The Generalized Additive Model for the Assessment of the Direct, Diffuse, and Global Solar Irradiances Using SEVIRI Images, With Application to the UAE

Taha B. M. J. Ouarda, *Member, IEEE*, Christian Charron, Prashanth R. Marpu, *Senior Member, IEEE*, and Fateh Chebana

***Abstract*—Generalized additive models (GAMs) can model the nonlinear relationship between a response variable and a set of explanatory variables through smooth functions. GAM is used to assess the direct, diffuse, and global solar components in the United Arab Emirates (UAE), a country which has a large potential for solar energy production. Six thermal channels of the spinning enhanced visible and infrared imager (SEVIRI) instrument onboard Meteosat second generation (MSG) are used as explanatory variables along with the solar zenith angle, solar time, day number, and eccentricity correction. The proposed model is fitted using reference data from three ground measurement stations for the full year of 2010 and tested on two other stations for the full year of 2009. The performance of the GAM model is compared to the performance of the ensemble of artificial neural networks (ANNs) approach. Results indicate that GAM leads to improved estimates for the testing sample when compared to the bagging ensemble. GAM has the advantage over ANN-based models that we can explicitly define the relationships between the response variable and each explanatory variable through smooth functions. Attempts are made to provide physical explanations of the relations between irradiance variables and explanatory variables. Models in which the observations are separated as cloud-free and cloudy and treated separately are evaluated along with the combined dataset. Results indicate that no improvement is obtained compared to a single model fitted with all observations. The performance of the GAM is also compared to the McClear model, a physical-based model providing estimates of irradiance in clear sky conditions.**

***Index Terms*—Assessment of solar radiation, generalized additive models (GAMs), satellite images, solar energy, spline functions.**

NOMENCLATURE

DNI	direct normal irradiance ($\mathrm{W/m^2}$)
DHI	diffuse horizontal irradiance ($\mathrm{W/m^2}$)
GHI	global horizontal irradiance ($\mathrm{W/m^2}$)
θ_z	solar zenith angle (degrees)
ϵ	eccentricity correction
δ	total optical depth of the atmosphere
I_0	solar constant ($1367\ \mathrm{W/m^2}$)
m	air mass
T04	SEVIRI T04 channel (3.9 μm) observed brightness temperature (K)
T05	SEVIRI T05 channel (6.2 μm) observed brightness temperature (K)
T06	SEVIRI T06 channel (7.3 μm) observed brightness temperature (K)
T07	SEVIRI T07 channel (8.7 μm) observed brightness temperature (K)
T09	SEVIRI T08 channel (10.8 μm) observed brightness temperature (K)
T10	SEVIRI T10 channel (12.0 μm) observed brightness temperature (K)
ANN	artificial neural network
GAM	generalized additive model
GLM	generalized linear model
RMSE	root-mean-squared error
MBE	mean bias error
rRMSE	relative RMSE (%)
rMBE	relative MBE (%)
$\mathbf{X}$	matrix of explanatory or independent variables
$\mathbf{Z}$	model matrix for the basis functions
$\mathbf{A}$	influence matrix
Y	response or dependent random variable
X	explanatory or independent random variable
$\mathbf{y}$	vector of observed values of Y
g	the link function in GAM and GLM
$\boldsymbol{\beta}$	unknown parameters of the linear model
$\boldsymbol{\theta}$	vector of unknown parameters of the basis functions
f	smooth functions
b	spline basis functions
λ	smoothing parameter

Manuscript received October 06, 2015; revised December 17, 2015; accepted January 22, 2016. Date of publication February 18, 2016; date of current version March 11, 2016.

T. B. M. J. Ouarda is with the Institute Center for Water and Environment (iWater), Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, United Arab Emirates, and also with the INRS-ETE, National Institute of Scientific Research, Quebec City, QC, G1K9A9, Canada (e-mail: touarda@masdar.ac.ae).

C. Charron and P. R. Marpu are with the Institute Center for Water and Environment (iWater), Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, United Arab Emirates (e-mail: ccharron@masdar.ac.ae; pmarpu@masdar.ac.ae).

F. Chebana is with the INRS-ETE, National Institute of Scientific Research, Quebec City, QC, G1K9A9, Canada (e-mail: fateh.chebana@ete.inrs.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTARS.2016.2522764

I. INTRODUCTION

SOLAR radiation reaching the earth is divided into different components. Direct normal irradiance (DNI) refers to the radiation received from a straight beam of light from the

direction of the sun at its current position to a surface that is always normal to that solar beam. Diffuse horizontal irradiance (DHI) is the radiation received by a horizontal surface from radiation scattered by the atmosphere and coming from all directions. Global horizontal irradiance (GHI) is the total amount of radiation received on a surface parallel to the ground. Assessment of solar radiation on the earth's surface is of primary importance for many applications in solar energy. For instance, the accurate assessment of DNI is needed for concentrating solar power systems or other installations that track the position of the sun. To model the global tilt irradiance for fixed flat plate collectors, the assessment of DNI, DHI, and GHI is required [1], [2].

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

Data acquired from satellite images have been extensively used for estimation of solar radiation on the earth's surface. Several models classified as physics-based, empirical, and hybrid models were proposed with a good adaptation for the regions of interest. An example of the physics-based modeling is the model of Gautier *et al.* [5] to estimate the GHI in North America. It was later on adapted by Cogliani *et al.* [6] using Meteosat images to produce SOLARMET. The original Heliosat model of Cano *et al.* [7] was used to estimate GHI, DNI, and DHI over the USA. It was later adapted by Perez *et al.* [8] for GOES images. The operational physical model of Schillings *et al.* [3] was used to estimate DNI from Meteosat images. The Heliosat model has been modified and improved through different versions [9]–[14]. Heliosat-4 model is being currently validated [15], [16]. Another physics-based model for cloud-free conditions is the McClear model [17], which is based on look-up tables established with the radiative transfer model libRadtran [18].

On the other hand, data-driven statistical approaches have also been frequently used to perform solar radiation assessment. Artificial neural networks (ANNs) have been used successfully in a wide range of fields (see for instance [19]–[21]). They have been adapted for solar resources assessment in a number of studies [22]–[28]. In these studies, location-dependent parameters and meteorological parameters were used as inputs to model solar irradiance components. ANNs with an ensemble approach, which provide better generalization compared to a single ANN [29], [30], were used in Eissa *et al.* [23] to retrieve irradiance components over the United Arab Emirates (UAE). A simple bagging-like approach was used to develop the ensemble models. Alobaidi *et al.* [22] further improved on this model, by introducing a novel ensemble framework which significantly improved the results compared to the results obtained in the previous studies. The model employs a two-stage resampling process to build ensemble architectures for nonlinear regression. Although the model performs well, it involves an ensemble of ensembles' framework resulting in high computation load apart from the number of computationally expensive optimization steps while training the architecture. Another biggest drawback of ANN type models is that the relations between the inputs and outputs cannot be explicitly presented.

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

The UAE presents a high potential for solar energy development due to the long-day light period and the marginal amount of cloud cover. Recently, Eissa *et al.* [23], [42] and Alobaidi *et al.* [22] developed models to accurately estimate irradiance components over the UAE territory in which they used images of the earth's surface acquired by the spinning enhanced visible and infrared imager (SEVIRI) onboard Meteosat second generation (MSG) satellite.

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

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

A comparison is also made with the McClear model, a physical-based model providing estimates of irradiance in clear sky conditions. The results of McClear model are available through a web service at the website of the monitoring atmospheric composition and climate project (MACC project) (http://www.gmes-atmosphere.eu). Estimates could be obtained by just providing the latitude, longitude, and altitude (optional) of the target site, and the period of interest.

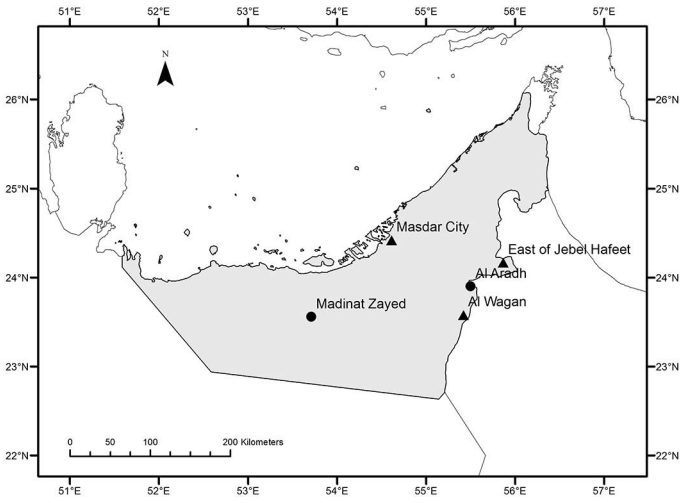

Fig. 1. Location of the ground measurement stations. Triangles represent stations of the training dataset and circles represent stations of the testing dataset.

II. Data

Ground measurements for DHI, DNI, and GHI consist of 10-min resolution data available at five stations over the UAE. At each station, data are collected using a rotating shadowband pyranometer (RSP). GHI is measured by the pyranometer when the shadowband is stationary. The shadowband makes a full rotation around the pyranometer. DHI is given by the lowest measured irradiance since at that moment DNI is completely blocked by the shadowband. DNI is deduced from GHI and DHI measured with the RSP. In the following, ground-measured DNI refers to DNI that is estimated from ground measured GHI and DHI. To match the 15-min resolution of the satellite data, successive ground measured data were interpolated. Data are available for the full year 2009 at the stations of Masdar City, Al Aradh, and Madinat Zayed, and for the full year 2010 at all stations. Fig. 1 presents the spatial distribution of the stations across the UAE.

Satellite images of the SEVIRI optical imager onboard MSG satellite were used in the present study. They provide continuous images of the earth in 12 spectral channels with a temporal resolution of 15 min and a spatial resolution of 3 km. Images from 6 thermal channels, T04 (3.9 μm), T05 (6.2 μm), T06 (7.3 μm), T07 (8.7 μm), T09 (10.8 μm), and T10 (12.0 μm), were collected and converted into brightness temperature. For each station, 3×3 pixels, with the station located in the center pixel, were extracted from satellite data. The other variables, solar zenith angle (θ_Z), Time, Day, and eccentricity correction ϵ, were computed for each pixel. The choice of the selected thermal channels is justified in Eissa *et al.* [23] by their sensitivity to the different constituents of the atmosphere: channels T05 and T06 are known to be affected by water vapor and T07, T08, and T09 are frequently used for dust detection. T04 was also selected in Eissa *et al.* [23] because it had improved their model accuracy.

The dataset is divided into training and testing datasets. The model is developed using the training dataset and tested using the testing dataset. The training dataset includes data from the stations of Masdar City, East of Jebel Hafeet, and Al Wagan for the full year 2010. The testing dataset includes data from the stations of Al Aradh and Madinat Zayed for the full year 2009. The training and testing datasets are further divided, respectively, into cloud-free and cloudy datasets. For this, a cloud mask was applied in which each pixel was classified as cloud-free or cloudy. The thin cirrus test [43], employing the T09 and T10 channels of SEVIRI, was used as a cloud mask following [23]. In all, the cloud-free and cloudy training datasets contain 29 193 and 7086 observations, respectively, and the cloud-free and cloudy testing datasets contain 16 864 and 2856 observations, respectively.

III. Methodology

A. Generalized Additive Model

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

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

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

The GAM model [45] is an extension of the GLM in which the linear predictor is replaced by a set of nonparametric functions of the explanatory variables. GAM can then be expressed by

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

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

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

where $b_{ji}(x_j)$ is the ith basis function of the jth explanatory variable evaluated at x_j, q_j is the number of basis functions for the jth explanatory variable, and θ_{ji} are unknown parameters.

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

$$g(\mathrm{E}(\mathbf{y})) = \mathbf{Z}\boldsymbol{\theta} \tag{4}$$

where $\mathbf{y}$ is a vector of observed values of the response variable Y, $\mathbf{Z}$ is a matrix including all the model matrix $\mathbf{Z}_j$, and $\boldsymbol{\theta}$ is a vector including all the smooth coefficient vectors $\boldsymbol{\theta}_j$. Parameters $\boldsymbol{\theta}$ could be estimated by the maximum likelihood method, but if q_j is large enough, the model will generally overfit the data. For that reason, GAM is usually estimated by penalized likelihood maximization. The penalty is typically a measure of the wiggliness of the smooth functions and is given by $\boldsymbol{\theta}_j^{\mathrm{T}}\mathbf{S}_j\boldsymbol{\theta}_j$ for the jth smooth function where $\mathbf{S}_j$ is a matrix of known coefficients. The penalized likelihood maximization objective is then given by

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

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

$$\nu_g = \frac{n\left\|\mathbf{y} - \mathbf{Z}\hat{\theta}\right\|^2}{[n - \mathrm{tr}(\mathbf{A})]^2} \tag{6}$$

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

The smooth functions used in this study are cubic regression splines. Cubic splines are constructed with piecewise cubic polynomials joined together at points called knots. The definition of the cubic smoothing spline basis arises from the solution of the following optimization problem [47]. Among all functions $f(x)$, with two continuous derivatives, find one that minimizes the penalized residual sum of squares

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

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

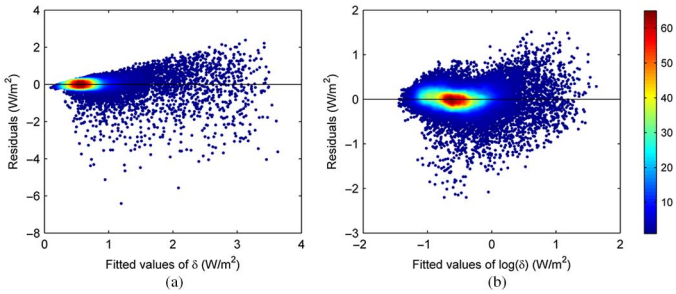

Fig. 2. Density scatter plots of residuals versus model fitted values for (a) δ and (b) $\log(\delta)$.

B. Model Configurations

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

$$\mathrm{DHI} = \alpha + \sum_{j=1}^{r} f_j(X_j) \tag{8}$$

where X_j is the jth explanatory variable, r is the number of explanatory variable included in the model, and α is the intercept.

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

$$\mathrm{DNI} = I_0\,\varepsilon\exp(-m\delta) \tag{9}$$

where I_0 is the solar constant with an approximate value of 1367 $\mathrm{W/m^2}$, and m is the air mass. The values of δ were computed from ground measured DNI. Parameters m and ϵ can be easily computed for any location on a given day by knowing θ_Z. For the estimation of δ with GAM, the following model is used:

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

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

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

TABLE I
RESULTS OBTAINED FOR THE MODELS FITTED ON THE SEPARATE CLOUD-FREE AND CLOUDY SKY CONDITIONS TRAINING DATASETS AND TESTED ON THE SEPARATE CLOUD-FREE AND CLOUDY SKY CONDITIONS TESTING DATASETS

Sky conditions	Statistic	GAM					ANN (Eissa *et al.* [23])			ANN (Alobaidi *et al.* [22])			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			

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

$$\mathrm{GHI} = \mathrm{DNI}\cos\theta_Z + \mathrm{DHI}. \tag{12}$$

In this study, we also propose to estimate GHI directly with GAM fitted on ground-measured GHI. Estimated GHI is then denoted by $\mathrm{GHI}^{\mathrm{D}}$, and the following model is used:

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

C. Validation Method

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

The performances are evaluated in terms of the root-mean-squared error (RMSE), mean bias error (MBE), relative RMSE (rRMSE), and relative MBE (rMBE). The rRMSE and rMBE are defined here by

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

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

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

IV. RESULTS

A. Models Trained and Tested on the Cloud-Free and Cloudy Sky Datasets

This section presents the results of the estimation of irradiance variables with GAM. Two separate models for each irradiance variable were fitted on the cloud-free and cloudy training datasets with all the explanatory variables included.

TABLE II
RESULTS OBTAINED FOR THE MODELS FITTED ON THE ALL SKY CONDITIONS TRAINING 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

Finally, each model was validated with either the cloud-free or the cloudy testing dataset. Table I presents the results obtained for the irradiance variables in terms of RMSE, MBE, rRMSE, and rMBE for cloud-free and cloudy conditions, and for both GAM and ANN models.

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

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

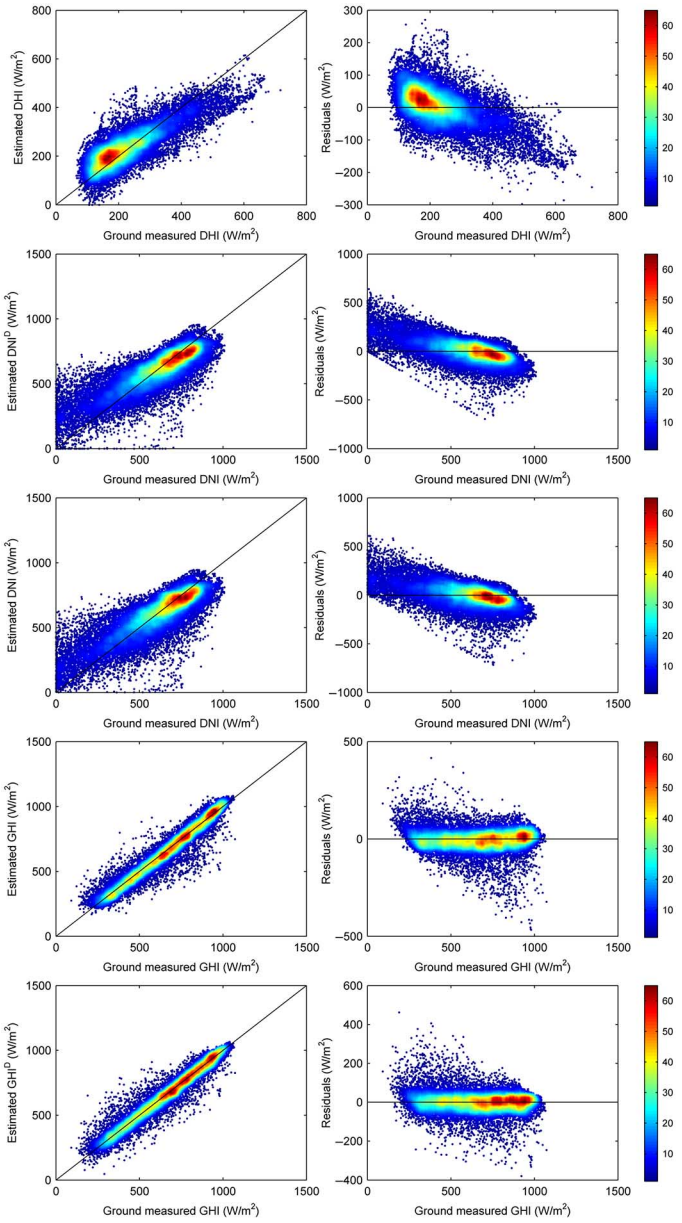

Fig. 3. Density scatter plots of estimated versus ground measured irradiance and residuals versus ground measured irradiance for the models fitted and tested on the all sky conditions training and testing datasets.

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

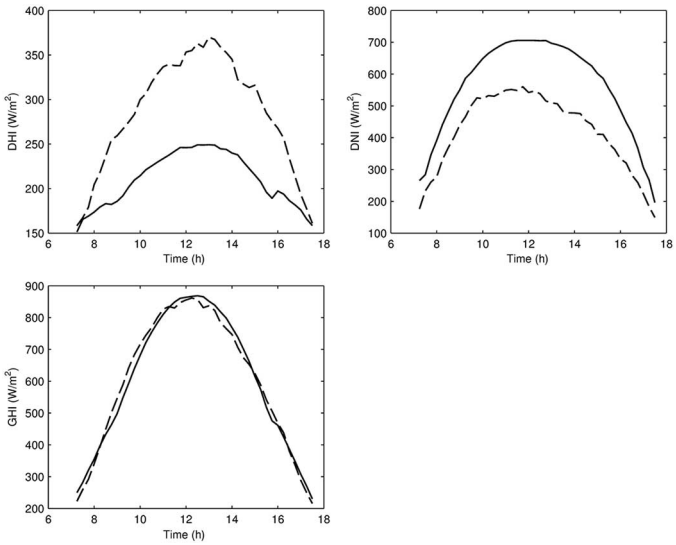

Fig. 4. Mean ground measured DHI, DNI, and GHI as function of time for the training dataset. Solid lines represent cloud-free conditions and dashed lines represent cloudy conditions.

TABLE III
RESULTS OBTAINED WITH MODELS WITHOUT ε

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

The models are fitted and tested on the all sky conditions training and testing datasets.

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

B. Single Model Trained on All Sky Dataset and Tested on Cloud-Free, Cloudy, and All Sky Datasets

In Eissa *et al.* [23] and Alobaidi *et al.* [22], two different ANN ensemble models were trained and tested separately for cloud-free and cloudy datasets. The impact of using separate datasets based on sky conditions is evaluated here. For that, a global model was fitted to the all sky conditions dataset and tested separately on the cloud-free, cloudy, and all sky testing datasets. Results obtained with the global model are presented in Table II. In the following, they are compared to the results of Table I. For the cloud-free case, RMSEs are in most cases slightly higher with the global model and MBEs equivalent for both approaches. For the cloudy case, no general conclusion can be made concerning RMSEs and MBEs. However, MBEs are significantly reduced for DNI and DNI^{D} with the global model. For the all sky conditions case, relative statistics represent a tradeoff between results when tested

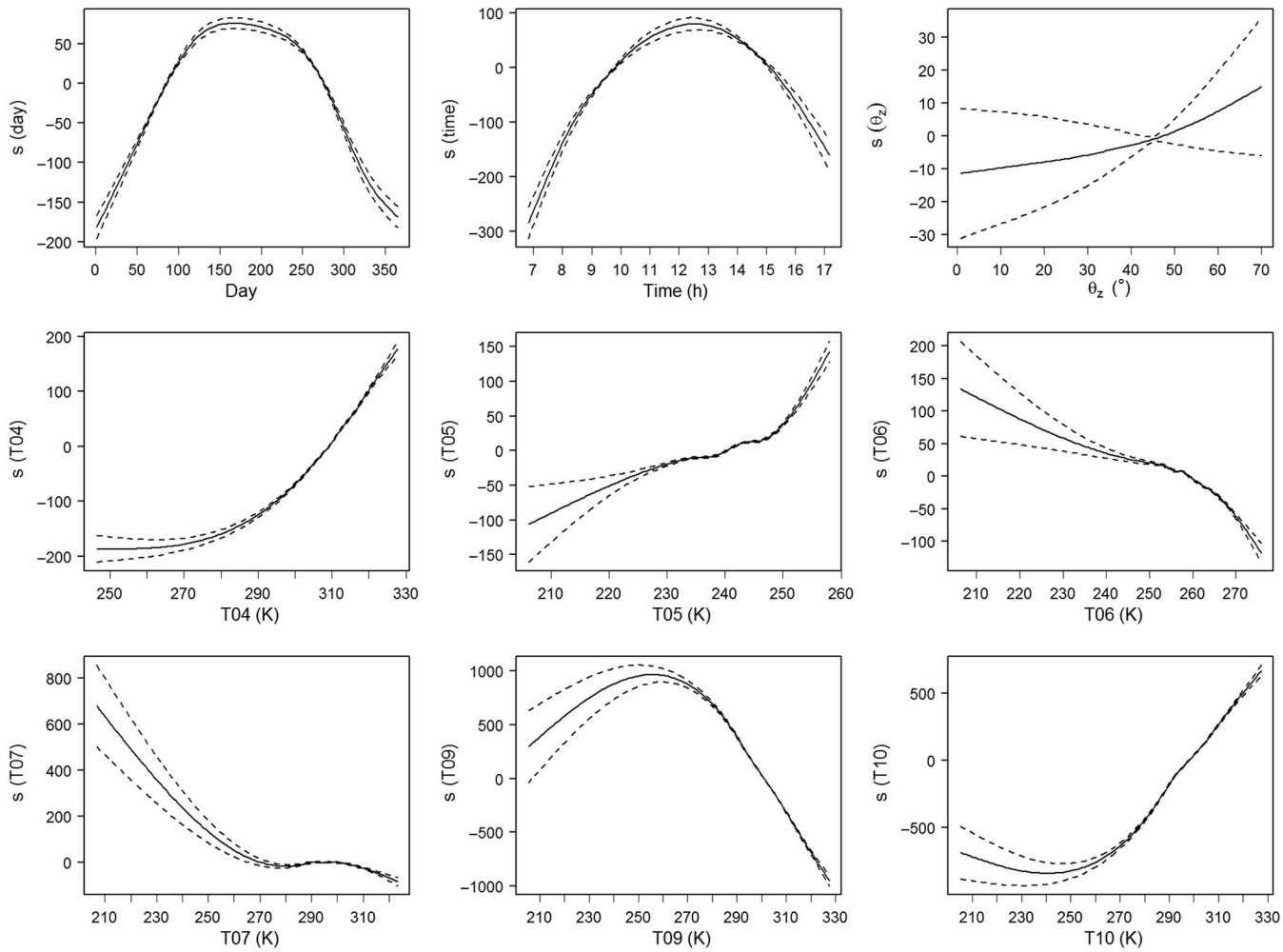

Fig. 5. Smooth functions of explanatory variables for the model estimating DHI fitted on the all sky conditions dataset. The dotted lines represent the limits of the 5% confidence interval.

on the cloud-free testing dataset and when tested on the cloudy testing dataset. This reflects the fact that both sky conditions testing datasets are mixed together. These overall results show that using separate models trained on cloud-free and cloudy conditions do not have a significant positive impact on the performances.

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

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

C. Interpretation of Smooth Functions

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

Attempts to obtain simpler models were carried out through stepwise regression methods. However, in most cases, the best model ends up being the model with all variables. Nevertheless, with GAM, it is hypothesized that the inclusion of ϵ is

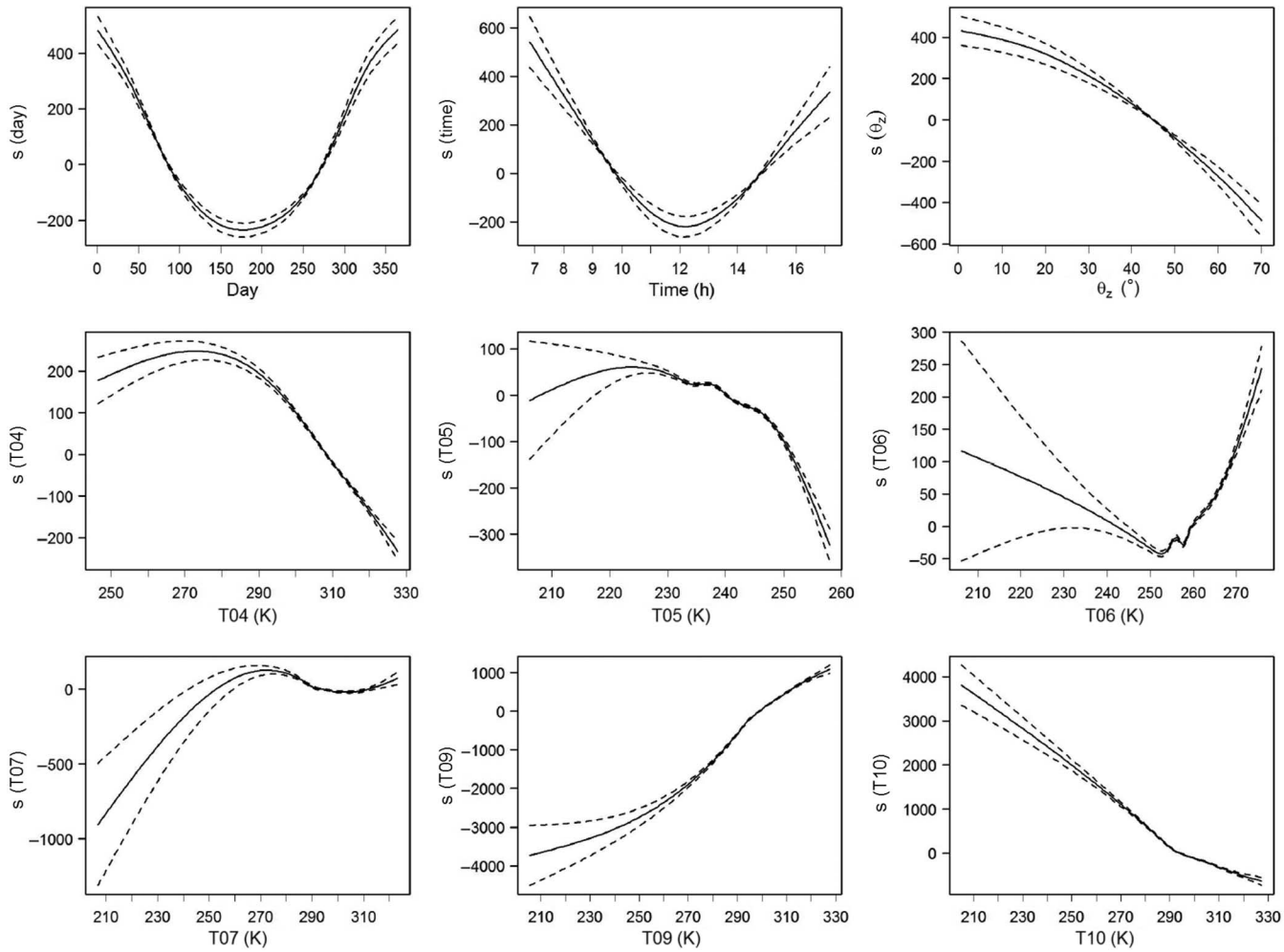

Fig. 6. Smooth functions of explanatory variables for the model estimating DNI^D fitted on the all sky conditions dataset. The dotted lines represent the limits of the 5% confidence interval.

unnecessary. Indeed, ϵ is computed at each location with a formula that depends only on day number, which is already included as an explanatory variable in the model. Table III presents the results obtained for the estimation of radiation variables with models using all explanatory variables except ϵ. The results obtained with and without ϵ are very similar and show that ϵ is redundant.

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

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

The smooth function of DHI versus Time increases with time to reach a maximum at around noon and decreases afterward. Because time is related to the sun height and therefore to irradiance intensity, it is expected to observe a similar shape of smooth curve for every irradiance variable. However, for DNI^D and GHI^D, an inverse relation is observed where the minimum irradiance is reached at around noon. This behavior is explained by the fact that the explanatory variable θ_Z, included in the model, also explains the sun position. In the case of DHI, the smooth function of θ_Z is strictly increasing. In this case, the time explains the sun position, and θ_Z explains a complementary portion of the total variance. In the case of DNI^D and GHI^D, θ_Z rather explains the sun position as the smooth functions are strictly decreasing with θ_Z.

The interpretation of the smooth functions of the predictors related to thermal channels is difficult because of their number and the fact that they are not independent. In all cases, a change in the slope of the curve occurs in midtemperatures. Confidence intervals are larger for low temperatures and decrease to become very small with increasing temperatures.

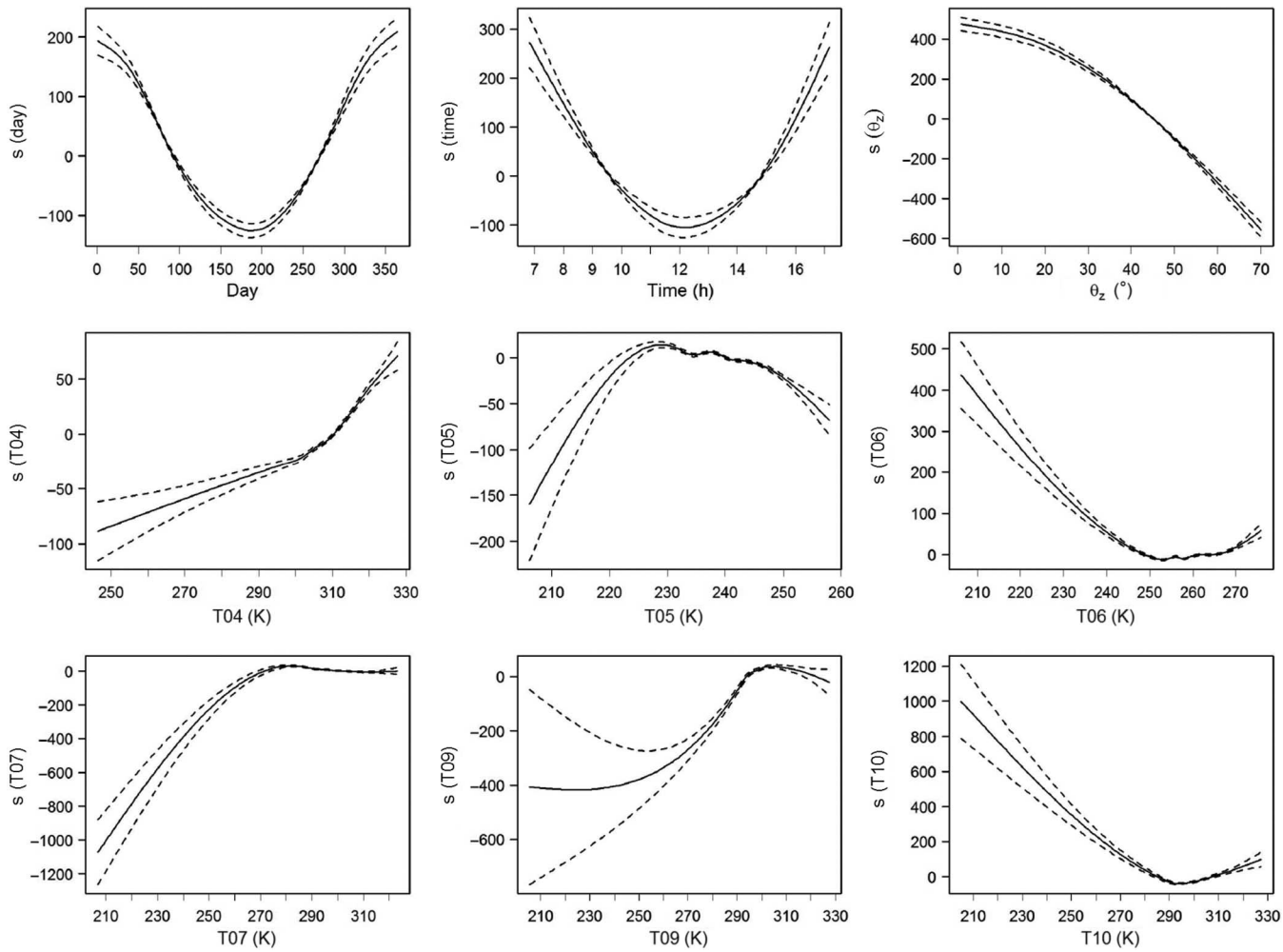

Fig. 7. Smooth functions of explanatory variables for the model estimating GHI^D fitted on the all sky conditions dataset. The dotted lines represent the limits of the 5% confidence interval.

This is explained by the fact that there are fewer observations for small temperatures as seen in the scatter plots of Figs. 8–10.

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

A strong seasonality is observed in both water vapor and AOT. It is, therefore, important to verify whether this seasonal behavior propagates also into the performance statistics. For this purpose, the year of the testing sample was divided in four seasons of three months, and the performance statistics were computed for each season. Table IV presents the performances for each season with the models without ϵ, fitted and tested on the all sky conditions training and testing datasets (i.e., the same models used in Table III). The results of Table IV show that biases are in general higher during the summer (AMJ and JAS), and RMSEs are higher during the winter season of JFM. The high bias values associated to the summer season can be

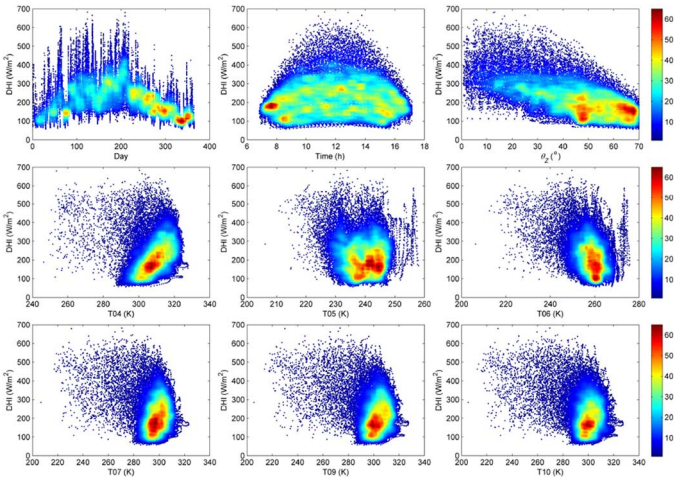

Fig. 8. Scatter plots of ground measured DHI versus explanatory variables for the all sky training conditions dataset.

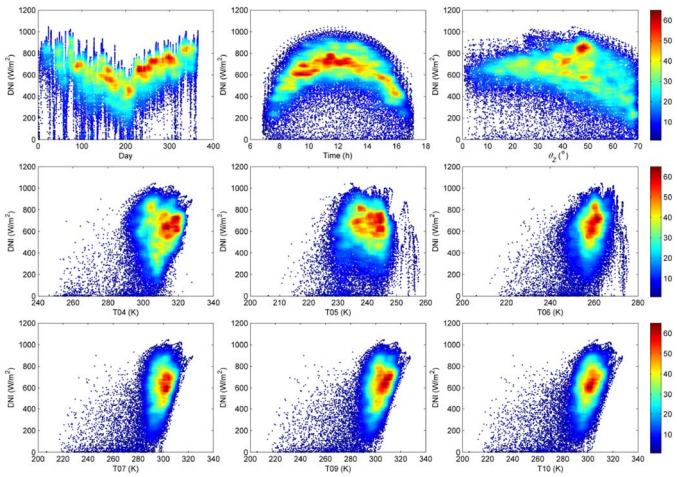

Fig. 9. Scatter plots of ground measured DNI versus explanatory variables for the all sky conditions training dataset.

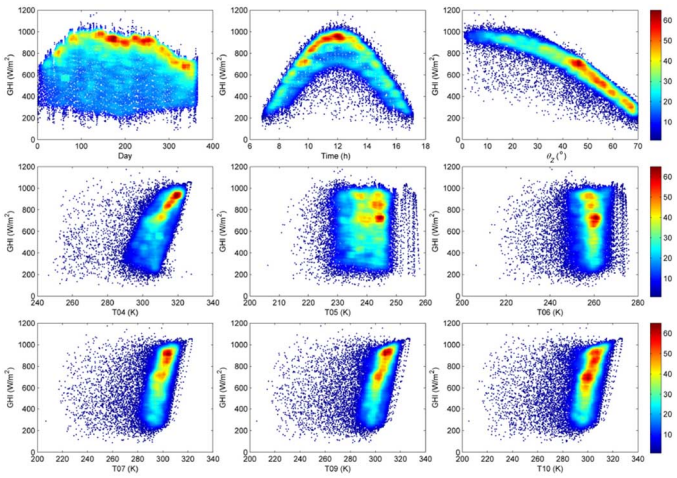

Fig. 10. Scatter plots of ground measured GHI versus explanatory variables for the all sky conditions training dataset.

explained by the scattering by aerosol constituents. It is also observed that the biases of DHI are generally of opposite sign than DNI^D and DNI. GHI biases are generally very small due to the canceling effect of the DHI and DNI biases.

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

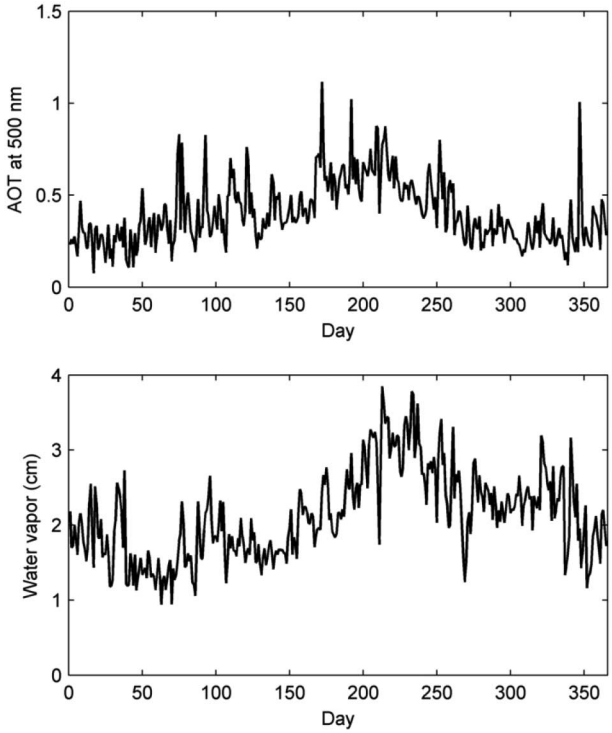

Fig. 11. Mean daily AOT at 500 nm and mean daily water vapor in Abu Dhabi.

TABLE IV
SEASONALITY IN THE PERFORMANCE STATISTICS

Season	Statistic	DHI	DNI	DNID	GHI	GHID
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

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

Smooth functions of variables Day, Time, and θ_Z have similar relationships with response variables than those obtained with the model with more variables. There is an exception in the case of DHI for θ_Z where the smooth function is now strictly decreasing. For the thermal channels, most observations occur after a certain threshold temperature which is channel dependent. This can be clearly seen in the scatter plots. Consequently, a change in slope occurs generally around this threshold temperature in the smooth functions of thermal channels. As the number of observations is negligible for the temperatures below the threshold, the analysis is restricted on temperatures higher than this threshold. For DHI, T09 is the most important thermal channel. Its smooth function has a strong negative slope. On the other hand, the smooth function for T04 increases

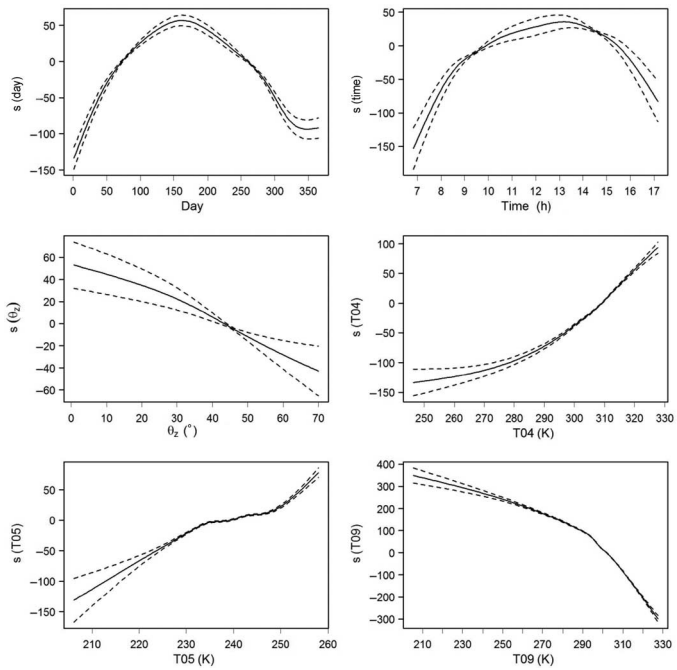

Fig. 12. Smooth functions of explanatory variables for the model estimating DHI fitted on the all sky conditions dataset (simplified model). The dotted lines represent the limits of the 5% confidence interval.

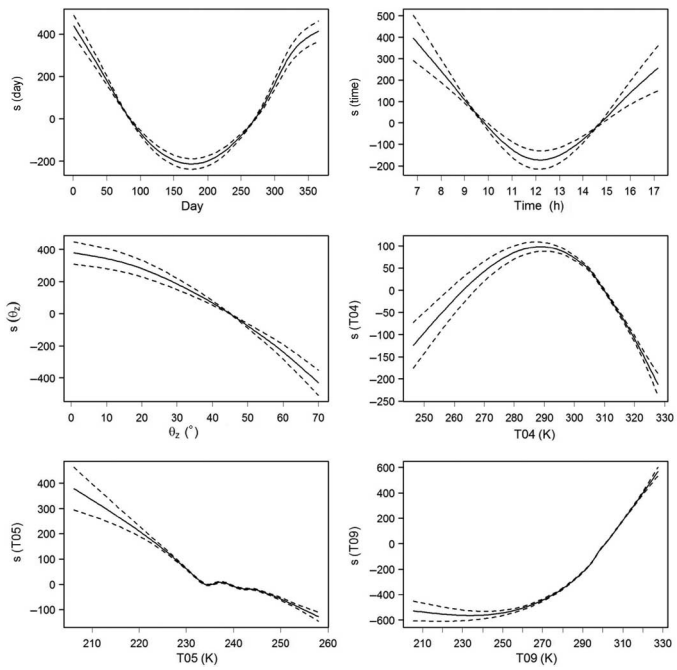

Fig. 13. Smooth functions of explanatory variables for the model estimating DNI^D fitted on the all sky conditions dataset (simplified model). The dotted lines represent the limits of the 5% confidence interval.

continuously. The smooth function of T05 increases continuously with a light slope. The scatter plots of Fig. 8 reveal that DHI has a positive relation with temperature for T04. For DNI^D, T09 is the most important thermal channel. Its smooth function increases constantly with a strong slope. The smooth function of T04 decreases continuously with a strong slope for high temperatures (Fig. 13). The smooth function of T05 has a light

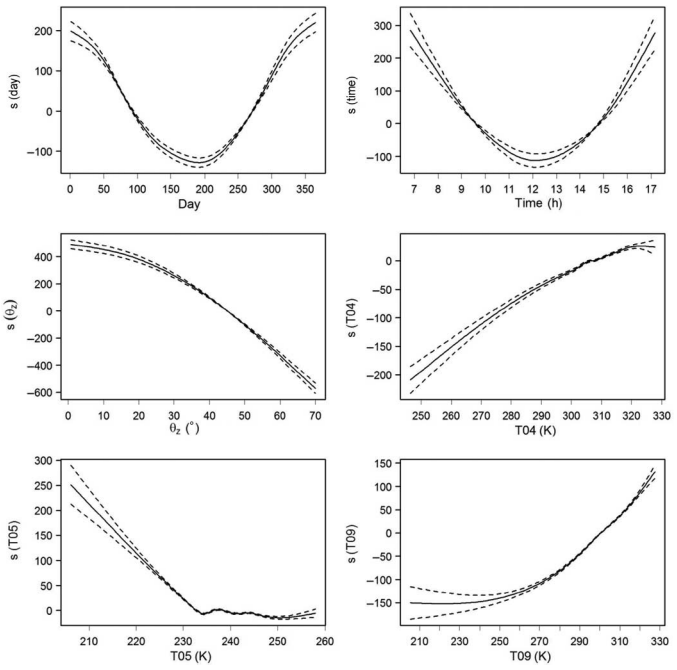

Fig. 14. Smooth functions of explanatory variables for the model estimating GHI^D fitted on the all sky conditions dataset (simplified model). The dotted lines represent the limits of the 5% confidence interval.

TABLE V
RESULTS OBTAINED WITH MODELS INCLUDING THE EXPLANATORY VARIABLES DAY, TIME, θ_Z, T04, T05, AND T09

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

The models are fitted and tested on the all sky conditions training and testing datasets.

decreasing slope. A strong positive relation of DNI with temperature for T09 is also observed in Fig. 9 while being less important for T04. For GHI, the smooth functions of T04 and T09 are both strictly increasing (Fig. 14). Strong positive relations are also observed in the scatter plots of thermal channels T04 and T09 in Fig. 10. The smooth function of T05 has a slope of about zero and is thus not very important.

The thermal channels T05 and T09 were chosen to represent, respectively, water vapor and dust in the atmosphere. We aim to evaluate to which extent these thermal channels capture the seasonality of the airborne constituents. For this, the individual thermal channel components of the linear predictor are displayed as a function of the day of the year. The simplified models DHI and DNI^D fitted on the all sky conditions dataset are considered. Fig. 15 presents the mean daily predicted DHI and DNI as a function of the day. It can be observed that the curves for T04 follow the seasonal evolution of the ground temperature with a peak during summer. For T05, a strong attenuation due to water vapor is observed where no noticeable seasonality can be observed. For T09, the same seasonal pattern than T04 is notice but with a small attenuation during summer due to dust. Fig. 16 presents the daily mean ground-measured

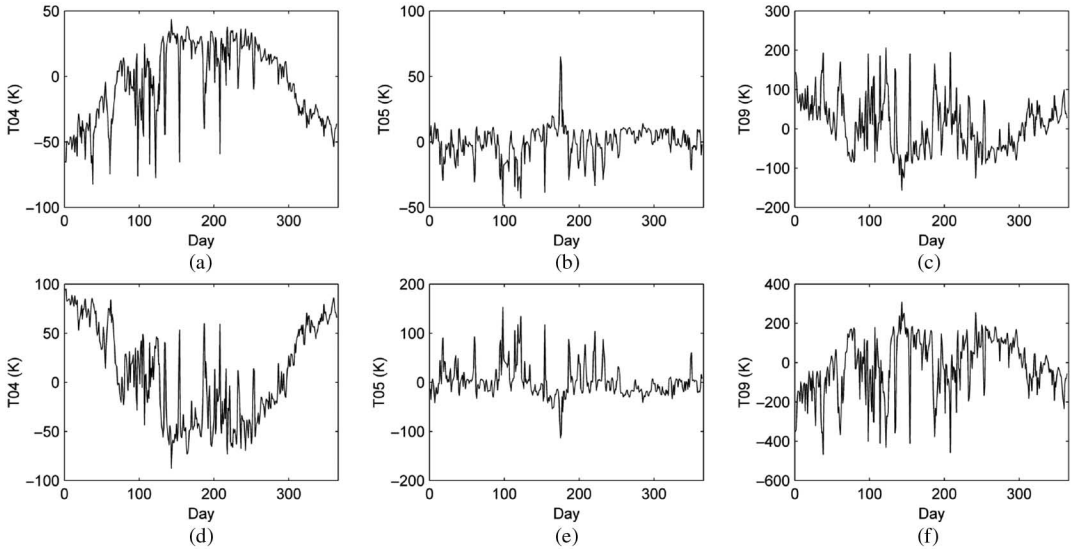

Fig. 15. Components of the linear predictor related to the thermal channels as a function of the day. The models DHI (a,b,c) and DNI^D (d,e,f) fitted on the all sky conditions dataset (simplified model) are considered.

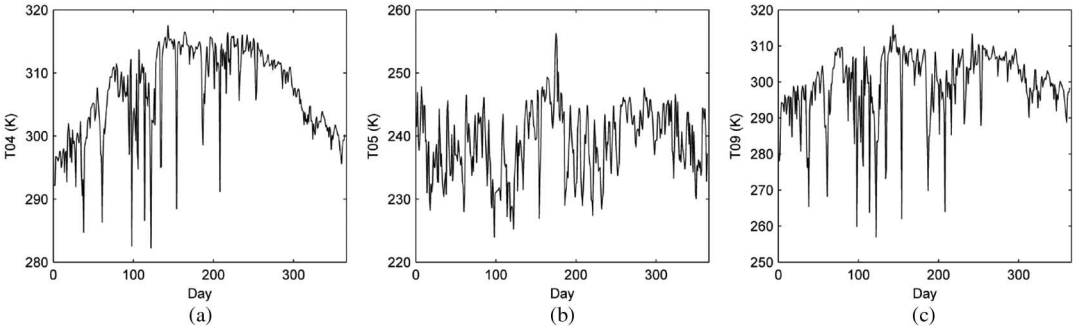

Fig. 16. Daily mean ground measured thermal channels T04, T05, and T09 for the all sky conditions training dataset.

thermal channels T04, T05, and T09 as a function of the day for the all sky conditions training dataset.

D. Comparison With McClear

Using the web service for McClear, estimates of irradiances were obtained at the two stations included in the testing sample during the same time period. Performance statistics computed for the cloud-free condition testing sample are added in Table I. Fig. 17 presents the density scatter plots of estimated variables with McClear versus ground measured variables. Scatter plots are rather similar to GAM. Same trends are observed in the residuals. One small difference that can be observed is that more observations of DHI are underestimated with McClear for very high irradiances. There is also more positive bias with McClear for very low DNI. Performances presented in Table I show that McClear, compared to GAM, has higher RMSEs for all variables and higher biases for DNI and GHI. In Eissa *et al.* [49], the McClear model was validated for the same stations as in the present study and better performances were obtained. This can be explained by the fact that the two publications used different methods to discriminate the cloud-free samples from the cloudy samples. Indeed, the algorithm of Long and Ackerman [50] was used in Eissa *et al.* [49] instead of the thin cirrus method used in the present work and in Eissa *et al.* [23]. The application of the Long and Ackerman method has resulted in a much lower proportion of retained cloud-free instants where only 65% of the data were considered cloud-free compared to 85% in the case of the present work. The algorithm of Long and Ackerman is more restrictive in its discrimination and might have removed some instants that were in fact cloudy.

V. Conclusion

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

Results were compared with those obtained with an ANN ensemble approach in Eissa *et al.* [23] and Alobaidi *et al.* [22]

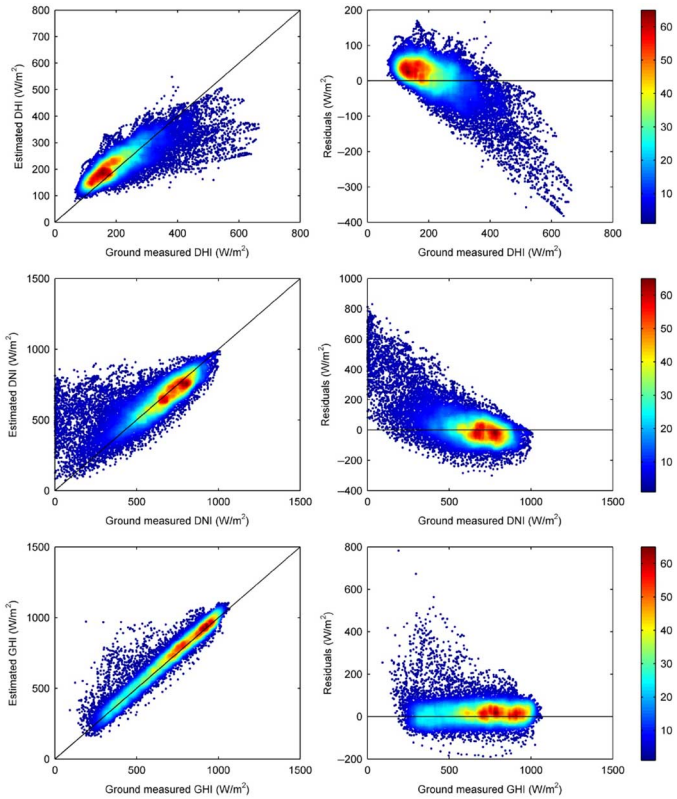

Fig. 17. Density scatter plots of estimated versus ground measured irradiance and residuals versus ground measured irradiance for the McClear model fitted and tested on the cloud-free conditions training and testing datasets.

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

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

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

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

ACKNOWLEDGMENT

The authors would like to thank the editor, Dr. Q. Du, the associate editor, and the two anonymous reviewers for their judicious comments. They also would like to thank the staff responsible maintaining the AERONET stations in the UAE.

REFERENCES

[1] C. A. Gueymard, "Direct and indirect uncertainties in the prediction of tilted irradiance for solar engineering applications," *Solar Energy*, vol. 83, pp. 432–444, 2009.

[2] B. Y. H. Liu and R. C. Jordan, "The long-term average performance of flat-plate solar-energy collectors: With design data for the U.S., its outlying possessions and Canada," *Solar Energy*, vol. 7, pp. 53–74, 1963.

[3] C. Schillings, H. Mannstein, and R. Meyer, "Operational method for deriving high resolution direct normal irradiance from satellite data," *Solar Energy*, vol. 76, pp. 475–484, 2004.

[4] A. Zelenka, R. Perez, R. Seals, and D. Renné, "Effective accuracy of satellite-derived hourly irradiances," *Theor. Appl. Climatol.*, vol. 62, pp. 199–207, Apr. 1, 1999.

[5] C. Gautier, G. Diak, and S. Masse, "A simple physical model to estimate incident solar radiation at the surface from GOES satellite data," *J. Appl. Meteorol.*, vol. 19, pp. 1005–1012, 1980.

[6] E. Cogliani, P. Ricchiazzi, and A. Maccari, "Physical model SOLARMET for determinating total and direct solar radiation by meteosat satellite images," *Solar Energy*, vol. 81, pp. 791–798, 2007.

[7] D. Cano, J. M. Monget, M. Albuisson, H. Guillard, N. Regas, and L. Wald, "A method for the determination of the global solar radiation from meteorological satellite data," *Solar Energy*, vol. 37, pp. 31–39, 1986.

[8] R. Perez *et al.*, "A new operational model for satellite-derived irradiances: Description and validation," *Solar Energy*, vol. 73, pp. 307–317, 2002.

[9] H. G. Beyer, C. Costanzo, and D. Heinemann, "Modifications of the heliosat procedure for irradiance estimates from satellite images," *Solar Energy*, vol. 56, pp. 207–212, 1996.

[10] A. Hammer *et al.*, "Solar energy assessment using remote sensing technologies," *Remote Sens. Environ.*, vol. 86, pp. 423–432, 2003.

[11] R. W. Mueller *et al.*, "Rethinking satellite-based solar irradiance modelling: The SOLIS clear-sky module," *Remote Sens. Environ.*, vol. 91, pp. 160–174, 2004.

[12] J. Polo, L. Martín, and M. Cony, "Revision of ground albedo estimation in Heliosat scheme for deriving solar radiation from SEVIRI HRV channel of meteosat satellite," *Solar Energy*, vol. 86, pp. 275–282, 2012.

[13] C. Rigollier, M. Lefèvre, and L. Wald, "The method Heliosat-2 for deriving shortwave solar radiation from satellite images," *Solar Energy*, vol. 77, pp. 159–169, 2004.

[14] M. Schroedter-Homscheidt, J. Betcke, G. Gesell, D. Heinemann, and T. Holzer-Popp, "Energy-specific solar radiation data from MSG: Current status of the HELIOSAT-3 project," in *Proc. 2nd MSG RAO Workshop*, 2004, p. 131.

[15] Z. Qu *et al.*, "Assessment of Heliosat-4 surface solar irradiance derived on the basis of SEVIRI-APOLLO cloud products," in *Proc. EUMETSAT Meteorol. Satellite Conf.*, 2012, pp. S2–S6.

[16] Z. Qu *et al.*, "A new method for assessing surface solar irradiance: Heliosat-4," in *Proc. EGU Gen. Assembly Conf. Abstracts*, 2012, p. 10228.

[17] M. Lefèvre *et al.*, "McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions," *Atmos. Meas. Techn.*, vol. 6, pp. 2403–2418, 2013.

[18] B. Mayer and A. Kylling, "Technical note: The libRadtran software package for radiative transfer calculations–description and examples of use," *Atmos. Chem. Phys.*, vol. 5, pp. 1855–1877, 2005.
[19] B. Khalil, T. B. M. J. Ouarda, and A. St-Hilaire, "Estimation of water quality characteristics at ungauged sites using artificial neural networks and canonical correlation analysis," *J. Hydrol.*, vol. 405, pp. 277–287, 2011.
[20] C. Shu and T. B. J. M. Ouarda, "Flood frequency analysis at ungauged sites using artificial neural networks in canonical correlation analysis physiographic space," *Water Resour. Res.*, vol. 43, p. W07438, Jul. 2007.
[21] I. Zaier, C. Shu, T. B. M. J. Ouarda, O. Seidou, and F. Chebana, "Estimation of ice thickness on lakes using artificial neural network ensembles," *J. Hydrol.*, vol. 383, pp. 330–340, Mar. 30, 2010.
[22] M. H. Alobaidi, P. R. Marpu, T. B. M. J. Ouarda, and H. Ghedira, "Mapping of the solar irradiance in the UAE using advanced artificial neural network ensemble," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 8, pp. 3668–3680, Aug. 2014.
[23] Y. Eissa, P. R. Marpu, I. Gherboudj, H. Ghedira, T. B. M. J. Ouarda, and M. Chiesa, "Artificial neural network based model for retrieval of the direct normal, diffuse horizontal and global horizontal irradiances using SEVIRI images," *Solar Energy*, vol. 89, pp. 1–16, 2013.
[24] A. Mellit and A. M. Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy," *Solar Energy*, vol. 84, pp. 807–821, 2010.
[25] K. Moustris, A. G. Paliatsos, A. Bloutsos, K. Nikolaidis, I. Koronaki, and K. Kavadias, "Use of neural networks for the creation of hourly global and diffuse solar irradiance data at representative locations in Greece," *Renew. Energy*, vol. 33, pp. 928–932, 2008.
[26] J. Mubiru and E. J. K. B. Banda, "Estimation of monthly average daily global solar irradiation using artificial neural networks," *Solar Energy*, vol. 82, pp. 181–187, 2008.
[27] O. Şenkal, "Modeling of solar radiation using remote sensing and artificial neural network in Turkey," *Energy*, vol. 35, pp. 4795–4801, 2010.
[28] F. S. Tymvios, C. P. Jacovides, S. C. Michaelides, and C. Scouteli, "Comparative study of Ångström's and artificial neural networks' methodologies in estimating global solar radiation," *Solar Energy*, vol. 78, pp. 752–762, 2005.
[29] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proc. Int. Conf. Mach. Learn.*, 1996, pp. 148–156.
[30] T. B. M. J. Ouarda and C. Shu, "Regional low-flow frequency analysis using single and ensemble artificial neural networks," *Water Resour. Res.*, vol. 45, p. W11428, 2009.
[31] D. L. Borchers, S. T. Buckland, I. G. Priede, and S. Ahmadi, "Improving the precision of the daily egg production method using generalized additive models," *Can. J. Fish. Aquat. Sci.*, vol. 54, pp. 2727–2742, 1997.
[32] L. Wen, K. Rogers, N. Saintilan, and J. Ling, "The influences of climate and hydrology on population dynamics of waterbirds in the lower Murrumbidgee river floodplains in Southeast Australia: Implications for environmental water management," *Ecol. Model.*, vol. 222, pp. 154–163, 2011.
[33] S. N. Wood and N. H. Augustin, "GAMs with integrated model selection using penalized regression splines and applications to environmental modelling," *Ecol. Model.*, vol. 157, pp. 157–177, 2002.
[34] F. Chebana, C. Charron, T. B. M. J. Ouarda, and B. Martel, "Regional frequency analysis at ungauged sites with the generalized additive model," *J. Hydrometeorol.*, vol. 15, pp. 2418–2428, Dec. 1, 2014.
[35] M. Durocher, F. Chebana, and T. B. M. J. Ouarda, "A nonlinear approach to regional flood frequency analysis using projection pursuit regression," *J. Hydrometeorol.*, vol. 16, pp. 1561–1574, Jul. 1, 2015.
[36] L. Bayentin, S. El Adlouni, T. Ouarda, P. Gosselin, B. Doyon, and F. Chebana, "Spatial variability of climate effects on ischemic heart disease hospitalization rates for the period 1989-2006 in Quebec, Canada," *Int. J. Health Geographics*, vol. 9, p. 5, 2010.
[37] C. Cans and C. Lavergne, "De la régression logistique vers un modèle additif généralisé: Un exemple d'application," *Rev. Stat. Appl.*, vol. 43, pp. 77–90, 1995.
[38] S. Clifford, S. Low Choy, T. Hussein, K. Mengersen, and L. Morawska, "Using the generalised additive model to model the particle number count of ultrafine particles," *Atmos. Environ.*, vol. 45, pp. 5934–5945, 2011.
[39] A. M. Leitte *et al.*, "Respiratory health, effects of ambient air pollution and its modification by air humidity in Drobeta-Turnu Severin, Romania," *Sci. Total Environ.*, vol. 407, pp. 4004–4011, 2009.
[40] J. Rocklöv and B. Forsberg, "The effect of temperature on mortality in Stockholm 1998 2003: A study of lag structures and heatwave effects," *Scand. J. Public Health*, vol. 36, pp. 516–523, 2008.
[41] V. Vieira, T. Webster, J. Weinberg, and A. Aschengrau, "Spatial analysis of bladder, kidney, and pancreatic cancer on upper Cape Cod: An application of generalized additive models to case-control data," *Environ. Health*, vol. 8, p. 3, 2009.
[42] Y. Eissa, M. Chiesa, and H. Ghedira, "Assessment and recalibration of the Heliosat-2 method in global horizontal irradiance modeling over the desert environment of the UAE," *Solar Energy*, vol. 86, pp. 1816–1825, 2012.
[43] J. Hocking, P. N. Francis, and R. Saunders, "Cloud detection in Meteosat second generation imagery at the met office," *Meteorol. Appl.*, vol. 18, pp. 307–323, 2011.
[44] J. A. Nelder and R. W. M. Wedderburn, "Generalized linear models," *J. R. Stat. Soc. A (Gen.)*, vol. 135, pp. 370–384, 1972.
[45] T. Hastie and R. Tibshirani, "Generalized additive models," *Stat. Sci.*, vol. 1, pp. 297–310, 1986.
[46] S. N. Wood, *Generalized Additive Models: An Introduction with R*. London, U.K.: Chapman & Hall, 2006.
[47] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*. London, U.K.: Chapman & Hall, 1990.
[48] B. N. Holben *et al.*, "AERONET—A federated instrument network and data archive for aerosol characterization," *Remote Sens. Environ.*, vol. 66, pp. 1–16, 1998.
[49] Y. Eissa *et al.*, "Validating surface downwelling solar irradiances estimated by the McClear model under cloud-free skies in the United Arab Emirates," *Solar Energy*, vol. 114, pp. 17–31, 2015.
[50] C. N. Long and T. P. Ackerman, "Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects," *J. Geophys. Res.: Atmos.*, vol. 105, pp. 15609–15626, 2000.
[51] J. M. Cunderlik, T. B. M. J. Ouarda, and B. Bobée, "On the objective identification of flood seasons," *Water Resour. Res.*, vol. 40, p. W01520, 2004.

Taha B. M. J. Ouarda (M'14) received the Ph.D. degree in civil engineering from Colorado State University, Fort Collins, CO, USA, in 1991.

He was the Chairman of the Canada Research Chair on the Estimation of Hydro-Meteorological Variables and is currently the Head of the Institute Center for Water and Environment with Masdar Institute of Science and Technology, Abu Dhabi, UAE.

Christian Charron received the Bachelor degree in computer engineering from the Laval University, Quebec City, QC, Canada, in 2001, and the M.Sc. degree in water science from the National Institute of Scientific Research (INRS), Quebec City, QC, Canada, in 2007.

Currently, he is a Research Engineer with the Institute Center for Water and Environment (iWater), Masdar Institute of Science and Technology, Abu Dhabi, UAE.

Prashanth R. Marpu (M'04–SM'14) received the M.Sc. degree in wireless engineering from the Technical University of Denmark, Lyngby, Denmark, in 2006, and the Ph.D. degree from TU Freiberg, Freiberg, Germany, in 2009.

He is an Assistant Professor with Masdar Institute of Science and Technology, Abu Dhabi, UAE. His research interests include machine learning, image processing, environmental data analysis, high-performance computing, remote sensing, and GIS.

Fateh Chebana received the Ph.D. degree in statistics from the University Paris 6, Paris, France, in 2003.

He is a Professor of Statistical Hydrology with the National Institute of Scientific Research (INRS), Quebec City, QC, Canada. His research interests include statistics with a focus on applications in hydrology and environmental health.