly changes UVER. The other is altitude. Both these processes affect both total UVER and the direct/diffuse ratio. It should at least be mentioned that these processes are not included in the dataset and that the proposed models thus have not been tested for these processes.

R11: Following the referee's comment, the text has been rewritten to arrive at (lines 89-95):

"The period analyzed in this study comprises years 2011 and 2012, which ensures that a large variety of seasonal processes and meteorological conditions are sampled. The large variety of sun-geometry and meteorological situations that occur during a year guarantees the representativeness of the dataset for the proposal and assessment of empirical models for our location. However, it must be mentioned that snow and altitude are additional factors that have not been considered in this study. They are not represented in the dataset and the proposed models have not been tested for the processes they involve. These factors can significantly affect total UVER and the direct/diffuse ratio and, therefore, should be included for high and snowed locations."

C12 (Page 4, lines 109-110): On line 84 you argue that the dataset includes a "large variety of situations", then you reduce the variety significantly by hourly averaging this dataset. Please provide a sound justification that the reduced dataset also includes a "large variety of situations". And it should be shown that the reduced dataset with hourly resolution produces similar results as the original higher time resolution dataset. Should not the proposed model also account for short-term fluctuations? If not, why are not these important for users of these type of models? If short-term fluctuations are not included, this should be stated as a limitation of the model. Did you hourly average the solar zenith angle? It would seem more appropriate to average the cosine of the solar zenith angle. May you please comment on this?

R12: Line 84 refers to the large variety of sun-geometry and meteorological situations that are sampled by the two year period of study. By including two years of hourly data, the diurnal and annual cycles are suitably described. The text has been rewritten as follows:

"The period of study comprises years 2011 and 2012, which ensures that a large variety of seasonal processes and meteorological conditions are sampled. The large variety of sun-geometry and meteorological situations that occur during a year guarantees the representativeness of the dataset for the proposal and assessment of empirical models for our location."

Additionally, the selection of the hourly scale has been justified as follows (lines 123-129):

" In this study, hourly data has been used similarly similarly to the majority of previous studies [Reindl et al., 1990; González and Calbó, 1999; Boland et al., 2001 and 2008; Ridley et al., 2010, Ruiz-Arias et al., 2010]. According to Ruiz-Arias [2010], while random errors are much lower than shorter intervals, it offers an appropriate agreement between data availability and the inherent solar radiation temporal variability. Thus, this temporal frequency is the one used by many applications, such as house energy ratings scheme software [Boland et al., 2001]. As a consequence, most of the statistical models are based on the hourly interval of the solar radiation data [Ruiz-Arias 2010, Gueymard and Ruiz-Arias, 2016]."

Additionally, in order, three parameters proposed by Gonzalez and Calbo [1999] to account for the short-term fluctuations and another two proposed by Ruiz-Arias et al. [2010] have been included in some models. This way, the relevance of the short-term scales can be tested by comparing models with and without these parameters. The results of this comparison have been included as follows (lines 359-371):

"Main results of the fitting of each empirical model to the fitting subset are summarized in Table 1. Ordinary least squares fitting for models REU, GCU1, GCU2, GCU3, BOU, RIU and KUU, and non-linear fitting for models RAU1, RAU2 and RAU3 have been calculated. As mentioned in Section 3, some models involve parameters accounting for the short-term fluctuation. In particular, models based on Gonzalez and Calbo [1999], that is, GCU1, GCU2 and GCU3, include parameters $\Delta_{UVER,1}$, $\Delta_{UVER,2}$, $\Delta_{UVER,3}$, respectively. In the case of model RAU3, inspired from Ruiz-Arias et al. [2010], the parameters kt^2 and m^2 are introduced to account for the short-term variability. Results in Table 1 show that model GCU2, with a short-term variability parameter, presents a better performance than model REU, which has a same functional form without a short-term variability parameter. Conversely, models GCU1 and GCU3 do not show a better performance than REU despite to include short-term variability parameters. This primacy of $\Delta_{UVER,2}$ agrees with the case of total diffuse fraction as reported by Gonzalez and Calbó [1999]. In the case of models inspired from Ruiz-Arias et al. [2010], RAU1, RAU2 and RAU3, a notable improvement is observed in the values of R^2 and rRMSE when the parameters kt^2 and m^2 are introduced in RAU3. This result is in line with results reported by Ruiz-Arias et al. [2010] for the total diffuse fraction. Therefore, hereafter, only GCU2 and RAU3 will be hereinafter considered."

Regarding the averaging, it was indeed cos(SZA) what was averaged and not SZA. The text has been rewritten to clarify it as follows (lines 120-123):

"Global and diffuse UVER measurements were recorded every minute by a Campbell Scientific CR-1000 data logger. Based on these data and the time of each measurement, a one-minute dataset consisting in the UVER diffuse fraction, UVER transmissivity, relative optical mass and cosine of the solar zenith angle was built. Subsequently these quantities were averaged hourly."

C13 (Page 4, line 119): "fraction determines the" → "fraction describes the".

R13: The text has been corrected.

C14 (Page 5, line 149): It should be mentioned that the solar constant is not a constant and that it may vary over the solar cycle, especially in the UV, see for example Lean et al. [1992] and Kopp and Lean [2011]. Please mention the uncertainties in the estimated erythemally-weighted solar constant due to the variations in the solar constant.

R14: This point has been included in the text as follows (lines 168-172):

"S_{UVER} is the erythemally-weighted solar constant, with an estimated value of 10.031 W/m². It must be mentioned that the solar constant may vary over the solar cycle, mainly in the very short UV wavelengths [Lean et al, 1992; Kopp and Lean, 2011]. In the case of the erythemally-weighted irradiance at surface, the wavelength interval of interest starts at 290 nm and, therefore, the variation is lower than 1% [Floyd et al, 2002; DeLand and Cebula 2012; Yeo et al, 2015]."

C15 (Page 5, line 160): "radiometric magnitudes" → "radiometric quantities".

R15: The text has been corrected.

C16 (Page 6, line 185): What is meant by "adding the a term"? Please clarify sentence.

R16: The text has been corrected. It should have said: "... adding a term...".

C17 (Page 6, line 203): "in clear" → "on clear".

R17: The text has been corrected.

C18 (Page 10, line 319): The description of xi* and xi should follow immediately after Eqs. (19) and (20) and not later in a different paragraph, lines 325-326.

R18: The definition of xi* and xi has been now written just after Eqs. (19) and (20) .

C19 (Page 11, line 334): When reading this sentence I thought results from all models would be included in Table 1. However, the results for GCU1, RAU1 and RAU2 are excluded. It is stated that results for these models were calculated. Hence please include these results in Table 1 to make the manuscript complete.

R19: The results of the statistics R^2 and rRMSE for models GCU1, GCU3, RAU1 and RAU2 have been included in Table 1.

C20 (Page 12, Conclusions): It should be mentioned in the Conclusions that the models have not been tested for high albedo (snow) conditions nor for high altitudes. Also, the testing have been limited to solar zenith angles less than 70º. Hence the models may not be suitable for the large solar zenith angles encountered at high latitudes.

R20: This points has been mentioned in the Conclusions section as follows (157-164):

"*It should be mentioned that factors affecting the UVER diffuse fraction such as the altitude or surface albedo have not been tested in this study. Moreover, since only solar zenith angles below 70º have been considered, the models may not be suitable for the large solar zenith angles encountered at high latitudes. Therefore, similar research must be conducted to assess the general validity of the proposed models and/or establish their possible adaptation to other locations. Although these results apply mainly to regions with similar characteristics than those analyzed in this study, the methodology and comparisons described in this paper can be used to develop similar analysis for other locations.*"

C21 (Page 22, lines 485-486): It should be stated in the caption that the black dot is the observation. Also describe what value the green lines represents.

R21: This information has been included in the caption of Figure 2 as follows:

"*Figure 2. Taylor diagram showing the performance of the models proposed to estimate the diffuse fraction, as compared to experimental measurements. This diagram summarizes different aspects of the performance of a model such as the centered root-mean-square error (green lines), the correlation and the standard deviation with respect to the reference data set (black dot).*"

C22 (Page 23, Table 2, caption): This tables does not only list the "Empirical fitting coefficients and their corresponding standard error" as stated in the caption, but the full functional form of the models. This should be mentioned in the caption. It should also be mentioned in the caption that these coefficients are only applicable to the Badajoz, Spain, site. This to avoid that others misuse these equations for other locations.

R22: This information has been included in the caption of Table 2 as follows:

"*Table 2. Functional form of the models and empirical fitting coefficients with their corresponding standard error for Badajoz, Spain.*"