Article

A Hybrid Technique for Day-Ahead PV Generation Forecasting Using Clear-Sky Models or Ensemble of Artificial Neural Networks According to a Decision Tree Approach

Stefano Massucco [ID], **Gabriele Mosaico** [ID]**, Matteo Saviozzi** [ID] **and Federico Silvestro *** [ID]

Department of Electrical, Electronic, Telecommunication Engineering and Naval Architecture (DITEN),
University of Genova, Via all'Opera Pia 11a, 16145 Genova, Italy; stefano.massucco@unige.it (S.M.);
gabriele.mosaico@edu.unige.it (G.M.); matteo.saviozzi@unige.it (M.S.)
* Correspondence: federico.silvestro@unige.it

check for
updates

Abstract: PhotoVoltaic (PV) plants can provide important economic and environmental benefits to electric systems. On the other hand, the variability of the solar source leads to technical challenges in grid management as PV penetration rates increase continuously. For this reason, PV power forecasting represents a crucial tool for uncertainty management to ensure system stability. In this paper, a novel hybrid methodology for the PV forecasting is presented. The proposed approach can exploit clear-sky models or an ensemble of artificial neural networks, according to day-ahead weather forecast. In particular, the selection among these techniques is performed through a decision tree approach, which is designed to choose the best method among those aforementioned. The presented methodology has been validated on a real PV plant with very promising results.

Keywords: PV forecasting; hybrid method; clear-sky model; artificial neural networks; basic ensemble method; decision trees; CART tree; weather type partition; weather classification

1. Introduction

In recent years, global energy demand has increased dramatically. Several factors have contributed to this rise: the growth of the world population, the industrialization of developing countries, and the worldwide process of urbanization [1]. The exploitation of the main conventional sources, fossil fuels, has proven to be detrimental for the environment and, therefore, alternative Renewable Energy Sources (RESs) have gained wide interest. Among all the RESs, PhotoVoltaic (PV) systems have gained a lot of attention for their availability, low maintenance and operational cost, lifetime, ease of application, and environmental benefits.

This has implied a growth of the global solar energy production from 3.7 GW (2007) to 402 GW (2017) [2]. In this context, high PV penetration provides many environmental and economic benefits, but the stochastic behavior of the solar power may also introduce technical issues (e.g., generation schedule, operating reserve, market regulation, etc.) without robust and precise forecast [3].

A reliable forecast is the key for several smart-grid applications [4], such as optimal dispatch, [5], active demand response, grid regulation [6], and intelligent energy management [7].

PV forecasting represents a large research topic which can be characterized by the time horizon related to the prediction [8]:

- Very short/short-term forecasting, wherein the time horizon varies from seconds to 24–48 h;
- Medium-term forecasting, which analyzes periods up to one month;

- Long-term forecasting, wherein the prediction horizon can be set to 1–10 years.

Among these, the 24 h-ahead horizon is crucial for the scheduling of the conventional generation, and many national grid codes (e.g., [9,10]) require punctual and precise power forecasting. In addition, in countries with a day-ahead electricity market, large RES plants can act as producers providing sale bids, wherein the actual production must follow a scheduler offer that is provided through a forecasting approach. For these reasons, this paper focuses on the day-ahead PV forecasting topic.

1.1. Short-Term PV Forecasting-State of the Art

PV generation forecasting is an important topic within the scientific community that has proposed a large variety of solutions. The different approaches to this problem allow for a general classification.

A first distinction is made between direct and indirect forecasting methods [8]. In an indirect forecasting approach, the solar irradiance is forecasted and then exploited in commercial PV simulation software to predict the PV power generation. Direct methodologies on the other hand aim to directly predict the PV output. A comparison between the two strategies can be found in [11], wherein the results show that direct methods perform better.

A more common classification of the forecasting strategies is the following:

- physical models;
- statistical/artificial intelligence approaches;
- hybrid methods.

The physical methodologies model the PV production considering mainly the two variables that mostly affect the power generation: irradiance reaching the panel and PV modules temperature.

There are many techniques to model these two variables, for each possible physical model. In general, the irradiance depends mainly on sun position, ambient temperature, relative humidity, albedo, possible shadings of near objects, and clouds.

On the other hand, the temperature of the module is typically affected by irradiance, ambient temperature and wind speed [12].

The main advantage of physical models is that they can be employed even without a set of historical data. In addition, they can be adopted to generate input variables for statistical models, hence allowing the definition of hybrid models. Two examples of physical methods for the day-ahead horizon are described in [13]. These two methodologies are based on two physical representations of a PV cell through an electrical circuit.

Statistical approaches are very popular in the field of short-term forecasting. They are based on the use of historical data related to weather and PV production defining any type of statistical method, be it classical, such as time series [14], or advanced, like machine learning [15] or, recently, deep learning techniques [16]. Statistical methodologies typically outperform physical models and they are easier to implement.

The most used statistical methods, according to [17], are Artificial Neural Network (ANN)-based. A reason behind this popularity is the ANN capability to capture the high nonlinearity of PV production. In this context, Reference [18] shows that also variables characterized by a poor linear correlation with PV output can help the ANN to improve the forecasting accuracy.

The similar day approach represents another type of statistical method. In [19] a technique for the identification of the past days with the highest probability of having similar meteorological characteristics to the day under forecasting is presented. Past days data are then used to perform the prediction.

Any kind of mixture of physical and statistical approach is called hybrid method. For this reason, it is possible to find a big number of hybrid methods in literature. Physical models can become hybrid if statistical techniques are used to correct systematic errors. On the other hand, statistical approaches that exploit physical methods for the design of input variables can be considered hybrid. Examples of these typologies of hybrid procedures are described in [20,21].

The literature also presents methods that are called hybrid because of a particular combination of two techniques [22–24], even if they can be considered data-driven approaches.

In [13] the proposed physical approaches are compared with a hybrid technique, specifically an ANN that uses another physical model as input. The results prove that the hybrid methodology outperforms all the physical models.

In general, hybrid methodologies are designed to improve the performance of physical or statistical techniques.

1.2. Contributions

This article describes the design and the implementation of an innovative hybrid forecasting technique for the power output of a PV system. The proposed procedure has been validated on a real PV plant.

The innovation of the proposed hybrid forecasting method consists in the combination modality of physical and statistical approaches. In particular, two physical models and one statistical method are developed, while the proposed hybrid technique chooses among them for the PV prediction according to the day-ahead weather forecasting.

Several works present a methodology based on the selection of different models, trained with data coming from different types of days.

For example, Reference [25] describes a method composed of four support vector regressors, one for each identified meaningful weather condition. In [26] five neural networks coupled with a harmony search algorithm are used according to a fuzzy *k*-means clustering technique. In that work, a fuzzy inference approach is used according to the weather prediction. In [27], days are divided into 2 groups: for each of them an ensemble of ANNs is developed. In the forecasting phase one of the two ensembles is used. In [28], data are grouped in 6 clusters using the variance of five differential sequences of weather Key Performance Indicators (KPIs). Each cluster is used to train, through a back-propagation algorithm, a neural network which is employed for the PV forecasting. Recently, in [29] *k*-means clustering and gated recurrent unit are employed respectively for classification and prediction tasks. Finally, the specific problem related to the classification of the day-ahead weather condition is addressed in [30] and faced through *k*-nearest neighbor and support vector machines.

In all these cases the weather prediction is used to select among different statistical methods. The innovative contribution of this work lies on the selection between a physical model (Clear-Sky Model (CSM)), a statistically corrected physical model (Corrected Clear-Sky Model (CCSM)) and a statistical approach (Basic Ensemble Method (BEM) of neural networks). The proposed approach represents a novel hybrid method for the PV forecasting because it is neither a corrected physical approach nor a statistical technique that uses inputs from a physical model. It is a methodology that, according to the day-ahead weather forecast, may use a physical or a statistical approach, differently from all the above-mentioned hybrid strategies, which select among techniques of the same type.

The weather forecast can be used for the selection of the most appropriate method in different modalities. A decision tree algorithm has been adopted in this work, because of its easy implementation and straightforward interpretation. Through the analysis of the resulting decision rule, it is possible to verify the rationality of the proposed approach.

The rest of the paper is organized as follows. In Section 2 the considered KPIs are defined; Section 3 shows the various components of the proposed hybrid forecasting methodology; the test site is described in Section 4; Section 5 is dedicated to the results on a real PV plant; finally, conclusions are drawn in Section 6.

2. Key Performance Indicators-KPI

This section collects all the KPIs used to assess the quality of the proposed forecasting procedure and for the selection of all the parameters of the methodologies described in Section 3.

The base for all the considered KPIs is the prediction error. It is defined as the difference between the forecast and the measured variable at time t:

$$\varepsilon(t) = x_{forec}(t) - x_{meas}(t) \tag{1}$$

From this equation, it is possible to observe that positive errors correspond to over-estimation of the actual value.

The second KPI proposed in this work is the Root Mean Square Error (RMSE) [12]. It is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\varepsilon(t))^2} \tag{2}$$

where N is the samples number. The RMSE can be normalized, obtaining an estimation of the percentage error. This KPI is called normalized Root Mean Square Error (nRMSE) and can be evaluated as:

$$nRMSE = \frac{RMSE}{\sqrt{\frac{1}{N}\sum_{t=1}^{N}(x_{meas}(t))^2}} \tag{3}$$

The Mean Bias Error (MBE) is defined as the mean difference between the prediction and the measurement:

$$MBE = \frac{1}{N}\sum_{t=1}^{N}\varepsilon(t) \tag{4}$$

It represents the systematic part (bias) of the error: if it is positive, the model has the tendency to overestimate the actual value; if negative it underestimates it.

Finally, the Skill Score (SS) [31] measures the accuracy of a forecasting technique with respect to the precision of a reference methodology. The SS can be defined for different KPIs. In this work it has been calculated as:

$$SS = 1 - \frac{|KPI_{proposed}|}{|KPI_{reference}|} \tag{5}$$

where $KPI_{proposed}$ represents an estimation of the accuracy of the proposed approach, while $KPI_{reference}$ is the same KPI evaluated on a reference method.

In this paper, SS is used to compare the presented technique with other methodologies.

The range of the SS is $[-\infty, +1]$. A positive value of SS implies that the proposed technique provides a better result with respect to the other approach, while a negative value corresponds to the opposite situation. Notice that a $SS = 1$ represents the perfect forecast.

3. Methodology

The method proposed in this work consists of two Decision Rules (DRs) and three Sub-Methodologies (SM). Figure 1 describes the flowchart of the proposed hybrid approach. In this figure diamonds represent the DRs, while rectangles are the available forecasting approaches. The implemented SMs are:

- CSM, based on well-known sun equations;
- CCSM, a linear model which combines CSM and cloud cover index;
- BEM, that uses outputs of multiple ANNs.

As can be seen from Figure 1, first, a decision based on day-ahead weather forecast is made on whether to use the BEM or a deterministic model. In case the choice ends up being a deterministic model, a second decision must be made between the CSM and the CCSM.

This hybrid approach should achieve an improved prediction, with respect to the single SMs, because physical models have higher accuracy on clear-sky days and when the weather conditions are stable, while ANNs are typically preferable in other cases (cloudy/rainy days) [32].

The forecasting has been designed with a 24 h-horizon and a granularity equal to 15 min (which represents the standard monitoring interval adopted by Italy). Notice that the output of the proposed method is composed of 96 values representing the PV output for the next 24 h.
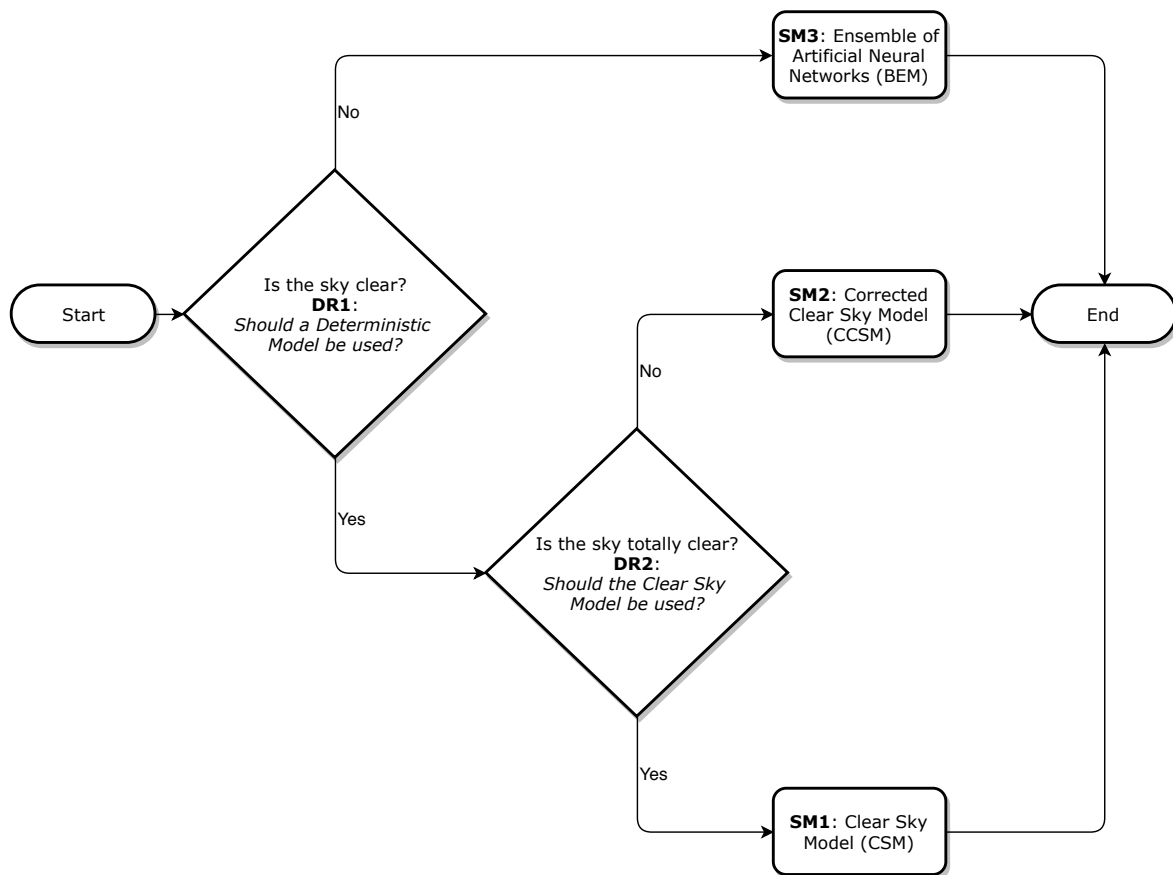

Figure 1. General scheme of the proposed technique.

The rest of this section describes the various part of the proposed strategy. In particular, in Sections 3.1 and 3.2 the deterministic approaches (CSM and CCSM) are described; in Section 3.3 the BEM is presented; finally, in Section 3.4 the proposed hybrid methodology is described.

3.1. Clear-Sky Model-CSM

When the sky is clear the PV system is not shaded by any cloud. In this case, there is little to no uncertainty in the PV output profile. Thus, a deterministic model can be set up for covering scenarios with this weather condition [8,32]. In this work the predicted PV output ($P_{system}(t)$) is modeled as follows [33] (notice that (t) indicates the time dependency):

$$P_{system}(t) = \frac{E_{g,pv}(t) \cdot P_{peak} \cdot \eta_{pan}(t) \cdot \eta_{inv} \cdot \omega_{DEG}(t)}{E_{STD}} \qquad (6)$$

where: $E_{g,pv}(t)$ is the global irradiance on the plane of the array $\left[\frac{W}{m^2}\right]$, P_{peak} represents the total rated peak power of the solar panel [kW], $\eta_{pan}(t)$ is the relative efficiency factor of the panels [p.u.], η_{inv} indicates the relative efficiency factor of the inverter [p.u.], $\omega_{DEG}(t)$ represents the coefficient of degradation [p.u.] and E_{STD} is the irradiance of standard test conditions $\left[\frac{W}{m^2}\right]$.

P_{peak}, η_{inv} are parameters related to technical data of the PV plant, while E_{STD} is a constant equal to $1000 \frac{W}{m^2}$. The parameter $\omega_{DEG}(t)$ can be either set to prescribed values taken from large scientific reviews (e.g., [34]) or estimated through field measurements [35].

Thus, the main variables of this model are $E_{g,pv}(t)$ and $\eta_{pan}(t)$. The global irradiance is the sum of three components, weakened by the shading factors. These take into account the possible shadows due to the surrounding buildings [36]:

$$E_{g,pv}(t) = E_{b,pv}(t) \cdot (1 - S_{dir}(\alpha(t), \gamma(t))) + E_{d,pv}(t) \cdot (1 - S_{diff}) + E_{r,pv}(t) \tag{7}$$

where: $E_{b,pv}(t)$ is the beam irradiance reaching the plane of the array $\left[\frac{W}{m^2}\right]$, $E_{d,pv}(t)$ represents the diffuse irradiance reaching the plane of the array $\left[\frac{W}{m^2}\right]$, $E_{r,pv}(t)$ is the reflected irradiance reaching the plane of the array $\left[\frac{W}{m^2}\right]$, $\alpha(t)$ denotes the sun azimuth [degree], $\gamma(t)$ represents the sun elevation [degree], $S_{dir}(\alpha(t), \gamma(t))$ is the direct component shading factor [p.u.] and S_{diff} denotes the diffuse component shading factor [p.u.]. In particular, $E_{r,pv}(t)$ is calculated as follows:

$$E_{r,pv}(t) = \begin{cases} E_{g,hor}(t) \cdot \rho_g \cdot \left(\frac{1-\cos(\beta)}{2}\right) & for \quad \frac{1-\cos(\beta)}{2} > 0 \\ 0 & o/w \end{cases} \tag{8}$$

where: $E_{g,hor}(t)$ represents the global horizontal irradiance $[\frac{W}{m^2}]$, ρ_g is the ground albedo [p.u.] and β denotes the tilt angle of the PV array [degree].

The shading factors calculation $(S_{dir}(\alpha(t), \gamma(t)), S_{diff})$ and the ground albedo (ρ_g) are specific to the site and therefore described in Section 4.1, while the irradiance components are calculated as in [37].

The efficiency of the panel is mainly affected by modules temperature. In this work the following equation has been used for its estimation [23,33]:

$$\eta_{pan}(t) = 1 + \beta_c(T_m(t) - 25°C) \tag{9}$$

where β_c is the module temperature coefficient.

For the module temperature estimation, several models have been considered [38–41]. The performance of these models have been compared on a test set composed of clear-sky days and the following relation has been selected [38]:

$$T_m(t) = T_a(t) + \frac{E_{g,pv}(t)}{E_{STD}} \cdot (0.0712 \cdot W_s(t)^2 - 2.411 \cdot W_s(t) + 32.96) \tag{10}$$

where $T_a(t)$ is the ambient temperature [°C] and $W_s(t)$ is the wind speed $[\frac{m}{s}]$.

3.2. Corrected Clear-Sky Model-CCSM

The model described in the previous paragraph supposes that the sky is completely clear. This means that its performance can be improvable if the presence of clouds is considered. Thus, a modified version of CSM has been developed. It is a simple Stepwise Linear Regression (SLR) model [42], trained on clear-sky or almost clear-sky days, with regressors composed of CSM output and Cloud Cover index (CC).

The CC index is a number that measures the percentage of the considered sky portion which is covered by the clouds at a given time. Throughout this work, it ranges from 0 to 100, where CC = 0 indicates a cloudless sky, while CC = 100 indicates a weather condition of completely covered sky.

To estimate nonlinear behaviors, the variables have been taken up to the fifth power.

The SLR is a linear regression where the regressors are selected through an automated procedure that iteratively adds and removes regressors by testing their statistical significance through a hypothesis test on their corresponding coefficients, measured by the *p*-value of an F-statistic [42].

The proposed procedure is the following:

1. Fit an initial model with only the constant term;
2. Add to the model the candidate regressor with the smallest *p*-value, provided that it is smaller than a predetermined entrance tolerance. Repeat this step until no regressor can be added in the model;
3. Subtract to the model the regressor with the highest *p*-value, provided that it is higher than a predetermined exit tolerance. If there is no regressor with such high *p*-value, end; otherwise return to step 2.

The final result depends on the initial model and the predetermined tolerances. For this reason, there is no guarantee that the final result is the best possible model. However, checking all the possible model candidates would typically take a large amount of time (for p candidate regressors, there are 2^p possible models) and therefore this procedure is used as a compromise between optimality and feasibility.

In the proposed approach the entrance and exit tolerance have been set respectively to 0.05 and 0.1.

3.3. Artificial Neural Network-Ensemble Approach

ANNs are a wide class of logical structures freely inspired by the human brain. They are vastly used in PV forecasting. This is testified by the fact that almost 25% of the papers proposed in the literature on this topic are ANN-based [17].

The architecture adopted in this article is the Multi-Layer Perceptron (MLP) [43]. Figure 2 provides the general structure of an MLP.
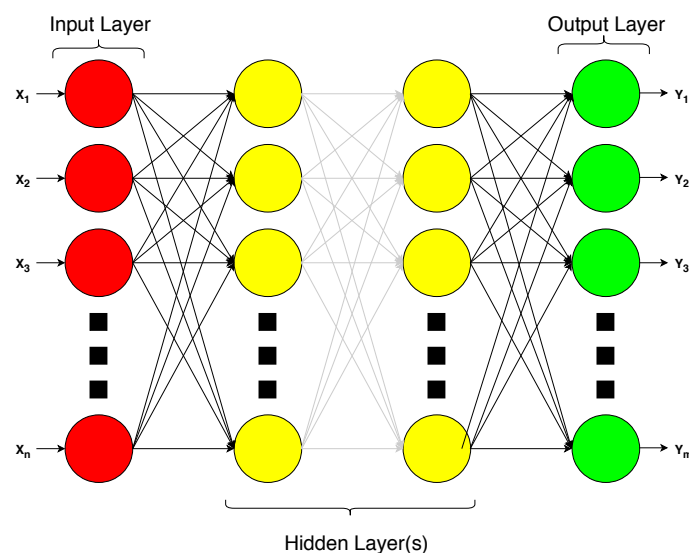

Figure 2. A multi-layer perceptron.

Its architecture consists of three parts: input layer, at least one hidden layer and output layer. Each layer receives the inputs from the preceding layer and, by means of weighting, translation, and a nonlinear transformation, passes them to the next layer. The input layer processes the original input vector, while the output layer passes the processed values to the user.

In this work an ensemble technique has been exploited within the ANN approach.

3.3.1. Basic Ensemble Method-BEM

Ensemble averaging methods are usually implemented to achieve more accurate results than a single ANN. The basic principle is to combine outputs of several ANNs to have a better forecast of the PV generation. Two main aspects can help to achieve a better prediction:

- the combined effect of different ANNs compensates for the different random initializations;
- each MLP employs a slightly different number of hidden units.

There are several typologies of ensemble methods. In this work the BEM has been analyzed and implemented [44]. The BEM output is defined by:

$$f_{BEM}(t) = \frac{1}{n} \sum_{i=1}^{n} f_i(t);$$ (11)

where n is the total number of ANNs and $f_i(t)$ are the single networks outputs defined as a function of time index t.

3.3.2. Input Variables

A crucial part for the design and the implementation of a reliable forecasting algorithm based on ANNs is represented by the input selection. The set of inputs chosen for the PV day-ahead forecasting is the following:

- quarter of hour in the day (number from 1 to 96);
- day of the year (number from 1 to 366);
- ambient temperature [°C];
- relative Humidity [%];
- wind Speed [$\frac{m}{s}$];
- CC index [%].

The format of days and quarters of hour is chosen to take into account temporal autocorrelations of the target variable, as suggested in [45]. Temperature and wind speed are selected because they are involved in the panel efficiency estimation (see Section 3.1). Humidity is included because it influences temperature and irradiance [46], and it is exploited with interesting results in several literature works ([19,20,45]). Finally, CC represents a numerical index for the estimation of the sky covering. Notice that all the meteorological inputs must be provided by weather predictions.

An example of inputs for a single ANN within the BEM is reported in Table 1.

Table 1. BEM inputs used for the PV output forecasting (1 November 2018).

Quarter of Day (1–96)	Day of Year (1–366)	Ambient Temperature [°C]	Relative Humidity [%]	Wind Speed [$\frac{m}{s}$]	Cloud Cover [%]
1	305	11.73	0	0	0
2	305	11.78	8.33	0.21	0
3	305	11.83	16.67	0.42	0
⋮	⋮	⋮	⋮	⋮	⋮
96	305	13.42	100	1.30	7.33

3.3.3. Parameters Selection

After the definition of the inputs for a single ANN, another fundamental step for the implementation of the BEM is represented by the parameter selection of the single MLPs. A generic MLP is characterized by the following parameters:

- Number of hidden layers;

- Neurons number in the hidden layers;
- Transfer functions between each layer. These functions define the relationship of inputs and outputs between each layer;
- Training algorithm. Any ANN must be trained on a knowledge dataset. This database is composed of an input vector and a score vector. The training algorithm defines the training method.

For the selection of all these parameters, the approaches proposed in [47–49] have been used. Table 2 reports the results of this phase. These parameters have been used in all the ANNs within the BEM.

Table 2. Selected hyperparameters for the single ANNs within the BEM.

Hidden layers number	1
Transfer functions among all the layers	Hyperbolic Tangent Sigmoid Function $tansig(x) = \frac{2}{[(1+e^{-2x})]} - 1$
Training algorithm	Resilient Back-propagation

In addition to these parameters related to a single ANN, for the BEM methodology is necessary to determine the number of MLP (i.e., parameter n in (11)) within the ensemble. The selection of this parameter has been performed using the strategy presented in [49]. The selected number of ANNs is $n = 6$: three with 52 neurons, two with 50 neurons, and one with 88 neurons.

3.4. Hybrid Technique for the Selection of the Most Appropriate Methodology

The accuracy of any of the previous forecasting techniques is strongly related to the weather conditions on the prediction window. Thus, the available information related to weather data could be exploited before the forecast execution to select the best methodology.

The selection steps can be two or one (see Figure 1):

1. DR1 consists in assessing whether the ANN approach (BEM) or a deterministic model is more convenient;
2. DR2 takes place whenever in DR1 the ANN methodology is not chosen. In that case, the second decision rule consists in the selection between two deterministic models: the CSM (see Section 3.1) or the CCSM (see Section 3.2).

The main idea is to compare, on a training set, the performance in terms of nRMSE of the two models under consideration in different climate conditions, evaluated through the forecast mean on the prediction horizon of several weather variables (CC, temperature, humidity, wind speed, and pressure). This is performed to determine which methodologies must be selected (CSM/CCSM or BEM) according to the different combinations of the considered weather variables. In this work the two choices have been implemented through a decision tree technique, but in principle any binary classifier that takes as inputs multiple numerical variables could be used.

3.4.1. Decision Tree Technique

Decision trees are composed of a series of If/Else rules on the regressors that lead to the output of the model. To predict a response, the user must follow the decisions in the tree from the root node down to a leaf node. This last node contains the response. The If/Else rules are also known as splits, while the regressors are often called attributes in this context.

There are several techniques for the design and implementation of a decision tree. In this work CART (Classification and Regression Trees) methodology has been employed [50].

CART can process nominal and continuous attributes both as targets and predictors. Given a training set, the algorithm grows the tree to its full size and then prunes it by eliminating the splits that give a little contribution to the overall performance and could produce overfitting [50].

The splits are chosen by inspecting all the possible cases on each attribute. Each possible splitting value divides the data that has reached the node into two groups.

CART produces a sequence of nested pruned trees that are candidate final trees. The final tree must be chosen by a comparison on a separate validation set [51].

3.4.2. Hybrid Procedure Implementation

Since the time horizon has been set to 24-h, also the two datasets built for the selection technique are composed of days. For the choice between the two deterministic approaches (DR2), only clear or almost clear-sky days are included.

The general procedure for the dataset definition and the implementation of the proposed hybrid technique is depicted in Figure 3 and can be summarized as follows:

1. the daily mean values of the aforementioned weather variables of all day types are computed;
2. taking into account only clear-sky days:

    (A) performance in terms of nRMSE of the deterministic models are computed;
    (B) their difference in terms of nRMSE is computed: the sign of the difference tells which model has performed better. In this block the numerical performance is transformed into a categorical label
    (C) exploitation of weather variables and labels for the implementation of DR2 through a decision tree approach;

3. considering all day types:

    (A) performance in terms of nRMSE of the BEM and the deterministic model selected by Decision 2 is calculated;
    (B) the difference between the two performances for each day is computed: the sign of the difference tells whether the ensemble or the deterministic model has performed better. Also, in this case the numerical performance is transformed into a categorical label;
    (C) exploitation of weather variables and labels for the implementation of DR1 through a decision tree approach.

Table 3 collects an example of the results obtained with steps (A), (B) related to DR2 for the definition of a dedicated database, which is used for the definition of the decision tree. Table 4 reports the same operation for DR1.

Table 3. Example of database definition for the decision tree—DR2.

Day	Atmospheric Pressure (Average) [hPa]	Ambient Temperature (Average) [°C]	Relative Humidity (Average) [%]	Wind Speed (Average) [$\frac{m}{s}$]	Cloud Cover (Average) [%]	Label
8 May	1018	17.44	72.93	0.27	18.83	CCSM
14 May	1016	19.26	35.65	1.86	10.42	CSM
29 May	1012	18.75	70.93	0.45	24.69	CSM
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Table 4. Example of database definition for the decision tree—DR1.

Day	Atmospheric Pressure (Average) [hPa]	Ambient Temperature (Average) [°C]	Relative Humidity (Average) [%]	Wind Speed (Average) [$\frac{m}{s}$]	Cloud Cover (Average) [%]	Label
9 May	1017	17.91	74.06	0.30	21.54	DR2
10 May	1012	17.90	81.91	0.87	61.41	BEM
11 May	1007	18.61	53.82	1.55	25.63	BEM
⋮	⋮	⋮	⋮	⋮	⋮	⋮
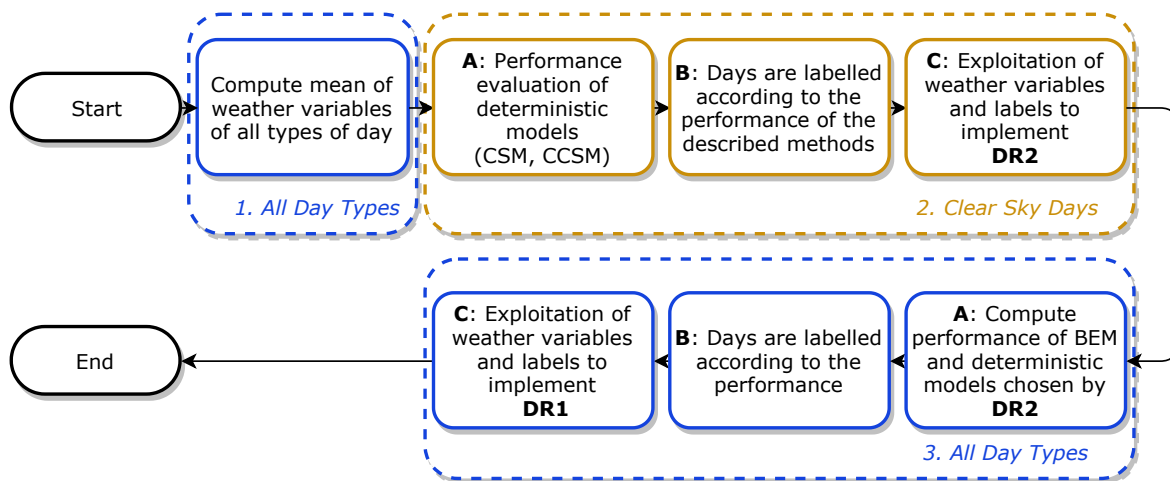
Figure 3. General procedure for the implementation of the hybrid selection method.

3.4.3. Selection of the Optimal Tree

A decision tree is grown on different training sets, i.e., one for the clear-sky choice (DR2, see Figure 1) and another one for the selection among a deterministic approach or BEM (DR1, see Figure 1).

To understand the process for the selection of the optimal pruning level, it is necessary to introduce two fundamental concepts:

- trivial tree, which is the tree that always labels the observations with the most frequent class. In DR1, for example, the trivial tree always selects the same method among the three proposed techniques (CSM, CCSM, and BEM). For this reason, it is not a suitable tree: it makes useless the hybrid technique of this section;
- pruning levels. These represent the orders of the nested pruned trees produced by CART. Pruning level 0 is the complete tree, which achieves perfect performance on the training set (and therefore is affected by overfitting problems). The maximum pruning level corresponds to the trivial tree.

The methodology adopted in this work for the selection of the best pruning level is to use the tree corresponding to the smallest pruning level that improves the trivial tree on a validation set. Moreover, only the first split of the resulting tree is considered.

3.4.4. Final CART Trees

The proposed decision tree technique applied to the considered test site (see Section 4) provides the hybrid method reported in this section. Figure 4 illustrates the final rules for the selection of the methodologies described in this paper.

Notice that despite different weather variables have been considered during the tree implementation, both the decisions are only based on the CC index. From Figure 4 it can be noticed that in DR1 the BEM is chosen when the CC index is particularly high, confirming that the ANN

technique is better in cases wherein the sky is far from being clear (cloudy/rainy days), confirming that the intuition behind the use of deterministic models is correct.
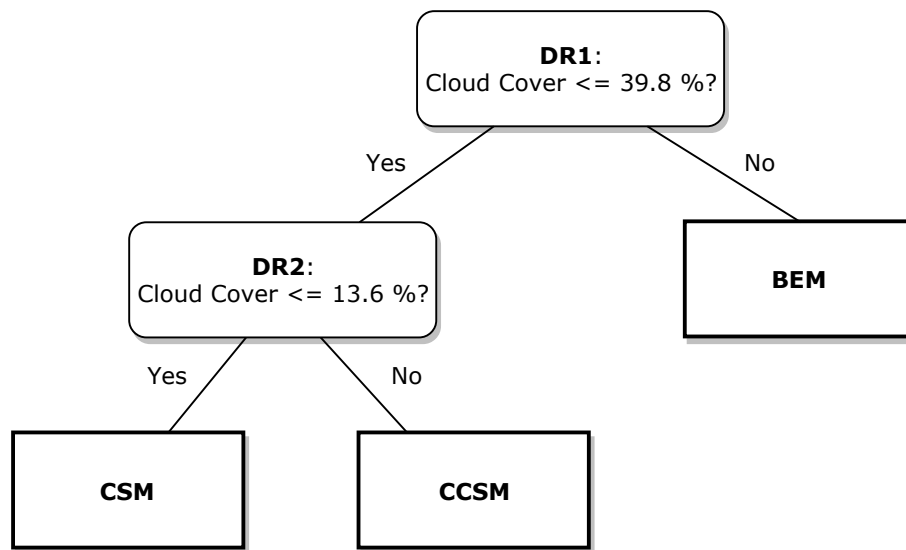

Figure 4. The final hybrid technique for the selection of the most appropriate model.

4. Test Site

The test system considered in this work for the validation of the proposed hybrid forecasting procedure is a PV plant located in the harbor of Genova. In particular, the PV system is positioned on the rooftop of the Economics School of the University of Genova (see Figure 5). The building is oriented with respect to the south of about 30° towards west.

Figure 5. The considered PV system. Picture from the rooftop of the Economics School.

The considered PV system presents a peak power of about 20 kW$_p$ and it is directly connected to the electric system of the underneath building. The photovoltaic modules are supported by an aluminum structure of 51 m × 3.3 m, which has a 30° inclination (tilt angle) with respect to the horizon. The modules are composed of multi-crystalline silicon and each of them can produce 180 W. The dimension of each panel is 1.3 m^2. A total of 108 panels are installed on the structure. The modules are supplied by 2 inverters, with nominal power equal to 12.5 kW.

Table 5 collects all the main parameters of the test site considered in this work. This table reports also the symbols related to the parameters.

Table 5. Technical parameters of the test site.

Parameter	Value
Peak Power (P_{peak})	20 kW$_p$
Efficiency of the Inverter (η_{inv})	0.971
Temperature Coefficient (β_c)	$-0.045 \frac{1}{°C}$
Tilt Angle (β)	30°
Latitude	44.4141°
Longitude	8.9221°
Local Time Zone	15°
Orientation regarding S, positive W	30°
Number of Inverters	2
Inverters Nominal power	12.5 kW
Number of Panels	108
Nominal Power of each Panel	180 W
Surface of each Panel	1.3 m^2
Coefficient of Degradation (ω_{DEG})	1

4.1. Shading Modeling

The PV modules are positioned on the roof to minimize the losses of irradiance due to the shadows of the higher surrounding buildings. Nevertheless, the considered PV system is slightly shaded by surrounding buildings (see Figure 6), especially in the morning and in the late afternoon.

Figure 6. Top view of the considered PV system.

The software used in this work for the shadows modeling is PVSyst 6.6.4 [52]. This software allows the user to draw the shape and the dimension of the buildings surrounding the PV plant (Figure 7). The measurements needed for the modeling are obtained through Google Earth Pro. It is possible to have good measurements of both lengths and angles through the ruler functionality.

The shapes of the near buildings are various and therefore the most appropriate geometrical model must be carefully chosen for each of them. In addition, also the colors of the building are important to set because they influence the outcome of the ground albedo, which in turn influences the value of $E_{r,pv}$ (see (8)).

Once the drawing has been made, the software is able to estimate all the shadow-related parameters involved in (7) (S_{dir}, S_{diff} and ρ_g). In particular, the software provides a table that presents values of S_{dir} for different combinations of the sun position (see Table 6). Using linear

interpolation, these values can be exploited to compute the parameters for any combination of azimuth and solar altitude.
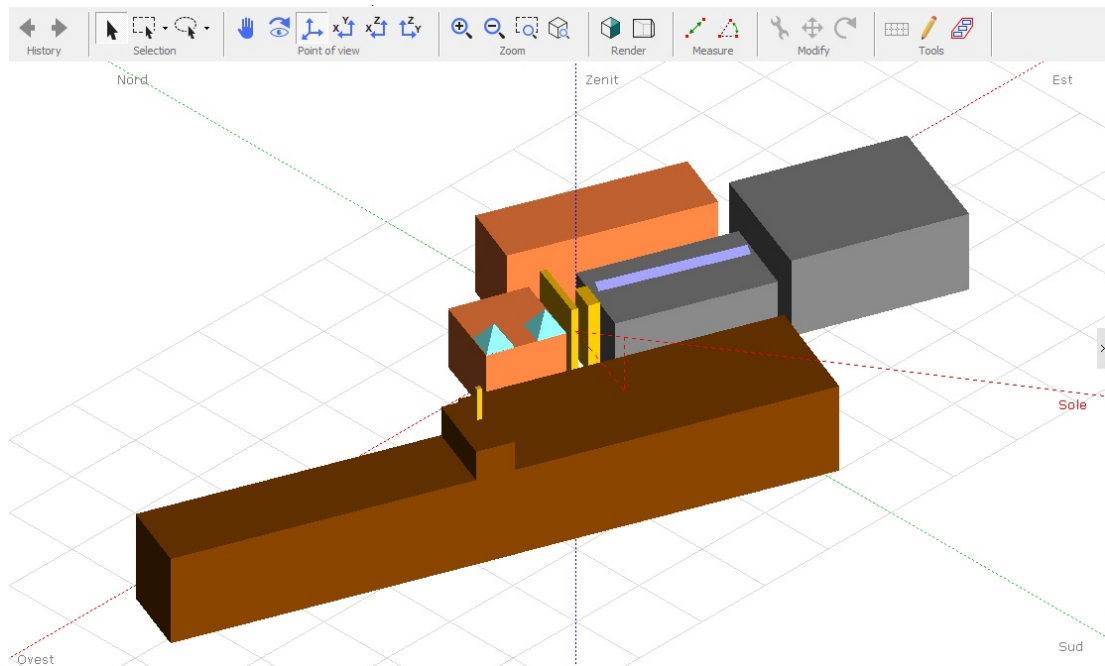

Figure 7. Shadows modeling of the surrounding buildings.

Table 6. S_{dir} approximation as a function of Azimuth (Az.) and Height (Hgt.).

Hgt. Az.	−180°	−100°	−80°	−60°	−40°	−20°	0°	20°	40°	60°	80°	100°	180°
90°	0	0	0	0	0	0	0	0	0	0	0	0	0
80°	0	0	0	0	0	0	0	0	0	0	0	0	0
70°	0	0	0	0	0	0	0	0	0	0	0	0	0
60°	0	0	0	0	0	0	0	0	0	0	0	0	0
50°	0	0	0	0	0	0	0	0	0	0	0	0	0
40°	0	0	0	0	0	0	0	0	0	0	0	0	0
30°	1	0	0.1	0.1	0	0	0	0	0	0	0	0	1
20°	1	0.1	0.2	0.2	0.1	0.1	0	0	0	0	0	0	1
10°	1	1	0.4	0.4	0.4	0.2	0.1	0.4	0.3	0	0.1	0.1	1
2°	1	1	1	1	0.5	0.2	0.3	0.8	1	1	1	1	1

4.2. Available Data

This section describes the available data for the considered test site.

An historical database of the PV plant, described in this section, collects data from 2014 related to the power production. For the weather variables, two different sources have been considered:

1. A weather station located just outside the PV array. This device can provide measurement data related to the actual temperature, humidity, and wind speed with a granularity of 15 min. In addition, the weather station has its own historical database that collects measurements since 2014;

2. A web weather provider [53]. From this website it is possible to download a historical bulk dataset that contains all the weather information and, in particular, the crucial data related to the CC index. This weather provider has been employed for the weather forecasting of the variables used by the proposed hybrid procedure. Weather data have been imported from the provider through the dedicated Application Programming Interface (API) to be quickly stored

and analyzed. The data obtained from [53] are related to a position which is one kilometer away from the PV system.

To have a reliable dataset, several preprocessing actions have been performed (such as outlier identification and missing data management). The result of the preprocessing stage is a database with measurements related to more than 70,000 quarters of hour (2 years) of complete, reliable data. According to literature, this represents a robust dataset for the implementation of an accurate forecasting procedure [20]. The preprocessed, historical data have been used in the training phase of all the described methodologies.

5. Results

In this Section, the hybrid model, the ANN Ensemble (BEM), the CCSM and the CSM models described in Section 3 are tested on two months (November–December 2018) for a total of 61 days.

All the proposed methodologies need weather forecast, provided by [53]. Thus, for each day of the testing period a MATLAB routine has launched at 23:45 p.m. to retrieve the meteorological predictions. The PV forecasting is then executed automatically at midnight by making use of the weather forecast.

Table 7 reports the performance of the base methodologies (CSM, CCSM, BEM), as well as the hybrid approach.

Table 7. Hybrid and non-hybrid methods performance on online test set.

Method	KPIs			Days Number		
	nRMSE	RMSE [kW]	MBE [kW]	BEM	CCSM	CSM
BEM	0.6090	2.4384	−0.8092	61	0	0
CCSM	0.6702	2.6833	0.7194	0	61	0
CSM	0.8143	3.2601	1.1996	0	0	61
Hybrid Tree	**0.3892**	**1.5583**	**−0.0694**	20	13	28
Ideal Hybrid	0.2761	1.1053	−0.0127	21	13	27

The best stand-alone method is the BEM, which outperforms the two clear-sky methods. However, thanks to the decision rule, the BEM can be enhanced by the two CSMs, whose individual performance are worse than the ANN approach. The last line of Table 7 reports the results of an ideal hybrid method, which is a hypothetical technique that chooses always the most accurate methodology. Since its nRMSE is lower with respect to the error committed by the proposed approach, it is possible to understand that there is a margin for an improvement of the described decision rule. This can lead to an increment of the accuracy of the proposed hybrid methodology.

The proposed hybrid methodology can be seen as a three categories classification problem. Tables 8 and 9 report respectively the confusion matrices related to the hybrid model and to the BEM (which can be viewed as a hybrid model that always chooses the ANN approach).

Numbers on the main diagonal identify the days wherein the actual ideal model is chosen. The BEM is included because it is the best trivial classifier (i.e., a classifier that chooses the most populated category).

As a classifier, the BEM has been right 24 times out of 61 (39% accuracy, see Table 9) while the proposed hybrid method has been right twice more: 50 out of 61 times (85% accuracy, Table 8).

Table 8. Confusion matrix for the hybrid technique.

Method		Ideal		
		CSM	CCSM	BEM
Predicted	CSM	18	2	0
	CCSM	2	8	3
	BEM	1	3	24

Table 9. Confusion matrix for the BEM method.

Method		Ideal		
		CSM	CCSM	BEM
Predicted	CSM	0	0	0
	CCSM	0	0	0
	BEM	21	13	27

This analysis is very important because it provides a performance estimation of the DRs that define the proposed hybrid approach.

Table 10 reports the Skill Scores for the comparison of the proposed hybrid approach with the single SMs implemented in this work and other recent methodologies. In addition, this table collects also the different KPIs used in (5) for the evaluation of the SS. Notice that for the literature comparison it has been considered the KPI average obtained in all the experimental tests.

Table 10. Skill Scores evaluation for a comparison with different day-ahead methodologies.

Reference Method	SS	KPI
CSM	+0.5220	nRMSE
CCSM	+0.4192	nRMSE
BEM	+0.3609	nRMSE
[54]	+0.1147	nRMSE
[55]	+0.0137	$RMSE/P_{peak}$
[56]	+0.1018	$(Average(|\varepsilon(t)|))/P_{peak}$

In [54] a hierarchical approach based on machine learning methods has been implemented, while in [55] a similar day PV forecasting technique has been adopted. The authors in [56] present an ensemble of five methods (Grey-Box Model, ANN, K-Nearest Neighbors, Quantile Random Forest and Support Vector Regression), proving that their strategy provides a more accurate forecast with respect to the single approaches.

As can be seen from Table 10, all the SSs are positive highlighting the precision of the forecast approach.

This literature comparison, even if on different test sites/sets, suggests that the proposed hybrid technique is a robust and an accurate procedure, representing a useful and reliable functionality for the uncertainty management.

Focusing on the hybrid model, Figures 8–10 are useful to inspect the improvement of this approach with respect to the versatile ANN-based approach. As can be seen from these Figures the results of the proposed methodology are very satisfying.

As testified by the Figures 8–10, the BEM can attain a large error in clear days, wherein the deterministic models provide better results. For this reason, the proposed hybrid technique gives an important contribution to improve the overall accuracy of the PV forecast.
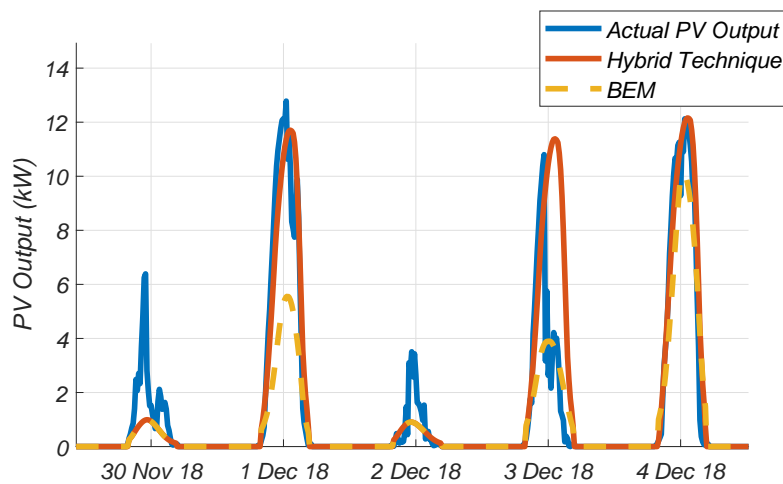
Figure 8. PV output, BEM (yellow dotted line) and proposed hybrid technique (red solid line) for five days of December 2018.
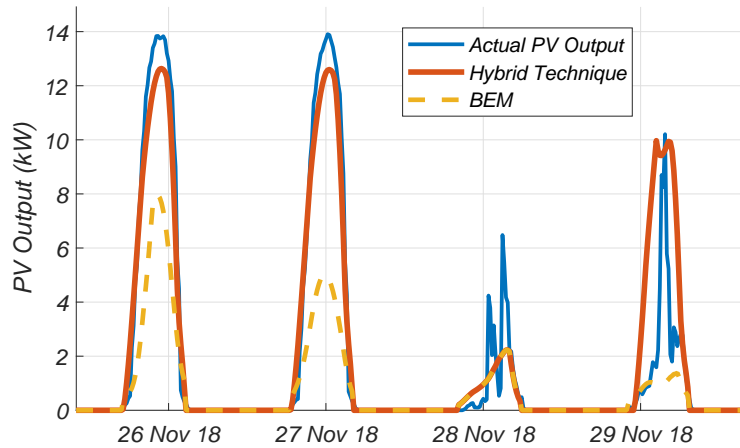

Figure 9. PV output, BEM (yellow dotted line) and proposed hybrid technique (red solid line) for four days of November 2018.
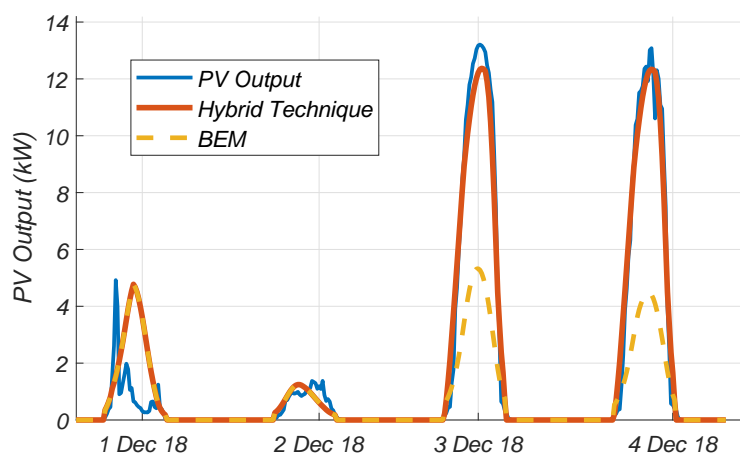

Figure 10. PV output, BEM (yellow dotted line) and proposed hybrid technique (red solid line) for the first four days of December 2018.

6. Conclusions

A novel hybrid methodology for the day-ahead forecasting has been implemented and validated on a real PV plant. The presented approach can exploit CSMs or an ensemble of ANNs, according to day-ahead weather forecast. The selection among these techniques is performed through a decision tree approach.

The novel hybrid procedure looks promising, as testified from all the results of Section 5.

The proposed methodology presents all the good properties of an ensemble method, coupled with a robust performance on clear-sky days. In particular, it has outperformed the BEM method in realistic forecasting tests, giving an accurate prediction in all weather conditions. In addition, the performance of the presented approach has been very satisfactory, as indicated by a comparison in terms of accuracy with the literature.

The final decision tree provides good results for the selection of the most appropriate method and it reflects the intuition behind the use of deterministic models or an ANN-based approach.

For all these reasons, the presented approach can be employed as an input of optimization or advanced algorithms within generation scheduling/unit commitment applications, energy management strategies, grid regulation procedures, etc.

Several parts of the proposed algorithms could be investigated to improve the forecasting accuracy:

1. the CCSM procedure could include more inputs (currently only CC is considered to be a weather regressor) to be more competitive with respect to the other methodologies;
2. more sophisticated ensemble techniques can be considered;
3. different architecture of ANNs could be analyzed;
4. the hybrid technique could be designed to select different methods within the same day;
5. a three-category classifier could substitute the two decisions that compose the hybrid approach;
6. Model Output Statistics (MOS) could be performed on weather forecasts to understand if they are affected by any bias;
7. Irradiance data could be used instead of CC index by the proposed hybrid approach.

In particular, points 1, 5, and 7 have potentially the greatest impact on the performance, and therefore they would be the first to be investigated in future works.

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

References

1. *World Energy Outlook 2017—Executive Summary*; Technical Report; International Energy Agency: Pairs, France, 2017. Available online: https://www.iea.org/publications/freepublications/publication/WEO_2017_Executive_Summary_English_version.pdf (accessed on 30 March 2019).
2. Sawin, J.L.; Sverrisson, F.; Rutovitz, J.; Dwyer, S.; Teske, S.; Murdock, H.E.; Adib, R.; Guerra, F.; Blanning, L.H.; Hamirwasia, V.; et al. *Renewables 2018 Global Status Report*; Technical Report; REN 21: Pairs, France, 2018.
3. Tuhoy, A.; Zack, J.; Haupt, S.H.; Sharp, J.; Ahlstrom, M.; Dise, S.; Grimit, E.; Mohrlen, C.; Lange, M.; Garcia Casado, M.; et al. Solar Forecasting: Methods, Challenges and Performance. *IEEE Power Energy Mag.* **2015**, *13*, 50–59. [CrossRef]
4. Adinolfi, F.; Baccino, F.; D'Agostino, F.; Massucco, S.; Saviozzi, M.; Silvestro, F. An architecture for implementing state estimation application in Distribution Management System (DMS). In Proceedings of the Innovative Smart Grid Technologies Conference (ISGT 2013), Lyngby, Denmark, 6–9 October 2013.

5.   Ming, B.; Liu, P.; Guo, S.; Cheng, L.; Zhou, Y.; Gao, S.; Li, H. Robust hydroelectric unit commitment considering integration of large-scale photovoltaic power: A case study in China. *Appl. Energy* **2018**, *228*, 1341–1352. [CrossRef]

6.   Namor, E.; Sossan, F.; Cherkaoui, R.; Paolone, M. Control of Battery Storage Systems for the Simultaneous Provision of Multiple Services. *IEEE Trans. Smart Grid* **2018**. [CrossRef]

7.   Van Der Meer, D.; Chandra Mouli, G.; Morales, G.; Elizondo, L.; Bauer, P. Energy Management System with PV Power Forecast to Optimally Charge EVs at the Workplace. *IEEE Trans. Ind. Inform.* **2018**, *14*, 311–320. [CrossRef]

8.   Das, U.K.; Tey, K.S.; Seyedmahmoudian, M.; Mekhilef, S.; Idris, M.Y.I.; Deventer, W.V.; Horan, B.; Stojcevski, A.   Forecasting of photovoltaic power generation and model optimization: A review. *Renew. Sustain. Energy Rev.* **2018**, *81*, 912–928. [CrossRef]

9.   Conte, F.; Massucco, S.; Saviozzi, M.; Silvestro, F. A Stochastic Optimization Method for Planning and Real-Time Control of Integrated PV-Storage Systems: Design and Experimental Validation. *IEEE Trans. Sustain. Energy* **2018**, *9*, 1188–1197. [CrossRef]

10. French Electricity Regulations. July 2018. Available online: https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000027262791 (accessed on 30 March 2019).

11. Kudo, M.; Takeuchi, A.; Nozaki, Y.; Endo, H.; Sumita, J. Forecasting electric power generation in a photovoltaic power system for an energy network. *Electr. Eng. Jpn.* **2009**, *167*, 16–23. [CrossRef]

12. Pelland, S.; Remund, J.; Kleissl, J.; Oozeki, T.; De Brabandere, K. *Photovoltaic and Solar Forecasting: State of the Art*; Technical Report; IEA: Pairs, France, 2013.

13. Ogliari, E.; Dolara, A.; Manzolini, G.; Leva, S. Physical and hybrid methods comparison for the day ahead PV output power forecast. *Renew. Energy* **2017**, *113*, 11–21. [CrossRef]

14. Li, Y.; Su, Y.; Shu, L. An ARMAX model for forecasting the power output of a grid connected photovoltaic system. *Renew. Energy* **2014**, *66*, 78–89. [CrossRef]

15. Rana, M.; Koprinska, I.; Agelidis, V. 2D-interval Forecasts for Solar Power Production. *Sol. Energy* **2015**, *122*, 191–203. [CrossRef]

16. Xie, T.; Zhang, G.; Liu, H.; Liu, F.; Du, P. A Hybrid Forecasting Method for Solar Output Power Based on Variational Mode Decomposition, Deep Belief Networks and Auto-Regressive Moving Average. *Appl. Sci.* **2018**, *8*, 1901. [CrossRef]

17. Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R.; Martinez-de Pison, F.; Antonanzas-Torres, F. Review of photovoltaic power forecasting. *J. Sol. Energy* **2016**, *136*, 78–111. [CrossRef]

18. De Giorgi, M.G.; Congedo, P.M.; Malvoni, M. Photovoltaic power forecasting using statistical methods: Impact of weather data. *IET Sci. Meas. Technol.* **2014**, *8*, 90–97. [CrossRef]

19. Zhang, X.; Jiang, B.; Zhang, X.; Fang, F.; Gao, Z.; Feng, T. Solar Photovoltaic Power Prediction Based on Similar Day Approach. In Proceedings of the 36th Chinese Control Conference (CCC 2017), Dalian, China, 26–28 July 2017; pp. 10634–10639.

20. Ogliari, E.; Gandelli, A.; Grimaccia, F.; Leva, S.; Mussetta, M. Neural Forecasting of the Day-Ahead Hourly Power Curve of a Photovoltaic Power Plant. In Proceedings of the International Joint Conference on Neural Networks (IJCNN 2016), Vancouver, BC, Canada, 24–29 July 2016; pp. 654–659.

21. Cristaldi, L.; Leone, G.; Ottoboni, R. A Hybrid Approach for Solar Radiation and Photovoltaic Power Short-Term Forecast. In Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (I2MTC 2017), Turin, Italy, 22–25 May 2017; pp. 1252–1257.

22. Asrari, A.; Wu, T.X.; Ramos, B. A Hybrid Algorithm for Short-Term Solar Power Prediction—Sunshine State Case Study. *IEEE Trans. Sustain. Energy* **2017**, *8*, 582–591. [CrossRef]

23. Yang, H.T.; Huang, C.M.; Huang, Y.C.; Yi, S.P. A Weather-Based Hybrid Method for 1-Day Ahead Hourly Forecasting of PV Power Output. *IEEE Trans. Sustain. Energy* **2014**, *5*, 917–926. [CrossRef]

24. Sánchez-Garcia, J.L.; Espinosa-Juárez, E.; Flores, J.J. Short Term Photovoltaic Power Production Using a Hybrid of Nearest Neighbor and Artificial Neural Networks.   In Proceedings of the IEEE PES Transmission & Distribution Conference and Exposition-Latin America (PES T& D–LA), Morelia, Mexico, 20–24 September 2016; pp. 1–6.

25. Shi, J.; Liu, Y.; Yang, Y. Forecasting Power Output of Photovoltaic Systems Based on Weather Classification and Support Vector Machines. *IEEE Trans. Ind. Appl.* **2012**, *48*, 1064–1069. [CrossRef]

26. Huang, C.M.; Chen, S.J.; Yang, S.P.; Kuo, C.J. One-day-ahead hourly forecasting for photovoltaic power generation using an intelligent method with weather-based forecasting models. *IET Gener. Transm. Distrib.* **2015**, *9*, 1874–1882. [CrossRef]

27. Rana, M.; Koprinska, I.; Agelidis, V.G. Solar Power Forecasting Using Weather Type Clustering and Ensembles of Neural Networks. In Proceedings of the International Joint Conference on Neural Networks (IJCNN 2016), Vancouver, BC, Canada, 24–29 July 2016; pp. 4962–4969.

28. Li, Z.; Lu, Z.; Qiao, Y.; Wang, N.; Ding, K. Weather type partition method considering sequential features in photovoltaic forecasting. *J. Eng.* **2017**, *13*, 1259–1263. [CrossRef]

29. Wang, Y.; Liao, W.; Chang, Y. Gated Recurrent Unit Network-Based Short-Term Photovoltaic Forecasting. *Energies* **2018**, *11*, 2163. [CrossRef]

30. Wang, F.; Zhen, Z.; Wang, B.; Mi, Z. Comparative Study on KNN and SVM Based Weather Classification Models for Day Ahead Short Term Solar PV Power Forecasting. *Appl. Sci.* **2017**, *8*, 28. [CrossRef]

31. Coimbra, C.; Kleissl, J.; Marquez, R. Overview of solar forecasting methods and a metric for accuracy evaluation. In *Solar Resource Assessment and Forecasting*; Elsevier: Amsterdam, The Netherlands, 2013; pp. 171–194.

32. Raza, M.Q.; Nadarajah, M.; Ekanayake, C. On recent advances in PV output power forecast. *Sol. Energy* **2016**, *136*, 125–144. [CrossRef]

33. Wagner, A. *Photovoltaik Engineering*; Springer: Berlin, Germany, 2009.

34. Jordan, D.; Kurtz, S.; VanSant, K.; Newmiller, J. Compendium of Photovoltaic Degradation Rates. *Prog. Photovolt. Res. Appl.* **2016**, *24*, 978–989. [CrossRef]

35. Quansah, D.; Adaramola, M.; Takyi, G.; Edwin, I. Reliability and Degradation of Solar PV modules—Case Study of 19-Year-Old Polycrystalline Modules in Ghana. *Technologies* **2017**, *5*, 22. [CrossRef]

36. Quaschning, V.; Hanitsch, R. Irradiance calculation on shaded surfaces. *Sol. Energy* **1998**, *62*, 369–375. [CrossRef]

37. Liu, B.Y.H.; Jordan, R.C. The interrelationship and characteristic distribution of direct, diffuse and total solar radiation. *Sol. Energy* **1960**, *4*, 1–9. [CrossRef]

38. King, D. Photovoltaic Module and Array Performance Characterization Methods for All System Operating Conditions. In Proceedings of the NREL/SNL Photovoltaics Program ReviewMeeting, Lakewood, CO, USA, 18–22 November 1996.

39. Skoplaki, E.; Boudouvis, A.G.; Palyvos, J.A. A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting. *Sol. Energy Mater. Sol. Cells* **2008**, *92*, 1393–1402. [CrossRef]

40. Muzathik, A.M. Photovoltaic Modules Operating Temperature Estimation Using a Simple Correlation. *Int. J. Energy Eng.* **2014**, *4*, 151–158.

41. Koehl, M.; Heck, M.; Wiesmeier, S.; Wirth, J. Modeling of the nominal operating cell temperature based on outdoor weathering. *Sol. Energy Mater. Sol. Cells* **2011**, *95*, 1638–1646. [CrossRef]

42. Draper, N.R.; Smith, H. *Applied Regression Analysis*; CRC Press: Boca Raton, FL, USA, 1998.

43. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. *Learning Internal Representations by Error Propagation*; MIT Press: Cambridge, MA, USA, 1980.

44. Perrone, M.P.; Cooper, L.N. When networks disagree: Ensemble methods for hybrid neural networks. In *Neural Networks for Speech and Image Processing*; Mammone, R.J., Ed.; CRC Press: Boca Raton, FL, USA, 1993.

45. Ceci, M.; Corizzo, R.; Fumarola, F.; Malerba, D.; Rashkovska, A. Predictive Modeling of PV Energy Production: How to Set Up the Learning Task for a Better Prediction? *IEEE Trans. Ind. Inform.* **2017**, *13*, 956–966. [CrossRef]

46. Mekhilef, S.; Saidur, R.; Kamalisarvestani, M. Effect of dust, humidity and air velocity on efficiency of photovoltaic cells. *Renew. Sustain. Energy Rev.* **2012**, *16*, 2920–2925. [CrossRef]

47. Bagnasco, A.; Saviozzi, M.; Silvestro, F.; Vinci, A.; Grillo, S.; Zennaro, E. Artificial Neural Network Application to Load Forecasting in a Large Hospital Facility. In Proceedings of the International Probabilistic Methods Applied to Power Systems Conference (PMAPS 2014), Durham, UK, 7–10 July 2014; pp. 1–6.

48. Adinolfi, F.; D'Agostino, F.; Massucco, M.; Morini, A.; Saviozzi, M.; Silvestro, F. Pseudo-Measurement Modeling Using Neural Network and Fourier Decomposition for Distribution State Estimation. In Proceedings of the Innovative Smart Grid Technologies Conference (ISGT 2014), Istanbul, Turkey, 12–15 October 2014; pp. 1–6.

49. Saviozzi, M.; Massucco, S.; Silvestro, F. Implementation of Advanced Functionalities for Distribution Management Systems: Load Forecasting and Modeling through Artificial Neural Networks ensembles. *Electr. Power Syst. Res.* **2019**, *167*, 230–239. [CrossRef]

50. Breiman, L.; Freidman, J.H.; Olsen, R.A.; Stone, C.J. *Classification and Regression Trees*; CRC Press: Boca Raton, FL, USA, 1984.

51. Wu, X.; Kumar, V.; Ross Quinlan, J.; Gosh, J.; Yang, Q.; Motoda, H.; McLachlan, G.J.; Ng, A.; Liu, B.; Yu, P.S.; et al. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* **2008**, *14*, 1–37. [CrossRef]

52. *PVSyst 6.6.4 Software*; PVsystSA: Satigny, Switzerland, 2019. Available online: http://www.pvsyst.com/en/ (accessed on 30 March 2019).

53. OpenWeatherMap Inc. Available online: http://www.openweathermap.org/ (accessed on 30 March 2019).

54. Li, Z.; Rahman, S.; Vega, R.; Dong, B. A Hierarchical Approach Using Machine Learning Methods in Solar Photovoltaic Energy Production Forecasting. *Energies* **2016**, *9*, 55. [CrossRef]

55. Zhang, Y.; Beaudin, M.; Taheri, R.; Zareipour, H.; Wood, D. Day-Ahead Power Output Forecasting for Small-Scale Solar Photovoltaic Electricity Generators. *IEEE Trans. Smart Grid* **2015**, *6*, 2253–2263. [CrossRef]

56. Gigoni, L.; Betti, A.; Crisostomi, E.; Franco, A.; Tucci, M.; Bizzarri, F.; Mucci, D. Day-Ahead Hourly Forecasting of Power Generation From Photovoltaic Plants. *IEEE Trans. Sustain. Energy* **2018**, *9*, 831–842. [CrossRef]