One day-ahead forecasting of energy production in solar photovoltaic installations: An empirical study

Article *in* Intelligent Decision Technologies · August 2012

DOI: 10.3233/IDT-2012-0136

CITATIONS
21

READS
1,213

3 authors:

Marco Cococcioni
University of Pisa
150 PUBLICATIONS **1,912** CITATIONS
SEE PROFILE

Eleonora D'Andrea
University of Pisa
26 PUBLICATIONS **1,312** CITATIONS
SEE PROFILE

B. Lazzerini
University of Pisa
123 PUBLICATIONS **1,858** CITATIONS
SEE PROFILE

Intelligent Decision Technologies 6 (2012) 197–210
DOI 10.3233/IDT-2012-0136
IOS Press

One day-ahead forecasting of energy production in solar photovoltaic installations: An empirical study

Marco Cococcioni, Eleonora D'Andrea and Beatrice Lazzerini*
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, University of Pisa, Pisa, Italy

Abstract. This paper presents a flexible and easy-to-use methodological approach to the forecasting of energy production in solar photovoltaic (PV) installations, using time series analysis and neural networks. The aim is to develop and validate a one day-ahead forecasting model by adopting an artificial neural network with tapped delay lines.
The main novelty of our approach is the proposal of a general methodology, consisting of a sequence of steps to perform in order to find, based on heuristic criteria, the optimal structure of the neural network (particularly, number of hidden neurons and number of delay elements) and the best configuration of the neural predictor (namely, the training window width and the sampling frequency).
The best experimental results have been obtained using as inputs the irradiation and the sampling hour to predict the daily accumulated energy. Considering a dataset of 15-minute measurements pertinent to one year, despite the presence of 77 missing days (scattered through the whole year in correspondence with system slowdown), we achieved seasonal mean absolute percentage errors ranging from a minimum of 12.2% (Spring) to a maximum of 26% (Autumn). Moreover the achieved results are significantly better than those obtained by the persistence method, a benchmark frequently used in this kind of applications.

Keywords: Artificial neural networks, forecasting, NARX, solar photovoltaic energy, time series

1. Introduction

Renewable energy refers to energy generated by natural sources, such as solar energy, wind energy and tidal energy.

As conventional fossil fuel energy sources are diminishing [13], renewable energy, and in particular solar energy, is becoming a valid alternative to traditional energy since it is considered economical and non-polluting [32] besides practically inexhaustible. For these reasons photovoltaic (PV) installations have spread in recent years [23].

A PV installation consists of a series of solar panels that using sunlight energy generates directly usable electricity thanks to the PV effect. Usually, in an installation, several panels are connected together to form a system called PV array, to which an inverter is connected that measures the production power of that array, and converts the DC power in AC power, as requested by the electrical network.

With the diffusion of PV systems the monitoring of the performance of solar panels has become a key issue, so as to detect efficiency losses or effectively plan the energy distribution, for instance in smart grid installations. This can be done by estimating the forecasted energy and comparing it with the real produced energy.

Although a lot of people work in this research area, they are mainly concerned with forecasting solar radiation [2,18,19,21,22], whereas only a few are concerned with the forecasting of solar energy production directly.

The literature related to the renewable energy field presents many approaches to forecasting load, wind speed or solar irradiation. The most widely used tech-

*Corresponding author: Beatrice Lazzerini, Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, University of Pisa, Largo Lucio Lazzarino 1, 56122 Pisa, Italy. E-mail: b.lazzerini@iet.unipi.it.

ISSN 1872-4981/12/$27.50

niques include regression methods [7,9,14], neural networks [1,19,28], and time series analysis [10]. However most of the existing methodologies present some drawbacks such as high average accuracy error, dependence on the particular design of the PV installation, and inability to provide real-time prediction [3,24,26].

The proposed methodology aims to find the most appropriate model structure together with the optimal width of the prediction window used to forecast the energy production of the following day. The potential benefits of having energy production predictability are obvious: knowing the energy production ahead of time is useful in automatic power dispatch and load scheduling, and energy control [29].

The goal of the proposed approach is twofold. On the one hand, energy production forecasting is important for improving the efficiency of the PV installation, as well as for finding faults. On the other hand, energy production forecasting is vital for efficiently planning the energy distribution [8]. In particular, the chance to forecast the energy production up to 24 hours can become of the utmost importance in decision-making processes, with particular reference to grid-connected photovoltaic plants. Moreover, this approach could be particularly useful in smart grid systems, which are able to make better operational decisions using ahead predictions [25].

Variability of weather, in particular solar irradiation, is maybe the main difficulty faced by PV installation operators so that good forecasting tools are required for the appropriate integration of renewable energy into the power system [14].

Among various prediction models, such as analytic, stochastic, and empirical, neural networks are fairly able to correctly model the nonlinear nature of dynamic processes. Actually, an artificial neural network is able to reproduce an empirical, possibly nonlinear, relationship between some inputs and one or more outputs [4].

Forecasting is one of the most interesting nonlinear applications of neural networks. Indeed, the measurements of environmental parameters are generally provided in the form of time series which are suitable to use neural networks for prediction purposes [5]. Moreover, neural networks are fault tolerant, i.e., are able to handle noisy or incomplete data [12].

The paper has the following structure. Section 2 describes the experimental data. Section 3 introduces the time series analysis model employed in this work and reviews the main concepts of artificial neural networks. Section 4 presents the proposed methodology to correctly set up the model for forecasting the solar energy production. Sections 5 and 6, respectively, present and discuss the achieved results. Finally, concluding remarks and future work are provided in Section 7.

2. Experimental dataset

The data used in this work were collected from two PV ground installations of solar panels 500 meters far away from each other, located in Apulia, Italy. We employed data only from one installation for the development of the model, as, under conditions of proper working, we can assume that the two installations are correlated. We then successfully tested the developed model also on the second installation.

According to the literature, the two most significant environmental parameters are *temperature* and *irradiation* on the solar panels. The first one refers to the surface of the panel exposed to the sunlight, while the second one is the quantity of sun radiation incident on the panel with respect to the whole surface of the panel and to all the electromagnetic spectrum frequencies.

Actually, we employed only the *irradiation* input, similarly to what happens in PV plant sizing, since the total solar radiation is considered as the most important parameter in the performance prediction of renewable energy systems [20]. The first input, instead, was used only for offline checks, as explained in Section 5.

The output parameter is the *produced energy* from each PV installation.

Data were collected every 15 minutes during 1 year, from October 2009 to September 2010.

3. NARX time series analysis model and neural networks

3.1. Nonlinear Autoregressive with Exogenous Input (NARX) model

Prediction uses past values of one or more time series to predict future values of the unknown time series.

One of the most general time series analysis models is called NARX (*Nonlinear Autoregressive with Exogenous Input*) [16,17]. This model allows to forecast the future values of the output time series $y(t)$, knowing the past values of the same endogenous series $y(t)$ and the past values of the exogenous series $x(t)$. The equation model is the following:

$$\begin{aligned} y(t) = \; & f(y(t-1), \ldots y(t-d_1), x(t-1), \\ & \ldots x(t-d_2)). \end{aligned} \tag{1}$$

In Eq. (1), d_1 and d_2 represent the number of delay elements for the two series employed for the prediction. Though these values can be different for the two time series involved, usually $d_1 = d_2 = d$ is the preferred choice.

Figure 1 shows the model block diagram in closed loop mode. This means that the output of the system is fed back as input.

Finally, two different forecasting schemes exist: i) an *iterative* scheme, which aims to predict a variable, e.g., the energy production, at a given time step and the obtained prediction is used as input for the forecasting of the next time step; ii) a *direct* scheme, which aims to predict the next n (e.g., 24 hours) time steps from the same input data [31]. In the paper we adopted the iterative scheme.

3.2. An overview of artificial neural networks

Artificial Neural Networks (ANNs) are data-driven intelligent systems having the capability to learn, remember and generalize. ANNs reproduce the learning process of the human brain by learning the relationship between input parameters and output variables based on previously recorded data [3].

The elementary processing element of an ANN is the neuron. In its basic computational form, a neuron appropriately processes a set of input signals coming from other neurons or sources [34]. ANN models may be used as alternative methods in engineering analyses and predictions [12]. Neural networks are particularly suited to implement NARX models.

The most used architecture and training algorithm for predicting purposes are, respectively, the multilayer feed-forward neural network, which includes one or more hidden layers, and the Levenberg-Marquardt (LM) backpropagation training algorithm, which shows good generalization capability and simplicity [30]. We therefore decided to adopt this type of ANN trained with the LM algorithm.

4. The proposed methodology

The proposed methodology aims to provide an easy-to-use tool for correctly configuring the best predictor for energy production in a PV installation based only on the collected historical data, with no concern about possible relationships between plant attributes (like positions, construction, etc.) and predictor parameters. More in detail, we aim to forecast the energy production

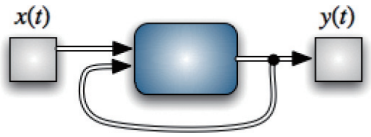

Fig. 1. NARX model block diagram. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IDT-2012-0136)

up to 24 hours, given the environmental parameters of an appropriate number of previous days that compose the *training window*.

We decided to implement the NARX model using a feed-forward neural network with tapped delay lines, having one hidden layer. The hidden neurons are characterized by a hyperbolic tangent sigmoid function while the output neuron has a linear transfer function.

The tasks to be performed are basically the following: i) choice of the training window width; ii) choice of the sampling frequency; iii) choice of the number of delays; iv) choice of the number of hidden neurons. In the following we will use the term *structural* parameters to refer to the number of hidden neurons and the number of delay elements, and the term *configuration* parameters to mention the training window width and the sampling frequency.

Based on heuristic considerations, we propose i) to consider the training window width as a multiple of the day, ii) to use balanced training sets obtained by randomly extracting the same number of training samples from each day of a window, and iii) to assess the performance of the NARX model based on the mean performance obtained on all the predicted days within periods of one month. This last choice aims to avoid the strong dependence of the performance results, achieved by a given neural network, on the particular day to be predicted (which might be atypical).

In the following we will describe the operation steps for the development of the neural model.

4.1. Choice of the structural and configuration parameters

Based on the previous considerations, we make use of training windows having width w from a minimum of 7 days to a maximum of 30 days before the predicted day. Furthermore, for the sake of simplicity, we kept the 15 minute sampling frequency.

As previously stated, we decided to take into account the mean performance obtained on all the predicted days within periods of one month. We considered one month for each season. More in detail, we performed the experiments on the days of the first complete month

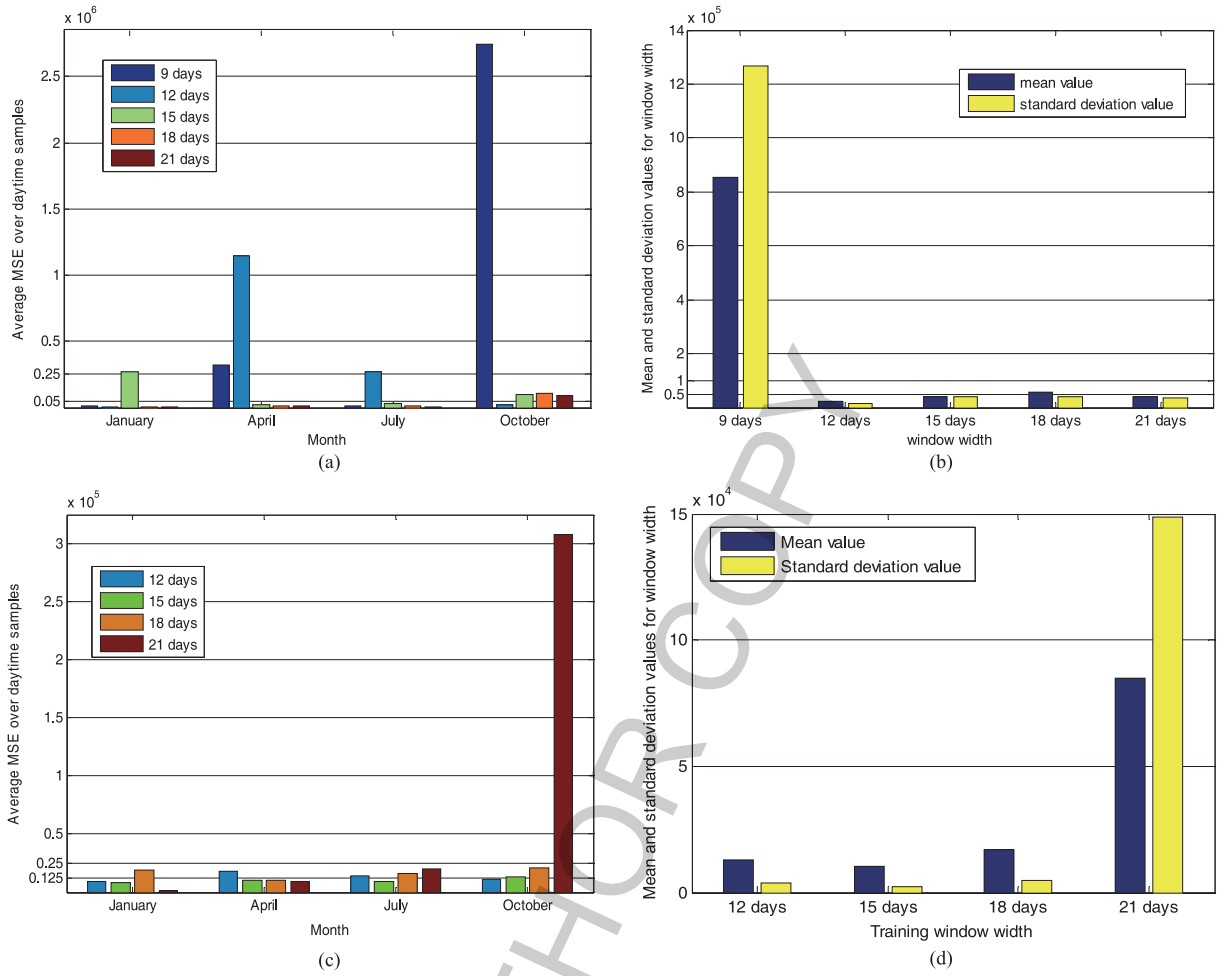

Fig. 2. (a) Average MSE over the daytime samples for whole months (January, April, July and October), using irradiance only. (b) Mean and standard deviation of the average MSEs, related to the four considered months and for each window width, using irradiance only. (c) Average MSE over the daytime samples for the same months, using irradiance and hour input. (d) Mean and standard deviation of the average MSEs, related to the four considered months and for each window width, using irradiance and hour. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IDT-2012-0136)

of each season (e.g., being the beginning of winter on December 21$^{\text{st}}$, January is chosen as the first complete month of winter), having a total of four months.

We heuristically decided to try a number of hidden neurons (h) ranging from 8 to 20 with step 1, and a number of delays (d) ranging from 3 to 10 with step 1.

Each kind of experiment has been repeated 30 times, once fixed the neural network configuration and structure, to mitigate the effects, in terms of convergence, of the random initialization typical of neural networks.

For each month considered, we found a set of windows that appear most often, i.e., at least in a given percentage (80% in our case) of the trials of each experiment, as the best windows among the 24 possible windows. This set is $w = 9, 12, 15, 18, 21$.

The best results were obtained with a neural network with 10 hidden neurons, and 3 delay elements on both the exogenous (*irradiation*) and endogenous (*produced energy*) variables. These values represent the best compromise between efficiency and simplicity. In particular, to set the number of delays to 3 means to use the information pertinent to the three quarters of hour immediately before each predicted sample.

We used the Mean Square Error (MSE) defined in Eq. (2), where t_i and o_i are, respectively, the target and predicted instantaneous energy values of sample i, and S is the number of samples of the considered day:

$$MSE = \frac{1}{S}\sum_{i=1}^{S}(t_i - o_i)^2. \quad (2)$$

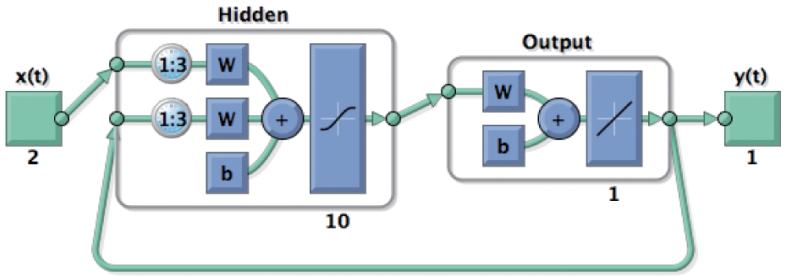

Fig. 3. Network model employed in the experiments. The model has 2 inputs, 10 hidden neurons with hyperbolic tangent sigmoid transfer function and 1 output with linear transfer function; the output is fed back as input; 3 delays are used on all the inputs. W and b represent, respectively, the weight matrix and the bias. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IDT-2012-0136)

Figure 2(a) shows the results related to window widths equal to 9, 12, 15, 18 and 21 days, respectively, for the four months. For each window width the figure shows the MSE made during the daytime (hours of daylight, from 6 a.m. to 8 p.m.) of each predicted day, averaged over all the days of the considered month.

To compare the performances of the different neural models related to the five considered windows, we have evaluated the mean value and the standard deviation of the previously computed average MSEs for each window width over the four months. Figure 2(b) shows the mean value and the standard deviation of the five neural models (corresponding to the five windows) over the four months (January, April, July and October). As we can easily observe in Fig. 2(b), the worst window is that with 9 days width, so we decided to discard it.

Regarding the number of delays, as stated before, we chose to use 3 delays. Actually, we have experimentally found that, by increasing the number of delays, we can slightly improve the performance of the network. Anyhow, this depends on the window width value chosen and on the month considered, therefore it makes no sense to increase the complexity of the network. In a sense, we may say that the window width w lets the network be "aware" of the specific season or more in general the temporal context, while the number of delays lets the network detect the particular real-time climatic variation.

Moreover, we have experimentally found that there does not exist an optimum value of w effective for all the seasons. Nevertheless, we have realized that in most cases, (about) 10 days is the minimal window width able to provide the network with the correct temporal context (season). In fact, the typical way to provide data for solar energy climatology has monthly, annual and 10-days granularity [11].

The previous result has been confirmed by repeating the experiments on the remaining months of each season.

4.2. Model refinement

The goal of this phase is to propose a way to improve the model performance by trying to resolve possible irregularities present in the data collected from the specific PV installation.

During the experiments we noticed the presence of error spikes corresponding to days having a mean MSE sensibly higher than the other days of the same month. The position of these spikes (and the related days) is not fixed as the window width varies. More precisely, if a given day shows a high error for a specific window maybe the following day shows a similar high error for a different window.

So, to resolve this irregularity of the model, we decided to add one more input parameter to improve the performance, and, at the same time, to favor the regularization of the occurrence of the previous error spikes.

As we have a timestamp associated with each sample, we decided to add the *hour* input, maintaining fixed the structural parameters of the network. In fact, given a temporal context (e.g., a season or a month), the daily irradiation values at the same time of the day tend to be very similar, especially considering the values far from the maximum (approximately corresponding to midday).

Stated in other terms, by adding the hour input, we want to let the network be more aware of the concept of succession of samples during the day, with the hour representing the specific time at which the samples have been collected, so as to exploit some kind of regularity in the sun irradiation cycle.

So we performed the previous experiments with one more input, i.e., the hour input.

The resulting network model is depicted in Fig. 3 (produced in the Matlab® environment).

We found that the network that employs irradiation and hour inputs performs better compared to the one

Table 1
Parameters for the final neural model

Parameter	Value
Structural parameters	
Number of hidden layers	1
Number of hidden neurons (h)	10
Number of delay elements (d)	3
Transfer functions	Hyperbolic tangent sigmoid (hidden layer), linear (output layer)
Training algorithm	Levenberg-Marquardt
Maximum number of learning epochs	30
Early stopping criterion	6 validation failures
Exogenous input variables	*Irradiation*, *hour*
Endogenous input variable	*Produced energy*
Configuration parameters	
Training window width (w)	15 days
Sampling frequency (s)	15 minutes

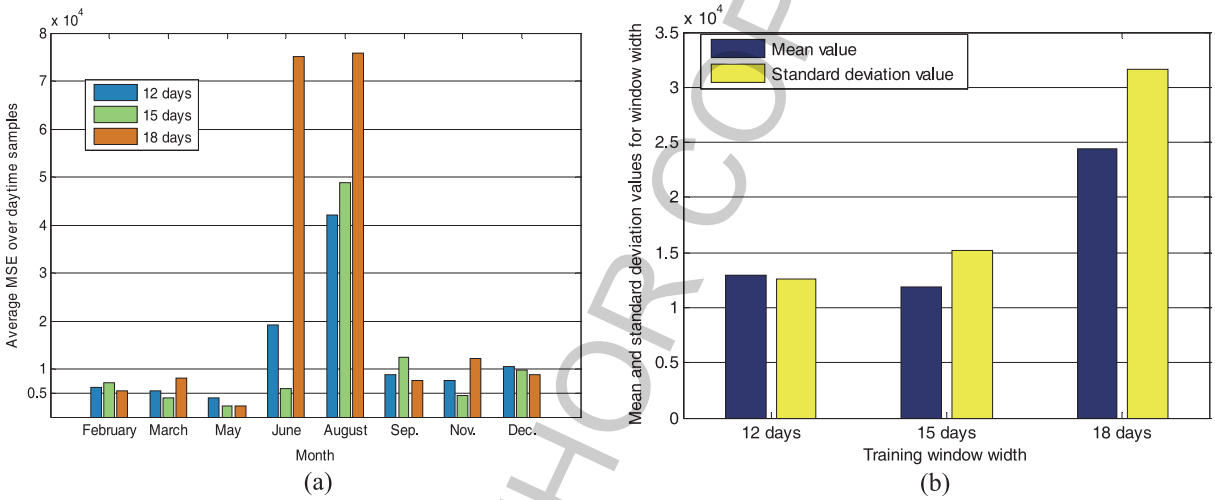

Fig. 4. (a) Average MSEs over the daytime samples for a whole month: February, March, May, June, August, September, November, December and (b) the corresponding mean and standard deviation values for each window width. This neural model uses both the irradiation and hour inputs. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IDT-2012-0136)

that uses only the irradiation input (see Figs 2(c) and 2(d)). We have measured a performance improvement of about one order of magnitude for all window widths.

Figure 2(c) shows the daytime MSEs of each predicted day, averaged over all the days of the considered month and related to window widths equal to 12, 15, 18 and 21 days, respectively, while Fig. 2(d) shows the mean value and the standard deviation of the average MSEs for the four months (January, April, July and October). From Fig. 2(d) we can notice that the best window width is $w = 15$ as it corresponds to the lowest mean value and standard deviation, although the three window widths 12, 15 and 18 are comparable with each other. On the other hand, these three windows are sensibly better than $w = 21$. For this reason we decided to eliminate $w = 21$ from further analysis.

We performed a following assessment phase to attempt to select the best window width w, among the previous three found (12, 15, 18), for the particular temporal context (season). We carried out our experiments on the remaining months of each season.

Figure 4(a) shows the goodness of the adopted methodology on the remaining months. The achieved performance is in line with the previous results. As we can see, the optimal window width results to be 15 days. Figure 4(b) shows the mean value and the standard deviation of the average MSEs for the three windows over the remaining eight months (February, March, May, June, August, September, November, and December). From Fig. 4(b) we can notice that the best window width is $w = 15$ as it corresponds to the lowest mean value although the standard deviation is slightly higher than that of 12 days. On the other hand, the

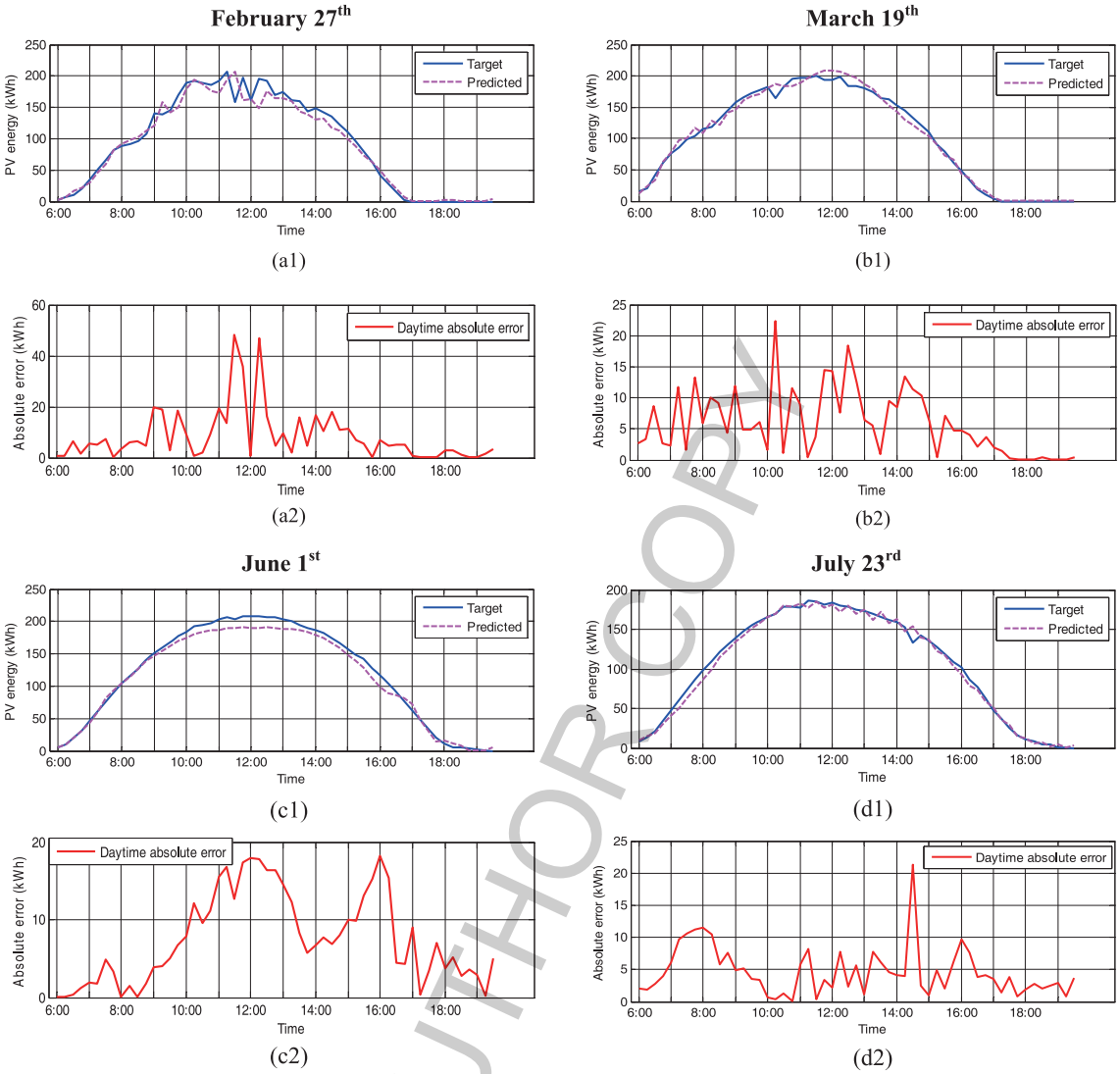

Fig. 5. Results on four *well performing days* chosen at random. (a1), (b1), (c1), (d1) Comparison between the real energy and the predicted energy. (a2), (b2), (c2), (d2) Associated daytime absolute instantaneous error. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IDT-2012-0136)

differences among the three window widths are in fact negligible. For practical reasons, in the experiments described in the next section we decided to adopt $w = 15$.

Table 1 summarizes the parameter values of the chosen final configuration.

5. Experimental results

The analysis described in the previous section shows that the neural network with 10 neurons in the hidden layer and 3 delay elements is the best structure for the proposed problem. Moreover, the best performance was obtained employing the hour input along with the irradiation input.

Finally, we have verified in Section 4 that the best results were achieved using training window widths of 12, 15 or 18 days. As already stated, we adopted $w = 15$.

In the following two sub-sections we analyze, respectively, the prediction of the instantaneous energy on the continuous daily horizon and the prediction of the total (accumulated) energy produced over the whole day.

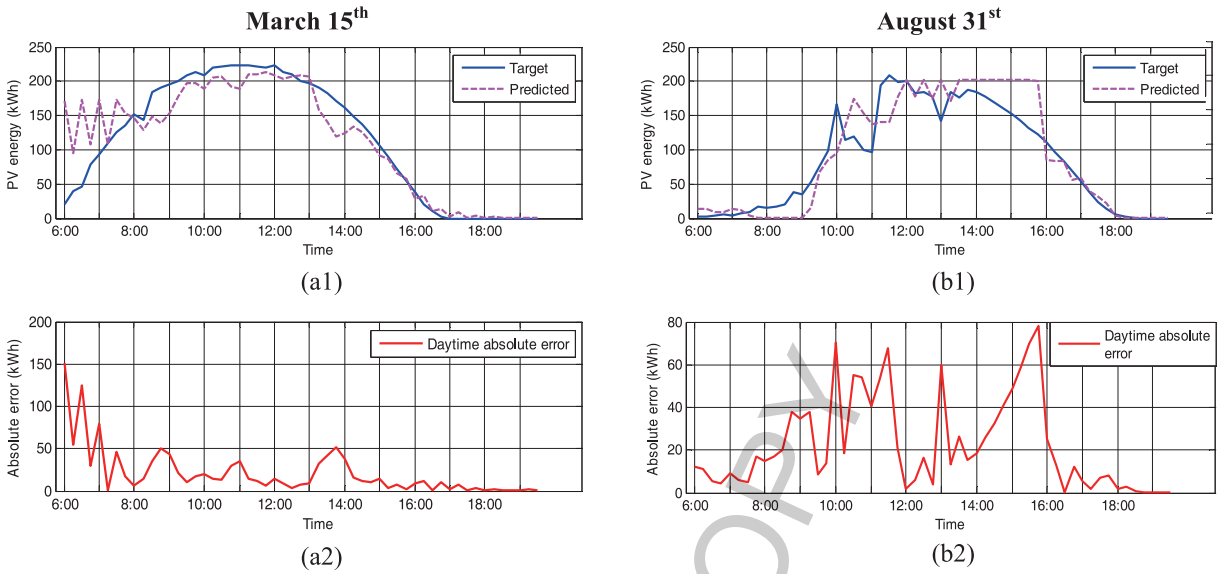

Fig. 6. Results on two *atypical days* chosen at random. (a1), (b1) Comparison between the real energy and the predicted energy. (a2), (b2) Associated daytime absolute instantaneous error. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IDT-2012-0136)

5.1. Prediction of the instantaneous energy

Since the goal of the following experiments is the comparison between the target produced energy and the predicted energy on the continuous daily horizon consisting of samples taken every 15 minutes, we considered the unsigned absolute instantaneous error for all samples of a day.

For the aim of these experiments we need to take into account the presence of "missing" days in the used dataset. A "missing" day is a day in which the system was down or under maintenance. Of course, meteorological changes possibly occurred during missing days cannot be correctly modeled by the forecasting system with the consequence that the prediction of one or more days following a missing one may produce high errors. In the following, we will distinguish two different categories of badly performing days, called, respectively, *atypical* days and *unpredictable* days. The difference between the two types of bad days is, respectively, the absence or presence of at least one missing day in the training window.

Figures 5–7 show the comparison, sample by sample, between the real energy (target) and the predicted energy of some randomly chosen days of the year. More in detail, Fig. 5 regards four days chosen at random among well performing days, Fig. 6 regards two days chosen at random among atypical days, while Fig. 7 concerns two randomly extracted unpredictable days.

Table 2
Comparison between real and predicted accumulated energy

Predicted day	Accumulated energy (kWh)		Absolute percentage error (%)
	Real	Predicted	
February 27th	5110	4915.59	3.80
March 19th	5999	6062.04	1.05
June 1st	6556	6277.71	4.24
July 23rd	5779	5652.25	2.19
March 15th	6455	6414.31	0.63
August 31st	4784	5006.53	4.65
August 20th	5821	7788.51	33.8
October 26th	3254	4617.11	41.9

Each of these figures shows also the unsigned absolute instantaneous error, made on each sample of the day.

We can notice that the performance error in Figs 5(a2), (b2), (c2), (d2) is quite small with respect to the nominal energy production values shown in Figs 5(a1), (b1), (c1), (d1).

In the following Fig. 6 we can see two examples of badly performing days (atypical days), having the predicted energy curve with a quite irregular trend. With the aim of interpreting these results, we checked the data at our disposal and we found out that the bad performance of the atypical days under consideration might be due to a rapid change in the panel temperature values. In fact, if rapid changes in solar radiation or temperature occur during the predicted day, the produced power can sensibly change and the prediction error might increase [6,27].

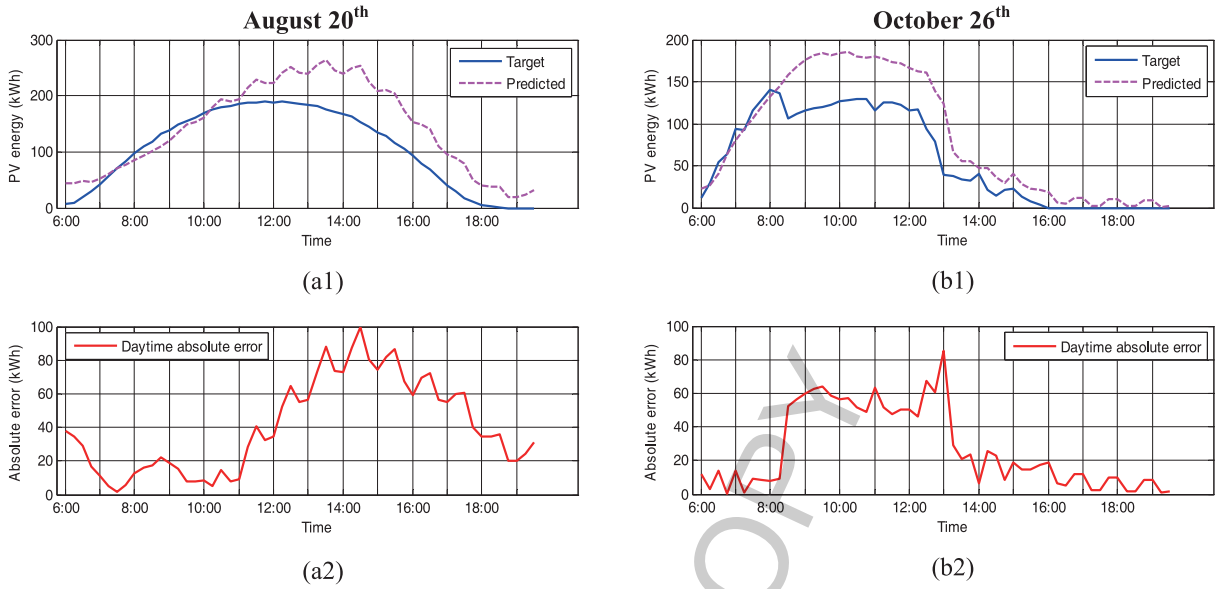

Fig. 7. Results on two *unpredictable days* chosen at random. (a1), (b1) Comparison between the real energy and the predicted energy. (a2), (b2) Associated daytime absolute instantaneous error. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IDT-2012-0136)

Figure 7 shows two examples of badly performing days (unpredictable days), which present a noticeably evident difference between real and predicted energies.

By analyzing the results achieved in this sub-section, we can conclude that, regarding the atypical days, the forecasting performance could be sensibly improved only if we had a trustworthy prediction of changes in the panel temperature values at our disposal. On the other hand, as far as unpredictable days are concerned, improved forecasting performance could be easily achieved by using a complete dataset without missing days.

5.2. *Prediction of the accumulated energy*

Usually in the renewable energy field, and in particular in PV plants, the total (accumulated) produced energy is considered instead of the instantaneous produced energy. So we take into account also the daily accumulated produced energy. Among other things, this typically allows to achieve an error reduction at the end of the day due to a compensation effect.

Table 2 compares the daily accumulated energy values predicted by the network with the real ones for the eight test days referred to above. In addition, we have considered the Absolute Percentage Error (APE) for each test day.

As we can see from Table 2, using the accumulated energy we obtain an acceptable error for all the considered (predictable) days, independently of the dynamic behavior of the curve representing the instantaneous predicted energy, with a maximum value of 4.65% for August 31st.

We can observe that an apparently quite bad day, e.g., the 15th of March (Fig. 6(a)), has actually achieved the best performance among all the eight days (absolute percentage error less than 1%) thanks to the compensation effect.

To evaluate the accuracy of the proposed model, we computed the accumulated energy for each day of the four seasons and we compared it with the target accumulated energy.

Figure 8 shows the comparison between the real and predicted daily accumulated energy, with reference to all days of the four seasons. Please note that the figures may refer to a different number of days due to the lack of data pertinent to days in which the system was down or under maintenance.

From Fig. 8, we can see that the predicted energy generally fits quite well the target energy although error spikes may appear. Once again, the bad performance of the badly performing days might be due to rapid changes in solar radiation or temperature during the predicted day [11,27], or to the presence of missing days in the training window.

Furthermore, we computed the error made on each season as the average of the errors made on the daily accumulated energy values pertinent to the days of that season.

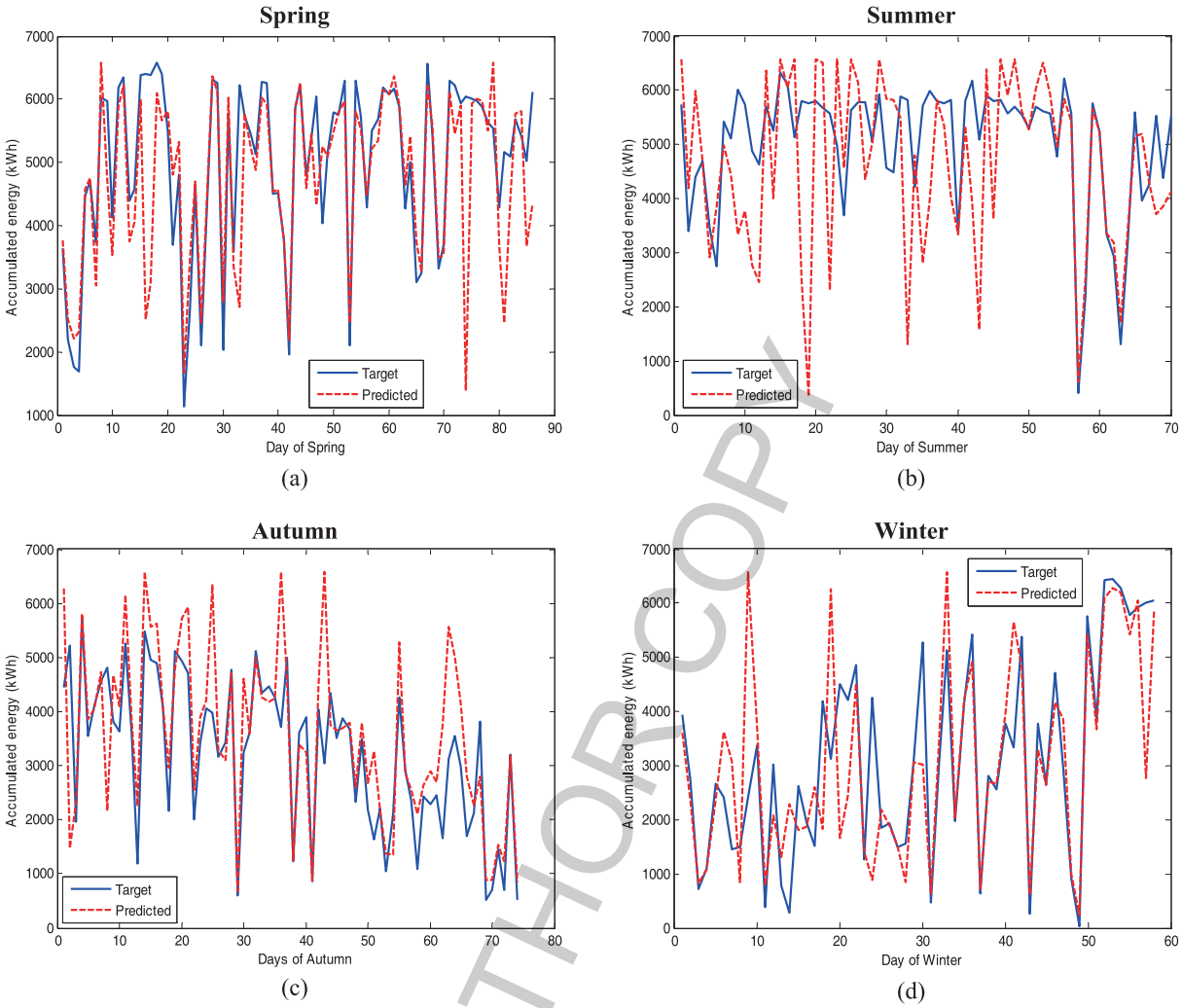

Fig. 8. Target and predicted daily accumulated produced energy for (a) Spring, (b) Summer, (c) Autumn, and (d) Winter. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IDT-2012-0136)

We used the Mean Absolute Percentage Error (MAPE) defined in Eq. (3), where t_i and o_i are, respectively, the target and predicted accumulated energy values of day i, and N is the number of forecasted days of the season:

$$MAPE = \frac{100}{N} \cdot \sum_{i=1}^{N} \frac{|t_i - o_i|}{t_i}. \quad (3)$$

Table 3 compares our method with the classical persistence method, which provides as forecasting value the last known value of the time series. The table shows the seasonal MAPEs made by the persistence method and by the neural model in the prediction of the daily accumulated energy values over all days of each season. It can be seen that the persistence method achieves results significantly worse than our neural network-based NARX model for all the four seasons.

The results achieved by the neural model compare similarly (even though obtained on different datasets and with a different technique) with those obtained in [33], where the average prediction error per day from April to September is about 26% of the measured power.

Finally, with the aim of investigating the seasonal MAPE in Table 3, in Fig. 9 we show the error histograms related to the APE made on the daily accumulated energy for the four seasons.

Each seasonal histogram shows the frequency with which an error value is made in the considered season, i.e., how many days of that season collected that error

Table 3
Comparison between the persistence method and the NARX neural network

Predicted season	MAPE (%) Persistence method	MAPE (%) NARX neural network
Spring	31.16	12.2
Summer	38.13	21.1
Autumn	84.59	26
Winter	77.84	23.9

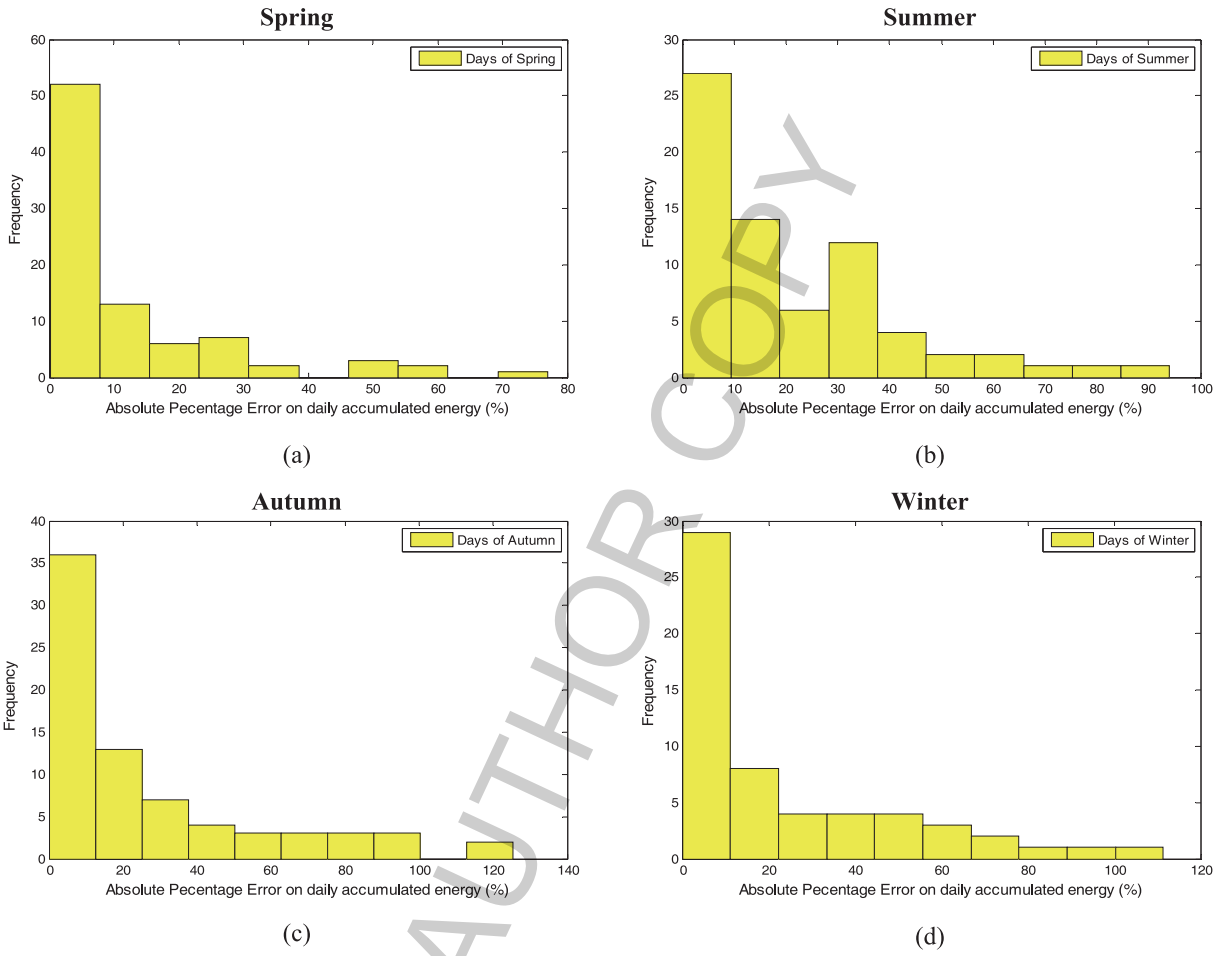

Fig. 9. Error histograms for (a) Spring, (b) Summer, (c) Autumn, and (d) Winter. (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IDT-2012-0136)

value.

Figure 9(a) shows that in 52 Spring days, which represent more than 60% of all the considered Spring days (see Fig. 8(a)), the error made is lower than 10%, while only a few days have errors significantly higher. Similarly, from Figs 9(b), (c), (d), we can see that, respectively, 27, 36 and 29 days (i.e., 39%, 51% and 50% of the considered days for that season, respectively (see Figs 8(b), (c), (d))) produce an error lower than 10%. Regarding higher errors the same considerations as those made for Spring hold.

In order to correctly interpret the results shown above we have tried to identify which are the days that produce the worst errors, e.g., errors higher than 30% for Spring, or higher than 40% for Autumn. We have found out that almost all these days follow immediately (after one or two days) a "missing" day, that is a day in which the system was down or under maintenance. As already stated, missing days are not included in the used dataset, so that their information cannot be used by the forecasting system. As a consequence, the prediction of one or more days following a missing one may

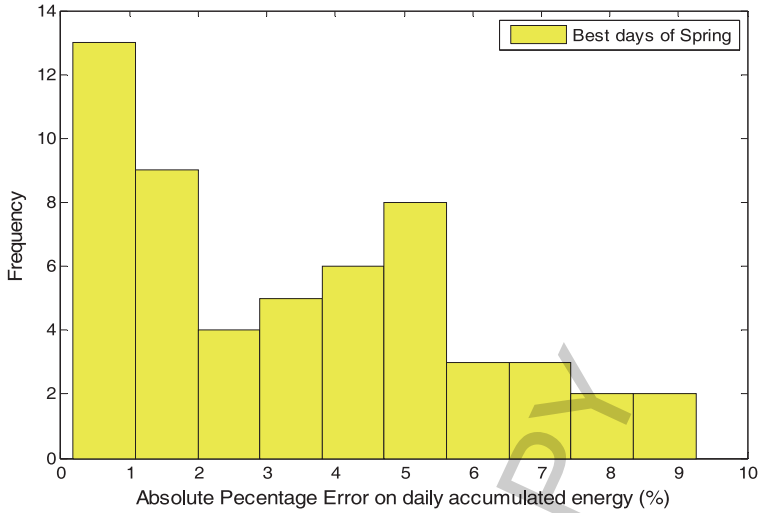

Fig. 10. Error histogram for the best days of Spring (daily error lower than 10%). (Colours are visible in the online version of the article; http://dx.doi.org/10.3233/IDT-2012-0136)

produce high errors, especially when there have been meteorological changes that result not to be correctly modeled.

In Fig. 10 we show the error histogram related to the APE made on the daily accumulated energy for the best days of Spring (those corresponding to an error lower than 10%). Similar histograms (not shown) can be obtained for Summer, Autumn and Winter.

Finally, we tested the developed forecasting system on the second PV installation obtaining practically the same results as those shown above (the differences are so negligible to make the presentation of such results unnecessary).

Based on the previous considerations, we can observe that the compensation effect resulting from the use of the daily accumulated energy makes the developed forecasting system suitable to be effectively and profitably used for one day-ahead forecasting of energy production.

6. Discussion

In this paper we have shown that a powerful nonlinear forecasting technology (neural networks), employed within a NARX model, is able to predict the energy production in a PV installation, provided that (see Table 1):

- the right exogenous inputs are used (irradiance and hour of the day);
- the right NARX model is used (i.e., the correct number of tapped delays d);
- the right neural network structure is used (particularly, the number of hidden neurons h);
- the right training window size (w) is adopted;
- the right sampling frequency (s) is used.

In our experience the other neural network parameters (namely, training algorithm, maximum number of learning epochs, early stopping criterion, etc.) have a lower impact on the forecasting performance.

The proposed methodology does not depend on the characteristics of the specific PV installation, such as geographical location, panels inclination, etc. Its degrees of freedom make it suitable to be applied to any other PV installation.

The forecasting system resulting from the proposed methodology could be profitably used to control the energy distributing grid. Indeed, since PV-based power generators are discontinuous, being influenced by weather conditions, this discontinuity has to be mitigated using alternative power sources, like gas turbines and thermal power plants, which have short start-up time (of the order of a few hours). This means that knowing the energy produced by a grid-connected PV installation 24 hours ahead of time is enough to prepare the start-up of such alternative power sources. The importance of the present study stems exactly from this consideration. In addition, it is widely recognized that an accurate forecasting tool for energy production is a key component of a smart grid, especially when coupled with an energy consumption predictor.

7. Conclusions and future work

In this paper we have presented a general methodology to solve a problem of one day-ahead forecasting of solar energy production. We implemented the time series analysis model NARX by means of a feed-forward neural network with tapped delay lines. In the training process we use the solar irradiation data and the hour as exogenous inputs, and the PV energy production data as endogenous input.

Experimental results have showed that the proposed neural network can faithfully reproduce the curve of daily produced energy so as to predict the daily accumulated energy with seasonal mean absolute percentage errors ranging from a minimum of 12.2% (Spring) to a maximum of 26% (Autumn). These results were achieved despite the presence of missing days (about 21% of the days of one year) in the used training windows. The proposed methodology has been validated by showing that it significantly outperforms the persistence method, a frequently used benchmark in this kind of applications.

Future work will focus on integrating the proposed system with a weather forecasting system, so as to estimate directly the environmental variables. This could be useful when the input data are not available and anyway could increase the prediction performance.

Moreover, due to the good results obtained, we may extend the forecasting horizon to a multiple of the day, so as to strengthen the long-term forecasting analysis.

Finally, we may change the actual forecast time-step of 15 minutes to hourly time-steps.

Acknowledgements

We wish to acknowledge Samares, Via Giuntini, 25, 56023 Navacchio di Cascina, Pisa (Italy), for having provided the set of experimental data used in the present paper.

References

[1] I. Ashraf and A. Chandra, Artificial neural network based models for forecasting electricity generation of grid connected solar PV power plant, *International Journal of Global Energy* **21**(1–2) (2004), 119–130.

[2] A. Azadeh, A. Maghsoudi and S. Sohrabkhani, An integrated artificial neural networks approach for predicting global radiation, *Energy Conversion and Management* **50**(6) (2009), 1497–1505.

[3] T.G. Barbounis, J.B. Theocharis, M.C. Alexiadis and P.S. Dokopoulos, Long-term wind speed and power forecasting using local recurrent neural network models, *IEEE Transactions on Energy Conversion* **21**(1) (2006), 273–284.

[4] M. Benghanem, A. Mellit and S.N. Alamri, ANN-based modelling and estimation of daily global solar radiation data: A case study, *Energy Conversion and Management* **50**(7) (2009), 1644–1655.

[5] E. Cadenas and W. Rivera, Short term wind speed forecasting in La Venta, Oaxaca, Mexico, using artificial neural networks, *Renewable Energy* **34**(1) (2009), 274–278.

[6] D. Caputo, F. Grimaccia, M. Mussetta and R.E Zich, Photovoltaic plants predictive model by means of ANN trained by a hybrid evolutionary algorithm, in *Proc. of the 2010 International Joint Conference on Neural Networks* (*IJCNN '10*), 2010, 1–6.

[7] M. Cococcioni, E. D'Andrea, B. Lazzerini and S.L. Volpi, Short-time forecasting of renewable production energy in solar photovoltaic installations, in *Proc. of 2010 International Conference on Competitive and Sustainable Manufacturing, Products and Services* (*APMS '10*), Como, Italy, 2010.

[8] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen and E. Feitosa, A review on the young history of the wind power short-term prediction, *Renewable and Sustainable Energy Reviews* **12** (2008), 1725–1744.

[9] T. Haida and S. Muto, Regression based peak load forecasting using a transformation technique, *IEEE Transactions on Power Systems* **9**(4) (1994), 1788–1794.

[10] J.D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, NJ, 1994.

[11] J. Ivancheva and E. Koleva, An estimation of the global solar radiation over Bulgaria in *Proc. of International Conference on Water Observation and Information System for Decision Support*, Ohris, Republic of Macedonia, 2008.

[12] S.A. Kalogirou, Applications of artificial neural-networks for energy systems, *Applied Energy* **67**(1–2) (2000), 17–35.

[13] T.L. Kottas, Y.S. Boutalis and A.D. Karlis, New maximum power point tracker for PV arrays using fuzzy controller in close cooperation fuzzy cognitive networks, *IEEE Transactions on Energy Conversion* **21**(3) (2006), 793–803.

[14] M. Kudo, A. Takeuchi, Y. Nozaki, H. Endo and J. Sumita, Forecasting electric power generation in a photovoltaic power system for an energy network, *Electrical Engineering in Japan* **167**(4) (2009).

[15] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling and Z. Yan, A review on the forecasting of wind speed and generated power, *Renewable and Sustainable Energy Reviews* **13**(4) (2009), 915–920.

[16] I.J. Leontaritis and S.A. Billings, Input-output parametric models for non-linear systems Part I: deterministic non-linear systems, *International Journal of Control* **41**(2) (1985), 303–328.

[17] I.J. Leontaritis and S.A. Billings, Input-output parametric models for non-linear systems Part II: stochastic non-linear systems, *International Journal of Control* **41**(2) (1985), 329–344.

[18] R. Marquez and C.F.M. Coimbra, Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database, *Solar Energy* **85**(5) (2011), 746–756.

[19] L. Martin, L.F. Zarzalejo, J. Polo, A. Navarro, R. Marchante and M. Cony, Prediction of global solar irradiance based on time series analysis: Application to solar thermal pow-

er plants energy production planning, *Solar Energy* **84**(10) (2010), 1772–1781.
[20] A. Mellit, M. Benghanem and S.A. Kalogirou, An adaptive wavelet-network model for forecasting daily total solar-radiation, *Applied Energy* **83**(7) (2006), 705–722.
[21] A. Mellit and 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, *Solar Energy* **84**(5) (2010), 807–821.
[22] J. Mubiru and E.J.K.B. Banda, Estimation of monthly average daily global solar irradiation using artificial neural networks, *Solar Energy* **82**(2) (2008), 181–187.
[23] M. Oliver and T. Jackson, The market for solar photovoltaics, *Energy Policy* **27** (1999), 371–385.
[24] C. Paoli, C. Voyant, M. Muselli and M. Nivet, Solar radiation forecasting using ad-hoc time series preprocessing and neural networks, *Solar Energy* **84**(12) (2010), 2146–2160.
[25] C.W. Potter, A. Archambault and K. Westrick, Building a smarter smart grid through better renewable energy information, in *Proc. of 2009 Power Systems Conference and Exposition* (*PSCE '09*), Seattle, WA, U.S., IEEE/PES, 2009, 1–5.
[26] H.M.I. Pousinho, V.M.F. Mendes and J.P.S. Catalão, Neuro-Fuzzy Approach to Forecast Wind Power in Portugal, in *Proc. of 2010 International Conference on Renewables Energies and Power Quality* (*ICREPQ'10*), Spain, 2010, 1–4.
[27] T. Senjyu, H. Takara, K. Uezato and T. Funabashi, One-hour-ahead load forecasting using neural network, *IEEE Transactions on Power Systems* **17**(1) (2002), 113–118.
[28] A. Sfetsos, A novel approach for the forecasting of mean hourly wind speed time series, *Renewable Energy* **27**(2) (2002), 163–174.
[29] Y.D. Song, A new approach for wind speed prediction, *Wind Engineering* **24**(1) (2000), 35–47.
[30] S.I. Sulaiman, T.K.A. Rahman, I. Musirin and S. Shaari, Performance analysis of two-variate ANN models for predicting the output power from grid-connected photovoltaic system, *International Journal of Power, Energy and Artificial Intelligence* (IJPEAI) **2**(1) (2009), 72–76.
[31] A. Troncoso Lora, J.C. Riquelme, J.L. Martínez Ramos, J.M. Riquelme Santos and A. Gómez Expósito, *Influence of kNN-Based Load Forecasting Errors on Optimal Energy Production,* in Lecture Notes in Computer Science, Springer-Verlag Berlin, Germany, 2003, **2902**, Progress in Artificial Intelligence, 189–203.
[32] H. Xie, L. Liu, F. Ma and H. Fan, Performance Prediction of Solar Collectors Using Artificial Neural Networks, in *Proc. of 2009 International Conference on Artificial Intelligence and Computational Intelligence* (*AICI '09*), Shanghai, China, 2009, 573–576.
[33] A. Yona, T. Senjyu, A.Y. Saber, T. Funabashi, H. Sekine and K. Chul-Hwan, Application of neural network to 24-hour-ahead generating power forecasting for PV system, in *Power and Energy Society General Meeting – Conversion and Delivery of Electrical Energy in the 21st Century*, 2008, 1–6.
[34] S.S. Panda, D. Chakraborty and S.K. Pal, Drill wear prediction using neural network architectures, *International Journal of Knowledge-based and Intelligent Engineering Systems* **12**(5–6) (2008), 327–338.