

1 The generalized additive model for the assessment of the direct, diffuse and global
2 solar irradiances using SEVIRI images, with application to the UAE

3

4 T.B.M.J. Ouarda^{1, 2*}, C. Charron¹, P.R. Marpu¹ and F. Chebana²

5

6 ¹Institute Center for Water and Environment (iWATER), Masdar Institute of Science
7 and Technology, P.O. Box 54224, Abu Dhabi, UAE

8

9 ²INRS-ETE, National Institute of Scientific Research, Quebec City (QC), G1K9A9,
10 Canada

11

12

13 *Corresponding author:

14 Email: touarda@masdar.ac.ae

15 Tel: +971 2 810 9107

16

17

18

December 2015

19 **Abstract**

20 Generalized additive models (GAMs) can model the non-linear relationship between a
21 response variable and a set of explanatory variables through smooth functions. GAM
22 is used to assess the direct, diffuse and global solar components in the United Arab
23 Emirates, a country which has a large potential for solar energy production. Six
24 thermal channels of the SEVIRI instrument onboard Meteosat Second Generation are
25 used as explanatory variables along with the solar zenith angle, solar time, day
26 number and eccentricity correction. The proposed model is fitted using reference data
27 from three ground measurement stations for the full year of 2010 and tested on two
28 other stations for the full year of 2009. The performance of the GAM model is
29 compared to the performance of the ensemble of artificial neural networks (ANN)
30 approach. Results indicate that GAM leads to improved estimates for the testing
31 sample when compared to the bagging ensemble. GAM has the advantage over ANN-
32 based models that we can explicitly define the relationships between the response
33 variable and each explanatory variable through smooth functions. Attempts are made
34 to provide physical explanations of the relations between irradiance variables and
35 explanatory variables. Models in which the observations are separated as cloud-free
36 and cloudy and treated separately are evaluated along with the combined dataset.
37 Results indicate that no improvement is obtained compared to a single model fitted
38 with all observations. The performance of the GAM is also compared to the McClear
39 model, a physical based model providing estimates of irradiance in clear sky
40 conditions.

41 **1. Introduction**

42 Solar radiation reaching the earth is divided into different components. Direct
43 normal irradiance (DNI) refers to the radiation received from a straight beam of light
44 from the direction of the sun at its current position to a surface that is always normal
45 to that solar beam. Diffuse horizontal irradiance (DHI) is the radiation received by a
46 horizontal surface from radiation scattered by the atmosphere and coming from all
47 directions. Global horizontal irradiance (GHI) is the total amount of radiation received
48 on a surface parallel to the ground. Assessment of solar radiation on the earth's
49 surface is of primary importance for many applications in solar energy. For instance,
50 the accurate assessment of DNI is needed for concentrating solar power systems or
51 other installations that track the position of the sun. To model the global tilt irradiance
52 for fixed flat plate collectors, the assessment of DNI, DHI and GHI is required [1, 2].

53 Solar resource assessment is crucial for efficient realization of solar energy
54 applications, but is often limited by the lack of sufficient ground measurements which
55 incur high costs [3]. Infrared images acquired by satellites at different frequencies can
56 characterize earth's emission and the atmospheric constituents, which can be used to
57 obtain estimates of solar radiation information in areas where there are no ground
58 measurements. Since satellite data are continuous in time and space, it would be
59 possible to perform solar resource assessment over the entire region. Solar maps
60 derived from satellite based methods have been proven to be more efficient than
61 interpolation of solar data from ground measurements [4].

62 Data acquired from satellite images have been extensively used for estimation
63 of solar radiation on the earth's surface. Several models classified as Physics-based,
64 empirical and hybrid models were proposed with a good adaptation for the regions of

65 interest. An example of the physics-based modeling is the model of Gautier *et al.* [5]
66 to estimate the GHI in North America. It was later on adapted by Cogliani *et al.* [6]
67 using Meteosat images to produce SOLARMET. The original Heliosat model of Cano
68 *et al.* [7] was used to estimate GHI, DNI and DHI over the USA. It was later adapted
69 by Perez *et al.* [8] for GOES images. The operational physical model of Schillings *et*
70 *al.* [3] was used to estimate DNI from Meteosat images. The Heliosat model has been
71 modified and improved through different versions [9-14]. Heliosat-4 model is being
72 currently validated [15, 16]. Another Physics-based model for cloud-free conditions is
73 the McClear model [17], which is based on look-up-tables established with the
74 radiative transfer model libRadtran [18].

75 On the other hand, data-driven statistical approaches have also been frequently
76 used to perform solar radiation assessment. Artificial neural networks (ANN) have
77 been used successfully in a wide range of fields (See for instance [19-21]). They have
78 been adapted for solar resources assessment in a number of studies [22-28]. In these
79 studies, location dependent parameters and meteorological parameters were used as
80 inputs to model solar irradiance components. ANNs with an ensemble approach,
81 which provide better generalization compared to a single ANN [29, 30], were used in
82 Eissa *et al.* [23] to retrieve irradiance components over the UAE. A simple bagging-
83 like approach was used to develop the ensemble models. Alobaidi *et al.* [22] further
84 improved on this model, by introducing a novel ensemble framework which
85 significantly improved the results compared to the results obtained in the previous
86 studies. The model employs a two-stage resampling process to build ensemble
87 architectures for non-linear regression. Though the model performs well, it involves
88 an ensemble of ensembles framework resulting in high computation load apart from
89 the number of computationally expensive optimization steps while training the

90 architecture. Another biggest drawback of ANN type models is that the relations
91 between the inputs and outputs cannot be explicitly presented.

92 In this work, we propose to use the generalized additive model (GAM), which
93 is an extension of the generalized linear model (GLM) which uses non-parametric
94 smooth functions to relate explanatory variables to the response variable. This flexible
95 method represents an interesting approach to model the complex relation between
96 irradiance and explanatory variables. GAMs have been applied widely in
97 environmental studies [31-35], and in public health and epidemiological studies [36-
98 41]. However, GAMs have never been used for solar irradiance assessment. An
99 advantage of GAM over ANN is that the relationship between each predictor and the
100 response variable is made explicit through a set of smooth functions.

101 The United Arab Emirates (UAE) presents a high potential for solar energy
102 development due to the long day light period and the marginal amount of cloud cover.
103 Recently, Eissa *et al.* [23, 42] and Alobaidi *et al.* [22] developed models to accurately
104 estimate irradiance components over the UAE territory in which they used images of
105 the earth's surface acquired by the Spinning Enhanced Visible and Infrared Imager
106 (SEVIRI) onboard Meteosat Second Generation (MSG) satellite.

107 The aim of the present paper is to use GAM for the assessment of the
108 irradiance components DHI, DNI and GHI using SEVIRI satellite images. Following
109 previous work, six SEVIRI thermal channels along with the solar zenith angle (θ_z),
110 solar time (Time), day number (Day) and eccentricity correction (ϵ) are used as
111 explanatory variables in the model.

112 In Eissa *et al.* [23], DHI was directly estimated with the ANN but DNI was
113 deduced from the ANN estimated optical depth (δ) and GHI was deduced from DNI

114 and DHI estimates. In the present study, we propose also to estimate directly the DNI
115 and GHI with GAM. In Eissa *et al.* [23] and Alobaidi *et al.* [22], an algorithm was
116 used to separate the training and the testing dataset as cloud-free and cloudy sub-
117 datasets. ANN models were then trained and tested separately on the two sky
118 condition samples. While this approach is also considered in the present work, we
119 additionally propose to develop a global model to the all sky training dataset and to
120 validate it on the cloud-free, cloudy and all sky testing datasets. GAM allows
121 explicitly defining the relationship between the response variable and each
122 explanatory variable through smoothing functions. Attempts to find physical
123 interpretations of the shape of these curves are made in the present work.

124 A comparison is also made with the McClear model, a physical based model
125 providing estimates of irradiance in clear sky conditions. The results of McClear
126 model are available through a web service at the website of the MACC project
127 (Monitoring Atmospheric Composition and Climate project) ([http://www.gmes-](http://www.gmes-atmosphere.eu)
128 [atmosphere.eu](http://www.gmes-atmosphere.eu)). Estimates could be obtained by just providing the latitude, longitude
129 and the altitude (optional) of the target site, and the period of interest.

130

131 **2. Data**

132 Ground measurements for DHI, DNI and GHI consist of 10 min resolution
133 data available at 5 stations over the UAE. At each station, data are collected using a
134 Rotating Shadowband Pyranometer (RSP). GHI is measured by the pyranometer
135 when the shadowband is stationary. The shadowband makes a full rotation around the
136 pyranometer. DHI is given by the lowest measured irradiance since at that moment
137 DNI is completely blocked by the shadowband. DNI is deduced from GHI and DHI

138 measured with the RSP. In the following, ground measured DNI refers to DNI that is
139 estimated from ground measured GHI and DHI. To match the 15 min resolution of the
140 satellite data, successive ground measured data were interpolated. Data are available
141 for the full year 2009 at the stations of Masdar City, Al Aradh and Madinat Zayed,
142 and for the full year 2010 at all stations. Fig. 1 presents the spatial distribution of the
143 stations across the UAE.

144 Satellite images of the SEVIRI optical imager onboard MSG satellite were
145 used in the present study. They provide continuous images of the earth in 12 spectral
146 channels with a temporal resolution of 15 min and a spatial resolution of 3 km.
147 Images from 6 thermal channels, T04 (3.9 μm), T05 (6.2 μm), T06 (7.3 μm), T07 (8.7
148 μm), T09 (10.8 μm) and T10 (12.0 μm) were collected and converted into brightness
149 temperature. For each station, 3-by-3 pixels, with the station located in the center
150 pixel, were extracted from satellite data. The other variables, solar zenith angle (θ_z),
151 Time, Day and eccentricity correction ε were computed for each pixel. The choice of
152 the selected thermal channels is justified in Eissa *et al.* (2013) by their sensitivity to
153 the different constituents of the atmosphere: channel T05 and T06 are known to be
154 affected by water vapor and T07, T08 and T09 are frequently used for dust detection.
155 T04 was also selected in Eissa *et al.* (2013) because it had improved their model
156 accuracy.

157 The dataset is divided into training and testing datasets. The model is
158 developed using the training dataset and tested using the testing dataset. The training
159 dataset includes data from the stations of Masdar City, East of Jebel Hafeet and Al
160 Wagan for the full year 2010. The testing dataset includes data from the stations of Al
161 Aradh and Madinat Zayed for the full year 2009. The training and testing datasets are

162 further divided respectively into cloud-free and cloudy datasets. For this, a cloud
163 mask was applied in which each pixel was classified as cloud-free or cloudy. The thin
164 cirrus test [43], employing the T09 and T10 channels of SEVIRI, was used as a cloud
165 mask following [23]. In all, the cloud-free and cloudy training datasets contain 29193
166 and 7086 observations respectively, and the cloud-free and cloudy testing datasets
167 contain 16864 and 2856 observations respectively.

168

169 **3. Methodology**

170 **3.1 Generalized Additive Model**

171 GLMs [44] generalize the linear model with a response distribution other than
172 normal and a link function relating the linear predictor with the expectation of the
173 response variable. Let us define Y , a random variable called response variable, and \mathbf{X} ,
174 a matrix whose columns are a set of r explanatory variables X_1, X_2, \dots, X_r . The GLM
175 model is defined by:

$$176 \quad g[E(Y | \mathbf{X})] = \alpha + \sum_{j=1}^r \beta_j X_j, \quad (1)$$

177 where g is the link function and β_j and α are unknown parameters. With GLM, the
178 distribution of Y is generalized to have any distribution within the exponential family.
179 The role of the link function is used to transform Y to a scale where the model is
180 linear.

181 The GAM model [45] is an extension of the GLM in which the linear predictor
182 is replaced by a set of non-parametric functions of the explanatory variables. GAM
183 can then be expressed by:

184 $g[E(Y | \mathbf{X})] = \alpha + \sum_{j=1}^r f_j(X_j),$ (2)

185 where f_j are smooth functions of X_j . This model is more flexible by allowing non-
 186 linear relations between the response variable and the explanatory variables through
 187 the smooth functions. Because of the additive structure of GAM, the effect of each
 188 explanatory variable on Y can be easily interpreted. A smooth function can be
 189 represented by a linear combination of basis functions:

190 $f_j(x_j) = \sum_{i=1}^{q_j} \theta_{ji} b_{ji}(x_j),$ (3)

191 where $b_{ji}(x_j)$ is the i th basis function of the j th explanatory variable evaluated at
 192 x_j , q_j is the number of basis functions for the j th explanatory variable and θ_{ji} are
 193 unknown parameters.

194 Given a basis function, we define a model matrix \mathbf{Z}_j for each smooth
 195 function where the columns of \mathbf{Z}_j are the basis functions evaluated at the values of
 196 the j th explanatory variable. Eq. (2) can be rewritten as a GLM in a matrix form as:

197 $g(E(\mathbf{y})) = \mathbf{Z}\boldsymbol{\theta},$ (4)

198 where \mathbf{y} is a vector of observed values of the response variable Y , \mathbf{Z} is a matrix
 199 including all the model matrix \mathbf{Z}_j and $\boldsymbol{\theta}$ is a vector including all the smooth
 200 coefficient vectors $\boldsymbol{\theta}_j$. Parameters $\boldsymbol{\theta}$ could be estimated by the maximum likelihood
 201 method, but if q_j is large enough, the model will generally overfit the data. For that
 202 reason, GAM is usually estimated by penalized likelihood maximization. The penalty
 203 is typically a measure of the wiggleness of the smooth functions and is given by:

204 $\boldsymbol{\theta}_j^T \mathbf{S}_j \boldsymbol{\theta}_j$ for the j th smooth function where \mathbf{S}_j is a matrix of known coefficients. The
 205 penalized likelihood maximization objective is then given by:

$$206 \quad l_p(\boldsymbol{\theta}) = l(\boldsymbol{\theta}) - \frac{1}{2} \sum_j \lambda_j \boldsymbol{\theta}_j^T \mathbf{S}_j \boldsymbol{\theta}_j. \quad (5)$$

207 where $l(\boldsymbol{\theta})$ is the likelihood of $\boldsymbol{\theta}$ and λ_j are the smoothing parameters which
 208 control the degree of smoothness of the model. For given values of the parameters λ_j
 209 , the GAM penalized likelihood can be maximized by penalized iterative re-weighted
 210 least squares (P-IRLS) to estimate $\boldsymbol{\theta}$ (see [46]). However, λ_j should be estimated by
 211 an iterative method like Newton's method [46]. For each trial of λ_j , the P-IRLS is
 212 iterated to convergence. In this study, λ_j are optimized by minimizing the
 213 generalized cross validation score (GCV), which is based on the leave-one-out
 214 method. This method ends up being computationally less expensive as it can be
 215 shown that the GCV score equals:

$$216 \quad v_g = \frac{n \|\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\theta}}\|^2}{[n - \text{tr}(\mathbf{A})]^2}, \quad (6)$$

217 where $\mathbf{A} = \mathbf{Z}(\mathbf{Z}\mathbf{Z} + \lambda\mathbf{Z})^{-1}\mathbf{Z}^T$ is the influence matrix. In this study, all GAM model
 218 parameters are estimated with the R package mgcv [46].

219 The smooth functions used in this study are cubic regression splines. Cubic
 220 splines are constructed with piecewise cubic polynomials joined together at points
 221 called knots. The definition of the cubic smoothing spline basis arises from the
 222 solution of the following optimization problem [47]: Among all functions $f(x)$, with

223 two continuous derivatives, find one that minimizes the penalized residual sum of
224 squares:

$$225 \quad \sum_{i=1}^n \{y_i - f(x_i)\}^2 + \lambda \int_a^b f''(x)^2 dx, \quad (7)$$

226 where y_1, y_2, \dots, y_n is a set of observed values of the response variable and
227 x_1, x_2, \dots, x_n a set of observed values of an explanatory variable, λ is the smoothing
228 parameter, and $a \leq x_1 \leq x_2 \leq \dots \leq x_n \leq b$. The first term of (7) measures the degree of
229 fit of the function to the data, while the second term adds a penalty for the curvature
230 of the function, and the smoothing parameter controls the degree of penalty given for
231 the curvature in the function. With regression splines, the numbers of knots can be
232 considerably reduced, and the position of the knots needs to be chosen. In fact, with
233 cubic penalized splines, the exact location of the knots and their numbers are not as
234 important as the smoothing parameters. In this study, the positions of the knots will be
235 evenly spaced along the dimension of each explanatory variable.

236 **3.2 Model configurations**

237 For the GAM models of this study, the identity link function and the Gaussian
238 error with mean zero and a constant variance σ^2 are assumed. In each model,
239 residuals obtained are checked for any trends in the variance and for normality to
240 confirm the model assumptions. In Eissa *et al.* [23], DHI was estimated directly with
241 the ANN trained with ground measured DHI. The model for the GAM estimated DHI
242 is given by the following expression:

$$243 \quad \text{DHI} = \alpha + \sum_{j=1}^r f_j(X_j), \quad (8)$$

244 where X_j is the j th explanatory variable, r is the number of explanatory variable
245 included in the model and α is the intercept.

246 DNI estimations in Eissa *et al.* [23] were deduced from the ANN estimated δ .
247 DNI estimations were then computed using the Beer-Bouguer-Lambert law which
248 relates δ to DNI by the following equation:

$$249 \quad \text{DNI} = I_0 \varepsilon \exp(-m\delta), \quad (9)$$

250 where I_0 is the solar constant with an approximate value of 1367 W/m², and m is the
251 air mass. The values of δ were computed from ground measured DNI. Parameters m
252 and ε can be easily computed for any location on a given day by knowing θ_z . For
253 the estimation of δ with GAM, the following model is used:

$$254 \quad \log(\delta) = \alpha + \sum_{j=1}^r f_j(X_j). \quad (10)$$

255 The logarithmic transformation of δ in (10) is used to meet the model assumptions.
256 Fig. 2 presents the residuals against the fitted values for the models with and without a
257 logarithmic transformation. Fig. 2b clearly improves the residual constant variance
258 assumption. In this study, we also propose to estimate DNI directly with GAM fitted
259 on ground measured DNI. Estimated DNI is then denoted by DNI^D and the following
260 model similar to that of DHI, is used:

$$261 \quad \text{DNI}^D = \alpha + \sum_{j=1}^r f_j(X_j). \quad (11)$$

262 The GHI is deduced from the estimated DHI and DNI using the following
263 relation:

264 $GHI = DNI \cos \theta_z + DHI.$ (12)

265 In this study, we also propose to estimate GHI directly with GAM fitted on ground
 266 measured GHI. Estimated GHI is then denoted by GHI^D and the following model is
 267 used:

268 $GHI^D = \alpha + \sum_{j=1}^r f_j(X_j).$ (13)

269

270 **3.3 Validation method**

271 For comparison with the results of Eissa *et al.* [23] and Alobaidi *et al.* [22], the
 272 same validation method is used in the present study. On the 5 ground stations in the
 273 UAE, the data from 3 stations for the full year 2010 are used for fitting the model and
 274 the data from the 2 remaining stations for the full year 2009 are used for testing the
 275 model. With this approach, the model is trained and tested on completely independent
 276 conditions with different locations and a different year.

277 The performances are evaluated in terms of the root mean square error
 278 (RMSE), mean bias error (MBE), relative root mean square error (rRMSE) and
 279 relative mean bias error (rMBE). The rRMSE and rMBE are defined here by:

280 $rRMSE = \left(\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \right) \cdot \frac{100}{\bar{y}},$ (14)

281 $rMBE = \left(\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) \right) \cdot \frac{100}{\bar{y}}$ (15)

282 where y_i is the measured irradiance, \hat{y}_i is the estimated irradiance and \bar{y} is the
283 mean of the measured irradiance.

284

285 **4. Results**

286 **4.1. Models trained and tested on the cloud-free and cloudy sky datasets**

287 This subsection presents the results of the estimation of irradiance variables
288 with GAM. Two separate models for each irradiance variable were fitted on the
289 cloud-free and cloudy training datasets with all the explanatory variables included.
290 Finally, each model was validated with either the cloud-free or the cloudy testing
291 dataset. Table 1 presents the results obtained for the irradiance variables in terms of
292 RMSE, MBE, rRMSE, and rMBE for cloud-free and cloudy conditions, and for both
293 GAM and ANN models.

294 The comparison of the relative statistics obtained with GAM indicates that
295 best estimations are obtained for GHI and GHI^D in both sky conditions. The rRMSEs
296 reach their lowest values for GHI and GHI^D (7.1% and 6.5% for cloud-free conditions
297 and 15.3% and 13.5% for cloudy conditions respectively). In cloud-free conditions,
298 the worst estimations are obtained for DHI with an rRMSE of 23.8%. In the cloudy
299 case, the worst estimations are obtained for DNI and DNI^D with rRMSEs equal to
300 36.7% and 35.9% respectively. When comparing results for cloud-free and cloudy
301 conditions, the worst estimations are systematically obtained for cloudy conditions.
302 The rRMSE and rMBE values are significantly higher for cloudy conditions for most
303 irradiance variables compared to cloud-free conditions.

304 Results for DNI^{D} and GHI^{D} , directly estimated with GAM, are compared with
305 results for DNI and GHI. In cloud-free conditions, GHI^{D} results are slightly better
306 than GHI, while DNI results are slightly better than DNI^{D} . In cloudy conditions,
307 absolute and relative RMSEs are improved slightly with directly estimated DNI and
308 GHI. More important improvements are observed for absolute and relative MBEs: For
309 instance, absolute MBEs obtained for DNI and DNI^{D} in cloudy conditions are -50.2
310 and -25.3 W/m^2 respectively. For GHI and GHI^{D} , they are -36.7 and -13.4 W/m^2
311 respectively.

312 For comparison purposes, results obtained in Eissa *et al.* [23] and Alobaidi *et*
313 *al.* [22] with the ANN approach using the same case study and validation procedure
314 are presented in Table 1. Comparison of GAM and Bagging ANN results of Eissa *et*
315 *al.* [23] shows that significant improvements are generally obtained with GAM for
316 DNI, DNI^{D} , GHI and GHI^{D} with respect to absolute and relative RMSE and MBE for
317 both sky conditions. For instance, in the case of cloud-free conditions, the RMSE for
318 DNI is 140.0 W/m^2 with ANN compared to 115.1 W/m^2 with GAM. For DHI,
319 RMSEs are relatively similar in both sky conditions but MBEs are significantly better
320 for GAM in both sky conditions. Overall, the results indicate a clear advantage of
321 GAM over ensemble ANN model of Eissa *et al.* [23].

322 The results of Alobaidi *et al.* [22] are comparable for the cloud-free conditions
323 and are slightly better for the cloudy conditions. For the cloud-free conditions, the
324 RMSE of the proposed GAM model is slightly higher for DHI, but the results of
325 GAM model have lower MBE. The DNI results are very similar. The GAM model
326 however produces better estimates of the GHI for cloud free conditions which implies
327 that the errors in DHI and DNI cancel each other.

328 **4.2. A single model trained on all sky dataset and tested on cloud-free, cloudy**
329 **and all sky datasets**

330 In Eissa *et al.* [23] and Alobaidi *et al.* [22], two different ANN ensemble
331 models were trained and tested separately for cloud-free and cloudy datasets. The
332 impact of using separate datasets based on sky conditions is evaluated here. For that, a
333 global model was fitted to the all sky conditions dataset and tested separately on the
334 cloud-free, cloudy and all sky testing datasets. Results obtained with the global model
335 are presented in Table 2. In the following, they are compared to the results of Table 1.
336 For the cloud-free case, RMSEs are in most cases slightly higher with the global
337 model and MBEs equivalent for both approaches. For the cloudy case, no general
338 conclusion can be made concerning RMSEs and MBEs. However, MBEs are
339 significantly reduced for DNI and DNI^{D} with the global model. For the all sky
340 conditions case, relative statistics represent a tradeoff between results when tested on
341 the cloud-free testing dataset and when tested on the cloudy testing dataset. This
342 reflects the fact that both sky conditions testing datasets are mixed together. These
343 overall results show that using separate models trained on cloud-free and cloudy
344 conditions do not have a significant positive impact on the performances.

345 Fig. 3 presents the density scatter plots of estimated variables versus ground
346 measured variables. For DHI, a downward trend in residuals is observed and a
347 positive bias is visible in the zone with the highest density. DNI and DNI^{D} present
348 similar scatter plots. A downward trend in residuals is also observed for these
349 variables. Residuals in the scatter plot of for GHI and GHI^{D} are similar. They are
350 evenly distributed around the line representing zero bias and no trend is observed.

351 Mean ground measured DHI, DNI and GHI were computed for separate cloud-
352 free and cloudy conditions. Fig. 4 presents the mean ground measured DHI, DNI and
353 GHI as a function of time for the training dataset. Cloud-free and cloudy conditions
354 were computed separately. For DHI, the received irradiance is superior for cloudy
355 conditions. For DNI, the inverse occurs where the irradiance received is superior for
356 cloud-free conditions. For GHI, both curves confound each other. These curves are
357 explained by the fact that under cloudy sky conditions, the scatter irradiance is
358 increased, resulting in an increased DHI and a reduced DNI. However, the total
359 irradiance received is not affected by sky conditions as GHI is equal for both
360 conditions. These results advocate the use a single model for both sky conditions for
361 GHI.

362 **4.3 Interpretation of smooth functions**

363 In GAM, the sum of the smooth functions of one or more explanatory
364 variables and the intercept give a function of the response variable (See (2)). Each
365 smooth function then represents the effect on the response variable of one predictor in
366 relation with the effect of the other predictors. Smooth functions are graphically
367 presented here and attempts to provide physical explanations are made. The global
368 model fitted on the all sky conditions training dataset is used here for illustration as no
369 important improvement was obtained by using two separate models for both sky
370 conditions as shown in the last subsection.

371 Attempts to obtain simpler models were carried out through stepwise
372 regression methods. However, in most cases, the best model ends up being the model
373 with all variables. Nevertheless, with GAM, it is hypothesized that the inclusion of ε
374 is unnecessary. Indeed, ε is computed at each location with a formula that depends

375 only on day number, which is already included as an explanatory variable in the
376 model. Table 3 presents the results obtained for the estimation of radiation variables
377 with models using all explanatory variables except ε . The results obtained with and
378 without ε are very similar and show that ε is redundant.

379 The smooth functions of each explanatory variable are represented in Figs. 5-7
380 for DHI, DNI^{D} and GHI^{D} respectively using the model without ε and fitted on the all
381 sky conditions training dataset. The dotted line represents the 5% confidence interval.
382 To help interpreting the smooth functions, Figs. 8-10 present the scatter plots of
383 measured DHI, DNI and GHI versus each explanatory variable respectively for the all
384 sky conditions training dataset.

385 The smooth function of DHI versus Day increases with Day until summer then
386 decreases until the end of the year. The scatter plot of DHI with Day in Fig. 8 shows a
387 similar relation. For DNI^{D} and GHI^{D} , an inverse relation in the smooth functions is
388 observed where the irradiance reaches its minimum during summer. The scatter plot
389 of DNI with Day in Fig. 9 reveals a similar relation. This result is counterintuitive
390 because irradiance is expected to increase during summer. A possible explanation
391 could be the significantly higher air humidity during summer and/or more dust
392 scattering the solar radiation during the summer season.

393 The smooth function of DHI versus Time increases with time to reach a
394 maximum at around noon and decreases afterwards. Because time is related to the sun
395 height and therefore to irradiance intensity, it is expected to observe a similar shape of
396 smooth curve for every irradiance variable. However, for DNI^{D} and GHI^{D} , an inverse
397 relation is observed where the minimum irradiance is reached at around noon. This
398 behavior is explained by the fact that the explanatory variable θ_z , included in the

399 model, also explains the sun position. In the case of DHI, the smooth function of θ_z
400 is strictly increasing. In this case, the time explains the sun position and θ_z explains a
401 complementary portion of the total variance. In the case of DNI^D and GHI^D , θ_z rather
402 explains the sun position as the smooth functions are strictly decreasing with θ_z .

403 The interpretation of the smooth functions of the predictors related to thermal
404 channels is difficult because of their number and the fact that they are not
405 independent. In all cases, a change in the slope of the curve occurs in mid-
406 temperatures. Confidence intervals are larger for low temperatures and decrease to
407 become very small with increasing temperatures. This is explained by the fact that
408 there are fewer observations for small temperatures as seen in the scatter plots of Figs.
409 8-10.

410 The analysis of the smooth curves seems to indicate that the seasonal pattern
411 may be caused by the solar scattering by airborne particles. To further study this
412 hypothesis we quantified the aerosol particle content over the UAE, using data from
413 the AERONET map (AErosol RObotic NETwork), a ground-based aerosol
414 monitoring network initiated by NASA [48]. The dataset includes the aerosol optical
415 thickness (AOT) for different wavelengths and the total water vapor in the column.
416 Fig. 11 presents the mean daily aerosol optical thickness (AOD) at the wavelength of
417 500 nm and the mean daily water vapor at the Abu Dhabi station (24.44 °N, 54.62 °E).
418 This figure shows an important seasonality in the dust and the water vapor peaking
419 during summer.

420 A strong seasonality is observed in both water vapor and aerosol optical
421 thickness. It is therefore important to verify whether this seasonal behavior propagates

422 also into the performance statistics. For this purpose, the year of the testing sample
423 was divided in four seasons of three months and the performance statistics were
424 computed for each season. Table 4 presents the performances for each season with the
425 models without ε , fitted and tested on the all sky conditions training and testing
426 datasets (i.e., the same models used in Table 3). The results of Table 4 show that
427 biases are in general higher during the summer (AMJ and JAS) and RMSEs are higher
428 during the winter season of JFM. The high bias values associated to the summer
429 season can be explained by the scattering by aerosol constituents. It is also observed
430 that the biases of DHI are generally of opposite sign than DNI^{D} and DNI. GHI biases
431 are generally very small due to the canceling effect of the DHI and DNI biases.

432 In the second stage, the number of thermal channels was reduced in order to
433 ease the physical interpretation of the smooth functions related to the thermal
434 channels. In this way, only three thermal channels, T04, T05 and T09, in addition to
435 the other variables were included in the models. T05 and T09 were chosen to
436 represent the water vapor and dust constituents of the atmosphere and T04 was
437 selected because it was shown to be an important channel in the models. Smooth
438 functions obtained for each explanatory variable are presented in Figs. 12-14 for
439 variables DHI, DNI^{D} and GHI^{D} . Performances obtained with this configuration are
440 shown in Table 5. Because the number of explanatory variables has been reduced,
441 most performance indicators decreased. However, RMSE values are similar for GHI
442 and GHI^{D} and absolute and relative MBE values for DNI have improved for the
443 model with fewer explanatory variables.

444 Smooth functions of variables Day, Time and θ_z have similar relationships
445 with response variables than those obtained with the model with more variables.

446 There is an exception in the case of DHI for θ_z where the smooth function is now
447 strictly decreasing. For the thermal channels, most observations occur after a certain
448 threshold temperature which is channel dependent. This can be clearly seen in the
449 scatter plots. Consequently, a change in slope occurs generally around this threshold
450 temperature in the smooth functions of thermal channels. As the number of
451 observations is negligible for the temperatures below the threshold, the analysis is
452 restricted on temperatures higher than this threshold. For DHI, T09 is the most
453 important thermal channel. Its smooth function has a strong negative slope. On the
454 other hand, the smooth function for T04 increases continuously. The smooth function
455 of T05 increases continuously with a light slope. The scatter plots of Fig. 8 reveal that
456 DHI has a positive relation with temperature for T04. For DNI^D , T09 is the most
457 important thermal channel. Its smooth function increases constantly with a strong
458 slope. The smooth function of T04 decreases continuously with a strong slope for
459 high temperatures (Fig. 13). The smooth function of T05 has a light decreasing slope.
460 A strong positive relation of DNI with temperature for T09 is also observed in Fig. 9
461 while being less important for T04. For GHI, the smooth functions of T04 and T09
462 are both strictly increasing (Fig. 14). Strong positive relations are also observed in the
463 scatter plots of thermal channels T04 and T09 in Fig. 10. The smooth function of T05
464 has a slope of about zero and is thus not very important.

465 The thermal channels T05 and T09 were chosen to represent respectively
466 water vapor and dust in the atmosphere. We aim to evaluate to which extent these
467 thermal channels capture the seasonality of the airborne constituents. For this, the
468 individual thermal channel components of the linear predictor are displayed as a
469 function of the day of the year. The simplified models DHI and DNI^D fitted on the all
470 sky conditions dataset are considered. Fig. 15 presents the mean daily predicted DHI

471 and DNI as a function of the day. It can be observed that the curves for T04 follow the
472 seasonal evolution of the ground temperature with a peak during summer. For T05, a
473 strong attenuation due to water vapor is observed where no noticeable seasonality can
474 be observed. For T09, the same seasonal pattern than T04 is notice but with a small
475 attenuation during summer due to dust. Fig. 16 presents the daily mean ground-
476 measured thermal channels T04, T05 and T09 as a function of the day for the all sky
477 conditions training dataset.

478 **4.5 Comparison with McClear**

479 Using the web service for McClear, estimates of irradiances were obtained at
480 the two stations included in the testing sample during the same time period.
481 Performance statistics computed for the cloud-free condition testing sample are added
482 in Table 1. Fig. 17 presents the density scatter plots of estimated variables with
483 McClear versus ground measured variables. Scatter plots are rather similar to GAM.
484 Same trends are observed in the residuals. One small difference that can be observed
485 is that more observations of DHI are underestimated with McClear for very high
486 irradiances. There is also more positive bias with McClear for very low DNI.
487 Performances presented in Table 1 show that McClear, compared to GAM, has higher
488 RMSEs for all variables and higher biases for DNI and GHI. In Eissa *et al.* [49], the
489 McClear model was validated for the same stations as in the present study and better
490 performances were obtained. This can be explained by the fact that the two
491 publications used different methods to discriminate the cloud-free samples from the
492 cloudy samples. Indeed, the algorithm of Long and Ackerman [50] was used in Eissa
493 *et al.* [49] instead of the thin cirrus method used in the present work and in Eissa *et al.*
494 [49]. The application of the Long and Ackerman method has resulted in a much lower
495 proportion of retained cloud-free instants where only 65% of the data was considered

496 cloud-free compared to 85% in the case of the present work. The algorithm of Long
497 and Ackerman is more restrictive in its discrimination and might have removed some
498 instants that were in fact cloudy.

499

500 **5. Conclusions**

501 In this study, GAM was used to estimate the irradiance components DHI, DNI
502 and GHI in the UAE. Ground irradiance measurements were available at 5 stations
503 over the UAE. The data from three stations for the full year of 2010 were used to fit
504 the model and the data of the two remaining stations for the full year of 2009 were
505 used for the validation. In this way, the model was trained and tested in completely
506 independent temporal and spatial conditions. For the purpose of estimating irradiance
507 throughout the UAE, six SEVIRI thermal channels were used along with other
508 variables including the solar zenith angle θ_z , Day, Time and the eccentricity
509 correction ε . These variables can be calculated for any location over the UAE.

510 Results were compared with those obtained with an ANN ensemble approach
511 in Eissa *et al.* [23] and Alobaidi *et al.* [22] where the same database and validation
512 procedure were used. Results indicate clearly that GAM leads to an improved
513 estimation when compared with the bagging ensemble, and is similar or better for
514 cloud-free conditions and slightly lower for cloudy conditions compared to the two-
515 stage ensemble architecture proposed in Alobaidi *et al.* [22]. However, the simplicity
516 of the GAM models and their ability to provide explicit expressions unlike the ANN
517 ensemble is a clear advantage.

518 In Eissa *et al.* [23], the training and testing datasets were separated into cloud-
519 free and cloudy sub-datasets and models were fitted and tested separately for these
520 two datasets. The same approach was used in Alobaidi *et al.* [22] as well. The
521 obtained estimations were weaker in the case of cloudy conditions. In the present
522 study, a single model was also fitted using the training data for all sky conditions and
523 was tested on the cloud-free and cloudy testing datasets. Results have shown that
524 similar performances were obtained for both sky conditions with the global model.
525 This suggests that using two different models is not necessary.

526 As mentioned before, the advantage of the GAM approach over the ANN
527 approach is that relations between irradiance variables and explanatory variables can
528 be defined explicitly. The smoothing curves for each explanatory variable were
529 graphically represented and analyzed to provide physical explanations to the modeled
530 relations.

531 It is proposed in future work to add more variables such as relative humidity
532 as explanatory covariates. Relative humidity has a high variability throughout the
533 year, with large values during the summer. Its inclusion as covariate may help explain
534 an additional percentage of the variance, especially in the summer season. The
535 development of specific summer and winter models based on a rational definition of
536 the seasons (see for instance [51]) should also lead to improved models. The usage of
537 coarse resolution aerosol maps normally used in the physics based approaches can
538 also be integrated into the proposed framework. Future efforts can also focus on
539 testing more advanced basis functions in the GAM model.

540

541 **Acknowledgment**

542 The authors thank the Editor, Dr. Qian Du, the associate editor and the two anonymous
543 reviewers for their judicious comments. The authors also thank the staff responsible for
544 maintaining the AERONET stations in the UAE.

545

546 **Nomenclature**

547	DNI	direct normal irradiance (W/m^2)
548	DHI	diffuse horizontal irradiance (W/m^2)
549	GHI	global horizontal irradiance (W/m^2)
550	θ_z	solar zenith angle (degrees)
551	ε	eccentricity correction
552	δ	total optical depth of the atmosphere
553	I_0	solar constant ($1367 \text{ W}/\text{m}^2$)
554	m	air mass
555	T04	SEVIRI T04 channel ($3.9 \mu\text{m}$) observed brightness temperature (K)
556	T05	SEVIRI T05 channel ($6.2 \mu\text{m}$) observed brightness temperature (K)
557	T06	SEVIRI T06 channel ($7.3 \mu\text{m}$) observed brightness temperature (K)
558	T07	SEVIRI T07 channel ($8.7 \mu\text{m}$) observed brightness temperature (K)
559	T09	SEVIRI T08 channel ($10.8 \mu\text{m}$) observed brightness temperature (K)
560	T10	SEVIRI T10 channel ($12.0 \mu\text{m}$) observed brightness temperature (K)
561	ANN	artificial neural network
562	GAM	generalized additive model
563	GLM	generalized linear model
564	RMSE	root mean square error
565	MBE	mean bias error
566	rRMSE	relative RMSE (%)
567	rMBE	relative MBE (%)
568	X	matrix of explanatory or independent variables

569	\mathbf{Z}	model matrix for the basis functions
570	\mathbf{A}	influence matrix
571	Y	response or dependent random variable
572	X	explanatory or independent random variable
573	\mathbf{y}	vector of observed values of Y
574	g	the link function in GAM and GLM
575	β	unknown parameters of the linear model
576	θ	vector of unknown parameters of the basis functions
577	f	smooth functions
578	b	spline basis functions
579	λ	smoothing parameter

580 **References**

581

- 582 [1] C. A. Gueymard, "Direct and indirect uncertainties in the prediction of tilted
583 irradiance for solar engineering applications," *Solar Energy*, vol. 83, pp. 432-
584 444, 2009.
- 585 [2] B. Y. H. Liu and R. C. Jordan, "The long-term average performance of flat-
586 plate solar-energy collectors: With design data for the U.S., its outlying
587 possessions and Canada," *Solar Energy*, vol. 7, pp. 53-74, 1963.
- 588 [3] C. Schillings, H. Mannstein, and R. Meyer, "Operational method for deriving
589 high resolution direct normal irradiance from satellite data," *Solar Energy*, vol.
590 76, pp. 475-484, 2004.
- 591 [4] A. Zelenka, R. Perez, R. Seals, and D. Renné, "Effective Accuracy of
592 Satellite-Derived Hourly Irradiances," *Theoretical and Applied Climatology*,
593 vol. 62, pp. 199-207, 1999/04/01 1999.
- 594 [5] C. Gautier, G. Diak, and S. Masse, "A Simple Physical Model to Estimate
595 Incident Solar Radiation at the Surface from GOES Satellite Data," *Journal of*
596 *Applied Meteorology*, vol. 19, pp. 1005-1012, 1980.
- 597 [6] E. Cogliani, P. Ricchiazzi, and A. Maccari, "Physical model SOLARMET for
598 determining total and direct solar radiation by meteosat satellite images,"
599 *Solar Energy*, vol. 81, pp. 791-798, 2007.
- 600 [7] D. Cano, J. M. Monget, M. Albuisson, H. Guillard, N. Regas, and L. Wald, "A
601 method for the determination of the global solar radiation from meteorological
602 satellite data," *Solar Energy*, vol. 37, pp. 31-39, 1986.
- 603 [8] R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, and F.
604 Vignola, "A new operational model for satellite-derived irradiances:
605 description and validation," *Solar Energy*, vol. 73, pp. 307-317, 2002.
- 606 [9] H. G. Beyer, C. Costanzo, and D. Heinemann, "Modifications of the Heliosat
607 procedure for irradiance estimates from satellite images," *Solar Energy*, vol.
608 56, pp. 207-212, 1996.
- 609 [10] A. Hammer, D. Heinemann, C. Hoyer, R. Kuhlemann, E. Lorenz, R. Müller,
610 and H. G. Beyer, "Solar energy assessment using remote sensing
611 technologies," *Remote Sensing of Environment*, vol. 86, pp. 423-432, 2003.
- 612 [11] R. W. Mueller, K. F. Dagestad, P. Ineichen, M. Schroedter-Homscheidt, S.
613 Cros, D. Dumortier, R. Kuhlemann, J. A. Olseth, G. Piernavieja, C. Reise, L.
614 Wald, and D. Heinemann, "Rethinking satellite-based solar irradiance
615 modelling: The SOLIS clear-sky module," *Remote Sensing of Environment*,
616 vol. 91, pp. 160-174, 2004.
- 617 [12] J. Polo, L. Martín, and M. Cony, "Revision of ground albedo estimation in
618 Heliosat scheme for deriving solar radiation from SEVIRI HRV channel of
619 Meteosat satellite," *Solar Energy*, vol. 86, pp. 275-282, 2012.
- 620 [13] C. Rigollier, M. Lefèvre, and L. Wald, "The method Heliosat-2 for deriving
621 shortwave solar radiation from satellite images," *Solar Energy*, vol. 77, pp.
622 159-169, 2004.
- 623 [14] M. Schroedter-Homscheidt, J. Betcke, G. Gesell, D. Heinemann, and T.
624 Holzer-Popp, "Energy-Specific Solar Radiation Data from MSG: Current
625 Status of the HELIOSAT-3 Project," in *Second MSG RAO Workshop*, 2004, p.
626 131.

- 627 [15] Z. Qu, A. Oumbe, P. Blanc, M. Lefevre, L. Wald, M. S. Homscheidt, G.
628 Gesell, and L. Klueser, "Assessment of Heliosat-4 surface solar irradiance
629 derived on the basis of SEVIRI-APOLLO cloud products," in *2012*
630 *EUMETSAT Meteorological Satellite Conference*, 2012, pp. s2-06.
- 631 [16] Z. Qu, A. Oumbe, P. Blanc, M. Lefevre, L. Wald, M. Schroedter-Homscheidt,
632 and G. Gesell, "A new method for assessing surface solar irradiance: Heliosat-
633 4," in *EGU General Assembly Conference Abstracts*, 2012, p. 10228.
- 634 [17] M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald,
635 M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola, A. Benedetti, J. W.
636 Kaiser, and J. J. Morcrette, "McClear: a new model estimating downwelling
637 solar radiation at ground level in clear-sky conditions," *Atmos. Meas. Tech.*,
638 vol. 6, pp. 2403-2418, 2013.
- 639 [18] B. Mayer and A. Kylling, "Technical note: The libRadtran software package
640 for radiative transfer calculations - description and examples of use," *Atmos.*
641 *Chem. Phys.*, vol. 5, pp. 1855-1877, 2005.
- 642 [19] B. Khalil, T. B. M. J. Ouarda, and A. St-Hilaire, "Estimation of water quality
643 characteristics at ungauged sites using artificial neural networks and canonical
644 correlation analysis," *Journal of Hydrology*, vol. 405, pp. 277-287, 2011.
- 645 [20] C. Shu and T. B. J. M. Ouarda, "Flood frequency analysis at ungauged sites
646 using artificial neural networks in canonical correlation analysis physiographic
647 space," *Water Resources Research*, vol. 43, p. W07438, Jul 2007.
- 648 [21] I. Zaier, C. Shu, T. B. M. J. Ouarda, O. Seidou, and F. Chebana, "Estimation
649 of ice thickness on lakes using artificial neural network ensembles," *Journal of*
650 *Hydrology*, vol. 383, pp. 330-340, Mar 30 2010.
- 651 [22] M. H. Alobaidi, P. R. Marpu, T. B. M. J. Ouarda, and H. Ghedira, "Mapping
652 of the Solar Irradiance in the UAE Using Advanced Artificial Neural Network
653 Ensemble," *Selected Topics in Applied Earth Observations and Remote*
654 *Sensing, IEEE Journal of*, vol. 7, pp. 3668-3680, 2014.
- 655 [23] Y. Eissa, P. R. Marpu, I. Gherboudj, H. Ghedira, T. B. M. J. Ouarda, and M.
656 Chiesa, "Artificial neural network based model for retrieval of the direct
657 normal, diffuse horizontal and global horizontal irradiances using SEVIRI
658 images," *Solar Energy*, vol. 89, pp. 1-16, 2013.
- 659 [24] A. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using artificial
660 neural network: Application for performance prediction of a grid-connected
661 PV plant at Trieste, Italy," *Solar Energy*, vol. 84, pp. 807-821, 2010.
- 662 [25] K. Moustris, A. G. Paliatsos, A. Bloutsos, K. Nikolaidis, I. Koronaki, and K.
663 Kavadias, "Use of neural networks for the creation of hourly global and
664 diffuse solar irradiance data at representative locations in Greece," *Renewable*
665 *Energy*, vol. 33, pp. 928-932, 2008.
- 666 [26] J. Mubiru and E. J. K. B. Banda, "Estimation of monthly average daily global
667 solar irradiation using artificial neural networks," *Solar Energy*, vol. 82, pp.
668 181-187, 2008.
- 669 [27] O. Şenkal, "Modeling of solar radiation using remote sensing and artificial
670 neural network in Turkey," *Energy*, vol. 35, pp. 4795-4801, 2010.
- 671 [28] F. S. Tymvios, C. P. Jacovides, S. C. Michaelides, and C. Scouteli,
672 "Comparative study of Ångström's and artificial neural networks'
673 methodologies in estimating global solar radiation," *Solar Energy*, vol. 78, pp.
674 752-762, 2005.
- 675 [29] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm,"
676 in *International Conference on Machine Learning*, 1996, pp. 148-156.

- 677 [30] T. B. M. J. Ouarda and C. Shu, "Regional low-flow frequency analysis using
678 single and ensemble artificial neural networks," *Water Resources Research*,
679 vol. 45, p. W11428, 2009.
- 680 [31] D. L. Borchers, S. T. Buckland, I. G. Priede, and S. Ahmadi, "Improving the
681 precision of the daily egg production method using generalized additive
682 models," *Canadian Journal of Fisheries and Aquatic Sciences*, vol. 54, pp.
683 2727-2742, 1997.
- 684 [32] L. Wen, K. Rogers, N. Saintilan, and J. Ling, "The influences of climate and
685 hydrology on population dynamics of waterbirds in the lower Murrumbidgee
686 River floodplains in Southeast Australia: Implications for environmental water
687 management," *Ecological Modelling*, vol. 222, pp. 154-163, 2011.
- 688 [33] S. N. Wood and N. H. Augustin, "GAMs with integrated model selection
689 using penalized regression splines and applications to environmental
690 modelling," *Ecological Modelling*, vol. 157, pp. 157-177, 2002.
- 691 [34] F. Chebana, C. Charron, T. B. M. J. Ouarda, and B. Martel, "Regional
692 Frequency Analysis at Ungauged Sites with the Generalized Additive Model,"
693 *Journal of Hydrometeorology*, vol. 15, pp. 2418-2428, 2014/12/01 2014.
- 694 [35] M. Durocher, F. Chebana, and T. B. M. J. Ouarda, "A Nonlinear Approach to
695 Regional Flood Frequency Analysis Using Projection Pursuit Regression,"
696 *Journal of Hydrometeorology*, vol. 16, pp. 1561-1574, 2015/08/01 2015.
- 697 [36] L. Bayentin, S. El Adlouni, T. Ouarda, P. Gosselin, B. Doyon, and F.
698 Chebana, "Spatial variability of climate effects on ischemic heart disease
699 hospitalization rates for the period 1989-2006 in Quebec, Canada,"
700 *International Journal of Health Geographics*, vol. 9, p. 5, 2010.
- 701 [37] C. Cans and C. Lavergne, "De la régression logistique vers un modèle additif
702 généralisé : un exemple d'application," *Revue de Statistique Appliquée*, vol.
703 43, pp. 77-90, 1995.
- 704 [38] S. Clifford, S. Low Choy, T. Hussein, K. Mengersen, and L. Morawska,
705 "Using the Generalised Additive Model to model the particle number count of
706 ultrafine particles," *Atmospheric Environment*, vol. 45, pp. 5934-5945, 2011.
- 707 [39] A. M. Leitte, C. Petrescu, U. Franck, M. Richter, O. Suci, R. Ionovici, O.
708 Herbarth, and U. Schlink, "Respiratory health, effects of ambient air pollution
709 and its modification by air humidity in Drobeta-Turnu Severin, Romania,"
710 *Science of The Total Environment*, vol. 407, pp. 4004-4011, 2009.
- 711 [40] J. Rocklöv and B. Forsberg, "The effect of temperature on mortality in
712 Stockholm 1998-2003: A study of lag structures and heatwave effects,"
713 *Scandinavian Journal of Public Health*, vol. 36, pp. 516-523, 2008.
- 714 [41] V. Vieira, T. Webster, J. Weinberg, and A. Aschengrau, "Spatial analysis of
715 bladder, kidney, and pancreatic cancer on upper Cape Cod: an application of
716 generalized additive models to case-control data," *Environmental Health*, vol.
717 8, p. 3, 2009.
- 718 [42] Y. Eissa, M. Chiesa, and H. Ghedira, "Assessment and recalibration of the
719 Heliosat-2 method in global horizontal irradiance modeling over the desert
720 environment of the UAE," *Solar Energy*, vol. 86, pp. 1816-1825, 2012.
- 721 [43] J. Hocking, P. N. Francis, and R. Saunders, "Cloud detection in Meteosat
722 Second Generation imagery at the Met Office," *Meteorological Applications*,
723 vol. 18, pp. 307-323, 2011.
- 724 [44] J. A. Nelder and R. W. M. Wedderburn, "Generalized Linear Models,"
725 *Journal of the Royal Statistical Society. Series A (General)*, vol. 135, pp. 370-
726 384, 1972.

- 727 [45] T. Hastie and R. Tibshirani, "Generalized Additive Models," *Statistical*
728 *Science*, vol. 1, pp. 297-310, 1986.
- 729 [46] S. N. Wood, *Generalized Additive Models: An Introduction with R*: Chapman
730 and Hall/CRC Press, 2006.
- 731 [47] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*. New York
732 (USA): Chapman & Hall, 1990.
- 733 [48] B. N. Holben, T. F. Eck, I. Slutsker, D. Tanré, J. P. Buis, A. Setzer, E.
734 Vermote, J. A. Reagan, Y. J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak,
735 and A. Smirnov, "AERONET—A Federated Instrument Network and Data
736 Archive for Aerosol Characterization," *Remote Sensing of Environment*, vol.
737 66, pp. 1-16, 1998.
- 738 [49] Y. Eissa, S. Munawwar, A. Oumbe, P. Blanc, H. Ghedira, L. Wald, H. Bru,
739 and D. Goffe, "Validating surface downwelling solar irradiances estimated by
740 the McClear model under cloud-free skies in the United Arab Emirates," *Solar*
741 *Energy*, vol. 114, pp. 17-31, 2015.
- 742 [50] C. N. Long and T. P. Ackerman, "Identification of clear skies from broadband
743 pyranometer measurements and calculation of downwelling shortwave cloud
744 effects," *Journal of Geophysical Research: Atmospheres*, vol. 105, pp. 15609-
745 15626, 2000.
- 746 [51] J. M. Cunderlik, T. B. M. J. Ouarda, and B. Bobée, "On the objective
747 identification of flood seasons," *Water Resources Research*, vol. 40, p.
748 W01520, 2004.

749

750

751

752 Table 1. Results obtained for the models fitted on the separate cloud-free and cloudy sky conditions training datasets and tested on the separate
 753 cloud-free and cloudy sky conditions testing datasets.

Sky conditions	Statistic	GAM					ANN (Eissa <i>et al.</i> , 2013)			ANN (Alobaidi <i>et al.</i> , 2014)			McClear		
		DHI	DNI	DNI ^D	GHI	GHI ^D	DHI	DNI	GHI	DHI	DNI	GHI	DHI	DNI	GHI
Cloud-free	RMSE	55.7	115.1	117.3	47.5	43.4	58.0	140.0	76.9				67.3	149.6	62.9
	MBE	3.8	1.1	2.8	-2.0	1.3	12.2	-33.7	-14.3				0.2	38.9	21.5
	rRMSE (%)	23.8	19.4	19.7	7.1	6.5	24.7	23.6	11.4	21.8	19.5	8.4	28.7	25.5	9.4
	rMBE (%)	1.6	0.2	0.5	-0.3	0.2	5.2	-5.7	-2.1	-3.2	-0.2	-1.5	0.1	6.6	3.2
Cloudy	RMSE	75.5	173.9	170.3	90.1	79.2	76.9	201.0	105.0						
	MBE	5.8	-50.2	-25.3	-36.7	-13.4	-12.2	-40.5	-49.4						
	rRMSE (%)	28.8	36.7	35.9	15.3	13.5	29.3	42.4	17.8	26.8	34.7	13.5			
	rMBE (%)	2.2	-10.6	-5.3	-6.3	-2.3	-4.7	-8.6	-8.4	2.7	1.3	2.1			

754

755 Table 2. Results obtained for the models fitted on the all sky conditions training
 756 dataset and tested on the cloud-free, cloudy and all sky conditions testing datasets.

Sky conditions	Statistic	DHI	DNI	DNI ^D	GHI	GHI ^D
Cloud-free	RMSE	57.1	119.0	122.2	46.6	44.4
	MBE	5.6	-1.7	-0.2	-0.4	1.8
	rRMSE (%)	24.4	20.0	20.6	6.9	6.6
	rMBE (%)	2.4	-0.3	-0.0	-0.1	0.3
Cloudy	RMSE	73.9	170.7	175.1	90.5	79.8
	MBE	-13.8	-10.3	16.6	-34.8	-16.3
	rRMSE (%)	28.2	36.0	37.0	15.4	13.6
	rMBE (%)	-5.3	-2.2	3.5	-5.9	-2.8
All sky conditions	RMSE	59.9	127.8	131.2	55.1	51.1
	MBE	2.8	-3.0	2.3	-5.4	-0.8
	rRMSE (%)	25.1	22.2	22.8	8.4	7.7
	rMBE (%)	1.2	-0.5	0.4	-0.8	-0.1

757

758

759 Table 3. Results obtained with models without ϵ . The models are fitted and tested on
760 the all sky conditions training and testing datasets.

Statistic	DHI	DNI	DNI ^D	GHI	GHI ^D
RMSE	59.2	125.8	129.2	54.1	50.7
MBE	2.9	-2.9	2.2	-5.6	-0.8
rRMSE (%)	24.8	21.8	22.4	8.2	7.7
rMBE (%)	1.2	-0.5	0.4	-0.8	-0.1

761

762

763 Table 4. Seasonality in the performance statistics. Results are obtained with models
 764 without ε . The models are fitted and tested on the all sky conditions training and
 765 testing datasets.

Season	Statistic	DHI	DNI	DNI ^D	GHI	GHI ^D
JFM	RMSE	67.51	147.83	153.24	64.79	64.78
	MBE	6.20	-14.73	-1.42	-6.49	4.24
	rRMSE (%)	30.13	25.04	25.96	10.38	10.38
	rMBE (%)	2.77	-2.49	-0.24	-1.04	0.68
AMJ	RMSE	59.07	113.30	115.33	47.73	40.81
	MBE	20.78	-18.03	-19.35	4.46	5.13
	rRMSE (%)	22.07	20.27	20.63	6.62	5.66
	rMBE (%)	7.76	-3.23	-3.46	0.62	0.71
JAS	RMSE	60.03	119.68	118.22	45.04	43.81
	MBE	-17.77	35.56	34.67	1.31	-0.13
	rRMSE (%)	21.99	22.96	22.68	6.57	6.39
	rMBE (%)	-6.51	6.82	6.65	0.19	-0.02
OND	RMSE	49.29	121.07	127.99	56.69	51.00
	MBE	-2.64	-7.47	2.39	-21.61	-12.52
	rRMSE (%)	25.91	19.24	20.34	9.40	8.45
	rMBE (%)	-1.39	-1.19	0.38	-3.58	-2.07

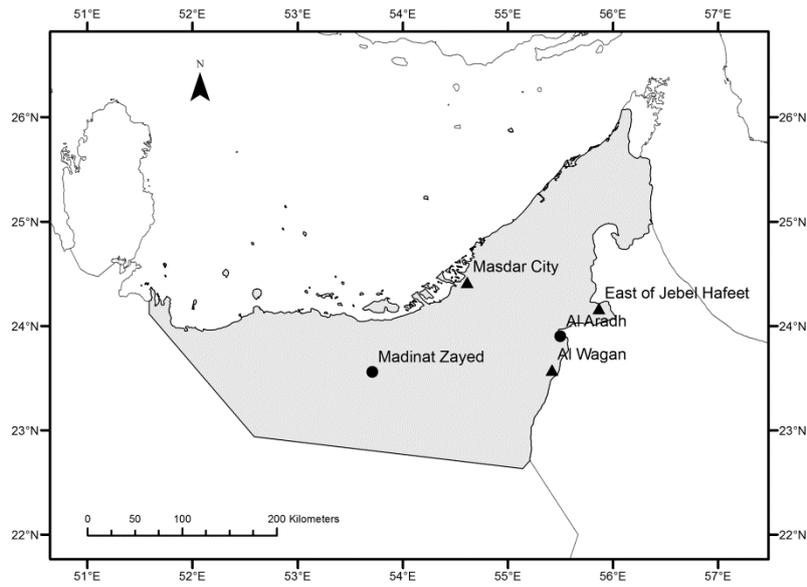
766

767 Table 5. Results obtained with models including the explanatory variables Day, Time,
 768 θ_z , T04, T05 and T09. The models are fitted and tested on the all sky conditions
 769 training and testing datasets.

Statistic	DHI	DNI	DNI ^D	GHI	GHI ^D
RMSE	67.7	132.5	136.1	54.2	51.1
MBE	-17.1	7.2	14.4	-17.2	-0.8
rRMSE (%)	28.4	23.0	23.6	8.2	7.7
rMBE (%)	-7.2	1.3	2.5	-2.6	-0.1

770

771



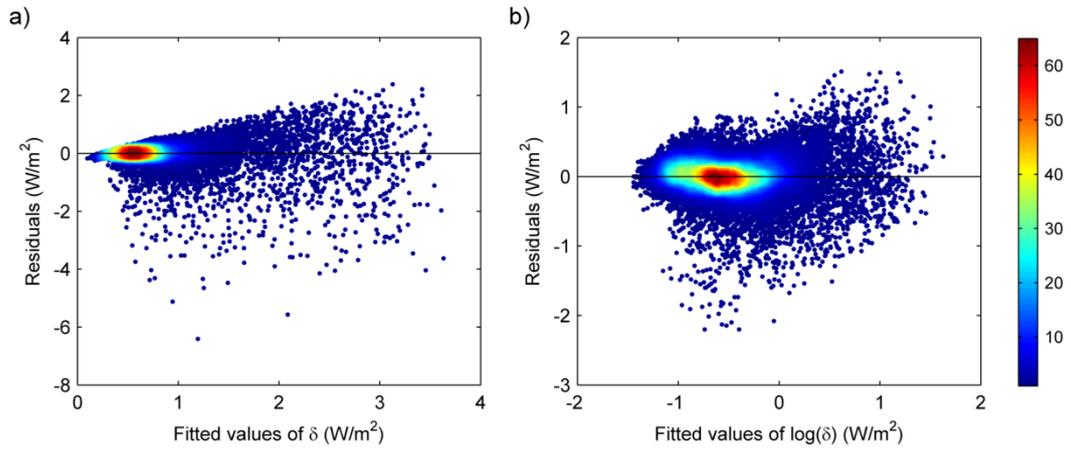
772

773

774

Fig. 1. Location of the ground measurement stations. Triangles represent stations of the training dataset and circles represent stations of the testing dataset.

775

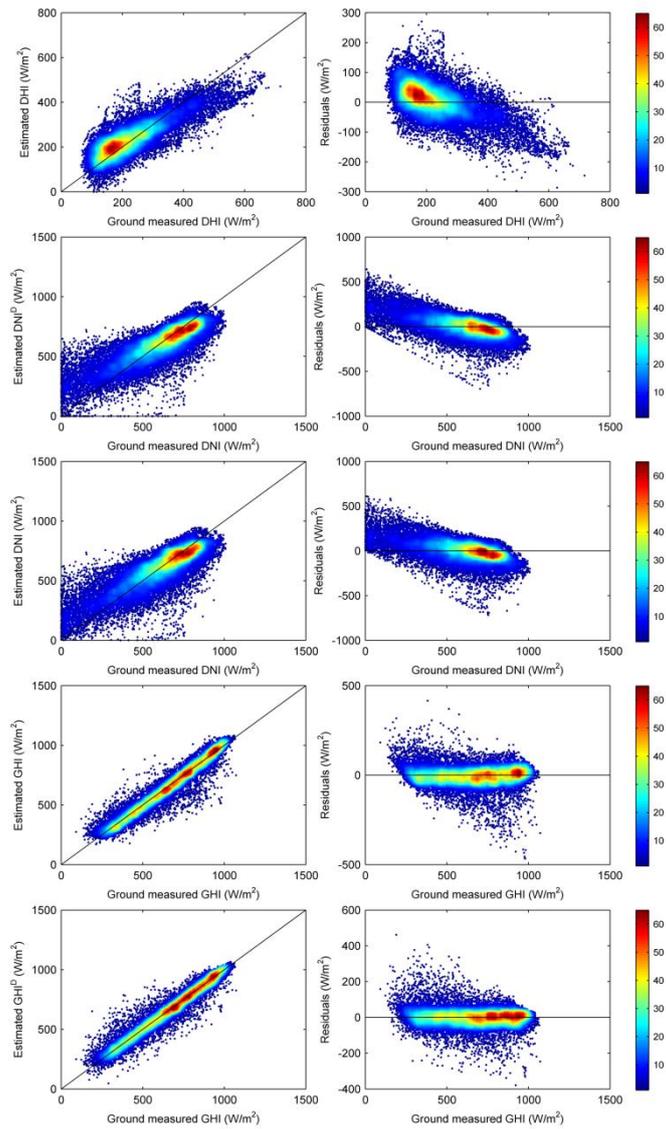


776

777 Fig. 2. Density scatter plots of residuals versus model fitted values for a) δ and b) $\log(\delta)$.

778

779



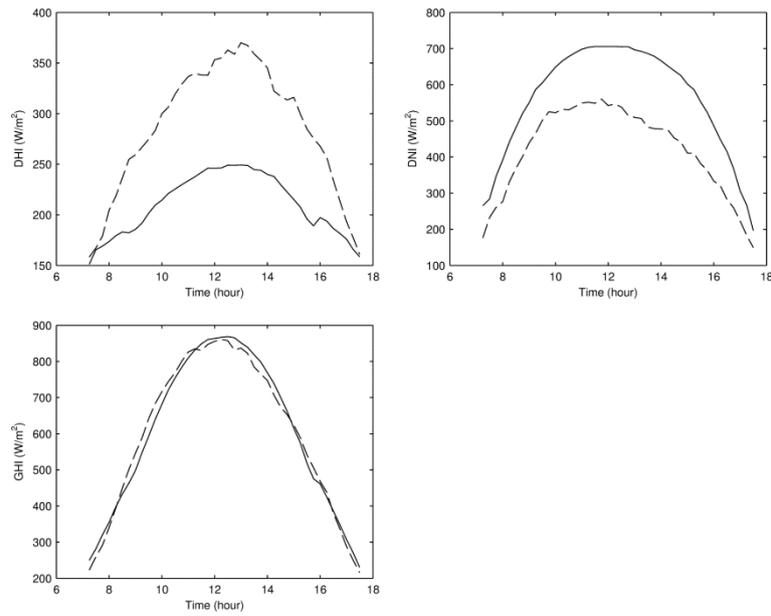
780

781

782

783

Fig. 3. Density scatter plots of estimated versus ground measured irradiance and residuals versus ground measured irradiance for the models fitted and tested on the all sky conditions training and testing datasets.



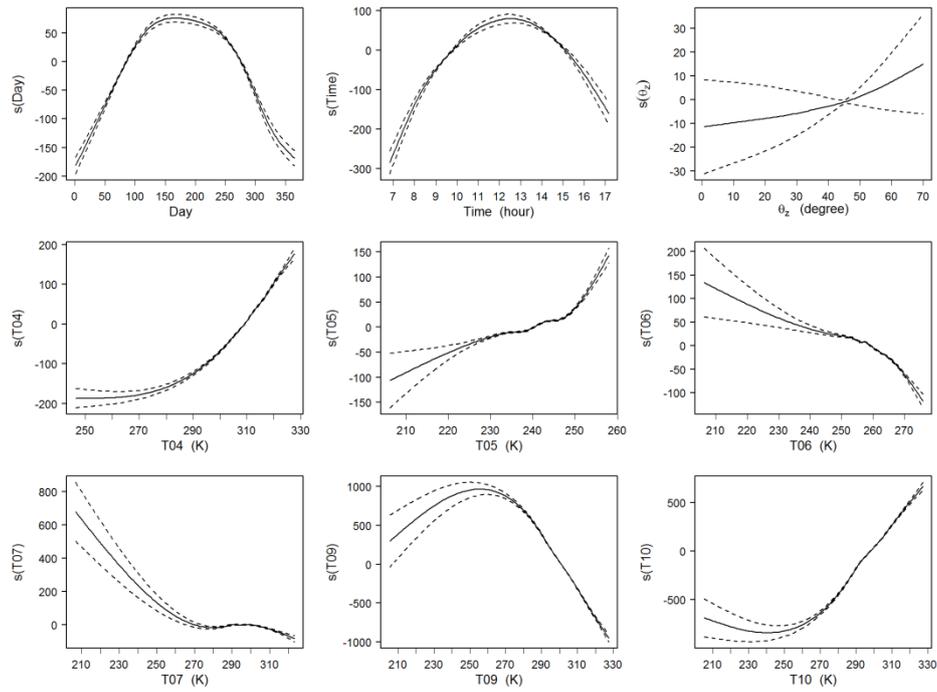
784

785 Fig. 4. Mean ground measured DHI, DNI and GHI as function of time for the training dataset.

786 Solid lines represent cloud-free conditions and dashed lines represent cloudy conditions.

787

788

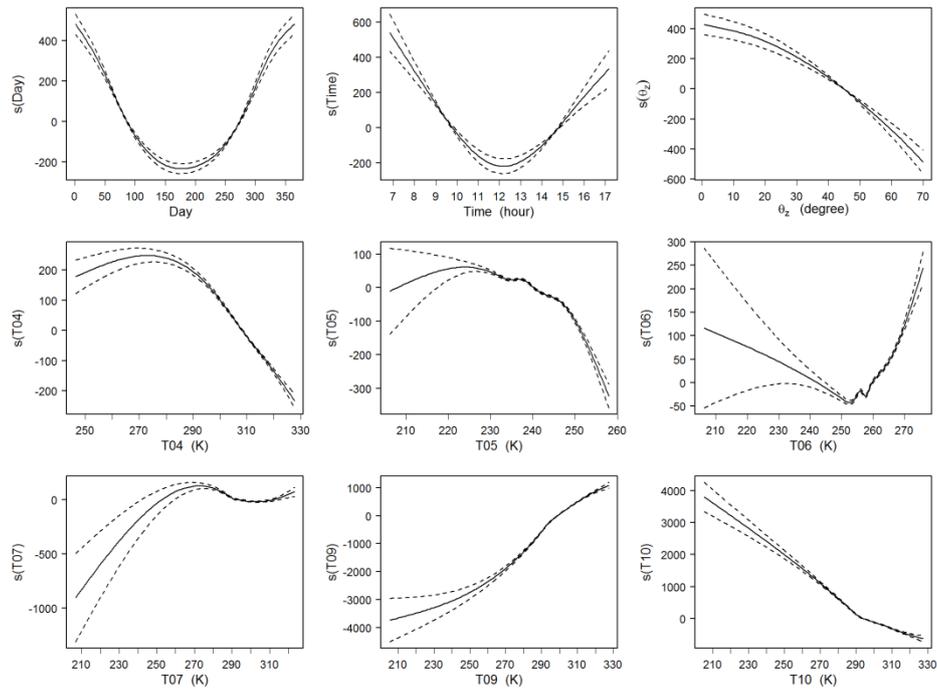


789

790

791

Fig. 5. Smooth functions of explanatory variables for the model estimating DHI fitted on the all sky conditions dataset. The dotted lines represent the limits of the 5% confidence interval.

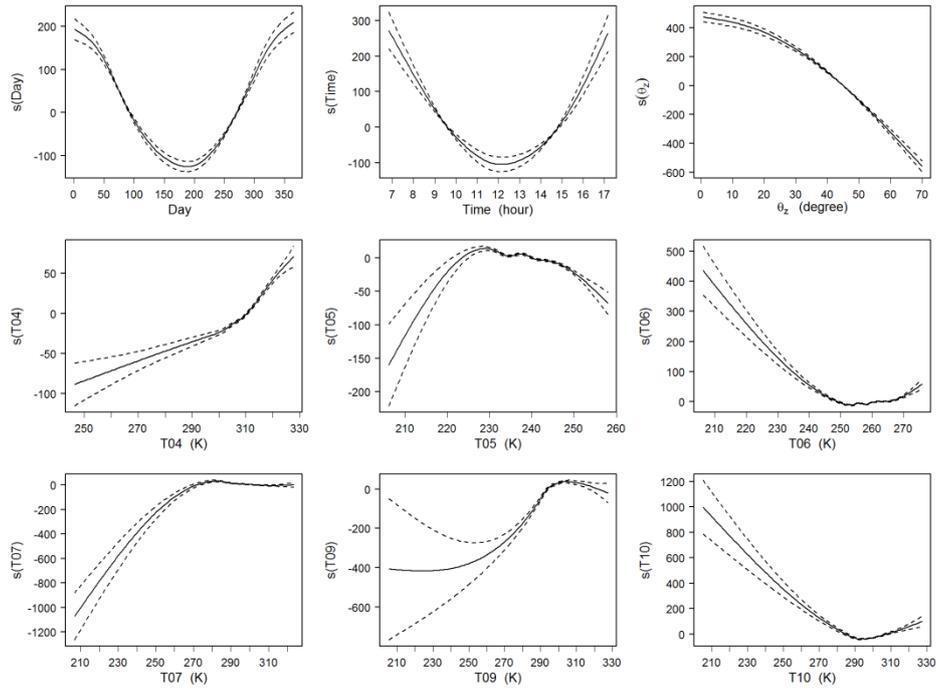


792

793

794

Fig. 6. Smooth functions of explanatory variables for the model estimating DNI^D fitted on the all sky conditions dataset. The dotted lines represent the limits of the 5% confidence interval.

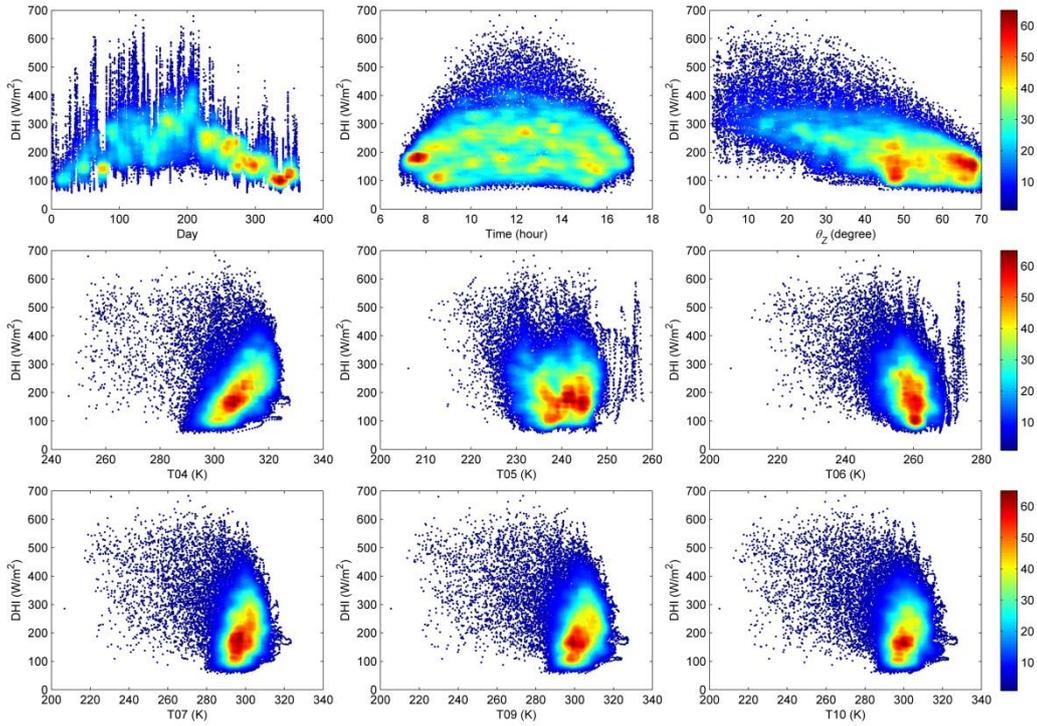


795

796

797

Fig. 7. Smooth functions of explanatory variables for the model estimating GHI^D fitted on the all sky conditions dataset. The dotted lines represent the limits of the 5% confidence interval.

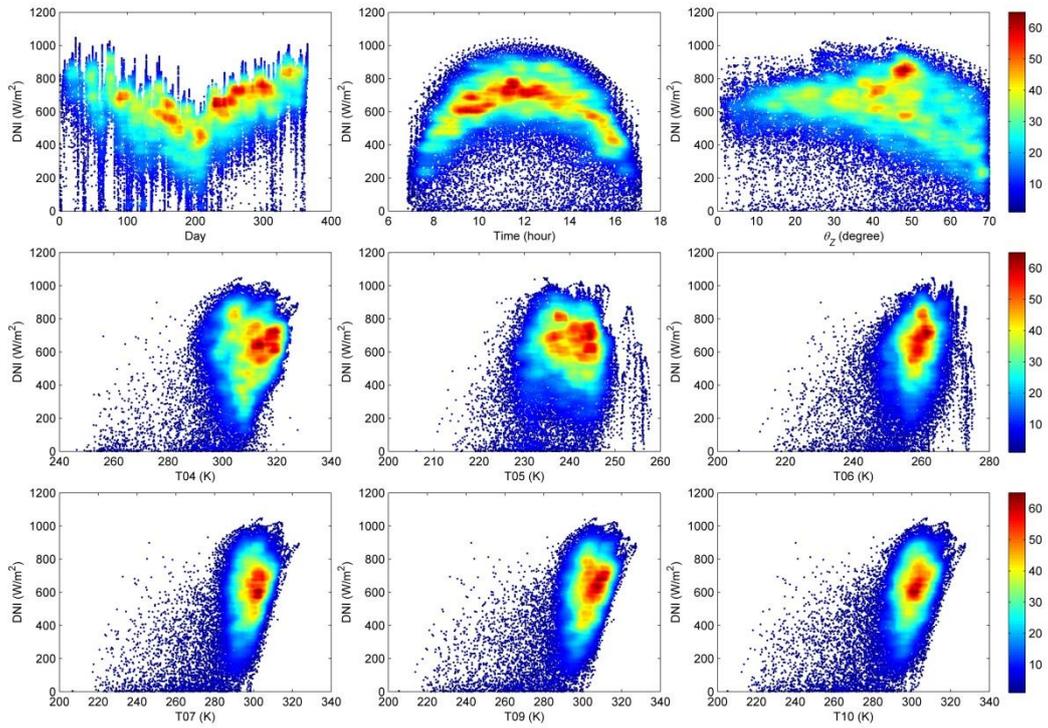


798

799

800

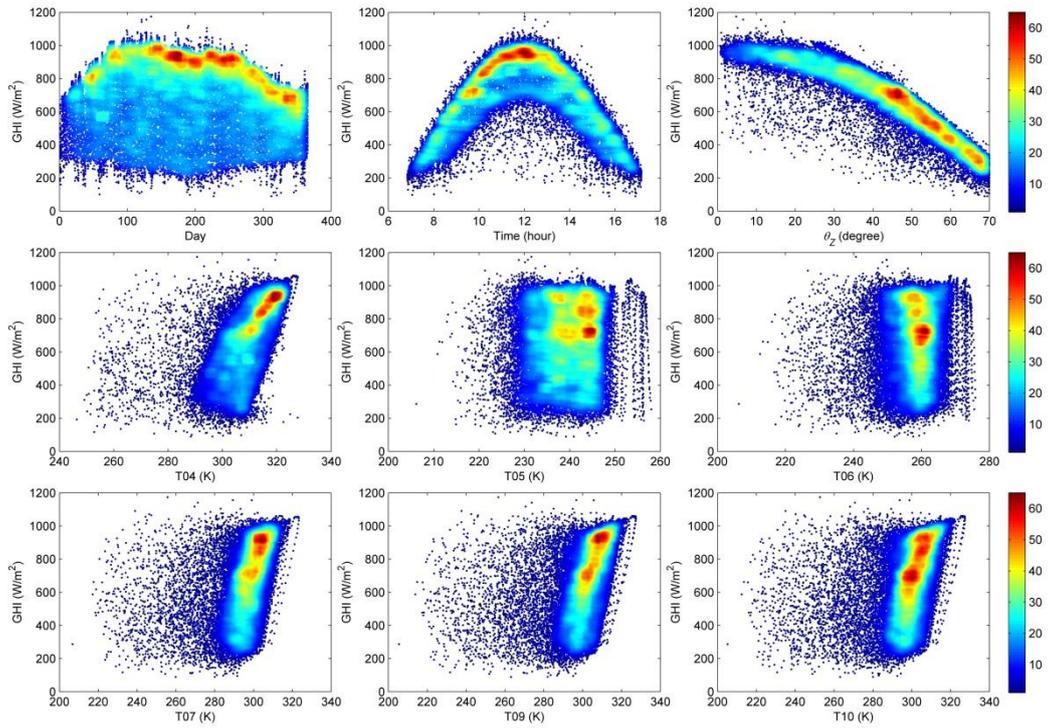
Fig. 8. Scatter plots of ground measured DHI versus explanatory variables for the all sky training conditions dataset.



801

802
803

Fig. 9. Scatter plots of ground measured DNI versus explanatory variables for the all sky conditions training dataset.



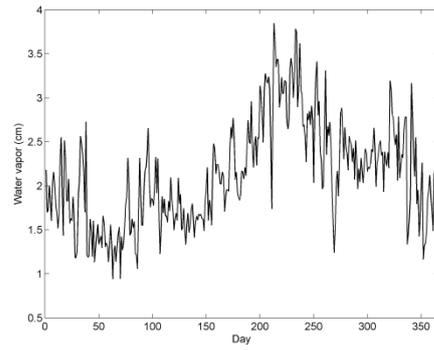
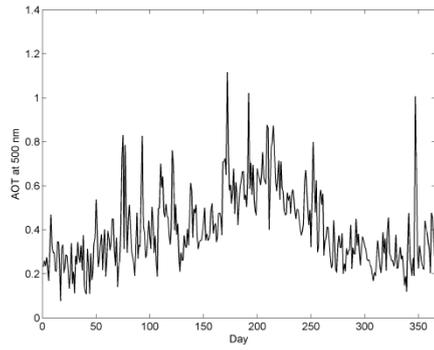
804

805

806

Fig. 10. Scatter plots of ground measured GHI versus explanatory variables for the all sky conditions training dataset.

807

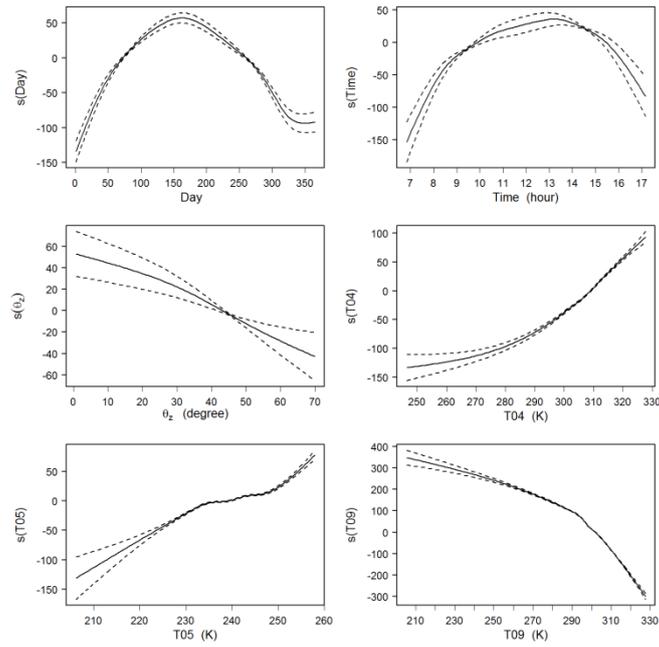


808

Fig 11. Mean daily AOT at 500 nm and mean daily water vapor in Abu Dhabi.

809

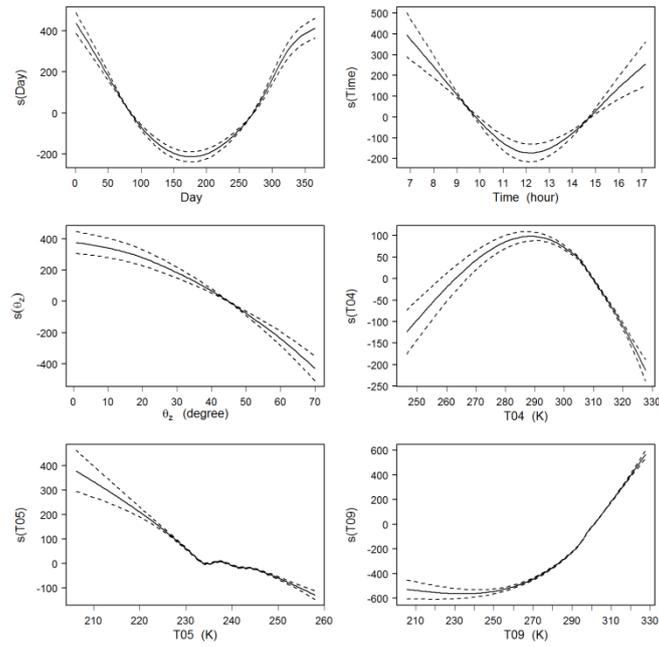
810



811

812 Fig. 12. Smooth functions of explanatory variables for the model estimating DHI
 813 fitted on the all sky conditions dataset (simplified model). The dotted lines represent
 814 the limits of the 5% confidence interval.

815



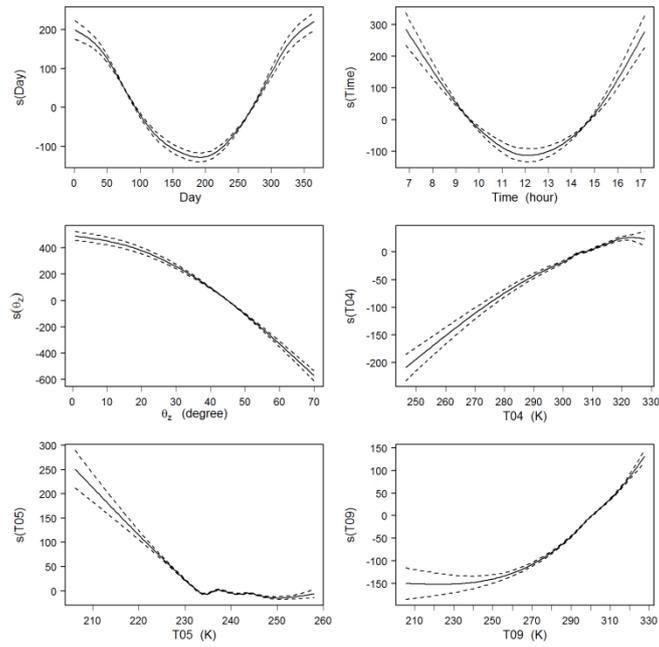
816

817

818

819

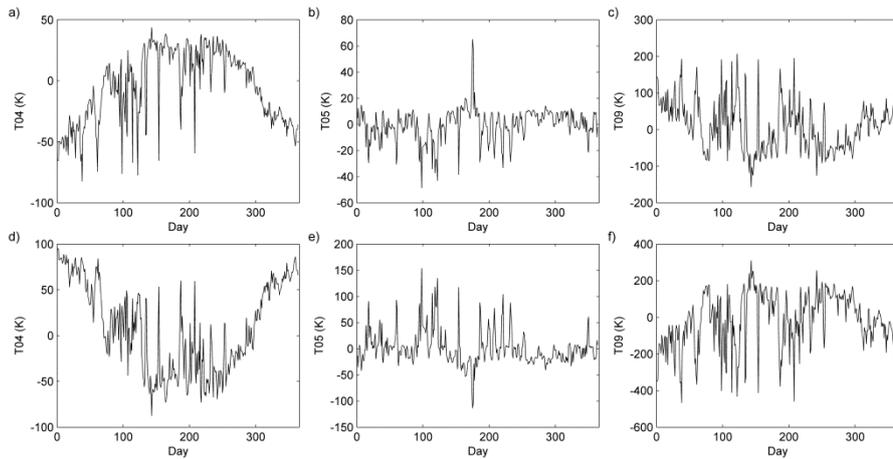
Fig. 13. Smooth functions of explanatory variables for the model estimating DNI^{D} fitted on the all sky conditions dataset (simplified model). The dotted lines represent the limits of the 5% confidence interval.



820

821 Fig. 14. Smooth functions of explanatory variables for the model estimating GHI^D
 822 fitted on the all sky conditions dataset (simplified model). The dotted lines represent
 823 the limits of the 5% confidence interval.

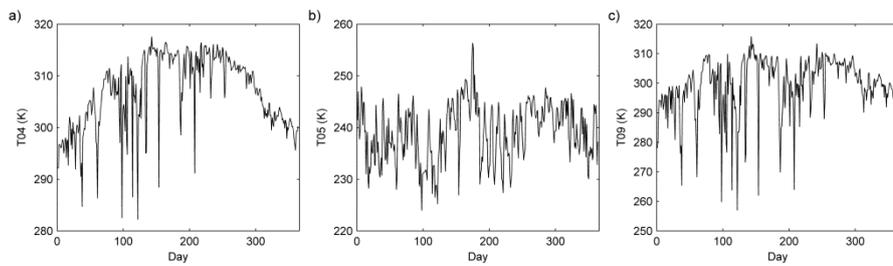
824



825

826 Fig. 15. Components of the linear predictor related to the thermal channels as a
 827 function of the day. The models DHI (a,b,c) and DNI^D (d,e,f) fitted on the all sky
 828 conditions dataset (simplified model) are considered.

829

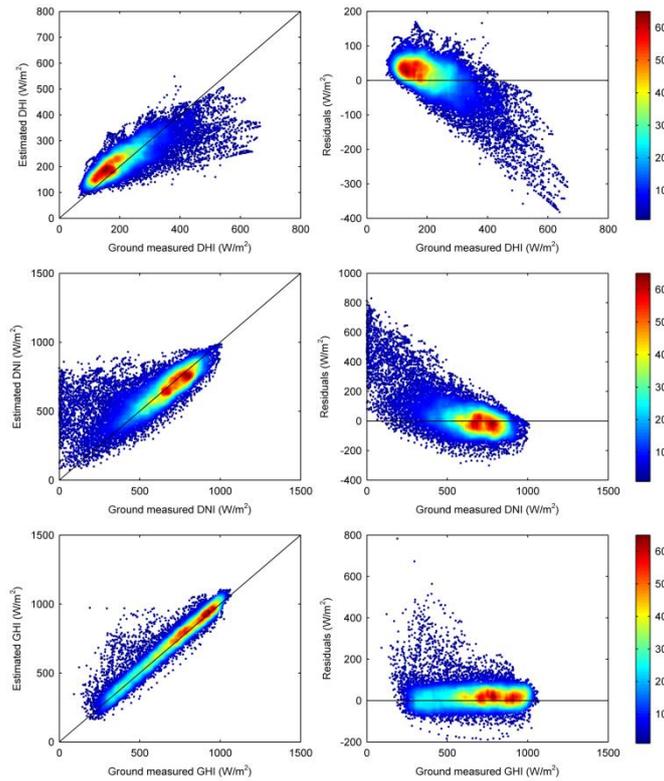


830

831 Fig. 16. Daily mean ground measured thermal channels T04, T05 and T09 for the all
 832 sky conditions training dataset.

833

834



835

836 Fig. 17. Density scatter plots of estimated versus ground measured irradiance and
 837 residuals versus ground measured irradiance for the McClear model fitted and tested on the cloud-free
 838 conditions training and testing datasets.

839

840