Article

Multiple Site Intraday Solar Irradiance Forecasting by Machine Learning Algorithms: MGGP and MLP Neural Networks

Gabriel Mendonça de Paiva [1,*] [ID], **Sergio Pires Pimentel** [1] [ID], **Bernardo Pinheiro Alvarenga** [1] [ID], **Enes Gonçalves Marra** [1] [ID], **Marco Mussetta** [2*] [ID] and **Sonia Leva** [2] [ID]

[1]   School of Electrical, Mechanical, and Computer Engineering, Federal University of Goias (UFG), Goiania 74605-010, Brazil; sergio_pimentel@ufg.br (S.P.P.); bernardo_alvarenga@ufg.br (B.P.A.); enes.gm@gmail.com (E.G.M.)
[2]   Department of Energy, Politecnico di Milano, 20156 Milano, Italy; sonia.leva@polimi.it
[*]   Correspondence: gabrielpaiva@discente.ufg.br (G.M.d.P.); marco.mussetta@polimi.it (M.M.)

Abstract: The forecasting of solar irradiance in photovoltaic power generation is an important tool for the integration of intermittent renewable energy sources (RES) in electrical utility grids. This study evaluates two machine learning (ML) algorithms for intraday solar irradiance forecasting: multigene genetic programming (MGGP) and the multilayer perceptron (MLP) artificial neural network (ANN). MGGP is an evolutionary algorithm white-box method and is a novel approach in the field. Persistence, MGGP and MLP were compared to forecast irradiance at six locations, within horizons from 15 to 120 min, in order to compare these methods based on a wide range of reliable results. The assessment of exogenous inputs indicates that the use of additional weather variables improves irradiance forecastability, resulting in improvements of 5.68% for mean absolute error (MAE) and 3.41% for root mean square error (RMSE). It was also verified that iterative predictions improve MGGP accuracy. The obtained results show that location, forecast horizon and error metric definition affect model accuracy dominance. Both Haurwitz and Ineichen clear sky models have been implemented, and the results denoted a low influence of these models in the prediction accuracy of multivariate ML forecasting. In a broad perspective, MGGP presented more accurate and robust results in single prediction cases, providing faster solutions, while ANN presented more accurate results for ensemble forecasting, although it presented higher complexity and requires additional computational effort.

Keywords: solar irradiance forecasting; multigene genetic programming; multilayer perceptron; artificial neural networks; short-term forecasting; intraday forecasting

1. Introduction

The increased penetration of renewable energy sources (RES) in power systems has created a complex challenge from the point of view of electric grid management [1–3], mainly due to high intermittence energy sources such as sun irradiation and wind [4,5]. Climatic variations instantly influence the electric power generation of wind farms and photovoltaic (PV) systems and may put the balance between load demand and power supply in the electrical power grid at risk. In this context, weather forecasting stands out as an important tool to guide the operation of electric power utility grids in the presence of intermittent RES [6,7].

Solar energy forecasting is normally classified in terms of two different forecasting horizons, namely intraday forecasting, from a few minutes to 6 h ahead, and day-ahead forecasting, where predictions are

performed for the next day, as defined by the International Energy Agency (IEA) in their Solar Forecasting State of the Art Report [8]. The IEA report also states that statistical techniques such as time-series machine learning provide good results in the intraday context, while physical models based on numerical weather prediction (NWP) provide good results in the day-ahead context.

Numerous contributions by different authors have reviewed the field of solar forecasting research [9,10]. Recurrently, the performance of a given method varies depending on different circumstances; a variety of machine learning (ML) and statistical methods have been studied for intraday solar predictions. Currently, ML is probably the most employed approach [9]. Research specifically done with ML techniques was reviewed by Voyant et al. [11] and has pointed out that the accuracy and robustness of ML forecasts depend on the training method and the metric used to evaluate predictions [9]. Ranking these methods in the literature is a complex mission due to the influence of the distinct data sets studied, time steps, forecasting horizons and performance indicators [11]. As a result, also considering significant errors noticed in solar forecast results, this work addresses the investigation of additional ML forecast methods at multiple locations in order to achieve further results to support the solar forecasting research field [12].

Concerning ML methods applied to solar forecasting, artificial neural networks (ANN), support vector machines (SVM), k-nearest neighbors (kNN), random forest (RF) and gradient boosted regression (GBR) are the most employed techniques [13–19]. In particular, in short-term intraday solar forecasting applications, autoregressive methods of the autoregressive integrated moving average (ARIMA) class are also frequently used as statistical method to forecast normalized indexes of irradiance [20]; moreover, frequency domain models are commonly used in solar irradiance forecasting [21].

Studies on intraday solar forecasting also differ from each other in terms of the category of data input. The most frequent approaches are time-series point forecasts, where meteorological measurements are used as input [13–15,17,21], sky imagers from satellite or from ground [22–24], experiments combining multiple time-series based on spatio-temporal forecasts [25,26] and hybrid approaches combining different statistical methods or multiple data acquisition systems [27].

This study proposes multigene genetic programming (MGGP) as a novel state-of-the-art ML method applied to solar irradiance forecasting. A comparison between MGGP and the multilayer perceptron (MLP) artificial neural network (ANN) is carried out on the same basis for intraday irradiance forecasting at multiple locations. Forecasts are based on meteorological historical data to execute predictions in horizons from 15 to 120 min ahead. The accuracy, robustness, advantages and disadvantages of each ML method are highlighted in order to support future research in the field. The results used in the comparative study were achieved by the implementation of both ML methods in the Matlab® programming platform.

2. Databases

Six locations were evaluated in this study, as shown in Figure 1 and reported in the following paragraphs.

2.1. Goiania, Brazil

The database from Goiania was acquired using a meteorological station setup at the Federal University of Goias (UFG) in Goiania—the capital city of the State of Goias, Midwest Brazil—whose coordinates are latitude $-16.67°$ (Southern Hemisphere), longitude $-9.24°$ (west); the station is located 749 m above sea level [28]. The three-year-long database sampled each 60 s from August 2015 to July 2018 is available at the webpage https://sites.google.com/site/sfvemcufg/weather-station [28]. The database has a rigorous data quality control with monthly inspection of equipment, and data are acquired and stored in a datalogger. Table 1 presents the configuration of the weather station setup for solar applications and research.

Measured global horizontal irradiation values were, respectively, 1929 kWh/m^2, 1913 kWh/m^2 and 1924 kWh/m^2 for each sequence of 12 months from August 2015 to July 2018.
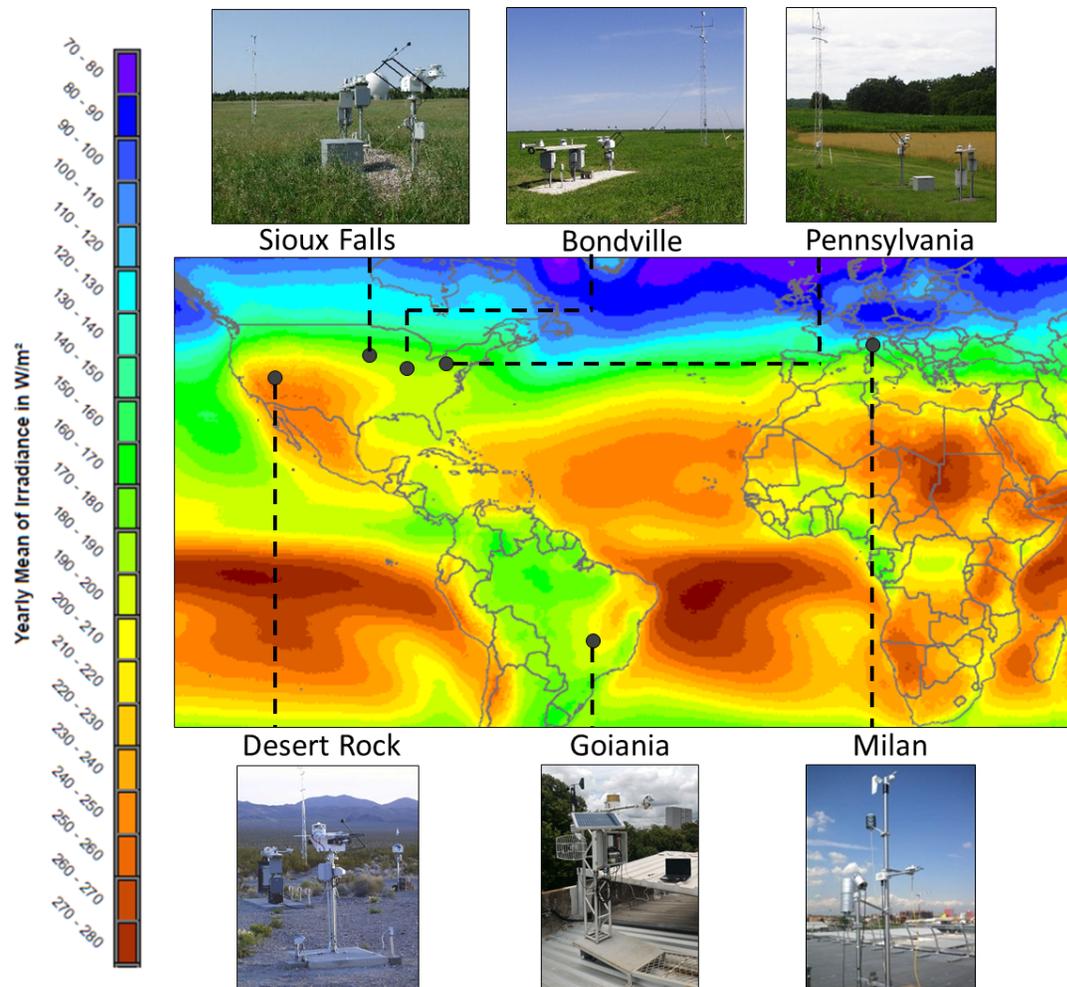
Figure 1. Locations of the weather stations under analysis presented on a world map, adapted from [29].

Table 1. Description of the equipment used at the Federal University of Goias (UFG) weather station, the parameters they measure and their accuracy and range of operation.

Equipment	Parameter Measured	Information
Pyranometer Hukseflux LP02 calibrated	Global horizontal irradiance	Second class ISO 9060: in-field uncertainty of $\pm 5\%$, calibration uncertainty $< 1.8\%$
R. M. Young Wind 03002	Wind speed	Range 0 to 50 m/s and accuracy of ± 0.5 m/s
	Wind direction	Accuracy of $\pm 5\%$
Texas Electronics TB-2012M	Atmospheric pressure	Calibration range 878 to 1080 mBars Uncertainty of ± 1.3 mBar
Texas Electronics TTH-1315	Ambient temperature	Operating ranges -40 °C–+60 °C and 0–100%
	Relative humidity	Accuracy of ± 0.3 °C and $\pm 1.5\%$ RH
Texas Electronics TR-525I	Rainfall	Accuracy of $\pm 1\%$
Datalogger Campbell Scientific CR800X	Automatic data acquisition	

2.2. Milan, Italy

The database from Milan was acquired using a meteorological station setup installed at the SolarTech Lab at the Politecnico di Milano in Milan, Italy, whose coordinates are latitude +45.50° (Northern Hemisphere), longitude +49.24° (east); the station is located 120 m above sea level [30]. A database was sampled each 60 s for 26 months from September 2016 to October 2018 with the station at the SolarTech Lab [30].

2.3. SURFRAD-US

The other four databases were obtained from sites based in the United States of America and collected from the USA National Oceanic and Atmospheric Administration (NOAA) Surface Radiation Network (SURFRAD). The coordinates of these sites are as follows: Desert Rock, latitude +36.62° (north), longitude −116.02° (west), altitude of 1007 m; Pennsylvania State University (PSU), latitude +40.72° (north), longitude −77.93° (west), altitude of 376 m; Bondville, latitude +40.05° (north), longitude −88.37° (west), altitude of 213 m; and Sioux Falls, latitude +43.73° (north), longitude −96.62° (west), altitude of 473 m. These databases are public domain and have also been analyzed in a previous study of irradiance forecasting [21]. This study used weather data sets from NOAA's SURFRAD averaged per minute from January 2013 to December 2015. The quality control of measurements from such databases was performed by the identification and removal of inconsistent values. The yearly global horizontal irradiations from the weather stations from January 2013 to December 2015 were, respectively, 2025, 2055 and 1916 kWh/m^2 for Desert Rock; 1317, 1350 and 1366 kWh/m^2 for PSU; 1438, 1458 and 1438 kWh/m^2 for Bondville; and 1410, 1384 and 1437 kWh/m^2 for Sioux Falls.

3. Data Processing

3.1. Normalization

Independent of whether forecasts are performed with the use of artificial intelligence methods or classical regressions, the data processing strategy and input–output scheme play a key role in developing improved forecasts. The first data processing strategy considered global horizontal irradiance (*G*) as a target value, combining past values of irradiance and weather variables in addition to deterministic variables (in order to capture temporal trends in datasets) [31].

The proposed approach was refined by adopting a data processing strategy that forecasts normalized indexes in order to remove seasonality in solar data, yielding prompter ML algorithm convergence for irradiance forecasting. Values measured at night and during solar elevations (*h*) less than 5° were neglected. Normalization of solar data can be performed by the application of Equation (1), where k_t^* is the so-called clear sky index, *G* is the observed global horizontal plane irradiance (GHI) and G_{clr} is the theoretical clear sky irradiance.

$$k_t^* = \frac{G}{G_{clr}} \tag{1}$$

Clear sky irradiance models used in the literature range from simple functions of extraterrestrial irradiance models to complex approaches that take numerous measured atmospheric parameters into account. It was found that Haurwitz clear sky irradiance and Ineichen–Perez models are simple and sufficiently accurate models that were systematically employed to evaluate meteorological data from a wide number of sites in the USA [32].

The Haurwitz clear sky irradiance model was developed in 1945 and is given by Equation (2), where θ_z is the solar zenith angle (complementary to the solar elevation angle *h*). The constants 1098 and −0.057 were obtained by the least-squares method in order to fit measured cloudless sky irradiance data from

a site in the USA to a theoretical curve based on a zenith angle exponential function. The exponential function is decreased by a factor proportional to $\cos\theta_z$ from sunrise to sunset.

$$G_{clr} = 1098\cos\theta_z \exp\frac{-0.057}{\cos\theta_z} \tag{2}$$

The solar zenith angle is defined as the angle between the zenithal axis and the line to the sun. Thus, this angle varies instantly, according to the rotation movement of the Earth. The cosine of the solar zenith angle is obtained from Equation (3), where δ is the declination angle, ϕ is the latitude of the weather station location, and ω is the sun hour angle. A detailed definition and calculation of solar geometry variables is provided in [33].

$$\cos\theta_z = \cos\phi\cos\delta\cos\omega + \sin\phi\sin\delta \tag{3}$$

Ineichen–Perez clear sky irradiance uses optical air mass ratio (AM), atmospheric turbidity and altitude of location in clear sky irradiance modeling [34]. Ineichen–Perez G_{clr} is calculated by Equation (4), where G_o is the extraterrestrial irradiance, h is the solar elevation angle, a_1, a_2, f_{h1} and f_{h2} are constant functions of local altitude, T_L is the Linke turbidity factor and AM is the optical air mass ratio. The constants in Equations (5) and 6 were added empirically by Ineichen and Perez to improve previous clear sky models which were logarithmically dependent on the Linke turbidity factor and limited to specific location and zenith angles. T_L was obtained in this study from a map of monthly averaged values for each site [29]. In order to avoid discontinuities in T_L and G_{clr} calculations, a daily fitness procedure was used as presented in [35,36]. Figure 2 presents an example of daily T_L fitness for Desert Rock.

$$G_{clr} = a_1 \cdot G_o \cdot \sin h \cdot \exp[-a_2 \cdot AM \cdot (f_{h1} + f_{h2} - (T_L - 1))] \tag{4}$$

$$a_1 = 5.09 \cdot 10^{-5} \cdot altitude + 0.868 \tag{5}$$

$$a_2 = 3.92 \cdot 10^{-5} \cdot altitude + 0.0387 \tag{6}$$
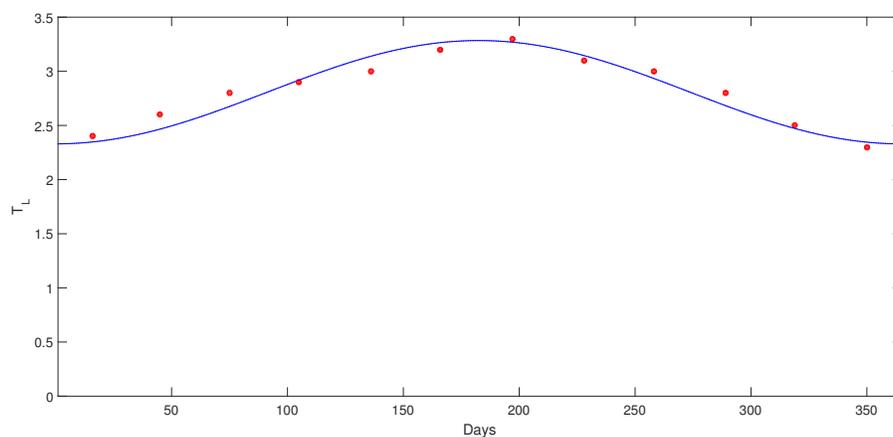
Figure 2. Linke turbidity daily fitness (blue line) of monthly averaged values (red dots) for Desert Rock.

Figure 3 presents an example of the data normalization. It is possible to observe how the normalized index removes daily and yearly seasonality and emphasizes the influence of both clouds and solar potential instantaneous variabilities.
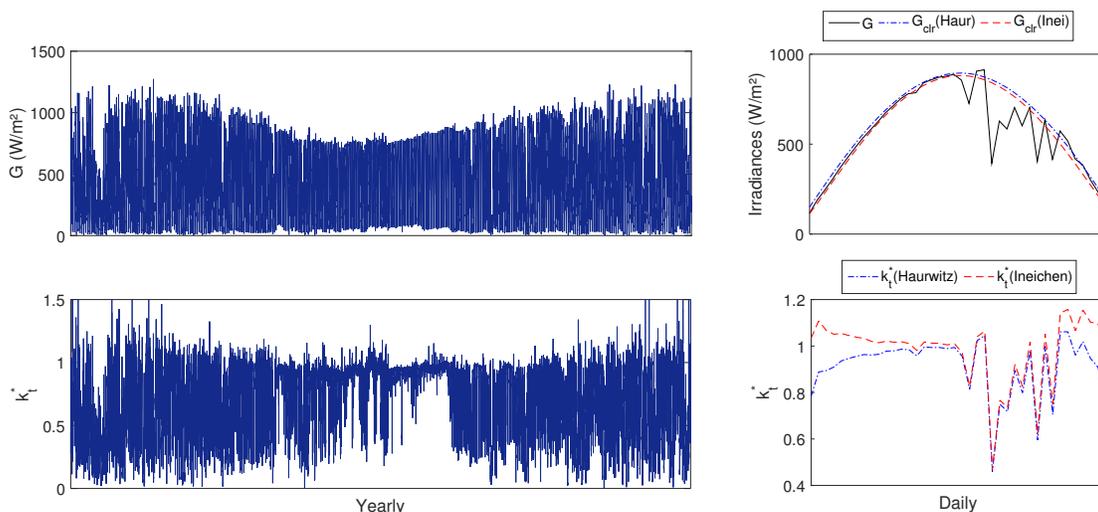
Figure 3. Data normalization from yearly (**left**) and daily (**right**) perspectives for Goiania.

3.2. Data Statistics

A general overview of the solar variability and statistics of each site is presented in Table 2, achieved by applying the Ineichen clear sky model for 15 min averaged point databases. Results in Table 2 show that training, validation and testing datasets present similar mean and standard deviations for k_t^*—an important requirement to implement ML forecasting models. Results from Desert Rock present a behavior with more clear sky conditions as opposed to other locations, thus presenting the highest mean k_t^* with lowest standard deviations, while results from Milan present the highest variabilities (σ).

Table 2. Data statistics of training, validation and test datasets for each location: N_{samp} is the number of samples of each dataset, μ is the average Ineichen k_t^* and σ is the standard deviation of k_t^*.

	Training			**Validation**			**Testing**		
	N_{samp}	μ	σ	N_{samp}	μ	σ	N_{samp}	μ	σ
Goiania	25,813	0.7379	0.3042	11367	0.7458	0.2983	11,163	0.7423	0.3022
Milan	17,828	0.8544	0.3843	7969	0.8069	0.3897	7944	0.7999	0.4194
Desert Rock	25,959	0.9139	0.2380	10,929	0.9133	0.2451	10,865	0.9025	0.2458
Pennsylvania	25,706	0.6741	0.3534	10,998	0.6260	0.3492	11,177	0.6604	0.3572
Bondville	25,935	0.7246	0.3593	10,818	0.6974	0.3660	11,005	0.7197	0.3478
Sioux Falls	25,839	0.7579	0.3455	10,898	0.7476	0.3594	10,708	0.7638	0.3353

3.3. Data Relations

The ML methods consists of a "multivariate" data structure of inputs, as defined in [17], to forecast k_t^* (single output), using relations among data based on output past values, past values of weather variables and deterministic solar variables. Irradiances are then obtained by multiplying back the normalized index outputs by respective clear sky irradiances.

The inputs are as follows:

- $k_t^*(-5)\ldots(-60)$: the 12 past values of k_t^* in time windows of 5 min averages.
- $T_a(-5)\ldots(-60)$: the 12 past values of ambient temperature in $^\circ$C.
- $W_s(-5)\ldots(-60)$: the 12 past values of wind speed in m/s.
- $H_r(-5)\ldots(-60)$: the 12 past values of relative humidity in %.

- p_a(-5)...(-60): the 12 past values of atmospheric pressure in mBar.
- h is the elevation angle of the forecast time window in radians, varying from around 0.0873 (5°) to 1.5708 (90°).
- t_s is the time difference in respect to sunrise in minutes.
- ω_s is the solar time angle in radians.
- "Day" is the day of forecast interval. The days of the year are counted starting one day after the winter solstice and ending on the winter solstice of the next year. We decided to adopt this definition to follow the solar cycle starting from the day of lowest irradiance levels, since the traditional day counting does not have a direct mathematical relation to the evolution of solar variables throughout the year.
- "Month" is the month of the forecast interval, varying from 1 to 12.

Solar deterministic variables are calculated by deterministic mathematical models that represent the solar time-based behavior of solar quantities. These variables are computed in solar time instead of legal time and are directly proportional to the irradiance and its indexes. The data processing methods presented in this section yielded the most accurate results when used by the authors for the analyzed databases.

4. Forecast Methods

4.1. Genetic Programming

Genetic programming (GP) is an artificial intelligence technique which was originally proposed by Koza [37] in the evolutionary computation field; it is considered as an extension of genetic algorithms. GP is inspired by population genetics and biological evolution at the population level [38] (Algorithm 1). GP has proved to be competitive in time series forecasting in relation to other statistical techniques based on artificial intelligence, such as ANN and the support vector machine (SVM) [39–41]. It has been applied in numerous studies of predictions of natural resources—e.g., hydrology [42,43]—and has also been applied to daily or monthly solar irradiance forecasting in PV power systems [44–46].

MGGP and MLP neural networks were analyzed in comparison to other different methods of forecasting, considering that ANN comprises adjustable parameters that can be trained using optimization techniques to solve classification and regression problems and GP is a stochastic optimization method. When GP is used to build a mathematical model based on sampled data with the aim of predicting future values, it is named symbolic regression (SR). GP models are typically described as in Equation (7), where y is the observed output variable, $\hat{y}$ is the predicted output, and $x_1 \dots x_n$ are the observed input variables. In contrast to other soft computing methodologies, such as feed-forward ANNs and SVMs, trained GP models are basic constitutive equations that can be implemented without a specific software environment in any modern programming language.

$$\hat{y} = f(x_1, ..., x_n) \tag{7}$$

GP models can be classified into three different categories according to their mathematical model complexity: naive SR, when the model requires only one gene to relate input data with output data; scaled SR, when the model employs one gene associated to a bias term to relate input and output data; and multigene SR, when the GP uses multiple genes and a bias term to relate input and output data (Figure 4).

$$\hat{y} \;=\; d_0 \,+\, d_1 \,\cdot\, \left(\text{tree} \right) \;+\; d_2 \,\cdot\, \left(\text{tree} \right)$$
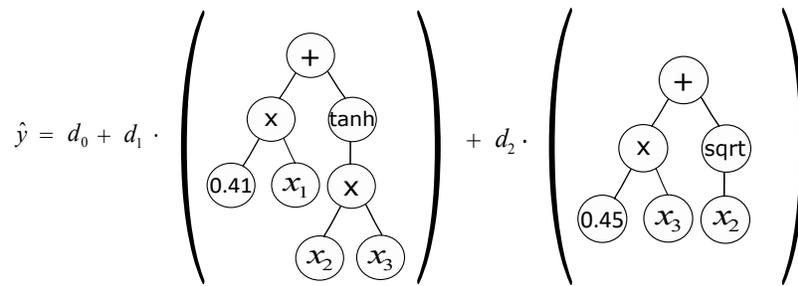
Figure 4. Example of a multigene symbolic regression (SR) model presented in a tree structure.

Algorithm 1: Genetic programming pseudocode

---

1 Input: $Population_{size}$, $Node_{functions}$, $Max_{generations}$, Max_{genes}, Max_{depth}, κ, p_c, p_m, p_r, $elit_{rate}$;
2 Output: S_{best}
3 Population $\leftarrow$ *InitializePopulation*($Population_{size}$, $Node_{functions}$, $Max_{generations}$, Max_{genes},
4 Max_{depth});
5 EvaluatePopulation(Population);
6 $S_{best} \leftarrow$ GetBestSolution(Population);
7 **while** *StopCondition()* **do**
8   Children $\leftarrow \varnothing$;
9   **while** *Size(Children) < $Population_{size}$* **do**
10   Operator $\leftarrow$ SelectGeneticOperator(p_c, p_m, p_r)
11   **if** *Operator = CrossoverOperator* **then**
12             $Parent_1$, $Parent_2 \leftarrow$ SelectParents(Population,$Population_{size}$)
13             $Child_1$, $Child_2 \leftarrow$ Crossover($Parent_1$, $Parent_2$)
14   Children $\leftarrow Child_1$
15   Children $\leftarrow Child_2$
16   **end**
17   **if** *Operator = MutationOperator* **then**
18             $Parent_1 \leftarrow$ SelectParents(Population,$Population_{size}$)
19             $Child_1 \leftarrow$ Mutate($Parent_1$)
20   Children $\leftarrow Child_1$
21   **end**
22   **if** *Operator = ReproductionOperator* **then**
23             $Parent_1 \leftarrow$ SelectParents(Population,$Population_{size}$)
24             $Child_1 \leftarrow$ Reproduce($Parent_1$)
25   Children $\leftarrow Child_1$
26   **end**
27   **end**
28   EvaluatePopulation(Children);
29     $S_{best} \leftarrow$ GetBestSolution(Children,S_{best});
30   Population $\leftarrow$ Children;
31 **end**
32 return S_{best}

Figure 4 illustrates a population individual and a multigene GP model, usning Equation (8), where a bias term d_0 is added to two genes with weights d_1 and d_2 in a tree structure. The terms "plus", "times", "square

"root" and "tanh" are known as node functions. Both weights and nodes are obtained in a GP training procedure.

$$\hat{y} = d_0 + d_1(0.41x_1 + \tanh(x_2 x_3)) + d_2(0.45x_3 + \sqrt{x_2}) \tag{8}$$

GP evolves a population of candidate solutions (population size) in multiple generations by the application of genetic operators with a tournament selection of best individuals. A crossover operation exchanges genes between individuals to assess possible structural improvements of individuals. Mutation is a fine adjustment operation that changes pieces or entire genes into new, random ones to evaluate a possible structural improvement in terms of fitness. Bias and gene weights of individuals are then optimized in terms of least root mean square errors applied to training data according to Equation (9). Applying an elitism strategy with a given elitism rate, a percentage of best fitness solutions is stored over generations. Based on these procedures, GP evaluates thousands of possible regression structures with optimized weights to relate inputs and outputs. Table 3 summarizes the parameters adopted in GP, which are considered again in Section 6.

$$s^* = min\sqrt{\frac{\sum_{i=1}^{N_{samp}}(y_i - \hat{y}_i)^2}{N_{samp}}} \tag{9}$$

Table 3. Summary of genetic programming (GP) simulation parameters.

Parameter	Adopted Setting
Node functions	$+, -, \cdot, /, x^2$, tanh, exp $\sqrt{x}$, exp $-x$, sin, cos
Population size	300
Maximum generations	150
Maximum number of genes	5
Maximum tree depth	4
Tournament size (κ)	6
Lexicographic selection	True
Elitism fraction	0.3
Fitness function	Root mean squared error (RMSE)
Crossover probability (p_c)	0.88
Mutation probability (p_m)	0.12
High-level crossover probability	0.5
Ephemeral random constants range	from -10 to $+10$
ERC probability at creating nodes	0.2

The dynamics of GP solutions are characterized by generalization ability, providing both accurate and robust solutions in training and for other datasets. On the other hand, ANN is highly influenced by overfitting, which is usually controlled by a validation step named early stopping, while GP does not require a validation step during the SR model training stage. Figure 5 presents the performance of the best individuals which evolved over generations for GP forecasts. It is possible to observe the robustness of solutions repeating from training to validation datasets. MGGP models were implemented on GPtips 2—an open-source GP platform for Matlab® [47].
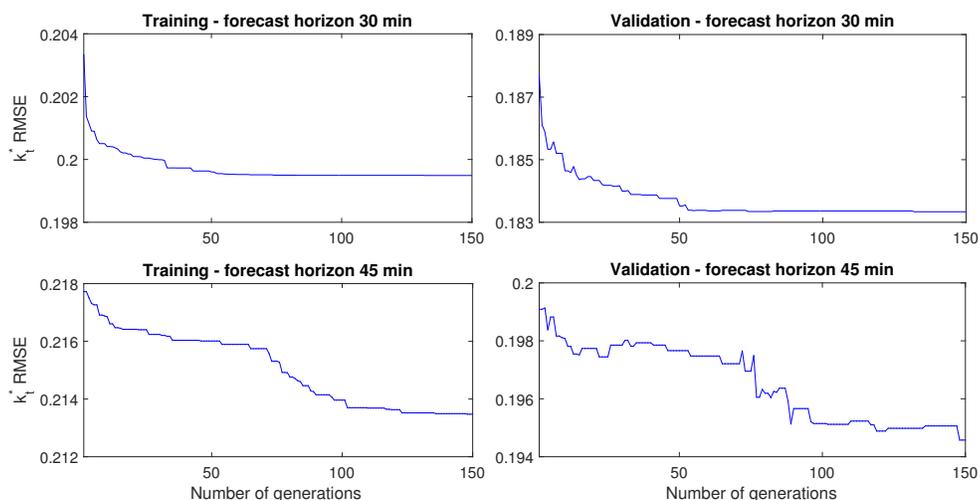
Figure 5. Fitness of best GP solution s^* measured by k_t^* RMSE for training and validation datasets.

4.2. Artificial Neural Networks

A feed-forward multilayer perceptron neural network (MLP) architecture was applied to this analysis, containing one hidden layer of 10 neurons using the hyperbolic tangent sigmoid transfer function. The neural networks were trained with the Levenberg–Marquadt algorithm including early stopping implemented in Matlab$^{\circledR}$ using the neural networks toolbox. The employed ANN set of attributes was previously validated for intraday solar forecasting [14].

4.3. Ensemble Forecasts

Ensemble forecast models are convenient to build with multiple ML simulations and tend to improve forecast accuracy [48]. The ensemble forecast is given by Equation (10), where N_{trial} is the number of trials by the given ML method. In this analysis, the internal parameters of GP and ANN do not vary in each trial, and 10 trials were performed to produce each ensemble according to the methodology described in [48].

$$\hat{G}_{ens} = \frac{\sum_{i=1}^{N_{trial}} \hat{G}_i}{N_{trial}} \tag{10}$$

4.4. Iterative Forecasts

Rana et al. [17] evaluated a forecast method where predictions of instant t+1 are iteratively added as inputs to predictions of instant t+2. As a conclusion, the iterative method did not improve forecasts in their study on PV power forecasting using ANN ensemble and SVM; however, the iterative GP method was tested in this work and yielded improvements on forecasting results. Results comparable to [17] were obtained, and no significant improvement was achieved by using iteration for ANNs. Therefore, the results reported here were obtained using iterative predictions for MGGP.

4.5. Persistence

Persistence forecasting is a common benchmark technique used to evaluate intraday solar forecasting. Persistence forecasting can be computed by Equation (11), where $\hat{G}(t + \Delta T)$ is the persistence forecast

and ΔT is the forecast horizon, which varies from 15 to 120 min; $k_t^*(t)$ is the present clear sky index; and $G_{clr}(t + \Delta T)$ is the clear sky irradiance at the horizon of the forecast.

$$\hat{G}(t + \Delta T) = k_t^*(t) \cdot G_{clr}(t + \Delta T) \tag{11}$$

5. Error Metrics

Although there are many error metrics used in the field of solar forecasting, this study assumed the most traditional metrics in the literature: the root mean squared error (RMSE) given by Equation (12), the mean absolute error (MAE) computed by Equation (13) and the forecast skill given by Equation (14).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_{samp}}(y_i - \hat{y}_i)^2}{N_{samp}}} \tag{12}$$

$$MAE = \frac{\sum_{i=1}^{N_{samp}}|(y_i - \hat{y}_i)|}{N_{samp}} \tag{13}$$

$$s = 1 - \frac{RMSE_{forecast}}{RMSE_{persistence}} \tag{14}$$

While MAE is a linear error metric, RMSE is a quadratic error metric that penalizes inaccurate forecast values due to quadratic factors. RMSE is of particular interest when evaluating RES forecasting since ramp behavior is a relevant issue in PV power system operation.

6. Results and Discussion

6.1. GP Tuning

Initial simulations were intended to analyze the influence of GP parameters in forecast accuracy and robustness. The analysis of parametric influence is known as the parameter tuning of evolutionary algorithms (EAs), as described in [49]. Parameter tuning is by nature an optimization task comprising multiple variables (parameters). In current analyses of multiple horizon forecasts, each forecast horizon at each location consists of a different problem to be tuned. In order to reduce the number of simulations to assess GP parameters, this study considered prior knowledge from other studies to seek good parameter choices to perform a lower number of simulations. Therefore, parameter assessment was carried out only for one forecast horizon using the dataset from Goiania station. Thus, parameter settings from Goiania were used in forecasting models for other sites.

Lima et al. [50] performed a systematic analysis of GP that indicated the population size, number of generations and tree size as the main parameters which influence fitness, while genetic operators have a lower influence. Increases in the size limit of regression functions tend to improve fitness; however, when the size limit is excessively large, this leads to a bloat (function size growth without fitness improvement) [51]. Bloat can be relieved by using realistic elitism rates [52]. In summary, lower tournament sizes and lower elitism rates lead to a higher diversity of solutions.

According to the literature review and some former analyses of irradiance forecasting, the maximum number of genes was set at 5, the tree depth at 4, the number of generations at 150 and population size at 300. These parameters presented a good trade-off between complexity and fitness improvement. Figure 6 presents the improvement of solution fitness in the validation dataset from Goiania station versus the increase in complexity (increasing the maximum number of genes).
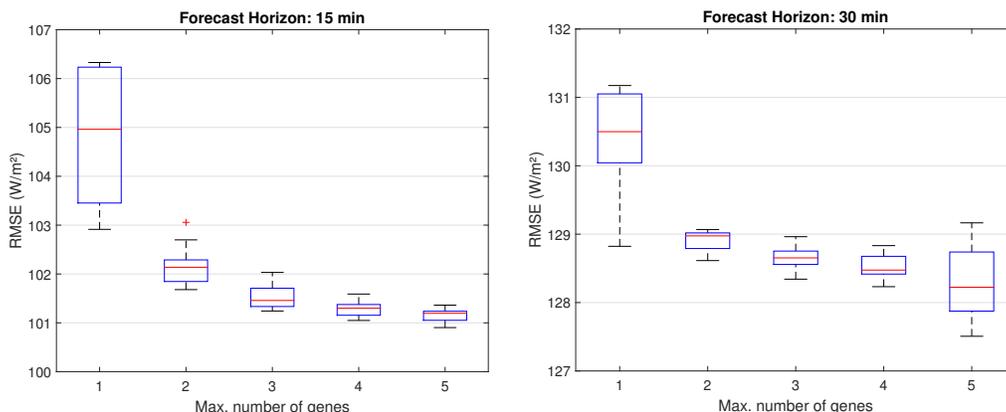
Figure 6. Influence of maximum number of genes on the fitness of best solutions, evaluated for the Goiania validation dataset.

Genetic operators were analyzed by multiple simulations for a forecast horizon of 90 min ahead, as this is a demanding time window for prediction and consequently presents high variability in the different algorithm simulations. The results for accuracy and robustness are given in Figure 7. The number of generations was lowered to 50 during tests in order to obtain a higher variability of results. It is possible to conclude that the best accuracy and robustness (standard deviation for multiple simulations) were those accomplished using higher mutation rates, lower tournament sizes and lower elitism rates. Therefore, we selected the setting with lowest RMSE: $\kappa = 6$, $p_m = 0.12$ and elit $= 0.30$.

Figure 7. Influence of tournament size (κ), mutation rate (p_m) and elitism rate (elit) on the accuracy and robustness (RMSE standard deviation) of the validation dataset from the Goiania site.

6.2. Assessment of Exogenous Input

ANN and GP were executed for all formerly defined locations and forecast horizons both considering and neglecting weather variables Hr, Ta, Ws and pa. The error improvement index, $Improv_{error}$, was defined in Equation (15) in order to assess the improvement yielded by the addition of weather variables at a given error metric, where $error_{univ}$ is the forecast error obtained based on past values of k_t^* with the sole addition of deterministic variables, and $error_{multiv}$ is the forecast error obtained by including weather variables. It is worth highlighting that deterministic variables are able to improve forecasts based merely on past values of k_t^*.

$$Improv_{error} = \frac{(error_{univ} - error_{multiv})}{error_{univ}} \cdot 100\% \qquad (15)$$

Improvements were calculated both in terms of MAE and RMSE, as described in Figure 8. The graphs represent typical behaviors, where weather variables generally improve forecastability for all locations by up to 5.68% in terms of MAE and 3.41% in terms of RMSE; in some locations, negative improvements were obtained for shorter forecast horizons from 15 to 60 min. Mostly, the addition of weather variables tends to improve forecastability for all locations; thus, the results obtained by the multivariate forecasts are reported.
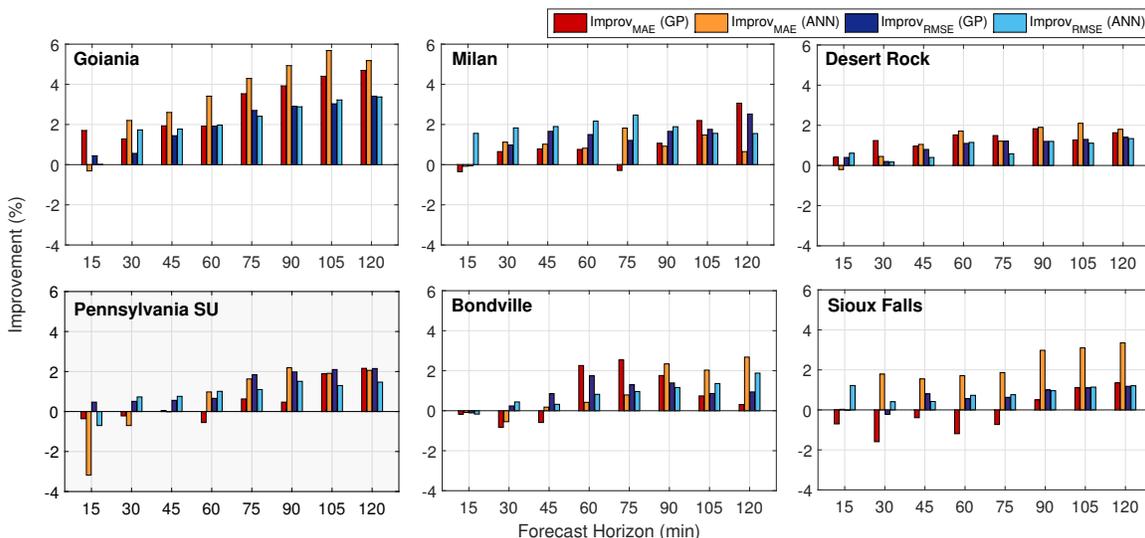
Figure 8. Improvements (%) of multivariate forecasting using GP according to mean absolute error (MAE) (dark red bars) and according to RMSE (dark blue bars), and improvements (%) of multivariate forecasting using an artificial neural network (ANN) according to MAE (orange bars) and according to RMSE (light blue bars).

6.3. Specific Results

Complete results for each forecast horizon and location are presented in Appendix A. The most accurate results are in bold characters for both single and ensemble forecast comparisons. Model accuracy dominance depends on the location, forecast horizon and error metric, as summarized in Figure 9. The accomplished results point toward ANN as the most accurate for short horizons and GP as the most accurate for longer horizons, which also predominantly improves robustness. Furthermore, location attributes have been proven to affect model dominance. Figure 10 presents forecast accuracies for both methods applied to the Goiania station, where the most accurate results were obtained by ANN, and Figure 11 displays the results for the Desert Rock station, where the most accurate results were obtained by GP.

Both GP and ANN methods were consistently improved considering both error metrics by employing an ensemble strategy for each forecast horizon and location. ANN presented more significant improvement and superior accuracy using the ensemble strategy in most cases, as summarized for model accuracy dominance in Figure 12 using ensemble forecasting. GP_{ens} led to the most accurate results in eight cases out of 48, while ANN_{ens} yielded the most accurate results in 23 cases out of 48. GP_{ens} achieved the most accurate results for the Milan station for horizons from 15 to 45 min and from 105 to 120 min using MAE as a reference metric. At Desert Rock station, GP_{ens} attained the lowest RMSE for horizons from 30 to 120 min. At Bondville station, GP_{ens} accomplished the lowest RMSE for horizons from 90 to 120 min and the lowest MAE for horizons from 75 to 120 min. At PSU station, GP_{ens} led to the lowest MAE and RMSE for horizons from 105 to 120 min. At Sioux Falls station, GP_{ens} yielded the lowest RMSE for horizons from 105 to 120 min and the lowest MAE for horizons of 45 min and from 90 to 120 min.

	Forecast Horizon							
	15	30	45	60	75	90	105	120
Goiania	ANN	ANN	ANN	GP/ANN	GP/ANN	ANN	GP/ANN	GP/ANN
Milan	GP/ANN	GP/ANN	GP	GP	GP	GP	GP	GP
Desert Rock	GP/ANN	GP	GP	GP	GP	GP	GP	GP
Pennsylvania	GP/ANN	ANN	GP/ANN	ANN	ANN	GP	GP	GP
Bondville	ANN	ANN	ANN	GP	GP	GP	GP	GP
Sioux Falls	GP/ANN	ANN	GP	GP	GP/ANN	GP	GP	GP
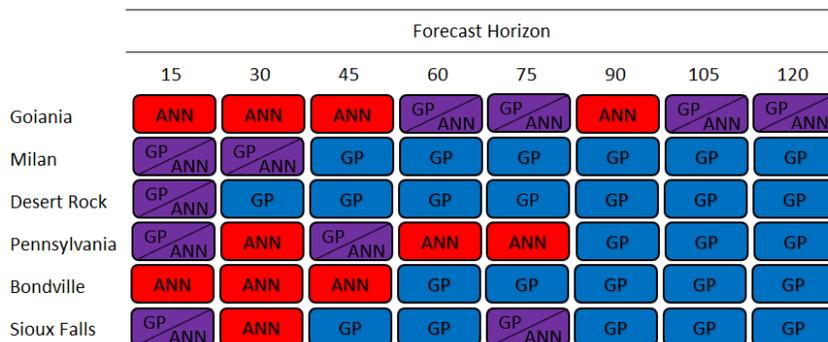
Figure 9. Model accuracy dominance by location and forecast horizon in single forecasts. GP/ANN indicates cases in which accuracy dominance differs from the error metric evaluated.
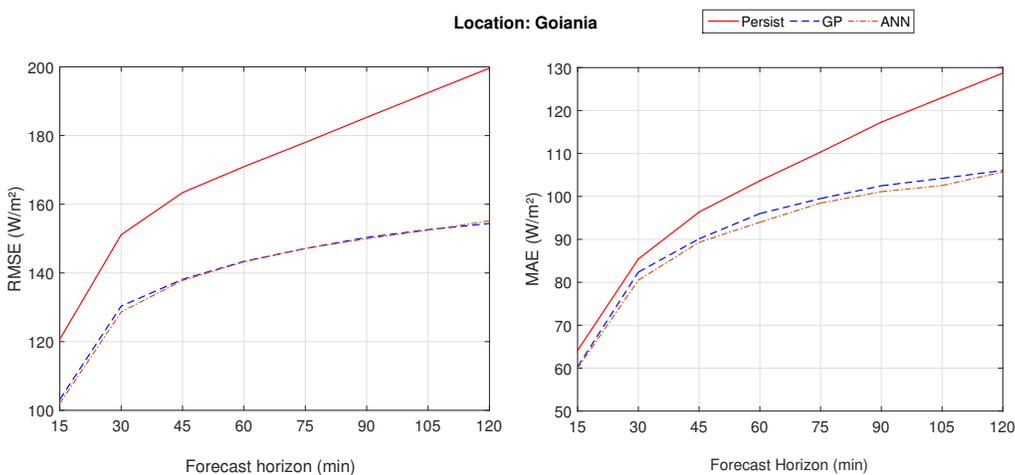
Figure 10. Accuracy of persistence, GP and ANN according to RMSE (**left**) and MAE (**right**) for Goiania, showing the dominance of ANN.

ANN_{ens} has proved to be consistently effective in the forecasts carried out for Goiania weather station, as expected, because the lower variations in sunshine duration along the year lead to a less biased dataset in terms of overfitting, as night period points are excluded from the dataset during the processing stage.

From a comparison of the results obtained by Haurwitz and Ineichen for clear sky index forecasts, it is possible to conclude that Ineichen k_t^* persistence produces lower errors than results obtained by Haurwitz for most of the locations and horizons of prediction. Nevertheless, as the AI methods used here are improved by employing exogenous inputs, a trend of clear sky model dominance over results from GP and ANN techniques was not achieved.
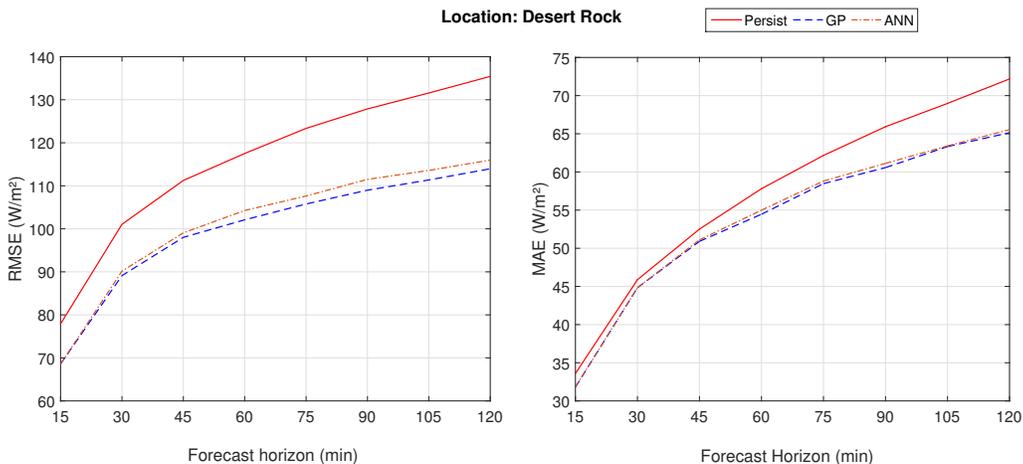
Figure 11. Accuracy of persistence, GP and ANN according to RMSE (**left**) and MAE (**right**) for Desert Rock, showing the dominance of GP.
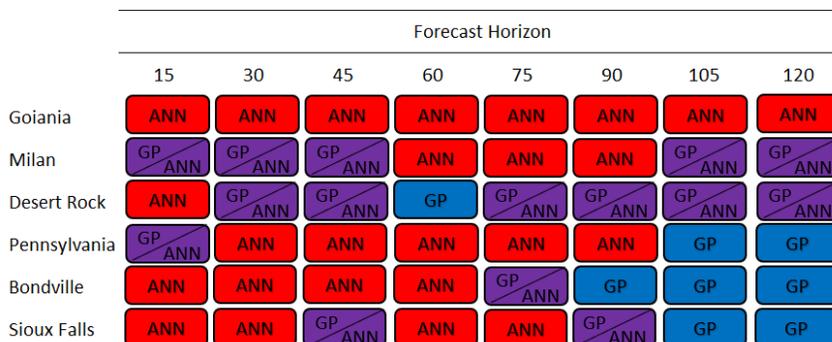
Figure 12. Model accuracy dominance by location and forecast horizon in ensemble forecasts. GP/ANN indicates cases in which accuracy dominance differs from the error metric evaluated.

6.4. Generic Results

The computation of averages based on multiple results is widely employed as a procedure to achieve reliable generalized results according to Rana et al. [17], although the use of averages does not disregard the importance of specific results. MAE and RMSE averages of all forecast horizons and locations were calculated in order to carry out a generic evaluation of accuracy for GP and ANN, and the results are presented in Figure 13. The average robustness of MAE and RMSE were similarly determined, and results are presented in Figure 14. From the generalized results, it is possible to assume that GP presents more accurate and robust forecast results in comparison to ANN for single forecasts; the ensemble strategy improves ANN forecasts more significantly than GP; the ANN ensemble generally presents the most accurate results; and both models produce similar forecastability, with little difference in terms of accuracy, indicating that GP can provide faster, more reliable and accurate predictions with lower computing complexity, while ANN can provide more accurate predictions using higher complexity and a time-demanding strategy.
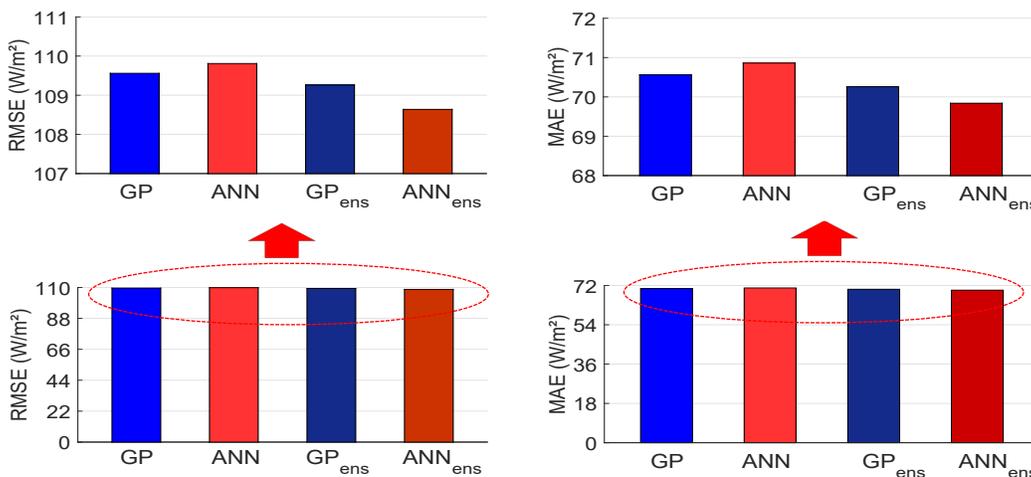
Figure 13. General accuracies of GP, ANN, GP ensemble and ANN ensemble for all sites according to RMSE values (left-hand graphs) and MAE values (right-hand graphs).
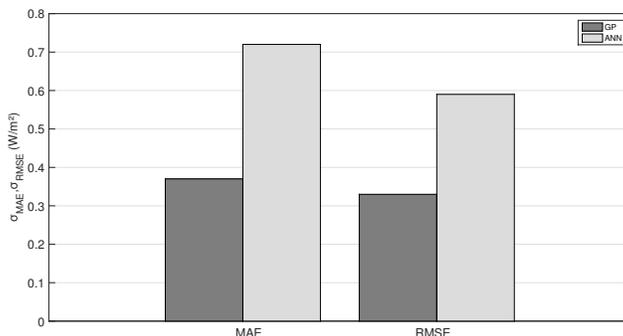
Figure 14. Comparison of general robustness of GP and ANN single forecasts according to MAE and RMSE.

A general comparison of clear sky indexes from multiple sites is exhibited in Table 4. From the analysis of results, it is possible to observe that the difference between Haurwitz k_t^* and Ineichen k_t^* forecast results is negligible, showing the low influence of the clear sky model on the accuracy of multivariate forecast results.

Table 4. Generalized accuracies for Haurwitz k_t^* and Ineichen k_t^* forecasts.

	RMSE	σ_{RMSE}	**MAE**	σ_{MAE}
Haurwitz	111.87	0.44	70.22	0.54
Ineichen	111.93	0.47	70.33	0.55

6.5. Regression Functions

Equation (16) presents an example of a regression function developed to forecast k_t^* (15), comprising a combination of the deterministic variable ω_s with previous values of k_t^* and the weather variables T_a and H_r. The algorithm has been proven to be efficient in selecting suitable variables to achieve accurate and robust models with generalization ability. Selected variables to develop regressions for Goiania station are expressed in Table 5.

$$k_t^*(15) = 0.535 + 0.98 \tanh k_t^*(-5) - 0.0049[T_a(-45) + \omega_s \cdot k_t^*(-20)] - 0.142[e^{-k_t^*(-35)\cdot k_t^*(-50)} + \cos k_t^*(-20)]$$

$$- 0.00141 H_r(-5) + 0.0244[e^{-\omega_s} - k_t^*(-5)k_t^*(-35)] + 0.00249\omega_s k_t^*(-20)e^{e^{-\omega_s}} \tag{16}$$

Table 5. Variables selected by GP regression models according to the forecast horizon for Goiania.

Forecast Horizon	Variables Selected
15 min	ω_s, $k_t^*(-5, -20, -35, -50)$, $H_r(-5)$, $T_a(-40)$
30 min	t_s, ω_s, $\hat{k_t^*}(15)$, $k_t^*(-5)$, $p_a(-25)$, $H_r(-40)$
45 min	ω_s, $\hat{k_t^*}(30)$, $H_r(-5, -35, -40)$, $T_a(-20)$, $p_a(-60)$
60 min	t_s, h, $\hat{k_t^*}(15, 45)$, $H_r(-15)$, $T_a(-40)$, $k_t^*(-45)$
75 min	ω_s, h, $\hat{k_t^*}(60)$, $p_a(-5, -10, -20)$, $T_a(-10, -55)$, $H_r(-10, -15)$, $W_s(-60)$
90 min	ω_s, $\hat{k_t^*}(30, 45, 75)$
105 min	ω_s, $\hat{k_t^*}(45, 60, 90)$
120 min	$Month$, $\hat{k_t^*}(105)$, $k_t^*(-25)$, $H_r(-30, -35)$, $T_a(-40)$

6.6. Comparison with the State-of-the-Art

A recent analysis of intraday solar irradiance forecasting at the SURFRAD weather stations has been carried out using regression and frequency domain models [21]. A direct comparison of the results obtained by regression, frequency domain and MGGP is presented in Table 6. Reikard et al. [21] analyzed forecasts for the same years, based on the same historical data and datasets used here. Although pieces of datasets used in each analysis are not guaranteed to be the same, a direct comparison of the results is able to ensure the suitability of the results of GP prediction.

Table 6. Comparison of state-of-the-art methods applied to intraday solar irradiance forecasting for Surface Radiation Network (SURFRAD) weather stations (best values in bold).

F.H.	Method	Desert Rock		Pennsylv. SU		Bondville		Sioux Falls	
		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
15	Regression	84.4	51.4	89.1	55.3	81.1	49.3	70.9	44.9
	Freq. Domain	84.2	51.0	91.0	56.1	82.5	50.1	73.9	46.5
	GP_{ens}	**68.3**	**31.6**	**81.7**	**46.9**	**72.0**	**40.8**	**67.6**	**37.7**
30	Regression	105.6	66.6	112.6	74.1	102.3	67.6	91.5	59.7
	Freq. Domain	108.1	63.0	112.0	73.2	102.2	66.9	92.1	60.3
	GP_{ens}	**89.0**	**44.7**	**105.4**	**65.8**	**90.6**	**56.2**	**86.5**	**52.8**
45	Regression	119.9	76.5	127.3	87.1	116.9	80.3	106.3	71.3
	Freq. Domain	119.1	71.7	125.1	86.1	114.5	78.8	106.6	69.4
	GP_{ens}	**97.4**	**50.5**	**115.2**	**74.0**	**100.9**	**64.2**	**96.5**	**60.5**

6.7. Machine Learning Algorithm Training Speed

Training machine learning algorithms to optimize results and accuracy is normally a time-consuming task. Table 7 presents a comparison of the average training times (in minutes) assessed for Goiania station according to each forecast horizon. Similar results were obtained for the other previously mentioned stations. Although MGGP has been demonstrated to be more robust for single forecasts, the training speed of this method is lower than that for ANN. Improvements of MGGP parameter tuning strategies should be considered in future studies in order to increase the speed of MGGP training.

Table 7. Comparison of training time required for each machine learning (ML) method, evaluated for the Goiania dataset (best values in bold).

ML Method	F.H.							
	15	30	45	60	75	90	105	120
GP	3.62	3.36	3.24	3.40	3.50	3.71	3.43	3.42
ANN	**0.89**	**0.47**	**0.44**	**0.34**	**0.35**	**0.45**	**0.39**	**0.35**

7. Conclusions

Machine learning algorithms are extensively adopted techniques for solar forecasting. This study proposed and evaluated multigene genetic programming (MGGP) as a novel machine learning algorithm, which is classified as a white box, with the aim of performing intraday solar irradiance forecasting. MGGP derives analytical regression functions that can be implemented without a specific software environment in any modern programming language using fundamental hardware. MGGP has been proven to consistently possess data generalization ability, providing robust and reliable solutions. The MGGP algorithm and another state-of-the-art MLP artificial neural network (ANN) algorithm were applied to datasets from six different locations from three different countries in order to compare results for forecast horizons from 15 to 120 min.

Data processing strategies were carefully analyzed in terms of input and output alternatives. Initial simulations were carried out for solar irradiance forecasting, using 15 minute time-windows as input data. Five minute time-window data, Haurwitz and Ineichen clear sky indexes were considered and combined with solar deterministic variables and weather variables in order to yield accurate forecast accuracies in terms of the data processing strategy.

The computation of MAE and RMSE as error metrics showed that the location, forecast horizon and error of evaluation impact the selection of the dominant model in terms of accuracy. MGGP and ANN typically yielded similar and consistent results. MGGP's utilization for single forecasts led to more accurate and robust results as opposed to ANN. Predictions were significantly improved for MGGP and ANN by adopting ensemble forecast, while the ensemble strategy improved ANN more extensively than MGGP. Regarding ensemble forecasts, MGGP was more accurate for a lower number of locations and evaluated forecast horizons in comparison to ANN, presenting the best forecast skills for Desert Rock station. MGGP predominantly accomplished more accurate prediction results for longer forecast horizons from 90 to 120 min ahead for different localities.

Based on a direct comparison with other state-of-the-art methods of forecasting applied to the same locations in USA, MGGP presented a relevant reduction in error and was proven to be a reliable and accurate approach for the analyzed localities.

The attributes of a locality have been proved to affect model dominance, showing that both MGGP and ANN can be applied to different locations. As suggestions for future investigation, studies may address hybridization strategies, ML algorithm improvement, advanced data processing strategies applied to MGGP forecasting and improvements in parameter tuning in order to enhance MGGP's training speed; furthermore, additional analyses of other solar parameters with the aim of improving the accuracy and performance of forecasting techniques could be undertaken.

Abbreviations

The following abbreviations are used in this manuscript:

AM	Air mass
ANN	Artificial neural network
ARIMA	Autoregressive integrated moving average
EA	Evolutionary algorithms
GBR	Gradient boosted regression
GP	Genetic programming
IEA	International energy agency
kNN	k-Nearest neighbors
MAE	Mean absolute error
MGGP	Multigene genetic programming
ML	Machine learning
MLP	Multilayer perceptron
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical weather prediction
PSU	Pennsylvania State University
PV	Photovoltaic
RES	Renewable energy sources
RF	Random forest
RMSE	Root mean squared error
SURFRAD	Surface Radiation Network
SVM	Support vector machine

Appendix A. Errors Obtained for Each Location, ML Algorithm and Forecast Horizon

Table A1. Forecast errors for Goiania (best values in bold). *Persist*: persistence.

FH	Method	Haur. k_t^* s	RMSE	σ_{RMSE}	MAE	σ_{MAE}	Inei. k_t^* s	RMSE	σ_{RMSE}	MAE	σ_{MAE}
15	*Persist*		120.67		64.13			120.64		63.85	
	GP	14.51	103.16	0.28	60.32	**0.29**	14.29	103.40	**0.37**	**60.09**	0.62
	ANN	**15.46**	**102.02**	0.27	**59.97**	0.44	**15.08**	**102.45**	**0.37**	60.19	**0.45**
	GP_ens	14.90	102.68		59.89		14.53	103.12		59.83	
	ANN_ens	**16.05**	**101.30**		59.09		15.87	101.50		59.16	
30	*Persist*		151.11		85.42			150.81		84.59	
	GP	13.75	130.34	**0.52**	82.36	**0.61**	13.56	130.36	**0.29**	82.31	**0.37**
	ANN	**14.87**	**128.64**	0.59	**80.49**	0.72	**14.49**	**128.95**	0.39	**81.02**	0.90
	GP_ens	14.11	129.79		81.84		13.94	129.78		81.81	
	ANN_ens	**15.52**	**127.65**		79.34		15.28	127.76		79.54	
45	*Persist*		163.39		96.35			162.72		94.92	
	GP	15.46	138.13	**0.31**	90.12	**0.36**	14.93	138.43	**0.44**	90.94	**0.28**
	ANN	**15.65**	**137.82**	0.37	**89.29**	0.40	**15.03**	**138.26**	0.50	**89.86**	0.65
	GP_ens	15.56	137.96		89.67		15.06	138.21		90.56	
	ANN_ens	**16.40**	**136.60**		87.99		16.03	136.64		88.43	
60	*Persist*		170.94		103.64			169.82		101.44	
	GP	16.13	143.36	**0.40**	96.00	**0.42**	15.92	**142.78**	**0.26**	95.69	**0.46**
	ANN	**16.21**	**143.22**	0.75	**93.98**	1.21	15.77	143.04	0.50	**94.52**	0.69
	GP_ens	16.45	142.82		95.61		16.07	142.53		95.44	
	ANN_ens	**17.03**	**141.83**		92.46		16.36	142.04		93.90	
75	*Persist*		178.00		110.33			176.44		107.51	
	GP	**17.36**	**147.11**	0.45	99.50	**0.41**	16.70	146.98	**0.25**	99.56	**0.41**
	ANN	17.34	147.15	0.66	**98.45**	0.79	16.42	147.47	0.52	**98.81**	0.50
	GP_ens	17.62	146.65		98.97		16.80	146.79		99.35	
	ANN_ens	**18.14**	**145.71**		96.98		17.25	146.00		97.42	
90	*Persist*		185.30		117.29			183.28		113.91	
	GP	18.86	150.35	**0.40**	102.47	**0.45**	18.18	**149.96**	0.40	102.16	**0.63**
	ANN	**19.08**	**149.95**	0.43	**101.11**	0.68	18.03	150.23	**0.28**	**101.49**	0.88
	GP_ens	19.09	149.91		102.13		18.35	149.65		101.73	
	ANN_ens	**19.88**	**148.46**		99.57		18.84	148.76		99.91	
105	*Persist*		192.48		123.01			190.04		119.03	
	GP	20.71	152.61	**0.35**	104.21	**0.53**	**19.91**	**152.21**	**0.36**	104.13	**0.35**
	ANN	**20.79**	**152.45**	0.50	**102.52**	0.56	19.45	153.09	0.81	**103.82**	1.30
	GP_ens	20.94	152.16		103.84		20.07	151.91		103.82	
	ANN_ens	**21.46**	**151.18**		101.29		20.20	151.65		102.47	
120	*Persist*		199.62		128.73			196.80		124.08	
	GP	**22.65**	**154.40**	0.27	106.04	**0.32**	21.58	154.33	**0.37**	**105.85**	**0.49**
	ANN	22.27	155.17	0.92	**105.71**	1.01	21.12	155.24	0.77	106.49	1.00
	GP_ens	22.80	154.10		105.79		**21.77**	**153.96**		105.50	
	ANN_ens	**23.09**	**153.52**		104.23		21.73	154.03		**105.26**	

Table A2. Forecast errors for Milan (best values in bold).

FH	Method	Haur. k_t^*					Inei. k_t^*				
		s	RMSE	σ_{RMSE}	MAE	σ_{MAE}	s	RMSE	σ_{RMSE}	MAE	σ_{MAE}
15	*Persist*		76.02		37.43			75.97		37.33	
	GP	11.63	67.18	**0.63**	**34.22**	0.38	11.35	67.34	**0.63**	**34.06**	0.30
	ANN	**12.54**	**66.48**	1.09	35.11	**0.31**	**12.06**	**66.80**	1.04	36.11	**0.25**
	GP$_{ens}$	12.23	66.72		**33.84**		11.95	66.89		**33.69**	
	ANN$_{ens}$	**14.81**	**64.76**		34.11		**14.01**	**65.32**		34.18	
30	*Persist*		99.41		51.12			99.29		50.80	
	GP	10.80	88.67	**0.32**	**48.95**	0.44	9.62	89.74	**0.35**	**50.18**	0.38
	ANN	**10.86**	**88.61**	0.71	50.23	0.51	9.24	90.12	1.80	51.49	1.09
	GP$_{ens}$	11.12	88.36		**48.65**		10.12	89.24		49.73	
	ANN$_{ens}$	**12.62**	**86.87**		48.90		**11.82**	**87.55**		48.92	
45	*Persist*		110.31		59.34			110.12		58.76	
	GP	**10.70**	**98.50**	**0.20**	56.46	0.42	9.18	**100.01**	0.38	56.92	0.52
	ANN	10.40	98.83	0.64	58.27	0.64	8.51	100.75	0.81	59.04	0.65
	GP$_{ens}$	10.87	98.32		**56.25**		9.46	99.70		**56.63**	
	ANN$_{ens}$	**12.01**	**97.06**		56.94		**11.02**	**97.99**		57.17	
60	*Persist*		120.03		66.08			119.82		65.28	
	GP	**11.58**	**106.12**	**0.22**	**62.04**	0.49	**11.09**	**106.52**	**0.35**	**62.78**	0.39
	ANN	11.55	106.16	0.61	63.80	0.79	10.48	107.27	0.78	63.56	0.47
	GP$_{ens}$	11.83	105.83		**61.75**		11.52	106.02		62.70	
	ANN$_{ens}$	**12.87**	**104.58**		62.50		**11.91**	**105.54**		61.69	
75	*Persist*		128.66		72.05			128.54		71.04	
	GP	**12.80**	**112.18**	**0.34**	**67.43**	**0.32**	**12.36**	**112.66**	0.65	67.20	**0.50**
	ANN	12.67	112.36	0.90	67.97	0.72	11.76	113.42	1.48	69.28	1.15
	GP$_{ens}$	12.44	112.65		67.22		12.50	112.47		**67.03**	
	ANN$_{ens}$	**13.94**	**110.72**		**66.63**		**13.35**	**111.38**		67.63	
90	*Persist*		136.10		77.66			136.15		76.32	
	GP	**13.60**	**117.59**	**0.40**	**70.97**	**0.29**	13.93	**117.18**	0.66	**71.16**	**0.49**
	ANN	13.52	117.69	0.53	72.62	1.02	13.39	117.91	0.85	72.66	1.37
	GP$_{ens}$	13.87	117.22		**70.78**		14.35	116.61		70.86	
	ANN$_{ens}$	**14.79**	**115.97**		71.25		**15.08**	**115.62**		70.74	
105	*Persist*		142.26		82.92			142.50		81.48	
	GP	**14.60**	**121.49**	**0.52**	**74.05**	**0.35**	**14.60**	**121.69**	**0.44**	**74.92**	**0.45**
	ANN	13.64	122.85	2.41	76.34	1.21	14.16	122.33	0.89	76.22	0.70
	GP$_{ens}$	14.77	121.26		**73.85**		14.77	121.46		74.71	
	ANN$_{ens}$	**15.45**	**120.29**		74.41		**15.71**	**120.11**		**74.56**	
120	*Persist*		147.60		87.74			148.21		86.14	
	GP	**15.70**	**124.43**	**0.36**	**76.90**	**0.28**	15.95	**124.58**	**0.36**	78.46	1.07
	ANN	15.03	125.42	1.21	79.81	1.64	15.57	125.14	1.01	79.72	1.26
	GP$_{ens}$	15.83	124.24		**76.73**		16.13	124.30		78.28	
	ANN$_{ens}$	**16.25**	**123.61**		78.47		**16.97**	**123.06**		**78.15**	

Table A3. Forecast errors for Desert Rock (best values in bold).

FH	Method	Haur. k_t^*					Inei. k_t^*				
		s	RMSE	σ_{RMSE}	MAE	σ_{MAE}	s	RMSE	σ_{RMSE}	MAE	σ_{MAE}
15	*Persist*		78.00		34.46			77.90		33.59	
	GP	11.96	68.68	**0.14**	**31.64**	**0.18**	11.94	68.60	**0.23**	**31.81**	**0.23**
	ANN	**11.98**	**68.66**	0.25	31.85	0.56	**11.98**	**68.57**	0.29	31.84	0.51
	GP$_{ens}$	12.26	68.44		31.29		12.28	68.34		31.58	
	ANN$_{ens}$	13.24	67.68		31.02		12.96	67.81		31.27	
30	*Persist*		101.32		47.74			101.05		45.88	
	GP	**11.45**	**89.72**	**0.57**	**45.24**	**0.41**	**11.80**	**89.12**	**0.09**	**44.82**	0.45
	ANN	10.74	90.44	0.71	44.93	0.43	10.78	90.16	0.53	44.83	**0.37**
	GP$_{ens}$	11.77	89.39		44.73		11.91	89.01		44.68	
	ANN$_{ens}$	11.60	89.57		44.15		11.62	89.31		44.33	
45	*Persist*		111.75		55.36			111.24		52.51	
	GP	**12.09**	**98.24**	**0.28**	**50.87**	**0.35**	**11.89**	98.02	1.41	**50.91**	1.08
	ANN	10.98	99.49	0.46	50.94	1.00	10.95	99.06	**0.22**	51.10	**0.67**
	GP$_{ens}$	12.21	98.11		50.72		12.47	97.37		50.47	
	ANN$_{ens}$	11.85	98.51		50.00		11.53	98.41		50.51	
60	*Persist*		118.32		61.63			117.51		57.80	
	GP	**13.05**	**102.88**	**0.32**	**55.24**	**0.38**	**13.13**	**102.09**	**0.15**	**54.46**	**0.16**
	ANN	11.81	104.34	1.21	56.51	1.01	11.28	104.26	0.29	54.99	0.47
	GP$_{ens}$	13.17	102.74		55.11		13.19	102.02		54.41	
	ANN$_{ens}$	12.77	103.20		55.35		11.90	103.53		54.47	
75	*Persist*		124.48		67.03			123.35		62.16	
	GP	**14.46**	**106.48**	**0.11**	**59.12**	**0.10**	**14.25**	**105.77**	**0.11**	**58.47**	0.34
	ANN	13.34	107.88	0.33	59.70	1.50	12.74	107.63	0.47	58.80	0.60
	GP$_{ens}$	14.56	106.36		58.42		14.32	105.69		57.81	
	ANN$_{ens}$	14.06	106.98		58.31		13.41	106.81		57.42	
90	*Persist*		129.33		71.66			127.87		65.93	
	GP	**15.18**	**109.70**	**0.20**	**61.55**	**0.36**	**14.77**	**108.98**	**0.15**	**60.57**	**0.24**
	ANN	13.87	111.40	0.76	61.60	0.91	12.81	111.50	0.80	61.12	0.65
	GP$_{ens}$	15.25	109.60		61.47		14.84	108.90		60.44	
	ANN$_{ens}$	14.72	110.30		60.67		13.55	110.55		59.99	
105	*Persist*		133.39		75.55			131.57		68.98	
	GP	**16.06**	**111.97**	**0.14**	**63.15**	**0.22**	**15.36**	**111.36**	**0.15**	**63.13**	**0.27**
	ANN	14.22	114.42	0.70	63.20	0.89	13.68	113.57	0.39	63.42	0.60
	GP$_{ens}$	16.11	111.91		63.49		15.45	111.24		62.52	
	ANN$_{ens}$	14.99	113.40		62.11		14.34	112.70		62.02	
120	*Persist*		137.62		79.56			135.42		72.19	
	GP	**16.83**	**114.46**	**0.20**	66.09	**0.25**	**15.85**	**113.96**	**0.15**	**65.14**	**0.29**
	ANN	15.60	116.15	0.56	**65.59**	0.52	14.36	115.98	0.42	65.54	1.16
	GP$_{ens}$	16.91	114.35		65.98		15.91	113.88		65.07	
	ANN$_{ens}$	16.24	115.27		64.86		15.07	115.01		64.83	

Table A4. Forecast errors for Pennsylvania State University (best values in bold).

FH	Method	Haur. k_t^* s	RMSE	σ_{RMSE}	MAE	σ_{MAE}	Inei. k_t^* s	RMSE	σ_{RMSE}	MAE	σ_{MAE}
	Persist		94.43		51.76			94.37		51.49	
	GP	12.78	82.36	0.44	**47.47**	**0.34**	12.99	82.11	**0.28**	**47.23**	**0.14**
15	*ANN*	**14.42**	**80.82**	0.29	47.49	0.45	**13.86**	81.29	0.29	48.17	0.53
	GP_{ens}	13.31	81.86		47.10		13.39	81.73		**46.92**	
	ANN_{ens}	**14.98**	**80.28**		47.00		**14.49**	80.69		47.57	
	Persist		118.31		68.46			118.13		67.85	
	GP	10.71	105.64	0.26	66.56	**0.28**	10.46	105.78	**0.26**	**66.02**	**0.28**
30	*ANN*	**11.70**	**104.47**	0.17	65.93	0.61	**11.54**	104.49	0.36	66.47	0.37
	GP_{ens}	11.05	105.23		66.23		10.77	105.40		65.78	
	ANN_{ens}	**12.34**	**103.72**		**65.28**		**12.29**	103.62		65.66	
	Persist		131.38		78.48			131.05		77.51	
	GP	12.32	115.19	**0.23**	**74.60**	**0.40**	11.85	115.52	**0.21**	**74.30**	**0.29**
45	*ANN*	**13.07**	**114.21**	0.28	74.80	0.63	**12.71**	114.39	0.23	74.52	0.45
	GP_{ens}	12.61	114.82		74.26		12.06	115.24		74.01	
	ANN_{ens}	**13.76**	**113.30**		**74.07**		**13.44**	113.44		73.69	
	Persist		138.63		84.75			138.19		83.45	
	GP	12.54	121.25	**0.06**	79.96	**0.17**	12.17	121.37	**0.29**	80.29	**0.31**
60	*ANN*	**13.11**	**120.45**	0.36	**79.94**	0.87	**12.67**	120.69	0.41	**80.03**	0.72
	GP_{ens}	12.68	121.06		79.78		12.39	121.06		80.00	
	ANN_{ens}	**13.68**	**119.66**		**79.28**		**13.42**	119.64		79.17	
	Persist		144.19		90.43			143.68		88.84	
	GP	12.65	125.95	**0.11**	84.92	**0.21**	12.40	125.86	0.86	84.80	0.97
75	*ANN*	**12.88**	**125.62**	0.48	**84.87**	0.98	**12.57**	125.63	**0.42**	**84.71**	**0.40**
	GP_{ens}	12.86	125.65		84.71		12.77	125.33		84.45	
	ANN_{ens}	**13.57**	**124.63**		**84.10**		**13.27**	124.61		83.81	
	Persist		150.58		96.09			150.03		94.27	
	GP	**13.34**	**130.50**	0.36	**89.32**	**0.20**	13.63	**129.59**	**0.29**	**89.07**	**0.37**
90	*ANN*	13.27	130.59	0.61	89.37	0.98	13.06	130.43	0.50	89.50	0.68
	GP_{ens}	13.52	130.23		89.09		13.77	129.37		88.89	
	ANN_{ens}	**14.13**	**129.30**		**88.42**		**13.89**	129.19		88.57	
	Persist		156.99		101.08			156.42		99.04	
	GP	**14.97**	**133.48**	**0.13**	**92.24**	**0.21**	14.69	**133.44**	0.38	**92.10**	**0.43**
105	*ANN*	14.38	134.42	0.23	92.99	0.94	13.85	134.75	0.45	93.37	0.53
	GP_{ens}	**15.14**	**133.22**		**92.13**		**14.90**	**133.11**		91.71	
	ANN_{ens}	15.08	133.31		92.23		14.67	133.46		92.44	
	Persist		164.37		106.50			163.82		104.31	
	GP	**16.23**	**137.70**	**0.21**	**95.99**	**0.15**	16.21	**137.27**	**0.22**	**95.95**	**0.22**
120	*ANN*	15.50	138.90	0.82	97.39	0.93	15.08	139.12	0.74	97.70	1.02
	GP_{ens}	**16.36**	**137.48**		**95.85**		**16.37**	**137.00**		95.73	
	ANN_{ens}	16.25	137.66		96.57		15.91	137.75		96.73	

Table A5. Forecast errors for Bondville (best values in bold).

FH	Method	Haur. k_t^* s	RMSE	σ_{RMSE}	MAE	σ_{MAE}	Inei. k_t^* s	RMSE	σ_{RMSE}	MAE	σ_{MAE}
15	*Persist*		81.28		43.45			81.20		43.01	
	GP	10.43	72.81	0.77	41.51	**0.35**	10.85	72.39	**0.41**	**41.05**	**0.15**
	ANN	**11.74**	**71.74**	**0.49**	40.87	0.59	**11.10**	72.18	**0.41**	41.26	0.34
	GP_{ens}	11.31	72.09		41.05		11.30	72.02		40.84	
	ANN_{ens}	**12.71**	**70.95**		**40.22**		**11.92**	**71.52**		**40.71**	
30	*Persist*		101.49		57.67			101.25		56.79	
	GP	9.97	91.37	**0.47**	56.08	**0.56**	10.23	**90.89**	0.28	**56.38**	0.29
	ANN	**10.16**	91.18	0.48	**55.92**	0.73	9.57	91.56	0.29	56.50	**0.21**
	GP_{ens}	10.35	90.98		55.75		**10.52**	**90.61**		56.15	
	ANN_{ens}	**10.94**	**90.38**		**55.32**		10.41	90.71		**55.78**	
45	*Persist*		113.22		66.80			112.82		65.39	
	GP	9.83	102.09	0.52	64.38	**0.48**	10.28	**101.22**	0.42	**64.37**	0.53
	ANN	**10.61**	**101.20**	0.32	63.27	0.68	9.82	101.74	0.89	64.49	1.34
	GP_{ens}	10.26	101.60		63.92		10.56	100.91		64.15	
	ANN_{ens}	**11.37**	**100.35**		**62.57**		**10.74**	**100.71**		**63.72**	
60	*Persist*		121.55		73.53			120.99		71.70	
	GP	**11.65**	**107.39**	0.35	**69.37**	**0.42**	**11.16**	**107.48**	0.46	**69.76**	0.79
	ANN	11.38	107.73	**0.33**	69.37	0.48	10.72	108.02	**0.37**	69.98	**0.54**
	GP_{ens}	12.06	106.90		68.94		**11.58**	**106.97**		69.30	
	ANN_{ens}	**12.15**	**106.79**		**68.64**		11.53	107.04		**69.15**	
75	*Persist*		127.95		79.19			127.21		77.04	
	GP	**12.01**	**112.58**	**0.11**	**73.78**	**0.20**	**11.19**	**112.97**	0.30	**73.98**	0.32
	ANN	11.52	113.21	0.49	74.30	0.83	10.62	113.71	1.01	74.83	1.01
	GP_{ens}	12.22	112.31		**73.47**		11.43	112.67		**73.70**	
	ANN_{ens}	**12.29**	**112.23**		73.56		**11.62**	**112.43**		73.77	
90	*Persist*		134.49		84.73			133.56		82.22	
	GP	**13.22**	**116.71**	0.19	**77.69**	**0.21**	**12.15**	**117.33**	0.22	**78.67**	0.28
	ANN	12.24	118.03	0.51	78.32	0.74	11.40	118.33	0.71	78.93	0.48
	GP_{ens}	**13.37**	**116.51**		**77.49**		12.29	117.14		78.49	
	ANN_{ens}	12.99	117.02		77.53		**12.44**	**116.94**		**77.77**	
105	*Persist*		140.34		89.83			139.23		86.79	
	GP	**13.52**	**121.37**	0.32	**81.88**	**0.39**	**12.21**	**122.23**	0.22	**82.32**	0.45
	ANN	12.56	122.71	0.69	82.06	0.52	11.90	122.66	0.55	82.70	0.45
	GP_{ens}	**13.87**	**120.87**		**81.01**		12.43	121.93		82.10	
	ANN_{ens}	13.40	121.53		81.18		**12.96**	**121.19**		**81.49**	
120	*Persist*		145.88		95.01			144.58		91.41	
	GP	**14.42**	**124.84**	0.35	**85.95**	**0.31**	**12.96**	**125.85**	0.72	**86.30**	0.55
	ANN	13.31	126.46	0.77	86.47	1.35	12.54	126.46	**0.55**	86.82	0.76
	GP_{ens}	**14.59**	**124.59**		**85.07**		13.25	125.42		85.79	
	ANN_{ens}	14.45	124.80		85.30		**13.59**	**124.93**		**85.66**	

Energies **2020**, *13*, 3005
25 of 28

Table A6. Forecast errors for Sioux Falls (best values in bold).

FH	Method	Haur. k_t^* s	RMSE	σ_{RMSE}	MAE	σ_{MAE}	Inei. k_t^* s	RMSE	σ_{RMSE}	MAE	σ_{MAE}
15	*Persist*		75.47		40.51			75.40		40.06	
	GP	10.09	67.85	0.30	38.17	0.24	10.13	67.76	**0.23**	**37.82**	**0.19**
	ANN	**12.25**	**66.23**	**0.20**	**38.02**	**0.18**	**11.70**	**66.58**	1.04	38.41	0.25
	GP$_{ens}$	10.46	67.58		37.91		10.32	67.61		**37.70**	
	ANN$_{ens}$	**13.04**	**65.63**		**37.55**		**12.76**	**65.78**		37.78	
30	*Persist*		95.39		54.00			95.22		53.11	
	GP	9.22	86.60	**0.10**	**52.86**	**0.17**	8.71	86.93	0.28	52.97	**0.25**
	ANN	**9.51**	**86.32**	0.45	53.15	0.51	**9.17**	**86.49**	**0.26**	**52.82**	0.42
	GP$_{ens}$	9.31	86.51		52.77		9.20	86.46		52.77	
	ANN$_{ens}$	**10.38**	**85.49**		**52.56**		**9.95**	**85.75**		**52.27**	
45	*Persist*		107.46		62.87			107.17		61.44	
	GP	9.56	97.18	**0.21**	**60.86**	**0.42**	**9.72**	**96.75**	**0.09**	**60.69**	**0.42**
	ANN	**9.84**	**96.89**	0.51	61.69	0.69	9.57	96.91	0.30	61.03	0.48
	GP$_{ens}$	9.79	96.94		**60.61**		10.00	96.45		**60.46**	
	ANN$_{ens}$	**10.62**	**96.05**		61.13		**10.23**	**96.20**		60.52	
60	*Persist*		116.51		69.81			116.12		67.95	
	GP	**10.59**	**104.17**	**0.11**	**66.92**	**0.23**	**10.42**	**104.03**	**0.17**	67.09	**0.30**
	ANN	10.16	104.68	0.38	67.47	0.33	10.20	104.27	0.45	**67.05**	0.46
	GP$_{ens}$	10.72	104.02		66.80		10.52	103.90		66.96	
	ANN$_{ens}$	**10.98**	**103.72**		**66.74**		**10.98**	**103.37**		**66.39**	
75	*Persist*		123.64		75.45			123.21		73.35	
	GP	**11.18**	**109.82**	**0.20**	**71.76**	**0.33**	**11.01**	**109.64**	**0.14**	71.93	**0.39**
	ANN	10.92	110.14	0.58	72.21	0.86	10.76	109.95	0.33	**71.58**	0.67
	GP$_{ens}$	11.37	109.58		71.56		11.18	109.43		71.73	
	ANN$_{ens}$	**11.75**	**109.11**		**71.49**		**11.51**	**109.03**		**70.92**	
90	*Persist*		130.98		81.05			130.57		78.71	
	GP	**12.47**	**114.65**	**0.19**	**75.83**	**0.15**	**12.31**	**114.49**	**0.11**	**75.46**	**0.35**
	ANN	11.84	115.48	0.42	76.45	1.00	11.60	115.42	0.27	76.37	0.76
	GP$_{ens}$	12.65	114.42		**75.62**		12.42	114.35		**75.33**	
	ANN$_{ens}$	**12.71**	**114.34**		75.66		**12.43**	**114.34**		75.59	
105	*Persist*		138.49		86.65			138.10		83.89	
	GP	**13.92**	**119.20**	**0.15**	**79.62**	**0.34**	**13.70**	**119.18**	**0.25**	**79.31**	**0.21**
	ANN	12.90	120.62	0.20	80.50	0.59	12.99	120.17	0.50	80.39	0.79
	GP$_{ens}$	**14.12**	**118.94**		**79.41**		**13.94**	**118.85**		**79.10**	
	ANN$_{ens}$	13.81	119.36		79.65		13.88	118.94		79.52	
120	*Persist*		143.72		91.37			143.38		88.30	
	GP	**14.59**	**122.75**	**0.32**	**82.76**	**0.20**	**13.98**	**123.34**	0.92	**83.06**	1.03
	ANN	13.71	124.02	0.63	84.05	0.48	13.34	124.26	**0.60**	83.99	**0.81**
	GP$_{ens}$	**14.76**	**122.50**		**82.52**		**14.34**	**122.82**		**82.70**	
	ANN$_{ens}$	14.52	122.85		83.29		14.23	122.97		83.11	

References

1. Liang, H.; Tamang, A.K.; Zhuang, W.; Shen, X.S. Stochastic information management in smart grid. *IEEE Commun. Surv. Tutor.* **2014**, *16*, 1746–1770. [CrossRef]
2. Bagheri, A.; Zhao, C.; Qiu, F.; Wang, J. Resilient transmission hardening planning in a high renewable penetration era.*IEEE Trans. Power Syst.* **2019**, *34*, 873–882. [CrossRef]
3. Lahon, R.; Gupta, C.P. Energy management of cooperative microgrids with high-penetration renewables. *IET Renew. Power Gener.* **2018**, *12*, 680–690. [CrossRef]

4.   Lauret, P.; Perez, R.; Aguiar, L.M.; Tapachès, E.; Diagne, H.M.; David, M. Characterization of the intraday variability regime of solar irradiation of climatically distinct locations. *Sol. Energy* **2016**, *125*, 99–110. [CrossRef]

5.   Shahriari, M.; Blumsack, S. Scaling of wind energy variability over space and time. *Appl. Energy* **2017**, *195*, 572–585. [CrossRef]

6.   Bakirtzis, E.A.; Biskas, P.N. Multiple time resolution stochastic scheduling for systems with high renewable penetration. *IEEE Trans. Power Syst.* **2017**, *32*, 1030–1040. [CrossRef]

7.   Du, E.; Zhang, N.; Hodge, B.; Wang, Q.; Lu, Z.; Kang, C.; Kroposki, B.; Xia, Q. Operation of a high renewable penetrated power system with csp plants: A look-ahead stochastic unit commitment model. *IEEE Trans. Power Syst.* **2019**, *34*, 140–151. [CrossRef]

8.   Pelland, S.; Remund, J.; Keissl, J.; Oozeki, T.; Brabandere, K.D. *Photovoltaic and Solar Forecasting: State of the Art*; Tech. rep.; International Energy Agency: Paris, France, 2013.

9.   Yang, D.; Kleissl, J.; Gueymard, C.A.; Pedro, H.T.; Coimbra, C.F. History and trends in solar irradiance and pv power forecasting: A preliminary assessment and review using text mining. *Sol. Energy* **2018**, *168*, 60–101, [CrossRef]

10.   Diagne, M.; David, M.; Lauret, P.; Boland, J.; Schmutz, N. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renew. Sustain. Energy Rev.* **2013**, *27*, 65–76. [CrossRef]

11.   Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.-L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* **2017**, *105*, 569–582. [CrossRef]

12.   Sperati, S.; Alessandrini, S.; Pinson, P.; Kariniotakis, G. The "weather intelligence for renewable energies" benchmarking exercise on short-term forecasting of wind and solar power generation. *Energies* **2015**, *8*, 9594–9619. [CrossRef]

13.   Li, J.; Ward, J.K.; Tong, J.; Collins, L.; Platt, G. Machine learning for solar irradiance forecasting of photovoltaic system. *Renew. Energy* **2016**, *90*, 542–553. [CrossRef]

14.   Pedro, H.T.; Coimbra, C.F. Short-term irradiance forecastability for various solar micro-climates. *Sol. Energy* **2015**, *122*, 587–602. [CrossRef]

15.   Benali, L.; Notton, G.; Fouilloy, A.; Voyant, C.; Dizene, R. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renew. Energy* **2019**, *132*, 871–884. [CrossRef]

16.   Persson, C.; Bacher, P.; Shiga, T.; Madsen, H. Multi-site solar power forecasting using gradient boosted regression trees. *Solar Energy* **2017**, *150*, 423–436. [CrossRef]

17.   Rana, M.; Koprinska, I.; Agelidis, V.G. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. *Energy Convers. Manag.* **2016**, *121*, 380–390. [CrossRef]

18.   Zeng, J.; Qiao, W. Short-term solar power prediction using a support vector machine. *Renew. Energy* **2013**, *52*, 118–127. [CrossRef]

19.   Dolara, A.; Grimaccia, F.; Leva, S.; Mussetta, M.; Ogliari, E. A Physical Hybrid Artificial Neural Network for Short Term Forecasting of PV Plant Power Output. *Energies* **2015**, *8*, 1138–1153. [CrossRef]

20.   Nobre, A.M.; Severiano, C.A.; Karthik, S.; Kubis, M.; Zhao, L.; Martins, F.R.; Pereira, E.B.; Rüther, R.; Reindl, T. Pv power conversion and short-term forecasting in a tropical, densely-built environment in Singapore. *Renew. Energy* **2016**, *94*, 496–509. [CrossRef]

21.   Reikard, G.; Hansen, C. Forecasting solar irradiance at short horizons: Frequency and time domain models. *Renew. Energy* **2019**, *135*, 1270–1290. [CrossRef]

22.   Yang, H.; Kurtz, B.; Nguyen, D.; Urquhart, B.; Chow, C.W.; Ghonima, M.; Kleissl, J. Solar irradiance forecasting using a ground-based sky imager developed at uc san diego. *Sol. Energy* **2014**, *103*, 502–524. [CrossRef]

23.   Chow, C.W.; Urquhart, B.; Lave, M.; Dominguez, A.; Kleissl, J.; Shields, J.; Washom, B. Intra-hour forecasting with a total sky imager at the uc san diego solar energy testbed. *Sol. Energy* **2011**, *85*, 2881–2893. [CrossRef]

24.   Dong, Z.; Yang, D.; Reindl, T.; Walsh, W.M. Satellite image analysis and a hybrid esss/ann model to forecast solar irradiance in the tropics. *Energy Convers. Manag.* **2014**, *79*, 66–73. [CrossRef]

25.   Bessa, R.J.; Trindade, A.; Miranda, V. Spatial-temporal solar power forecasting for smart grids. *IEEE Trans. Ind. Informat.* **2015**, *11*, 232–241. [CrossRef]

26. Gutierrez-Corea, F.-V.; Manso-Callejo, M.-A.; Moreno-Regidor, M.-P.; Manrique-Sancho, M.-T. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. *Sol. Energy* **2016**, *134*, 119–131. [CrossRef]

27. 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]

28. Solar lab EMC/UFG. Federal University of Goias. Brazil. Available online: https://sites.google.com/site/sfvemcufg/ (accessed on 4 June 2020).

29. SoDa. Solar Energy Services for Professionals. France. 2017. Available online: http://www.soda-pro.com/home;jsessionid=B032D33B0AD3460B741E14E41CC46BE2 (accessed on 4 June 2020).

30. SolarTech Lab. Politecnico di Milano. Italy. Available online: http://www.solartech.polimi.it/ (accessed on 4 June 2020).

31. de Paiva, G.M.; Pimentel, S.P.; Leva, S.; Mussetta, M. Intelligent approach to improve genetic programming based intra-day solar forecasting models. In Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8. [CrossRef]

32. Reno, M.J.; Hansen, C.W.; Stein, J.S. *Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis*; Tech. rep.; Sandia National Laboratories: Albuquerque, NM, USA, 2012.

33. Duffie, J.A.; Beckman, W.A. *Solar Engineering of Thermal Processes*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013.

34. Ineichen, P.; Perez, R. A new airmass independent formulation for the linke turbidity coefficient. *Sol. Energy* **2002**, *73*, 151–157. [CrossRef]

35. Ineichen, P. Comparison of eight clear sky broadband models against 16 independent data banks. *Sol. Energy* **2006**, *80*, 468–478. [CrossRef]

36. Engerer, N.; Mills, F. Kpv: A clear-sky index for photovoltaics. *Sol. Energy* **2014**, *105*, 679–693. [CrossRef]

37. Koza, J.R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*; MIT Press: Cambridge, MA, USA, 1992.

38. Brownlee, J. *Clever Algorithms: Nature-Inspired Programming Recipes*, 1st ed.; Lulu Press, Inc.: Morrisville, NC, USA, 2011.

39. Lee, Y.-S.; Tong, L.-I. Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming. *Knowl.-Based Syst.* **2011**, *24*, 66–72. [CrossRef]

40. Garg, A.; Sriram, S.; Tai, K. Empirical analysis of model selection criteria for genetic programming in modeling of time series system. In Proceedings of the 2013 IEEE Conference on Computational Intelligence for Financial Engineering Economics (CIFEr), Singapore, Singapore, 16–19 April 2013; IEEE: Piscataway, NJ, USA; pp. 90–94. [CrossRef]

41. Mehr, A.D.; Kahya, E.; Olyaie, E. Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. *J. Hydrol.* **2013**, *505*, 240–249. [CrossRef]

42. Ghorbani, M.A.; Khatibi, R.; Mehr, A.D.; Asadi, H. Chaos-based multigene genetic programming: A new hybrid strategy for river flow forecasting. *J. Hydrol.* **2018**, *562*, 455–467. [CrossRef]

43. Mehr, A.D.; Jabarnejad, M.; Nourani, V. Pareto-optimal mpsa-mggp: A new gene-annealing model for monthly rainfall forecasting. *J. Hydrol.* **2019**, *571*, 406–415. [CrossRef]

44. Russo, M.; Leotta, G.; Pugliatti, P.; Gigliucci, G. Genetic programming for photovoltaic plant output forecasting. *Sol. Energy* **2014**, *105*, 264–273. [CrossRef]

45. Pan, I.; Pandey, D.S.; Das, S. Global solar irradiation prediction using a multi-gene genetic programming approach. *J. Renew. Sustain. Energy* **2013**, *5*, 063129. [CrossRef]

46. Ghimire, S.; Deo, R.C.; Downs, N.J.; Raj, N. Global solar radiation prediction by ann integrated with european centre for medium range weather forecast fields in solar rich cities of queensland australia. *J. Clean. Prod.* **2019**, *216*, 288–310. [CrossRef]

47. Searson, D.; Leahy, D.; Willis, M. GPTIPS: An open source genetic programming toolbox for multigene symbolic regression. In Proceedings of the International Multiconference of Engineers and Computer Scientists, Hong Kong, 17–19 March 2010; IAENG: Hong Kong, 2010; pp. 77–80.
48. Leva, S.; Dolara, A.; Grimaccia, F.; Mussetta, M.; Ogliari, E. Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. *Math. Comput. Simul.* **2017**, *131*, 88–100, [CrossRef]
49. Eiben, A.; Smit, S. Parameter tuning for configuring and analyzing evolutionary algorithms. *Swarm Evol. Comput.* **2011**, *1*, 19–31. [CrossRef]
50. Lima, E.B.; Pappa, G.L.; Almeida, J.M.; Gonçalves, M.A.; Meira, W. Tuning genetic programming parameters with factorial designs. In Proceedings of the IEEE Congress on Evolutionary Computation, Barcelona, Spain, 18–23 July, 2010; IEEE: Piscataway, NJ, USA; pp. 1–8. [CrossRef]
51. Poli, R.; Langdon, W.B.; McPhee, N.F. (with contributions by J. R. Koza). *A Field Guide to Genetic Programming*; Lulu Press, Inc.: Morrisville, NC, USA, 2008.
52. Poli, R.; McPhee, N.F.; Vanneschi, L. Elitism reduces bloat in genetic programming. In Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation, Atlanta, GA, USA, 8–12 July, 2008; Association for Computing Machinery: New York, NY, USA; pp. 1343–1344. [CrossRef]