1 Analysis and validation of 24 hours ahead neural 2 network forecasting of photovoltaic output power

3 S. Leva, A. Dolara, F. Grimaccia, M. Mussetta, E. Ogliari

4 *Dipartimento di Energia, Politecnico di Milano*
5 *Piazza Leonardo da Vinci 32, 20133, Milano, Italy*
6 *e-mail : sonia.leva@polimi.it*

7 **Abstract**

8 In this paper an artificial neural network for photovoltaic plant energy fore-
9 casting is proposed and analyzed in term of its sensitivity with respect to the
10 input data sets.

11 Furthermore, the accuracy of the method has been studied as a function
12 of the training data sets and error definitions. The analysis is based on exper-
13 imental activities carried out on a real photovoltaic power plant accompanied
14 by clear sky model.

15 In particular, this paper deals with the hourly energy prediction for all
16 the daylight hours of the following day, based on 48 hours ahead weather
17 forecast. This is very important due to the predictive features requested
18 by smart grid application: renewable energy sources planning, in particular
19 storage system sizing, and market of energy.

20 *Keywords:* Artificial neural network, Energy forecasting, Photovoltaic
21 system

22 **1. Introduction**

23 The electricity produced by renewable energy sources (RES) is constantly
24 world-wide increasing thanks to government policies and technical progress.
25 Europe has experienced one of the largest growths: in the last five years the
26 electricity generation by RES, and in particular by photovoltaic (PV) and
27 wind plants, is doubled. However, the RES energy productions are charac-
28 terized by fluctuating output, because they are influenced by meteorological
29 conditions.

Challenges of controlling and maintaining energy from inherently intermittent sources in grid-connected systems involve many features: efficiency, reliability, safety, stability of the grid and ability to forecast energy production. In particular, PV and wind power forecasting, as an estimation of the expected power production, is crucial to help the grid operators to better manage the electric balance between power demand and supply, and to improve the penetration of distributed renewable energy sources. Furthermore, in countries with a day-ahead electricity market, large power plants based on RES can act, as any other electricity producer, providing power generation sale offers (bids) to the market. In electricity markets, when a power producer does not follow the scheduled bid it will be penalized with retributions lower than those established in the market for those hours with deviation between the electric energy actually produced and that presented in the bid [15, 24].

These technical and economic reasons have driven the development of power forecasting models for wind farms and relatively large grid-connected PV plants, with the aim to predict the hourly output power up to 24 hours ahead and even more.

In recent years several power forecasting models related to PV plants have been published. The existing solutions can be classified into the categories of physical, statistical and hybrid methods. Some of these models were at first oriented to obtain solar radiation predictions [14, 18] while other works present models specifically dedicated to the forecasting of the hourly power output from PV plants [12, 20]. Nowadays the most applied techniques to model the stochastic nature of solar irradiance at the ground level and thus the power output of PV installations are the statistical methods; in particular regression methods are often employed to describe complex non-linear atmospheric phenomena for few-hours ahead forecast and specific soft-computing techniques based on artificial neural network (ANN) are used for few-hours power output forecast [17]. Some other papers use physical methods [15, 23, 21]. Some papers report the comparison of the results obtained with different models based on two or more forecasting techniques [17, 18, 15]. Nowadays the most important forecasting horizon is 24 hours of the next days. Only a few papers describe forecasting models used to predict the daily irradiance or directly energy production of the PV plant for all the daylight hours of the following day [24, 15, 25].

In order to define the accuracy of the prediction, some error indexes are introduced to evaluate the performances of the forecasting models. Some of

Nomenclature

G	global solar radiation on module surface (W/m^2)
P	PV output power (W)
REL	Reliability Coefficient
e	error (W)
$NMAE$	Normalized Mean Absolute Error
$WMAE$	Weighted Mean Absolute Error
$nRNSE$	Normalized root mean square error
C	Rated power (W)
k	Generic time sample
$CSRM$	Clear Sky Radiation Model
STC	Standard Test Conditions
m	measured
p	predicted
$1/4h$	quarter of hour
h	hour sample
N	total number of considered samples (daylight hours)
i	single trial of ensemble method of ANN
n	number of trials of ensemble method of ANN

these definitions come from statistics while others originate from regulatory authority for market issues [1].

This paper uses a model based on ANN accompanied by clear sky model for input data validation for next-day energy forecasting of a PV plant with the aim to evaluate its sensitivity. It has been assessed by changing the size of the training data sets in input, the number of iterations and launching single-/multiple-runs. Different error definitions are also calculated and analyzed in order to evaluate the results. The analysis is based on experimental activities carried out by a real PV power plant.

The paper is organized as follows: in Section 2 a brief review of the hourly energy production forecasting methods is presented. In Section 3, the applied method is described giving emphasis to the pre-processing data analysis. In Section 4 some error indexes are defined in order to evaluate the performances of the forecasting models. In Section 5 the hourly energy prediction for all the daylight hours of the following day of a real plant is presented, in terms of hourly error, normalized mean absolute error, weighted mean absolute error, and normalized root mean square error. In the end some conclusions are stated.

2. Energy forecasting methods

RES energy production forecasting methods are commonly divided in different categories: physic, stochastic and hybrid. An analysis of the state-of-the-art approaches is proposed in [23]. Physical models are based on mathematical equations which describe the ability of PV systems to convert the introduced meteorological resources into electrical power [14], [15]. These models can be very simple, if based only to the global solar radiation, or more complicated if they include additional parameters. As a matter of fact, it is not easy to predict PV module energy production since it depends on several parameters. The conversion process is affected by solar radiation, cell temperature, presence of shadow [8] and the load resistance. Moreover, information provided by manufacturer is usually limited and only at nominal operating conditions. The major disadvantage of these models is that they have to be designed specifically for a particular plant and location.

Statistical methods are based on the concept of persistence or stochastic time series. Regression methods often employed to describe complex nonlinear atmospheric phenomena include the Auto-Regressive Moving Averages (ARMAs) method, as well as its variations, such as the Auto-Regressive Integrated Moving Averages (ARIMAs) method [18, 17]. The performance of these models is very good for few-minutes to few-hours ahead forecasts [18, 3]. Nonlinear methods, such as the Takagi-Sugeno (TS) fuzzy model [11] and wavelet-based methods [5], have been shown superior to linear models.

Nowadays the most common way to forecast the future values of a time series is the use of machine learning methods [6]. Reviewed literature shows that ANN methods have been successfully applied for forecasts of fluctuating energy supply. These methods learn to recognize patterns in data using training data sets. This is the main drawback: historical data about weather forecast and the real power production as well as environmental quantities are necessary to train the ANN and start the forecast of energy production by RES. Furthermore the ANN methods are iterative procedures with a stochastic base: in fact, at the first iteration, weighted links among neurons are randomly set; then they are optimized during iterations in order to minimize the error. For this reason the resulting forecasts depend on the specific trial. Therefore different trials can provide slightly different results. Usually the final profile is the average of the different trials led in a single run. This is called “ensemble method”. Some studies showed that ANN models using multivariate, such as sun duration, temperature, wind speed, and relative

humidity, can achieve much better performance than that using univariate [19].

Any combination of two or more of the previously described methods is a hybrid model. The idea is to combine different models with unique features to overcome the single negative performance and finally improve the forecast. Recently, some papers show that all these methods need a phase of pre-processing the input data sets in order to increase the forecasting accuracy [16].

Table 1 shows a possible time scale of RES energy forecasting. It includes very short-term, short-term, medium-term and long-term forecasting [22]. The forecast up to 24-h ahead or even more is needed for the power dispatching plans, the optimization operations of grid-connected RES plants and control of energy storage devices. Usually the medium term forecast is requested for the electricity market. The most common forecast horizon term for PV systems is 24-h ahead. Table 1 has been set up with reference to wind, but it can be also applied with reference to PV. Anyway, forecasting term limits are not strictly defined and some different specifications may be granted depending on the application of the forecasting model [24].

Table 1: Time scale classification for RES Forecasting

Term	Range	Application
Very short	Few sec.–30'	Control & adjustment actions
Short	30'–6h	Dispatch planning; load gain/drop
Medium	6h–1 day	Generator on/off; operational security; electricity market
Long	1 day–1 week	Unit commitment; reserve requirement; maintenance schedule

140

141 3. The proposed method

A method based on ANN was developed in order to make the hourly prediction of the production of a RES plant. Fig. 1 illustrates the different phases carried out for the proposed procedure. As regards the training phase, both the weather and the output power measured on the PV systems historical data sets are required in order to lead a supervised learning of the ANN. Once the ANN is trained and tuned, it can be used to provide predictions of the PV system output power by supplying only the weather forecasts as input. After this phase, the accuracy assessment of the results should be

150 carried out. Again, this step needs the output power measured on the PV
 151 systems. Furthermore, to evaluate the accuracy of the method, as well as
 152 the accuracy of the weather forecast, the real parameters measured on the
 153 plant (such as the solar radiation) are really important.

Figure 1: Block diagram of the PV forecasting method based on ANN.

154 3.1. Pre-processing and data validation

155 Before any other step, historical measured data must be always validated,
 156 since unreliable data increase the odds of higher errors in the forecast. The
 157 pre-processing block initially includes the control of the coherence among
 158 the main variables measured in the PV plant, such as the solar radiation
 159 and the PV output power, and a theoretical model of the solar radiation
 160 mathematically computed according to the geographical coordinates of the
 161 PV plant site, by means of a clear sky solar radiation model (CSRM) [4].
 162 Thus the aim of using CSRM in this preliminary step is not only to determine
 163 the time span of the forecast between the sunrise and the sunset of each day,
 164 but also to validate the reliability of each fifteen minutes sample. Fig. 2
 165 shows the flow chart used for this step. First it starts acquiring the values
 166 of $G_{CSRM,1/4h}^k$, $G_{m,1/4h}^k$ and $P_{m,1/4h}^k$ which are the clear sky solar radiation,
 167 the measured solar radiation and the PV output power respectively in the
 168 k -th quarter of an average hour sample. Then, by comparing the other two

variables, when the CSRM is positive, the reliability coefficient of the sample $REL_{1/4h}^k$, is equal to 1 if all the conditions occur at the same time, otherwise it is equal to 0. When $G_{m,1/4h}^k$ is greater than 0 at the same time when $P_{m,1/4h}^k$ is equal to 0 the reliability coefficient of the sample $REL_{1/4h}^k$, is equal to 1. In this case the “failure in the PV plan” condition or the “snow on the PV module” could occur.

Lastly if REL_h , the hourly average of the four reliability coefficients in the same hour, is greater than 0, the h -th hourly sample is considered in the next training and forecasting steps, otherwise not. Secondly, there must be always correspondence between the number of samples and the time-instant of the measured data and those provided by the meteorological service.

Figure 2: Flow chart of the Pre-Processing and data Validation block.

3.2. The forecasting procedure

Among the several existing methods in literature [6], the method here used is based on a statistical approach (specifically ANN) trained with classical algorithms. In this case, the implemented ANN has a classical structure

184 called *multi-layer perceptron* (MLP), while the chosen training procedure for
 185 the neural network is the *error back-propagation* (EBP). The inputs of the
 186 tool are the weather forecasts provided by the meteorological service, the ge-
 187 ographical coordinates of the site as well as the date and time to determine
 188 the correct sun position. The output of the tool is the predicted value of the
 189 hourly power produced by the PV plant for a given time. The training of
 190 the ANN consists in the updating of the weights between the neurons in the
 191 different layers in several iterations comparing the expected data with the
 192 historical-actual ones. On each iteration the existing links between the dif-
 193 ferent input variables are evaluated and the weights are updated accordingly.

194 3.3. Error assessment

195 According to the error definitions of the output power forecast, in com-
 196 parison with the measured data, an error assessment is led.

197 4. Error definitions

198 In order to correctly define the accuracy of the prediction and the related
 199 error, it is necessary to define the indexes that can be used to evaluate the
 200 performances of the forecasting model. Some of these definitions come from
 201 statistics and are well-known [25]. Others are introduced by regulatory au-
 202 thority for market issues: in Italy, for instance, the Authority for electricity
 203 and gas (AEEG) [1]. The error definitions are really different among each
 204 other. Also technical papers present a lot of these indexes, therefore here we
 205 report some of the most commonly used error definitions. The starting refer-
 206 ence point is the hourly error e_h defined as the difference between the average
 207 power produced (measured) in the h -th hour $P_{m,h}$ and the given prediction
 208 $P_{p,h}$ provided by the forecasting model [15], [16]:

$$e_h = P_{m,h} - P_{p,h} \quad (\text{W}) \quad (1)$$

209 From this basic definition, then other definitions can be introduced. The
 210 *absolute hourly error* $e_{h,abs}$ which is the absolute value of the previous defi-
 211 nition (e_h can give both positive and negative values):

$$e_{h,abs} = |e_h| \quad (\text{W}) \quad (2)$$

212 The *hourly error percentage* could be $e_{\%,p}$, if it is based on the hourly
 213 output expected power hour $P_{p,h}$:

$$e_{\%,p} = \frac{|e_h|}{P_{p,h}} \cdot 100 \quad (3)$$

214 or, if it is based on the hourly output measured power hour $P_{m,h}$:

$$e_{\%,m} = \frac{|e_h|}{P_{m,h}} \cdot 100 \quad (4)$$

215 These two errors will be compared in section 5.

216 The *normalized mean absolute error* $NMAE_{\%}$, based on net capacity of
 217 the plant C :

$$NMAE_{\%} = \frac{1}{N} \sum_{h=1}^N \frac{|P_{m,h} - P_{p,h}|}{C} \cdot 100 \quad (5)$$

218 where N represents the number of samples (hours) considered: usually it
 219 is referred to a day, a month or a year. For this indicator the rated power of
 220 the PV system was considered as C .

221 The *weighted mean absolute error* $WMAE_{\%}$, based on total energy pro-
 222 duction:

$$WMAE_{\%} = \frac{\sum_{h=1}^N |P_{m,h} - P_{p,h}|}{\sum_{h=1}^N P_{m,h}} \cdot 100 \quad (6)$$

223 The *normalized root mean square error* $nRMSE$, based on the maximum
 224 observed power output $P_{m,h}$:

$$nRMSE_{\%} = \frac{\sqrt{\frac{\sum_{h=1}^N |P_{m,h} - P_{p,h}|^2}{N}}}{\max(P_{m,h})} \cdot 100 \quad (7)$$

225 $NMAE_{\%}$ is largely used to evaluate the accuracy of predictions and trend
 226 estimations. In fact, often relative errors are large because they are di-
 227 vided by small power values (for instance the low values associated to sunset

and sunrise): in such cases, $WMAE_{\%}$ could result very large and biased, while $NMAE_{\%}$, by weighting these values with respect to the capacity C , is more useful. The $nRMSE_{\%}$ measures the average magnitude of the absolute hourly errors $e_{h,abs}$. In fact it gives a relatively higher weight to larger errors, thus allowing to emphasize particularly undesirable results.

5. Case study

This section describes the 1-day ahead hourly forecast achieved by using the ANN method. The prediction is carried out every day at the same time in the morning for all the daylight hours of the following day. The considered real plant is a 264kWp rated power, facing South; it is composed of polycrystalline silicon photovoltaic panels, 19° tilted, fixed on the roof of a factory located in the North of Italy. Different simulations have been run in order to compare the errors (3) and (4) both in terms of single trial and ensemble profile (based on 10 trials) of the ANN. Besides a simple sensitivity analysis of the method has been performed as a function both of the training set and the period of the year. This analysis consisted in 30 days forecasting, varying the above mentioned settings and evaluating the errors of the whole estimated period in all the daylight hours. Finally the results of some peculiar days are presented as typical cases which more frequently may occur.

5.1. Characteristics of data and models

While the hourly PV plant electric power generation data are recorded with measurement equipment on site, the weather data are provided by a forecasting service with 72 hours in advance. With reference to the prediction of hourly production relative to a day ahead, the analysis is performed using an ANN with the following characteristics: 240 days (8 months) data set, 9 neurons in the first layer, 7 neurons in the second layer, and 3000 iterations for each trial. Different structures of ANN have been tested both in terms of number of neurons [2] and training iterations [9], and the above described configuration proved to be a good compromise in terms of efficiency and computational time effectiveness. The training period over the same forecasting span instead was changed both in terms of number of days and starting point.

5.2. Results

The results of this analysis concern specifically the error definitions previously exposed. They are analyzed following different approaches, as better

clarified in the following subsections, in order to underline the efficiency of the ANN method due to its setting changes. In this paper all the errors are referred only to the prediction in the daylight hours.

A. Comparison of error definitions

The starting point of the error analysis is the hourly error definition (1). Its absolute value can be related, for the same computed forecast, both to the hourly average produced (measured) power $P_{m,h}$, as expressed by (4), and to the forecast power in the h -th hour $P_{p,h}$ (3). Since the ANN is a stochastic method, the forecast activity is performed several times (or by several neural network's casts in parallel at the same time) for the same time period, resulting in different predicted PV output power profiles. Each predicted profile, which is referred to the i -th simulation, is therefore called a *trial*.

Thus the final hourly power forecast $P_{p,h}$ is the average of power $P_{p,h}^i$ over i samples referred to the same h hour. We obtain for n different trials, the following expression:

$$P_{p,h} = \sum_{i=1}^n \frac{P_{p,h}^i}{n} (W) \quad (8)$$

which is the power forecast by the so-called *ensemble method*. Therefore the hourly percentage errors (3) and (4), can be redefined respectively as:

$$e_{\%,p} = \frac{|P_{m,h} - \sum_{i=1}^n \frac{P_{p,h}^i}{n}|}{\sum_{i=1}^n \frac{P_{p,h}^i}{n}} \cdot 100(\%) \quad (9)$$

and:

$$e_{\%,m} = \frac{|P_{m,h} - \sum_{i=1}^n \frac{P_{p,h}^i}{n}|}{P_{m,h}} \cdot 100(\%) \quad (10)$$

where n is the number of trials as already explained.

Fig. 3 shows the absolute hourly error of each trial (thin lines) and the absolute hourly error of the ensemble forecast (thicker lines). This format is

Figure 3: Absolute hourly error based both on predicted (blue) and measured (red) output power for each trial (thin lines) and ensemble (thicker lines).

286 shown both for the hourly errors based on the predicted output power $e_{%,p}$
 287 (in blue) and the hourly error based on the measured output power $e_{%,m}$ (in
 288 red).

289 In our analysis we used the first 90 days of historical data for training and
 290 the last 150 days for forecast evaluation. All the curves reported in Fig. 3
 291 are ordered based on hourly forecasting error magnitude starting from the
 292 largest to the smallest, truncating at 1500 samples, for an easier comparison.

293 Following this approach we can draw some considerable results. First of
 294 all it can be observed that the hourly error based on the predicted power
 295 $e_{%,p}$ is generally smaller than the one based on the measured power $e_{%,m}$.

296 Secondly the hourly ensemble error based on the predicted power is
 297 smaller than each relative trials. For this reason later on the forecasted pro-
 298 files are evaluated by using the ensemble method based on predicted power.

299 Finally, the ensemble error based on the measured power, by definition
 300 (from eq. 10), is just the average value over the n trials.

301 *B. Training time period analysis*

302 Figure 4 shows the errors evaluated as a function of size of the training
 303 set (60, 90 and 120 days) and of the different 30 days' predicted period in
 304

various times of the year. The error definitions are applied to this entire period with reference to daylight hours. Fig. 5 shows, for instance, how this process works for the PV output forecasting in the period from 181th to 210th day.

Generally the longer is the training set size, the lower are the errors. But some peculiar weather conditions play an important role: for instance, in the considered 240 days dataset, there is snow from day 27 to 53, 66 and 67. Also few days after, the actual weather conditions were severe (day 101, 104 and 105). It can be noticed how these conditions affected the forecasting reliability. In fact, even if the number of the training days increased, if they included very peculiar weather conditions, the global forecasting accuracy was worse.

In order to identify those days which may affect the training set reliability, it can be useful to sort the dataset according to the daily *clearness index* k_t , defined as the ratio of the horizontal global irradiance to the corresponding irradiance available out of the atmosphere [10]. In [13] an example of clustering days with these criteria is provided and, in this research, the dataset has been classified, according to the typical values of k_t for the PV plant site, into three partitions as reported in Table 2.

Table 2: Clearness index k_t partitions.

Weather condition	k_t range
Clear	$k_t > 0.45$
Partially cloudy	$0.25 > k_t > 0.45$
Cloudy	$k_t < 0.25$

Table 3: Classification of the training periods.

Days	Clear	Part.cloudy	Cloudy
1-30	57%	30%	13%
31-60	60%	17%	23%
61-90	87%	7%	7%
91-120	63%	13%	23%
121-150	77%	7%	17%
151-180	83%	17%	0%
181-210	87%	13%	0%
211-240	93%	7%	0%
TOTAL	76%	14%	10%

In particular, as reported in Table 3, according to the *clearness index*

Figure 4: Errors as a function of the size of the training set and of the different period of the year.

Figure 5: Scheme of the 30 consecutive forecasted days with a different size of the training period.

classification the number of day-type is different in each of the 30-days period considered, especially in those forecast periods after May. Moreover the next forecast periods are close to summer, which reasonably means a higher number of “clear” or “partially cloudy” days in comparison to the previous periods. Indeed, during the 30 days periods after the 121-150, the number of “clear days” increases, whereas the number of “cloudy days” is lowered to zero. In fact the first forecasting period (121-150) has a comparable number of day-type mix, along with the increasing number of days employed in the training. And, as it is shown in Figure 4, evaluation indexes are largely constant. While, considering the day-type composition of the 151-180 period, it has no cloudy days at all, and the evaluation indexes are greatly affected with the 60 days training forecast not only for the lower number of days employed in the training, but also because the different mix of the day-type presents the highest number of “cloudy days”. Therefore it is clear that both quantity and quality of the samples in the training set are critical to the forecasting reliability and accuracy.

A direct comparison of these results with previous ones presented in literature is not easy: often error definitions and considered time frames differ from paper to paper. However, if we consider the above presented $nRMSE$ values, we can see a significant agreement with results shown in [23] and [17], which were obtained on more favorable conditions, i.e. from 1 to 12 hours ahead.

C. Analysis of significant days

Some typical days have been taken into account in order to evaluate the method forecast accuracy applied to a reduced number of hourly samples. The simulations have been carried out with the same settings listed before,

352 applied to three significant days with different weather conditions recorded
353 in the month of May:

- 354 (a) sunny day with sunny weather forecasts (Fig. 6),
- 355 (b) partially cloudy day with variable weather forecasts (Fig. 7),
- 356 (c) cloudy day (Fig. 8).

357 These figures show the trends of the PV plant predicted power P_p and the
358 measured power P_m based on the rated power of the plant C ; the irradiance
359 provided by the weather service G_p and the measured irradiance G_m based
360 on the irradiance at the standard test conditions $G_{stc} = 1000W/m^2$. Fur-
361 thermore the $NMAE\%$, $nRMSE\%$ and $WMAE\%$ forecasting errors referred
362 only to the daylight hours are reported.

363 It can be noticed that the error is highly related to the solar irradiance
364 forecasting accuracy. Furthermore $WMAE\%$ becomes high during the un-
365 stable days. The best case is represented by the typical sunny day (case (a)):
366 the measured irradiance G_m is totally in agreement with the weather forecast
367 provided by the meteorological service.

368 Instead, in case (b) the measured irradiance G_m is only partially in agree-
369 ment with the weather forecast provided by the meteorological service. In
370 fact, the weather service was not able to accurately forecast the exact time
371 when the instability appeared. Thus the forecasting error was mainly due
372 to the time shift between predicted and measured data. Of course in this
373 light further improvements are achievable improving the accuracy of weather
374 service forecast, and on the other hand reducing the relative time interval
375 for their predictions, namely from the cited 72 hours used in this work to at
376 least 48 hours in advance.

377 Finally, the worst case is represented by (c). In this case the measured
378 power P_m is really low (full cloudy day) and consequently percentage errors
379 are quite relevant even if the overall produced energy in that day is quite
380 negligible. However, it is important to underline that this specific day repre-
381 sents one of the worst cases over the entire data set analyzed in this paper.
382 In all the considered cases, we noticed that the most relevant errors occur
383 during sunrise and sunset; therefore, possible enhancement to our method
384 can be performed by improving the way sunset and sunrise are taken into
385 account, for instance by adopting hybrid methods, as shown in [7].

Figure 6: Predicted, measured power and irradiance curves, and errors in a sunny day.

Figure 7: Predicted, measured power and irradiance curves, and errors in a partially cloudy day.

Figure 8: Predicted, measured power and irradiance curves, and errors in an cloudy day.

6. Conclusion

In this paper a PV energy forecasting method based on ANN is presented. The error assessment, according to the error definitions here introduced, shows that the ensemble error is smaller than those obtained by the single trials. Besides it has been highlighted that the method accuracy is strictly related to the historical data pre-process step and to the accuracy of the historical data set used for the training step. The trends of the errors clearly show how the accuracy in the sunny days is higher, while in partially cloudy and cloudy days the overall efficiency is slightly different. Some improvements are therefore connected to the reliability of the weather forecasting and to the pre-processing of the raw data to train the network.

Acknowledgements

The authors are thankful to the *Solar Tech^{LAB}* of Department of Energy, Politecnico di Milano, <http://www.solartech.polimi.it/>.

References

- [1] AEEG, Orientamenti in materia di regolazione del servizio di dispacciamento da applicarsi alle unità di produzione di energia elettrica, con particolare riferimento a quelle non programmabili, Tech. Rep. 35/2012/R/efr, Autorità per l'energia elettrica e il gas, Italy (2012).
- [2] M. Aghaei, A. Dolara, F. Grimaccia, S. Leva, M. Mussetta, E. Ogliari, PV plant planning and operations by neural network analysis and validation, in: 29th European Photovoltaic Solar Energy Conference and Exhibition, 2014.
- [3] P. Bacher, H. Madsen, H. A. Nielsen, Online short-term solar power forecasting, *Solar Energy* 83 (10) (2009) 1772 – 1783.
- [4] R. E. Bird, A simple, solar spectral model for direct-normal and diffuse horizontal irradiance, *Solar Energy* 32 (4) (1984) 461 – 471.
- [5] G. Capizzi, F. Bonanno, C. Napoli, A wavelet based prediction of wind and solar energy for long-term simulation of integrated generation systems, in: *Power Electronics Electrical Drives Automation and Motion (SPEEDAM)*, 2010 International Symposium on, 2010.
- [6] S. H. Chen, A. J. Jakeman, J. P. Norton, Artificial intelligence techniques: An introduction to their use for modelling environmental systems, *Mathematics and Computers in Simulation* 78 (23) (2008) 379 – 400.
- [7] 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 (2) (2015) 1138–1153.
- [8] A. Dolara, G. C. Lazaroiu, S. Leva, G. Manzolini, Experimental investigation of partial shading scenarios on PV (photovoltaic) modules, *Energy* 55 (0) (2013) 466 – 475.
- [9] A. Gandelli, F. Grimaccia, S. Leva, M. Mussetta, E. Ogliari, Hybrid model analysis and validation for pv energy production forecasting, in: *Neural Networks (IJCNN)*, 2014 International Joint Conference on, 2014.

- 431 [10] M. Iqbal, An Introduction To Solar Radiation, Academic Press, 1983.
- 432 [11] R. Iqdour, A. Zeroual, A rule based fuzzy model for the prediction of
433 daily solar radiation, in: Industrial Technology, 2004. IEEE ICIT '04.
434 2004 IEEE International Conference on, vol. 3, 2004.
- 435 [12] E. Izgi, A. Oztopal, B. Yerli, M. K. Kaymak, A. D. Sahin, Shortmid-
436 term solar power prediction by using artificial neural networks, Solar
437 Energy 86 (2) (2012) 725 – 733.
- 438 [13] A. Kudish, A. Ianetz, Analysis of daily clearness index, global and beam
439 radiation for beer sheva, israel: Partition according to day type and
440 statistical analysis, Energy Conversion and Management 37 (4) (1996)
441 405 – 416.
- 442 [14] A. Mellit, A. M. Pavan, A 24-h forecast of solar irradiance using arti-
443 ficial neural network: Application for performance prediction of a grid-
444 connected pv plant at trieste, italy, Solar Energy 84 (5) (2010) 807 –
445 821.
- 446 [15] C. Monteiro, L. A. Fernandez-Jimenez, A. Ramirez-Rosado, I. J. ans
447 Munoz-Jimenez, P. M. Lara-Santillan, Short-term forecasting models
448 for photovoltaic plants: Analytical versus soft-computing techniques,
449 Mathematical Problems in Engineering 2013 (2013) 9.
- 450 [16] E. Ogliari, F. Grimaccia, S. Leva, M. Mussetta, Hybrid predictive mod-
451 els for accurate forecasting in pv systems, Energies 6 (4) (2013) 1918–
452 1929.
- 453 [17] H. T. Pedro, C. F. Coimbra, Assessment of forecasting techniques for
454 solar power production with no exogenous inputs, Solar Energy 86 (7)
455 (2012) 2017 – 2028.
- 456 [18] G. Reikard, Predicting solar radiation at high resolutions: A comparison
457 of time series forecasts, Solar Energy 83 (3) (2009) 342 – 349.
- 458 [19] A. Sfetsos, A. Coonick, Univariate and multivariate forecasting of hourly
459 solar radiation with artificial intelligence techniques, Solar Energy 68 (2)
460 (2000) 169 – 178.

- 461 [20] J. Shi, W.-J. Lee, Y. Liu, Y. Yang, P. Wang, Forecasting power output of
462 photovoltaic system based on weather classification and support vector
463 machine, in: Industry Applications Society Annual Meeting (IAS), 2011
464 IEEE, 2011.
- 465 [21] M. Simonov, M. Mussetta, F. Grimaccia, S. Leva, R. Zich, Artificial
466 intelligence forecast of pv plant production for integration in smart en-
467 ergy systems, *International Review of Electrical Engineering* 7 (1) (2012)
468 3454–3460.
- 469 [22] S. Soman, H. Zareipour, O. Malik, P. Mandal, A review of wind power
470 and wind speed forecasting methods with different time horizons, in:
471 North American Power Symposium (NAPS), 2010, 2010.
- 472 [23] R. Ulbricht, U. Fischer, W. Lehner, H. Donker, First steps towards a
473 systematical optimized strategy for solar energy supply forecasting, in:
474 European Conference on Machine Learning and Principles and Practice
475 of Knowledge Discovery in Databases (ECMLPKDD 2013), 2013.
- 476 [24] F. Wang, Z. Mi, S. Su, H. Zhao, Short-term solar irradiance forecasting
477 model based on artificial neural network using statistical feature param-
478 eters, *Energies* 5 (5) (2012) 1355–1370.
- 479 [25] H.-T. Yang, C.-M. Huang, Y.-C. Huang, Y.-S. Pai, A weather-based
480 hybrid method for 1-day ahead hourly forecasting of pv power output,
481 *Sustainable Energy, IEEE Transactions on* 5 (3) (2014) 917–926.