

## Article

# Modelling and Prediction of Monthly Global Irradiation Using Different Prediction Models

Cecilia Martinez-Castillo<sup>1</sup>, Gonzalo Astray<sup>2,3,\*</sup> and Juan Carlos Mejuto<sup>2,\*</sup> 

<sup>1</sup> Department of Analytical and Food Chemistry, Nutrition and Bromatology, Faculty of Sciences, University of Vigo, 32004 Ourense, Spain; martinez.castillo.cecilia@gmail.com

<sup>2</sup> Department of Physical Chemistry, Faculty of Sciences, University of Vigo, 32004 Ourense, Spain

<sup>3</sup> CITACA, University of Vigo, Campus Auga, 32004 Ourense, Spain

\* Correspondence: gastray@uvigo.es (G.A.); xmejuto@uvigo.es (J.C.M.)

**Abstract:** Different prediction models (multiple linear regression, vector support machines, artificial neural networks and random forests) are applied to model the monthly global irradiation (MGI) from different input variables (latitude, longitude and altitude of meteorological station, month, average temperatures, among others) of different areas of Galicia (Spain). The models were trained, validated and queried using data from three stations, and each best model was checked in two independent stations. The results obtained confirmed that the best methodology is the ANN model which presents the lowest RMSE value in the validation and querying phases 1226 kJ/(m<sup>2</sup>·day) and 1136 kJ/(m<sup>2</sup>·day), respectively, and predict conveniently for independent stations, 2013 kJ/(m<sup>2</sup>·day) and 2094 kJ/(m<sup>2</sup>·day), respectively. Given the good results obtained, it is convenient to continue with the design of artificial neural networks applied to the analysis of monthly global irradiation.

**Keywords:** prediction; solar irradiation; artificial neural network; random forest; vector support machine



**Citation:** Martinez-Castillo, C.; Astray, G.; Mejuto, J.C. Modelling and Prediction of Monthly Global Irradiation Using Different Prediction Models. *Energies* **2021**, *14*, 2332. <https://doi.org/10.3390/en14082332>

Academic Editor: Silvio Simani

Received: 25 March 2021

Accepted: 16 April 2021

Published: 20 April 2021

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



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Solar radiation exerts its influence over all Earth's processes related to the environment, plant growing and even over human activities development [1]. At ground level, the solar radiation data are important for a large number of applications related to agricultural hydrology, plant growth and others [1]. Besides these, global solar irradiation is a significant parameter in renewable energy applications (for example to determine size and model photovoltaic systems) [2].

Global solar irradiation measurements can be obtained using specific devices [3] which are limited to a small number of meteorological stations, probably due to their high cost and other inconveniences such as a need for regular calibration and maintenance [3,4]. Besides this, these data may not be accessible because the meteorological observatories that include measurement series of solar irradiation are still rarely distributed and these data present, sometimes, a problematic spatial interpolation in areas of intricate orography [5]. According to Notton et al. [6], there are only some 1000 continental stations over the world that can measure solar radiation (number reported by Notton et al., from the World Radiation Data Center (WRDC) [7]) although, currently this number is probably higher.

According to different authors, the shortage, difficulties and uncertainties of these measurements can be estimated from other more abundant variables (climatological properties) such as cloudiness, among others [3,5].

Take into account the increase in energy demand and consumption worldwide, and the search for alternatives to the decrease in fossil fuel reserves [8], it can be understood that the determination of the global solar irradiation can be very important in solar energy conversion systems. Solar photovoltaic energy has presented an important growth in the last years due to its cost reduction [9]. In this energetic context, the Spanish climatic

conditions can obtain a high performance using photovoltaic solar energy [10]. Within this territory, the Galicia area is suitable for solar installations [8].

Taking into account all the above, to the design of solar energy conversion systems, it is necessary to have solar irradiation monthly average data, and these need to be reliable [5]. Due to this, different techniques have been developed to find a correlation between solar irradiation and other variables such as relative humidity, air temperature, among others [2]. Traditionally, the estimation has been carried out using parametric-empirical models which obtained a relatively high certainly level [9].

Possibly, the three principal groups of models used to forecast solar radiation are machine learning, physical (or numerical weather prediction) and sky imaging [11]. Machine learning is an artificial intelligence subfield that studies and develop mathematical algorithms intended to comprehend data and obtain data without a prearranged model algorithm [12]. The machine learning models can find the relationship between inputs and outputs variables, which allow that these models can be used, sometimes, in classification problems, forecasting problems, among others [13].

Different methodologies such as multiple linear regression (MLR), support vector machines (SVM), artificial neural networks (ANN and random forest (RF) are available to prediction purposes.

- Multiple linear regression can be used in environmental science to determine fine particles ( $PM_1$ ) concentration using environmental, meteorological and physical eventualities variables [14], or to determine diffuse pollutant discharge using different basin's environmental parameters [15].
- Support vector machines can also be used in different fields such as Bioinformatics to identify single-nucleotide polymorphisms [16], to predict dihedral angle regions [17] or to forecast interface residue pairs of protein trimers [18].
- Artificial neural networks can be used in food science to model and optimize the extraction of cashew apple juice [19], to optimize an enzymatic approach to obtain modified artichoke pectin and pectic oligosaccharides [20] or to determine the broccoli buds loss green color velocity using hyperspectral camera combined with artificial neural networks [21].
- Finally, random forest can be used in economics to early prediction of university dropouts [22] or to predict the clean energy stock prices [23].

These types of approximations can be carried out independently or, as in some studies, simultaneously to be compared in their predictive capacity. This is the case of the study carried out by Torkashvand et al. [24] that compared two of these models, MLR and ANN, for the prediction of fruit firmness after six months using different input variables (nutrients concentrations alone, nutrient concentrations combination, among others) and show that, in general, the ANN model showed bigger potential to determine six month kiwifruit firmness [24]. On the other hand, Niu et al. [25] carried out a comparison between three kind of models proposed in this research, multiple linear regression, artificial neural network and support vector machine in the hydropower operation field and they concluded that the artificial neural networks and the support vector machine provide better performances than the MLR model [25]. A similar treatment can be seen in the research carried out by Lee et al. [26] in which these researchers developed simple/multiple LR, RF and SVM models to predict canopy nitrogen weight in corn using multispectral images obtained by an unmanned aerial vehicle. Authors concluded that the RF models presented the best results for the validation set and verified that when more spectral variables were used the model improved the accuracy and make longer the overall processing time [26].

The aim of this research was to develop different prediction models (linear regression, artificial neural networks, support vector machines and random forest) to model the monthly global irradiation (*MGI*) from three meteorological stations located in the Autonomous Community of Galicia (Spain) and then generalize the knowledge to other two nearby stations. These models will allow determining the monthly global irradiation in different areas of the Galicia based on the meteorological variables used, thus being able

to obtain the value of this variable in places where it had not previously been measured, which could facilitate the photovoltaic installations development. This work corresponds to the beginnings of a more ambitious study to model the monthly solar irradiation, and subsequently predict the monthly solar irradiation with a month in advance. This work is a summary of the final degree project developed by the first author of this research [27].

## 2. Related Works

In this work, four models have been chosen to predict the monthly global irradiation. Multiple linear regression models have been chosen because they are probably the simplest and fastest models that can be developed to try to determine the monthly global irradiation. The models based on artificial neural networks have been chosen because our Department has several years of experience applying these models to different areas of science such as hydrology, palynology, etc., and, based on accumulated experience, these models offer good results. Furthermore, Diez et al. [1] reported, according to the review of Qazi et al. [28], that this type of approach to determine solar radiation is usually accurate and shows errors less than 20% (depending on the input data and the different architectures of the neural networks developed). It must also be taken into account that this error percentage it's different if the goal is to do a solar irradiation model (that can be used to determine the *MGI* values in stations where this variable has not been registered, like our research) or to do a prediction model, for example, 3-week ahead. Nevertheless, ANNs have the disadvantage of their high computational cost and time required to obtain the model. The development of the other two models (SVM and RF) was in based on that they are generally between, according to the computational time, the multiple linear regression model and the artificial neural network model and show relatively good results according to the literature.

Related to this study, these kind of models can be used, together or separately, for different purposes.

- Multiple linear regression models can be used to predict the net radiation using meteorological data such as global solar radiation, temperature, relative humidity, etc. [29]. The researchers developed 8 different equations to estimate the daily net radiation and the results showed good adjustments and low errors on a daily scale, especially in the models that include the variables of relative humidity of the air, temperature, solar radiation and the inverse of the distance between the earth and the sun. Despite the simplicity of the multiple linear regression models, the authors showed good adjustments compared to the Rn FAO 56 OM model, which allows to conclude that the MLR models developed are an alternative to improve the evapotranspiration estimation.
- According to Diez et al. [1], artificial neural networks have been used to predict the solar irradiation at different time windows (hourly, daily and monthly) from different meteorological variables (temperature, atmospheric pressure, among others) or even including geographical coordinates such as latitude, longitude and altitude. In this sense, these authors developed ANNs to predict the global solar irradiation of the day after using data from one agrometeorological station located in Mansilla Mayor (León, Castilla y León). The authors concluded that artificial neural networks models provide better results compared to classical methods and require less input variables [1]. This kind of models can be used to determine the average monthly, the average weekly and the daily global solar radiation in Fortaleza (Brazilian Northeast region) using 14-year-long data set to train three different ANNs models [3]. ANNs can also be used to determine different parameters such as the global horizontal irradiation (from meteorological data), the global tilted irradiation (from the horizontal global irradiation and others) and to forecast the hourly direct normal and the global horizontal irradiation from one to six hours horizon [6].
- Support vector machines models can be used to generate the daily global solar irradiation using a general (non-locally dependent) model [9]. The model (which used temperatures, wind speed, relative humidity and rainfall, among other variables) presented a high capacity of generalization for the different studied locations and

improved, in terms of mean absolute error, the locally trained models in some locations [9]. SVM models can even be used to forecast photovoltaic power (and be compared with other models) [30].

- Random forest can be used to estimate the solar radiation using air pollution index in three different sites [31] or to forecast solar radiation and compared their result with other methods such as multivariate adaptive regression splines (MARS), classification and regression tree (CART) and M5 [32].

In many research articles, it can be also possible to see comparisons between this kind of models to predict solar irradiation and even other interesting variables related to the subject under study.

- MLR and ANN models can be compared in the estimation of monthly-average daily solar radiation over different locations in Turkey [33]. Different variables (latitude, longitude, altitude, land surface temperature and month) were used as input variables. According to the authors, the results showed that the ANN model could obtain good performance compared to the multiple linear regression model.
- SVM and ANN models were used in a comparative study of different methods carried out by da Silva et al. [34] to estimate the daily global solar irradiation. Four different kinds of architecture combining different input parameters were studied. According to the authors, statistical indicators showed that the SVM technique has better performance than ANN models for the study location (Botucatu/SP/Brazil). Neural models can be compared to random forest models (among other model) to forecast the normal beam, horizontal diffuse and global components [35]. SVM, ANN and deep neural network models can even be used to forecast photovoltaic power [30], to estimate electricity demand (using multiple linear regression, artificial neural network and support vector machine, among other) [12] or to estimate the surface downward longwave radiation (using ANN, SVR and RF, among others) [36]
- Random forest to model the daily variability of solar irradiance can be compared to other methods such as multiple linear regression, obtaining the best results between both [37].

### 3. Materials and Methods

#### 3.1. Study Area

According to Vázquez [38], Galicia can be divided into four climatic zones based on their solar radiation. To carried out this research, five meteorological stations were selected, all of them belonging to climatic zone II. This zone is characterized to present values between  $13.7 \text{ MJ/m}^2 \cdot \text{day}$  ( $3.8 \text{ kWh/m}^2 \cdot \text{day}$ ) and  $15.1 \text{ MJ/m}^2 \cdot \text{day}$  ( $4.2 \text{ kWh/m}^2 \cdot \text{day}$ ) [38]. The selected meteorological stations were: (i) Amiudal in the municipality of Avión, (ii) Serra do Faro in Rodeiro, (iii) Monte Medo in the municipality of Baños de Molgas, (iv) Ourense-Estacións in the city of Ourense and (v) Pazo de Fontefiz in Coles (Figure 1). The meteorological stations were selected taking into account the conditions and the quantity of available data to create useful and accurate models for the prediction of *MGI*.



### 3.4. MLR Models

Multiple linear regression analysis is a conventional method that relates different independent variables with a dependent one [41]. This method provides a linear input-output model for a specific data set [42]. Unlike the simple regression analysis, MLR analysis is closer to real situations because the phenomena are complex and must be explained using different variables that intervene in its existence [43].

It can be expressed mathematically as follows (Equation (1)):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (1)$$

being  $y$  the desired variable,  $\beta_0$  the constant,  $\beta_1$ – $\beta_n$  the regression coefficients,  $x_1$ – $x_n$  the input variables and  $\varepsilon$  is the error.

### 3.5. ANN Models

Artificial neural networks are a type of artificial intelligence (AI) model that simulates the human brain processes information [44]. ANNs presents different interesting aspects such as their fault tolerance or their generalization capabilities, among others [45,46].

The most used artificial neural models are the multilayer feedforward neural networks where the artificial neurons (also called nodes) are distributed into three different layers named; input, hidden and output layer [47]. The optimum number of hidden layer neurons, and the structure, can be defined by trial and error procedure [48,49]. The input layer receives the data provided by the user (in our case, the different variables from the Meteogalicia meteorological stations). During the model training, this information flow within the neural network in only one direction, in this case, from the input neurons to the output neuron going the hidden layer (each node). This flow is made up of two phases, the first one, called propagation, which the processed information is carries to the output layer where it is compared with the expected values and the error is calculated, and the second one, the weight update phase, where using the previous information, the model try to reduce the error made.

The ANN model implemented in this research has been tested using different parameters combination such as (i) the number of cycles (from 1 to 524,288 in 19 steps with a logarithmic scale), (ii) learning rate (0.1, 0.2 and 0.3), (iii) momentum (0.1, 0.2 and 0.3) and decay (true or false).

### 3.6. SVM Models

Support vector machines were introduced in the 1990s by different authors [50], to resolve classification problems and had a great reception and use due to its capacity to deal with non-linear data [9]. This method can be also used for regression purposes [12,50]. These approximations are a type of linear classifiers, which induce linear or hyperplane separators using a kernel function [50].

Support vector machine minimize the error of the training data trying to maximize the separation between classes and, when it comes to regression purposes, the goal is to find a function to approximate the nonlinear relationship between the used variables, that is, between inputs and output [30]. The basic mathematical ideas underlying SVM for function estimation can be analyzed in Smola and Schölkopf [51] which is a good introduction to support vector regression models.

A large combination of parameters to develop an SVM model is possible. To facilitate the development of these models, the combination of  $\gamma$  (represents the influence of a single training case) and  $C$  (represents the penalty factor) must be studied [36]. Different ranges can be chosen, in this research the range values for  $\gamma$  and  $C$  were chosen taking into account the “A Practical Guide to Support Vector Classification” proposed by Hsu et al. for classification problems [52]. Therefore, the combination of parameters used for SVM models’ development is (i) SVM type ( $\varepsilon$ -SVR and  $\nu$ -SVR), (ii)  $\gamma$  (from  $2^{-15}$  to  $2^3$  in 18 steps, with a logarithmic scale) and (iii)  $C$  (from  $2^{-5}$  to  $2^{15}$  in 20 steps, with a logarithmic scale).

### 3.7. RF Models

Random forests are non-parametric method proposed in 2001 by Breiman [31,53]. A random forest model is a set of random trees that can be used for regression and classification [54].

The random forest method offers better regression accuracy than the other methods such as MARS or M5 [32]. According to Srivastava et al., the RF model develop a large number of decorrelated decision trees which each generate an individual output, and then the final output value is obtained by averaging the individual output values.

In this research, the RF models were implemented using combinations of (i) number of trees (from 1 to 100 in 99 steps with linear scale), (ii) criterion (least square), (iii) maximal depth (from -1 to 100 in 101 steps with linear scale) and (iv) apply prepruning (true or false).

### 3.8. Statistics of the Developed Models

The statistics used to analyze the models were the squared correlation coefficient ( $r^2$ ), the root mean square error (RMSE, Equation (2)) and the average absolute relative error (Error, Equation (3)). The best model was chosen according to the lowest RMSE in the validation phase:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - x_i)^2}{N}} \tag{2}$$

$$Error = \frac{\sum_{i=1}^N \left( \left| \frac{y_i - x_i}{x_i} \right| \right) \cdot 100}{N} \tag{3}$$

### 3.9. Equipment and Software Used

The different models were implemented in the server available at the Department of Physical Chemistry of the University of Vigo, Campus of Ourense (Intel® Core™ i7-8700 processor at 3.20 GHz, with 16 GB of RAM). All models were run on Windows 10 Pro 64-bit operating system. Data were collected and processed using the software Microsoft Excel 2016, from Microsoft Office Professional Plus 2016 package, (Microsoft, Albuquerque, NM, USA). MLR, ANN, SVM and RF models were developed using a trial/free version of RapidMiner Studio 9.0.993 software (RapidMiner, Inc., Boston, MA, USA). Figures were made with SigmaPlot v. 13.0 (Systat Software, Inc., San Jose, CA, USA).

## 4. Results and Discussion

Table 2 show the bests models for each model type and the combination of variables used for that model. Next, the best models obtained for each of the studied approaches will be described.

**Table 2.** Adjustment parameters for each best approximation model developed according to its selected input variables. Latitude (*Lat*), longitude (*Long*), altitude (*Alt*), month, average ( $T_{av}$ ), average of the maximum ( $T_{av-max}$ ) and the average of the minimum temperature ( $T_{av-min}$ ), average ( $RH_{av}$ ), the average of the maximum ( $RH_{av-max}$ ) and the average of the minimum ( $RH_{av-min}$ ) relative humidity and precipitation (*P*). RMSE is the root mean square error (10 kJ/(m<sup>2</sup>·day)) and  $r^2$  is the squared correlation coefficient.

Combination Type	Model	Lat	Long	Alt	Month	$T_{av}$	$T_{av-max}$	$T_{av-min}$	$RH_{av}$	$RH_{av-max}$	$RH_{av-min}$	<i>P</i>	T		V		Q		
													RMSE	$r^2$	RMSE	$r^2$	RMSE	$r^2$	
Type 1	MLR													226.3	0.892	241.1	0.904	245.8	0.885
Type 4	ANN													127.1	0.967	122.6	0.975	113.6	0.980
Type 5	SVM													105.6	0.977	153.1	0.961	156.7	0.967
Type 5	RF													94.8	0.982	159.5	0.962	227.9	0.933

### 4.1. MLR Models

For the seven MLR models with different combination types, the one that presented the worst adjustment, based on the RMSE for the validation phase, was the model with

combination 7. This model presented an RMSE of 5674 kJ/(m<sup>2</sup>·day) for the validation phase, which corresponds to a low  $r^2$  (0.468). These bad adjustments for the validation phase are extensible to all phases of the model, training and querying. Thus, for these phases, the RMSE values are 5215 kJ/(m<sup>2</sup>·day) and 6300 kJ/(m<sup>2</sup>·day) which together with the low squared correlation, 0.426 and 0.343, make this model a model that cannot be used for modelling the *MGI*. The rest of the models present better adjustments than the previous model, with RMSE values for the validation phase, between 2924 kJ/(m<sup>2</sup>·day) and 2411 kJ/(m<sup>2</sup>·day). These models offer for the querying phase some RMSE similar to those provided for the training and/or the validation phase and an average absolute relative error between 18.1% and 19.6%. The best MLR model corresponds to a model with combination type 1 (Table 2), that is, an MLR that uses all the input available variables to model the behaviour of the *MGI*.

#### 4.2. ANN Models

The worst ANN model developed was the model with combination type 7. This model presented an RMSE value for the validation phase around 1526 kJ/(m<sup>2</sup>·day) which corresponds to an average absolute relative error of 12.1%. This value is close to the 10% that it is considered as, to our understanding, a good error percentage for this kind of modelling. Nevertheless, some authors suggest that prediction error less than 20% could be good accuracy in terms of solar radiation prediction [28]. The training and querying phase present similar adjustments to the validation phase with squared correlation coefficients of 0.943 and 0.953 for training and querying, respectively. These adjustments make the worst ANN an almost usable method to model the *MGI*, however, the other developed combination types improve the worst ANN model, presenting RMSE values between 1225 kJ/(m<sup>2</sup>·day) and 1494 kJ/(m<sup>2</sup>·day) for the validation phase. The best ANN model (Table 2) corresponds to a model with combination type 4 (input variables; latitude, longitude, altitude, month and the three humidities).

#### 4.3. SVM Models

For the different SVM models developed, the model that presented the worst adjustment, based on the RMSE for the validation phase, was, again, the model with combination 7. It seems clear that in all the models seen, those models that only have the precipitation variable, in addition to the other four fixed variables, do not present good results. The combination type 7 SVM model presents for validation phase an RMSE of 1704 kJ/(m<sup>2</sup>·day) which corresponds to a good  $r^2$  of 0.956. These adjustments for the validation phase are extensible to the training and querying phase where the RMSE are 1525 kJ/(m<sup>2</sup>·day) and 1743 kJ/(m<sup>2</sup>·day) with high squared correlation values, 0.951 and 0.962 which make this SVM a model that could be used for modelling the *MGI*. The rest of the models present better adjustments being the RMSE value in the validation phase dropped to 1556 kJ/(m<sup>2</sup>·day) for the second-best model. The best SVM model corresponds to a model with combination type 5, that is, an SVM that uses eight input variables to model the *MGI* response (Table 2).

#### 4.4. RF Models

Finally, the last kind of model is the RF. In this case, the worst model developed was, unlike to the other ML models, a model with combination type 2. This model presented an RMSE value for the validation phase around 2124 kJ/(m<sup>2</sup>·day) with an average absolute relative error of 15.0%. During the training and the querying phase, the model presents very different adjustments, 925 kJ/(m<sup>2</sup>·day) and 1651 kJ/(m<sup>2</sup>·day), respectively. The other combination types slightly improve this model and a better model is obtained when configuration 5 is used (Table 2).

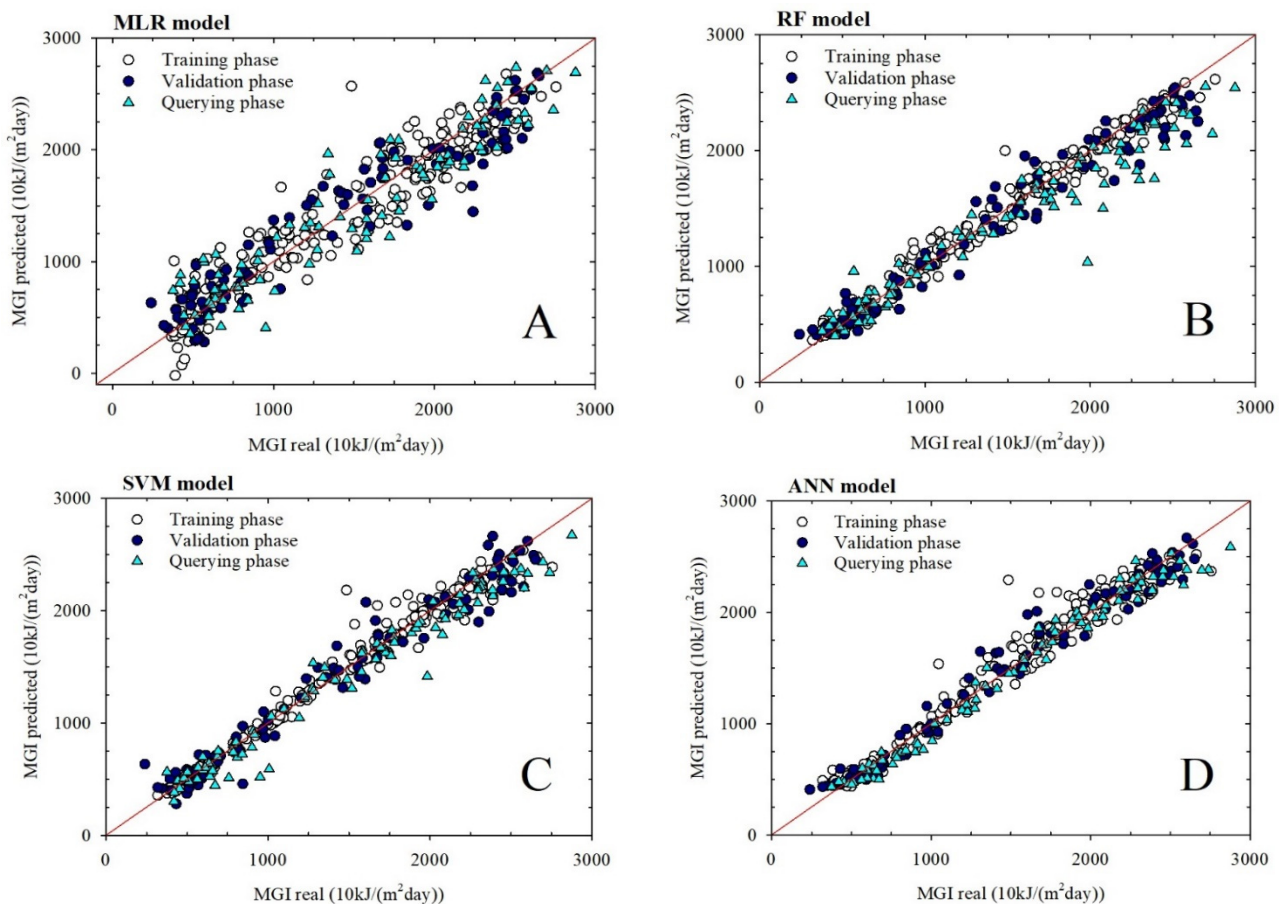


#### 4.5. Best Models Developed

Taking into account the previously chosen models (Table 2), we will now proceed to the analysis as a whole. It can be seen that the RMSE values obtained for the validation phase are included between  $1226 \text{ kJ}/(\text{m}^2 \cdot \text{day})$  and  $2411 \text{ kJ}/(\text{m}^2 \cdot \text{day})$ .

According to this, the multiple linear regression model is the model that obtains the worst RMSE value in the validation phase with a value of  $2411 \text{ kJ}/(\text{m}^2 \cdot \text{day})$  and the worst squared correlation coefficient (0.904). This model obtained an average absolute relative error around 19.2%. Regarding the training phase, the model presents lower RMSE value of  $2263 \text{ kJ}/(\text{m}^2 \cdot \text{day})$  compared with the validation phase, nevertheless, the  $r^2$  also present lower value (0.892).

Figure 2A shows the experimental and modelled MGI values by the MLR model. It can be seen how both, the training and validation phase cases, follow the line with slope one (red line), however, a great dispersion is observed in them, this fact can be intuited by the high values of absolute average relative error for both phases (17.3% and 19.2% for training and validation, respectively). These high errors are increased by the existence of some points that are distant from the line with slope one.



**Figure 2.** Graphical representation for the real and modelled values of MGI during the training phase (white dots), validation phase (black dots) and querying phase (turquoise triangles) for each select model: (A) multiple linear regression -MLR-, (B) random forest -RF-, (C) support vector machine -SVM- and (D) artificial neural network -ANN-. Redline is the line with slope one.

Given the results shown for both phases, it is expected that the results for the querying phase will also be the worst compared to the rest of the models. The RMSE is greater than in the validation phase ( $2458 \text{ kJ}/(\text{m}^2 \cdot \text{day})$ ) and the adjustments, in terms of squared correlation, was the lowest for the three phases (0.885).

In Figure 2A it can be seen that the querying cases also follow the line with slope one, however, as happened with the training and validation cases, these do not adjust the line, observing the existence of some point that is far away.

As expected, the MLR model is not capable of learning correctly and then generalizing that knowledge afterwards. A possible explanation for the poor adjustments of the MLR model may be based on the use of the month variable, which does not present a linear relationship with the *MGI*.

Given the results shown in the three phases, it can be concluded that the MLR model is not a suitable model for *MGI* modelling. It has concluded that this model is not usable for the prediction of monthly global irradiation because it presents a high percentage error for all phases (between 17.3% and 19.3%), although, since its error is less than 20% in the model, it could be considered good (taking into account bibliography reported above).

The next model in terms of low RMSE value in the validation phase is the RF model that presents a value of 1595 kJ/(m<sup>2</sup>·day). This value is improved in the model's training phase (948 kJ/(m<sup>2</sup>·day)). In both phases, the RF model improves the MLR model, both in RMSE values and in its squared correlation values (0.982 and 0.962 vs. 0.892 and 0.904, for the training and validation phase, respectively). Besides this, the model presents good behaviour in terms of average absolute relative error.

Figure 2B shows the experimental and modelled *MGI* values by the RF model. It can be seen how the training phase; the cases follow better the line with slope one than the cases predicted by the MLR. This behaviour is similar to the validation cases. The behaviour of both phases is good and reaches the average absolute relative error values of 5.9% and 10.5% for training and validation, respectively.

If we analyze the adjustments for the querying phase it can be seen how the RF model presents, for this phase, the worst adjustments in terms of RMSE (2279 kJ/(m<sup>2</sup>·day)) although the average absolute relative error remains at similar levels to those of the validation phase (10.7%).

In Figure 2B it can be seen that the querying cases also follow the line with slope one, however, a similar dispersion than provided by the MLR model is observed. It can be seen some cases that deviate more from the trend line one, although in the area of low *MGI* values it can be seen that the RF model adjusts much better than the MLR model. Due to this the average absolute relative error is good (around 10.7% for querying phase).

Given the results, it can be concluded that the RF model is a suitable approach for *MGI* modelling due to the fact that its errors for the validation and querying phases remain close to 10%.

The second-best model, taking into account the RMSE value in the validation phase, is the model developed based on support vector machine. The adjustments for the validation phase are kept close to the RF model, in fact, the RMSE value for the SVM model is 1531 kJ/(m<sup>2</sup>·day) compared to 1595 kJ/(m<sup>2</sup>·day) for the RF model and the squared correlation values are the same (0.961 vs. 0.962 for SVM and RF, respectively). The same happens with the error, which remains for both around 11%. For the training phase, a slight worsening of the fit for the SVM model reaching an RMSE of 1056 kJ/(m<sup>2</sup>·day) is observed.

Figure 2C shows the experimental and modelled *MGI* values by the SVM model. Training and validation phase cases follow the line with slope one (red line) nevertheless the model provided worse adjustment for the training cases in comparison with the RF model. This may be due to the existence of some points in the middle area that move away from the line with slope one. Both phases showed good adjustments in terms of average absolute relative error reaching values of 4.9% and 11.0% for training and validation, respectively.

Given the results provided by the SVM model, it can be assuming that for the querying phase the model will work well. According to the adjustment parameters for the querying phase, it can be said that both, the RMSE and the  $r^2$  values remain close to the RMSE of the validation phase.

This fact can be seen in Figure 2C where querying cases are close to the line with slope one and provided better fits than the RF model, although it can be seen some points in the lower and upper area that stray from the line with slope one.

Taking into account that the model offers 8.7% of absolute average relative error for the querying phase, it can be affirmed that the SVM model is a suitable model for MGI modelling.

Finally, for all the models developed, the best model is the ANN model taking into account the criterion of the lowest RMSE value in the validation phase. This model obtained for validation phase a RMSE value of 1226 kJ/(m<sup>2</sup>·day) that corresponds with the highest squared correlation coefficient (0.975) for all validation phases. Regarding the training phase, the RMSE value is around 1271 kJ/(m<sup>2</sup>·day) which supposes an absolute average relative error of 7.3%.

Figure 2D shows the experimental and modelled MGI values by the ANN model. It can be seen how for the training phase some points distance from the line with slope one. This fact explain that the ANN model did not obtain the best adjustments for the training phase, in comparison with the SVM and the RF model. This behaviour is reversed for the validation phase where it can be seen how this model is the one with the best fits to line with slope one. This behaviour is reversed for the validation phase where it can be seen how this model is the one with the best fits to line with slope one (obtaining an average absolute relative error of 8.8%).

Given the good results provided by the ANN model for both phases, good results for querying are expected. In this case, the RMSE value is lower than in both training and validation phases (1136 kJ/(m<sup>2</sup>·day)) and corresponds with a squared correlation of 0.980 (the highest for all the models in this phase).

In Figure 2D it can be seen the querying behaviour for this model where the querying cases are very close to the line with slope one. Some small dispersion is observed in the area with high MGI values, but this behaviour is an exception in the model.

Finally, taking into account all the adjustments provided by the model and the low absolute average relative error for the querying phase (6.6%) it can be affirmed that the ANN model is a suitable model for the MGI modelling.

Regarding the variables used by each of the best models, it can be seen in Table 2 that all the selected models have as input variables (apart from the latitude, longitude, altitude and month) all the variables of humidity and precipitation. This fact is only broken by the ANN model that does not use the precipitation. Regarding the MLR model, it can be seen how it includes temperature variables among its input variables. The inclusion of these variables may be because the MLR model, being a linear model, does not work properly with non-linear variables, as is the case of the month variable. Due to this fact, the authors understand that this variable can be counteracted by the MLR model with the inclusion of the temperature variables.

#### 4.6. ANN Generalization to Different Locations

After analyzing all the machine learning models developed in the previous section it will proceed to check how the best models work in the two reserved stations (Pazo de Fontefiz in Coles and Ourense-Estaci3ns in Ourense) which have not been used in any of the previous phases. The adjustments for the best models applied to these stations are presented in Table 3.

It can be seen how the support vector machines model is the one that offers worse modelling values for both stations; in fact, it presents errors in terms of RMSE much higher than the other selected models (Table 3). It can be seen how for the Pazo de Fontefiz station the error, in terms of root mean square error, is practically double (4029 kJ/(m<sup>2</sup>·day)) that the error presented by the best-selected model (ANN); while for the Ourense-Estaci3ns station, the error (8079 kJ/(m<sup>2</sup>·day)) is almost four times greater than that presented by the best model (ANN). As expected, these high errors affect the average relative absolute error presented by each station, so the Pazo de Fontefiz station presents an error of 24.8%, being

overcome by the error obtained in Ourense-Estaci3ns, 47.2%. The SVM model presents good adjustments in terms of squared correlation (upper than 0.940); however, taking into account the adjustments of the root mean square error and the average absolute relative error it can be concluded that the SVM model is not a suitable model for the MGI modelling.

**Table 3.** Adjustment parameters for each of the best models applied to the stations of Pazo de Fontefiz and Ourense-Estaci3ns. RMSE is the root mean square error (10 kJ/(m<sup>2</sup>·day)), Error is the average absolute relative error (%) and r<sup>2</sup> is the squared correlation coefficient.

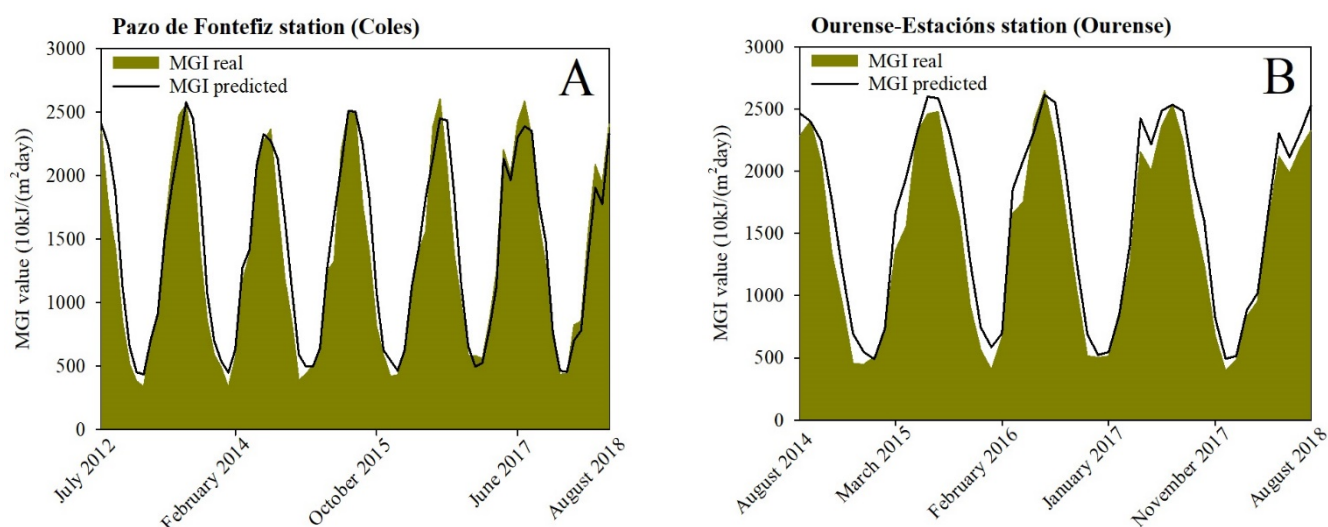
Model	Q <sub>PF</sub>			Q <sub>Ou</sub>		
	RMSE	Error	r <sup>2</sup>	RMSE	Error	r <sup>2</sup>
MLR	285.2	19.5	0.865	233.4	18.1	0.915
ANN	201.3	13.1	0.935	209.4	14.7	0.971
SVM	402.9	24.8	0.949	807.9	47.2	0.971
RF	246.1	21.2	0.920	216.5	19.6	0.950

The remaining three models have better fits than the SVM model. The one with the worst fit is the MLR model that presents an error, in terms of root mean square error, of 2852 kJ/(m<sup>2</sup>·day) and 2334 kJ/(m<sup>2</sup>·day) for the stations of Pazo de Fontefiz and Ourense-Estaci3ns, respectively (Table 3). Compared to the SVM model, this model improves its adjustments in terms of RMSE and error, although not in terms of squared correlation. The errors of this model are around 19% for each station. According to this error level, we can say that the model shows good behaviour, but shows a higher error percentage than desired, especially for the Pazo de Fontefiz station.

The second-best model is the random forest model. This model improves the previous models in terms of RMSE, for each of the analyzed stations, but the percentage errors remain high (21.2% and 19.6% for Pazo de Fontefiz and Ourense-Estaci3ns, respectively) (Table 3). Despite that the model presents high squared correlations (greater than 0.91), the use of the model should be limited.

Finally, the ANN model, which had been chosen in the previous section as the model with the best adjustments for each development phases, has emerged as the model with the best predictions for these two independent stations (Table 3). The model improves each statistic (except r<sup>2</sup> for the Pazo de Fontefiz station), reporting errors in terms of RMSE, around 2013 kJ/(m<sup>2</sup>·day) and 2094 kJ/(m<sup>2</sup>·day) for the station of Pazo de Fontefiz and Ourense-Estaci3ns, respectively. Likewise, for this model, the squared correlation is high (0.935 and 0.971) and the average absolute relative error remains lower than 15% error for each stations (which is considered as a good error percentage). These good adjustments are reflected in Figure 3. The first thing to note is the different size in the database between the two stations. The Pazo de Fontefiz station has data from July 2012 to August 2018 (a total of 73 months), while the Ourense-Estaci3ns station has data from June 2014 to August 2018 (a total of 49 monthly measurements). Figure 3 shows the time series for the real MGI values (olive colour) and the values predicted by the ANN model.

Figure 3A shows the time series for the Pazo de Fontefiz station. It can be seen the MGI's cycles with their maximums in the summer months and their minimums in the winter months (range from 3340 kJ/(m<sup>2</sup>·day) to 26,050 kJ/(m<sup>2</sup>·day)). The ANN predictions are shown in the Figure 3A as a black line. It can be seen how the modellings fit, almost perfectly, to the real-time series, which means (as we have already seen in the adjustments) that the ANN model can accurately predict the behaviour of the MGI for Pazo de Fontefiz station. It can be seen how for the low-value areas of MGI the prediction overestimates the values while the model behaves, in general, well for high-value areas of MGI (although also some underestimation is observed). Given the good adjustments, and taking into account the Figure 3A modelling time series, it can be said that this model is capable of generalizing the knowledge of the previous phase to other nearby geographical stations (Pazo de Fontefiz).



**Figure 3.** Real and modelled time series for (A) Pazo de Fontefiz and (B) Ourense-Estaci3ns stations. The olive shade corresponds to the actual values, and the black line corresponds to the values modelled by the ANN model.

Figure 3B shows the time series for the Ourense-Estaci3ns station. In this case, the time series has a real range from 3930 kJ/(m<sup>2</sup>·day) to 26,470 kJ/(m<sup>2</sup>·day). It can be seen how the modelling fits the real-time series, however in this case the adjustments show a worse behaviour than in the case of the Pazo de Fontefiz station. Again, it can be seen how for the low areas of *MGI* the prediction is overestimated, in general, the *MGI* values (can even see how this behaviour is observed in some measurement in the maximum area) although for high *MGI* values the ANN model generally predict well. Taking into account the adjustments and the Figure 3B, it can be said that the ANN model is usable on other nearby geographical stations.

Given the adjustments provided by the Artificial Neural Networks model both in the modeling phase and during the prediction phase (Pazo de Fontefiz and Ourense-Estaci3ns), it can be concluded that the ANN model is a useful model which can be used to model and predict the monthly global irradiation in areas bordering the studied stations. This statement is supported not only by the low errors committed in terms of the root mean square error, but also by the percentage of error associated with these predictions, which are maintained for the case of prediction around 15% which we could consider acceptable.

## 5. Conclusions

Based on the goodness of statistics, the modelling carried out by MLR, ANN, SVM and RF methodologies can model and predict the *MGI*, generally, in an appropriate way for the stations used for its development (Amiudal in Avi3n, Serra do Faro in Rodeiro and Monte Medo in Ba3os de Molgas). The results vary where these models are applied to other locations, Pazo de Fontefiz in Coles and Ourense-Estaci3ns in Ourense. Attending to the adjustments obtained for each station it can be affirmed that the best model is the ANN which presents the lowest RMSE value in the validation and querying phases 1226 kJ/(m<sup>2</sup>·day) and 1136 kJ/(m<sup>2</sup>·day), respectively, and predict conveniently for Coles and Ourense station 2013 kJ/(m<sup>2</sup>·day) and 2094 kJ/(m<sup>2</sup>·day), respectively. These good RMSE values are reinforced by the low percentage error obtained during the prediction phase at the two stations reserved for this purpose.

For all this, it can be concluded that, given the good results obtained, it is convenient to continue with the design of artificial neural networks applied to the analysis of monthly global irradiation in different areas of the Autonomous Community of Galicia to obtain a general model for the entire region. Due to this, this work may be the beginning of a more ambitious global study to model the monthly solar irradiation, and, subsequently, predict the monthly solar irradiation in advance in the Autonomous Community of Galicia (Spain).

All the models developed in this research could be improved with the inclusion of more stations, using different random split datasets, taking into account new meteorological input variables, among others.

**Author Contributions:** Conceptualization, C.M.-C. and G.A.; methodology, C.M.-C.; formal analysis, C.M.-C. and G.A.; writing—original draft preparation, C.M.-C. and G.A.; writing—review and editing, C.M.-C., G.A. and J.C.M.; supervision, G.A. and J.C.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Restrictions apply to the availability of these data. Data was obtained from Meteogalicia (Consellería de Medio Ambiente, Territorio e Vivenda da Xunta de Galicia) and are available <https://www.meteogalicia.gal/> (accessed on 19 April 2021) with the permission of Meteogalicia (Consellería de Medio Ambiente, Territorio e Vivenda da Xunta de Galicia).

**Acknowledgments:** G.A. thanks to the University of Vigo for his contract supported by “Programa de retención de talento investigador da Universidade de Vigo para o 2018” budget application 0000 131H TAL 641. Authors thank Meteogalicia and the Consellería de Medio Ambiente, Territorio e Vivenda of Xunta de Galicia for the database used in this research. G.A. thanks Xunta de Galicia, Consellería de Cultura, Educación e Ordenación Universitaria, for the computer equipment financed in 2017 from his postdoctoral grant B, POS-B/2016/001, K645 P.P.0000 421S 140.08. Authors thank RapidMiner Inc. for the Trial/Free license of RapidMiner Studio 9.0.993 software. This work is a summary of the final degree project developed by the first author of this research.

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

## References

1. Diez, F.J.; Navas-Gracia, L.M.; Chico-Santamarta, L.; Correa-Guimaraes, A.; Martínez-Rodríguez, A. Prediction of Horizontal Daily Global Solar Irradiation Using Artificial Neural Networks (ANNs) in the Castile and León Region, Spain. *Agronomy* **2020**, *10*, 96. [CrossRef]
2. Yacef, R.; Benghanem, M.; Mellit, A. Prediction of Daily Global Solar Irradiation Data Using Bayesian Neural Network: A Comparative Study. *Renew. Energy* **2012**, *48*, 146–154. [CrossRef]
3. Rocha, P.A.C.; Fernandes, J.L.; Modolo, A.B.; Lima, R.J.P.; da Silva, M.E.V.; Bezerra, C.A.D. Estimation of Daily, Weekly and Monthly Global Solar Radiation Using ANNs and a Long Data Set: A Case Study of Fortaleza, in Brazilian Northeast Region. *Int. J. Energy Environ. Eng.* **2019**, *10*, 319–334. [CrossRef]
4. Hunt, L.A.; Kuchar, L.; Swanton, C.J. Estimation of Solar Radiation for Use in Crop Modelling. *Agric. For. Meteorol.* **1998**, *91*, 293–300. [CrossRef]
5. Prieto, J.I.; Martínez-García, J.C.; García, D. Correlation between Global Solar Irradiation and Air Temperature in Asturias, Spain. *Sol. Energy* **2009**, *83*, 1076–1085. [CrossRef]
6. Notton, G.; Voyant, C.; Fouilloy, A.; Duchaud, J.L.; Nivet, M.L. Some Applications of ANN to Solar Radiation Estimation and Forecasting for Energy Applications. *Appl. Sci.* **2019**, *9*, 209. [CrossRef]
7. World Radiation Data Center (WRDC). WRDC Online Archive, National Renewable Energy Laboratory, US Department of Energy. 2012. Available online: <https://www.re3data.org> (accessed on 3 May 2017).
8. Vázquez Vázquez, M. *Atlas de Radiación Solar de Galicia*; Vázquez Vázquez, M., Ed.; Universidade de Vigo: Vigo, Spain, 2005; ISBN 84-609-7101-5.
9. Antonanzas-Torres, F.; Urraca, R.; Antonanzas, J.; Fernandez-Ceniceros, J.; Martinez-de-Pison, F.J. Generation of Daily Global Solar Irradiation with Support Vector Machines for Regression. *Energy Convers. Manag.* **2015**, *96*, 277–286. [CrossRef]
10. Espejo Marín, C. La Energía Solar Fotovoltaica en España. *Nimbus Rev. Climatol. Meteorol. Paisaje* **2004**, *13–14*, 5–31.
11. Fouilloy, A.; Voyant, C.; Notton, G.; Motte, F.; Paoli, C.; Nivet, M.-L.; Guillot, E.; Duchaud, J.-L. Solar Irradiation Prediction with Machine Learning: Forecasting Models Selection Method Depending on Weather Variability. *Energy* **2018**, *165*, 620–629. [CrossRef]
12. Solyali, D. A Comparative Analysis of Machine Learning Approaches for Short-/Long-term Electricity Load Forecasting in Cyprus. *Sustainability* **2020**, *12*, 3612. [CrossRef]
13. 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]

14. Morantes-Quintana, G.R.; Rincón-Polo, G.; Pérez-Santodomingo, N.A. Multiple Linear Regression Model to Estimate PM1 Concentration | Modelo de Regresión Lineal Múltiple para Estimar Concentración de PM1. *Rev. Int. Contam. Ambient.* **2019**, *35*, 179–194. [CrossRef]
15. Cho, J.H.; Lee, J.H. Multiple Linear Regression Models for Predicting Nonpoint-source Pollutant Discharge from a Highland Agricultural Region. *Water* **2018**, *10*, 1156. [CrossRef]
16. O’Fallon, B.D.; Woolderchak-Donahue, W.; Crockett, D.K. A Support Vector Machine for Identification of Single-nucleotide Polymorphisms from Next-generation Sequencing Data. *Bioinformatics* **2013**, *29*, 1361–1366. [CrossRef] [PubMed]
17. Zimmermann, O.; Hansmann, U.H.E. Support Vector Machines for Prediction of Dihedral Angle Regions. *Bioinformatics* **2006**, *22*, 3009–3015. [CrossRef]
18. Lyu, Y.; Gong, X. A Two-Layer SVM Ensemble-Classifer to Predict Interface Residue Pairs of Protein Trimers. *Molecules* **2020**, *25*, 4353. [CrossRef] [PubMed]
19. Abdullah, S.; Pradhan, R.C.; Pradhan, D.; Mishra, S. Modeling and Optimization of Pectinase-assisted Low-temperature Extraction of Cashew Apple Juice Using Artificial Neural Network Coupled with Genetic Algorithm. *Food Chem.* **2021**, *339*, 127862. [CrossRef]
20. Sabater, C.; Blanco-Doval, A.; Montilla, A.; Corzo, N. Optimisation of an Enzymatic Method to Obtain Modified Artichoke Pectin and Pectic Oligosaccharides Using Artificial Neural Network Tools. In silico and in vitro Assessment of the Antioxidant Activity. *Food Hydrocoll.* **2021**, *110*, 106161. [CrossRef]
21. Makino, Y.; Kousaka, Y. Prediction of Degreening Velocity of Broccoli Buds Using Hyperspectral Camera Combined with Artificial Neural Networks. *Foods* **2020**, *9*, 558. [CrossRef]
22. Behr, A.; Giese, M.; Teguim, H.; Theune, K. Early Prediction of University Dropouts-A Random Forest Approach. *Jahrb. Natl. Okon. Stat.* **2020**, *240*, 743–789. [CrossRef]
23. Sadorsky, P. A Random Forests Approach to Predicting Clean Energy Stock Prices. *J. Risk Financ. Manag.* **2021**, *14*, 48. [CrossRef]
24. Torkashvand, A.M.; Ahmadi, A.; Nikraves, N.L. Prediction of Kiwifruit Firmness Using Fruit Mineral Nutrient Concentration by Artificial Neural Network (ANN) and Multiple Linear Regressions (MLR). *J. Integr. Agric.* **2017**, *16*, 1634–1644. [CrossRef]
25. Niu, W.-J.; Feng, Z.-K.; Feng, B.-F.; Min, Y.-W.; Cheng, C.-T.; Zhou, J.-Z. Comparison of Multiple Linear Regression, Artificial Neural Network, Extreme Learning Machine, and Support Vector Machine in Deriving Operation Rule of Hydropower Reservoir. *Water* **2019**, *11*, 88. [CrossRef]
26. Lee, H.; Wang, J.; Leblon, B. Using Linear Regression, Random Forests, and Support Vector Machine with Unmanned Aerial Vehicle Multispectral Images to Predict Canopy Nitrogen Weight in Corn. *Remote Sens.* **2020**, *12*, 2071. [CrossRef]
27. Martínez Castillo, C.A. *Modelado de la Irradiación Global Mensual Usando Estaciones de la Red de Meteogalicia*; Universidad de Vigo: Ourense, Spain, 2019.
28. Qazi, A.; Fayaz, H.; Wadi, A.; Raj, R.G.; Rahim, N.A.; Khan, W.A. The Artificial Neural Network for Solar Radiation Prediction and Designing Solar Systems: A Systematic Literature Review. *J. Clean. Prod.* **2015**, *104*, 1–12. [CrossRef]
29. Ocampo, D.; Rivas, R. Estimating Daily Net Radiation from Multiple Linear Regression Models | Estimación de la Radiación Neta Diaria a Partir de Modelos de Regresión Lineal Múltiple. *Rev. Chapingo Ser. Ciencias For. Ambient.* **2013**, *19*, 263–271. [CrossRef]
30. Kim, M.; Song, H.; Kim, Y. Direct Short-term Forecast of Photovoltaic Power through a Comparative Study Between Coms and Himawari-8 Meteorological Satellite Images in a Deep Neural Network. *Remote Sens.* **2020**, *12*, 2357. [CrossRef]
31. Sun, H.; Gui, D.; Yan, B.; Liu, Y.; Liao, W.; Zhu, Y.; Lu, C.; Zhao, N. Assessing the Potential of Random Forest Method for Estimating Solar Radiation Using Air Pollution Index. *Energy Convers. Manag.* **2016**, *119*, 121–129. [CrossRef]
32. Srivastava, R.; Tiwari, A.N.; Giri, V.K. Solar Radiation Forecasting Using MARS, CART, M5, and Random Forest Model: A Case Study for India. *Heliyon* **2019**, *5*, e02692. [CrossRef]
33. Şahin, M.; Kaya, Y.; Uyar, M. Comparison of ANN and MLR Models for Estimating Solar Radiation in Turkey Using NOAA/AVHRR data. *Adv. Sp. Res.* **2013**, *51*, 891–904. [CrossRef]
34. Da Silva, M.B.P.; Francisco Escobedo, J.; Juliana Rossi, T.; dos Santos, C.M.; da Silva, S.H.M.G. Performance of the Angstrom-Prescott Model (A-P) and SVM and ANN techniques to estimate daily global solar irradiation in Botucatu/SP/Brazil. *J. Atmos. Solar-Terrestrial Phys.* **2017**, *160*, 11–23. [CrossRef]
35. 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]
36. Feng, C.; Zhang, X.; Wei, Y.; Zhang, W.; Hou, N.; Xu, J.; Jia, K.; Yao, Y.; Xie, X.; Jiang, B.; et al. Estimating Surface Downward Longwave Radiation Using Machine Learning Methods. *Atmosphere* **2020**, *11*, 1147. [CrossRef]
37. Huang, J.; Troccoli, A.; Coppin, P. An Analytical Comparison of Four Approaches to Modelling the Daily Variability of Solar Irradiance Using Meteorological Records. *Renew. Energy* **2014**, *72*, 195–202. [CrossRef]
38. Vázquez Vázquez, M. *Radiación Solar e Severidade Climática en Galicia*; Vázquez Vázquez, M., Ed.; Universidade de Vigo: Vigo, Spain, 2008; ISBN 978-84-612-4469-0.
39. Derivative Work from ME2000raster 2020 CC-BY 4.0 ign.es. Mapa de España 1:2.000.000 Ráster. Instituto Geográfico Nacional, Gobierno de España. Available online: <http://www.ign.es/web/ign/portal> (accessed on 11 April 2021).
40. Meteogalicia. Consellería de Medio Ambiente, Territorio e Vivenda. Xunta de Galicia. Observacións. Rede Meteorolóxica. Available online: <https://www.meteogalicia.gal/> (accessed on 17 September 2018).

41. Elbayoumi, M.; Ramli, N.A.; Fitri Md Yusof, N.F. Development and Comparison of Regression Models and Feedforward Backpropagation Neural Network Models to Predict Seasonal Indoor PM<sub>2.5</sub>–10 and PM<sub>2.5</sub> Concentrations in Naturally Ventilated Schools. *Atmos. Pollut. Res.* **2015**, *6*, 1013–1023. [[CrossRef](#)]
42. Al-Alawi, S.M.; Abdul-Wahab, S.A.; Bakheit, C.S. Combining Principal Component Regression and Artificial Neural Networks for More Accurate Predictions of Ground-level Ozone. *Environ. Model. Softw.* **2008**, *23*, 396–403. [[CrossRef](#)]
43. Rodríguez-Jaume, M.-J.; Mora Catalá, R. Análisis de Regresión Múltiple. In *Estadística Informática: Casos y Ejemplos con el SPSS*; Publicaciones de la Universidad de Alicante: Alicante, Spain, 2001; pp. 109–123, ISBN 84-7908-638-6.
44. Agatonovic-Kustrin, S.; Beresford, R. Basic Concepts of Artificial Neural Network (ANN) Modeling and Its Application in Pharmaceutical Research. *J. Pharm. Biomed. Anal.* **2000**, *22*, 717–727. [[CrossRef](#)]
45. Balas, C.E.; Koç, M.L.; Tür, R. Artificial Neural Networks Based on Principal Component Analysis, Fuzzy Systems and Fuzzy Neural Networks for Preliminary Design of Rubble Mound Breakwaters. *Appl. Ocean Res.* **2010**, *32*, 425–433. [[CrossRef](#)]
46. Basheer, I.A.; Hajmeer, M. Artificial Neural Networks: Fundamentals, Computing, Design, and Application. *J. Microbiol. Methods* **2000**, *43*, 3–31. [[CrossRef](#)]
47. Yolmeh, M.; Habibi Najafi, M.B.; Salehi, F. Genetic Algorithm-artificial Neural Network and Adaptive Neuro-fuzzy Inference System Modeling of Antibacterial Activity of Annatto Dye on Salmonella Enteritidis. *Microb. Pathog.* **2014**, *67*, 36–40. [[CrossRef](#)]
48. Lee, K.Y.; Chung, N.; Hwang, S. Application of an Artificial Neural Network (ANN) Model for Predicting Mosquito Abundances in Urban Areas. *Ecol. Inform.* **2016**, *36*, 172–180. [[CrossRef](#)]
49. Sutariya, V.; Groshev, A.; Sadana, P.; Bhatia, D.; Pathak, Y. Artificial Neural Network in Drug Delivery and Pharmaceutical Research. *Open Bioinform. J.* **2013**, *7*, 49–62. [[CrossRef](#)]
50. Carmona Suárez, E.J. *Tutorial sobre Máquinas de Vectores Soporte (SVM)*; Universidad Nacional de Educación a Distancia (UNED): Madrid, Spain, 2016; Available online: [http://www.ia.uned.es/~eucarmona/publicaciones/\[2013-Carmona\]%20SVM.pdf](http://www.ia.uned.es/~eucarmona/publicaciones/[2013-Carmona]%20SVM.pdf) (accessed on 17 April 2021).
51. Smola, A.J.; Schölkopf, B. A tutorial on Support Vector Regression. *Stat. Comput.* **2004**, *14*, 199–222. [[CrossRef](#)]
52. Hsu, C.-W.; Chang, C.-C.; Lin, C.-J. A Practical Guide to Support Vector Classification. Available online: <https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf> (accessed on 9 November 2020).
53. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
54. RapidMiner Documentation. Random Forest. Available online: [https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/parallel\\_random\\_forest.html](https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/trees/parallel_random_forest.html) (accessed on 3 November 2020).