Physical and hybrid methods comparison for the day ahead PV output power forecast

Emanuele Ogliari*, Alberto Dolara, Giampaolo Manzolini, Sonia Leva

Politecnico di Milano, Dipartimento di Energia, Via Lambruschini 4, 20156 Milano, Italy

An accurate forecast of the exploitable energy from Renewable Energy Sources, provided 24 h in advance, is becoming more and more important in the context of the smart grids, both for their stability issues and the reliability of the bidding markets. This work presents a comparison of the PV output power day-ahead forecasts performed by deterministic and stochastic models aiming to find out the best perfor-mance conditions. In particular, we have compared the results of two deterministic models, based on three and five parameters electric equivalent circuit, and a hybrid method based on artificial neural network. The forecasts are evaluated against real data measured for one year in an existing PV plant located at SolarTech^{lab} in Milan, Italy. In general, there is no significant difference between the two deterministic models, being the three-parameter approach slightly more accurate (NMAE three-parameter 8.5% vs. NMAE five-parameter 9.0%). The artificial neural network, combined with clear sky solar radiation, generally achieves the best forecasting results (NMAE 5.6%) and only few days of training are necessary to provide accurate forecasts.

Keywords: PV forecast power production, Artificial neural network, PV equivalent electrical circuit, NMAE, SolarTech^{lab}

1. Introduction

The renewable energy power production has been significantly increasing world-wide thanks to technological advancements, cost reduction and government policies: the world-wide installed capacity of renewable energy sources (RES) has almost doubled since 2006 (1036 GW in 2006 vs. 1985 GW in 2015) [31]. A significant part of this increase is connected with the larger capacity of photovoltaic (6 GW vs. 222 GW) and wind energy (73 GW vs. 440 GW). Therefore, the power generation sector is facing significant challenges in the reliability and control of the power transmission system. Because of the higher share of variable RES, grid operators must control a variable power supply together with the variable demand.

Furthermore, the "day-ahead electricity market" is a market for wholesale electricity trading where prices and exchanged quantities are defined for the following day, hour by hour. In Italy, the market session begins at 8 p.m. of the ninth day before the delivery day and it closes at 9.15 p.m. of the day before the delivery day.

Two options are available for controlling the grid while using the whole renewable energy produced: the first one is based on the installation of large storage systems (including batteries, pump-hydro, etc.) and the second one consists of developing accurate models for power generation forecast using the expected weather conditions. This work fits in the second group and will compare different methods for PV plant output power forecasts of the next 24 hourly samples, which are estimated the day before on the basis of the available weather forecasts (this approach is called "day-ahead forecast") [7].

The existing approaches of forecasting models can be classified into the following three categories: physical, statistical and hybrid [35].

Physical models describe the conversion processes, from solar to electricity, in PV modules and we adopted them to predict the daily generated power by using the expected weather conditions in a given day.

On the contrary, statistical methods are based on the concept of persistence, or stochastic time series, and are typically relying on machine learning methods. They can be applied in the field of the power forecast from renewable energy sources, because it is possible to train them through historical data (for example: a certain daily power measured in correspondence to certain weather forecasts, which is called "supervised learning") with the aim of returning a power value foreseen also starting from new

Article history:
Received 27 January 2017
Received in revised form 14 April 2017
Accepted 19 May 2017
Available online 20 May 2017

* Corresponding author.
E-mail address: emanuelegiovanni.ogliari@polimi.it (E. Ogliari).
URL: http://www.solartech.polimi.it

Nomenclature

AE_{max}	Maximum Absolute Error (W)
ANN	artificial neural network
C	Net Capacity of the plant (W)
DC	Direct Current (W)
e	Error (W)
G_{POA}	Plane Of Array total solar irradiance ($W \cdot m^{-2}$)
GHI	Global Horizontal Solar irradiance ($W \cdot m^{-2}$)
I	Current (A)
I_D	Current across the PN junction (A)
I_0	Reverse saturation current (A)
I_{PV}	Light-generated current (A)
K_d	Number of days included in the training and validation datasets
MAE	Mean Absolute Error (W)
MBE	Mean Biased Error (W)
MLP	Multilayer Perceptron
n	Diode ideality factor
N	number of hours in the time span (h)
N_d	maximum number of available days in the training and validation datasets
NMAE	Normalized mean absolute error (%)
NOCT	Nominal Operating Cell Temperature (K)
P	Output Power (W)
PHANN	Physical Hybrid Artificial Neural Network
PV	Photovoltaic
R_S	Series resistance (Ω)
R_{SH}	Shunt resistance (Ω)
T	Temperature (K)
V	Voltage (V)
V_t	Thermovoltage (V)
X_d	Forecast day

Greek symbols

β	Tilt angle (°)
σ	Pearson standard deviation
σ^2	variance

Subscript

amb	ambient
c	cell
p	forecasted value
m	measured value
M	module
h	hourly average sample
d	day
ref	parameter reference value

data. This is called: "generalization capability".

With respect to these statistical methods, ANNs have been successfully applied for the forecasts of fluctuating energy supply [17,21,25]. These methods learn to recognize patterns in data using training data sets, which is also their main disadvantage. Yet in a review [1] they are proved to be the most applied techniques. In Ref. [35] any combination of two or more previously described physical/statistical methods leads to a hybrid model. The idea is to combine different models with unique features to overcome the single negative performance and finally improve the forecast [6,30].

There are many papers dealing with a comparison among forecasting techniques, between physical/deterministic or statistical approach, but they are not including ANN and they mainly consider the short-term time frame. For example [36] focuses PV power forecasts by comparing the support vector regression (SVR) and a physical method, in the range of 15 min to 5 h ahead.

The aim of this study is to present comparison of the PV output power day-ahead forecasts performed by different methods by providing the same weather forecasts to different models:

- deterministic (3 and 5 parameters PV models, with NOCT PV cells thermal model);
- hybrid (PHANN – Physical Hybrid Artificial Neural Network)

We also found out, on average, the most effective forecasting model starting from the first day of the PV plant operation. While deterministic models are employable from the beginning, the hybrid method requires historical data to be trained.

For this reason we aim also to determine the most performing features, in terms of accuracy provided by different approaches in the training set composition as a function of the increasing training set size.

This comparison and analysis was realized on experimental data recorded from a PV module at SolarTech^{Lab} located in Milano [34] – Italy – assessing the day-ahead forecast with the most common error definitions.

This paper is structured as follows: from Section 2 to Section 4 the physical and stochastic models are described. Section 5 presents the test case selected for the analysis and the results are reported in Section 6. Finally, section 7 reports the main conclusion of the work and outlines additional research directions for the future activities.

2. Deterministic models

The deterministic models are based on a physical representation of a PV cell by equivalent electrical circuit and the PV module is an extension of the single cell considering the adequate number of cells in series. The sunlight conversion into electricity occurring in the PV cell can be represented either by a current generator. This simple representation is typically named three parameters model from the number of parameters required to define the I-V curve [9,10]. The electric equivalent representation can be more complex as in the five parameters equivalent circuit (see Fig. 1) [4,8–10,13,14,24,27,32,19]. In this case, there are two resistances, the light-generated current (I_{PV}) and the electrical behavior of the PN junction (I_D). The cell series resistance ($R_{S,c}$) and the cell shunt resistance ($R_{SH,c}$) are connected in series to the cell terminals and in parallel to the current generator, respectively. In particular, $R_{S,c}$ and

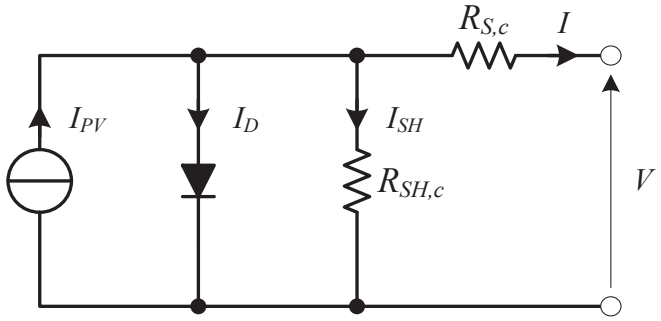

Fig. 1. Equivalent circuit representing the five-parameter model.

$R_{SH,c}$ represent the Ohmic losses and the recombination losses occurring in the cell. The five-parameter model is described in Eq. (1):

$$I = I_{PV} - I_0 \cdot \left(e^{\frac{V + R_{S,c} \cdot I}{n \cdot V_t}} - 1 \right) - \frac{V + R_{S,c} \cdot I}{R_{SH,c}} \quad (1)$$

In particular, V_t is the thermovoltage, I_0 and n represent the reverse saturation current of the pn-junction and the diode ideality factor, respectively.

Other more complex representations, namely six or seven parameters, are available in literature but not reported here for sake of brevity. In fact, in this work, the three parameters and the five parameters approaches are evaluated. The former is the simplest which requires a limited amount of information (OC, SC and MPP) for its characterization and the latter is a good compromise between complexity and accuracy as also demonstrated in a previous work [10].

The parameters for the three- and five-parameter models can be determined either by using the information reported in the datasheet, or starting from measured values. Being the three-parameter approach quite simple, the datasheet information is enough to characterized it, while the definition of the five-parameter model requires a full I-V curve measurements. In this work, the five parameters are fitted from the experimental I-V curves of the considered modules minimizing the square error. For further information about the approach, it can be referred to [10].

Once determined the parameters of the equivalent circuit representation, the I-V characteristic can be drawn and the generated power calculated accordingly, assuming that the PV module always operates at the maximum power point.

In this paper, on the basis of the weather forecasts provided at 11 a.m. of the previous day, the three-parameter (from datasheet) and the five-parameter (fitted from I-V curve) models are employed to perform the PV output power forecast. Previous work [10] demonstrated that the combination of three-parameter and five-parameter with datasheet and experimental I-V measurements guarantees the best results in terms of accuracy and computational efforts.

3. The stochastic artificial neural network model

Artificial Neural Networks (ANN) are employed in the field of the artificial intelligence with the aim of reproducing activities typical of the human brain, among which the perception of images and patterns, the comprehension of speech and the motor/sense coordination.

The structure of an ANN [22] imitates the biological neural networks typical of the nervous system, composed of billions of neurons. The artificial neuron, the basic unit of this network, has typically many inputs and only one output. Every input has a certain associated weight, which gives the conductivity of the corresponding input channel. The weighted total of the inputs of a neuron determines the neuron activation.

The ANN research activity mainly focuses on the learning phase which consists of the input set definition for its training. Once the training set is defined, the output processed by the network should be compared with the actual value and the model should be adjusted through selected weights to keep the error below a certain fixed threshold. For this "learning phase" there are various mechanisms which can be applied, each one characterized by certain mathematical rules regarding the weights adjustment. The Error Back Propagating (EBP) method suggests to calculate the error made by a neuron of the last hidden layer by propagating backwards the error calculated on the output neurons connected with such neuron (the same procedure is repeated for all the neurons of the penultimate hidden layer, and so on).

In this work, in order to perform the one day-ahead forecast, historical data are provided to the Neural Network Toolbox™ [3] in Matlab software.

The output of the neuron is calculated applying the activation function (also called "transfer function") to the weighted amount of the inputs (also called "net").

In this work layered neural networks are applied with double hidden layers and we refer to an "*x*-*y*" ANN (with *x* and *y* integers) as to a multilayer perceptron (MLP) with *x* neurons in the first hidden layer and *y* neurons in the second hidden layer. The correct functioning of a neural network depends first of all on the architecture of the network [5] (determination of the layers number and of the neurons number in every layer) and on the choice of the activation function.

This analysis is consistent with previous works where double layered MLP have been adopted with specific features which are the result of an accurate sensitivity analysis [11,16]. Therefore in this study, a MLP with 11 and 5 neurons respectively in the two hidden layers and the Levenberg-Marquardt training algorithm are employed. As a non-linear output function of the network neurons, we usually adopt a tan-sigmoid function which is derivable (requirement to be utilized in this training method) and produces values between 0 and 1.

In this paper a particular hybrid forecasting method (PHANN – Physical Hybrid Artificial Neural Network) is adopted as it is demonstrated that the outcome of the forecast is improved (A. [9,10]. Subsequently, we also inspected training data set composition as it is a key aspect in stochastic approaches [2,12,15,37].

4. The Physical Hybrid Artificial Neural Network model

Hybrid methods are any combination between the two groups of models (stochastic and deterministic) to overcome their weaknesses. Fig. 2 shows the PHANN which has been employed in this work. This method, which mixes the physical Clear Sky Solar Radiation algorithm (CSRM) by Hottel [18] and the stochastic ANN method, is able to improve the forecast of the PV output power.

Once the network is trained with the historical data (both of the PV plant power production and the weather forecast) by providing the same weather parameters, which are estimated 24 h in advance, it is able to give the hourly profile of the power output from the PV plant.

Furthermore, as this is a strongly stochastic method, the outputs (which are called "trials") of N networks in parallel are averaged in an "ensemble forecast". It is already proved [16] that the ensemble forecast provides better results and that the absolute hourly errors, calculated in the whole period, are lower than those from the single trials. Therefore, in this work 10 trials in the single day "ensemble forecast" are performed.

As one of the main objective of this work is to analyze the role of the training set in the PHANN output, two different setting

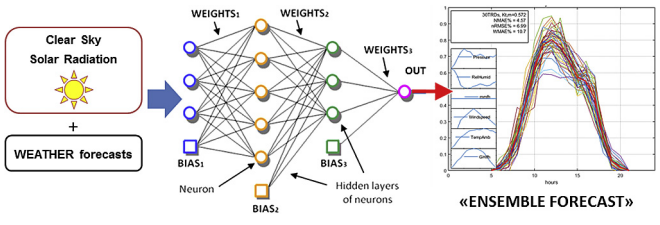

Fig. 2. PHANN model.

approaches of this crucial step have been examined in depth. Hence the results of the forecasts by these approaches have been reviewed and consequently compared.

In the proper Neural Network nomenclature they use to refer to the "test set" as to the group of samples which have to be forecasted after the training and the validation step.

The "training set" is the group of samples employed in the process of identifying the weights which minimize the error in the forecast performed on the same samples. Instead, the validation step proves the goodness of the training step weights on additional samples (the "validation set") which have not been previously included in the training set. The purpose of this step is to test on a new dataset the generalization capability of the neural network. In our case, as we are performing the one day-ahead forecast of a PV plant production in a specific "day X_d", the test set consists of 24 hourly samples; while both different dataset sizes and different approaches in the composition of the training and validation sets have been adopted.

4.1. Training method A — moving window with hourly samples randomly drawn from the previous K_d days

The first approach consists of a "moving window", namely to include in the training set the data of K_d days before the forecasted one. This rigid training time frame of K_d days is moving forward for all the N_d+1 available days in the dataset. For example, if the day after the forecasted day (X_d + 1 day) is going to be forecasted, the previous K_d days are grouped in the training and validation datasets, including also the historical data belonging to the former "day X_d" too.

Finally, the 90% of these hourly samples are randomly deployed in the training while the remaining 10% of the previous K_d days are randomly used for the validation. We refer with the expression of "equivalent day" to a generic group of 24 random hourly samples.

As a time limited database covering nearly one year of the PV output power hourly samples has been considered, a further clarification is necessary.

In fact, as it is shown in Fig. 3, a circular trend-wise training of the PHANN has been adopted. For example, in this way to forecast the PV power production of the last day, corresponding to the 14th December, the PHANN is trained with the previous available days in the dataset, which are not always following the chronological continuity because of the monitoring system or sensors failures resulting in missing data. This aspect reveals two important issues in reality:

1. Historical data are not always available and not chronologically continuous,
2. When available, it is reasonable to adopt historical data either of the same season even if they belong to a different year, or of an adjacent season if belonging to the same year.

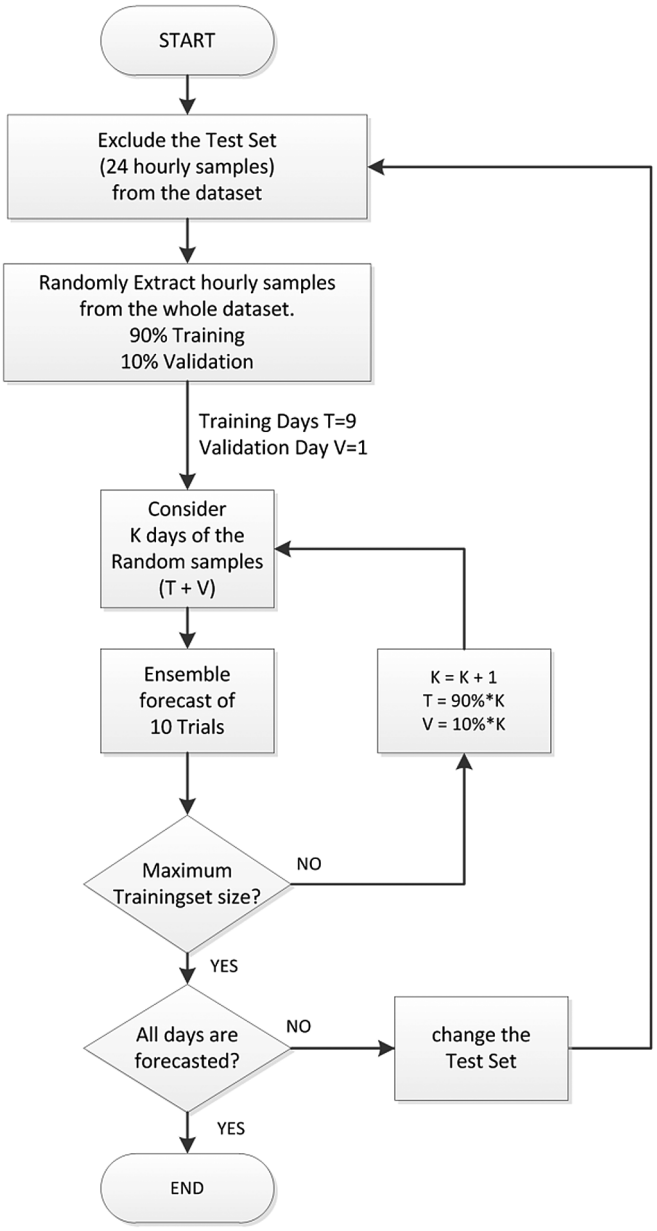

Fig. 4. Flow chart of the training method B employed.

For these reasons, when the forecasting day is the first one in the dataset (8th February), the corresponding training set, consisting in the previous N_d days, will be composed of the last days available in the same dataset which, in this case, ends on 14th December.

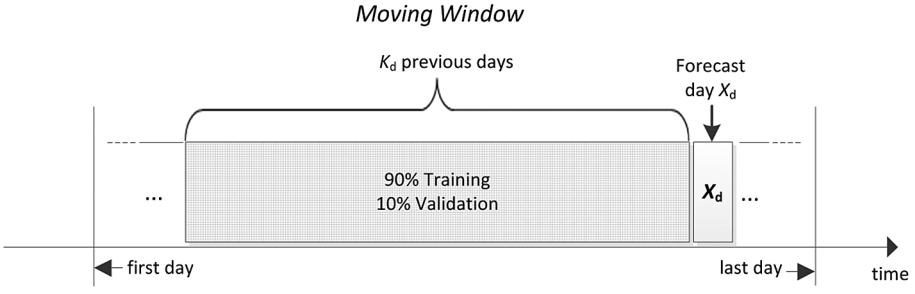

Fig. 3. "Moving window" training employed in the method A for every day forecasted in the dataset.

Table 1
Main features of the adopted training approaches.

Training method	A	B
Short explanation	"Moving window" training set with random hourly samples taken from the previous days	Growing training set with random hourly samples, taken from the whole database. Samples are kept constant in each trial
Training and Validation set size	From 10 to 215 days, with a rate of 5 days: - Training = 90% - Validation = 10%	From 10 to 215 days, with a rate of 5 days: - Training = 90% (fixed hourly samples) - Validation = 10% (fixed hourly samples)

4.2. Training method B — hourly samples randomly taken by the whole dataset

The second training set approach requires hourly samples randomly drawn from the whole dataset, in order to constitute a training set of N equivalent days. In fact the picked hourly samples easily belong to different days. These hourly samples are kept equal during the training of each trial, consequently providing an "ensemble forecast" derived from a common training set. Afterwards, the same day forecast is performed and the length of the training set is increased by adding further random hourly samples in the rate of 1 day each forecast until all the samples in the database are used up. In this case, the 90% of the randomly chosen hourly samples are employed for the training of the network and the remaining 10% are used for the validation (see Fig. 4). These shares of training and validation samples are chosen after a previous work [16,29] in which they were more effective. The main features of these methods are listed in Table 1 hereunder.

5. Case study and forecasting methods assessment

The different models were compared with experimental data recorded at the laboratory SolarTech^{Lab} [34], Politecnico di Milano, the coordinates of which are latitude 45° 30′ 10.588″ N and longitude 9° 9′ 23.677″ E. During the year 2014, the output power of a single PV module with the following characteristics was recorded:

- PV technology: Silicon mono crystalline
- Rated power: 245 Wp
- Net capacity[1] of the PV module (the maximum measured DC output power): 223 Wp
- Azimuth: −6°30' (assuming 0° as South direction and counting clockwise)
- Solar panel tilt angle (β): 30°
- PV module label: "F5-50"

The monitoring activity lasted from 8th February to 14th December 2014, but the employable data, without interruptions and discontinuities amount to 216 days. These hourly samples, night hours included, are used as the database for the forecasting methods comparison.

The PV module is linked to the electric grid by a micro inverter, guaranteeing the optimization of the production. Its operating parameters, DC power included, are transmitted to a computer desktop for storage using a ZigBee protocol wireless connection, in real time.

The weather forecasts employed are delivered by a weather service each day at 11 a.m. of the day before the forecasted one. The historical hourly database of these parameters is used to train the neural network and include the following parameters: ambient temperature (°C), GHI (W/m^2), G_{POA} (W/m^2), wind speed (m/s), wind direction (°), pressure (hPa), precipitation (mm), Cloud Cover (%) and Cloud Type (Low/Medium/High).

In addition to these parameters, in order to train the PHANN method, also the local time (hh:mm) of the day and the CSRM (W/m^2) are provided. As regards the physical models, only the G_{POA} solar irradiance and the ambient temperature are used for the forecasting.

In order to assess the forecasting methods accuracy, some of the most common error indexes in literature [7,26,35] have been considered in this work.

The common error definition for the assessment is the hourly error e_h, which is defined as:

$$e_h = P_{m,h} - P_{p,h} \tag{10}$$

where $P_{m,h}$ is the average actual power in the hour and $P_{p,h}$ is the prediction provided by one of the forecasting methods. Starting from the hourly error definition, the other error indexes adopted for the assessment can be derived as follows:

- Mean biased error (*MBE*)

$$MBE = \frac{1}{N}\sum_{i=1}^{N}\left(P_{m,h} - P_{p,h}\right) \tag{11}$$

- Mean absolute error (*MAE*)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|P_{m,h} - P_{p,h}\right| \tag{12}$$

- Normalized mean absolute error $NMAE_{\%}$ which is *MAE* based on net capacity of the plant *C*. In this analysis *C* is the maximum DC output power measured in the whole period and is expressed in Watt:

$$NMAE_{\%} = \frac{1}{N \cdot C}\sum_{i=1}^{N}\left|P_{m,h} - P_{p,h}\right| \cdot 100 \tag{13}$$

In all of these definitions, *N* is the number of hours considered in the evaluated period (i.e. 24 h in a daily error basis calculation).

As it was already mentioned in Ref. [16] the abovementioned error indexes are strongly correlated to each other and a single index can be equally adopted to describe the global trend of the others (even if in a different scale). For this reason, in this work, we adopted NMAE as the main error index for evaluating the forecast.

Finally, the coefficient of determination R^2 is also calculated [7]. This coefficient represents the accuracy of the forecasting models compared with the trends in measured values by dividing the relevant statistical variances σ^2:

[1] The PV plant "net capacity" is "gross capacity" which the parasitic loads have been subtracted from Ref. [28].

$$R^2 = 1 - \frac{\sigma^2\left(P_{p,h} - P_{m,h}\right)}{\sigma^2\left(P_{m,h}\right)} \tag{15}$$

For perfect forecasting, $R^2 = 1$.

6. Results

This section presents the results of the forecasts performed by the different methods and approaches previously exposed.

Table 2 shows the mean errors of the deterministic methods, calculated over the whole period (216 days). The indexes calculation considers all the 24 h in a day. Table 2 also gives the maximum absolute error occurred on 11th October 2014 at noon. In general, the three-parameters model shows slightly higher accuracy, but the trend of the two models is similar.

In Fig. 5 on the left, the global horizontal irradiance provided by the weather service 24 h in advance is shown together with the actual values. On the right, in the same figure the DC output power measured in blue and forecasted in red are respectively depicted. The maximum absolute error is marked as a dashed green line around 12 p.m.

The direct correlation between the irradiance and the PV output power, both in the forecasted and in the measured values, highlights how the inaccuracy in the weather forecasts affects the deterministic model prediction.

In Table 3, referring to the whole database, the MAE and the MBE values for the global horizontal irradiance (GHI) provided by the weather service and for the deterministic methods with the relative standard deviations are reported. What is more remarkable is the strong correlation between the two deterministic methods and the weather forecasts: there is a substantial overestimation of the forecasts compared to with the measured values of PV power and irradiance. Finally, the restrained standard deviation of about 35–36 W, particularly referred to the MBE, suggests that the hourly errors are generally close to their mean values.

In a previous work [9,10], the NMAE calculated on daylight hours of the same deterministic models based on actual weather measurements was below 5%.

Table 2
Maximum absolute error and mean error indicators of the deterministic forecasting methods.

Equivalent model	$AE_{max}(W)$	$MBE(W)$	$MAE(W)$	$\overline{NMAE}_{\%}$
3 parameters	199.5	−12.6	19.1	8.56
5 parameters	200	−14.3	20.2	9.05

Table 3
R^2, MAE and MBE with the relative standard deviation for the 3 and 5 parameters model and the global horizontal irradiance provided by the weather service 24 h in advance. Where P_{ref} equals 245 W for PV module and it equals 1000 W/m^2 for the GHI.

	3 parameters	5 parameters	Weather service GHI 24 h
R^2	0.658	0.645	0.776
MAE(W)	19.1	20.2	64
$MAE/P_{ref}\cdot 100$	7.80	8.24	6.4
Std Dev absolute error: $\sigma(\lvert e_h\rvert)$	32.7	33.6	103.5
$\sigma(\lvert e_h\rvert)/P_{ref}\cdot 100$	13.3	13.7	10.35
MBE(W)	−12.6	−14.3	−34.7
Std Dev error: $\sigma(e_h)$	35.8	36.5	116.7

6.1. Method A - moving window with limited random hourly samples

Method A is adopted for all the forecasted 216 days of the whole dataset. Besides two different sizes in the "training plus validation" datasets are considered: a small one of 10 days and a large one of 215 days. These different sized datasets are separately provided to the same neural network of 11 and 5 neurons in the hidden layers. Finally, 10 days have been selected as the smallest size of the training set in order to guarantee at least 1 equivalent day of validation after 9 equivalent days of training.

After the forecast has been performed with these settings, the results are compared with the ones obtained with the deterministic models (see Fig. 6).

Fig. 6 shows the NMAE% calculated for every forecasted day. On the left the errors of the output provided by the PHANN trained with 10 equivalent days are compared with the results obtained with the three-parameters model output.

In this case, the PHANN with the smaller training set (10 days) has lower daily errors in 76% of the forecasted days. Furthermore, Fig. 6 shows on the right, the comparison between the same layered PHANN trained with 215 days and the three-parameters deterministic model is shown. Also in this case almost 77% of the days shows a more accurate power forecast by the hybrid method if compared with the deterministic model.

In order to better understand global trends, Fig. 7 on the left shows again the daily NMAE% of the PHANN trained with different training set sizes and the relative average values calculated over the entire period. The comparison here depicted includes the two different time frames of 10 and 215 equivalent days before the forecasted one employed for the training setting. From this picture we can see that the larger training set provides slightly better

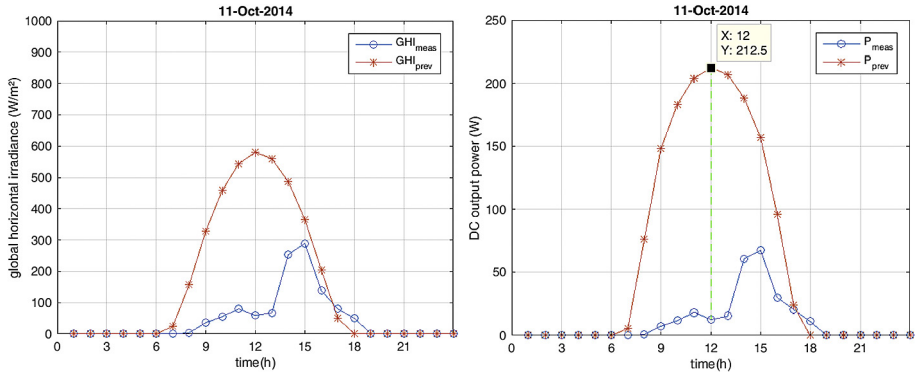

Fig. 5. Global horizontal irradiance on the left and DC output power on the right. Measured values are in blue, the forecasted values with the 5 parameter model in red. The maximum absolute error is the green dashed line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

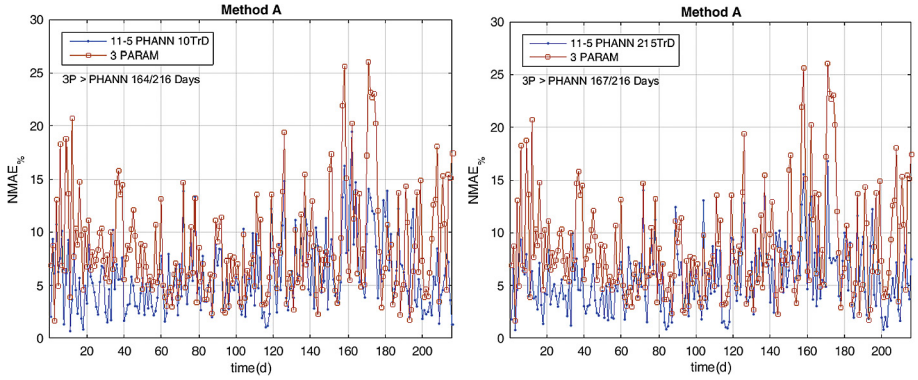

Fig. 6. Daily NMAE% of the 11-5 PHANN trained with 10 days (on the left) and with 215 days (on the right) compared with the five-parameters model.

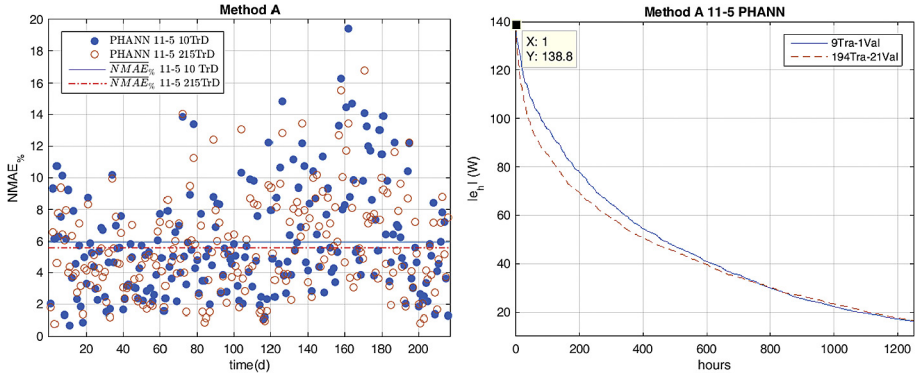

Fig. 7. On the left 11-5 PHANN trained by different dataset sizes daily NMAE% comparison and on the right sorted absolute hourly errors comparison committed in the whole period of 216 days.

results, but the two averages are pretty close.

In the same figure on the right, the absolute hourly errors $|e_h|$ calculated over the whole period, are shown in descending order of magnitude.

With respect to these hourly errors, the maximum absolute value of 138.8 W belongs to the PHANN which is trained with the largest set (215 days) and occurred during 19th September 2014 at 2 p.m. as it is shown in Fig. 8 on the right by a vertical yellow line. Instead, the PHANN trained with the smallest training set (10 days) committed the highest error on the same date (19th September 2014) at 1 p.m. with the similar value of 137 W. It can be highlighted that, only by looking at Fig. 7 on the right, the two maxima are very close, therefore it's impossible to claim which is the best training set size only on this basis. Nevertheless, it must be noted that the absolute hourly errors are generally lower for the PHANN with the largest training set, depicted with the dashed red line, on the overall period. If we consider the date in which the highest hourly error occurred, in Fig. 8 on the left, the global irradiance provided by the weather service for that day is depicted in red while the actual global irradiance is in blue.

If we compare these trends with the forecasted PV power output by the PHANN in the same figure on the right, it can be noted that the PHANN forecast with only 10 days of training and validation (the dashed red line) is closer to the forecasted irradiance G_p trend

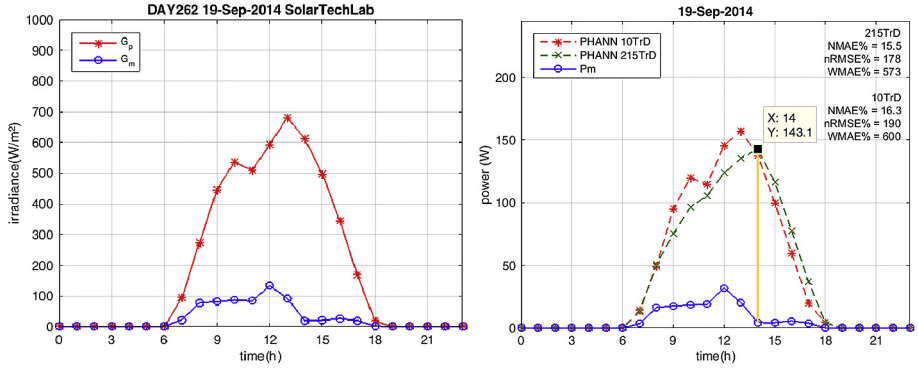

Fig. 8. Irradiance provided by the weather service G_p, and the actual one G_m measured on the plant during 19th September 2014 (left) and power output forecasted by the PHANN trained with 10 days in red, with 215 days in green and P_m measured in blue (right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

on the left (the solid red line). This sounds reasonable as the solar irradiance is the key parameter in a photovoltaic plant power production. Instead, with 215 days of training and validation, the forecasted output (the dashed green line) is more accurate. In fact, the daily error indexes in this case are lower than those of the 10 equivalent days of training and validation.

Therefore, the larger daily errors committed in "day 262" by the PHANN with 10 days of training are likely attributable to the weather service inaccuracy. In the case of the PHANN trained with 10 days, the inaccuracy of the PV output power forecast might be related to the PHANN which has not been properly trained.

As the maximum entity of the hourly error barely gives any clue of definitive evidence on which is the best training set size approach, the worst and the best cases of the daily $NMAE_{\%}$ should be considered.

The worst case for the 10 equivalent days of training occurred on 23rd September 2014 with the highest NMAE% daily value of 19.4 (see Fig. 9 on the left), while for the largest training set of 215 equivalent days it happened on 8th October 2014 (see the same figure below on the right).

For the PHANN trained with 10 equivalent days, the best case occurred on 19th February 2014 (see Fig. 10 on the left) and on 10th February 2014 the PHANN trained with 215 equivalent days (in the same Fig. 10 on the right) was more accurate than the other. The two small boxes within Fig. 10 show the forecasted plane of the array irradiance G_p in red and the measured one G_m in blue for the relative days.

First of all, it is strange to find such an accuracy in both of these cloudy day types. Actually, it would have been more likely to find such an accurate forecast on a sunny day rather than on a totally cloudy day. Mid-February for the given location is usually rainy or snowy as it is still winter time.

However, the irradiance forecasts employed to perform the PV power forecast in these cases are very accurate and the largest training set size provided comparable results to the smaller training set. This conclusion is strengthened by Table 4 showing the mean figures of the two training set sizes here adopted. The difference is not essential to discriminate which training set size is successful for an accurate forecast with Method A.

Finally, by looking at the hourly errors boxplots in Fig. 11, it is evident how the central daily hours (from 11 a.m. to 3 p.m.) are the most critical ones showing the lowest accuracy. The only difference related to the two set sizes is the value of the hourly absolute error which is slightly lower in the case of the PHANN trained with 215 days. Again we can assume that this trend is mainly affected by the inaccuracies of the weather forecasts provided as an input to the

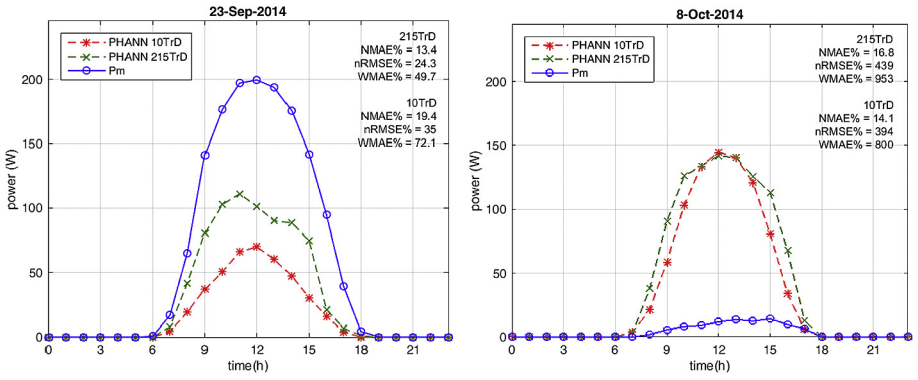

Fig. 9. Worst case with the maximum daily NMAE% error for the PHANN trained with 10 days (left) and 215 days (right).

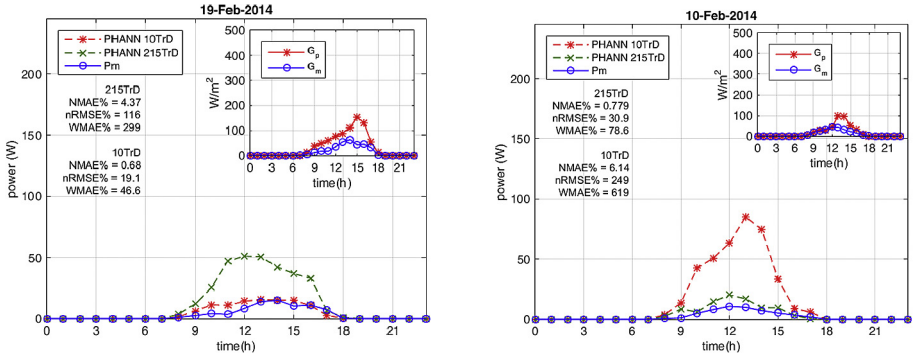

Fig. 10. Best case with the minimum daily NMAE% error for the PHANN trained with the with 10 days (on the left) and 215 days (on the right). In the small boxes the irradiance provided by the weather service 24 h in advance is red, while the actual values are blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4
Mean errors calculated over the whole period of 216 days according to the different training set size for "Method A" training.

Number of Neurons	(Number of days) Training/Validation	$AE_{max}(W)$	$MBE(W)$	$MAE(W)$	$\overline{NMAE}_{\%}$
11−5	(10) 9/1	137	0.9	13.25	5,94
	(215) 194/21	138.8	0.3	12.46	5,58

PHANN, which is a slightly mitigated with a large training set.

6.2. Method B - hourly samples randomly taken from the whole dataset

In this approach, the hourly samples are randomly drawn from the whole database of 216 days and grouped in a dataset kept fixed in every trial. At the beginning the length of the dataset accounts 10 equivalent days. After performing the ensemble forecast for one given day, the training set is increased by adding more hourly samples, which are again randomly drawn from the whole dataset. In this way the new data are added to the previous ones enlarging the training set size with 5 equivalent days.

Fig. 12 shows the results of the forecasts performed with this training method: on the left the shortest training set size of 10 equivalent days (in blue) and on the right the longest one of 215 days. Furthermore, the comparison with the daily errors of the three-parameter physical model are shown in red, and the relative overall period averages $\overline{NMAE}_{\%}$ of the two methods are traced as constant blue lines. It is evident that the hybrid method generally performs better, especially when trained with the largest dataset.

In addition, all the averages of the error indicators listed in Table 5 clearly show the better behavior of the largest dataset if compared with the smallest.

Focusing on $\overline{NMAE}_{\%}$ column, it is more evident that, in this case, the 11-5 PHANN forecast improves by increasing the number of training samples. Note that, if we compare these average values with the relative column in Table 4, which is related to Method A, numbers are decreasing while in the previous case, they are nearly constant. We will discuss more about this trend in the next paragraph.

7. Discussion and comparison of results

The main purpose of this analysis is to determine the most accurate forecast model and the most effective technique in the PHANN training.

Fig. 14 on the left shows the average of the daily error definition $\overline{NMAE}_{\%}$ calculated over the entire period of the 216 days forecasted with the analyzed methods, according to the number of days employed in the training (the deterministic methods are depicted by two horizontal lines).

Furthermore it is evident that PHANN always gives the lowest errors after 10 days of the PV plant operation.

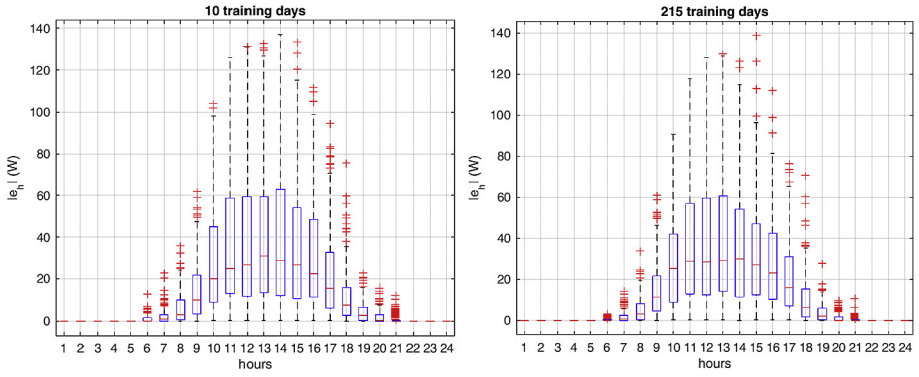

Fig. 11. Hourly absolute errors boxplots, according to the hour of the days for the 11-5 PHANN, trained with different training set size: 10 equivalent days on the left and 215 equivalent days on the right.

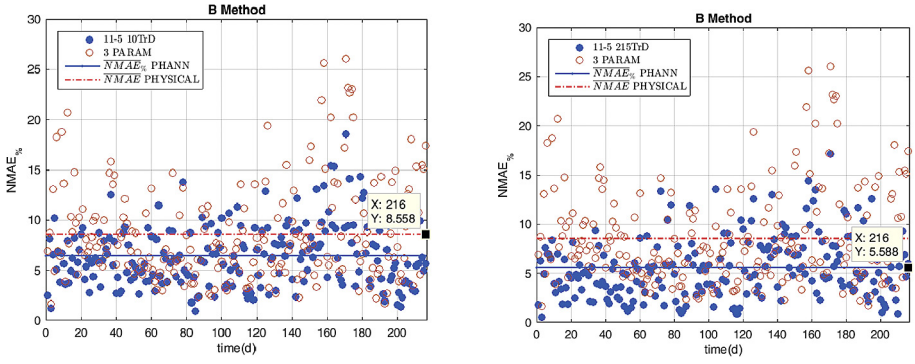

Fig. 12. Comparison of the daily NMAE% between the deterministic (empty red circles) and the PHANN (solid blue circles) forecasts. PHANN was trained with 10 equivalent days (on the left) and 215 equivalent days (on the right). The average values are also marked. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Mean errors calculated over the whole period of 216 days according to the different training set size for the training method B.

Number of Neurons	(Number of days) Training/Validation	$AE_{\max}(W)$	$MBE(W)$	$MAE(W)$	$\overline{NMAE}_{\%}$
11−5	(10) 9/1	140.1	0.6	14.3	6.40
	(215) 194/21	141.3	0.4	12.5	5.59

Thus the most interesting trend is provided by the comparison between different PHANN training approaches. By looking at Fig. 14 on the left it is evident that when only few days are available, for example when the PV plant has just started to operate, Method A is more effective. In fact, the proximity of the training samples with the forecasted day is stronger than in using Method B, which is randomly drawing the samples from a larger period. In this case the neural network with Method B is undertrained, if compared with Method A, but it has a significantly decreasing trend with the increasing of the days in the training set.

This is true until 50 days, but after 65 with Method B the PHANN is more capable of generalization and is generally committing a lower error, while Method A is substantially constant with some fluctuations around 5.9–6.

By reaching the maximum available number of training samples the two methods obviously converge to the same values, being finally Method A equivalent to B, both picking randomly from the whole available dataset of 215 days.

Another significant difference is evident by considering the probability densities of the maximum absolute errors committed daily by the different forecasting methods (see Fig. 14 on the right) despite of their magnitude.

The maxima calculated every day are statistically centered around 1 p.m. winter time with a peak in the probability density of nearly 14%. The deterministic methods are flatter and involve the early morning hours (from 5 a.m.) and late evening (to 9 p.m.).

In the case of the PHANN methods the probability densities of the maximum hourly errors are narrower and they are centered again in the mid-day hours with peaks near 20%. This means that within 45 days among the 216 forecasted, the daily maximum absolute error occurred at 1 p.m.

Furthermore, considering the longest training set size (215 days) the narrowest bell-shaped curve is obtained. This means that such training set size provides a more accurate forecast in the tail hours of the day.

Table 6 summarizes the main error indicators calculated over the whole period. Focusing on the MBE, it can be noted how the PHANN methods record the lowest values and that their standard deviations are more restrained, if compared to the deterministic methods. The coefficient of determination R^2 confirms the $\overline{NMAE}_{\%}$ trend.

These results are totally in line with the boxplots shown in Figs. 13 and 11.

Table 6
Summary of the main error indicators calculated for each forecasting method.

Forecasting method	(# of days) Tra/Val	AE_{max} (W)	MBE (W)	Std Dev $\sigma(e_h)$ (W)	MAE (W)	Std Dev $\sigma(\lvert e_h \rvert)$ (W)	$\overline{NMAE}_{\%}$	R^2
3 PARAM	–	199.5	−12.6	35.8	19.1	32.7	8.56	0.658
5 PARAM	–	200	−14.3	36.5	20.2	33.6	9.05	0.645
PHANN A	(10) 9/1	137	0.87	27.7	13.25	23.78	5.94	0.849
	(215) 194/21	138.8	0.25	25.1	12.46	21.78	5.58	0.873
PHANN B	(10) 9/1	140.1	0.6	27.7	14.3	23.7	6.4	0.796
	(215) 194/21	141.3	0.4	25.2	12.5	21,9	5.59	0.831

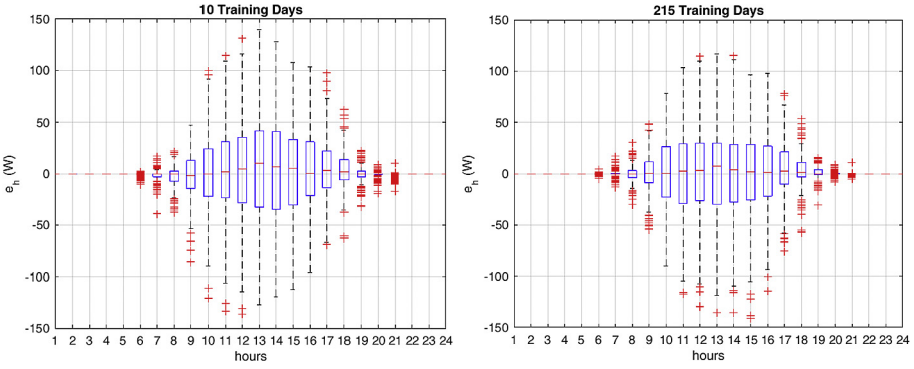

Fig. 13. Hourly error boxplots of PHANN trained by 10 days (left) and 215 days (right).

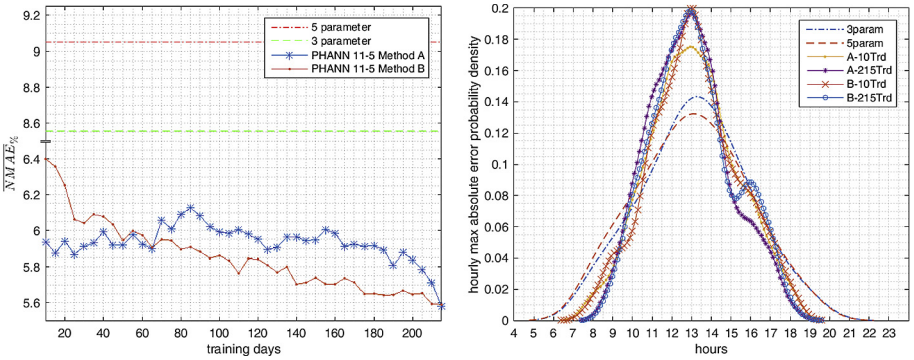

Fig. 14. $\overline{NMAE}_{\%}$ forecasting methods comparison according to the training set size (on the left) and daily maximum absolute errors probability density according to the hours (on the right).

8. Conclusions

This work provided an assessment between different forecasting methods for the day-ahead output power of the same PV module: physical-deterministic methods based on the NOCT thermal model of the PV module and a hybrid stochastic-deterministic one which merges the CSRM and ANN. Within the first category the three-parameter and the five-parameter electric circuit models were considered, whereas for the hybrid one (PHANN) different approaches in the training set composition and dimension were analyzed. In this last group, historical data recorded at SolarTech^{Lab} in Milan – Italy – in the year 2014, amounting to 216 available days, have been exploited to train the PHANN model. The daily PV power output was forecasted starting from the weather parameters supplied 24 h in advance by a weather service provider.

In the analysis carried out in this work, some of the most common error definitions were considered, but ultimately the normalized mean absolute error ($NMAE_{\%}$) was mainly adopted as it showed the same trend and a correlation with the other errors.

By comparing the mean $NMAE_{\%}$ ($\overline{NMAE}_{\%}$) calculated considering all the 216 available days forecasted with the different methods, the three-parameters method performs slightly better than the five-parameter one (8.5% vs. 9.0% respectively), on the other hand PHANN always shows higher accuracy in the range of 6%.

As regards PHANN, the training based on hourly samples randomly taken by the whole dataset (Method B) has scored better results starting from datasets of 50 days or more, while the use of a moving window with limited random hourly samples (Method A) better performed with shortest training set. In this last case a greater "proximity effect" of the training set on the forecasted day is inferable, as regards the better performances which have obtained on average. Furthermore, by increasing the training set size with any composition approach, forecasting errors are generally lowered. This result is reasonably due to the capability of the hybrid method to learn both the inaccuracies of the provided weather forecasts and the peculiarities of the PV plant (shadowing, etc.).

In conclusion, this study shows that the day-ahead output power forecast in a just commissioned PV plant should be first performed with the three-parameter model. After few days of operation (from 10 to 50 days) an accurate forecast can be performed by the PHANN model, trained with Method A. Starting from the 50th day on, the PV output power forecast could be reached by adopting Method B.

References

[1] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.J. Martinez-de-Pison, F. Antonanzas-Torres, Review of photovoltaic power forecasting, Sol. Energy (2016), http://dx.doi.org/10.1016/j.solener.2016.06.069.

[2] E.B. Baum, D. Haussler, What size net gives valid generalization? Neural Comput. 1 (1989) 151–160, http://dx.doi.org/10.1162/neco.1989.1.1.151.

[3] Beale, M.H., Hagan, M.T., Demuth, H.B., n.d. Neural Network Toolbox: User's Guide.

[4] A.N. Celik, Artificial neural network modelling and experimental verification of the operating current of mono-crystalline photovoltaic modules, Sol. Energy 85 (2011) 2507–2517, http://dx.doi.org/10.1016/j.solener.2011.07.009.

[5] S.H. Chen, A.J. Jakeman, J.P. Norton, Artificial Intelligence techniques: an introduction to their use for modelling environmental systems 78 (2008) 379–400, http://dx.doi.org/10.1016/j.matcom.2008.01.028.

[6] A. Chopde, D. Magare, M. Patil, R. Gupta, O.S. Sastry, Parameter extraction for dynamic PV thermal model using particle swarm optimization, Appl. Therm. Eng. 100 (2016) 508–517, http://dx.doi.org/10.1016/j.applthermaleng.2016.01.164.

[7] C.F.M. Coimbra, J. Kleissl, R. Marquez, Solar Energy Forecasting and Resource Assessment, Elsevier, 2013.

[8] W. De Soto, S.A. Klein, W.A. Beckman, Improvement and validation of a model for photovoltaic array performance, Sol. Energy 80 (2006) 78–88, http://dx.doi.org/10.1016/j.solener.2005.06.010.

[9] A. Dolara, F. Grimaccia, S. Leva, M. Mussetta, E. Ogliari, A physical hybrid artificial neural network for short term forecasting of PV plant power output, Energies 8 (2015a), http://dx.doi.org/10.3390/en8021138.

[10] A. Dolara, S. Leva, G. Manzolini, Comparison of different physical models for PV power output prediction, Sol. Energy 119 (2015b) 83–99, http://dx.doi.org/10.1016/j.solener.2015.06.017.

[11] A. Dolara, S. Leva, M. Mussetta, E. Ogliari, PV hourly day-ahead power forecasting in a micro grid context, in: EEEIC 2016-International Conference on Environment and Electrical Engineering, 2016, http://dx.doi.org/10.1109/EEEIC.2016.7555636.

[12] F.U. Dowla, L.L. Rogers, Solving Problems in Environmental Engineering and Geosciences with Artificial Neural Networks, Mit Press, 1995.

[13] a. Durgadevi, S. Arulselvi, S.P. Natarajan, Photovoltaic modeling and its characteristics, in: 2011 Int. Conf. Emerg. Trends Electr. Comput. Technol, 2011, pp. 469–475, http://dx.doi.org/10.1109/ICETECT.2011.5760162.

[14] F. a. Farret, J.M. Lenz, J.G. Trapp, New methodology to determinate photovoltaic parameters of solar panels, in: XI Brazilian Power Electron. Conf, 2011, pp. 275–279, http://dx.doi.org/10.1109/COBEP.2011.6085281.

[15] G.M. Foody, M.K. Arora, An evaluation of some factors affecting the accuracy of classification by an artificial neural network, Int. J. Remote Sens. 18 (1997) 799–810, http://dx.doi.org/10.1080/014311697218764.

[16] A. Gandelli, F. Grimaccia, S. Leva, M. Mussetta, E. Ogliari, Hybrid model analysis and validation for PV energy production forecasting, in: Proceedings of the International Joint Conference on Neural Networks, 2014, http://dx.doi.org/10.1109/IJCNN.2014.6889786.

[17] F. Hocaoglu, O. Gerek, M. Kurban, Hourly solar radiation forecasting using optimal coefficient 2-D linear filters and feed-forward neural networks, Sol. energy 8 (2008) 714–726.

[18] H.C. Hottel, A simple model for estimating the transmitance of direct solar radiation through clear atmospheres, Sol. energy 18 (1976) 129–134.

[19] M. Houabes, PV cell 5P modeling with shunt resistance correction, in: EFEEA 10th International Symposium on Environment Friendly Energies in Electrical Applications, 2010, pp. 1–5.

[21] E. Izgi, A. Oztopal, B. Yerli, M. Kaymak, A. Sahin, Short-mid-term solar power prediction by using artificial neural networks, Sol. Energy 86 (2012) 725–733.

[22] B. Krose, P. Van Der Smagt, An Introduction to Neural Networks, eight. ed., University of Amsterdam, 1996.

[24] A. Luque, S. Hegedus, Handbook of Photovoltaic Science and Engineering, 2013.

[25] A. Mellit, A. Massi Pavan, A 24-h forecast of solar irradiance using artificial neural network: application for performance prediction of a grid-connected PV plant at Trieste, Italy, Sol. Energy 84 (2010) 807–821.

[26] C. Monteiro, A. Fernandez-Jimenez, I. Ramirez-Rosado, A. Munoz-Jimenez, P. Lara-Santillan, Short-term forecasting models for photovoltaic plants: analytical versus soft-computing techniques, Math. Probl. Eng. 6 (5) (2013) 2624–2643. Article ID, 9.

[27] J. Nelson, The Physics of Solar Cells, Imperial college press, London, 2003.

[28] NREL, Solar Energy and Capacity Value, 2013.

[29] E. Ogliari, A. Gandelli, F. Grimaccia, S. Leva, M. Mussetta, Neural forecasting of the day-ahead hourly power curve of a photovoltaic plant, in: 2016 Int. Jt. Conf. Neural Networks, 2016, pp. 654–659, http://dx.doi.org/10.1109/IJCNN.2016.7727262.

[30] E. Ogliari, F. Grimaccia, S. Leva, M. Mussetta, Hybrid predictive models for accurate forecasting in PV systems, Energies 6 (2013), http://dx.doi.org/10.3390/en6041918.

[31] Renewable capacity statistics 2016, n.d.

[32] E. Rodrigues, R. Melício, V. Mendes, J. Catalão, Simulation of a solar cell considering single-diode equivalent circuit model, in: International Conference on Renewable Energies and Power Quality. Spain, 2011.

[34] SolarTech Lab [WWW Document], 2013.

[35] R. Ulbricht, F. Fisher, W. Lehner, H. Donker, First steps towards a systematical optimized strategy for solar energy supply forecasting, in: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Prague, September 23th to 27 Th, 2013.

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

[37] G. Zhang, B. Eddy Patuwo, Y. Hu, Forecasting with artificial neural networks: the state of the art, Int. J. Forecast 14 (1998) 35–62, http://dx.doi.org/10.1016/S0169-2070(97)00044-7.