

Article

Short-Term Forecasting of Energy Production for a Photovoltaic System Using a NARX-CVM Hybrid Model

Eduardo Rangel-Heras ¹, César Angeles-Camacho ^{2,*}, Erasmo Cadenas-Calderón ¹
and Rafael Campos-Amezcuca ³

¹ Facultad de Ingeniería Mecánica, Universidad Michoacana de San Nicolás de Hidalgo, Santiago Tapia No. 403, Centro, Morelia 58000, Mexico; phdeduardorangelheras@gmail.com (E.R.-H.); ecadenas@umich.mx (E.C.-C.)

² Instituto de Ingeniería, Universidad Nacional Autónoma de México, Avenida Universidad No. 3000 Coyoacán, Ciudad Universitaria, Ciudad de México 04510, Mexico

³ Tecnológico Nacional de México/Centro Nacional de Investigación y Desarrollo Tecnológico, Interior Internado Palmira S/N, Col. Palmira, Cuernavaca 62490, Mexico; rafael.ca@cenidet.tecnm.mx

* Correspondence: cangelesc@ii.unam.mx

Abstract: In this paper, a methodology for short-term forecasting of power generated by a photovoltaic module is reported. The method incorporates a nonlinear autoregressive with exogenous inputs (NARX) fed by the solar radiation and temperature time series, as well as an estimation of power time series obtained by implementing an ideal single diode model. This synthetic time series was validated against an actual photovoltaic module. The NARX model has been implemented in conjunction with the corrective vector multiplier (CVM) technique, which uses solar radiation under clear sky conditions to adjust the forecasting results. In addition, collinearity and the Granger causality tests were used to choose the input variables. The forecasting horizon was 24-h-ahead. The hybrid NARX-CVM model was compared to a nonlinear autoregressive neural network and persistence model using the typical forecasting error measures such as the mean bias error, mean squared error, root mean squared error and forecast skill. The results showed that the forecasting skills of the hybrid model are about 34% against the NAR model and about 42% against the Persistence model. The model was validated by blind forecasting. The results demonstrated evidence of the quality of the conformed forecasting model and the convenience of its implementation and building.

Keywords: solar energy; electrical power forecasting; artificial intelligence



Citation: Rangel-Heras, E.; Angeles-Camacho, C.; Cadenas-Calderón, E.; Campos-Amezcuca, R. Short-Term Forecasting of Energy Production for a Photovoltaic System Using a NARX-CVM Hybrid Model. *Energies* **2022**, *15*, 2842. <https://doi.org/10.3390/en15082842>

Academic Editors: Sonia Leva, Emanuele Ogliari and Alessandro Niccolai

Received: 14 March 2022

Accepted: 8 April 2022

Published: 13 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Research concerning solar radiation (SR) forecasting with the finality of estimating the electrical energy produced by photovoltaic systems has been increased in the last few years. SR estimation is of great importance because this parameter is used in the sizing of the photovoltaic and thermal systems; moreover, it has meteorological variables [1,2].

Even though it is difficult to classify them, they can generally be divided into three groups: in the first group, those who use only the clarity index [3–7]. In the second, the hybrid models, such as hybrid data grouping models [8], models that combine different types of artificial neural networks (ANN) [9,10], and models that use the clarity index [11]. Finally, in the third group, only ANN techniques are used.

There are models to predict the solar radiation or some of its components that only use historical data of a single variable; these models generally present a good performance; however, in some cases, there are used because there are no more meteorological variables. The most useful techniques for univariate data are statistical techniques [2,5] and neural networks.

Many of these works focused mainly on the architecture of the neural network and varying different parameters, such as the lags number, training functions, activation func-

tions, et cetera [2,3,11–14]. Among the most used models are the Feed Forward Neural Networks and Elman Recurrent [15].

For example, Tingting Zhu et al. [16] used a Siamese convolutional neural network to carry out the forecasting of the direct normal irradiance in a time horizon of 10 min; this is an interesting study due to the implemented neural network type since the more conventional used is deep learning. Other ANNs are also used in this forecasting process as genetic algorithms [17]. Besides the univariate models, there are approaches that take advantage of the characteristics of two or more techniques to develop hybrid models [8]. Abdel-Rahman Hedar et al. [18] proposed a hybrid model used to forecast solar radiation, using Weather Research Forecasting (WRF) and machine learning. They established a classification of the solar radiation, and then several models of classification were used to determine the corresponding class. The proposed methodology is evaluated using a real environment temporal environment data set collected from different regions of Saudi Arabia; nevertheless, the causality between the variables was not analyzed. Other models used different ANN techniques [9,10], diverse meteorological variables and additional variables such as extraterrestrial radiation [19]. It is also important to mention the models that only used the clear index to make predictions.

According to the above mentioned, many models use different types of ANN. In most works, the vectors that make up the input layer of the network are chosen arbitrarily; the same happens with the inputs that feed the network that represent the lags of the used variables (in the case of the NARX model, for example). Generally, the authors solve the problem of the input vector conformation by testing different combinations of variables and choosing the architecture that provides the best results, focusing on the development of a methodology for ANN training [20,21]; failing that, they use other types of methods to find the optimal inputs such as the k-nearest neighbors model [5,8,10], or other data grouping techniques.

Different combinations and types of input variables have been proposed for training the ANN for modeling, estimating, or forecasting the behavior of the SR. Kaplanis et al. [14], Yadav et al. [22] and Kaushika et al. [23] implemented a combination of meteorological and geographic data to improve the SR prediction. In contrast, Hocaoglu and Serttas [20] and Wang et al. [21] worked only with meteorological data. NARX models have been implemented successfully to forecast wind speed by Cadenas et al. [24]. These models have also been implemented successfully to predict SR by Ahmad et al. [25] and Alzahrani et al. [26].

This research reports a methodology for predicting short-term energy produced by a photovoltaic generating model (PGM) using a hybrid model. The methodology incorporates a NARX model fed with solar radiation and temperature time series, as well as a synthetic time series of the power of a photovoltaic module obtained from an idealized model of a single diode. Collinearity and the Engel–Granger tests were used to choose the significant variables of the input vector of the NARX model. The autocorrelation and partial autocorrelation functions were also implemented to estimate the input vector lags. In addition to the NARX model, the corrective vector multiplier (CVM) technique obtained from a solar model was implemented. The turbidity factor of the solar energy was used to adjust the forecasting results.

This work was divided into six sections: Section 1 exhibits a general view of the models used to forecast solar radiation as a first approach to predicting solar systems' output power. Section 2 presents the fundamental aspects of the models, techniques and tests used for solar radiation prediction. Section 3 describes the proposed methodology used to build the forecasting hybrid model. Section 4 shows the typical performance metrics employed to evaluate the forecasting models. In Section 5, an analysis of the forecasting results, as well as a discussion of the feasibility of the methodology used in the hybrid model construction, are included. Finally, the conclusions of the research work are presented in Section 6.

2. Mathematical Models

2.1. Collinearity Test

The collinear variables are those whose data vectors that represent them and are in the same line. In a general way, it is if the vectors that represent n variables are in the same subspace, that is, when one vector is a linear combination of the other vectors [27]. Exact collinearity is difficult to find in a real situation, obviously; the fact that exact collinearity is difficult to find does not mean that there are no collinearity troubles [28].

To find trouble collinearity, an analysis is required of the main components of the dependent variables to figure out if there is such a problem. The linear combination of the original variables is defined by the main components of a set of variables to other variables.

2.2. Augmented Dickey–Fuller Test

It is possible to apply the causality test only if the time series is stationary. This work implemented the augmented Dickey–Fuller (ADF) test. This test is based on the null hypothesis, which establishes that the analyzed time series is not stationary. Therefore, the time series is stationary if the null hypothesis is rejected [29–31]. One way to delete non-stationarity is through the differentiation method. This method is defined by the change between each of the observations in the original time series, as in the following equation:

$$Y'_t = Y_t - Y_{t-1} \quad (1)$$

where Y_t is the corresponding value in the time t and Y_{t-1} is the corresponding value in the time $t-1$.

2.3. Engle–Granger Causality Test

The Engle–Granger causality test has been used to know the relationship between diverse meteorological variables, as presented in [32]. This technique examines whether the lagged value of one variable helps to forecast other variables in a model, such as the Engle–Granger test [29]:

- H_0 : X does not Granger cause Y ;
- H_1 : X Granger causes Y .

Rule of decision: of p -value is:

- <0.01 X Granger causes Y at the 1%;
- >0.01 X does not Granger cause Y at the 1%.

The equations for two variables are as follows:

$$Y_t = \sum_{i=1}^n \alpha_i X_{t-i} + \sum_{j=1}^n \beta_j Y_{t-j} + u_{1t} \quad (2)$$

$$X_t = \sum_{i=1}^n \lambda_i Y_{t-i} + \sum_{j=1}^n \delta_j X_{t-j} + u_{2t} \quad (3)$$

where X represents mean global irradiation, Y represents the air temperature, u_{1t} is the uncorrelated white noise, α_i , β_j , λ_i and δ_j are parameters to be determined and n is the number of lags. Similarly, using the autoregressive vector technique, it is possible to apply this technique to several variables [29].

2.4. Simplified Single Diode Model

The method based on series resistance has been widely studied and characterized to model a PV generator [33–35]. It can obtain the characteristic curves and the generated electrical power for a PVG [27], as shown in Figure 1.

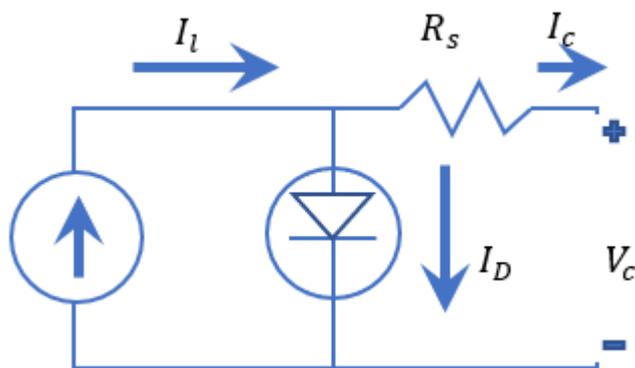


Figure 1. Modeling of a solar cell with series resistance.

The characteristic equation V-I that is used to model de PVG is defined by:

$$I_G = I_{scGE} - N_{pG} I_0 \left(e^{\frac{V_G + I_G R_{sG}}{n N_{sG} V_T}} - 1 \right) \tag{4}$$

where I_0 is determined by solving Equation (4) for the following open-circuit conditions: $I_G = 0$ and $V_G = V_{ocGE}$. I_{scGE} and V_{ocGE} are the short-circuit current and open-circuit voltages to the environmental conditions for the generator.

$$I_0 = \frac{I_{scGE}}{N_{pG} \left(e^{\frac{V_{ocGE}}{n N_{sG} V_T}} - 1 \right)} \tag{5}$$

Equation (5) can be solved by an iterative method. The resistance R_{sG} of the generator is defined as [36]:

$$R_{sG} = \frac{V_{ocG}}{I_{scG}} - \frac{P_{sG}}{FF_0 I_{scG}^2} \tag{6}$$

with,

$$FF_0 = \frac{v_{oc} - \ln(v_{oc} - 0.72)}{1 + v_{oc}} \tag{7}$$

$$v_{oc} = \frac{V_{ocG}}{N_{sG} n V_T} \tag{8}$$

where FF_0 is the filling factor of the generator without series resistance, and v_{oc} is the normalized value for the open-circuit voltage.

The short-circuit current of the PVG with temperature dependence was defined by [36] as:

$$I_{scGE} = \frac{I_{scGn}}{1000} + \left(\frac{\partial I_{scG}}{\partial T_c} \right) (T_c - T_{c0}) \tag{9}$$

where T_c is the cell temperature, T_{c0} is the cell temperature to conditions of nominal operation (25 °C), I_{scGE} is the short-circuit current of the generator for environmental conditions, $\left(\frac{\partial I_{scG}}{\partial T_c} \right)$ is the temperature coefficient for short-circuit current and I_{scGn} is the nominal short-circuit current of the generator, i.e., the maximum power point to nominal operation.

The open-circuit voltage of the generator, depending on the temperature, is defined as:

$$V_{ocGE} \approx V_{ocGn} + \left(\frac{\partial V_{ocG}}{\partial T_c} \right) (T_c - T_{c0}) + \frac{k T_{cK}}{q} \ln \left(\frac{I_{scGE}}{I_{scGn}} \right) \tag{10}$$

where $\left(\frac{\partial V_{ocG}}{\partial T_c} \right)$ is the temperature coefficient of the open-circuit voltage.

The cell temperature can be calculated with the environmental temperature and the nominal operating cell temperature (NOCT) provided by the manufacturer. In this way, the cell and module temperature will be given by:

$$T_c = T_E + \left(\frac{NOCT - 20}{800} \right) \cdot SR \quad (11)$$

where $NOCT$ is the nominal operating cell temperature provided by the manufacturer, T_c is the cell temperature and T_E is the environmental temperature.

2.5. Solar Radiation under Clear Sky Conditions

Solar radiation suffers losses because of atmospheric absorption and dispersion when this passes through the atmosphere. According to the Handbook of Solar Energy [37], after the atmospheric absorption phenomenon, the normal solar flow rate (solar radiation/normal irradiation) that reaches the Earth's surface can be estimated from Equation (12):

$$SR_{clr} = I_{ext} \cdot \exp \left[- \frac{T_R}{(0.9 + 9.4 \cdot \cos \theta_z)} \right] \quad (12)$$

where SR_{clr} is the SR under clear sky conditions, I_{ext} is the extraterrestrial radiation and $\cos \theta_z$ is defined as:

$$\cos \theta_z = \cos \phi \cos \delta \cos \omega + \sin \delta \sin \phi \quad (13)$$

where θ_z is the zenith angle, T_R is the turbidity factor, ϕ is the site latitude, δ is the declination angle of the Earth and ω is the hourly angle. The extraterrestrial radiation is obtained through the following equation [38]:

$$I_{ext} = I_{sc} E_0 (\sin \delta \sin \phi + 0.9972 \cdot \cos \delta \cos \phi \cos \omega_{0.5}) \quad (14)$$

where E_0 is the eccentricity correction factor, $I_{sc} = 1367 \text{ W/m}^2$ is the solar constant and $\omega_{0.5}$ is the hourly angle each half hour.

2.6. Calculation of the Turbidity Factor

The following criteria were considered for calculating the turbidity factor:

- At least one day of each month is completely clear and corresponds to the day that records the maximum SR measured for that month.
- For each month of the year, the maximum extraterrestrial SR value was calculated to each maximum extraterrestrial SR value corresponding to a value of $\cos \theta_{z_{month}}$.

By solving the turbidity factor from Equation (12), Equation (15) is obtained:

$$T_R = -\log \left(\frac{SR_{max_{month}}}{I_{ext_{month}}} \right) \cdot (0.9 + 9.4 \cdot \cos \theta_{z_{month}}) \quad (15)$$

where $SR_{max_{month}}$ is the monthly maximum SR, $I_{ext_{month}}$ is the monthly maximum extraterrestrial SR and $\cos \theta_{z_{month}}$; i.e., they depend on the maximum extraterrestrial radiation value.

Twelve $SR_{max_{month}}$ values were obtained; one for each month. In the same way, twelve $I_{ext_{month}}$ maximum values were calculated. Finally, twelve values of $\cos \theta_{z_{month}}$, each according to the maximum monthly of $I_{ext_{month}}$, were obtained. Once these values were obtained, the monthly turbidity factor could be calculated [27].

3. Methodology for Building the NARX-CVM Hybrid Model

During the first stage, the time series of the site's meteorological variables used for the hybrid NARX-CVM model construction were the temperature (T), solar radiation (SR), relative humidity (RH), wind speed (WS), and pressure (P). In addition, a synthetic time series of the electric power (EP) was implemented, which was obtained by modeling the photovoltaic equipment. The structure of the NARX model was performed by an input/output

vector. Collinearity, unit root and causality tests defined the input vector [24,27,29,32]. In addition, the autocorrelation and partial autocorrelation functions were implemented to estimate the lags. The final resulting vector used to feed up the NARX model was conformed by the time series of SR, T and EP, as well as 24 lags. The output vector was the time series of the electric power (CFP).

Figure 2 shows the flow chart to forecast the electric power of a photovoltaic system. First, the electrical power (EP) estimation was obtained by feeding up the ideal single diode model (PVM) with SR and T time series. Second, the NARX model was trained using the time series of the SR, T and EP to get the forecast of the electrical power (FP). Finally, FP was improved by implementing the corrective vector multiplier (CVM) technique and obtaining the corrected forecast of the power (CFP). The main steps of the methodology are described below.

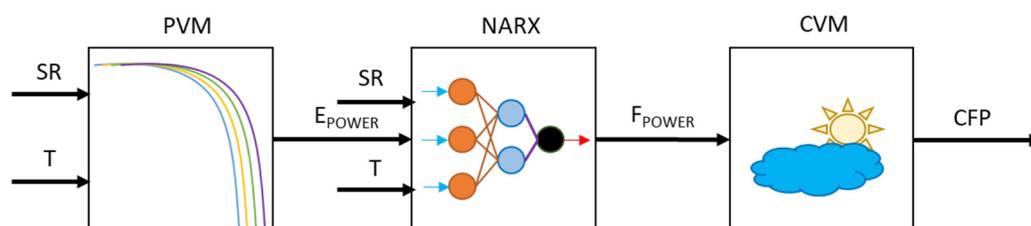


Figure 2. Flow diagram to carry out the electric power forecasting of a photovoltaic system.

3.1. Step 1: Databases (VARIABLES)

Temixco is a Mexican city located in the state of Morelos. It has warm, sub-humid weather and possesses excellent solar energy potential. The meteorological variables recorded in the ESOLMET station [39] are solar radiation, temperature, relative humidity, wind speed and pressure. This database was chosen thanks to its quality, its small sampling rate and its storage frequency.

The sensors record the data with an average of 10 min, and the measurement data correspond to two years. The database has 105,120 data on each meteorological variable. The data were transformed to an hourly scale, obtaining 17,520 data of each variable.

Table 1 shows the sensor characteristics used in the ELSOMET meteorological station.

Table 1. Characteristics of the meteorological sensors.

Probe	Sensor	Range	Accuracy
CS500 Temperature probe	1000 Ω platinum resistance, DIM43760B	-40.0 $^{\circ}\text{C}$ to $+60.0$ $^{\circ}\text{C}$	± 0.5 $^{\circ}\text{C}$
CS500 Relative humidity probe	Vaisala INTERCAP	0 to 100%	$\pm 3\%$
R.M. Young wind sentry anemometer	Cups Wheel Assembly	0.0 to 50.0 m/s	± 0.5 m/s
PTB110 Barometer	Vaisala BAROCAP	500.0–1100.0 hPa	± 0.3 hPa
WXT510 Weather transmitter	Ultrasonic Signal	0 to 60 m/s	3%
	BAROCAP	600 to 1100 hPa	± 0.5 hPa
	THERMOCAP Sensor	-52.0 $^{\circ}\text{C}$ to 60.0 $^{\circ}\text{C}$	± 0.3 $^{\circ}\text{C}$
	HUMICAP Sensor	0 to 100% RH	$\pm 3\%$ RH

3.2. Step 2: Selecting the Input Variables (INPUTS)

In this step, different practical techniques for choosing the multivariable forecasting model's input variables are explained: collinearity test, augmented Dickey–Fuller (ADF) test, time series differentiation and causality test. For practical purposes, the electrical power was not treated as one more variable among the input variables analysis because the EP was obtained from SR and T; besides, it has similar behavior to SR.

3.2.1. Collinearity Test

To avoid spurious results, a forecasting model should not be including redundant or irrelevant variables. The collinearity test helps to identify these kinds of inlet variables.

Figure 3 shows the collinearity test results; the analyzed meteorological variables are shown in the abscissa axis, while the ordinate axis presents the variance decomposition in a 0 to 1 range. The collinear variables have values above 0.5, and a red circle highlights them.

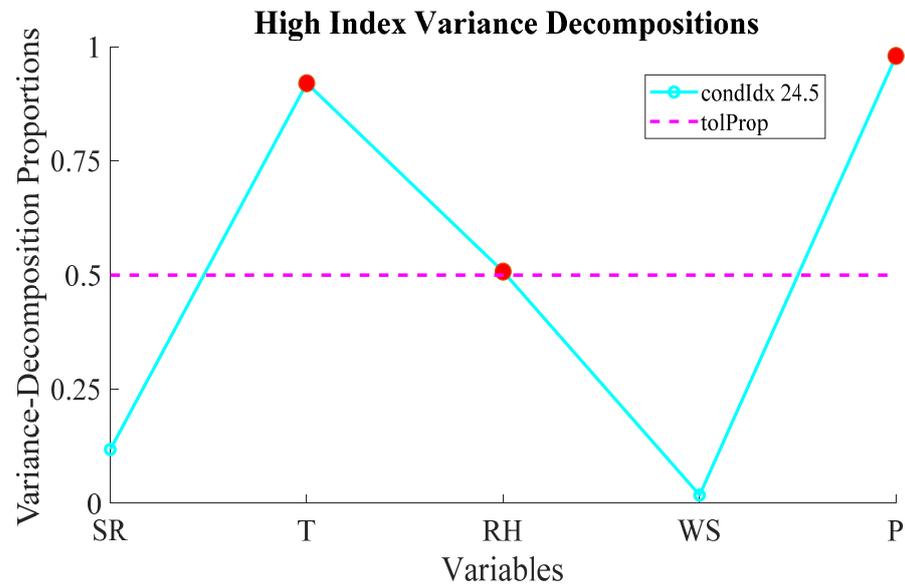


Figure 3. Collinearity test of the meteorological variables of Temixco.

From Figure 3, the collinear variables are temperature (T), relative humidity (RH) and atmospheric pressure (P). Therefore, three different input vectors have been obtained (Figure 4): input 1 = SR, T, wind speed (WS); input 2 = SR, RH, WS, and finally, input 3 = SR, WS, P. The collinear variables, highlighted in bold font, were not kept together in either combination.

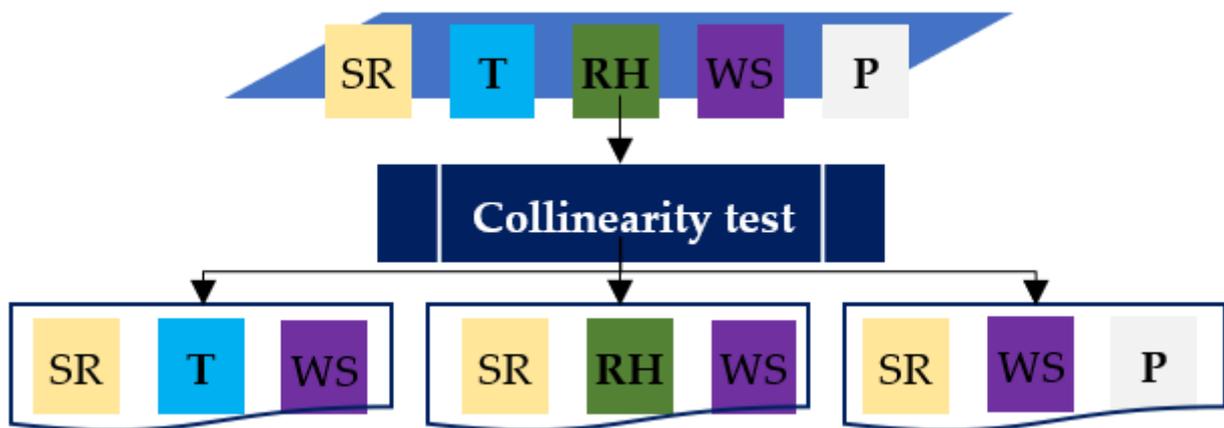


Figure 4. Meteorological variables combination.

3.2.2. Augmented Dickey–Fuller Test (ADF)

One of the conditions for applying the Engle–Granger causality test is that the time series must be stationary; therefore, the ADF test must be implemented before the causality test. The augmented Dickey–Fuller test determines whether the time series is stationary. Table 2 shows the results of the ADF test. First, the autocorrelation is verified by analyzing the Durbin–Watson statistic.

Table 2. ADF test results (significance test 5%).

Variable	Durbin–Watson Statistic	Critical Value	T-Statistic	p-Valor
SR	2.00	−1.94	−1.31	0.18
T	2.00	−1.94	−0.40	0.54
RH	2.00	−1.94	−1.46	0.13
WS	1.99	−1.94	−1.71	0.08
P	1.99	−1.94	−0.02	0.67

Durbin–Watson values, estimated using a 5% significance level, range from 1.85 to 2.15, and conclude that autocorrelation does not exist in the model [25,39]. Second, the *p*-value was verified to be higher than 0.05, which indicates that the time series has a unit root; thus, it is stationary. Consequently, there was no need to differentiate any time series, and the causality test could be applied directly.

Figure 5 shows the flow chart following the results obtained with the ADF test. In case some of the time series have not been stationary, the first difference is calculated, and the ADF test is applied again. Generally, it is enough to differentiate the time series once for passing the ADF test.

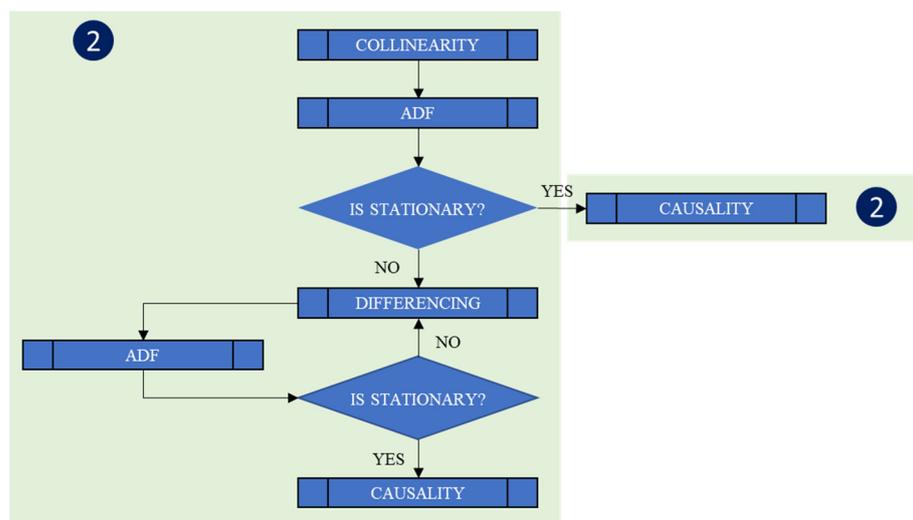


Figure 5. Flow chart to choose the input variables (Step 2).

3.2.3. Engle–Granger Causality Test Results

In this causality test, the nullity hypothesis establishes that the independent variables cause the dependent variable, so the independent variables contain helpful information for forecasting the behavior of the dependent variable. In these kinds of tests, the significance level is 1% [29]. Table 3 shows the causality test results applied to each group of variables obtained from the collinearity test. From the causality test results, only the WS value is statistically significant, as $0.12 > 0.01$. Therefore, in this case, the nullity is rejected. Thus, it is established that for the variable set (SR, T, WS), the independent variable WS does not contain useful information to forecast the dependent variable, SR. Then, this variable is discarded from the input variables, and the unit is used.

Figure 6 shows the variable combinations once the causality tests were applied. The red-dotted line indicates the inputs that best describe the behavior of the power, and the bold capital letters indicate the variables with strong collinearity. As previously reported, the first input combination resulting from the collinearity test was SR, T and WS. Once the causality test was applied, SR and T were obtained. The second variable set obtained from the collinearity test was SR, RH and WS; the causality test did not report any change. The causality test also shows no change for the third variable set (SR, WS, P). Therefore, it had

input 1 = SR, T; input 2 = SR, RH, WS; and input 3 = SR, WS, P. The proposed methodology focuses on finding the optimal input vector; thus, the NARX model includes only SR and T as input variables for this study case.

Table 3. Causality test results with a significance level of 1%.

(a) Group of variables (SR, T, WS) → Dependent variable SR	
Variable	Probability
T	0.00
WS	0.12

(b) Group of variables (SR, RH, WS) → Dependent variable SR	
Variable	Probability
RH	0.00
WS	0.00

(c) Group of variables (SR, P, WS) → Dependent variable SR	
Variable	Probability
P	0.00
WS	0.00

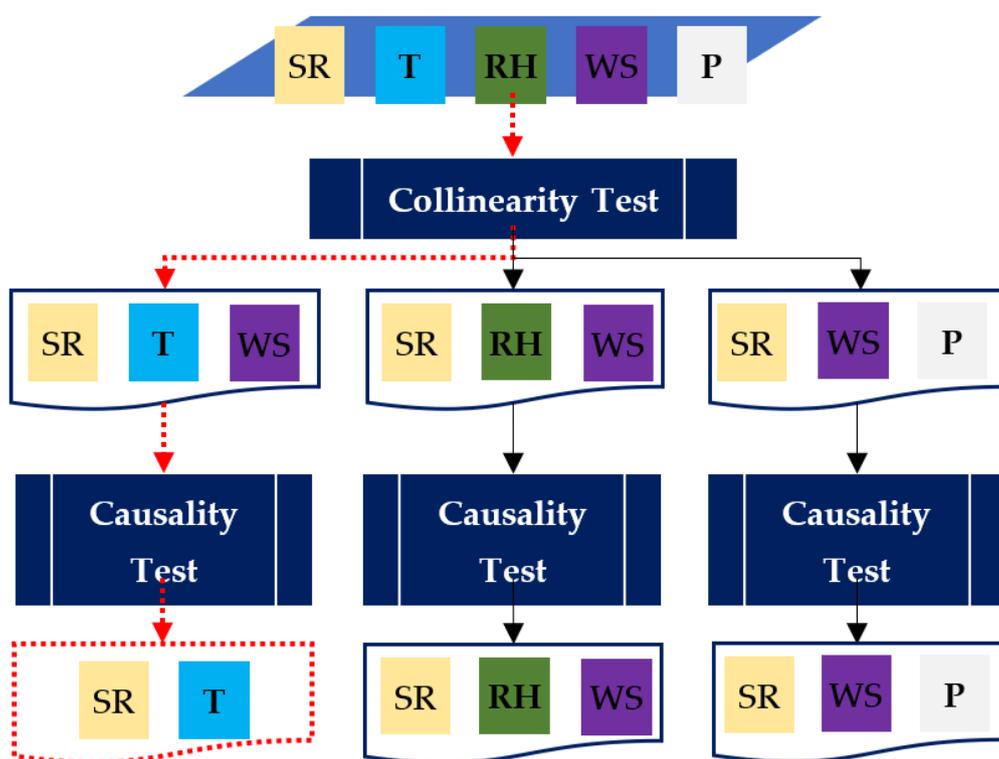


Figure 6. Output variables after applying the causality test.

3.3. Step 3: Lags for the NARX Model (LAGS)

The autocorrelation (ACF) and partial autocorrelation functions (PACF) were implemented to estimate the number of lags in the forecasting models. First, these functions were applied to the time series; then, they were analyzed to find the seasonal patterns that determine the lags number. The resulting interpretation of the ACF and PACF is through

their plots, which indicate seasonality. Seasonality is defined as a pattern that repeats itself over fixed intervals in time [28,30].

$$r_k = \sum_{t=k+1}^n \frac{(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad (16)$$

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_k Y_{t-k}. \quad (17)$$

Figure 7 shows the ACF and PACF plots. The ACF presents a sinusoidal behavior, which indicates the seasonality of the SR time series (Figure 7a). Simultaneously, the PACF plot (Figure 7b) shows peaks every 24 h, which means the time series has seasonality every 24 h. Therefore, the lags number defined for the NARX models was 24.

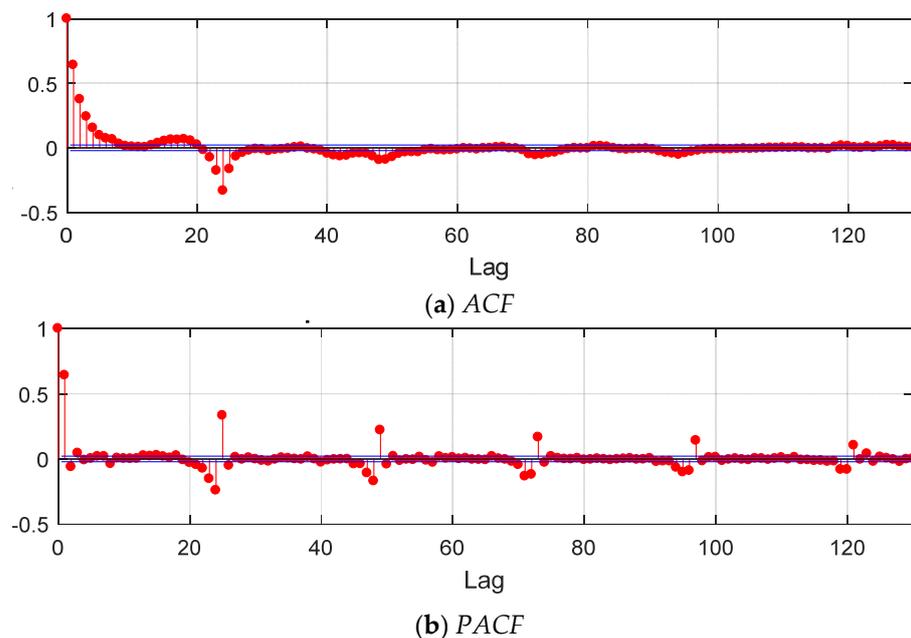


Figure 7. Autocorrelation and partial autocorrelation functions applied to SR.

3.4. Step 4: Modeling Photovoltaic Systems

The idealized single diode model was used to estimate the electric power from T (°C) and SR (W/m²) time series. A monocrystalline photovoltaic module ISF-250 black [40] was used in order to verify the model's validity. Mechanical and electrical characteristics are shown in Tables 4 and 5, respectively.

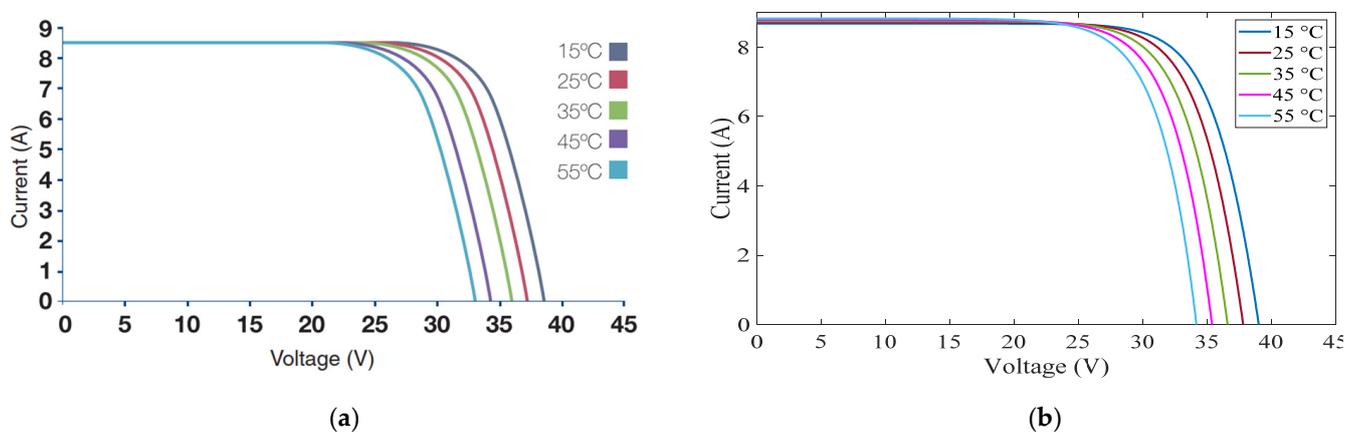
Table 4. Mechanical characteristics of the monocrystalline module ISF-250 [40].

Parameter	Characteristics
Solar cell	Monocrystalline silicon—156 mm × 156 mm (6 inches)
Number of cells	60 cells (6 × 10)
Dimensions	1667 × 994 × 45 mm (65.63 × 39.13 × 1.77 in)
Weight	19 kg (41.89 pounds)
Glass	High transmittance, patterned, tempered, 3.2 mm (EN-12150)
Frame	Anodized aluminium, grounding drills
Maximum mechanical load	5400 Pa (112.78 psf) (Snow load)
Junction box	IP 65 with three bypass diodes
Cables, plug	Solar cable 1 m (39.37 in), four mm ² (12 AWG). MC4 or LC4

Table 5. Electrical characteristics of the monocrystalline Module ISF-250 [40]. Performance at STC: Irradiance, 1000 W/m²; cell temperature, 25 °C (77 °F); AM, 1.5.

Parameter	Characteristics
Rated power (P_{max})	250 W
Open-circuit voltage (V_{oc})	37.8 V
Short-circuit current (I_{sc})	8.75 A
Maximum power point voltage (V_{max})	30.6 V
Maximum power point current (I_{max})	8.17 A
Efficiency	15.1%
Power tolerance (% P_{max})	0/+3%

Figure 8a shows the characteristic curves of the manufacturer of the monocrystalline module ISF-250, and Figure 8b shows the characteristic curves obtained by the idealized single diode model. A great similitude between both characteristic curves is observed, which indicates the reliability of the electric power time series calculated from the model.

**Figure 8.** Characteristic of the power curves of the monocrystalline module ISF-250. (a) Manufacturer curves [40], (b) Curves obtained with the algorithm.

A comparison was also made between the electrical data of the monocrystalline module ISF-250 reported in Table 5 (manufacturer) and Table 6 (calculated). The comparison result is presented in Table 7, where the most significant error was for the maximum current (I_{max}), with an error of 1.52% above the theoretical value. An essential factor to fit the single idealized diode model to the experimental model was the nonideality constant of the diode (n), which frequently is set as 1.2 [34]. But, for this case study, $n = 1.8$ was the value that the estimated results best fit the data provided by the manufacturer.

Table 6. Error rate between actual and estimated electrical characteristics.

Variable	Estimated	Actual	Error
P_{max} (W)	248.1	250.0	0.75%
V_{oc} (V)	37.5	37.8	0.83%
V_{max} (V)	30.8	30.6	−0.78%
I_{sc} (A)	8.8	8.8	−0.12%
I_{max} (A)	8.1	8.2	1.52%

Table 7. Architectures used to generate the NARX models.

Models	Lags (L)	Input $x(t)$	Output $y(t)$	Hidden Neurons (hn)	Output Neurons (m)	Tests
NARX I	24	SR, T, RH, WS, P	FPOWER	10	1	All variables
NARX II	24	SR, T, WS	FPOWER	10	1	Collinearity and causality
NARX III	24	SR, RH, WS	FPOWER	10	1	Collinearity and causality
NARX IV	24	SR, WS, P	FPOWER	10	1	Collinearity and causality
H-NARX	24	SR, T	FPOWER	10	1	Collinearity and causality

3.5. Step 5: Multivariable Forecasting Model (NARX)

In this work, a nonlinear autoregressive exogenous model (NARX) was used for the short-term prediction of solar radiation. NARX model is a nonlinear autoregressive model that has exogenous inputs. The artificial neural network training was carried out using the meteorological variables recorded during the measurement of the first year. From this database, 70% dataset was used to train the NARX model, 15% dataset for the validation, and 15% dataset for the test. The models to obtain the blind forecasting were then implemented; thus, the measurement of the second year that was not used to train the model was selected. Different inputs were generated according to the results obtained in Step 2. The model outputs were the electrical power generated by the photovoltaic module in Step 4.

Generally, a NARX model is formed of an input layer, a hidden layer, and the output layer. For this case, the input layer was conformed by the input vectors previously obtained in Step 2 and the electrical power calculated in Step 4 (EPower), while the output layer was the electrical power (FPower). Figure 9 shows a simplified representation of the NARX models developed from the input vectors and lags obtained in Steps 2 and 3. The letter “I” indicates inputs, whereas “O” denotes the outputs, and 24 is the lag number.

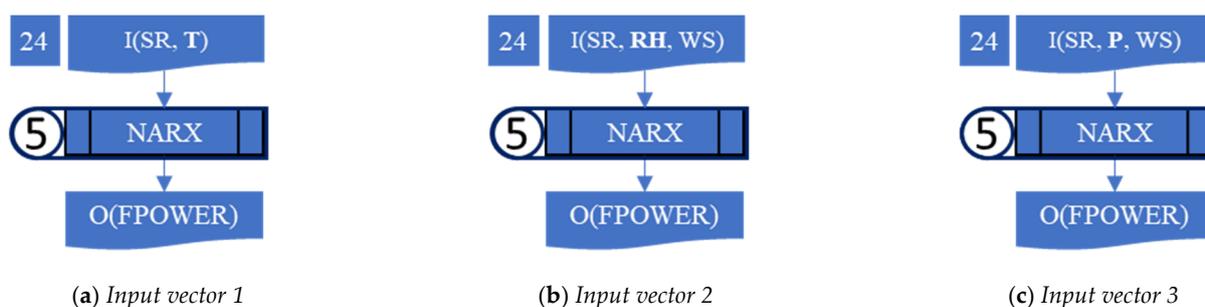


Figure 9. Development of the NARX models using different input vectors for electrical power prediction (FPOWER).

This work’s primary purpose is to improve the forecast of the power generated by a PV generator based on the appropriate selection of the input variables, the lags and the application of the corrective vector multiplier (CVM). Therefore, the default configuration proposed by the Matlab[®] program was used.

Figure 10 shows the general arrangement of the NARX model, where $x(t)$ represents the neural network inputs, n is the number of inputs variables and $y(t)$ is the output variable, L is the number of lags, w are the weights, b are the biases, hn is the number of hidden layers and m is the number of output variables.

A test was made first using all variables to determine the effectiveness of the proposed methodology. Then, using the input vectors and lags previously selected, the different architectures of the NARX models are described in Table 7. The simplest model was

the H-NARX model, and it is a NARX model with two input neurons (SR and T), ten hidden neurons, 24 lags and one output neuron (electrical power). This simple model was compared with model NARX I (all the variables are used as inputs), NARX II (input neurons with SR, T and WS), NARX III (input neurons with SR, RH and WS), and finally, NARX IV (input neurons with SR, WS and P).

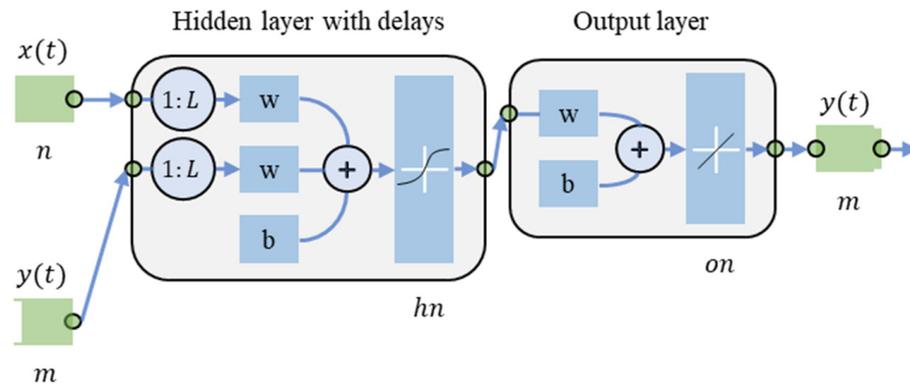


Figure 10. The general arrangement of the neural networks model.

3.6. Step 6: Output Data Depuration of the Forecasting Model (CVM)

Forecasting models of solar radiation and photovoltaic power sometimes return values that should not be considered in the final forecasting results. For example, a prediction model can forecast a positive value of the SR or output electric power of a PVG at night, which is wrong. For this reason, the use of the SR under clear sky conditions was proposed to improve the forecasting results of the electrical power through a corrective vector multiplier (CVM).

According to the performance tests, the forecast of the electrical power (CPOWER) results improved when the corrective vector multiplier was applied,

$$\text{CPOWER} = \text{FPOWER} \cdot \text{CVM} \quad (18)$$

where FPOWER is the forecast of electrical energy obtained from the NARX model, and the corrective vector multiplier was built from

$$\text{CVM}(SR_{clr}) = \begin{cases} 0 & \text{if } SR_{clr} = 0 \\ 1 & \text{if } SR_{clr} > 0 \end{cases} \quad (19)$$

4. Performance Tests

The NARX models were programmed using the *ntstool* library from Matlab[®]. The input vectors were reported in Table 8. According to the proposed methodology, the best-input vector was formed by SR and T, H-NARX without CVM and H-NARX-CVM once the CVM was applied (see Tables 8 and 9). The input number of neurons is defined by the input vectors, the hidden layer neurons are set up in 10, the output neuron is one and it is defined by the output vector. The number of lags was obtained using the ACF and PACF. The time series were pre/post-processing using the functions *removeconstantrows* and *mapminmax*. The first function removes the rows of the input vector that correspond to input elements that always have the same value because these input elements are not providing any useful information to the network, and the second function transforms input data so that all values fall into the interval $[-1, 1]$; this can speed up the learning networks. The division of data for training, validation and testing was carried out using *dividerand*; this function divided data randomly; the sample data were split up into 70%, 15% and 15% for training, validation and testing, respectively. The performance function is the mean squared error (*mse*). The transfer function is the tan-sigmoid defined as $\text{tansig}(n) = \frac{2}{1 + \exp(-2 \cdot n)} - 1$ and set up in Matlab[®] as *tansig*.

Table 8. Performance test results before applying the corrective vector multiplier.

Model	Lag	Input	Output	MBE (W)	MSE (W ²)	RMSE (W)	R ²
NARX I	24	SR, T, RH, WS, P	FPower	0.45	210.30	14.50	0.95
NARX II	24	SR, T, WS	FPower	0.70	147.83	12.16	0.97
NARX III	24	SR, RH, WS	FPower	−0.27	149.81	12.24	0.97
NARX IV	24	SR, WS, P	FPower	0.72	145.15	12.05	0.97
H-NARX	24	SR, T	FPower	−0.18	131.42	11.46	0.97

Table 9. Performance test results after applying the corrective vector multiplier.

Model	Lag	Input	Output	cMBE (W)	cMSE (W ²)	cRMSE (W)	cR ²
NARX-CVM I	24	SR, T, RH, WS, P	CPower	−0.45	184.80	13.59	0.96
NARX-CVM II	24	SR, T, WS	CPower	−0.01	142.78	11.95	0.97
NARX-CVM III	24	SR, RH, WS	CPower	−0.57	145.41	12.06	0.97
NARX-CVM IV	24	SR, WS, P	CPower	0.40	143.96	12.00	0.97
H-NARX-CVM	24	SR, T	CPower	−0.41	130.07	11.40	0.97

The NARX models trained with the 2017 dataset were used to forecast the 2018 data set from 24 h ahead and updated the existing data until the whole year was completed. Performance tests were applied to the results of the annual forecasting to estimate which ANN architecture was the one that best predicts the behavior of the electrical power. Some of the most used performance tests are the mean squared error (MSE), the root of the mean squared error (RMSE), the mean bias error (MBE) and the coefficient of determination (R²) [29,41–43]. Equations (20)–(23) describe all these performance metrics.

$$MSE = \frac{1}{n} \sum_{i=1}^n (P_{forecast} - P_{calculated})^2, \quad (20)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_{forecast} - P_{calculated})^2}, \quad (21)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (P_{forecast} - P_{calculated}) \quad (22)$$

$$R^2 = 1 - \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2}. \quad (23)$$

Table 8 shows the results of the performance tests for models before applying a corrective vector multiplier (output: FPower). Table 9 shows the results of the performance tests when the CVM (output: CPower).

5. Results and Discussion

5.1. Comparison between Models with and without CVM

Table 10 shows a comparison between the RMSE and cRMSE and the improvement rate of RMSE for each case. In all cases, forecasting results were improved when CVM was applied. The most significant improvement was obtained for the NARX-CVM I model, where the RMSE was enhanced by 6.7%. By contrast, the lowest improvement was obtained in NARX-CVM IV with a 0.4% improvement.

Another essential tool used for evaluating forecasting results is the linear regression plot, where the ordinate axis represents the forecast data while the abscise axis indicates the actual data. Figure 11 shows the linear regression plots of the five evaluated cases, where models without the CVM are at the top of the figure. At the bottom of the figure are the linear regressions that resulted from applying the corrective vector. This figure shows that the linear regression dataset was more homogeneous when the corrective vector was applied. The coefficient of correlation showed slight improvement when the CVM was

used. The linear regression dataset shows how the application of the CVM considerably improved the forecasting of the output power, especially in NARX I, II and III.

Table 10. RMSE comparison for NARX models with and without CVM.

Model	RMSE (W)	cRMSE (W)	Improvement
NARX I vs. NARX-CVM I	14.50	13.59	6.7%
NARX II vs. NARX-CVM II	12.16	11.95	1.8%
NARX III vs. NARX-CVM III	12.24	12.06	1.5%
NARX IV vs. NARX-CVM IV	12.05	12.00	0.4%
H-NARX vs. H-NARX-CVM	11.46	11.40	0.5%

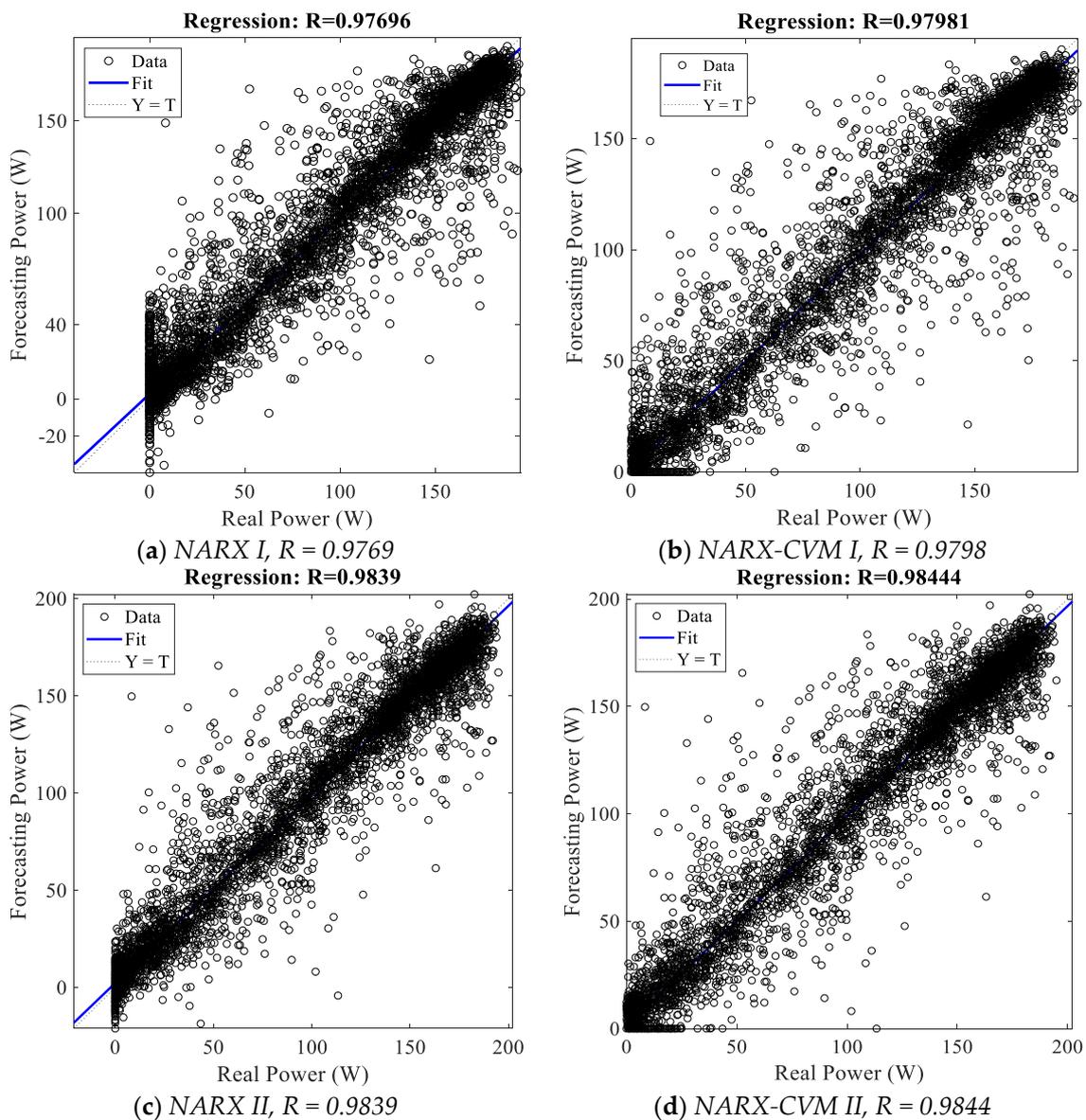


Figure 11. Cont.

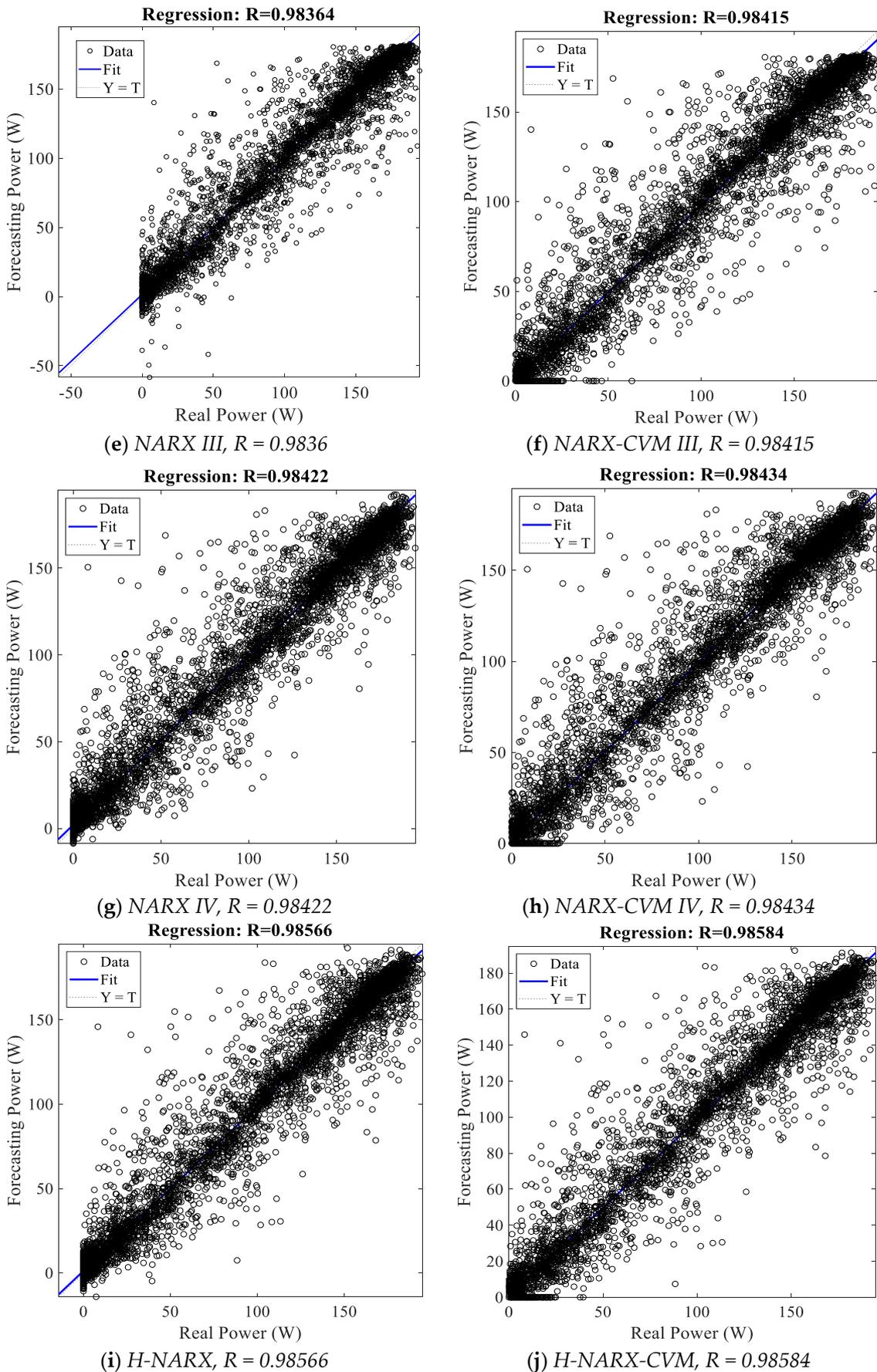


Figure 11. Linear regression analysis for the five study cases: models without CVM at the top and models with CVM at the bottom.

5.2. Comparison of the H-NARX-CVM Model against Other Models

The forecasting results are often compared with simpler models to identify the proposed model's ability to forecast. From Table 10, the H-NARX-CVM model was the one that provided the best results, that is, the model with the smallest RMSE. Therefore, this was the only model used to compare the results with other prediction models. One of the most widely used models for comparison purposes is the well-known persistence model (Equation (24)) [25,43–45].

$$S(t+h) = S(t) \quad (24)$$

Like the persistence model, the nonlinear autoregressive (NAR) neural network has been used by Benmouiza, Cheknane [9,46] and García-Tena et al. [47] to forecast time series or as a benchmark to compare more complex models.

Table 11 shows the results of the performance tests for the proposed methodology, NAR and the persistence model. The NARX models obtained with the proposed method outperformed the persistence and NAR models.

Table 11. Performance tests of the H-NARX-CVM, NAR and persistence models.

Models	Performance Tests			R ²
	MBE	MSE	RMSE	
H-NARX-CVM	−0.41	130.07	11.40	0.97
NAR	−1.12	300.57	17.34	0.94
Persistence	0.00	386.12	19.65	0.92

On the other hand, predictive modeling researchers use forecasting skills as one of the most widely used measures. This compares the developed model with a less complex model, such as persistence [41–43]. In this research, the NAR and persistence models have also been used as benchmarks:

$$Forecast_{skill}(\%) = \left(1 - \frac{RMSE_{H-NARX-CVM}}{RMSE_{Ref}} \right) \quad (25)$$

where $RMSE_{H-NARX-CVM}$ is the RMSE for the H-NARX-CVM model and $RMSE_{Ref}$ is the RMSE for the benchmark models.

Figure 12 (cyan-dotted line) shows the results of the forecasting skill using the persistence model as a benchmark, and a key performance indicator (KPI) of 35% was used as a goal. H-NARX-CVM I, II, III and IV models were higher than 35% of the KPI, with the proposed methodology being the one that obtained the best result with a skill forecasting of 42%. Compared with model 1, with a forecasting skill of 31%, it did not exceed the target value. But, when the exercise was performed using the NAR model as a benchmark, the NARX models did not exceed the KPI goal, H-NARX-CVM IV being the best result, with skill forecasting of 34%, as shown in Figure 12 (magenta-dotted line).

Results of the H-NARX-CVM, NAR and Persistence Models Versus the Real Data

This section compares the actual electrical power and the forecasting results calculated as follows:

- (1) The blind prediction of the power obtained using the proposed methodology;
- (2) The blind prediction using the NAR model;
- (3) The prediction using the persistence model.

A qualitative comparison was made using plots of time series of randomly chosen days each month (Figure 13), and quantitative analysis was performed to calculate the forecasting errors (Table 12). The days used to compare the mentioned models were obtained using the single diode model algorithm, which allows obtaining a power production daily of the PVG.

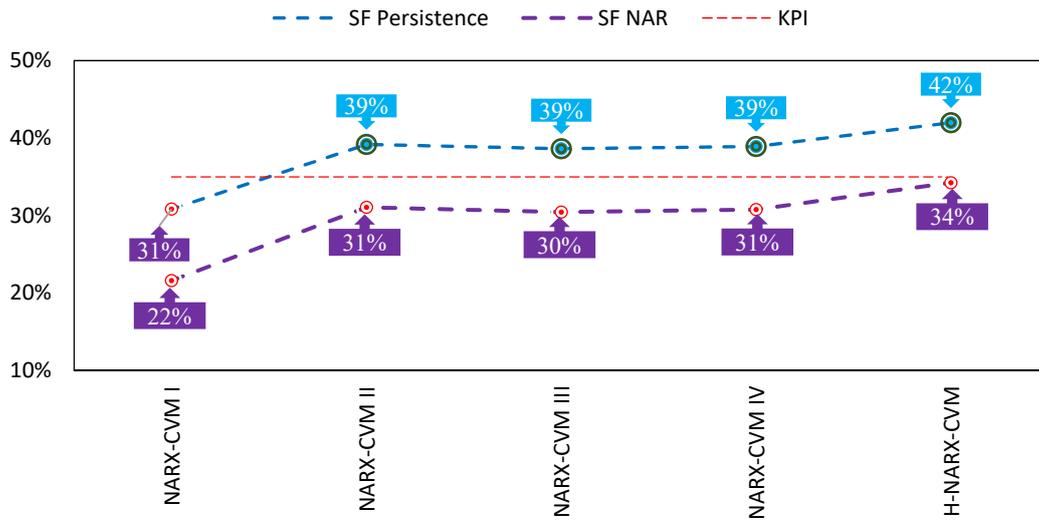
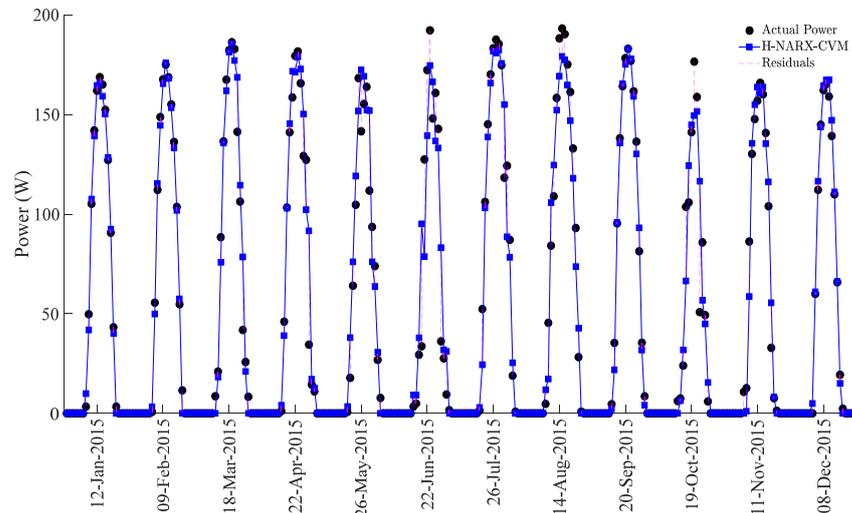
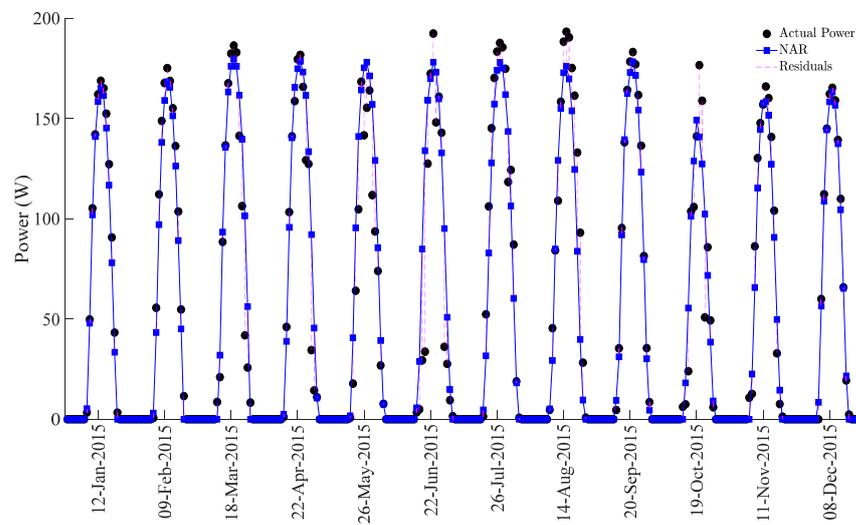


Figure 12. Skill forecasting using a 35% KPI.

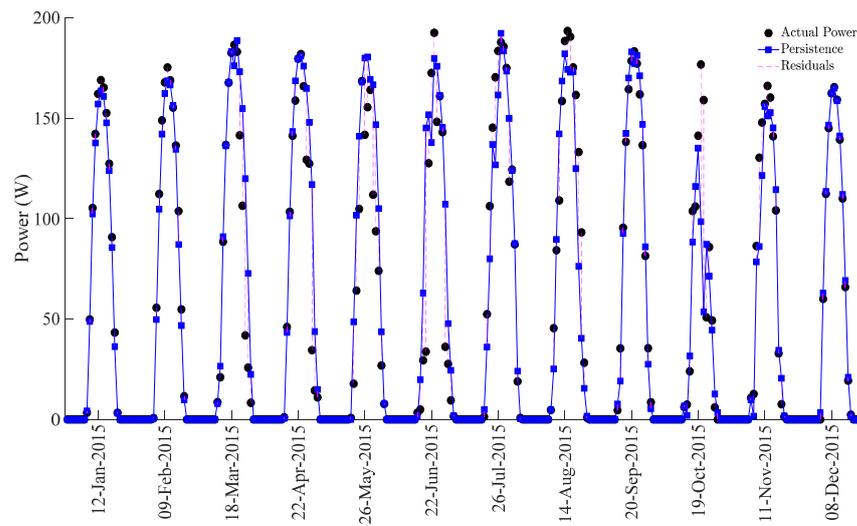


(a) NARX-CVM model



(b) NAR model

Figure 13. Cont.



(c) Persistence model

Figure 13. Comparison of real active power versus forecasting models.

Table 12. Results of performance tests.

Models	MBE	Performance Tests		R ²
		MSE	RMSE	
H-NARX-CVM	0.08	142.59	11.94	0.97
NAR	0.56	220.54	14.85	0.95
Persistence	1.48	330.01	18.17	0.93

Finally, a random day was chosen from the 12 days separated from the annual forecast to make a visual and more detailed analysis of the forecast’s behavior obtained from the H-NARX-CVM, NAR and persistence. In this case, it turned out to be October 19, as is shown in Figure 14. From this figure, the H-NARX-CVM and the NAR models are the best predictors of the behavior of electrical power. It can also be observed that at 3:00 p.m., the electrical power generated by the PV system was 50.84 W. In comparison, the persistence prediction was 87.22 W, the artificial neural network was 102.36 W and 116.58 W by the H-NARX-CVM model.

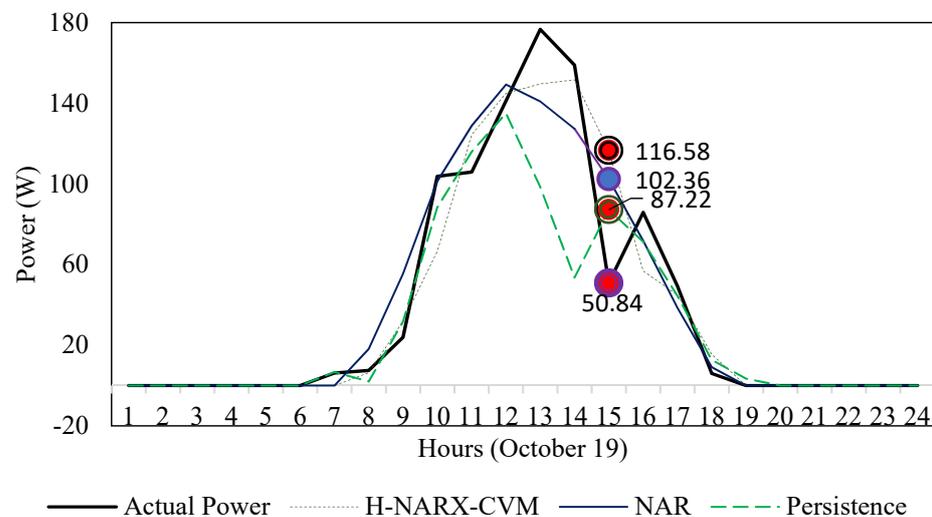


Figure 14. Actual active power model versus the proposed methodology, NAR and persistence models.

According to Figure 14, the best model that predicted the electrical power produced by the photovoltaic system was the NAR model, followed by the proposed methodology. The persistence model had a deficient performance. It is evident that on this particular day (October 19), the NAR model surpassed the NARX-CVM model. However, in general terms, the proposed methodology surpassed the NAR model, thus concluding that its use is more appropriate, as shown in Tables 11–13.

Table 13. Results of the performance tests for October 19.

Models	Performance Tests			R ²
	MBE	MSE	RMSE	
H-NARX-CVM	−0.29	329.36	18.15	0.90
NAR	1.12	291.60	17.08	0.91
Persistence	−6.89	803.77	28.35	0.78

6. Conclusions

In this work, the authors present a methodology to improve and simplify the NARX models, the input vector is the meteorological variable and the output vector is the electric power. The electric power is estimated with the simplified single diode model. The methodology is divided into two parts. The first one focuses on the input vector and implements the collinearity and Granger causality tests to build the input vector from the available variables. The second part focuses on the output vector and implements solar radiation models to build a CVM, which is applied to the output vector to treat with implying atypical results due to the solar radiation behavior. The collinearity test determines which variables are collinear, and the collinear variables are used to form three variable groups. The first group is formed with SR, T and WS, the second one with SR, RH and WS, and the third group is formed with SR, WS and P. The Granger causality test is applied to the three variables' group; this technique determines which variables have useful information to forecast the dependent variable. According to the collinearity and Granger causality tests, the simpler NARX model is when an input vector forms with SR and T in the H-NARX-CVM model is used.

Four NARX models were developed to validate the results with NARX-CVM I–IV. The first uses all meteorological variables as input vectors, and three more are used as input variables: the variables obtained from the collinearity test. The results indicate that the best model was H-NARX-CVM model, obtained with the proposed methodology, demonstrating the importance that the input vector plays in multivariable models. The skill forecasting using a 35% as a goal (KPI) was determined; as a benchmark, we proposed to use the NAR model and the persistence. The skill forecasting results indicate that the worse NARX model, NARX-CVM I, does not overcome the proposed KPI, whereas the NARX-CVM II, III and IV only overcame the KPI, taking it as a benchmark persistence. The H-NARX-CMV obtained from the proposed methodology overcame the proposed KPI with a value of 42%; using persistence as a benchmark and using the NAR model as a benchmark, the model obtains a result of 34%, 1% above the established goal. According to the previous report, we can conclude that even using only the collinearity tests to prove different vectors, we obtained good results; otherwise, it is important to point out that the usage of all variables gives worse results.

Finally, it is important to point out that the development methodology in this work can be applied anywhere. The only requirement is that the site counts with a considerable solar energy resource. However, due to the significant variability of the meteorological variables in different places, the causal relationship between the variables changes. It is necessary to carry out the whole methodology and fix the proposed NARX model.

Author Contributions: Conceptualization, E.C.-C. and C.A.-C.; methodology, E.R.-H., R.C.-A. and C.A.-C.; software, E.R.-H. and E.C.-C.; validation, E.C.-C., R.C.-A. and C.A.-C.; formal analysis, E.C.-C., E.R.-H. and R.C.-A.; investigation, E.C.-C. and E.R.-H.; resources, E.C.-C.; data cleaning, E.R.-H. and R.C.-A.; writing and original draft preparation, E.C.-C., E.R.-H. and R.C.-A.; writing, review and editing, E.C.-C. and R.C.-A.; visualization, R.C.-A.; supervision, C.A.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: The authors appreciate the Instituto de Energías Renovables de la Universidad Nacional Autónoma de México for providing the database through the “Sistema de Información de datos Solarimétricos y Meteorológicos” and sincerely thank Jesús Quiñones for the collaboration.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial neural network
ADF	Augmented Dickey–Fuller test
ACF	Autocorrelation function
CFP	Corrected forecasting power
CVM	Corrective vector multiplier
EP	Electric power
H-NARX-CVM	Híbrido NARX model
KPI	Key performance index
NOCT	Nominal operating cell temperature
NAR	Nonlinear autoregressive
NARX	Nonlinear autoregressive with exogenous inputs
PACF	Partial autocorrelation function
PGM	Photovoltaic generating model
PVG	Photovoltaic generator
PVM	Photovoltaic module
P	Pressure
RH	Relative humidity
SR	Solar radiation
T	Temperature
WS	Wind speed

References

1. Cao, J.; Lin, X. Study of Hourly and Daily Solar Irradiation Forecast Using Diagonal Recurrent Wavelet Neural Networks. *Energy Convers. Manag.* **2008**, *49*, 1396–1406. [\[CrossRef\]](#)
2. Hocaoglu, F.O.; Gerek, Ö.N.; Kurban, M. Hourly Solar Radiation Forecasting Using Optimal Coefficient 2-D Linear Filters and Feed-Forward Neural Networks. *Sol. Energy* **2008**, *82*, 714–726. [\[CrossRef\]](#)
3. Akarslan, E.; Hocaoglu, F.O. A Novel Adaptive Approach for Hourly Solar Radiation Forecasting. *Renew. Energy* **2016**, *87*, 628–633. [\[CrossRef\]](#)
4. Boland, J.; David, M.; Lauret, P. Short Term Solar Radiation Forecasting: Island versus Continental Sites. *Energy* **2016**, *113*, 186–192. [\[CrossRef\]](#)
5. Jiménez-Pérez, P.F.; Mora-López, L. Modeling and Forecasting Hourly Global Solar Radiation Using Clustering and Classification Techniques. *Sol. Energy* **2016**, *135*, 682–691. [\[CrossRef\]](#)
6. Martín, L.; Zarzalejo, L.F.; Polo, J.; Navarro, A.; Marchante, R.; Cony, M. Prediction of Global Solar Irradiance Based on Time Series Analysis: Application to Solar Thermal Power Plants Energy Production Planning. *Sol. Energy* **2010**, *84*, 1772–1781. [\[CrossRef\]](#)
7. Monjoly, S.; André, M.; Calif, R.; Soubdhan, T. Hourly Forecasting of Global Solar Radiation Based on Multiscale Decomposition Methods: A Hybrid Approach. *Energy* **2017**, *119*, 288–298. [\[CrossRef\]](#)
8. Azimi, R.; Ghayekhloo, M.; Ghofrani, M. A Hybrid Method Based on a New Clustering Technique and Multilayer Perceptron Neural Networks for Hourly Solar Radiation Forecasting. *Energy Convers. Manag.* **2016**, *118*, 331–344. [\[CrossRef\]](#)

9. Benmouiza, K.; Cheknane, A. Forecasting Hourly Global Solar Radiation Using Hybrid K-Means and Nonlinear Autoregressive Neural Network Models. *Energy Convers. Manag.* **2013**, *75*, 561–569. [[CrossRef](#)]
10. Chen, C.-R.; Kartini, U. K-Nearest Neighbor Neural Network Models for Very Short-Term Global Solar Irradiance Forecasting Based on Meteorological Data. *Energies* **2017**, *10*, 186. [[CrossRef](#)]
11. Voyant, C.; Muselli, M.; Paoli, C.; Nivet, M.L. Hybrid Methodology for Hourly Global Radiation Forecasting in Mediterranean Area. *Renew. Energy* **2013**, *53*, 1–11. [[CrossRef](#)]
12. Ji, W.; Chee, K.C. Prediction of Hourly Solar Radiation Using a Novel Hybrid Model of ARMA and TDNN. *Sol. Energy* **2011**, *85*, 808–817. [[CrossRef](#)]
13. Renno, C.; Petito, F.; Gatto, A. ANN Model for Predicting the Direct Normal Irradiance and the Global Radiation for a Solar Application to a Residential Building. *J. Clean. Prod.* **2016**, *135*, 1298–1316. [[CrossRef](#)]
14. Kaplanis, S.; Kaplani, E. Stochastic Prediction of Hourly Global Solar Radiation for Patra, Greece. *Appl. Energy* **2010**, *87*, 3748–3758. [[CrossRef](#)]
15. Sfetos, A.; Coonick, A.H. Univariate and Multivariate Forecasting of Hourly Solar Radiation with Artificial Intelligence Techniques. *Sol. Energy* **2000**, *68*, 169–178. [[CrossRef](#)]
16. Zhu, T.; Guo, Y.; Li, Z.; Wang, C. Solar Radiation Prediction Based on Convolution Neural Network and Long Short-Term Memory. *Energies* **2021**, *14*, 8498. [[CrossRef](#)]
17. Zhu, T.; Li, Y.; Li, Z.; Guo, Y.; Ni, C. Inter-Hour Forecast of Solar Radiation Based on Long Short-Term Memory with Attention Mechanism and Genetic Algorithm. *Energies* **2022**, *15*, 1062. [[CrossRef](#)]
18. Hedar, A.R.; Almaraashi, M.; Abdel-Hakim, A.E.; Abdulrahim, M. Hybrid Machine Learning for Solar Radiation Prediction in Reduced Feature Spaces. *Energies* **2021**, *14*, 7970. [[CrossRef](#)]
19. Jo, S.C.; Jin, Y.G.; Yoon, Y.T.; Kim, H.C. Methods for Integrating Extraterrestrial Radiation into Neural Network Models for Day-Ahead Pv Generation Forecasting. *Energies* **2021**, *14*, 2601. [[CrossRef](#)]
20. Hocaoglu, F.O.; Serttas, F. A Novel Hybrid (Mycielski-Markov) Model for Hourly Solar Radiation Forecasting. *Renew. Energy* **2017**, *108*, 635–643. [[CrossRef](#)]
21. Wang, F.; Mi, Z.; Su, S.; Zhao, H. Short-Term Solar Irradiance Forecasting Model Based on Artificial Neural Network Using Statistical Feature Parameters. *Energies* **2012**, *5*, 1355–1370. [[CrossRef](#)]
22. Yadav, A.K.; Malik, H.; Chandel, S.S. Selection of Most Relevant Input Parameters Using WEKA for Artificial Neural Network Based Solar Radiation Prediction Models. *Renew. Sustain. Energy Rev.* **2014**, *31*, 509–519. [[CrossRef](#)]
23. Kaushika, N.D.; Tomar, R.K.; Kaushik, S.C. Artificial Neural Network Model Based on Interrelationship of Direct, Diffuse and Global Solar Radiations. *Sol. Energy* **2014**, *103*, 327–342. [[CrossRef](#)]
24. Cadenas, E.; Rivera, W.; Campos-Amezcuca, R.; Cadenas, R. Wind Speed Forecasting Using the NARX Model, Case: La Mata, Oaxaca, México. *Neural Comput. Appl.* **2015**, *27*, 2417–2428. [[CrossRef](#)]
25. Ahmad, A.; Anderson, T.N.; Lie, T.T. Hourly Global Solar Irradiation Forecasting for New Zealand. *Sol. Energy* **2015**, *122*, 1398–1408. [[CrossRef](#)]
26. Alzahrani, A.; Kimball, J.W.; Dagli, C. Predicting Solar Irradiance Using Time Series Neural Networks. *Procedia Comput. Sci.* **2014**, *36*, 623–628. [[CrossRef](#)]
27. Rangel, E.; Cadenas, E.; Campos-Amezcuca, R.; Tena, J.L. Enhanced Prediction of Solar Radiation Using NARX Models with Corrected Input Vectors. *Energies* **2020**, *13*, 2576. [[CrossRef](#)]
28. Belsley, D.; Kuh, E.; Welsch, R. *Regression Diagnostics—Identifying Influential Data and Sources of Collinearity*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1980.
29. Gujarati, N.D.; Porter, D.C. *Basic Econometrics*, 5th ed.; Anne, E.H., Ed.; Douglas Reiner: New York, NY, USA, 2009.
30. Makridakis, S.G.; Wheelwright, S.C.; Hyndman, R.J. *Forecasting Methods and Applications*, 3rd ed.; Wiley: Hoboken, NJ, USA, 1997.
31. Montgomery, D.C.; Jennings, C.L.; Kulahci, M. *Introduction to Time Series Analysis and Forecasting*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2015.
32. Hocaoglu, F.O.; Karanfil, F. A Time Series-Based Approach for Renewable Energy Modeling. *Renew. Sustain. Energy Rev.* **2013**, *28*, 204–214. [[CrossRef](#)]
33. Salas, V.Á.; Olías, E.; Barrado, A.; Lázaro, A. Review of the Maximum Power Point Tracking Algorithms for Stand-Alone Photovoltaic Systems. *Sol. Energy Mater. Sol. Cells* **2006**, *90*, 1555–1578. [[CrossRef](#)]
34. Skoplaki, E.; Palyvos, J.A. On the Temperature Dependence of Photovoltaic Module Electrical Performance: A Review of Efficiency/Power Correlations. *Sol. Energy* **2009**, *83*, 614–624. [[CrossRef](#)]
35. Farivar, G.; Asaei, B. Photovoltaic Module Single Diode Model Parameters Extraction Based on Manufacturer Datasheet Parameters. In Proceedings of the PECon2010—2010 IEEE International Conference on Power and Energy, Kuala Lumpur, Malaysia, 29 November–1 December 2010; pp. 929–934. [[CrossRef](#)]
36. Castañer, L.; Silvestre, S. *Modelling Photovoltaic Systems Using PSpice*; John Wiley & Sons, Ltd.: Barcelona, Spain, 2002.
37. Tiwari, G.N.; Tiwari, A.; Shyam. *Handbook of Solar Energy*, 1st ed.; Springer: Singapore, 2016.
38. Behar, O.; Khellaf, A.; Mohammedi, K. Comparison of Solar Radiation Models and Their Validation under Algerian Climate—The Case of Direct Irradiance. *Energy Convers. Manag.* **2015**, *98*, 236–251. [[CrossRef](#)]
39. IER-UNAM ELSOMET-IER. Available online: <http://esolmet.ier.unam.mx/index.html> (accessed on 31 December 2016).

40. MONOCRYSTALLINE MODULE ISF-250 BLACK. Available online: http://www.solarypsi.org/repository/documents/445_Second/250-black_usa_.pdf (accessed on 1 August 2021).
41. Aguiar, L.M.; Pereira, B.; Lauret, P.; Díaz, F.; David, M. Combining Solar Irradiance Measurements, Satellite-Derived Data and a Numerical Weather Prediction Model to Improve Intra-Day Solar Forecasting. *Renew. Energy* **2016**, *97*, 599–610. [[CrossRef](#)]
42. Aryaputera, A.W.; Yang, D.; Zhao, L.; Walsh, W.M. Very Short-Term Irradiance Forecasting at Unobserved Locations Using Spatio-Temporal Kriging. *Sol. Energy* **2015**, *122*, 1266–1278. [[CrossRef](#)]
43. Chu, Y.; Pedro, H.T.C.; Coimbra, C.F.M. Hybrid Intra-Hour DNI Forecasts with Sky Image Processing Enhanced by Stochastic Learning. *Sol. Energy* **2013**, *98*, 592–603. [[CrossRef](#)]
44. Huang, R.; Huang, T.; Gadh, R.; Li, N. Solar Generation Prediction Using the ARMA Model in a Laboratory-Level Micro-Grid. In Proceedings of the 2012 IEEE 3rd International Conference on Smart Grid Communications, SmartGridComm, Tainan, Taiwan, 5–8 November 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 528–533. [[CrossRef](#)]
45. Voyant, C.; Muselli, M.; Paoli, C.; Nivet, M.-L. Optimization of an Artificial Neural Network Dedicated to the Multivariate Forecasting of Daily Global Radiation. *Energy* **2011**, *36*, 348–359. [[CrossRef](#)]
46. Benmouiza, K.; Cheknane, A. Small-Scale Solar Radiation Forecasting Using ARMA and Nonlinear Autoregressive Neural Network Models. *Theor. Appl. Climatol.* **2016**, *124*, 945–958. [[CrossRef](#)]
47. García Tena, J.L.; Cadenas Calderón, E.; Rangel Heras, E.; Morales Ontiveros, C. Generating Electrical Demand Time Series Applying SRA Technique to Complement NAR and SARIMA Models. *Energy Effic.* **2019**, *12*, 1751–1769. [[CrossRef](#)]