Energy Reports 12 (2024) 1710–1722

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

Experimental development of a method of short and medium-term photovoltaic generation forecasting using multivariate statistics and mathematical modeling

André Possamai Rosso [a,*], Giuliano Arns Rampinelli [b], Lirio Schaeffer [c]

[a] Federal University of Rio Grande do Sul, Porto Alegre, Brazil, Postgraduate Program in Mining, Metallurgical and Materials Engineering, Antônio Marcelo Ferreira Street 41, Criciúma, Santa Catarina 88810-016, Brazil

[b] Federal University of Santa Catarina, Araranguá, Brazil, Postgraduate Program in Energy and Sustainability, Pedro João Pereira Street 150, Araranguá, Santa Catarina 88905-120, Brazil

[c] Federal University of Rio Grande do Sul, Porto Alegre, Brazil, Postgraduate Program in Mining, Metallurgical and Materials Engineering, Aneron Correa de Oliveira Street 250, Porto Alegre, Rio Grande do Sul 91410-070, Brazil

ARTICLE INFO

Keywords:
Intermittent solar radiation
Forecasting
Photovoltaic
Weather forecast data

ABSTRACT

Uncertainties in photovoltaic solar energy production can make it challenging to dispatch energy into the electricity grid. Although photovoltaic generation storage solves this problem, a forecasting of the photovoltaic solar energy produced is necessary to control the energy injected into the grid. This article aims to develop the probabilistic methodology Reduced-Rank Regression (RRR) for forecasting photovoltaic generation in the short and medium terms. The RRR methodology forecasting uses the generation data of a grid-connected photovoltaic system. The proposed RRR model is simple, easy to access and apply, and does not use irradiance data. The model developed uses the multivariate statistical analysis technique. A advantage is that with a correlation with the performance indices of photovoltaic solar energy systems, the proposed method can be applied in any geographical location on the planet and with different photovoltaic solar energy systems. The application of the RRR methodology requires two searches/inputs. The first input is weather forecast data obtained from a weather forecasting platform, and the second is actual historical data on photovoltaic generation at the site where the method was developed. The proposed method was compared with the persistence method. Using a horizon of 1–10 h, the average monthly root mean square error for the RRR ranged from 7.3 % to 50.1 %. For the persistence method, the average monthly root mean square error ranged from 15.1 % to 65.0 %. Therefore, with the horizon of 24 h, the average monthly root mean square error for the RRR ranged from 4.5 % to 43.2 %. For the persistence method, the average monthly root mean square error ranged from 11.5 % to 75.0 %. We show experimentally that our method is competitive with the state-of-the-art in terms of obtaining photovoltaic generation forecasting without using solar radiation data.

Abbreviations: RRR, Reduced rank regression method; UFSC, Federal University of Santa Catarina; PV, Photovoltaics; DG, Distributed generation PV System; FPV, Forecasting photovoltaic generation; HRT, Time resolution horizon; PF, Final standard value; $\overline{PFP}$, Final profile standard; GHI, Global horizontal irradiance; AG, Actual generation of the photovoltaic system; P, Predicted; $\overline{AG}$, Average actual generation; PM, Persistence Method; NWP, Numerical weather prediction variables; WD, Wavelet decomposition; RMSE, Root mean squared error; MAE, Mean absolute error; MAPE, Mean absolute percent error; MBE, Mean bias error; NMAE, normalized mean absolute error; MeanC, Combined mean; MedianC, Combined median; NRMSE, Normalized root mean squared error; RRMSE, Relative root mean squared error; GA, Genetic algorithms; ANN, Artificial neural networks; AR, Autoregressive; ARMA, Autoregressive moving average; PHANN, Physical-hybrid-artificial neural network; NN, Neural networks; ELM, Extreme learning machine; SVM, Support vector machine; LM, Linear model; LRIC, combined linear regression interacts; LRC, Combined linear regression; LSTMC, Combined long short-term memory; SVRLC, Combined support vector regression with linear Kernel; SVRGC, Combined support vector regression with Gaussian Kernel; YF, Final Yield; $\overline{PFP}$PVS, Final standard future forecasting photovoltaic system profile; $\overline{PFP}$PSD, Final standard profile of the photovoltaic system where the method was implemented; $P_{\text{photovoltaic }PVS}$, Photovoltaic system power from future prediction; $P_{\text{photovoltaic }PSD}$, Power of the photovoltaic system of the developed method.

* Corresponding author.
E-mail addresses: possamairosso@yahoo.com.br (A.P. Rosso), giuliano.rampinelli@ufsc.br (G.A. Rampinelli), schaefer@ufrgs.br (L. Schaeffer).

https://doi.org/10.1016/j.egyr.2024.07.058
Received 3 April 2024; Received in revised form 6 July 2024; Accepted 29 July 2024

1. Introduction

Uncertainties in photovoltaic (PV) generation can make dispatching the energy produced into the electricity grid difficult. Although PV storage solves this problem, it is necessary to forecast the photovoltaic generation (FPV) produced to control the energy injected into the grid (Barbieri et al., 2017; Böök and Lindfors, 2020).

Integrating PV into the electricity matrix still presents some challenges that must be overcome. These are mainly related to the intermittent nature of PV generation (Gong et al., 2023). As PV generation begins to reach a considerable percentage of the electricity matrix, the development of FPVs becomes indispensable (Böök and Lindfors, 2020; Wirth, 2020).

The integration of PV systems requires care in managing and controlling the energy produced. The variability of PV generation results in voltage variations at the entrance to the transmission or distribution network. Voltage variations can damage protection equipment and power quality itself, and they may increase the cost of energy (Barbieri et al., 2017; Böök and Lindfors, 2020).

To maintain the reliability of the electrical distribution network, understanding and planning are required to mitigate the effect of the uncertainties of PV generation and the inherent variability of the system. Power generators and transmission and distribution lines are likely to fail due to the demand for electricity varying throughout the day. Generation variability also becomes a factor with the addition of generation sources such as solar PV. The transmission and distribution system must be prepared and physically capable of responding to this variability in power generation (Dyreson et al., 2014).

The FPV is important for maintaining the stability of electrical grids connected to PVs systems (Zhou et al., 2019). Next-day forecasting is crucial for managing storage systems in PV power plants (Sangrody et al., 2020).

The energy produced by PVs plants depends on a series of meteorological variables, such as solar irradiation, air temperature, cloud cover, wind speed, and relative humidity, among others (Zhang et al., 2021; Luque and Hegedus, 2011). Photovoltaic forecasting is a considerably tricky challenge (Ahmed et al., 2020; Das et al., 2018).

The FPV is the application of several steps with large databases, unreliable measurements, and multiple input-output observations. On the other hand, reliable forecasting makes it possible to offer plant managers/directors reliability. Forecasting minimizes the deviations between programmed energy and energy produced (Antonanzas et al., 2016).

Solar radiation forecasting provides information about the uncertainties in generating electricity from a solar photovoltaic plant. By knowing these uncertainties, it is possible to manage the electricity input into the grid without causing disturbances and power fluctuations (Bakker et al., 2019; Dimd et al., 2023).

Several studies have been carried out on solar photovoltaic energy generation forecasting in recent years. The most economically developed countries, such as Germany, the United States, China, and others, have carried out the most studies in this area (Mosavi et al., 2019; Blaga et al., 2019). Many studies use data from devices and sensors that capture information to forecasting photovoltaic generation, such as meteorological information and solar irradiance (Maciel, 2022; Yang and Dong, 2018).

Given the importance of studying FPV plants in Brazil, the investment situation, and the exponential growth of solar PV energy in the country, there is a need to develop and apply methodologies for forecasting solar PV energy generation in Brazil. The purpose of this article is to present a methodology for forecasting PV generation (RRR) using the generation data from a distributed generation PV system (DG) located at the Federal University of Santa Catarina, UFSC, Campus Araranguá - SC, in the southern region of the state of Santa Catarina, Brazil. The application of the RRR methodology requires two searches/entries. The first input is weather forecast data obtained from a weather forecasting platform, and the second input is actual historical PV generation data at the location where the RRR method was developed. Once the FPV in a given location has been completed, we can use correlations with the performance indices and apply the methodology to carry out the forecasting in any location on the planet for any configuration of PV systems.

The developed model utilizes the technique of multivariate statistical analysis. Multivariate statistical analysis generally involves exploring the variations in a set of interrelated variables or investigating the simultaneous relationships between two or more sets of variables (Reinsel and Velu, 1998). When analyzing weather forecasting, the meteorological platform analyzes data such as atmospheric pressure, rainfall, air humidity, temperature, among others. In the real data of photovoltaic system generation, variables such as solar radiation, temperature, shading, spectral mismatch, dirt, mismatch losses, cabling losses, among others, are embedded.

The classic multivariate regression model does not make direct use of the fact that the response variables are probably correlated. A more serious practical concern is that even for a moderate number of variables whose interrelationships are to be investigated, the number of parameters in the regression matrix can be large. Thus, in many practical situations, there is a need to reduce the number of parameters in the model and this problem can be addressed through the assumption of reduced rank of the matrix (Reinsel and Velu, 1998). The model developed in the article has some characteristics and adaptations of Reduced-Rank Regression.

The RRR model offers several advantages over other models by using the production data from a photovoltaic solar energy system instead of irradiance data in photovoltaic production forecasts. In the RRR model, due to the use of generation data for forecasting, some parameters are already incorporated and do not need to be dealt with, such as the soiling of the photovoltaic modules, shading, the spectral mismatch of the photovoltaic cells with the radiation spectrum, losses due to mismatch, losses due to cabling, losses due to series and parallel resistance, among others.

1.1. Forecasting method categories and models

Numerous PV forecasting methods using solar radiation forecasting have been developed in the literature (Table 1). According to (Diagne et al., 2013), the existing global horizontal irradiance (GHI) forecasting methods and models can be classified according to the input data, which also determines the forecasting horizon. The main methods used are the

Table 1
Methods and models used to forecasting global horizontal irradiance (GHI).

Method	Subcategory	Model	Time Scale
Persistence		Global irradiance at time t+1 best predicted by its value at time t ($\hat{Z}_{t+1} = Z_t$).	Short-term accuracy
Statistic	Serie Temporal Model	Autoregressive(AR)	Short-term accuracy
		Autoregressive moving average (ARMA)	Short-term accuracy
Statistic	Neural Network	Artificial Neural Networks (ANNs)	Short-term accuracy
Based on satellite data/ images	Geoestationary Satellite Image	Image processing	Good performance in intervals from 30 min to 6 h
Based on sky images	Total sky image	Image processing	Good performance in sub-hours interval
Basedo on numerical weather models	Numerical weather prediction variables (NWP)	Global Forecast System	Accurate for longer predictions

persistence method, the satellite data/image method, the numerical weather forecasting method, the sky image method, and statistical methods. An overview of the methods is shown in Table 1.

The main factors affecting the performance of prognostics used in PV forecasting are the time horizon and time resolution, weather conditions, geographical location, and the availability and quality of data (Nespoli et al., 2019).

About these parameters, different prognoses can be chosen for a particular need: physical methods (Wolff et al., 2016; Dolara et al., 2015), which are mainly based on the use of numerical weather forecasting (NWP) models (Larson et al., 2016; Zhang et al., 2019), or satellite images that can be used to develop regional models (Jang et al., 2016). In this case, meteorological data from satellite images is often used for long-term forecasting (Pelland et al., 2013). NWP models are widely used to forecasting the state of the atmosphere for the next 15 days and do not require historical data (Pelland et al., 2013).

These methods can provide good precision but depend mainly on the stability of the climatic conditions. However, the implementation of physical models is generally relatively difficult, as it requires a series of parameters and expensive equipment that is not always available in many areas of the world. In addition, for most of the NWPs available, the first few hours of forecasts are not particularly useful for solar forecasting (Tuohy et al., 2015).

1.2. Photovoltaic generation forecast studies

In (Mellit and Kalogirou, 2008) a review was carried out on a series of techniques that use artificial neural networks (ANNs) to forecasting the power produced by photovoltaic systems. The authors showed that techniques based on artificial intelligence have great potential for estimating photovoltaic power. Forecasting based on recurrent neural networks (RNNs) have been recognized as the most accurate.

In (Antonanzas et al., 2016), regression techniques and methods based on artificial intelligence were included. The authors found that statistical approaches perform better than parametric approaches. The authors mentioned that the latest techniques use linear model (LM) methods, including support vector machine (SVM), extreme learning machine (ELM), and so on. These techniques allow for easy modeling without the need to know the characteristics of the photovoltaic system.

In (Raza et al., 2016), a complete review was provided, including time series, ANNs, and some hybrid approaches. A comparison between ANNs-based models and classical time series models was also presented. The conclusions were that forecasting and accuracy can be improved by pre and post-processing the historical data. ANNs performed better than other classical time series approaches.

In (Sobri et al., 2018), it was concluded that ANN and SVM-based methods are widely used due to their ability to solve complex, non-linear forecasting problems. Ensemble methods capable of improving forecasting accuracy have been found; they can merge linear and non-linear methods.

A comparative study between physical and hybrid methods for photovoltaic day-ahead forecasting was presented in (Ogliari et al., 2017). The conclusion was that physical-hybrid-artificial neural networks (PHANN) always show the highest accuracy. Using the PHANN model, the authors obtained a normalized mean absolute error (NMAE) of 5.6 %.

In (Ahmed et al., 2020) a general review of recent studies on direct short-term forecasting methods for photovoltaic generation based on historical data was carried out. The authors pointed out that forecasting models based on ANN and SVM perform well under fast and varied environmental conditions. The optimized algorithm significantly increased forecasting accuracy. Genetic algorithms (GA) represented one of the most viable optimization techniques for photovoltaic energy forecasting.

In (Babalhavaeji et al., 2023), the authors explored how both spatial and temporal information can be considered through a deep learning approach. The authors proposed a photovoltaic generation forecasting that considers spatial and temporal information. A convolutional neural network is used as a pre-processing step to capture spatial information. The convolutional neural network is followed by a gated recurrent unit neural network to model temporal characteristics. The proposed model can forecast a horizon for which there is no available information on irradiance, humidity, or wind. The authors experimentally demonstrated that the method is competitive with the state of the art in terms of time and memory, resulting in better forecasting performance.

In (Dewangan et al., 2020), different combined photovoltaic day-ahead forecasting methods were explored for three plants located in Australia. Photovoltaic day-ahead forecasting is the most relevant when studying planning and strategies for electricity dispatch. The authors used twelve NWP variables and energy time series recorded from April 1, 2012 to June 30, 2014. The authors pointed out that combined forecasting are used when it is difficult to determine the best forecasting model among several forecasting models. Different forecasting models have different learning methods. Thus, the forecasting outputs have different information, which can be combined to obtain an accurate model. Candidate forecasting models must have satisfactory accuracy to participate in combined forecasting, and poorly performing forecasting models should not be considered when forming the combined forecasting model. Combined forecasting methods include simple averages with equal weights, median, linear, and non-linear regression. The authors obtained the best mean square error (RMSE) of 7.79 % for the combined Average C model at plant 1 for the first month. At plant 2, the best RMSE was 15.36 % for the combined linear regression interacts combined (LRIC) model. Plant 3 showed the combined linear regression (LRC) model as the best, with an RMSE of 8.71 %. The second month's best RMSE was 7.37 % for plant 2 using the combined support vector regression with linear kernel (SVRLC) model. For plant 1, the best RMSE was 9.80 % using the combined median (MedianC) model. For plant 3, the best RMSE was 9.19 % using the combined long short-term memory (LSTMC) model. In the third month, the best RMSE was 10.72 % using the combined LSTMC model for plant 3. Plant 1 had the best RMSE, 12.11 %, using the combined support vector regression with Gaussian (SVRGC) model. At plant 3, the best RMSE was 10.72 % for the LSTMC model.

In (Zhu et al., 2015), a hybrid method for forecasting photovoltaic generation was presented that combined the advantages of wavelet decomposition (WD) and artificial neural networks (ANN). The authors used theoretical solar irradiance and meteorological variables as input for the hybrid model based on WD and ANN. The output power of the photovoltaic plant was decomposed using WD to separate the helpful information from the disturbances. ANNs are used to build the models of the decomposed photovoltaic output power. Finally, the outputs of the ANNs models are reconstructed into the predicted power of the photovoltaic plant. The method presented is compared with the traditional ANN-based forecasting method. The results show that the method described in this paper requires less calculation time and has better forecasting accuracy. On clear days, the WD + ANN hybrid model showed an RMSE of 7.19 %, while the ANN model showed 9.31 %. The WD + ANN showed an RMSE of 19.66 % for rainy days, while the ANN showed 22.95 %. On cloudy days, the WD + ANN model showed an RMSE of 16.82 %, while the ANN method showed 18.51 %.

1.3. Persistence method

The persistence methods (PMs) are more straightforward forecasting methods and are used as a reference for more developed models. Classical PMs assume that the conditions (solar irradiance, generation, clear sky index, etc.) remain the same between the current time t and $t + fh$. The persistence model is recommended when the time series is stationary (Antonanzas et al., 2016). The forecasting of generation at time t is given by Eq. 1 (Antonanzas et al., 2016):

$$P_p\,(t + f_h) = P(t) \quad (1)$$

PMs have some desirable characteristics, including (i) simplicity, a method that does not require any training or intelligence; (ii) speed, a computational method that is quick to implement and easy to operate; and (iii) repeatability, a method that consistently produces an expected result for the same input (Maciel, 2022).

1.4. Contributions

This article aims to achieve high accuracy in forecasting short-and medium-term photovoltaic generation. We have made the following contributions.

1. A new photovoltaic forecasting methodology is proposed based on just two searches/inputs. The first input is weather forecast data from meteorological platforms, and the second is actual historical data on photovoltaic generation at the site where the methodology was developed.
2. The proposed RRR model is simple, easy to access and apply, and does not use irradiance data. Once the photovoltaic forecasting in one location has been completed, we can use correlations with the performance indices and apply the methodology to carry out the forecasting in any location on the planet and in any configuration of photovoltaic systems.
3. Some parameters are already built into the model and do not need to be dealt with, such as soiling of the photovoltaic modules, shading, spectral mismatch of the photovoltaic cells with the radiation spectrum, losses due to mismatch, losses due to cabling, losses due to series and parallel resistance, among others.

2. Methodology

This section refers to the development of the photovoltaic generation forecasting methodology (RRR) using the grid-connected photovoltaic system of Federal University of Santa Catarina (UFSC) and the Ventusky weather forecast platform. The grid-connected photovoltaic system used in this article for the development of the method is located at UFSC, whose campus is in the city of Araranguá, SC, latitude: 28°56'S and longitude: 49°29'W.

2.1. Distributed generation photovoltaic systems at the Federal University of Santa Catarina, Araranguá – SC

The photovoltaic system at UFSC – Araranguá Campus consists of 3 photovoltaic modules model 330PHK-36 from the manufacturer BYD with an individual nominal power of 330 Wp and 1 inverter model PHB1500-NS from the manufacturer PHB Solar with a nominal power of 1.5 kW. The system's power output is 0.99 kWp.

The photovoltaic array is installed on the roof of a bioclimatic pilot plant located at the Center for Science, Technology and Health at UFSC. The system data is shown in Table 2.

Table 2
UFSC photovoltaic system specifications – Araranguá – SC.

SFGD Specifications	
Nominal system power	0,99 kWp
Photovoltaic arrangement	1 *string* - 3 module
Tilt angle	20°
Azimuthal deviation (module orientation)	0° N
Module manufacturer/model	BYD/330PHK−36
Unit power of the modules	330 Wp
Invertor	1x PHB1500-NS
Nominal power of the invertor	1,5 kW

2.2. Reduced-rank regression methodology for forecasting photovoltaic generation

The proposed methodology is based on a relationship with a meteorological weather forecasting platform and photovoltaic generation data from the location where the forecasting is to be made. It was developed at UFSC's distributed generation photovoltaic systems, Araranguá Campus.

The RRR methodology forecasts future photovoltaic generation in a short and medium-term time resolution horizons (HRT). The short-term HRT is considered to be a forecasting of 1–10 h in the future (daily). The medium-term HRT forecasts 24 h in the future (the next day). The application of the RRR is simple and only requires two searches/entries. The first input is weather forecast data from the weather forecasting platform, and the second is a final standard value (PF) obtained from real historical photovoltaic generation data from the site where the RRR method was developed.

Using the PF obtained from historical data and a weather forecasting network platform, it is possible to predict future generation for the region in which the photovoltaic system is installed for each type of weather forecast available on the platform. The first step in forecasting photovoltaic generation is to obtain the PF value for each month of the year and for each type of weather forecast from a given search platform. For a preliminary analysis of the methodology, the weather forecasting platform used is Ventusky.

2.2.1. Ventusky platform

The Ventusky platform was developed at InMeteo. On it, weather forecasts were extracted every day. On the day of the search, the forecast was extracted from 8 a.m. until 6 p.m. For the following day (24 h ahead), the forecast was extracted every three hours (6 a.m., 9 a.m., 12 p.m., 3 p.m., and 6 p.m.) because the platform does not provide hourly data. The weather forecasts provided by the platform were assigned as sunny, sunny with a few clouds, sunny with few clouds, sunny with many clouds, sunny and rain showers, sunny and cirrus, foggy, totally cloudy, rainy, heavy rain, rain and thunderstorms, moon, moon with few clouds and moon with a few clouds.

2.2.2. Obtaining the final standard value of forecasting for different weather forecasts using the Ventusky platform

Based on the generation curves (Fig. 1) and the weather forecasts obtained using the Ventusky platform, a final standard value (PF) of forecasting was obtained for each interval (time) between 6 a.m. and 6 p.m. The hourly PF value was also obtained for each month of the year. The first PF forecasting value found was for all the months of the year with a forecast of sunny days (clear skies). The other PF values with different day forecasts were obtained in the same way as for sunny days. From the sunny generation curve for each month of the year, the generation was integrated for each hourly interval, including 6 a.m. to 6 p.m. The generation curves for rain and thunderstorm forecasts were not obtained. In the period analyzed, from March 2022 to February 2023, it was not possible to obtain a complete day with this forecast. Fig. 2 shows the diagram representing the steps to obtain the forecasting PF value.

2.2.3. Survey of the final profile standard ($\overline{\text{PFP}}$) used to forecasting RRR generation using the Ventusky platform

Because the Ventusky weather forecasting platform has a very broad legend for each type of day, the platform's weather forecast types have been divided into three profiles (Table 3).

The profiles shown in Table 3 were used to obtain the values for each profile (sunny, intermediate, and rainy). However, photovoltaic generation forecasting are obtained for each profile. This is obtained using Eq. 2, by averaging the PFs of the types of forecasting for each profile. The photovoltaic generation forecastings for the city of Araranguá-SC were made using each profile, using the Ventusky platform.

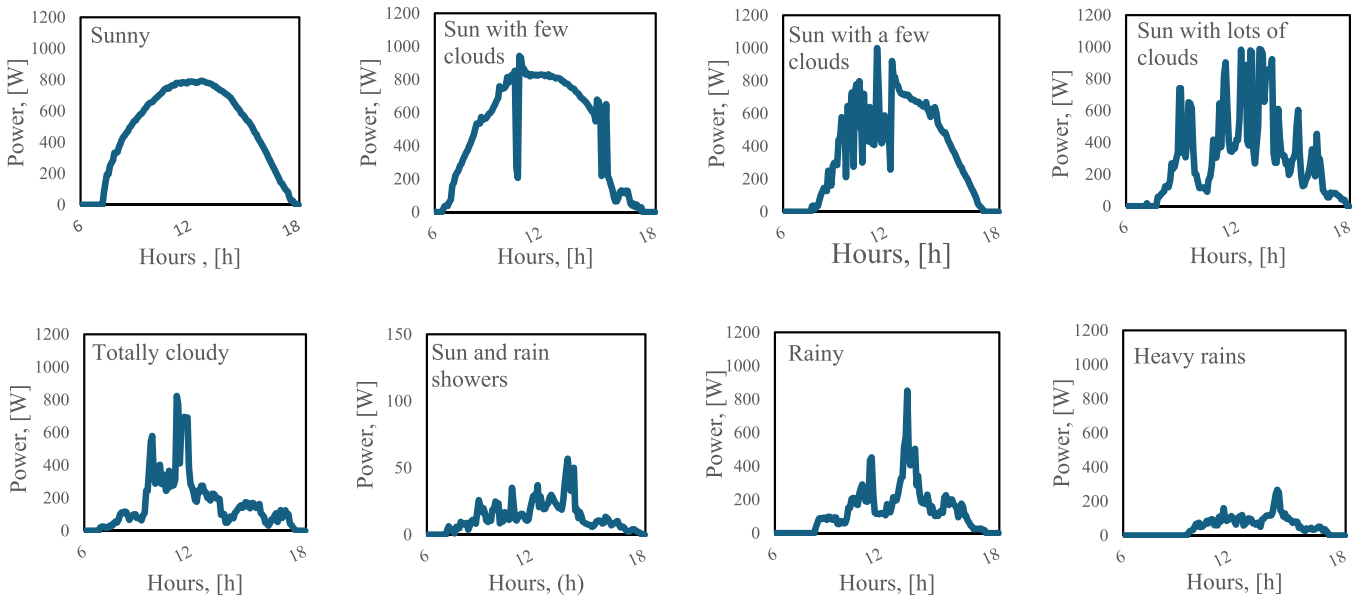

Fig. 1. Photovoltaic power curves of the UFSC photovoltaic system, Araranguá – SC.

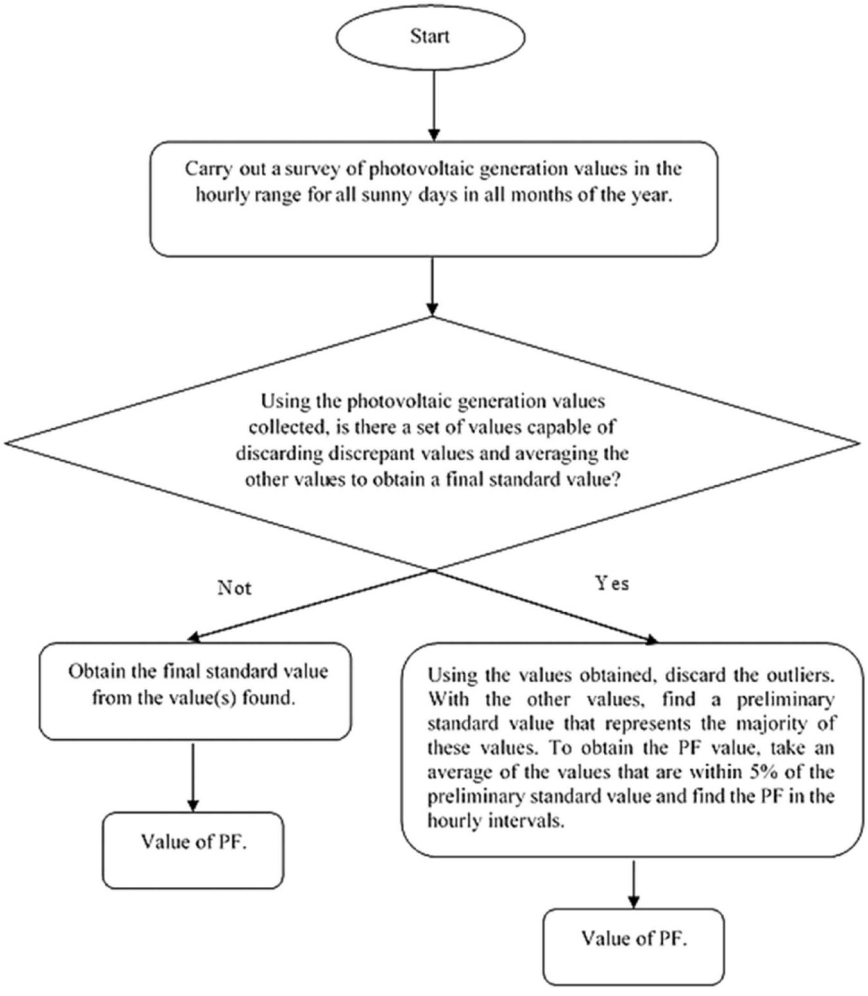

Fig. 2. Flowchart for obtaining the final forecasting standard value for each month of the year and for each weather forecast.

Table 3
Weather forecast profiles obtained from the Ventusky platform.

Profiles	Types of Ventusky Forecasts
Sunny	sunny, sun with few clouds, sun with a few clouds and sun and cirrus clouds
Intermediate	sunny with lots of clouds, sunny with rain showers and totally cloudy
Rainy	rain, showers and thunderstorms and heavy rains

$$\overline{PFP} = \frac{1}{n}\sum_{1}^{n} PF_i \tag{2}$$

After determining the average standard value of the final profile ($\overline{\text{PFP}}$) at the location of the photovoltaic system where the RRR methodology was applied, the ($\overline{\text{PFP}}$) can be correlated with the final yield (Y_F) to obtain forecasting in other geographical regions.

The first step is to find the value of the final standard profile of the photovoltaic system ($\overline{\text{PFP}}_{\text{PVS}}$) that will be used for the photovoltaic forecast at the other location. The value of ($\overline{\text{PFP}}_{\text{PVS}}$) is determined using the value of the average of the final profile standard of the photovoltaic system where the method was implemented ($\overline{\text{PFP}}_{\text{PSD}}$) and the power of both photovoltaic systems (Eq. 3).

$$\overline{PFP}_{PVS} = \frac{\overline{PFP}_{PSD} \quad x \quad P_{\text{photovoltaic}PVS}}{P_{\text{photovoltaic}PSD}} \tag{3}$$

The second step is to use the Y_F from the previous year at the location where the method was developed in the month for which the forecasting is to be determined. With this Y_F and the Y_F of the location for which the forecasting is to be determined, the ($\overline{\text{PFP}}_{\text{SPV}}$) of the site to be forecasted for photovoltaic generation is determined (Eq. 4).

$$\overline{PFP}_{PVS} = \frac{\overline{PFP}_{PVS} \quad x \quad Y_{F\ (PVS)}}{Y_{F\ (PSD)}} \tag{4}$$

The ($\overline{\text{PFP}}_{\text{PVS}}$) can be determined for hourly intervals of each month and for each profile of meteorological data.

2.2.4. Horizon of time resolution

The main factors affecting performance in forecasting photovoltaic generation are the time-horizon, time resolution (HRT), weather conditions, geographical location, availability, and quality of data (Nespoli et al., 2019).

The time resolution horizon used in this work is for short and medium-term forecasting. The short-term horizon can vary from 1 min to 24 h. The medium-term horizon can vary from 24 h to 1 week. In this study, a forecasting horizon of the day (1–10 h in the future) and 24 h in the future (next day) is being adopted. In the 1–10 h future horizon, generation forecastings were made between 8 a.m. and 6 p.m. at hourly intervals. In the 24 h future horizon, generation forecastings were made between 6 a.m. and 6 p.m. in 3 h intervals because the Ventusky platform offers weather forecasts in these intervals.

2.2.5. Statistical metrics for evaluating the performance of the proposed method

In (Gueymard, 2014), a detailed review was carried out on the main statistical metrics used for solar radiation forecasting. According to (Barbieri et al., 2017), a forecasting model's performance is often evaluated by root mean square error (RMSE), normalized root mean square error (nRMSE), relative root mean square error (rRMSE), mean absolute error (MAE), or mean bias error (MBE), among other means.

The root mean square error (RMSE), given by Eq. 5, is a quadratic scoring rule that estimates the average magnitude of error. It is the most standard function used to calculate the difference between predicted and observed values, since it reflects the level of differences between the actual and forecasted values (Piotrowski et al., 2022). The mean absolute error (MAE), Eq. 6, corresponds to the estimated level of absolute error. This level indicates the average magnitude of the actual value and the predicted value (Yildiz et al., 2021). The mean absolute percentage error (MAPE), Eq. 7, calculates the percentage error relative to the actual value, which is stated as the average ratio, and is also commonly used to compare different models (Zheng and Wu, 2019). The metrics used in this work are presented in Eqs. 5, 6, and 7.

$$RMSE[\%] = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{n}(AG - \quad P)^2}}{\overline{\text{AG}}}.100 \tag{5}$$

$$MAE[\%] = \frac{\frac{1}{N}\sum_{i-1}^{N}|AG - \quad P|}{\overline{\text{AG}}}.100 \tag{6}$$

$$MAPE[\%] = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{AG - \quad P}{AG}\right|.100 \tag{7}$$

In which: $AG = actualgeneration; P = predictedgeneration; \overline{AG} =$ average actual generation.

The errors used in the article were calculated in percentages. RMSE, MAE, and MAPE were obtained at hourly intervals from 8 a.m. to 6 p.m. for the three profiles with an HRT from 1 to 10 h in the future. Annual average errors were also obtained for the three profiles using RMSE, MAE, and MAPE. The monthly average error was obtained using RMSE, MAE, and MAPE for all the months in the interval analyzed. For the 24-h HRT, errors were found at intervals of 6–7 a.m., 9–10 a.m., 12–1 p.m., 3–4 p.m., and 6–7 p.m. The results were compared with PM.

3. Results and discussions

This section presents the results obtained from the proposed method, as well as relevant discussions about them.

3.1. Analysis of hourly photovoltaic generation forecasting for the profiles developed using horizon of time resolution of 1–10 h

Table 4 shows the hourly RMSE, MAE and MAPE values using RRR and PM for the sunny, intermediate, and rainy profiles. The RRR method showed lower RMSE, MAE and MAPE values than the PM method at all time intervals in the sunny, intermediate, and rainy profiles.

The sunny profile showed lower RMSE, MAE, and MAPE than the other three profiles. The sunny profile has the characteristic of not varying solar radiation over a short period of time. The RMSE for the RRR method was between 20.2 % and 28.4 %. The PM showed RMSE values of between 36.8 % and 57.7 %. In the sunny profile, the times with the highest solar radiation had the lowest RMSE, MAE, and MAPE. The smallest errors were between the 9–10 a.m. and 1–2 p.m. intervals.

The intermediate profile showed the highest RMSE, MAE and MAPE compared to the 3 profiles. This is the profile with the greatest challenge in terms of photovoltaic forecasting accuracy. It is the most challenging due to the increase in cloud circulations. The RMSE for the RRR method was between 45.0 % and 63.5 %. PM showed RMSE values between 76.6 % and 95.6 %. The lowest RMSE was between 4 p.m. and 5 p.m. MAE and MAPE followed the same profile as RMSE, maintaining lower values for the RRR method when compared to PM.

The rainy profile showed RMSE values between the sunny and intermediate profiles in most intervals. In the 11 a.m.–12 p.m., 2–3 p.m., and 3–4 p.m. intervals, the rainy profile showed lower RMSE than the sunny profile. The RMSE for the RRR method ranged from 21.5 % to 154.8 %. The PM showed RMSE values between 114.7 % and 263.5 %. The lowest RMSE was between 11 a.m. and 12 p.m. The MAE and MAPE followed the same profile as the RMSE, maintaining lower values for the RRR method when compared to the PM.

The RRR method showed more significant errors between 5 p.m. and 6 p.m. in the sunny and rainy profiles. These errors do not have a relative energetic impact because this interval has low irradiation. In the

Table 4
Errors in the three profiles at hourly intervals for the months of March 2022 to February 2023, using the RRR and MP methods.

	8–9 am	9–10 am	10–11 am	11–12 am	12–1 pm	1–2 pm	2–3 pm	3–4 pm	4–5 pm	5–6 pm
RMSE [%] - SUNNY										
RRR	25,9	22,8	20,8	23,3	20,2	23,7	25,5	26,7	27,8	28,4
PM	41,9	42,7	36,8	38,2	38,7	37,0	40,1	41,6	50,6	57,7
MAE [%] - SUNNY										
RRR	14,9	13,2	12,1	12,7	10,9	13,7	14,0	15,6	17,2	20,4
PM	28,8	28,7	24,7	26,0	25,8	23,6	26,0	26,3	34,5	36,9
MAPE [%] - SUNNY										
RRR	46,1	25,6	20,3	24,9	19,6	24,0	38,3	36,4	38,3	33,6
PM	68,0	71,0	29,8	30,2	31,5	27,0	45,9	44,1	51,6	48,9
RMSE [%] - INTERMEDIATE										
RRR	49,0	50,7	62,9	57,6	58,8	56,5	63,5	56,4	45,0	50,1
PM	85,1	85,9	86,1	80,4	79,5	95,6	91,1	81,0	82,7	76,6
MAE [%] - INTERMEDIATE										
RRR	34,8	40,6	47,0	43,5	42,0	39,9	45,5	41,7	35,4	32,0
PM	81,1	75,7	68,9	63,6	70,0	80,2	73,1	67,6	69,8	52,5
MAPE [%] - INTERMEDIATE										
RRR	48,1	44,2	51,8	57,6	55,0	43,6	76,6	67,9	67,5	84,4
PM	120,3	113,7	105,8	95,1	136,1	120,3	129,9	110,0	260,3	128,5
RMSE [%] - RAINY										
RRR	33,9	33,8	31,7	21,5	63,1	51,9	21,9	18,8	86,9	154,8
PM	123,1	168,1	172,3	128,9	119,8	126,7	114,7	128,4	263,5	177,4
MAE [%] - RAINY										
RRR	21,4	21,7	18,7	12,0	32,1	24,8	13,3	10,8	53,2	75,9
PM	102,3	383,9	133,7	102,6	82,9	105,8	87,6	110,2	211,2	144,1
MAPE [%] - RAINY										
RRR	32,6	37,3	27,6	16,8	30,8	23,6	27,0	17,0	170,3	241,8
PM	144,0	167,8	183,7	134,2	137,4	243,3	187,7	274,7	306,4	375,5

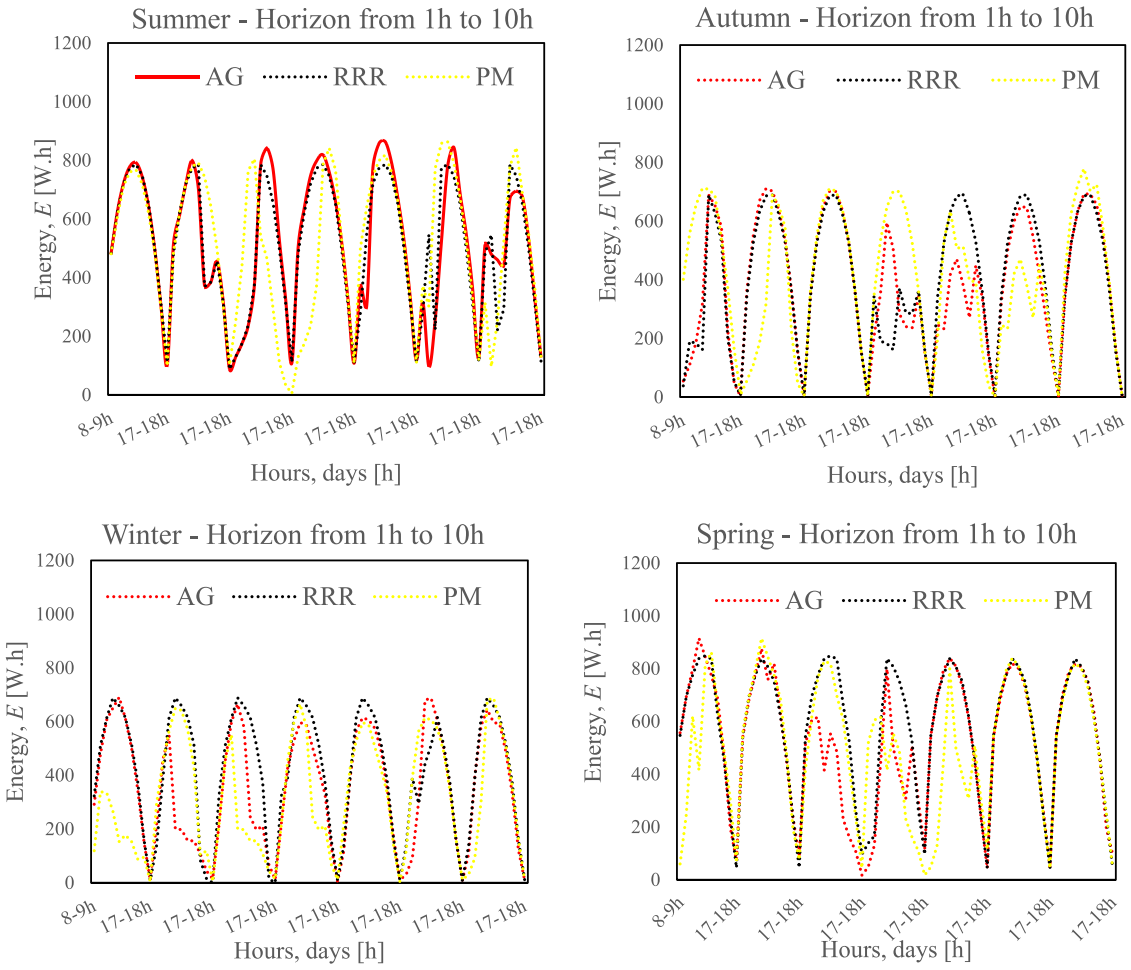

Fig. 3. Graph of RRR and MP forecasting with the 7-day GR for the seasons.

intermediate profile, the errors in the hourly intervals with the lowest irradiation were close to those of the other intervals. The RRR method proved to be a more effective and accurate model for the sunny and rainy profiles.

Fig. 3 also shows the performance of the photovoltaic generation forecast between the RRR and PM methods. The considered horizon was only from HRT = 1–10 h. Seven consecutive days were analyzed for each season, including the three weather forecast profiles. Prediction using the RRR method was compared with actual generation of the photovoltaic system (AG) and forecasting using the PM method. The performance graphs visually show that RRR performed better than PM in all seasons. In the summer season, the RRR method performed well over the seven consecutive days. The PM method showed poorer performance compared to both RRR and the actual generation. In the autumn season, the RRR method performed well on the first three days, as well as on the sixth and seventh days. On the fourth and fifth days, the RRR method's performance dropped due to the Ventusky platform not being as accurate. The PM method showed poorer performance compared to both RRR and the AG. In the winter season, the RRR method showed poorer performance on the second, third, fourth, fifth,

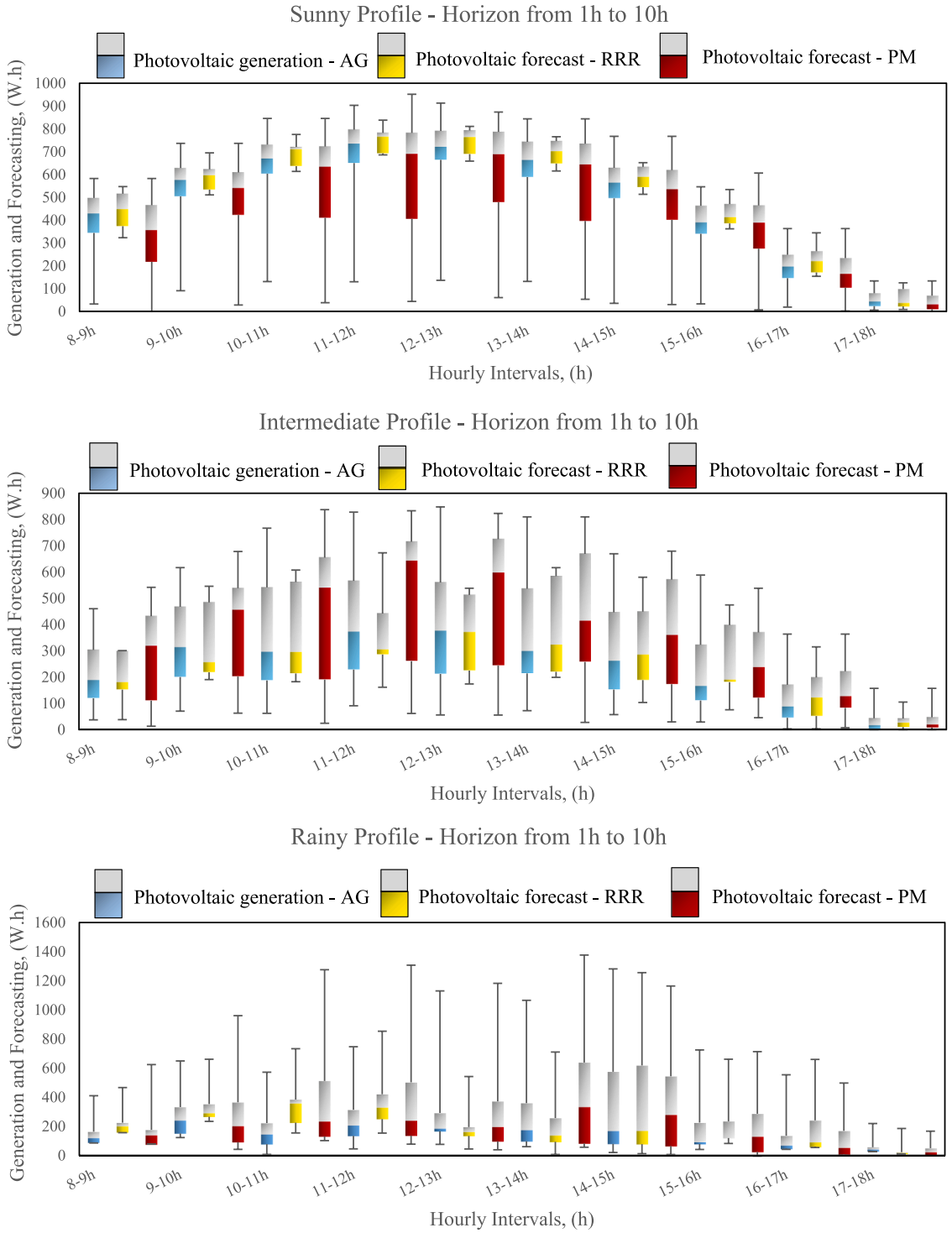

Fig. 4. Box Plot of the generation and forecasting for the profiles in the hourly intervals from March 2022 to February 2023 for an HRT = 1–10 h.

and sixth days. On the first and seventh days, the RRR method performed better. During the winter season, the consecutive days selected exhibited considerable instability, which posed challenges for photovoltaic forecasting. In the spring season, the RRR method performed well on the first, second, fifth, sixth, and seventh days. On the third and fourth days, the RRR method showed poorer performance. The PM method performed worse compared to both RRR and the AG on the first, second, and fourth days.

The forecasting of photovoltaic generation using the two forecasting methods shows more significant uncertainty on days with intermediate and rainy profiles in all seasons. Fig. 4 shows that RRR forecasting performs better than PM in all seasons, regardless of the profile. Fig. 4 shows that the RRR forecasting had a smaller amplitude when compared to the generation and PM forecasting for the sunny, intermediate, and rainy profiles.

In the sunny profile, the smaller amplitude is due to the values being close throughout the year for each hourly interval. The RRR showed less annual hourly variation in forecasting than the PM. RRR showed a higher correlation with generation at all intervals than PM.

For the intermediate profile (Fig. 4), the RRR method showed a lower annual hourly forecasting variation when compared to PM and a higher correlation with generation at all intervals when compared to PM.

In the rainy season profile (Fig. 4), the RRR method also showed a lower annual hourly forecasting variation than the PM. The RRR showed a higher correlation with generation at all intervals than the PM.

3.2. Analysis of hourly photovoltaic generation forecasting for the profiles developed using horizon of time resolution of 24 h

Table 5 shows the hourly RMSE, MAE, and MAPE values using RRR and PM for the sunny, intermediate, and rainy profiles. No photovoltaic generation was detected for the rainy profile in the 6–7 p.m. time interval. For HRT = 24 h, the RRR method also showed lower RMSE, MAE, and MAPE values than the PM method in all hourly intervals in the sunny, intermediate, and rainy profiles.

Table 5
Errors in the three profiles at hourly intervals for the months of March 2022 to February 2023, using the RRR and PM methods.

	6–7 am	9–10 am	12 am-1 pm	3–4 pm	6–7 pm
SUNNY					
RMSE [%]					
RRR	42,1	21,1	23,5	26,8	65,0
PM	75,7	38,9	40,1	38,8	67,1
MAE [%]					
RRR	26,1	12,4	13,1	15,9	48,1
PM	54,7	24,7	26,4	25,7	53,2
MAPE [%]					
RRR	52,8	22,3	22,7	29,1	70,5
PM	72,0	25,5	32,1	30,2	70,8
INTERMEDIATE					
RMSE [%]					
RRR	104,1	57,3	54,6	66,3	89,1
PM	122,4	74,9	77,6	81,1	119,5
MAE [%]					
RRR	81,1	44,0	39,8	52,4	54,1
PM	94,1	62,8	59,6	63,4	95,0
MAPE [%]					
RRR	109,3	74,4	59,8	95,7	19,2
PM	168,4	129,8	144,2	147,1	69,4
RAINY					
RMSE [%]					
RRR	87,0	30,6	34,9	73,0	-
PM	297,2	138,5	172,7	256,6	-
MAE [%]					
RRR	66,3	19,7	26,8	48,4	-
PM	216,3	116,3	147,8	202,1	-
MAPE [%]					
RRR	64,3	28,0	26,7	62,2	-
PM	151,6	400,7	175,6	386,4	-

The sunny profile again showed the lowest RMSE, MAE, and MAPE compared to the three profiles. The RMSE for the RRR method ranged from 21.1 % to 65.0 %. The PM showed RMSE values between 38.8 % and 75.7 %. Again, for the sunny profile, the times with the highest solar radiation had the lowest RMSE, MAE, and MAPE. The smallest errors were between the 9–10 a.m. and 12–1 p.m. intervals.

The intermediate profile showed higher RMSE, MAE, and MAPE than the three profiles. Only in the 3–4 p.m. interval the RRR method showed lower RMSE and MAPE than the rainy profile. The RMSE for the RRR method ranged from 54.6 % to 104.1 %. The PM showed RMSE values between 74.9 % and 122.4 %. The lowest RMSE was in the interval with the highest solar radiation between 12 p.m. and 1 p.m. The MAE and MAPE followed the same profile as the RMSE, maintaining lower values for the RRR method when compared to the PM.

Again, for HRT = 24 h, the rainy profile showed RMSE values between the sunny and intermediate profiles in most intervals. In the 3–4 p.m. interval, the rainy profile had a higher RMSE than the sunny and intermediate profiles. The RMSE for the RRR method ranged from 30.6 % to 87.0 %. The PM showed RMSE values between 138.5 % and 297.2 %. The lowest RMSE was between 9 a.m. and 10 a.m. The RMSE, MAE, and MAPE had lower values than the PM.

The future 24-h horizon had higher errors compared to the 1–10 h horizon. The RMSE for the RRR method was 0.9–36.6 % higher. The MAE for the RRR method was 1.5–27.7 % higher, and the MAPE was 2.7–24.4 % higher.

The RRR method showed greater errors in the interval between 5 p.m. and 6 p.m. in the sunny profile and between 6 a.m. and 7 a.m. for the intermediate and rainy profiles. Again, the larger errors in these intervals do not have a relative energy impact because these intervals have low photovoltaic generation.

RRR for the following day proved to be a much more efficient method than PM in the rainy profile. In the intermediate profile, RRR was closer to PM. In the sunny profile, the PM method is more assertive due to the lower intensity of cloud cover. However, the RRR method was more efficient than PM in the sunny profile.

Analyzing Fig. 5, it can be seen that the RRR forecasting again had a smaller amplitude than the generation and PM forecasting for the sunny profile. This smaller amplitude is due to the hourly values being close throughout the year for the sunny profile. The RRR method showed a lower annual hourly forecasting variation when compared to PM and a higher correlation with generation at all intervals when compared to PM.

For the intermediate profile (Fig. 5), the RRR method again had a smaller amplitude than the PM generation and forecasting. The RRR method showed less annual hourly variation in forecasting than the PM. Compared to PM, the RRR method had a higher correlation with photovoltaic generation at all intervals.

In the rainy profile (Fig. 5), the RRR method also showed a lower annual hourly forecasting variation than PM. Compared to PM, the RRR method obtained a higher correlation with photovoltaic generation in all hourly intervals.

3.3. Analysis of monthly photovoltaic generation forecasting through average errors using horizon of time resolution of 1–10 h and horizon of time resolution of 24 h

Table 6 shows the monthly average RMSE, MAE, and MAPE values for the RRR method and PM with HRT = 1–10 h and HRT = 24 h. Using a horizon of 1–10 h, the average monthly RMSE for the RRR ranged from 7.3 % to 50.1 %. For the PM method, the average monthly RMSE ranged from 15.1 % to 65.0 %. Therefore, with HRT = 24 h, the average monthly RMSE for the RRR ranged from 4.5 % to 43.2 %. For the PM method, the average monthly RMSE ranged from 11.5 % to 75.0 %.

In the HRT = 1–10 h, the RRR outperformed the PM in almost all the months in the interval analyzed. On the other hand, in April 2022, the RRR showed higher RMSE, MAE, and MAPE than the PM. Using HRT =

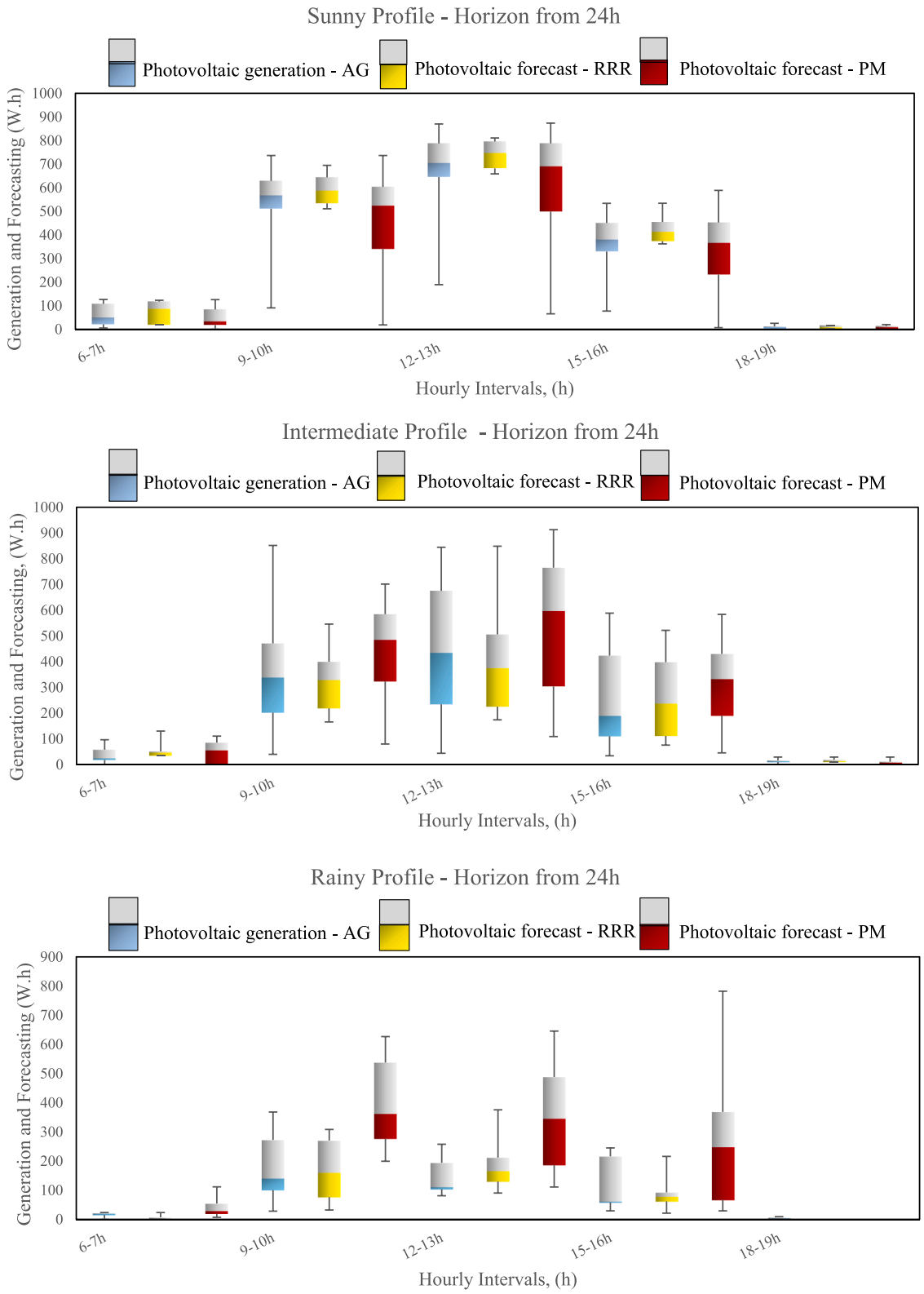

Fig. 5. Box Plot of the generation and forecasting for the profiles in the hourly intervals from March 2022 to February 2023 for an HRT = 24 h.

24 h, the RRR performed better in all months of the interval. April 2022 was a very unstable, cloudy, and rainy month. As a result, the Ventusky platform was less assertive regarding predicting the weather. In April 2022, the Ventusky platform performed better when forecasting the next day's weather than the day's forecast. However, the higher the assertiveness of the weather forecasting platform, the higher the performance of the RRR method.

When we used the RRR for HRT = 24 h, we observed that the RMSE was lower when compared to HRT = 1–10 h in Jan./23, Apr./22, Jun./22, Jul./22, Aug./22 and Nov./22. This result shows that the RRR method is reliable for short and medium-term horizons. The MAE error followed the same pattern as the RMSE in Jan./23, Apr./22, Jun./22,

Table 6
Monthly average daily RMSE, MAE and MAPE for the RRR and PM methods with HRT = 1–10 h and HRT = 24 h.

Horizon from 1 h to 10 h												
	Jan/23	Feb/23	Mar/22	Apr/22	May/22	Jun/22	Jul/22	Aug/22	Sep/22	Oct/22	Nov/22	Dec/22
RMSE [%]												
RRR	7,3	11,8	30,4	50,1	13,7	17,8	20,5	23,3	23,7	24,1	15,7	15,3
PM	15,1	20,4	65,0	42,9	41,2	36,5	39,3	38,9	63,0	36,1	37,3	31,3
MAE [%]												
RRR	5,0	7,9	20,3	38,9	11,5	12,3	14,7	16,3	20,2	17,2	8,9	12,1
PM	12,5	16,7	49,2	37,1	28,1	30,4	31,4	24,8	53,0	27,5	26,2	23,6
MAPE [%]												
RRR	6,1	9,6	27,8	49,4	12,6	13,5	20,3	30,2	28,7	27,6	11,7	16,5
PM	13,1	18,4	64,4	48,8	38,3	39,7	37,4	31,0	86,1	34,8	26,6	26,5
Horizon from 24 h												
	Jan/23	Feb/23	Mar/22	Apr/22	May/22	Jun/22	Jul/22	Aug/22	Sep/22	Oct/22	Nov/22	Dec/22
RMSE [%]												
RRR	4,5	15,0	43,2	34,4	22,4	15,3	19,9	21,8	30,1	32,5	18,2	20,5
PM	11,5	16,0	69,5	44,4	31,8	40,9	29,8	58,7	75,0	56,2	35,9	20,8
MAE [%]												
RRR	4,4	10,6	26,8	25,3	14,0	11,1	15,2	15,4	25,2	24,5	10,7	13,1
PM	8,2	13,8	48,5	38,2	22,9	31,2	21,8	40,8	59,9	46,8	23,8	15,8
MAPE [%]												
RRR	4,4	11,6	45,4	29,1	14,5	14,7	18,0	44,9	43,5	34,0	15,5	19,5
PM	8,2	14,7	89,0	68,4	35,6	32,5	24,2	114,0	97,3	88,3	29,1	21,2

and Aug./22, being smaller when comparing HRT = 24 h with HRT = 1–10 h. The MAPE error followed the same pattern as the RMSE in Jan./23, Apr./22, and Jul./23, being lower when comparing HRT = 24 h with HRT = 1–10 h.

The RRR method's performance is close to that of the methods in literature. Compared to the study by (Dewangan et al., 2020), January 2023 obtained a lower RMSE than all the methods combined for the three plants in the four months presented by (Dewangan et al., 2020).

In February 2023 and June and November 2022, the RRR method performed similarly to the combined methods of (Dewangan et al., 2020) for plant 2.

For January, the RRR method had a lower MAPE than all the methods evaluated by (Dewangan et al., 2020) in months one, three, and four. In the second month, (Dewangan et al., 2020) showed lower MAPE than the RRR method when using the combined SVRGC, SVRLC, MeanC, and MedianC methods. The RRR method showed the highest MAPE of the other 12 methods presented in the paper.

Many of these differences are due to photovoltaic generation forecasting being carried out in different locations, methodologies, premises, assumptions, and climates. The number of data points also favors better performance of the RRR method.

3.4. Analysis of the forecasting of annual photovoltaic generation through average daily errors using horizon of time resolution of 1–10 h and horizon of time resolution of 24 h

Annual average RMSE, MAE and MAPE were obtained for the three profiles. Table 7 shows the values obtained. The best performance of the RRR was for the rainy profile (HRT = 1–10 h) with an annual average RMSE of 9.9 %.

Table 7
Annual mean daily error RSME, MAE and MAPE for the three profiles.

PROFILES	RMSE (%) RRR	MAE (%) RRR	MAPE (%) RRR	RMSE (%) PM	MAE (%) PM	MAPE (%) PM
Sunny*	15,7	9,7	11,9	20,4	15,5	15,7
Intermediate*	25,0	19,4	19,7	47,8	37,2	44,7
Rainy*	9,9	8,3	8,7	85,6	63,5	78,5
Sunny**	18,4	10,9	18,5	31,6	21,5	26,9
Intermediate**	33,7	26,8	32,5	52,7	43,4	65,5
Rainy**	23,3	16,2	24,8	173,2	126,0	140,8

* Horizon from 1 h to 10 h.
** Horizon from 24 h

As expected, the intermediate profile had a higher average RMSE value, 25.0 %. This is the most challenging profile to forecasting. In this profile, solar radiation is intermittent due to a significant accumulation and passage of clouds. For the sunny profile, the average RMSE was 15.7 %. Comparing RRR with PM for all profiles (HRT = 1–10 h and HRT = 24 h), the RRR method outperformed PM.

In the study by (Zhu et al., 2015), the authors found an RMSE of 9.3 % (ANN) and 7.19 % (WD + ANN) for forecasting clear-sky weather. The RRR method obtained a lower performance of 6.4 % compared to ANN and 8.5 % compared to WD + ANN. One of the reasons for the lower performance of the RRR method is that the sunny profile considered forecasts of sun with some clouds and sun with few clouds together with clear skies. These forecasts have clouds passing through during the day, so it is not a totally sunny day.

When forecasting cloudy weather, the authors (Zhu et al., 2015) found an RMSE of 18.5 % (ANN) and 17.6 % (WD + ANN). The RRR method had a lower performance of 6.5 % compared to ANN and 7.4 % compared to WD + ANN. The RRR method considered the forecast of sunshine and rain showers as an intermediate profile, thus increasing the variability of irradiance throughout the day.

For the rainy profile, the RRR method showed an average RMSE of 9.9 %, while (Zhu et al., 2015), for the rainy forecast, showed an RMSE of 22.9 % for the ANN method and 19.6 % for the WD + ANN method. The RRR method performed better for the rainfall profile than the two methods presented by (Zhu et al., 2015).

For HRT = 24 h, the RRR method obtained higher RMSE, MAE, and MAPE than for HRT = 1–10 h. In the sunny profile, the RMSE, MAE, and MAPE were 2.7 %, 1.2 %, and 6.6 %, respectively, higher when compared to HRT = 1–10 h. In the intermediate profile, they were 8.7 %, 7.4 %, and 12.8 % higher when comparing HRT = 24 h with HRT = 1–10 h. For the rainy profile, it was 13.4 %, 7.9 %, and 16.1 % higher when comparing HRT = 24 h with HRT = 1–10 h.

Regarding photovoltaic forecasting for the following day (HRT = 24 h), the RRR method was performed closely with the forecasting for the day (HRT = 1–10 h). The difference in assertiveness when comparing the two forecasting horizons is due to the Ventusky platform's lower performance when forecasting the following day.

In general, the RRR method performed closely to the literature. These differences are due to the different methodologies, forecast sites, and model assumptions. If the number of data points increases and the

forecasting time is longer, the RRR model will perform better. Another way to increase the method's assertiveness is to use more accurate weather forecasting platforms. The advantage of using the RRR model is that there is no need to invest heavily in measuring equipment in conjunction with the operating photovoltaic system.

4. Conclusions

This paper has presented a new method for forecasting photovoltaic generation in the short and medium term. The advantage of the method is that it uses only two data inputs: the first is actual historical electricity generation data from a distributed generation system, and the second is weather forecast data from a weather forecasting platform. In predicting annual photovoltaic generation with HRT = 1–10 h and HRT = 24 h, the RRR method performed better than PM in the sunny, intermediate, and rainy profiles. When forecasting monthly photovoltaic generation with HRT = 1–10 h, the RRR method performed better than PM, except for April 2022. Using HRT = 24 h, the RRR method performed better than PM in all the months of the interval analyzed. The best performance of the RRR was for the rainy profile (HRT = 1–10 h) with an annual average RMSE of 9.9 %. As expected, the intermediate profile had a higher average RMSE value, 25.0 %. In this profile, solar radiation is intermittent due to a significant accumulation and passage of clouds. For the sunny profile, the average RMSE was 15.7 %. Comparing RRR with PM for all profiles (HRT = 1–10 h and HRT = 24 h), the RRR method outperformed PM. Using a horizon of 1–10 h, the average monthly RMSE for the RRR ranged from 7.3 % to 50.1 %. For the PM method, the average monthly RMSE ranged from 15.1 % to 65.0 %. Therefore, with HRT = 24 h, the average monthly RMSE for the RRR ranged from 4.5 % to 43.2 %. For the PM method, the average monthly RMSE ranged from 11.5 % to 75.0 %. In the HRT = 1–10 h, the RRR outperformed the PM in almost all the months in the interval analyzed. On the other hand, in April 2022, the RRR showed higher RMSE, MAE, and MAPE than the PM. Using HRT = 24 h, the RRR performed better in all months of the interval. April 2022 was a very unstable, cloudy, and rainy month.

The RRR method proved to be easy to access and inexpensive to run. Another advantage is that photovoltaic systems can be applied anywhere on the planet for any configuration, inclination, and orientation. The RRR method does not require additional equipment, such as pyranometers, reference cells, high-resolution digital cameras, etc. The RRR is as low-cost as the PM method and performs better. Another advantage is that if there is no generation data available at the forecasting site, we can use generation data from photovoltaic systems located close to the forecasting site. For better performance of the proposed method, more accurate weather forecast meteorological data can be sought. The errors presented by the proposed model are linked to the errors in the meteorological data. Another recommendation to improve the model's performance is to avoid creating the ($\overline{PFP}$) pattern for sunny, intermediate, and rainy profiles. Instead, we can create a PF for each type of weather forecast provided by the weather platform and use it for forecasting. We can also create the ($\overline{PFP}$) for every 15 days instead of for the entire month. In some months, the generation varies at the beginning and end of the month.

Author statement

André Possamai Rosso: Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Drafting the manuscript, Approval of the version of the manuscript to be published. **Giuliano Arns Rampinelli:** Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Revising the manuscript critically for important intellectual contente, Approval of the version of the manuscript to be published. **Lirio Schaeffer:** Drafting the manuscript, evising the manuscript critically for important intellectual content, Approval of the version of the manuscript to be published.

CRediT authorship contribution statement

André Possamai Rosso: Writing – original draft, Validation, Methodology, Investigation, Data curation, Conceptualization. **Giuliano Arns Rampinelli:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Data curation, Conceptualization. **Lirio Schaeffer:** Writing – review & editing, Visualization, Validation, Supervision, Methodology, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

The authors would like to thank the Federal University of Santa Catarina for making photovoltaic generation data available. The Ventusky platform for the availability of weather forecast data.

References

Ahmed, R., Sreeram, V., Mishra, Y., Arif, M.D., 2020. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. Renew. Sustain. Energy Rev. 124, 109792 https://doi.org/10.1016/j.rser.2020.109792.

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-pison, F.J., Antonanzas-Torres, F., 2016. Review of photovoltaic power forecasting. Sol. Energy 136, 78–111. https://doi.org/10.1016/j.solener.2016.06.069.

Babalhavaeji, A., Radmanesh, M., Jalili, M., Gonzalez, S.A., 2023. Photovoltaic generation forecasting using convolutional and recurrent neural networks. Energy Rep. 9, 119–123. https://doi.org/10.1016/j.egyr.2023.09.149.

Bakker, K., Whan, K., Knap, W., Schmetts, M., 2019. Comparison of statistical post-processing methods for probabilistic NWP forecasts of solar radiation. Sol. Energy 191, 138–150. https://doi.org/10.1016/j.solener.2019.08.044.

Barbieri, F., Rajakaruna, S., Ghosh, A., 2017. Very short-term photovoltaic power forecasting with cloud modeling: a review. Renew. Sustain. Energy Rev. 75, 242–263. https://doi.org/10.1016/j.rser.2016.10.068.

Blaga, R., Sabadus, A., Stefu, N., Dughir, C., Paulescu, M., Badescu, V., 2019. A current perspective on the accuracy of incoming solar energy forecasting. Prog. Energy Combust. Sci. 70, 119–144. https://doi.org/10.1016/j.pecs.2018.10.003.

Böök, H., Lindfors, A., 2020. Site-specific adjustment of a NWP-based photovoltaic production forecast. Sol. Energy 211, 779–788. https://doi.org/10.1016/j.solener.2020.10.024.

Das, U.K., Tey, K.S., Seyedmahmoudian, M., Mekhilef, S., Idris, M.Y.I., Deventer, W.V., Horan, B., Stojcevski, A., 2018. Forecasting of photovoltaic power generation and model optimization: a review. Renew. Sustain. Energy Rev. 81, 912–928. https://doi.org/10.1016/j.rser.2017.08.017.

Dewangan, C.L., Singh, S.N., Chakrabarti, S., 2020. Combining forecasts of day-ahead solar power. Energy 202, 117743. https://doi.org/10.1016/j.energy.2020.117743.

Diagne, M., David, M., Lauret, P., Boland, J., Schmutz, N., 2013. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renew. Sustain. Energy Rev. 27, 65–76. https://doi.org/10.1016/j.rser.2013.06.042.

Dimd, B.D., Völler, S., Midtgård, O.-M., Sevault, A., 2023. The effect of mixed orientation on the accuracy of a forecast model for building integrated photovoltaic systems. Energy Rep. 9, 202–207. https://doi.org/10.1016/j.egyr.2023.08.082.

Dolara, A., Leva, S., Manzolini, G., 2015. Comparison of different physical models for PV power output prediction. Sol. Energy 119, 83–99. https://doi.org/10.1016/j.solener.2015.06.017.

Dyreson, A.R., Morgan, E.R., Monger, S.H., Acker, T.L., 2014. Modeling solar irradiance smoothing for large PV power plants using a 45-sensor network and the Wavelet Variability Model. Sol. Energy 110, 482–495. https://doi.org/10.1016/j.solener.2014.09.027.

Gong, D., Chen, N., Ji, Q., Tang, Y., Zhou, Y., 2023. Multi-scale regional photovoltaic power generation forecasting method based on sequence coding reconstruction. Energy Rep. 9, 135–143. https://doi.org/10.1016/j.egyr.2023.05.128.

Gueymard, C.A., 2014. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: towards a better bankability of solar projects. Renew. Sustain. Energy Rev. 39, 1024–1034. https://doi.org/10.1016/j.rser.2014.07.117.

Jang, H.S., Bae, K.Y., Park, H.S., Sung, D.K., 2016. Solar power prediction based on satellite images and support vector machine. IEEE Trans. Sustain. Energy 7, 1255–1263. https://doi.org/10.1109/TSTE.2016.2535466.

Larson, D.P., Nonnenmacher, L., Coimbra, C.F., 2016. Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest. Renew. Energy 91, 11–20. https://doi.org/10.1016/j.renene.2016.01.039.

Luque, A., Hegedus, S., 2011. Handbook of Photovoltaic Science and Engineering. 2ª Ed.

Maciel, J.N. Hybrid Prediction Method with Image Processing and Artificial Intelligence Applicable to Photovoltaic Solar Energy Generation, 2022. Doctoral Thesis, UNILA, Foz do Iguaçu.

Mellit, A., Kalogirou, S.A., 2008. Artificial intelligence techniques for photovoltaic applications: a review. Prog. Energy Combust. Sci. 34, 574–632. https://doi.org/10.1016/j.pecs.2008.01.001.

Mosavi, A., Salimi, M., Ardabili, S.F., Rabczuk, T., Shamshirband, S., Varkonyi-koczy, A., 2019. R. State of the Art of Machine Learning Models in Energy Systems, a Systematic. Review. Energies 12 (7), 1301. https://doi.org/10.3390/en12071301.

Nespoli, A., Ogliari, E., Leva, S., Pavan, M.A., Mellit, A., Lughi, V., Dolara, A., 2019. Day-ahead photovoltaic forecasting: a comparison of the most e effective techniques. Energies 12, 1621. https://doi.org/10.3390/en12091621.

Ogliari, E., Dolara, A., Manzolini, G., Leva, S., 2017. Physical and hybrid methods comparison for theday ahead PV output power forecast. Renew. Energy 113, 11–21. https://doi.org/10.1016/j.renene.2017.05.063.

Pelland, S., Galanis, G., Kallos, G., 2013. Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model. Prog. Photovolt. Res. Appl. 21, 284–296. https://doi.org/10.1002/pip.1180.

Piotrowski, P., Rutyna, I., Baczyński, D., Kopyt, M., 2022. Evaluation metrics for wind power forecasts: a comprehensive review and statistical analysis of errors. Energies 15 (24), 9657. https://doi.org/10.3390/en15249657.

Raza, M.Q., Nadarajah, M., Ekanayake, C., 2016. On recent advances in PV output power forecast. Sol. Energy 136, 125–144. https://doi.org/10.1016/j.solener.2016.06.073.

Reinsel, G.C., Velu, P., 1998. Multivariate Reduced-Rank Regression: Theory and Applications. Springer, New York.

Sangrody, H., Zhou, N., Zhang, Z., 2020. Similarity-based models for day-ahead solar PV generation forecasting. IEEE Access 8, 104469–104478. https://doi.org/10.1109/ACCESS.2020.2999903.

Sobri, S., Koohi-Kamali, S., Rahim, N.A., 2018. Solar photovoltaic generation forecasting methods: a review. Energy Convers. Manag. 156, 459–497. https://doi.org/10.1016/j.enconman.2017.11.019.

Tuohy, A., Zack, J., Haupt, S.E., Sharp, J., Ahlstrom, M., Dise, S., Black, J., 2015. Solar forecasting: methods, challenges, and performance. IEEE Power Energy Mag. 13, 50–59. https://doi.org/10.1109/MPE.2015.2461351.

Wirth, H., 2020. Recent Facts about Photovoltaics in Germany, Fraunhofer ISE. ⟨https://www.ise.fraunhofer.de/en/publications/studies/recent-facts-about-pv-in-germany.html⟩.

Wolff, B., Kühnert, J., Lorenz, E., Kramer, O., Heinemann, D., 2016. Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data. Sol. Energy 135, 197–208. https://doi.org/10.1016/j.solener.2016.05.051.

Yang, D., Dong, Z., 2018. Operational photovoltaics power forecasting using seasonal time series ensemble. Sol. Energy 166, 529–541. https://doi.org/10.1016/j.solener.2018.02.011.

Yildiz, C., Acikgoz, H., Korkmaz, D., Budak, U., 2021. An improved residual-based convolutional neural network for very short-term wind power forecasting. Energy Convers. Manag. 228, 113731 https://doi.org/10.1016/j.enconman.2020.113731.

Zhang, R., Ma, H., Saha, T.K., Zhou, X., 2021. Photovoltaic nowcasting with bi-level spatio-temporal analysis incorporating sky images. IEEE Trans. Sustain. Energy 12 (3), 1766–1776. https://doi.org/10.1109/TSTE.2021.3064326.

Zhang, X., Li, Y., Lu, S., Hamann, H.F., Hodge, B.M., Lehman, B., 2019. A solar time based analog ensemble method for regional solar power forecasting. IEEE Trans. Sustain. Energy 10, 268–279. https://doi.org/10.1109/TSTE.2018.2832634.

Zheng, H., Wu, Y., 2019. A XGBoost model with weather similarity analysis and feature engineering for short-term wind power forecasting. Appl. Sci. 9, 3019 https://doi.org/10.3390/app9153019.

Zhou, H., Zhang, Y., Yang, L., Liu, Q., Yan, K., Du, Y., 2019. Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism. IEEE Access 7, 78063–78074. https://doi.org/10.1109/ACCESS.2019.2923006.

Zhu, H., Li, X., Sun, Q., Nie, L., Yao, J., Zhao, G., 2015. A power prediction method for photovoltaic power plant based on wavelet decomposition and artificial neural networks. Energies 9 (11), 1–15. https://doi.org/10.3390/en9010011.