

Engineering Applications of Computational Fluid Mechanics

ISSN: 1994-2060 (Print) 1997-003X (Online) Journal homepage: https://www.tandfonline.com/loi/tcfm20

Construction of functional data analysis modeling strategy for global solar radiation prediction: application of cross-station paradigm

Ufuk Beyaztas, Sinan Q. Salih, Kwok-Wing Chau, Nadhir Al-Ansari & Zaher Mundher Yaseen

To cite this article: Ufuk Beyaztas, Sinan Q. Salih, Kwok-Wing Chau, Nadhir Al-Ansari & Zaher Mundher Yaseen (2019) Construction of functional data analysis modeling strategy for global solar radiation prediction: application of cross-station paradigm, Engineering Applications of Computational Fluid Mechanics, 13:1, 1165-1181, DOI: 10.1080/19942060.2019.1676314

To link to this article: https://doi.org/10.1080/19942060.2019.1676314

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 17 Oct 2019.

Submit your article to this journal

Article views: 618

View related articles

View Crossmark data

Citing articles: 5 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=tcfm20

ENGINEERING APPLICATIONS OF COMPUTATIONAL FLUID MECHANICS
2019, VOL. 13, NO. 1, 1165–1181
https://doi.org/10.1080/19942060.2019.1676314

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Construction of functional data analysis modeling strategy for global solar radiation prediction: application of cross-station paradigm

Ufuk Beyaztas[a], Sinan Q. Salih[b], Kwok-Wing Chau[c], Nadhir Al-Ansari[d] and Zaher Mundher Yaseen[e]

[a]Department of Statistics, Bartin University, Bartin 74100, Turkey; [b]Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam; [c]Department of Civil and Environmental Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, People's Republic of China; [d]Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 97187 Lulea, Sweden; [e]Sustainable Developments in Civil Engineering Research Group, Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam

ABSTRACT

To support initiatives for global emissions targets set by the United Nations Framework Convention on climate change, sustainable extraction of usable power from freely-available global solar radiation as a renewable energy resource requires accurate estimation and forecasting models for solar energy. Understanding the Global Solar Radiation (GSR) pattern is highly significant for determining the solar energy in any particular environment. The current study develops a new mathematical model based on the concept of Functional Data Analysis (FDA) to predict daily-scale GSR in the Burkina Faso region of West Africa. Eight meteorological stations are adopted to examine the proposed predictive model. The modeling procedure of the regression FDA is performed using two different internal parameter tuning approaches including Generalized Cross-Validation (GCV) and Generalized Bayesian Information Criteria (GBIC). The modeling procedure is established based on a cross-station paradigm wherein the climatological variables of six stations are used to predict GSR at two targeted meteorological stations. The performance of the proposed method is compared with the panel data regression model. Based on various statistical metrics, the applied FDA model attained convincing absolute error measures and best goodness of fit compared with the observed measured GSR. In quantitative evaluation, the predictions of GSR at the Ouahigouya and Dori stations attained correlation coefficients of $R = 0.84$ and 0.90 using the FDA model, respectively. All in all, the FDA model introduced a reliable alternative modeling strategy for global solar radiation prediction over the Burkina Faso region with accurate line fit predictions.

ARTICLE HISTORY
Received 1 August 2019
Accepted 1 October 2019

KEYWORDS
Burkina Faso; functional data analysis; global solar radiation; energy harvesting; regional investigation

1. Introduction

The growth in electrical energy demand is becoming a critical issue, especially as regards promoting sufficient technologies for solar (and other renewable) energy utilization that must support United Nations Sustainable Development Goal 7. Over the past three decades, the main genuine channel of energy as being through because it can maintain and sustain every process and activity that enhance the lives of animals, plants and other materials on earth (Yang, 2019). The main source of energy that meets the environmental challenges related to limited reserves and fossil fuels is solar (Li, Bu, Long, Zhao, & Ma, 2012; Ulgen & Hepbasli, 2004). Other forms of non-renewable energy can generate significant environmental issues. Renewable energy, includes tidal, solar, wind and geothermal, are favored because they present reduced environment impact compared to traditional means like fossil fuels. Hence, solar energy can be a sustainable and promising energy source that can minimize environmental hazards (De Souza et al., 2016).

Solar radiation from the sun resulting in solar energy is an electromagnetic radiation with a varied wavelength from radio waves (10–8 μm) to U-rays (10–6 μm) (Adeyefa & Adedokum, 1991). Extra-terrestrial and terrestrial spectra deviate from each other owing to different absorptions in the atmosphere (Ugwuoke & Okeke, 2012). In general, power per unit area covered by the sun regarding electromagnetic radiation within the measuring instrument wavelengths. Solar radiation helps in improving energy efficiency, de-carbonizing the global economy and ameliorating greenhouse gas emitter costs (Besharat, Dehghan, & Faghih, 2013). A coherent understanding and precise evaluation of solar radiation is needed for several applications, including: the supply of energy to natural processes and photovoltaic cell electrons that are in existence such as

CONTACT Zaher Mundher Yaseen yaseen@tdtu.edu.vn

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

photovoltaic and thermal photosynthesis systems (Yakintepe & Genc, 2015); climatology, meteorology, energy budgets and radiation, water treatment processes, natural and heating lighting, use of renewable energy, forestry and agriculture (De Souza et al., 2016); building energy-conscious designers and air conditioning engineers (Li et al., 2012; Muneer & Munawwar, 2006).

Solar radiation changes from one geographical area to another. It depends on:

(i) meteorological variables including the effects of cloud cover, evaporation, relative humidity, temperature, precipitation, extra-terrestrial solar radiation and sunshine duration;
(ii) geographical variables including the elevation of the site, longitude and latitude;
(iii) geometrical variables including the orientation and inclination angles of solar receivers;
(iv) astronomical variables including hour angle, solar constant, solar declination and earth–sun distance; and
(v) physical variables including water vapour content, scattering due to air molecules, scattering due to dust, earth–sun distance, and other atmospheric components such as CO_2, N_2 and O_2.

Different methods and measurements have been employed in the various parts of the world to measure global solar radiation. These techniques required consistent measurements using meteorological measuring instruments including satellite remote sensing and Eppley pyranometer instruments such as Meteosat-images and Moderate-Resolution Imaging Spectroradiometer (MODIS) products. Because of the maintenance, cost and skill required in producing satellite-derived data and ground measurements, especially in developing and rural nations, several prediction models have been postulated to generate global solar radiation data that do not require a high initial outlay for the instrumental network (Sunday, Agbasi, & Samuel, 2016; Sunday, Samuel, Agbasi, & Sylvia, 2016).

Any technologically conscious developing country can get through conventional or renewable sources. This might be due to the enormous energy usage required by some developing countries, expertise, cost of installation, and required maintenance. Thus, combining both non-renewable and renewable sources will favor power supply in developing countries; nevertheless, renewable energy sources should be focused upon owing to their minimal environmental hazard. Most West African countries lack global solar radiation data. For the past 30 years, Global Solar Radiation (GSR) has been evaluated based on the horizontal interface on a monthly and daily basis. Different kinds of empirical models have been postulated in several West African countries. Owing to this, several input variables have been used to achieve many functional forms. The models that have been employed fall into six groups, depending on the input variables used. These models were categorized into several subgroups, depending on the year postulated. Overall, a total of 68 functional forms and 356 empirical models have been postulated in previous studies for evaluating GSR in West African countries. Soft and empirical models were compared for evaluating GSR across West Africa, and the results obtained reflected a better outcome for soft computer models.

It is not possible to gather solar radiation data in many regions/locations owing to the absence of solar power stations. Thus, the solar radiation data for such locations have to be predicted, and the accuracy of the predictions depends on the model used (Yagli, Yang, & Srinivasan, 2019). Several statistical and data-driven models have been proposed for solar radiation prediction. For example: Olatomiwa, Mekhilef, Shamshirband, and Petković (2015b) developed A Neuro-Fuzzy Inference System (ANFIS) to predict solar radiation; Aybar-Ruiz et al. (2016) proposed a grouping genetic and extreme learning machine algorithms to predict global solar radiation; Kaplani, Kaplani, and Mondal (2018) investigated a spatiotemporal model for predicting daily global solar radiation; Meenal and Selvakumar (2018) compared the accuracy of several data-driven methods in predicting solar radiation; Bahrooz, Mert, and Kisi (2018) compared four different heuristic regression methods for estimating solar radiation; Khosravi, Koury, Machado, and Pabon (2018) proposed two machine learning algorithms to predict the hourly solar irradiance; Cornejo-Bueno, Casanova-Mateo, Sanz-Justo, and Salcedo-Sanz (2019) compared several machine learning regression techniques for global solar radiation estimation; and Torres-Barran, Alonso, and Dorronsoro (2019) evaluated the accuracy of random forest, gradient boosted and extreme gradient boosting regression models in solar radiation prediction. Such models model the data observed from a single time point. Throughout the literature, multiple investigations have been conducted in this regard, for example: solar energy prediction using linear and nonlinear models by the American Meteorological Society (Aggarwal & Saini, 2014); operational and ground-based models developed for solar radiation prediction for multiple advanced daily-scales throughout Greece (Kosmopoulos, Kazadzis, Lagouvardos, Kotroni, & Bais, 2015); the feasibility of a support vector regression model examined for solar irradiance prediction throughout coastal Taiwan (Kosmopoulos et al., 2015); the prediction of

monthly-scale global solar radiation conducted based on a statistical distribution modeling strategy using a clearness index for Nigeria (Ayodele & Ogunjuyigbe, 2015); and hourly-scale solar irradiance prediction established using the potential of the Long Short-Term Memory (LSTM) model for Santiago Island, Cape Verde (Qing & Niu, 2018). The literature has demonstrated noticeable progress in solar radiation pattern prediction using diverse advanced methodologies.

Among the procedures for the generation of global solar data, the ideal method is the use of a proper radiometric instrument that will directly measure the solar data at a given solar farm. However, the cost demand and expertise required for on-site global solar radiation measurement have limited the availability of radiometric data in most African and Asian countries (Zou et al., 2019). Another problem is the hosting of solar radiation stations in urban areas while effectively neglecting rural areas, where the energy crisis is more predominant. In Burkina Faso, most stations owned by the government do not have the capacity to measure solar radiation data routinely (Azoumah, Ramde, Tabsoba, & Thiam, 2010), while monthly or daily radiometric data are missing in areas with readily available data due to poor calibration of equipment. Solar radiation can also be generated using a meteorological reanalysis technique called Meteoblue (David & Lauret, 2018). This involves the use of physical models to simulate meteorological parameters physically (5 km × 5 km). This simulation relies on Non-hydrostatic Meso-scale Modeling (NMM) technology, which depends on parameters such as topography, soil and coverage. One major problem of this approach is that the generated values are simulated rather than real. However, its major advantage is the incorporation of physical processes that influence ground-based solar radiation. Given that mathematical equations with predetermined initial and model boundary conditions are used during the simulation, the data observed physically at a station may differ significantly from the predicted data (Fabbri, Canuti, & Ugolini, 2017).

Most investigations on solar radiation prediction have generally been done using empirical datasets collected from a single time point. However, datasets that are repeatedly measured over discrete time points may provide more information. Also, recent technological developments lead to data collection processes having high-dimensional and complex structures. Traditional statistical/mathematical techniques may not be applicable for such data types because of some difficulties such as multicollinearity, high dimensionality, high correlation between sequential observations, etc. Analyzing such datasets using Functional Data Analysis (FDA) techniques may be more useful since FDA has several important advantages over traditional statistical techniques. For example, FDA does not suffer from the missing data problem and the high correlation problem between repeated measurements; by smoothing the data, it minimizes the noise present in the data; and it can be used for irregularly sampled data. Thus, the need for FDA techniques is gradually increasing.

Functional regression models are used among others to explore the relationship between the functional response and predictor variables, and these models have received substantial attention in the literature. Also, they have successfully been used in many areas; see for example Valderrama, Ocana, Aguilera, and Ocana-Peinado (2010), Ivanescu, Staicu, Scheipl, and Greven (2015) and Chiou, Yang, and Chen (2016). See also Ferraty and Vieu (2006), Horvath and Kokoszka (2012) and Cuevas (2014) for more information about functional regression models and their applications. In this paper, we propose a functional regression model to predict global solar radiation data using meteorological variables so as to improve prediction accuracy. In summary, the proposed model works as follows: first, Gaussian basis function expansion and two information criteria – Generalized Cross-Validation (GCV) and Generalized Bayesian Information Criteria (GBIC) – are used to convert discretely observed data into a functional form. Second, the penalized log-likelihood method is used to estimate the discretized version of the model parameter matrix. Finally, the coefficient function of the functional regression model is obtained by applying a smoothing step. To the best of our knowledge, this work is the first study to predict global solar radiation data using a functional regression model. For future work, the FDA procedure proposed in this study can be extended to other real-life problems as an alternative to the methods proposed by Chau and Muttil (2007), Ghorbani, Kazempour, Chau, Shamshirband, and Ghazvinei (2017), Yaseen, Sulaiman, Deo, and Chau (2018) and Moazenzadeh, Mohammadi, Shamshirband, and Chau (2018).

The rest of the paper is organized as follows. Section 2 presents the details of the proposed method and the panel data regression model. The performance of the proposed method is evaluated with real-world data and the results are given in Sections 3 and 4. Section 5 concludes the paper.

2. Methodology

2.1. Functional regression model

Let $\{t\}_{j=1}^{J} \in \mathcal{T}$ represent the discrete time points at which the data is observed. For $n = 1, \ldots, N$ and $m = 1, \ldots, M$, let $\left(x_{nm}(s), y_n(t); s \in \mathcal{T}_m, t \in \mathcal{T}\right)$ denote the m functional

predictors and a functional response with ranges $\mathcal{T}_m \subset \mathbb{R}$ and $\mathcal{T} \subset \mathbb{R}$, respectively. The functional relationship between the predictors and response can be modeled by the following functional regression model (Matsui, Kawano, & Konishi, 2009; Ramsay & Silverman, 2005):

$$y_n(t) = \beta_0(t) + \sum_{m=1}^{M} \int_{\mathcal{T}_m} x_{nm}(s)\beta_m(s,t)\, ds + \epsilon_n(t), \quad (1)$$

where $\beta_0(t)$, $\beta_m(s,t)$ and $\epsilon_n(t)$ represent the intercept function, bivariate coefficient functions and error functions, respectively. For the sake of clarity, the role of function $\beta_0(t)$ can be eliminated by centering the functional predictors and response. Let $x^*_{nm}(s) = x_{nm}(s) - \bar{x}_m(s)$ and $y^*_n(t) = y_n(t) - \bar{y}(t)$, where $\bar{x}_m(s) = N^{-1}\sum_{n=1}^{N} x_{nm}(s)$ and $\bar{y}(t) = N^{-1}\sum_{n=1}^{N} y_n(t)$, denote the centered functional predictors and response, respectively. Then the functional regression model (1) can be written as follows:

$$y^*_n(t) = \sum_{m=1}^{M} \int_{\mathcal{T}_m} x^*_{nm}(s)\beta_m(s,t)\, ds + \epsilon^*_n(t), \quad (2)$$

where $\epsilon^*_n(t) = \epsilon_n(t) - \bar{\epsilon}(t)$ is the centered error functions. Hereafter it is assumed that both functional response and predictors are centered.

The first step in FDA is to smooth the functional data using a suitable basis function system. Let $\mathbf{\Phi}_k(t) = \{\phi_1(t), \ldots, \phi_K(t)\}$ denote a system of k, $k = 1, \ldots, K$, basis functions; then a function, say $y(t)$, can be defined as $y(t) = \sum_{k=1}^{K} c_k\phi_k(t)$ where c_k is the coefficient vector of the kth basis function $\phi_k(y)$. Accordingly, the smooth functions of the (centered) functional predictors $x_{nm}(s)$ and functional response $y_n(t)$ are defined as follows:

$$\begin{aligned} y^*_n(t) &= \sum_{k=1}^{K_y} c_{nk}\phi_k(t) = \mathbf{c}'_n\mathbf{\Phi}(t) \quad \forall t \in \mathcal{T}, \\ x^*_{nm}(s) &= \sum_{j=1}^{K_{m,x}} d_{nmj}\psi_{mj}(s) = \mathbf{d}'_{nm}\mathbf{\Psi}(s) \quad \forall s \in \mathcal{T}_m, \end{aligned} \quad (3)$$

where $\mathbf{\Phi}(t) = \{\phi_1(t), \ldots, \phi_{K_y}(t)\}'$ and $\mathbf{\Psi}(s) = \{\psi_{m1}(s), \ldots, \psi_{mK_{m,x}}(s)\}$ are vectors of basis functions and $\mathbf{c}_n = \{c_{n1}, \ldots, c_{nK_y}\}'$ and $\mathbf{d}_{nm} = \{d_{nm1}, \ldots, d_{nmK_{m,x}}\}'$ are the corresponding vectors of coefficients. Choosing the right basis functions is one of the most crucial steps in FDA. Several types of basis function, such as the Fourier basis, the B-splines basis and the radial basis, have been proposed to smooth functional data; please see Ramsay and Silverman (2005) for more details. We consider the following Gaussian basis functions in our numerical analyses (see Matsui et al., 2009):

$$\phi_k(t) = \exp\left\{-\frac{\left(t - \tau_{k+2}\right)^2}{2\sigma^2}\right\}, \quad (4)$$

$$\psi_{mj}(s) = \exp\left\{-\frac{\left(s - \tau^{(m)}_{j+2}\right)^2}{2\sigma_m^2}\right\}, \quad (5)$$

where the equally spaced knots τ_k and $\tau_j^{(m)}$ determine the centers of the basis functions, and $\sigma = \left(\tau_{k+2} - \tau_k\right)/2$ and $\sigma_m = \left(\tau^{(m)}_{j+2} - \tau^{(m)}_j\right)/2$ are the widths. Another important task in smoothing functional data is to choose the optimum number of basis functions K. Generally, (*i*) the data is well fitted by the functions when K is large, but the noise present in the data may not be eliminated; on the other hand (*ii*) some key features of the smooth function could be ignored when K is too small. To select the optimal K, we consider the generalized cross validation and generalized Bayesian information criteria proposed by Matsui et al. (2009).

Using the basis function, the bivariate coefficient functions $\beta_m(s,t)$ in (1) can be written as follows:

$$\beta_m(s,t) = \sum_{j,k} \psi_{mj}(s)b_{mjk}\phi_k(t) = \mathbf{\Psi}'_m(s)\mathbf{B}_m\mathbf{\Phi}(t), \quad (6)$$

where $\mathbf{B}_m = (b_{mjk})_{j,k}$ is a coefficient matrix with dimension $K_{m,x} \times K_y$. From (2), (3) and (6), the functional regression model given in (1) can be written as follows:

$$\begin{aligned} \mathbf{c}'_n\mathbf{\Phi}(t) &= \sum_{m=1}^{M} \mathbf{d}'_{nm}\mathbf{\Upsilon}_{\psi_m}\mathbf{B}_m\mathbf{\Phi}(t) + \epsilon^*_n(t), \\ &= g'_n\mathbf{B}\mathbf{\Phi}(t) + \epsilon^*_n(t), \end{aligned} \quad (7)$$

where $\mathbf{\Upsilon}_{\psi_m} = \int_{\mathcal{T}_m} \psi_m(s)\psi'_m(s)\, ds$ is a matrix with dimension $K_{m,x} \times K_{m,x}$, $g_n = \left(\mathbf{d}'_{n1}\mathbf{\Upsilon}_{\psi_1}, \ldots, \mathbf{d}'_{nM}\mathbf{\Upsilon}_{\psi_M}\right)'$ is a vector with length $\sum_{m=1}^{M} K_{m,x}$ and $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_M)'$ is the coefficient matrix with dimension $\sum_{m=1}^{M} K_{m,x} \times K_y$. Accordingly, the functional linear model for the whole system can be expressed as follows:

$$\mathbf{C}\mathbf{\Phi}(t) = \mathbf{G}\mathbf{B}\mathbf{\Phi}(t) + \boldsymbol{\epsilon}^*(t). \quad (8)$$

Several techniques, including the least squares, maximum likelihood and penalized maximum likelihood methods, have been proposed to estimate the coefficient matrix $\mathbf{B}$, see for example Ramsay and Silverman (2005), Yao, Muller, and Wang (2005), Konishi and Kitagawa (2008) and Matsui et al. (2009). The least squares and/or maximum likelihood methods provide

unstable/unfavorable estimates for the model parameters (Matsui et al., 2009), and thus the penalized maximum likelihood method proposed by Matsui et al. (2009), which controls the degree of smoothness of the functions and provides more flexible results, has been considered to estimate the functional parameters of the regression model (8).

Suppose that the error function $\boldsymbol{\epsilon}_n^*(t)$ has the form $\boldsymbol{\epsilon}_n^*(t) = \mathbf{e}_n'\boldsymbol{\Phi}(t)$, where the K_y-dimensional error vectors $\mathbf{e}_n = (e_{n1}, \ldots, e_{nK})'$ are assumed to be independent and identically distributed Gaussian random variables with mean $\mathbf{0}$ and variance–covariance matrix $\boldsymbol{\Sigma}$. Define the functional regression model (7) by

$$\mathbf{c}_n'\boldsymbol{\Phi}(t) = g_n'\mathbf{B}\boldsymbol{\Phi}(t) + \mathbf{e}_n'\boldsymbol{\Phi}(t). \tag{9}$$

Multiplying both sides of (9) by $\boldsymbol{\Phi}'(t)$ and integrating with respect to $\mathcal{T}$ yields

$$\begin{aligned} \mathbf{c}_n'\boldsymbol{\Upsilon}_\phi &= g_n'\mathbf{B}\boldsymbol{\Upsilon}_\phi + \mathbf{e}_n'\boldsymbol{\Upsilon}_\phi, \\ \mathbf{c}_n &= \mathbf{B}'g_n + \mathbf{e}_n. \end{aligned} \tag{10}$$

Let $f(y_n|x_n;\theta)$ with parameter vector $\boldsymbol{\theta} = (\mathbf{B}, \boldsymbol{\Sigma})$ denote the probability density function of the model (10). Then the penalized log-likelihood function for θ is obtained as follows:

$$\ell_\lambda(\theta) = \sum_n \log f(y_n|x_n;\theta) - \frac{N}{2}\mathrm{tr}\left\{\mathbf{B}'\left(\boldsymbol{\Lambda}_M \odot \boldsymbol{\Omega}\right)\mathbf{B}\right\}, \tag{11}$$

where $\boldsymbol{\Lambda}_M = \boldsymbol{\lambda}_M\boldsymbol{\lambda}_M'$ with $\boldsymbol{\lambda}_M = (\sqrt{\lambda_1}\mathbf{1}_{K_{1,x}}', \ldots, \sqrt{\lambda_M}\mathbf{1}_{K_{M,x}}')$ being a $(\sum_{m=1}^M K_{m,x}) \times (\sum_{m=1}^M K_{m,x})$ -dimensional matrix of penalty parameters, $\odot$ and $\mathrm{tr}\{\cdot\}$ are, respectively, the Hadamart product and the trace of a matrix, and $\boldsymbol{\Omega}$ is a positive semi-definite matrix. Equating the derivatives of the penalized log-likelihood function given in (11) with respect to $\theta = (\mathbf{B}, \boldsymbol{\Sigma})$ to $\mathbf{0}$ gives the penalized maximum likelihood estimators of θ, $\hat{\theta} = \left(\hat{\mathbf{B}}, \hat{\boldsymbol{\Sigma}}\right)$, as follows:

$$\begin{aligned} \mathrm{vec}(\hat{\mathbf{B}}) &= \left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{G}'\mathbf{G} + N\mathbf{I}_{K_y} \otimes \left(\boldsymbol{\Lambda}_M \odot \boldsymbol{\Omega}\right)\right)^{-1} \\ &\quad \left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{G}'\right)\mathrm{vec}(\mathbf{C}), \\ \hat{\boldsymbol{\Sigma}} &= \frac{1}{N}\left(\mathbf{C} - \mathbf{G}\hat{\mathbf{B}}\right)'\left(\mathbf{C} - \mathbf{G}\hat{\mathbf{B}}\right). \end{aligned} \tag{12}$$

Finally, the penalized maximum likelihood estimator of $\mathbf{C}$ is obtained as

$$\begin{aligned} \mathrm{vec}\left(\hat{\mathbf{C}}\right) &= \mathrm{vec}\left(\mathbf{G}\hat{\mathbf{B}}\right) = \left(\mathbf{I}_{K_y} \otimes G\right) \\ &\quad \times \left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{G}'\mathbf{G} + N\mathbf{I}_{K_y} \otimes \left(\boldsymbol{\Lambda}_M \odot \boldsymbol{\Omega}\right)\right)^{-1} \\ &\quad \times \left(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{G}'\right]\mathrm{vec}(\mathbf{C}). \end{aligned} \tag{13}$$

In practice, the performance of the penalized maximum likelihood method vigorously depends on a suitable choice of the parameter λ values, since the estimated model parameters $\hat{\theta}$ and $\hat{\mathbf{C}}$ depend on the penalty matrix $\boldsymbol{\Lambda}_M$. Several information criteria have been proposed to select proper penalty terms, λ_m, that minimize the corresponding objective function. Two different tuning parameter techniques including GCV and GBIC are implemented to obtain the suitable prediction process, as follows:

$$\mathrm{GCV} = \frac{\mathrm{tr}\left\{\left(\mathbf{C} - \mathbf{G}\hat{\mathbf{B}}\right)'\left(\mathbf{C} - \mathbf{G}\hat{\mathbf{B}}\right)\right\}}{NK_y\left(1 - \mathrm{tr}\left(\mathbf{S}_\lambda\right)/\left(NK_y\right)\right)^2}, \tag{14}$$

$$\begin{aligned} \mathrm{GBIC} = &-2\sum_{n=1}^N \log f(y_n|x_n;\hat{\theta}) + N\mathrm{tr}\left\{\hat{\mathbf{B}}'\left(\boldsymbol{\Lambda}_M \odot \boldsymbol{\Omega}\right)\hat{\mathbf{B}}\right\} \\ &+ (r + Kq)\log N - (r + Kq)\log(2\pi) \\ &- K\log|\boldsymbol{\Lambda}_M \odot \boldsymbol{\Omega}|_+ + \log|\mathbf{R}_\lambda(\hat{\theta})|, \end{aligned} \tag{15}$$

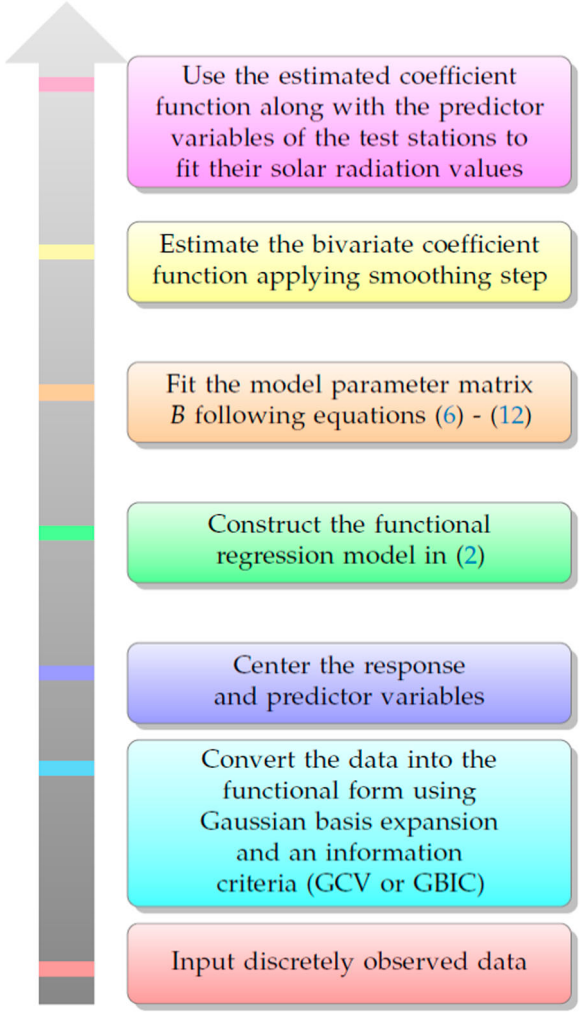


Figure 1. Flowchart of the proposed method.

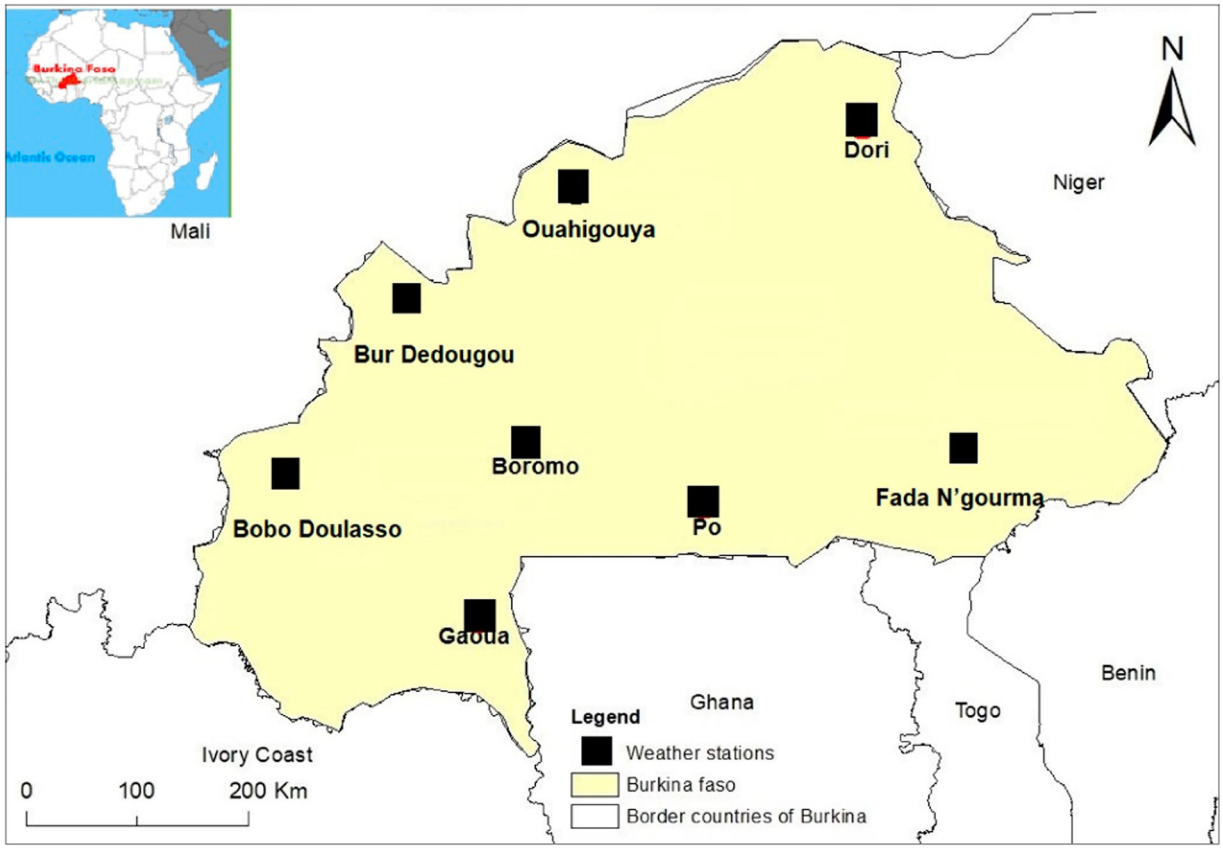


Figure 2. Study area; locations of stations.

where $q = p - \text{rank}(\boldsymbol{\Omega}), p = \sum_m K_{m,x}$ and $r = K_y(K_y + 1)/2$, please see Matsui et al. (2009) for the derivation of GCV and GBIC. To select the best prediction model, these two criteria work as follows: (1) the functional regression model is constructed based on the data, which are approximated by basis function expansion using several combinations of smoothing parameter and number of basis functions; then (2) both the GCV and GBIC select the best model according to a smoothing parameter – the number of basis function combinations that produces minimum GCV and GBIC values.

For the sake of clarity, a flowchart is presented in Figure 1 to show how the proposed method works to obtain the experimental results in this paper.

2.2. Panel data regression model

In this study, the finite sample performance of the proposed modeling strategy is compared with the linear panel data regression model with fixed effects. Let $i = 1, \ldots, N$ denote the individuals observed at time points $t = 1, \ldots, T$. Let also y_{it} and x_{it} denote the response and K-dimensional predictor variables. The linear panel data regression model is then defined as follows:

$$y_{it} = \alpha_i + \beta' x_{it} + u_{it}, \tag{16}$$

where α_i, β and u_{it} represent the individual effects, coefficient vector, and the error terms, respectively. The coefficient vector β is estimated using the Ordinary Least Squares (OLS) method. Briefly, let $\bar{y}_i, \bar{x}_i$ and $\bar{u}_i$ denote the averages of y_{it}, x_{it} and u_{it} for each individual $i = 1, \ldots, N$. Then, the OLS estimate of β is obtained as follows:

$$\hat{\beta} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}'_{it} \tilde{x}_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{x}'_{it} \tilde{y}_{it} \right), \tag{17}$$

where $\tilde{x}_{it} = x_{it} - \bar{x}_i$ and $\tilde{y}_{it} = y_{it} - \bar{y}_i$. Readers are referred to Baltagi (2005) for more information about the linear panel data regression model.

3. Case study

The solar radiation prediction model was developed for the Burkina Faso region, located in Sub-Saharan Africa (Figure 2). About 70% of the total power generation capacity in Burkina Faso is sourced from thermal-fossil fuel, while hydro-power accounts for the remaining 30% (REN21 2015, 2017). Owing to the incremental cost of production, the instability of the oil price, as well as the ever-increasing demand for electricity, the country recently installed 28 fossil-fuel powered stations with a

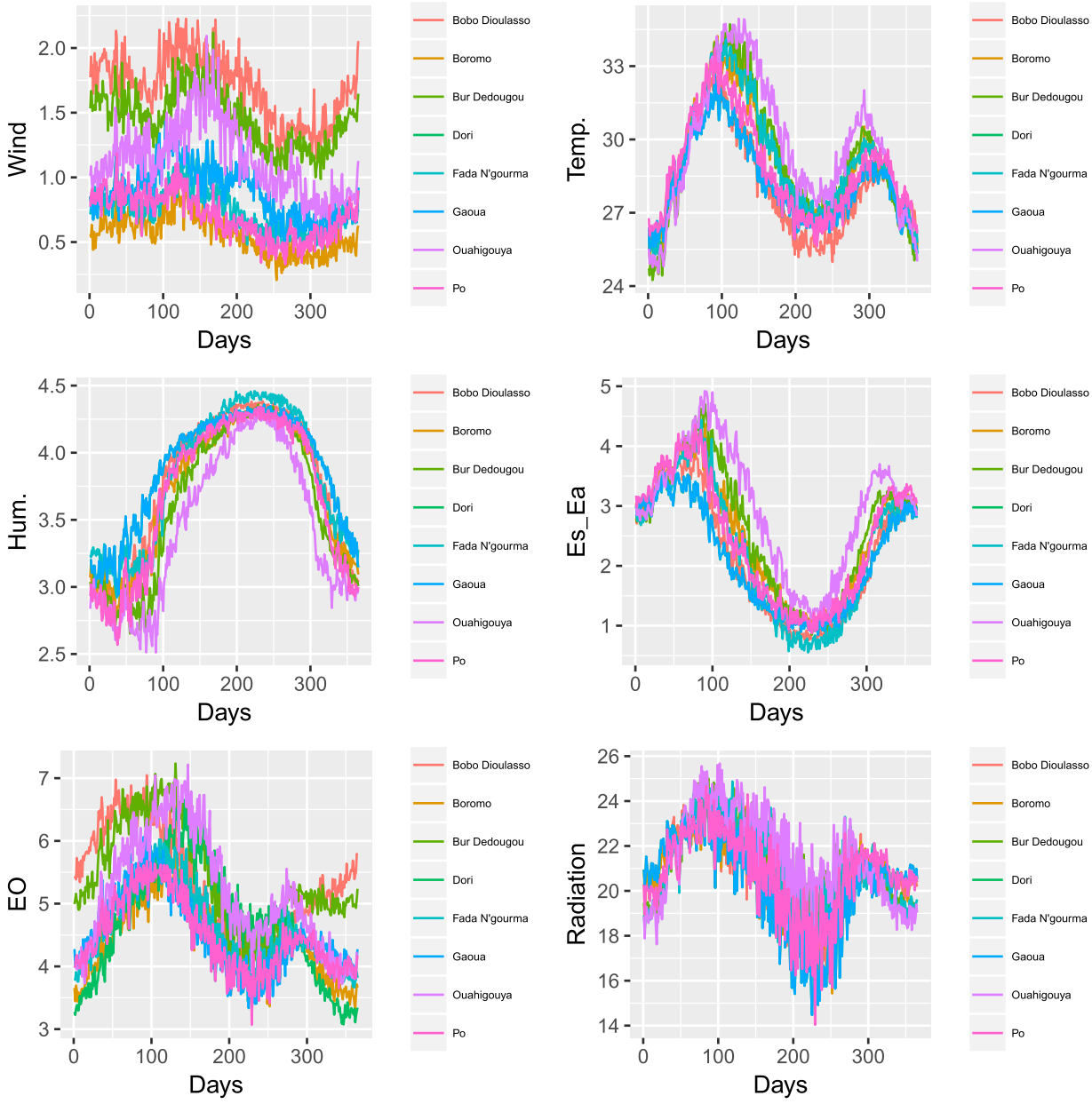


Figure 3. Time series plots of averaged datasets.

generating capacity of 247 MW. The net energy import of the country from its neighboring countries currently stands at about 20%. However, fuel-wood, charcoal, agricultural residues and animal dung are used as major sources of energy in remote villages.

In the present study, the proposed mathematical model was developed for the prediction of daily global solar radiation using eight meteorological stations distributed all over the Burkina Faso region, namely Bur Dedougou, Bobo Doulasso, Fada N'gourma, Ouahigouya, Bormo, Dori, Gaoua and Po. The daily-scale climatological data, obtained from 1 January 1998 to 31 December 2012, consist of six variables: wind power, temperature, *log* humidity, the difference between the saturation and the actual vapour pressure (Es−Ea), evaporation (Eo) and solar radiation. The datasets were averaged over the data points obtained from the whole time span to construct a functional regression model. The mean value of the climate variables in time series for each involved meteorological station are plotted in Figure 3.

4. Application and results

The current study has reported the feasibility of a newly developed mathematical model called the FDA technique

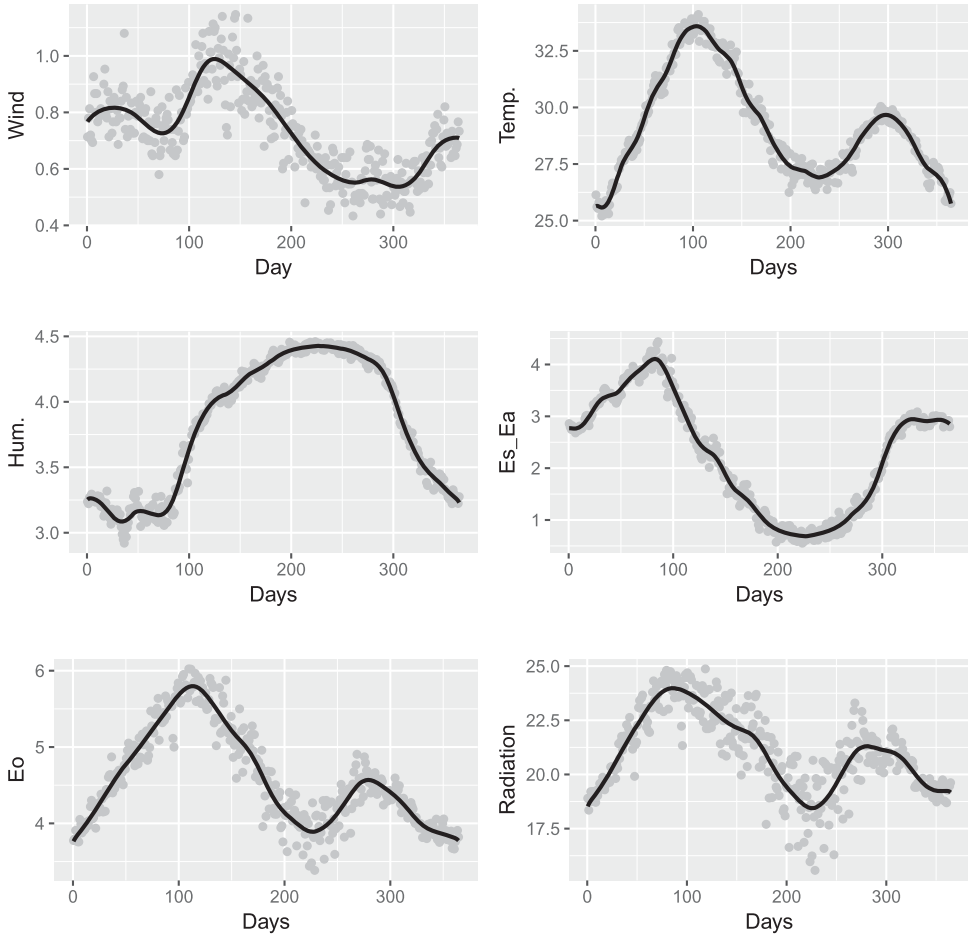


Figure 4. Functional datasets for Fada N'gourma city obtained using GBIC. Gray points represent the raw data and the solid (black) lines are the functions.

to predict daily-scale global solar radiation in the Burkina Faso region of West Africa. The global solar radiation was simulated based on various related climatological variables using a consistent timescale. At first, the datasets of all the climate variables were converted into functional form using penalized Gaussian basis function expansion taking into account the number of basis functions K and the penalty parameter λ estimated by GBIC and GCV. The modeling was conducted based on the distinguished modeling strategy cross-station paradigm simulation. Six meteorological stations were selected randomly to predict the solar radiation at two targeted meteorological stations (i.e. Ouahigouya and Dori). The main merit of this modeling archetype is the possibility of using the nearby maintained meteorological stations information as predictors for any particular station. This is highly significant and essential in the case where there is no consistency of monitoring measurements, lack of climate information over certain historical periods, and other reasons that might be experienced in such developing countries. For validation purposes, the predictability performance of the FDA technique was compared with one of the well-known regression models called panel regression.

To demonstrate the functionality of GCV and GBIC over the inspected meteorological dataset, the functions of the variables for the Fada N'gourma station (selected as an example) are illustrated using the $\hat{K}$ and $\hat{\lambda}$ values of GCV and GBIC (see Figures 4 and 5, respectively). Table 1 reports the tuning parameters of $\hat{K}$ and $\hat{\lambda}$ values in the form of a quantitative presentation. It is obvious, based on the tabulated values, that GCV performs the regression function with higher magnitudes of basis functions over those of GBIC to convert the raw data to functional form. Both functions provide a clear picture of the raw data as demonstrated in Figures 4 and 5.

The functional regression model was constructed using the variables of six randomly selected stations, i.e. Bobo Dioulasso, Boromo, Bur Dedougou, Fada

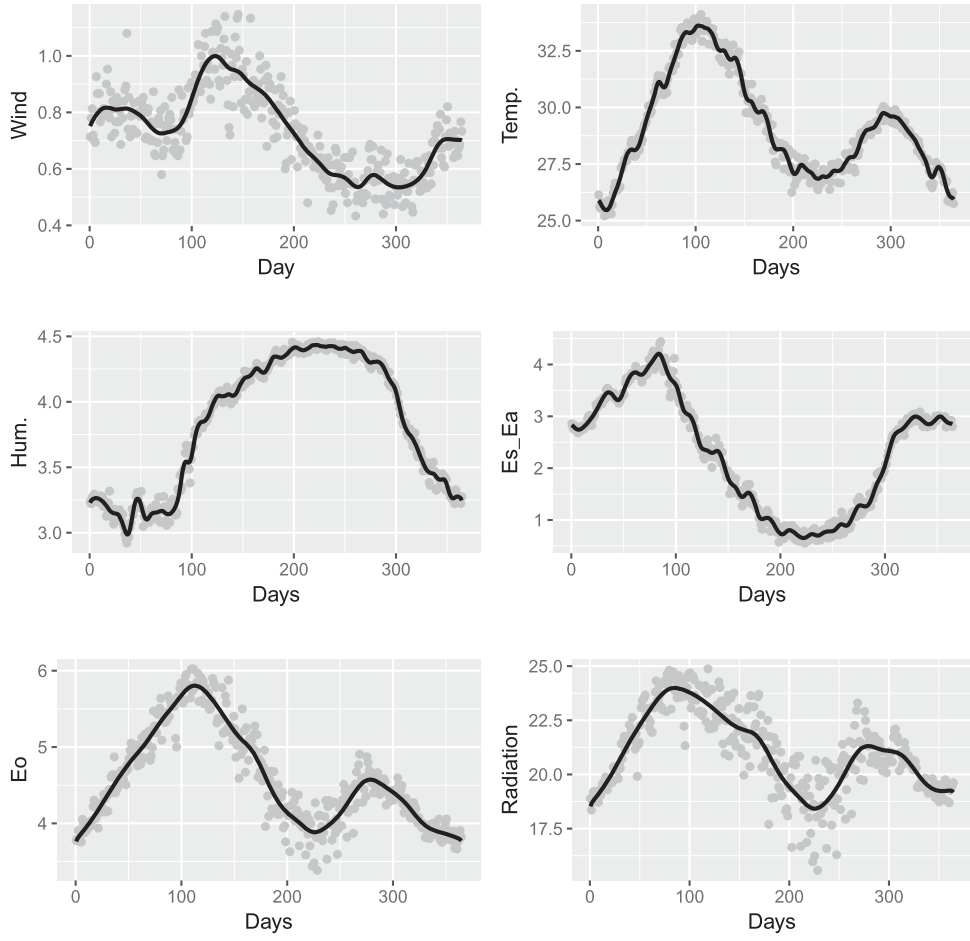


Figure 5. Functional datasets for the Fada N'gourma city obtained using GCV. Gray points represent the raw data and the solid (black) lines are the functions.

Table 1. Estimated number of basis functions and penalty parameters.

		Variable					
Method	Parameter	Wind	Temperature	Humidity	Es−Ea	Eo	Solar radiation
GBIC	$\hat{K}$	42	42	42	42	42	42
	$\hat{\lambda}$	0.7692	−1.7948	0.2564	−0.7692	−0.2564	−1.2820
GCV	$\hat{K}$	70	68	70	68	62	62
	$\hat{\lambda}$	0.7692	−2.8205	−1.2820	−1.7948	0.2564	−0.7692

N'gourma, Gaoua and Po, as follows:

$$y_n^*(t) = \int_{\mathcal{T}} x_{n1}^*(s)\beta_1(s,t) + \int_{\mathcal{T}} x_{n2}^*(s)\beta_2(s,t) + \int_{\mathcal{T}} x_{n3}^*(s)\beta_3(s,t), + \int_{\mathcal{T}} x_{n4}^*(s)\beta_4(s,t) + \int_{\mathcal{T}} x_{n5}^*(s)\beta_5(s,t) + \epsilon_n^*(t), \tag{18}$$

where y_n^*, x_{n1}^*, x_{n2}^*, x_{n3}^*, x_{n4}^* and x_{n5}^* are the centered functional variables for solar radiation, wind, temperature, log humidity, Es−Ea and Eo, respectively, and $\mathcal{T} = \{0.5, 1.5, 2.5, \ldots, 364.5\}$. The parameter matrix $\mathbf{B}$ was estimated using the penalized maximum likelihood method, and GBIC and GCV were used to select the best model.

Scatter plots of the observed average global solar radiation versus the fitted smooth function values are displayed in Figures 6 and 7. These scatter plots show that the observed solar radiation values were well fitted by the smooth functions obtained from the proposed model. On the other hand, Figure 8 presents the modeling performance of the panel regression model for the same modeled six stations. Based on the attained modeling performance of the six randomly selected meteorological stations, the model was platformed to predict the GSR at the two targeted stations (i.e. Ouahigouya and

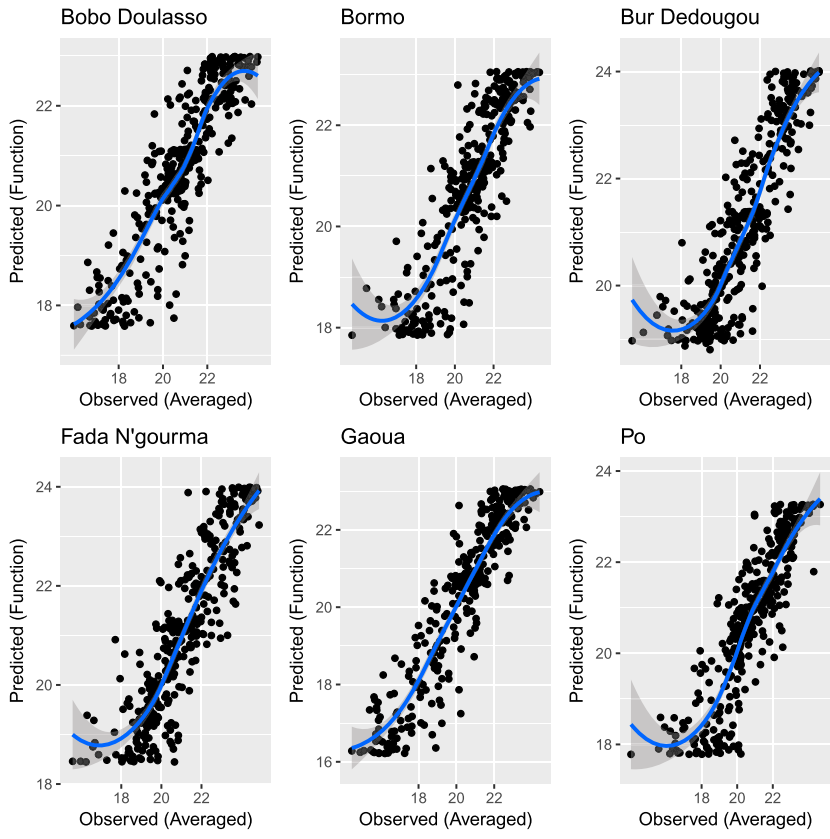


Figure 6. Scatter plots of the observed (average) and fitted (function) solar radiation values obtained from the model evaluated by GBIC.

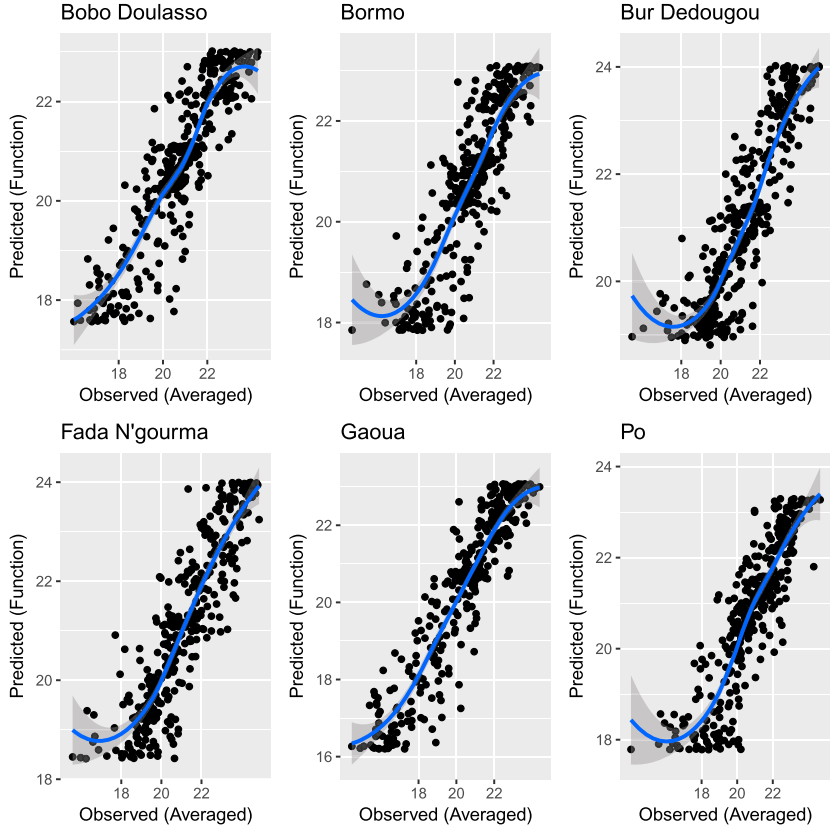


Figure 7. Scatter plots of the observed (average) and fitted (function) solar radiation values obtained from the model evaluated by GCV.

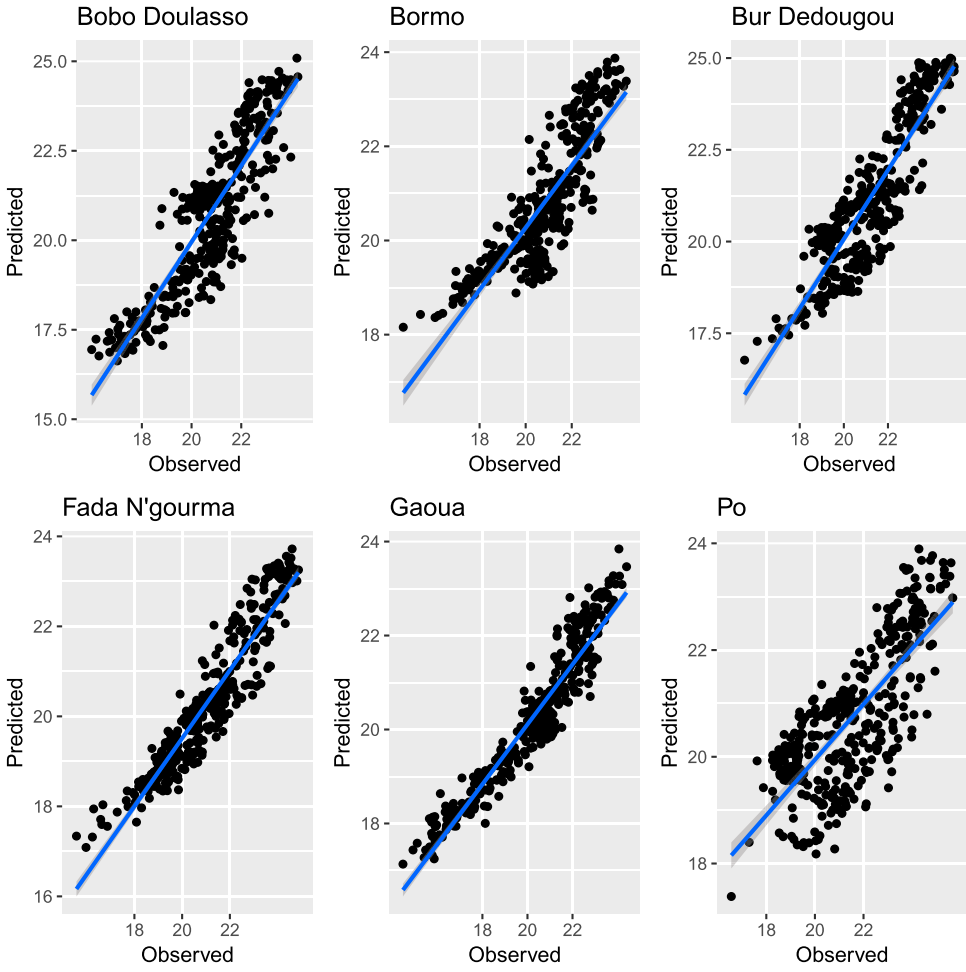


Figure 8. Scatter plots of the observed and fitted solar radiation values obtained from the panel data model.

Dori). The prediction results for the Ouahigouya and Dori stations for both the tuning parameter selection criterion and the panel model are presented in Figures 9 and 10. The figures show that the smoothed functions of the solar radiation data are well approximated by the predicted functions using the functional regression models in comparison with the panel regression model. The scatter plots of the observed and predicted solar radiation values are plotted in Figures 11 and 12.

Following several research works in the literature, the current research modeling was validated statistically using various performance metrics including root mean squared errors (RMSE), the determination coefficient (R^2) and the correlation coefficient (R) using the observed (average) solar radiation and fitted and/or predicted solar radiation functions. The mathematical formulation can be expressed as follows (Rodrigues & Henggeler Antunes, 2018; Yadav, Malik, & Chandel, 2015):

$$\text{RMSE} = J^{-1}\sum_{j=1}^{J}(y_j - \hat{y}(t_j))^2, \tag{19}$$

$$R^2 = \frac{\sum_{j=1}^{J}[(y_j - \bar{y})(\hat{y}(t_j) - \bar{\hat{y}}(t))]^2}{\sum_{j=1}^{J}(y_j - \bar{y})^2 \sum_{j=1}^{J}(\hat{y}(t_j) - \bar{\hat{y}}(t))^2}, \tag{20}$$

$$R = \frac{\sum_{j=1}^{J}(y_j - \bar{y})(\hat{y}(t_j) - \bar{\hat{y}}(t))}{\sqrt{\sum_{j=1}^{J}(y_j - \bar{y})^2}\sqrt{\sum_{j=1}^{J}(\hat{y}(t_j) - \bar{\hat{y}}(t))^2}}. \tag{21}$$

Values of all the performance metrics examined (i.e. RMSE, R^2 and R) are reported in Table 2. Note that the values given in columns three to eight belong to the stations used in the training modeling phase, whereas the values in the last two columns belong to the stations for which the performance metrics were predicted. These values indicate that the fitted/predicted functions evaluated by GCV provide slightly better approximations compared to those obtained by GBIC. However, the GCV and GBIC functions reveal a much better predictive capacity in comparison with the panel regression model.

The current research results are validated against established research in the literature and within the African region. Olatomiwa et al. (2015b) established an ANFIS method to predict monthly solar radiation at

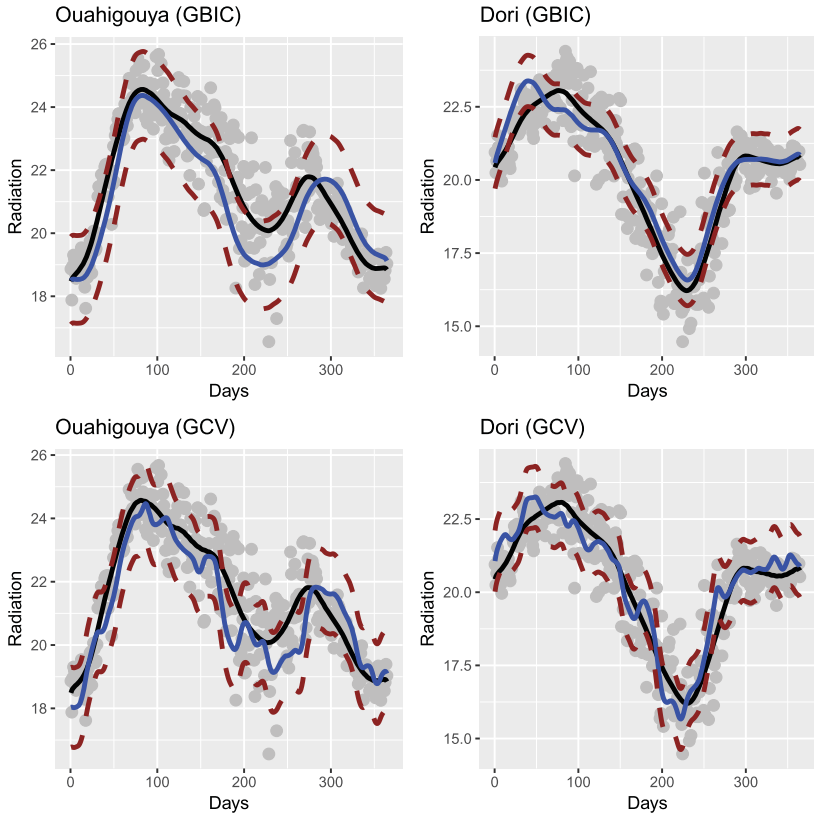


Figure 9. Results for the Ouahigouya and Dori stations. Gray points are the observed discrete solar radiation data points, black solid lines are the smoothed raw data, blue solid lines are the predicted functions obtained using the functional regression model, and the brown dashed lines are the approximate 95% confidence intervals of the predicted functions.

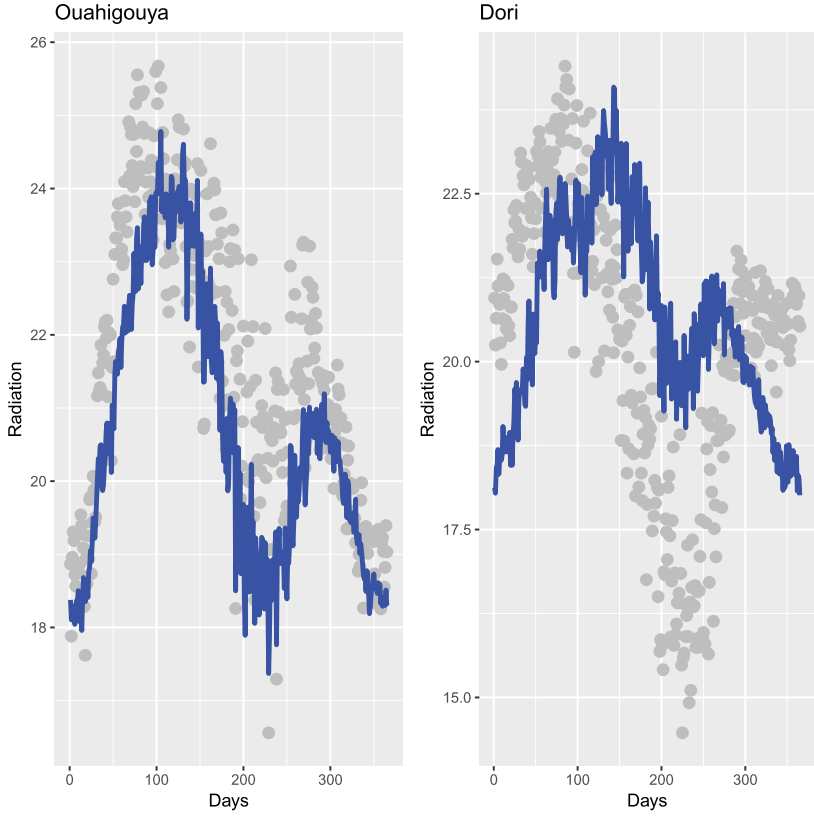


Figure 10. Results for the Ouahigouya and Dori stations. Gray points are the observed discrete solar radiation data points and blue solid lines are the predicted observations obtained using the panel data model.

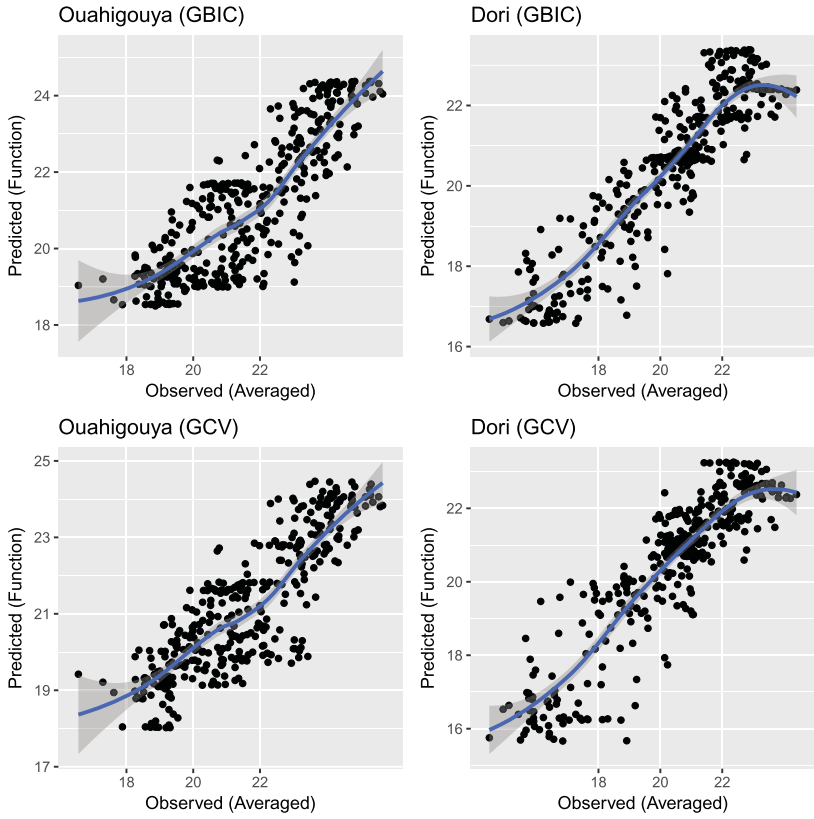


Figure 11. Scatter plots of the observed (average) and predicted (function) solar radiation values.

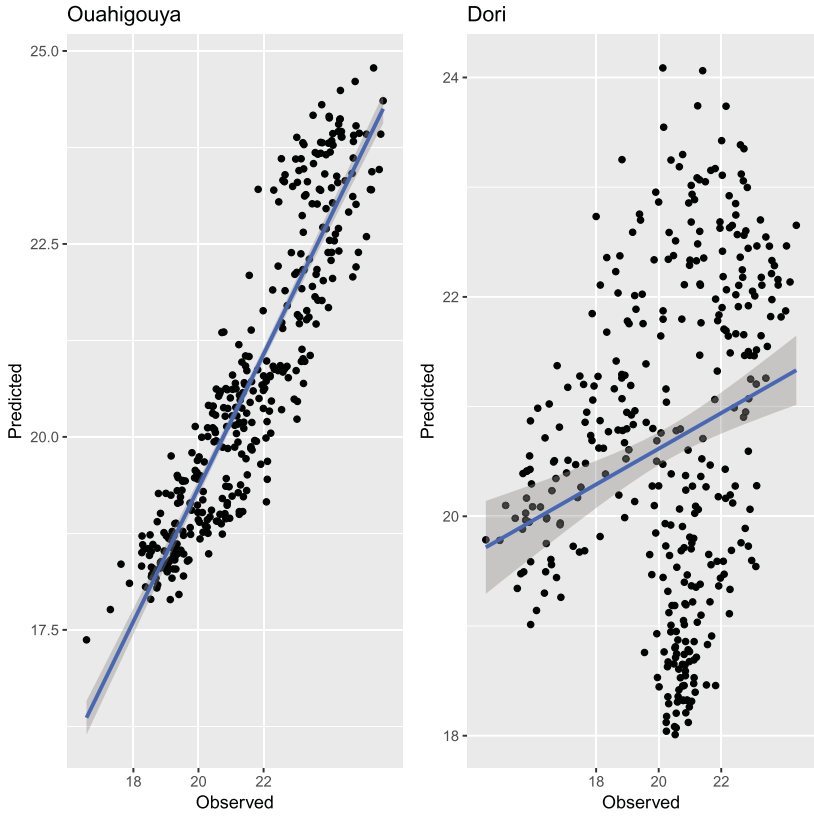


Figure 12. Scatter plots of the observed and predicted solar radiation values obtained by the panel data model.

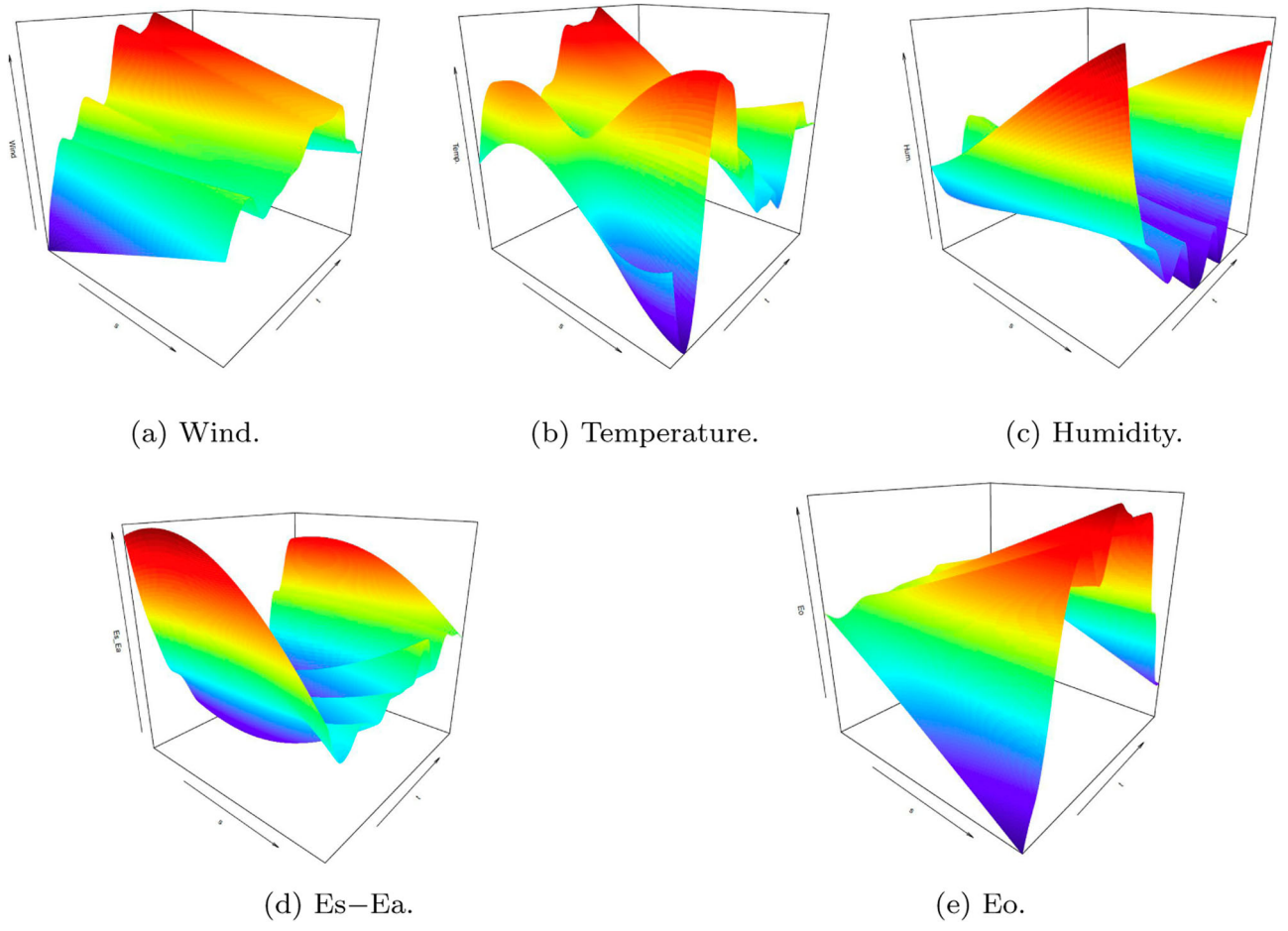
Table 2. Calculated performance metrics (RMSE, R^2 and R) at the various test stations.

Method	Performance metric	Station							
		Bobo Doulasso	Bormo	Bur Dedougou	Fada N'gourma	Gaoua	Po	Ouahigouya	Dori
GBIC	RMSE	0.8080	0.8761	0.8704	0.8827	0.8163	0.8778	1.1640	0.9460
	R^2	0.7910	0.7527	0.7799	0.7831	0.8587	0.7644	0.7065	0.8212
	R	0.8893	0.8675	0.8831	0.8849	0.9267	0.8743	0.8405	0.9062
GCV	RMSE	0.8044	0.8725	0.8674	0.8790	0.8129	0.8745	1.1008	0.9733
	R^2	0.7929	0.7547	0.7814	0.7849	0.8599	0.7644	0.7214	0.8028
	R	0.8904	0.8687	0.8840	0.8859	0.9273	0.8753	0.8494	0.8960
Panel regression	RMSE	1.0094	0.8989	0.8302	1.0398	0.8338	2.2758	1.4288	5.4919
	R^2	0.7552	0.7145	0.7886	0.8805	0.8992	0.5803	0.8268	0.0573
	R	0.8690	0.8452	0.8880	0.9383	0.9483	0.7618	0.9093	0.2395

Table 3. The statistical performance of the established research conducted by Olatomiwa et al. (2015a).

Station	Model	Performance indicator
Iseyin	SVM-FFA	RMSE = 0.493 and R^2 = 0.795
	ANN	RMSE = 0.550 and R^2 = 0.745
	GP	RMSE = 0.520 and R^2 = 0.767
Maiduguri	SVM-FFA	RMSE = 2.493 and R^2 = 0.209
	ANN	RMSE = 2.608 and R^2 = 0.133
	GP	RMSE = 2.549 and R^2 = 0.163
Maiduguri	SVM-FFA	RMSE = 2.611 and R^2 = 0.585
	ANN	RMSE = 2.979 and R^2 = 0.518
	GP	RMSE = 2.790 and R^2 = 0.623

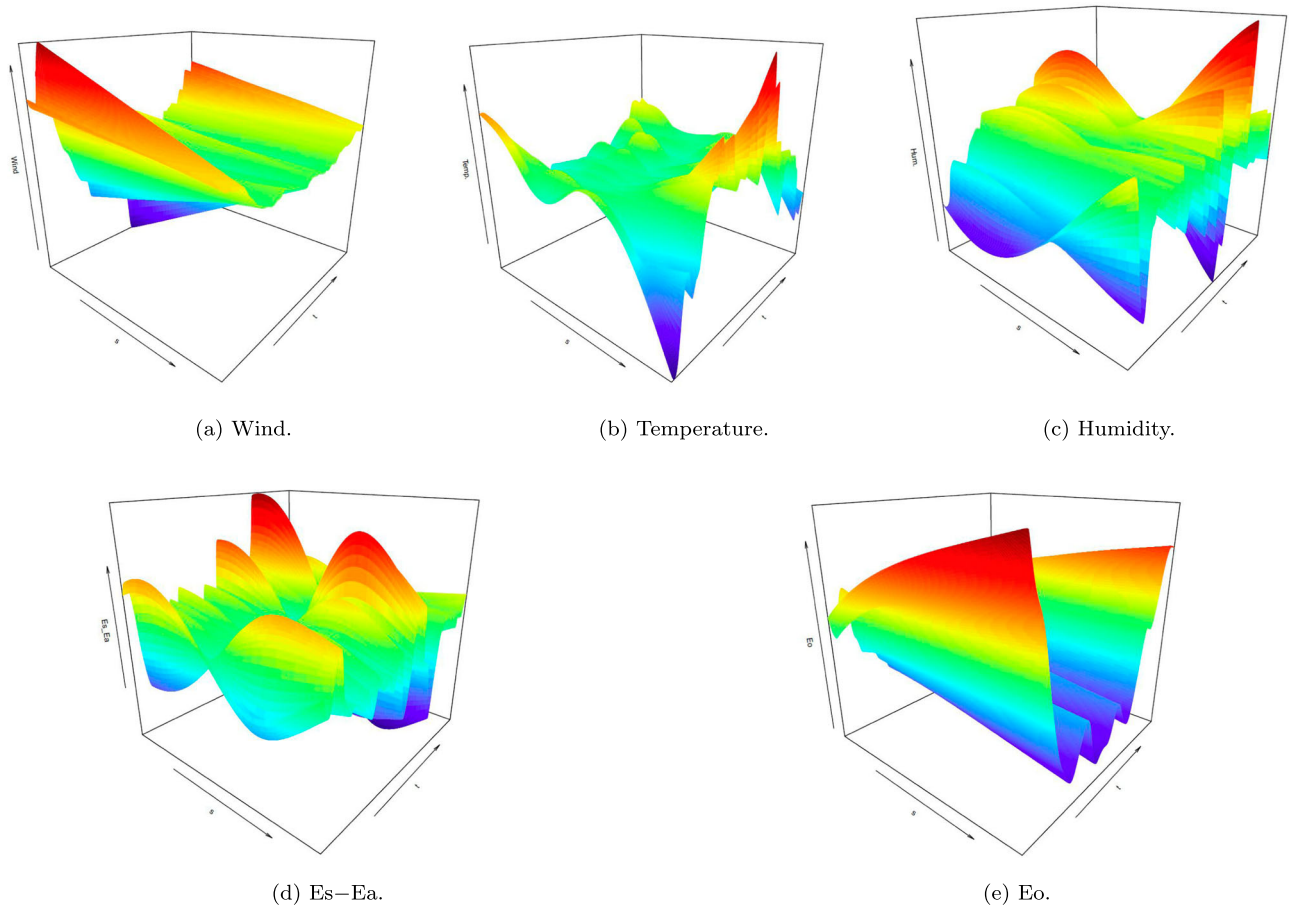
Iseyin, Nigeria. The best prediction attained using the developed ANFIS model achieved RMSE = 1.758 and R^2 = 0.6567. Recently, newly developed hybrid machine learning based on the integration of the Support Vector Machine (SVM) and the bio-inspired FireFly optimization Algorithm (FFA) and two stand-alone machine learning models – i.e. Artificial Neural Networks (ANNs) and Genetic Programming (GP) – to predict monthly mean solar radiation at three meteorological stations (i.e. Iseyin, Maiduguri and Jos) across the Nigeria region



(a) Wind. (b) Temperature. (c) Humidity.

(d) Es−Ea. (e) Eo.

Figure 13. Estimates of coefficient functions $\beta_i(s,t)$, for $i = 1,\ldots,5$, of the functional regression model given by Equation (18). Note that the coefficient functions were estimated based on GBIC.



(a) Wind. (b) Temperature. (c) Humidity.

(d) Es−Ea. (e) Eo.

Figure 14. Estimates of coefficient functions $\beta_i(s,t)$, for $i = 1,\ldots,5$, of the functional regression model given by Equation (18). Note that the coefficient functions were estimated based on GCV.

(Olatomiwa et al., 2015a). Another study was developed using the empirical formulation for diffuse solar radiation prediction by Khorasanizadeh and Mohammadi (2015). The results demonstrate six different empirical formulations with prediction accuracy achievement in the range RMSE = 0.9548–1.1698. Based on the statistical metrics performance of the prediction reported in Table 3, in comparison with the current research results, the performance metrics demonstrate superior prediction performance at the Ouahigouya and Dori stations.

The estimated coefficient functions of the functional linear model presented in Figures 13 (GBIC) and 14 (GCV). These figures show the effects of the meteorological variables on the predicted solar radiation. For example, panel (c) of Figure 13 indicates that, while the humidity has little effect on the predicted solar radiation during the early months of the year, it has a large effect in the last months of the year. Although the attained predictability performance of the functional data analysis technique on global solar radiation prediction is good, there is still room for modeling enhancement via the incorporation of the physical-based model established using Meteoblue (Fabbri et al. (2017)). Indeed, formulating such an integrative model based on functional data analysis and the mathematical formulation of the Meteoblue method could possibly enhance the prediction capacity performance further.

5. Conclusions

The development of a scientific, robust and reliable modeling strategy to predict global solar radiation in particular climatic regions could help climate change mitigation advocates and numerous energy decision-makers. This is to embrace renewable energy as a dynamic solution to mitigate the risk of the global warming and climate change phenomena. Converting global solar radiation into power grids entails an economical and intelligent model authenticated by the reliability of simulation. Hence, the exploration of newly friendly and robust mathematical models for comprehending the correlated available climate variables empowers research interest and innovation for the new era of energy engineering. The current study was devoted to exploring the

feasibility of a new mathematical model based on the functional data analysis modeling technique to simulate daily timescale global solar radiation in the Burkina Faso region of West Africa. Two different statistical modeling procedures were established (i.e. GCV and GBIC) for the prediction learning process. Fifteen years of daily-scale climate variables, including wind power, temperature, log humidity, the difference between the saturation and the actual vapour pressure (Es−Ea), evaporation (Eo) and solar radiation, were used to implement the prediction process. The findings of the current research are presented as follows.

- The conducted FDA modeling technique exhibited a reliable predictive model for GSR with a high and acceptable degree of accuracy based on the reported statistical metrics.
- Based on authentication against well-known machine learning predictive models conducted in the literature and within the same region, FDA proved to have greater prediction capacity based on RMSE and R^2.
- The predictability of the established modeling strategy was totally location dependent where the variance results can be observed. Hence, the idea of initiating a cross-station paradigm was an excellent proposition with which to gather more informative climate information from the nearby meteorological stations in order to enhance the learning procedure.
- Both of the applied learning procedures (i.e. GCV and GBIC) demonstrated an efficient computational methodology for solar radiation simulation based on various climate input variables. The merit of the results supports the possibility of embedding the model as a generalized predictive tool for the simulation of other meteorological stations.
- The investigated FDA predictive model provided a reasonable solar radiation prediction that is totally relying on the selected climate input attributes. Also, the appropriate internal parameters tuning that verified based on GCV and GBIC approaches were controlling the reliability of the modeling procedure.

Future investigations could be performed on the uncertainty analysis of data, model structure and input variability.

Acknowledgments

We should like to thank the four reviewers for their careful reading of our manuscript and their valuable suggestions and comments, which have helped us produce an improved version of our manuscript. We should also like to thank the provider of the climatological data: The National Agency of Meteorology, Burkina Faso.

Disclosure statement

No potential conflict of interest was reported by the authors .

ORCID

Zaher Mundher Yaseen http://orcid.org/0000-0003-3647-7137

References

Adeyefa Z. D., & Adedokum J. A. (1991). Pyrheliometric determination of atmospheric turbidity in the harmattan season over Ile-Ife, Nigeria. *Renewable Energy*, *1*, 555–566.

Aggarwal S. K., & Saini L. M. (2014). Solar energy prediction using linear and non-linear regularization models: A study on AMS (American Meteorological Society) 2013–14 Solar Energy Prediction Contest. *Energy*, *78*, 247–256. doi:10.1016/j.energy.2014.10.012

Aybar-Ruiz A., Jiménez-Fernandez S., Cornejo-Bueno L., Casanova-Mateo C., Sanz-Justo J., Salvador-González P., & Salcedo-Sanz S. (2016). A novel grouping genetic algorithm–extreme learning machine approach for global solar radiation prediction from numerical weather models inputs. *Solar Energy*, *132*, 129–142. doi:10.1016/j.solener.2016.03.015

Ayodele T. R., & Ogunjuyigbe A. S. O. (2015). Prediction of monthly average global solar radiation based on statistical distribution of clearness index. *Energy*, *90*, 1733–1742.

Azoumah Y., Ramde E. W., Tabsoba G., & Thiam S. (2010). Siting guidelines for concentrating solar power plants in the Sahel: Case study of Burkina Faso. *Solar Energy*, *84*, 1545–1553.

Bahrooz K., Mert C., & Kisi O. (2018). Comparison of four heuristic regression techniques in solar radiation modeling: Kriging method vs RSM, MARS and M5 model tree. *Renewable and Sustainable Energy Reviews*, *81*, 330–341.

Baltagi B. H. (2005). *Econometric analysis of panel data*. Chichester, UK: Wiley.

Besharat F., Dehghan A. A., & Faghih A. R. (2013). Empirical models for estimating global solar radiation: A review and case study. *Renewable and Sustainable Energy Reviews*, *21*, 798–821.

Chau K. W., & Muttil N. (2007). Data mining and multivariate statistical analysis for ecological system in coastal waters. *Journal of Hydroinformatics*, *9*, 305–317.

Chiou J. M., Yang Y. F., & Chen Y. T. (2016). Multivariate functional linear regression and prediction. *Journal of Multivariate Analysis*, *146*, 301–312.

Cornejo-Bueno L., Casanova-Mateo C., Sanz-Justo J., & Salcedo-Sanz S. (2019). Machine learning regressors for solar radiation estimation from satellite data. *Solar Energy*, *183*, 768–775.

Cuevas A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, *147*, 1–23.

David M., & Lauret P. (2018). Solar radiation probabilistic forecasting. In *Wind field and solar radiation characterization and forecasting – A numerical approach for complex terrain*, (p. 201–227). Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-76876-2_9.

De Souza J. L., Lyra G. B., Dos Santos C. M., Ferreira Juniour R. A., Tiba C., & Lyra G. B. (2016). Empirical models of

daily and monthly global solar irradiation using sunshine duration for Alagoas State, Northeastern Brazil. *Sustainable Energy Technologies and Assessments*, *14*, 35–45.

Fabbri K., Canuti G., & Ugolini A. (2017). A methodology to evaluate outdoor microclimate of the archaeological site and vegetation role: A case study of the Roman Villa in Russi (Italy). *Sustainable Cities and Society*, *35*, 107–133.

Ferraty F., & Vieu P. (2006). *Nonparametric functional data analysis – theory and practice*. New York: Springer. doi:10.1007/0-387-36620-2

Ghorbani M. A., Kazempour R., Chau K. W., Shamshirband S., & Ghazvinei P. T. (2017). Forecasting pan evaporation with an integrated artificial neural network quantum-behaved particle swarm optimization model: A case study in Talesh, Northern Iran. *Engineering Applications of Computational Fluid Mechanics*, *12*, 724–737.

Horvath L., & Kokoszka P. (2012). *Inference for functional data with applications*. New York: Springer.

Ivanescu A. E., Staicu A. M., Scheipl F., & Greven S. (2015). Penalized function-on-function regression. *Computational Statistics*, *30*, 539–568.

Kaplani E., Kaplani S., & Mondal S. (2018). A spatiotemporal universal model for the prediction of the global solar radiation based on Fourier series and the site altitude. *Renewable Energy*, *126*, 933–942.

Khorasanizadeh H., & Mohammadi K. (2015). Diffuse solar radiation on a horizontal surface: Reviewing and categorizing the empirical models. *Renewable and Sustainable Energy Reviews*, *52*, 1093–1096.

Khosravi A., Koury R. N. N., Machado L., & Pabon J. J. G. (2018). Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms. *Journal of Cleaner Production*, *176*, 63–75.

Konishi S., & Kitagawa G. (2008). *Information criteria and statistical modeling*. New York: Springer.

Kosmopoulos P. G., Kazadzis S., Lagouvardos K., Kotroni K., & Bais A. (2015). Solar energy prediction and verification using operational model forecasts and ground-based solar measurements. *Energy*, *93*, 1918–1930.

Li H., Bu X., Long Z., Zhao L., & Ma W. (2012). Calculating the diffuse solar radiation in regions without solar radiation measurements. *Energy*, *44*, 611–615.

Matsui H., Kawano S., & Konishi S. (2009). Regularized functional regression modeling for functional response and predictors. *Journal of Math-for-Industry*, *1*, 17–25.

Meenal R., & Selvakumar I. (2018). Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. *Renewable Energy*, *121*, 324–343.

Moazenzadeh R., Mohammadi B., Shamshirband S., & Chau K. W. (2018). Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran. *Engineering Applications of Computational Fluid Mechanics*, *12*, 584–597. doi:10.1080/19942060.2018.1482476

Muneer T., & Munawwar S. (2006). Improved accuracy models for hourly diffuse solar radiation. *Journal of Solar Energy Engineering*, *128*, 104–117. doi:10.1115/1.2148972

Olatomiwa L., Mekhilef S., Shamshirband S., Mohammadi K., Petković D., & Sudheer C. (2015a). A support vector machine–firefly algorithm-based model for global solar radiation prediction. *Solar Energy*, *115*, 632–644. doi:10.1016/j.solener.2015.03.015

Olatomiwa L., Mekhilef S., Shamshirband S., & Petković D. (2015b). Adaptive neuro-fuzzy approach for solar radiation prediction in Nigeria. *Renewable and Sustainable Energy Reviews*, *51*, 1784–1791. doi:10.1016/j.rser.2015.05.068

Qing X., & Niu Y. (2018). Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy*, *148*, 461–468.

Ramsay J. O., & Silverman B. W. (2005). *Functional data analysis*. New York: Springer.

Rodrigues E., & Henggeler Antunes C. (2018). Estimation of renewable energy and built environment-related variables using neural networks – A review. *Renewable and Sustainable Energy Reviews*, *94*, 959–988. doi:10.1016/j.rser.2018.05.060

Sunday E., Agbasi O., & Samuel N. (2016). Modelling and estimating photosynthetically active radiation from measured global solar radiation at Calabar, Nigeria. *Physical Science International Journal*, *12*, 1–12.

Sunday E., Samuel N., Agbasi O., & Sylvia J. J. (2016). Analysis of photosynthetically active radiation over six tropical ecological zones in Nigeria. *Journal of Geography, Environment and Earth Science International*, *7*, 1–15.

Torres-Barran A., Alonso A., & J. R. Dorronsoro (2019). Regression tree ensembles for wind energy and solar radiation prediction. *Neurocomputing*, *326–327*, 151–160.

Ugwuoke P., & Okeke C. (2012). Statistical assessment of average global and diffuse solar radiation on horizontal surfaces in tropical climate. *International Journal of Renewable Energy Research*, *2*, 2–6.

Ulgen K., & Hepbasli A. (2004). Solar radiation models. Part 1: A review. *Energy Sources*, *26*, 507–520.

Valderrama M. J., Ocana F. A., Aguilera A. M., & Ocana-Peinado F. M. (2010). Forecasting pollen concentration by a two-step functional model. *Biometrics*, *66*, 578–585.

Yadav A. K., Malik H., & Chandel S. S. (2015). Application of rapid miner in ANN based prediction of solar radiation for assessment of solar energy resource potential of 76 sites in Northwestern India. *Renewable and Sustainable Energy Reviews*, *52*, 1093–1106.

Yagli G. M., Yang D., & Srinivasan D. (2019). Automatic hourly solar forecasting using machine learning models. *Renewable and Sustainable Energy Reviews*, *105*, 487–498.

Yakintepe B., & Genc Y. A. (2015). Establishing new model for predicting the global solar radiation on horizontal surface. *International Journal of Hydrogen Energy*, *48*, 15278–15283.

Yang D. A. (2019). Universal benchmarking method for probabilistic solar irradiance forecasting. *Solar Energy*, *184*, 410–416.

Yao F., Muller H. G., & Wang J. L. (2005). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, *33*, 2873–2903.

Yaseen Z. M., Sulaiman S. O., Deo R. C., & Chau K. W. (2018). An enhanced extreme learning machine model for river flow forecasting: State-of-the-art, practical applications in water resource engineering area and future research direction. *Journal of Hydrology*, *569*, 387–408.

Zou L., Wang L., Li J., Lu Y., Gong W., & Niu Y. (2019). Global surface solar radiation and photovoltaic power from Coupled Model Intercomparison Project Phase 5 climate models. *Journal of Cleaner Production*, *224*, 304–324. doi:10.1016/j.jclepro.2019.03.268