An Approach to Solar Radiation Prediction Using ARX and ARMAX Models

Vinícius Leonardo Gadioli da Silva[1], Delly Oliveira Filho[1]*, Joyce Correna Carlo[2] and Patrícia Nogueira Vaz[1]

[1]Department of Agricultural Engineering, Federal University of Viçosa, Viçosa, Brazil, [2]Department of Architecture and Urbanism, Federal University of Viçosa, Viçosa, Brazil

In recent years, Brazilian meteorological networks have introduced numerous automatic stations to monitor global solar radiation at hourly intervals. Historically, large-scale climate data measurement has supported aviation and agricultural activities. The need for a good mathematical model to adequately describe a process is a great challenge, since the performance of control and simulation systems can significantly impact both system operation and/or automation and system planning. The design of control systems based on predictive models should allow for describing the dynamic behavior of the process or system under realistic conditions, as well as finding the simplest possible model to optimize the computational resources. The present work sought to predict solar radiation levels via ARX and ARMAX linear mathematical modeling. During the simulations, global horizontal radiation was defined as input, while the following parameters were outputs: extraterrestrial normal radiation, infrared horizontal radiation, extraterrestrial horizontal radiation, direct normal radiation, and diffuse horizontal radiation. It must be noted that a new simulation was performed for each variable. The use of linear modeling (ARX and ARMAX) to predict solar radiation data was efficient for extraterrestrial normal, infrared, and extraterrestrial horizontal radiation with the mean square error equal to 2.51, 1.40 and 7.15%, respectively.

Keywords: climatic data, predictors, solar radiation, energy, radiation levels

1 INTRODUCTION

According to Barnaby and Crawley (2011), large-scale climate data measurements historically supported aviation and agricultural activities. However, in architecture or engineering it has been used for only about 40 years. Up until the 1990's, climate data were mainly collected manually and then the procedures were automated. In Brazil, the availability of data from automatic stations is even more recent, and most automatic stations started acquiring measurements in 2007 (Roriz 2012).

Representativeness is the greatest difficulty related to the data collected, since meteorological data may significantly vary year after year. Thus, it is necessary to identify which data are representative

---

Abbreviations: ARX, AutoRegressive model with eXogenous inputs; ARMAX, AutoRegressive Model with Moving Average and eXogenous inputs; RMSE, Root Mean Square Error.

for the local climate (Rossi et al., 2009; Chan 2011). According to Rossi et al. (2009), this climatic representativeness can be obtained by means of statistical methods, such as the climatological normal, the identification of typical days of the project, or by specific methodologies, such as the construction of meteorological archives (Guimarães and Carlo 2015).

Although the methods provided meteorological data representative of the location, the selection of different methods tends to result in climate files with different data. In addition, according to Barnaby and Crawley (2011), meteorological parameters such as ambient temperature and wind speed may differ significantly according to the site of the measurement (Guimarães et al., 2014).

As a physical system, the atmosphere can be described by a system of mathematical equations derived from the application of Newton's second law and the development of differential calculus (18th century). However, the system of equations that determines atmospheric movements is overly complex. Thus, it cannot be solved in an exact and analytical way and demands a series of approximations (INPE 2016). With the exponential development of technology, computer simulation tools have been used in several areas of knowledge. The main objective of computational simulation is to provide users with results close to real situations by means of tools that seek to provide a solution similar to a realistic model, optimize projects and productive processes. At the same time errors must be minimized and results analyzed long before the start of the project prototyping phase (Carlo and Lamberts 2008; Hensen and Lamberts 2011).

The adjustment of the parameters of a mathematical model used to represent a system is called System Identification (G.J. Ríos-Moreno et al., 2007). According to Ljung (1999), System Identification is grounded in standard statistical techniques, and many of the basic routines have direct interpretations, for example, least squares and maximum likelihood. The control community had an important role in the development and application of these basic techniques to dynamic systems right after the birth and development of modern control theory (Ljung 1999).

A reliable forecast of incident solar radiation, both in buildings and in applications related to agriculture, photovoltaic and solar thermal energy generation, is crucial for projects of interest to these applications (Obukhov et al., 2018). The prediction of meteorological data via computational mathematical models aims to generate a set of numerical values with the same statistical characteristics of a historical series of collected data. A weather forecast is prepared for a certain period, location, or region. Forecasts are associated with a certain degree of uncertainty. These predictions are applied in several areas, including renewable energy generation systems and precision agriculture (Fruteira et al., 2011).

Therefore, according to Fruteira et al. (2011) and Ruslan et al. (2017), the simulated data must undergo a validation process for analysis of their reliability and representation of the actual climatic conditions of the location of interest. This can assure that the statistical properties contained in the historical series of each meteorological variable have been preserved.

In general, several models have been used in the literature to estimate and predict solar radiation. Thus, choosing the best model for each type of application becomes a growing challenge (Belmahdi et al., 2021; Suganthi and Samuel, 2012).

With the advancements in computing technology for data acquisition and processing, the parameters of the structural models can be updated from the responses measured under system excitation. This procedure is obtained using identification techniques as an inverse problem. The inverse problem can be defined as the determination of the internal structure of a physical system from either the measured behavior of the system or the estimation of an unknown input that gives rise to a measured output signal (Khanmirza et al., 2011).

Recent studies have used linear and non-linear algorithms intending to solve the problem of predicting solar radiation: Huang et al. (2021) used different machine learning algorithms and Implications for Extreme Climate Events; Belmahdi et al. (2021) used neural networks to predict daily global solar radiation for twenty-five Moroccan cities; Yang et al. (2021) used Kalman filter photovoltaic (PV) power prediction model based on forecasting experience.

In addition, several solar radiation prediction models have been developed over the years and are categorized into three main models: empirical or analytical models, machine learning models, which use computational intelligence techniques, and satellite remote sensing models (Yang et al., 2020). Thus, the present article sought to perform a study on the prediction of solar radiation using linear modeling (ARX, ARMAX).

2 METHODOLOGY

2.1 Meteorological Data and Climate Files

The automatic weather station collects meteorological information every minute (temperature, humidity, atmospheric pressure, precipitation, wind direction and speed, solar radiation), representative of the area in which it is located (Mellit and Pavan 2010; Zanetti et al., 2006). The set of received data is validated by quality control and stored in a database (INMET 2011). Literature analysis reveals that prediction methods are based on historical series of energy generation data. These prediction methods demand a large amount of historical data and computational effort (Yang et al., 2021).

Climate files are a set of meteorological data, usually composed of a typical representative year, expressed through several parameters, including temperature, relative air humidity, solar radiation, wind speed and direction (Guimarães and Carlo 2015). The most common climate files identified in Brazil can be found in the Test Reference Year (TRY) and the Test Meteorological Year (TMY) formats, whose statistical treatments select years and months, respectively, without hourly temperature extremes (Carlo and Lamberts 2008).

2.2 Modeling and System Identification

Control strategies based on predictive models are a class of computational algorithms based on the dynamic behavior of a

process explained by a mathematical model (Froisy 2006; Ren et al., 2006). The need for a mathematical model able to adequately describe a process is always a challenge, since the performance of control systems is significantly dependent on the modeling accuracy (Vasquez et al., 2008; Obukhov et al., 2018).

The development of an efficient mathematical model is not an ordinary issue, mainly because the criteria that make it applicable depend not only on its purpose and application (Ruslan et al., 2017). Design of control systems based on predictive models should allow for simulating the dynamic behavior of the process or system under the conditions as close to reality as possible, with the simplest possible model to optimize computational resources (Shakouri and Radmanesh 2009).

The representation theorem can be used to obtain several forms of expressing the mathematical models of a system. The linear mathematical representations most frequently used in system identification (Aguirre 2007) are presented below.

It is usual to describe a linear time-invariant system with a controlled input, U, and an uncontrolled input, i.e., noise, W. These inputs present the direct network gain, G(z) and H(z), respectively. In many cases the noise, also called disturbance, can be described by a steady-state stochastic process with rational spectral density. Thus, the Z-transform of the system output is given by the application of the overlapping theorem of linear systems and defined as the linear combination of the two inputs, **Eq. 1**:

$$Y(z) = G(z)U(z) + H(z)W(z) \qquad (1)$$

in which G(z) is the transfer function of the deterministic part of the system, and H (z) is the transfer function of the stochastic part of the system. Both are rational and stable functions.

Assuming that the transfer functions of the deterministic and stochastic parts generally have some poles in common, representation of the system output can be written as follows:

$$Y(z) = \frac{z^{-d}B(z)}{F(z)A(z)}U(z) + \frac{C(z)}{D(z)A(z)}W(z) \qquad (2)$$

where A(z), B(z), C(z), D(z), F(z) are polynomials as a function of z, whose roots of the denominators and numerators are the poles and zeros of the deterministic and stochastic parts of the system, respectively, and d is the system transport delay. Application of the inverse of the Z-transform and use of the back-shift operator provides:

$$A(q^{-1})y(k) = q^{-d}\frac{B(q^{-1})}{F(q^{-1})}u(k) + \frac{C(q^{-1})}{D(q^{-1})}w(k) \qquad (3)$$

where q^{-1} is the delay operator defined in a such a way that $y(k)q^{-1} = y(k-1)$, $w(k)$ is the white noise and $A(q^{-1})$, $B(q^{-1})$, $C(q^{-1})$, $D(q^{-1})$, $F(q^{-1})$ are polynomials defined as follows:

$$A(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a} \qquad (4)$$

$$B(q^{-1}) = b_0 + b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b} \qquad (5)$$

$$C(q^{-1}) = 1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c} \qquad (6)$$

$$D(q^{-1}) = 1 + d_1 q^{-1} + \ldots + d_{n_d} q^{-n_d} \qquad (7)$$

$$F(q^{-1}) = 1 + f_1 q^{-1} + \ldots + d_{n_f} q^{-n_f} \qquad (8)$$

The different identification models used in literature are obtained according to the particular values of the polynomials $A(q^{-1})$, $B(q^{-1})$, $C(q^{-1})$, $D(q^{-1})$, $F(q^{-1})$.

Identification of systems and processes has the following steps: data acquisition, data processing, selection of the model structure, parameter estimation and model validation. These steps will be discussed in more detail in the methodology section.

2.3 Autoregressive Model With Exogenous Inputs and Autoregressive Model With Moving Average and Exogenous Inputs Models

Several models can represent a system in diverse ways, depending on the perspective considered. Some of the models used to simulate linear systems are autoregressive, including the autoregressive model with exogenous inputs (ARX) and the autoregressive model with moving average and exogenous inputs (ARMAX), state variable and transfer functions models (Aguirre 2007; Jorgensen et al., 2011).

The present work explored the autoregressive models with exogenous inputs (ARX) and autoregressive model with moving average and exogenous inputs (ARMAX). The simplest model that can be adjusted to the data of a sample is an autoregressive model with the inclusion of exogenous variables (ARX–autoregressive with exogeneous inputs) (Piltan et al., 2017).

A time series y(t) generally follows an ARX model when it can be explained by the expression of **Eq. 9** (Shumway and Stoffer 2006; Aguirre 2007; Moura and Montini 2012):

$$y(t) + a_1 y(t-1) + a_2 y(t-2) + \ldots + a_{n_a} y(t-n_a)$$
$$= b_1 u(t-1) + b_2 u(t-2) + \ldots + b_{n_b} u(t-n_b) + e(t) \qquad (9)$$

where $e(t)$ refers to white noise.

The application of a unit delay operator $z^{-1}y(t) = y(t-1)$, in **Eqs 4,5** provides:

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_{n_a} z^{n_a} \qquad (10)$$

$$B(z) = b_1 z^{-1} + b_2 z^{-2} + \ldots + a_{n_b} z^{n_b} \qquad (11)$$

$$A(z)y(t) = B(z)u(t) + e(t) \qquad (12)$$

Thus, the input-output relationship of the model is given by:

$$y(t) = \frac{B(z)}{A(z)}u(t) + \frac{1}{A(z)}e(t) \qquad (13)$$

The ARX model can be improved with the use of a moving average applied to the disturbance. Thus, we obtain:

$$y(t) + a_1 y(t-1) + a_2 y(t-2) + \ldots + a_{n_a} y(t-n_a)$$
$$= b_1 u(t-1) + b_2 u(t-2) + \ldots + b_{n_b} u(t-n_b) + e(t)$$
$$+ c_1 e(t-1) + c_2 e(t-2) + \ldots + c_{n_b} e(t-n_c) \qquad (14)$$

Application of the unit delay in **Eqs 4–6**, similar to the previous case, yields:

$$A(z) = 1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_{n_a} z^{-n_a} \qquad (15)$$
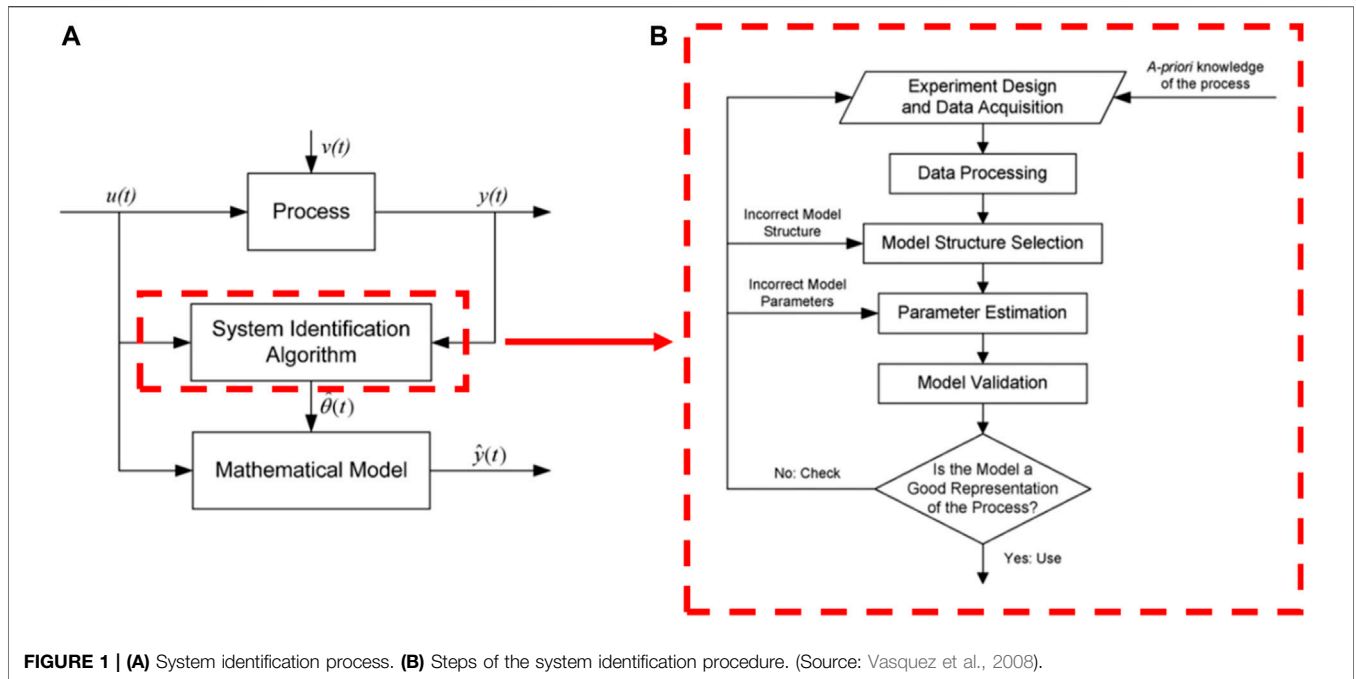
FIGURE 1 | (A) System identification process. **(B)** Steps of the system identification procedure. (Source: Vasquez et al., 2008).

$$B(z) = b_1 z^{-1} + b_2 z^{-2} + \ldots + a_{n_b} z^{-n_b} \quad (16)$$

$$C(z) = 1 + c_1 z^{-1} + c_2 z^{-2} + \ldots + c_{n_c} z^{-n_c} \quad (17)$$

where, n_a, n_b and n_c refer to the orders of the models.

Thus, the input-output ratio of the ARMAX (autoregressive–moving-average model with exogenous inputs model) is given by:

$$A(z)y(t) = B(z)u(t) + C(z)e(t) \quad (18)$$

$$y(t) = \frac{B(z)}{A(z)}u(t) + \frac{C(z)}{A(z)}e(t) \quad (19)$$

The present work used an experimental approach to determine systems for achieving a mathematical model that reproduces the dynamic characteristics of the process under study. It was based on observed variables such as: output signal or controlled variable y(t), the input signal x(t), and the disturbances e(t) (Ljung 1999). **Figure 1** shows a general outline of the process used for system identification (Vasquez et al., 2008).

In **Figure 1A**:

$u(t)$: System input;
$y(t)$: System output;
$v(t)$: Error signal;
$\hat{\theta}(t)$: Coefficients generated by the prediction model; and
$\hat{y}(t)$: Output generated by the prediction model.

(1) The process for system identification followed the steps below:
(2) Acquisition and modeling of meteorological data: the system must receive external excitation via the application of different input signals and record the evolution of its input and output signals during a predetermined fixed time interval.

(3) Pre-processing of the acquired meteorological data: the acquired data are usually accompanied by unwanted noises and other types of imperfections. Therefore, they must be treated before the start of the identification process.

(4) Definition of the model to be used: it is desirable to obtain a parametric model conditioned to the nature of the data used during the modeling process. Thus, the first step was to determine a structure suitable for the model. For such, it is necessary to acquire previous knowledge of the dynamic behavior of the process under study.

(5) The model parameters were then estimated (specify the names of the parameter here): to allow for determination of the parameter values of the structure that best adapted to the model response for the input and output experimental data.

(6) Then the model was validated, which is the final step. This step sought to determine if the obtained model satisfied the application with the accuracy required for the process. Otherwise, if the obtained model is considered invalid, the following aspects should be analyzed as probable causes:

(i) The input and output data acquired do not provide sufficient information about the dynamics of the system; and/or

(ii) The structure selected was unable to adequately describe the model; and/or

(iii) The criteria for determining the parameters were not properly adjusted.

Thus, depending on the reason an invalid template was obtained, the identification process should be repeated. The system identification process is an iterative process whose steps are illustrated in the flowchart shown in **Figure 1B**.

2.4 Modeling of the Solar Radiation Database Based on Systems Identification

Initially, the average input and output data were removed in order to normalize the database used. The data were then separated, so that one part was used for simulation and the other for model validation. The percentages used for both simulation and validation were defined according to the size of the database available for the work.

The parameters of the ARX and ARMAX models were simulated using different values of polynomial orders and delays. These values determine the degree of the ordinary differential equation computationally solved for determination of the coefficients of each model. This means that for each value of n_a, n_b and n_c, a simulation and validation model was created. In the present work the order of the model (n_a, n_b and n_c) ranged from 1 to 7, while the delay order (n_k) ranged from 1 to 3. These values are justified by the need for optimizing the use of computational resources, since it is necessary to expand them as the order of the models increases.

The model can be selected by means of the minimal Root Mean Square Error–RMSE. It is a criterion for selecting an appropriate estimator, which means selecting the order of the model to be adopted. The RMSE is the square root of the differences between the estimated value and the actual value of the squared data, weighted by the number of terms and by the estimated value, given by:

$$RMSE(\%) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y(t) - y_s(t)/y(t)\right)^2} \qquad (20)$$

Where: N = number of data; $y(t)$ = original value of each datum; and $y_s(t)$ = simulated value

This work also used F_{IT}, an estimator which according to Mustafaraj et al. (2010) and Rachad et al. (2015), is solely the variation of the output generated by the model. In other words, it measures how well the output "fits" the data used in validation of the model. The F_{IT} is calculated by the equation below:

$$F_{IT} = 1 - \frac{\sqrt{\sum_{t=1}^{N}\left[y(t) - y_s(t)\right]^2}}{\sqrt{\sum_{t=1}^{N}\left[y_s(t) - \frac{1}{N}\sum_{t=1}^{N}y(t)\right]^2}} \qquad (21)$$

In statistical modeling, these parameters are used to determine to what extent the model fits the data, or if the removal of some terms could simplify and benefit the model. That is, among other things, it can be used to help determine the variables of interest to be used in the work. It also provides a mechanism for selection of the best estimators (Belmahdi et al., 2021).

In this case, selection depends on the RMSE and F_{IT} in the ARX and ARMAX models and the degree of each model.

Initially, a treated meteorological database provided by the Laboratory of Technologies in Environmental Comfort and Energy Efficiency of the Department of the Architecture and Urban Planning, Federal University of Viçosa, was used for modeling and data processing. It is important to use a treated meteorological database, since the raw data presented several uncertainties according to the scenario of the measurement and the selection and assembling approach. Gross data tends to present discrepancies, such as measurement errors due to actions outside the meteorological station, sensor failures in data acquisition and equipment defects.

Hourly data was used from 2005 to 2015 including: air temperature (T_{air}), dew point temperature (T_{dp}), relative humidity (H_r), atmospheric pressure (P_{atm}), humidity ratio (h_r), air enthalpy (E_{air}), specific air mass (ρ), global horizontal radiation, infrared horizontal radiation, extraterrestrial normal radiation, extraterrestrial horizontal radiation, direct normal radiation, diffuse horizontal radiation, global horizontal illuminance, direct normal illuminance, diffuse horizontal illumination, wind direction, wind speed and precipitation.

3 RESULTS AND DISCUSSION

In the present work the variables of interest were the radiation compositions, given their importance both for agriculture and the generation of photovoltaic and solar thermal energy. **Table 1** shows a sample of the database for a typical day in January 2015. It should be noted that there is a range of variables that can be analyzed.

Table 1-Sample of hourly weather data for a typical day in January 2015. Continuation.

Hourly data from 2005 to 2015 were used, containing: Air Temperature (T_{air}), Dew Point Temperature (T_{dp}), Relative Air Humidity (H_r), Atmospheric Pressure (P_{atm}), Enthalpy, Air Density, Global Horizontal Radiation, Horizontal Infrared Radiation, Normal Extraterrestrial Radiation, Horizontal Extraterrestrial Radiation, Direct Normal Radiation, Horizontal Diffuse Radiation, Global Horizontal Illuminance, Direct Normal Illuminance, Horizontal Diffuse Illuminance, Wind Direction, Wind Speed and Precipitation.

Hourly data from all the years were grouped into a single file for simulation. From this, half of the data were used to create the models and the other half for validation. During the simulations, global horizontal radiation was defined as the input and the others as outputs. It must be noted that a new simulation was performed for each variable, **Table 2**.

In each simulation, a sample of 64,992 values was used, of which 32,496 were used to create the model and 32,496 for validation. In validation, the data generated from the models were compared with the existing data by the RMSE analysis and a performance coefficient (F_{IT}).

In creation of the ARX and ARMAX models, the order of the models (n_a, n_b and n_c) ranged from 1 to 7, while the delay order (n_K) ranged from 1 to 3. Therefore, the models were simulated and the RMSE and F_{IT} values for each model were obtained. Results related to the outputs with the best and worst RMSE were presented for discussion: infrared horizontal radiation and direct normal radiation, according to **Tables 3–6**.

TABLE 1 | Sample of hourly weather data for a typical day in January 2015. (Source: Guimarães et al., 2014).

Hour	T_{air} (°C)	T_{dp} (°C)	H_r (%)	P_{atm} (Pa)	h_r (kg$_{H20}$/kg$_{dried\ air}$)	E_{air} (kj/kg$_{dried\ air}$)	ρ (kg/m³)
09:00	27.4	19.4	62	93.610	0.01542	66.90631	1.05883
10:00	28.6	18.9	56	93.560	0.01493	66.90469	1.05486
11:00	30.3	19.9	54	93.500	0.01591	71.17780	1.04666
12:00	31.1	19.2	49	93.420	0.01511	69.94738	1.04433
13:00	32.2	19.4	47	93.300	0.01545	71.96627	1.03867
14:00	32.2	17.6	42	93.260	0.01378	67.67861	1.04096
15:00	31.7	19.2	47	93.240	0.01502	70.34427	1.04041

Hour	Extraterrestrial Hourly Radiation (Wh/m²)	Normal Extraterrestrial Radiation (Hourly) (Wh/m²)	Infrared Horizontal Hourly Radiation (Wh/m²)	Global Horizontal Radiation (Hourly) (Wh/m²)	Direct Normal Radiation (Hourly) (Wh/m²)	Diffuse Hourly Radiation (Wh/m²)
09:00	1257	1119	414	860	541	252
10:00	1374	1337	420	984	719	246
11:00	1411	1409	430	1045	811	234
12:00	1364	1318	434	1031	785	219
13:00	1238	1085	441	960	672	194
14:00	1041	767	438	828	490	164
15:00	786	437	437	492	166	193

TABLE 2 | Inputs and outputs used in the linear modeling.

Input x(t)	Output y(t)
Global horizontal radiation	Extraterrestrial normal radiation
Global horizontal radiation	Infrared horizontal radiation
Global horizontal radiation	Extraterrestrial horizontal radiation
Global horizontal radiation	Direct normal radiation
Global horizontal radiation	Diffuse horizontal radiation

Each model generates 32,496 hourly radiation values, but graphs with 100-h samples were presented in order to facilitate comparison of the original data with that generated by the models. In order to provide a notion of the dimension of the data, **Figure 2** shows a graph of 1000 h with all the data.

Work involving modeling and identification of systems generally uses analysis of errors to evaluate the quality of the predictors. In the paper published by Mateo et al. (2013), it is suggested to use the absolute mean error as a parameter for the comparison between linear and non-linear predictors. This error measures the average magnitude of the errors in a set of predictions, without considering its direction. Santos et al. (2007) used the comparison between prediction methods based on linear models with an exchange rate approach and used the RMSE as one of the criteria for evaluating the predictors.

Typically, the models of auto regressive prediction try to minimize the prediction errors of a function. The outputs are calculated considering that the system changes slowly over

TABLE 3 | RMSE and F_{IT} values for horizontal infrared horizontal radiation model ARX.

RMSE–ARX–Infra. Rad. (Wh/m²)–Without Zero Suppression										F_{IT}–ARX–Infra. Rad.–Without Zero Suppression								
Model Order			k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	Model Order			k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7
n_a = k	n_b = k	n_c = 1	6.25	5.37	5.24	5.18	5.14	5.11	5.09	n_a = k	n_b = k	n_c = 1	0.756	0.790	0.795	0.798	0.799	0.800	0.801
n_a = k	n_b = k	n_c = 2	6.63	5.31	5.23	5.20	5.16	5.14	5.12	n_a = k	n_b = k	n_c = 2	0.741	0.792	0.796	0.797	0.798	0.799	0.800
n_a = k	n_b = k	n_c = 3	6.53	5.31	5.22	5.20	5.18	5.15	5.13	n_a = k	n_b = k	n_c = 3	0.745	0.793	0.796	0.797	0.798	0.799	0.799
n_a = k	n_b = k	n_c = 1	1.69	1.45	1.42	1.40	1.39	1.38	1.38	n_a = k	n_b = k	n_c = 1	75.57	78.99	79.53	79.75	79.91	80.03	80.11
n_a = k	n_b = k	n_c = 2	1.79	1.44	1.41	1.40	1.39	1.39	1.38	n_a = k	n_b = k	n_c = 2	74.10	79.25	79.56	79.70	79.84	79.92	80.00
n_a = k	n_b = k	n_c = 3	1.77	1.43	1.41	1.40	1.40	1.39	1.39	n_a = k	n_b = k	n_c = 3	74.47	79.26	79.59	79.69	79.77	79.86	79.93

TABLE 4 | RMSE and F_{IT} values for horizontal infrared horizontal radiation model ARMAX.

RMSE - ARMAX - Infrared Rad. (Wh/m²) - Without Zero Suppression												F_{IT} - ARMAX - Infrared Rad. - Without Zero Suppression									
Model Order				k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7	Model Order				k = 1	k = 2	k = 3	k = 4	k = 5	k = 6	k = 7
n_a = k	n_b = k	n_c = k	n_k = 1	5.91	5.10	4.87	4.86	4.69	4.72	4.60	n_a = k	n_b = k	n_c = k	n_k = 1	0.769	0.801	0.810	0.810	0.817	0.815	0.820
n_a = k	n_b = k	n_c = k	n_k = 2	5.95	5.12	4.90	4.79	4.61	4.64	4.49	n_a = k	n_b = k	n_c = k	n_k = 2	0.768	0.800	0.808	0.813	0.820	0.819	0.825
n_a = k	n_b = k	n_c = k	n_k = 3	5.84	5.15	5.13	4.83	4.61	4.69	4.50	n_a = k	n_b = k	n_c = k	n_k = 3	0.772	0.799	0.799	0.811	0.820	0.817	0.824
n_a = k	n_b = k	n_c = k	n_k = 1	1.60	1.38	1.32	1.31	1.27	1.28	1.24	n_a = k	n_b = k	n_c = k	n_k = 1	76.90	80.07	80.98	80.99	81.67	81.54	82.01
n_a = k	n_b = k	n_c = k	n_k = 2	1.61	1.38	1.32	1.29	1.25	1.25	1.21	n_a = k	n_b = k	n_c = k	n_k = 2	76.76	79.98	80.85	81.28	81.98	81.88	82.47
n_a = k	n_b = k	n_c = k	n_k = 3	1.58	1.39	1.39	1.31	1.25	1.27	1.22	n_a = k	n_b = k	n_c = k	n_k = 3	77.16	79.88	79.95	81.11	81.97	81.65	82.40

TABLE 5 | RMSE and F_{IT} values for direct normal radiation model ARX.

| RMSE - ARX - Nor. Direct Rad. (Wh/m²) - Without Zero Suppression | | | | | | | | | | | F_{IT} - ARX - Nor. Direct Rad. - Without Zero Suppression | | | | | | | | | |
Model Order			$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$		Model Order			$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
$n_a=k$	$n_b=k$	$n_c=1$	78.33	75.36	73.81	73.37	73.19	73.02	72.72		$n_a=k$	$n_b=k$	$n_c=1$	0.440	0.461	0.472	0.475	0.476	0.478	0.480
$n_a=k$	$n_b=k$	$n_c=2$	81.83	75.57	74.87	74.85	74.66	74.29	73.99		$n_a=k$	$n_b=k$	$n_c=2$	0.415	0.459	0.464	0.465	0.466	0.469	0.471
$n_a=k$	$n_b=k$	$n_c=3$	81.70	79.50	76.64	76.56	76.21	75.87	75.75		$n_a=k$	$n_b=k$	$n_c=3$	0.415	0.431	0.452	0.452	0.455	0.457	0.458
$n_a=k$	$n_b=k$	$n_c=1$	126.34	121.55	119.04	118.3341	118.05	117.77	117.29		$n_a=k$	$n_b=k$	$n_c=1$	43.96	46.08	47.20	47.51	47.64	47.76	47.97
$n_a=k$	$n_b=k$	$n_c=2$	131.98	121.89	120.76	120.72	120.41	119.82	119.33		$n_a=k$	$n_b=k$	$n_c=2$	41.46	45.94	46.43	46.45	46.59	46.85	47.07
$n_a=k$	$n_b=k$	$n_c=3$	131.77	128.23	123.61	123.48	122.91	122.37	122.18		$n_a=k$	$n_b=k$	$n_c=3$	41.55	43.12	45.17	45.23	45.48	45.72	45.80
$n_a=k$	$n_b=k$	$n_c=1$	149.83	143.08	139.34	137.38	136.82	136.36	135.67		$n_a=k$	$n_b=k$	$n_c=1$	0.441	0.466	0.480	0.487	0.489	0.491	0.494
$n_a=k$	$n_b=k$	$n_c=2$	161.97	144.76	141.91	141.42	141.08	140.29	139.62		$n_a=k$	$n_b=k$	$n_c=2$	0.396	0.460	0.470	0.472	0.474	0.477	0.479
$n_a=k$	$n_b=k$	$n_c=3$	160.80	152.60	146.49	146.09	145.52	144.88	144.37		$n_a=k$	$n_b=k$	$n_c=3$	0.400	0.431	0.453	0.455	0.457	0.459	0.461
$n_a=k$	$n_b=k$	$n_c=1$	68.10	65.03	63.33	62.45	62.19	61.98	61.67		$n_a=k$	$n_b=k$	$n_c=1$	44.07	46.60	48.00	48.73	48.94	49.11	49.37
$n_a=k$	$n_b=k$	$n_c=2$	73.62	65.80	64.50	64.28	64.13	63.77	63.46		$n_a=k$	$n_b=k$	$n_c=2$	39.55	45.98	47.05	47.23	47.35	47.65	47.90
$n_a=k$	$n_b=k$	$n_c=3$	73.09	69.37	66.59	66.41	66.15	65.86	65.62		$n_a=k$	$n_b=k$	$n_c=3$	40.00	43.06	45.33	45.48	45.70	45.94	46.13

TABLE 6 | RMSE and F_{IT} values for direct normal radiation model ARMAX.

| RMSE - ARMAX - Nor. Direct Rad. (Wh/m²) – Without Zero Suppression | | | | | | | | | | | | F_{IT} - ARMAX - Nor. Direct Rad. - Without Zero Suppression | | | | | | | | | | |
Model Order				$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$		Model Order				$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=1$	76.55	73.47	72.32	72.14	70.84	70.83	70.55		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=1$	0.452	0.474	0.483	0.484	0.493	0.493	0.495
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=2$	79.45	75.08	73.55	73.00	72.20	71.34	69.48		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=2$	0.432	0.463	0.474	0.478	0.483	0.490	0.503
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=3$	81.20	76.45	75.94	75.36	74.20	70.93	71.51		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=3$	0.419	0.453	0.457	0.461	0.469	0.492	0.488
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=1$	123.47	118.50	116.65	116.36	114.26	114.25	113.79		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=1$	45.23	47.44	48.26	48.38	49.31	49.32	49.53
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=2$	128.14	121.09	118.64	117.74	116.46	115.06	112.06		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=2$	43.16	46.29	47.38	47.77	48.34	48.96	50.29
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=3$	130.96	123.31	122.48	121.54	119.68	114.41	115.34		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=3$	41.91	45.30	45.67	46.09	46.91	49.25	48.84
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=1$	145.90	137.51	134.92	134.16	130.75	129.84	130.16		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=1$	0.455	0.487	0.497	0.499	0.512	0.516	0.514
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=2$	155.54	142.77	138.99	138.72	134.82	134.85	135.73		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=2$	0.419	0.467	0.481	0.482	0.497	0.497	0.494
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=3$	160.33	146.76	143.46	143.36	142.80	132.56	128.91		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=3$	0.402	0.452	0.465	0.465	0.467	0.506	0.519
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=1$	66.32	62.50	61.33	60.98	59.43	59.02	59.16		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=1$	45.54	48.68	49.65	49.94	51.22	51.56	51.44
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=2$	70.70	64.89	63.18	63.05	61.28	61.29	61.69		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=2$	41.95	46.73	48.14	48.24	49.69	49.68	49.37
$n_a=k$	$n_b=k$	$n_c=k$	$n_k=3$	72.88	66.71	65.21	65.16	64.91	60.26	58.60		$n_a=k$	$n_b=k$	$n_c=k$	$n_k=3$	40.17	45.24	46.47	46.51	46.73	50.56	51.93

time, by a set of parameters that are estimated via system identification (Huang and Jane 2009).

Comparing the ARX and ARMAX models and graphs applied to the infrared horizontal radiation, it is concluded that in this case it becomes more feasible to use the ARX model, since it requires fewer computational resources than the ARMAX model.

Because zero-suppression infrared horizontal radiation data is all greater than zero, zero suppression was not used in this analysis. As can be seen in **Tables 3**, **4**, the percent difference between the RMSE and F_{IT} values in the ARX and ARMAX models is not very significant. Both maintain the error below 2%

and F_{IT} around 80%; and errors tend also to stabilize for $k = 3$. In **Figures 3** and **4** it is possible to graphically observe the behavior of mathematically generated data with respect to the data used for validation.

When analyzing **Table 5**, it can be observed that for the normal direct radiation, in the ARX model without zero suppression the error does not tend to decrease by increasing
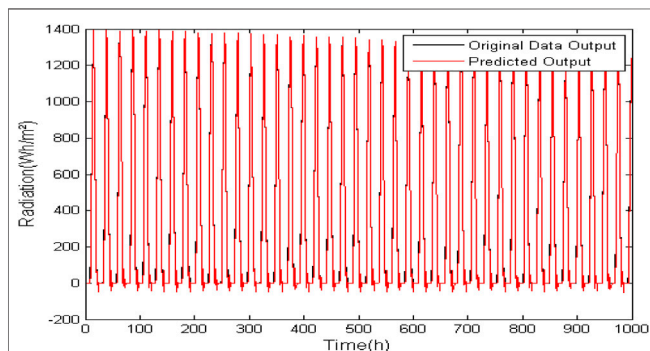

FIGURE 2 | Model ARX (na = 4, nb = 4, nc = 3) for prediction of normal extraterrestrial radiation data without zero suppression.
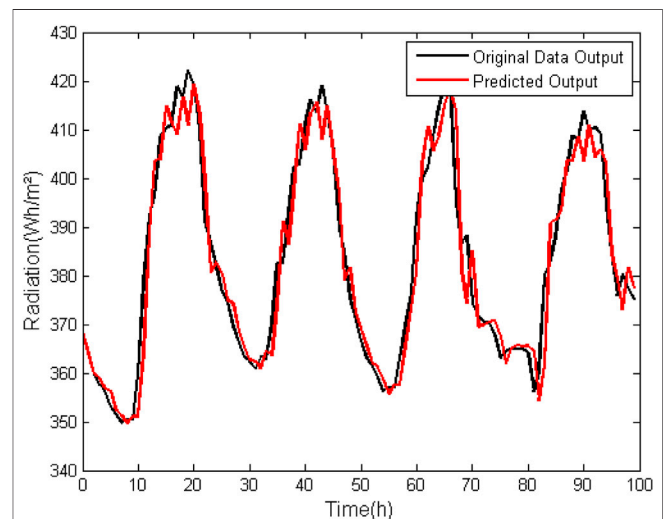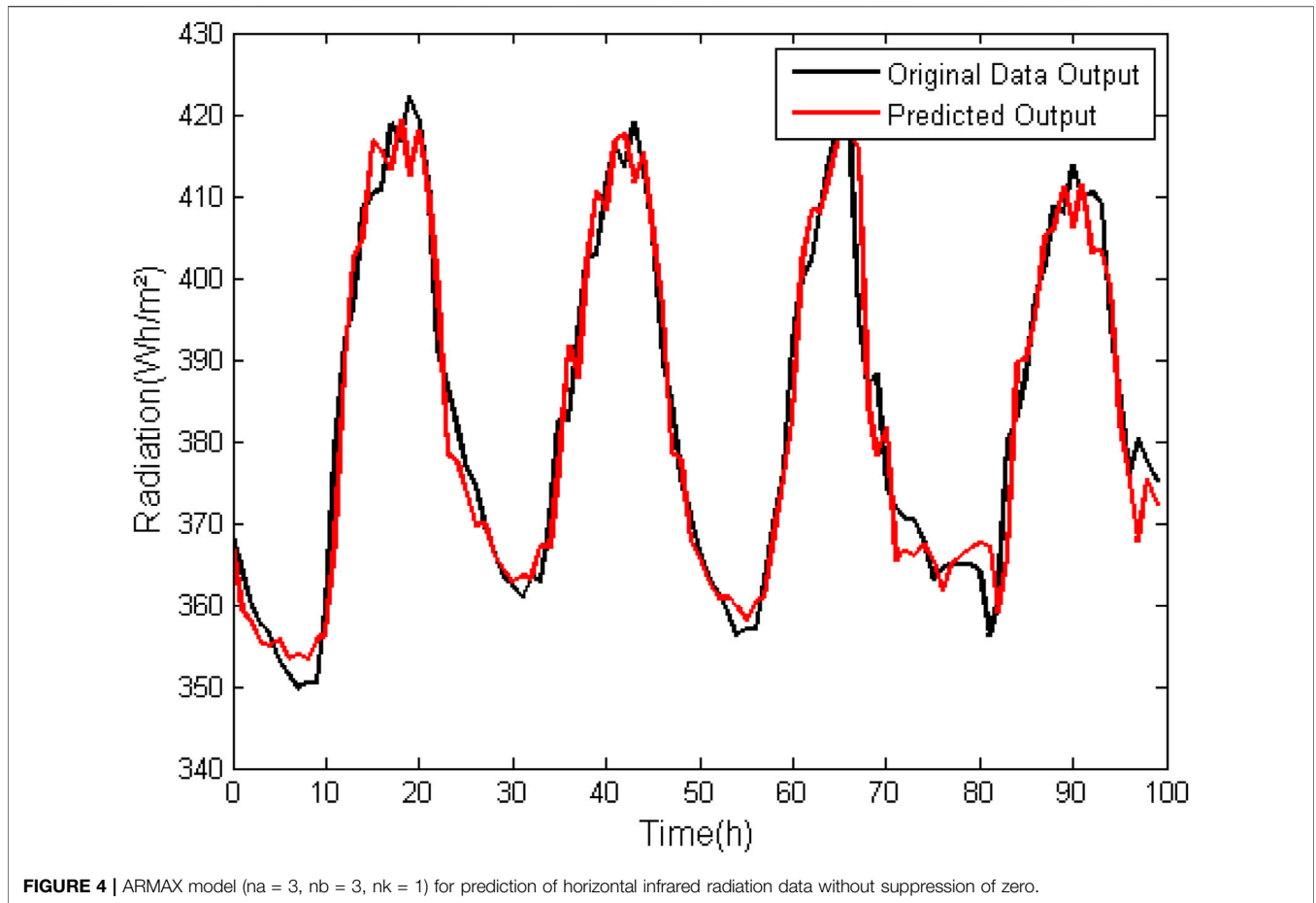

FIGURE 3 | ARX model (na = 3, nb = 3, nk = 1) for prediction of horizontal infrared horizontal radiation data without suppression of zero.

FIGURE 4 | ARMAX model (na = 3, nb = 3, nk = 1) for prediction of horizontal infrared radiation data without suppression of zero.

the order of the model, presenting order values of more than 100% in all degrees of the analyzed model. Furthermore, the F_{IT} is less than 50% in both cases. Normal direct radiation is of immense importance for simulations in the area of energy efficiency in buildings. Thus, it can be concluded that ARX

modeling is not effective in its use for prediction of normal direct radiation data.

In **Figures 5**, **6** it is possible to graphically observe the behavior of the data generated by the simulation with respect to the data used for validation.

FIGURE 5 | ARX model (na = 4, nb = 4, nk = 1) for prediction of normal direct radiation data without suppression of zero.

FIGURE 6 | ARX model (na = 4, nb = 4, nk = 1) for prediction of normal direct radiation data with suppression of zero.
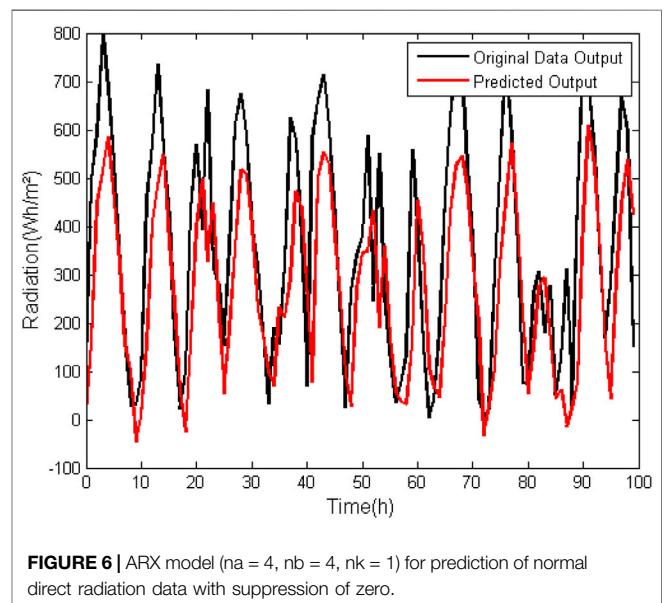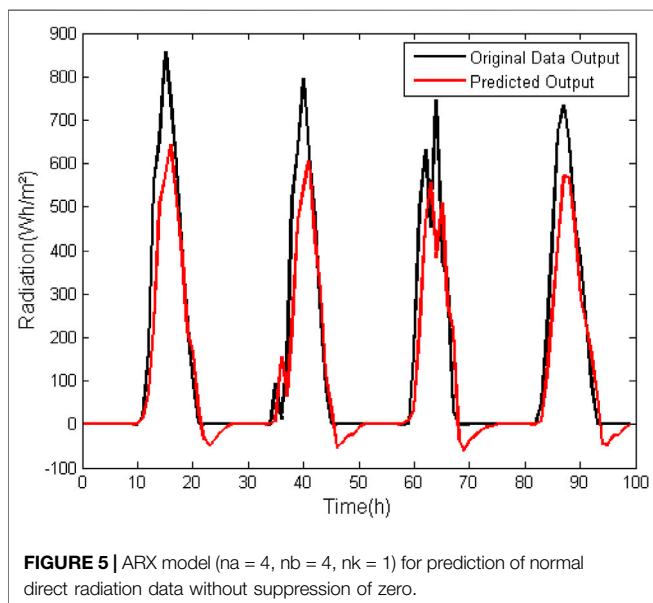
TABLE 7 | RMSE and F_{IT} of the most accurate models for each radiation component generated from horizontal global radiation.

	RMSE	F_{IT}
Normal extraterrestrial radiation	ARX (4,4,2) – 2.51%	97.33%
Infrared horizontal radiation	ARX (4,4,1) – 1.40%	79.75%
Horizontal extraterrestrial radiation	ARX (3,3,2) – 7.15%	91.34%
Normal direct radiation	ARMAX (5,5,5,1) - 59.43%	51.22%
Horizontal diffuse radiation	ARMAX (6,6,6,3) - 42.75%	53.83%
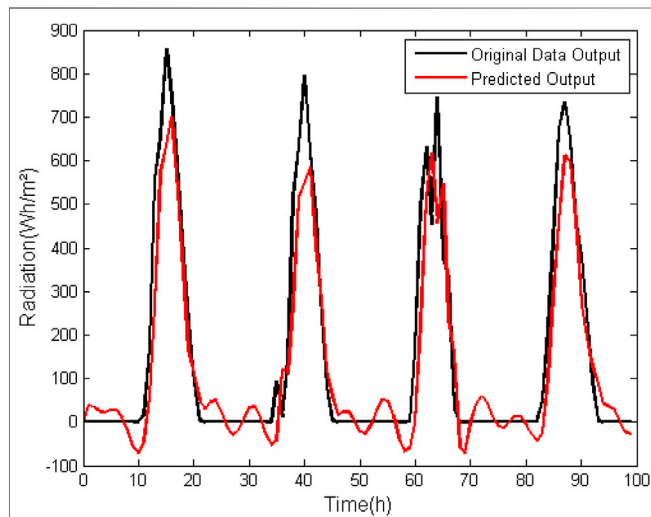
FIGURE 7 | ARMAX model (na = 5, nb = 5, nc = 5. nk = 1) for prediction of Normal Direct Radiation data without suppression of zero.
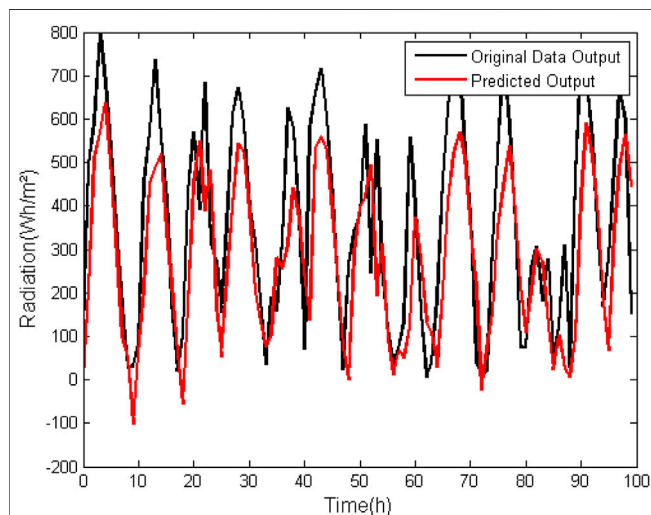
FIGURE 8 | ARMAX model (na = 5, nb = 5, nc = 5. nk = 1) for prediction of normal direct radiation data without suppression of zero.

In the ARMAX model (**Table 6**) the same effect observed in the ARX model for direct normal radiation composition occurred. Order values of more than 100% were obtained in all grades of the analyzed model. With the suppression of zero the

errors become much lower, but they are not yet within an acceptable range. Additionally, the F_{IT} is less than 50% in both cases. It can therefore be concluded that the ARMAX model generates a poorer approximation than those previously discussed when it comes to their use for prediction of normal direct radiation data.

In **Figures 7**, **8** it is possible to graphically observe the behavior of mathematically generated data with respect to the data used for validation.

The use of linear modeling (ARX and ARMAX) to predict solar radiation data was efficient for the components: normal, infrared horizontal and horizontal extraterrestrial. For these cases, the smallest errors found between the data used in validation and the data generated by the models, as well as their corresponding F_{IT}, took into consideration the computational cost-benefit in the choice of k. The best F_{IT} among the models can be seen in **Table 7**.

System identification is the procedure of mathematical construction of models, utilizing input and output data previously available for specific analyzes (Haykin 1999; Ljung 1999). Erdoğan and Gülal (2009) defined the system identification as "a matter of finding the numerical values of the model parameters that provide the best agreement between the computed and observed system outputs".

4 CONCLUSION

The description of the behavior of meteorological variables, more specifically of solar radiation, is of extreme importance in several areas including aviation, meteorology, power generation, agriculture, hydrology, and others. This description deserves better investigation such as the volume of meteorological data available, since advent of the installation of automatic meteorological stations is still less than expected to increase the level of research that depends on this variable.

Even if there is already a mathematical and physical definition that satisfactorily explains the behavior of solar radiation, it is also necessary to develop and apply predictive methods contribute to the improvement of research that uses solar radiation as the input data variable. The use of RMSE places greater weight on large errors than on small ones, thus emphasizing discrepant data inconsistently with the median of sample data. This explains why the analyzes of horizontal direct radiation and horizontal diffuse radiation have a high RMSE when compared to the others.

Furthermore, a non-sinusoidal arrangement of radiation compositions can be seen graphically in which high RMSE values and low F_{IT} values were found. This result demonstrates that linear modeling of the ARX or ARMAX type works well for variable databases that vary cyclically, as if it were a composition of sines and cosines.

It was noticed that the ARX or ARMAX linear models worked satisfactorily for databases including variables that vary cyclically, as is the case of most

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

VS conceived the analysis procedure, provided the analysis tools, the methodology, and the software. DF supervised it giving guidance. JC provided the database. PV has contributed checking the paper for errors and modifying the format of the paper. VS, DF, JC, and PV reviewed the manuscript. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Aguirre, L. A. (2007). *Introduction to Systems Identification*. 3rd ed. Brazil: UFMG. (In Portuguese).

Barnaby, C. S., and Crawley, D. B. (2011). "Weather Data for Building Performance Simulation," in *Building Performance Simulation for Design and Operation*. Editors J. Hensen and R. Lamberts (London, United Kingdom: SponPress), 37–55. chap.3.

Belmahdi, B., Louzazni, M., Akour, M., Cotfas, D. T., Cotfas, P. A., and El Bouardi, A. (2021). Long-Term Global Solar Radiation Prediction in 25 Cities in Morocco Using the FFNN-BP Method. *Front. Energ. Res.* 9, 733842. doi:10.3389/fenrg.2021.733842

Carlo, J., and Lamberts, R. (20082008). Development of Envelope Efficiency Labels for Commercial Buildings: Effect of Different Variables on Electricity Consumption. *Energy and Buildings* 40 (11), 2002–2008. doi:10.1016/j.enbuild.2008.05.002

Chan, A. L. S. (2011). Developing a Modified Typical Meteorological Year Weather File for Hong Kong Taking into Account the Urban Heat Island Effect. *Building Environ.* 46 (12), 2434–2441. doi:10.1016/j.buildenv.2011.04.038

Erdoğan, H., and Gülal, E. (2009). Identification of Dynamic Systems Using Multiple Input–Single Output (MISO) Models. *Nonlinear Anal. Real World Appl.* 10 (2), 1183–1196. doi:10.1016/j.nonrwa.2007.12.008

Froisy, J. B. (2006). Model Predictive Control Building a Bridge between Theory and Practice. *Comput. Chem. Eng.* 30 (10–12), 1426–1435. doi:10.1016/j.compchemeng.2006.05.044

Fruteira, R. S., Leite, M. L., and Virgens Filho, J. S. (2011). Performance of the Model PGECLIMA_R in the Simulation of Daily Synthetic Series of Global Solar Radiation for Different Localities of the State of Parana (In Portuguese). *Braz. J. Climatology* 9 (9), 35–47. Available at: ufpr.br/revistaabclima/article/view/27511/18331. doi:10.5380/abclima.v9i0.27511

Guimarães, I. B. B., Amorim, A. C., and Carlo, J. C. (2014). *Statistical and Simulative Comparison of TRY and TMY Climatic Archives Developed for the City of Viçosa-MG. (In Portuguese).* Brazil: UFV.

Guimarães, I. B. B., and Carlo, J. C. "Statistical Comparison between Climatic Archives Developed with Different Methods (In Portuguese)," in Proceedings of EURO ELECS - Connecting People and Ideas, Portugal, 2015.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. New Jersey: Prentice-Hall.

Hensen, J. L. M., and Lamberts, R. (2011). *Building Performance Simulation for Design and Operation*. Abingdon: Spon Press.

Huang, K. Y., and Jane, C.-J. (2009). A Hybrid Model for Stock Market Forecasting and Portfolio Selection Based on ARX, Grey System and RS Theories. *Expert Syst. Appl.* 36 (3), 5387–5392. doi:10.1016/j.eswa.2008.06.103

Huang, L., Kang, J., Wan, M., Fang, L., Zhang, C., and Zeng, Z. (2021). Solar Radiation Prediction Using Different Machine Learning Algorithms and Implications for Extreme Climate Events. *Front. Earth Sci.* 9, 596860. doi:10.3389/feart.2021.596860

INMET (2011). Monitoring of Automatic Stations. Technical Note: 001. (In Portuguese). Available at: www.inmet.gov.br/(Accessed Sept 25, 2016).

INPE (2016). CPTEC - Center for Weather Forecasting and Climatic Studies. Available at: www.cptec.inpe.br (Accessed Jan. 14, 2022).

Jorgensen, J. B., Huusom, J. K., and Rawlings, J. B. (2011). "Finite Horizon MPC for Systems in Innovation Form," in Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), Orlando, FL, December 2011. doi:10.1109/CDC.2011.6161509

Khanmirza, E., Khaji, N., and Majd, V. J. (2011). Model Updating of Multistory Shear Buildings for Simultaneous Identification of Mass, Stiffness and Damping Matrices Using Two Different Soft-Computing Methods. *Expert Syst. Appl.* 38 (5), 5320–5329. doi:10.1016/j.eswa.2010.10.026

Ljung, L. (1999). *System Identification, Theory for the User*. 2nd ed. Hoboken NJ: Prentice-Hall.

Mateo, F., Carrasco, J. J., Sellami, A., Millán-Giraldo, M., Domínguez, M., and Soria-Olivas, E. (2013). Machine Learning Methods to Forecast Temperature in Buildings. *Expert Syst. Appl.* 40 (4), 1061–1068. doi:10.1016/j.eswa.2012.08.030

Mellit, A., and Pavan, A. M. (2010). 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* 84 (84), 807–821. doi:10.1016/j.solener.2010.02.006

Moura, F. A., and Montini, A. A. (2012). Application of the Model ARX to Forecast Brazilian Consumption of Industrial Electricity. *FACEF Res. Dev. Manag.* 15 (2), 192–206. (In Portuguese)Available at: periodicos.unifacef.com.br/.

Mustafaraj, G., Chen, J., and Lowry, G. (2010). Development of Room Temperature and Relative Humidity Linear Parametric Models for an Open Office Using BMS Data. *Energy and Buildings* 42 (3), 348–356. doi:10.1016/j.enbuild.2009.10.001

Obukhov, S. G., Plotnikov, I. A., and Masolov, V. G. (2018). Mathematical Model of Solar Radiation Based on Climatological Data from NASA SSE. *IOP Conf. Ser. Mater. Sci. Eng.* 363, 012021. doi:10.1088/1757-899x/363/1/012021

Piltan, F., TayebiHaghighi, S., and Sulaiman, N. B. (2017). Comparative Study between ARX and ARMAX System Identification. *Ijisa* 9 (No.2), 25–34. doi:10.5815/ijisa.2017.02.04

Rachad, S., Nsiri, B., and Bensassi, B. (2015). System Identification of Inventory System Using ARX and ARMAX Models. *Ijca* 8 (No.12), 283–294. doi:10.14257/ijca.2015.8.12.26

Ren, Y., Cao, G.-y., and Zhu, X.-j. (2006). Particle Swarm Optimization Based Predictive Control of Proton Exchange Membrane Fuel Cell (PEMFC). *J. Zhejiang Univ. - Sci. A.* 7 (3), 458–462. doi:10.1631/jzus.2006.a0458

Ríos-Moreno, G. J., Trejo-Perea, M., Castañeda-Miranda, R., Hernández-Guzmán, V. M., and Herrera-Ruiz, G. (2007). Modelling Temperature in Intelligent Buildings by Means of Autoregressive Models. *Automation in Construction* 16, 713–722. doi:10.1016/j.autcon.2006.11.003

Rodriguez Vasquez, J. R., Rivas Perez, R., Sotomayor Moriano, J., and Peran Gonzalez, J. R. (2008). System Identification of Steam Pressure in a Fire-Tube Boiler. *Comput. Chem. Eng.* 32 (12), 2839–2848. doi:10.1016/j.compchemeng.2008.01.010

Roriz, M. (2012). Climate Files of Brazilian Municipalities. Available at: http://www.labeee.ufsc.br/downloads/arquivos-climaticos (Accessed Aug 24, 2016).

Rossi, F. A., Dumke, E., and Kruger, E. L. Update of the Reference Climatic Archive for Curitiba. Proceedings of X National Meeting and VI Latin American Meeting of Comfort in the Built Environment. Natal, Brazil. 2009. Available at: https://bv.fapesp.br/pt/auxilios.

Ruslan, F. A., HaronSamad, K., Samad, A. M., and Adnan, R. (2017). "Multiple Input Single Output (MISO) ARX and ARMAX Model of Flood Prediction System: Case Study Pahang," in Proceedings of the IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA 2017), Penang, Malaysia, 10 - 12 March 2017. doi:10.1109/CSPA.2017.8064947

Santos, A. A. P., da Costa, N. C. A., Jr., and Coelho, L. D. S. (2007). Computational Intelligence Approaches and Linear Models in Case Studies of Forecasting Exchange Rates. Expert Syst. Appl. 33 (4), 816–823. doi:10.1016/j.eswa.2006.07.008

Shakouri, G. H., and Radmanesh, H. R. (20092009). Identification of a Continuous Time Nonlinear State Space Model for the External Power System Dynamic Equivalent by Neural Network. Electr. Power Energ. Syst. 31 (7-8), 334–344. doi:10.1016/j.ijepes.2009.03.016

Shumway, R. H., and Stoffer, D. S. (2006). Time Series Analysis and its Applications with R Examples. New York: Springer.

Suganthi, L., and Samuel, A. A. (2012). Energy Models for Demand Forecasting-A Review. Renew. Sustain. Energ. Rev. 16 (2), 1223–1240. doi:10.1016/j.rser.2011.08.014

Yang, L., Cao, Q., Yu, Y., and Liu, Y. (2020). Comparison of Daily Diffuse Radiation Models in Regions of China without Solar Radiation Measurement. Energy 191, 116571. doi:10.1016/j.energy.2019.116571

Yang, Y., Yu, T., Zhao, W., and Zhu, X. (2021). Kalman Filter Photovoltaic Power Prediction Model Based on Forecasting Experience. Front. Energ. Res. 9, 682852. doi:10.3389/fenrg.2021.682852

Zanetti, S. S., Oliveira, V. D. P. S. D., and Pruski, F. F. (2006). Validação Do modelo ClimaBR em relação ao número de dias chuvosos e à precipitação total diária. Eng. Agríc. 26 (1), 96–102. doi:10.1590/s0100-69162006000100011