

Received February 18, 2022, accepted March 6, 2022, date of publication March 10, 2022, date of current version March 30, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3158364

A Selective Ensemble Approach for Accuracy Improvement and Computational Load Reduction in ANN-Based PV Power Forecasting

ALFREDO NESPOLI¹, (Graduate Student Member, IEEE),
SONIA LEVA¹, (Senior Member, IEEE), MARCO MUSSETTA¹, (Senior Member, IEEE),
AND EMANUELE GIOVANNI CARLO OGLIARI¹, (Member, IEEE)

Department of Energy, Politecnico di Milano, 20156 Milan, Italy

Corresponding author: Alfredo Nespoli (alfredo.nespoli@polimi.it)

This work was supported in part by the EU Horizon 2020 Project PLATOON under Grant 872592.

ABSTRACT Day-ahead power forecasting is an effective way to deal with the challenges of increased penetration of photovoltaic power into the electric grid, due to its non-programmable nature. This is significantly beneficial for smart grid and micro-grids application. Machine learning and hybrid approaches are well assessed techniques, able to provide effective forecasting with a data-driven approach based on previous measurements from existing power plants. Ensemble methods can be employed to increase solar power forecasting accuracy, by running several independent forecasting models in parallel. In this paper, a novel selective approach is proposed and assessed, where independently trained neural networks are evaluated in terms of accuracy, in order to properly select a suitable forecasting. Moreover, in order to reduce the associated computational burden, suitably developed new normalization approaches are proposed and evaluated. The considered experimental case study shows that the combination of the proposed procedures is able to increase accuracy and to mitigate the overall computational load, resulting in a simple and lightweight algorithm. Additionally, a comparison with other commonly used techniques has shown that the proposed approach is robust with respect to dataset limited size and discontinuities.

INDEX TERMS Artificial neural network, PV power forecasting, ANN sizing, ensemble method, renewable sources, short-term.

I. INTRODUCTION

In the last years, energy system have been progressively shifting towards Renewable Energy Sources (RES) integration in the Smart Grid (SG). In particular, in 2015 the United Nations included among the 17 Sustainable Development Goals (SDGs) one objective regarding energy [1], underlining the need to significantly increase the share of renewables in the global energy mix, together with the overall energy efficiency of the system [2].

Among the RES, Wind turbines and solar photovoltaic (PV) are the most relevant technologies exploiting resources belonging to this category and will be essential to

meet future energy needs while reducing the carbon emissions [3], [4]. Nevertheless, their intrinsic variability causes problems in balancing the energy supply and demand that still have to be fully addressed [5]: thus, more accurate energy forecasting approaches would be helpful for SG and stand-alone micro-grids stability, balancing of demand and supply, integration of distributed generation, monitoring, energy market [6].

PV power production, in particular, strongly relies on meteorological condition. Of particular importance is the solar radiation received by the solar module, that can generally be decomposed into two main contributions: one deterministic, depending on the position of Earth and the Sun, and one stochastic, due to atmospheric conditions and clouds [7].

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

Distributed Energy Management Systems (EMS) are envisaged for an efficient utilization of PV power in smart- and micro-grids, by suitable integration of forecasting and operation planning [8].

In order to perform the PV power forecast, different forecasting models can be implemented. Depending on their nature, they are generally divided into three classes: physical, statistical and hybrid [9].

Physical methods, also referred as parametric or “white box” methods, directly use available weather conditions as input features to a physical model of the PV systems; thus, they strongly rely on the accuracy of the weather forecast they are provided [10] and a proper selection of weather parameters. Furthermore, a complete model of both the system and the surrounding environment is hardly achievable with a high degree of precision.

Data-driven methods, such as Markov chains and exponential smoothing, are statistical approaches, where the underlying model is based on previous measurements rather than on a physical model of the system [11]; thus, they strongly rely on the accuracy of historical data collected. Among data-driven methods, “black box” approaches such as *e.g.* Machine Learning (ML) and Artificial Neural Networks (ANN) have been reported to guarantee high level of accuracy when adopted to the power forecast of PV system [12]. More recently, the use of recurrent neural network models has been also proposed for short-term residential load forecasting in SG applications [13].

Currently, among several data-driven approaches proposed in literature, a great attention was gained by Deep Learning, Convolutional Neural Networks [14] and Long Short-Term Memory (LSTM) networks [15].

Hybrid methods are usually able to suitably combine the previous approaches, actually improving solar forecasting performance based on the historical data by compensating deviations [16]. Several approaches have been successfully proposed in literature: in [17] an improvement in wind power forecasting is obtained by means of independent forecasting models hybridization; in [18], a physical-hybrid-based forecasting model is used to effectively monitor PV modules and to evaluate faults or anomalous trends that may affect the PV plant; in [19], different components of irradiance and additional weather features are employed to improve the input of ML models.

Nowadays, in addition more attention is given to the “ensemble methods” development which improve forecast accuracy and eliminate limitations of single models [20]. Generally, an “ensemble” consists of a set of individually trained models whose predictions are combined when classifying novel instances. [21]

Besides, these ensemble methods are primarily divided into cooperative ensemble where each of the ensemble members performs the same task and their predictions are aggregated to obtain the improved performance [22] or competitive ensemble where, on the contrary, the best prediction of a relevant ensemble member is selected for a particular

input. [23]. Hence, ensemble methods, could be obtained both by using different training sets and by having ANN with different structures and this proved to be effective both ambient temperature [24] and in solar radiation forecast [25].

Recently, ML methods for PV have been based also on LSTM models and aggregation functions, in order to achieve accurate predictions thanks to their recurrent architecture and memory units [26]. Effective strategies for aggregating the predictions of deep learning models are also addressed in [27], exploiting long-term information to properly model solar irradiance fluctuations.

In this view, the authors recently proposed a preliminary attempt to employ several parallel ML models, suitably selected, as an effective way to improve power forecasting accuracy [28]. However, the use of the ensemble approach generally implies reaching a tradeoff between the improved accuracy and the increased computational load.

In the present work, authors will focus on the day ahead PV power prediction by means of ML techniques with an ensemble approach. Here novel methodologies aiming at increasing the forecast accuracy and reducing the computational load are proposed. In particular, accuracy is addressed analyzing the new selective ensemble approach, where several independent models are assessed to properly choose the most effective ensemble. Moreover, in order to reduce the computational burden associated with the proposed ensemble procedure, new normalization approaches are presented and evaluated.

Finally, a suitable experimental case study is considered in order to show the forecasting accuracy improvement and the reduced computational load, thus resulting in a simple, robust and lightweight algorithm, capable of being implemented on an industrial micro-controller.

The present paper is structured as follows: in Section II the here employed physical hybrid approach is presented, with the adopted metrics for performance measurement and a presentation of the case study. In section III, the proposed parallel computing approach is analyzed, specifically addressing the accuracy of the selective ensemble technique. The computational burden issues are addressed in section IV, where suitably defined techniques for pre- and post-processing data are presented to improve the performance. Numerical results will be reported in Section V to validate the proposed approach. Section VI will conclude the paper with a brief discussion of perspective and future research.

II. THE FORECASTING MODEL

Among the previously mentioned approaches, hybrid predicting methods combining deterministic models with ANN have been demonstrated to improve forecasting accuracies [16]. In this work, a simple Multi-Layer Perceptron (MLP) with 12 neurons in the first hidden layer and 5 in the second one is hybridized with the physical Clear-sky Solar Radiation model (CSR) [29]. This method is called Physical Hybrid Artificial Neural Network (PHANN) and it is described in detail in [30], while a simple scheme of its functioning is reported in Figure 1.

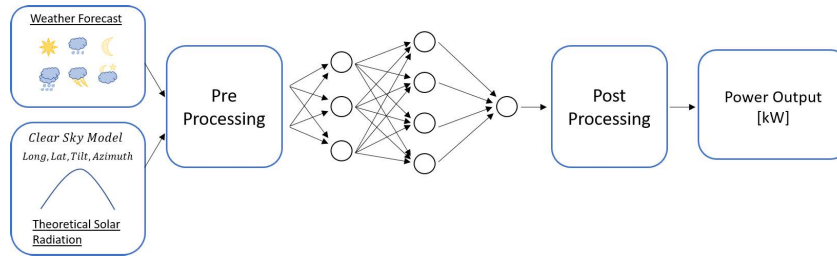


FIGURE 1. ANN forecasting scheme with pre- and post-processing operations.

This ML architecture has been chosen for its simplicity and lightweight in view of its implementation on an industrial micro-controller [18], [31]. This approach was already demonstrated to provide a significant accuracy improvement with respect to both deterministic and standard ML approaches [16], thus providing a reliable day-ahead prediction based on forecasted weather parameters, which are the input of the network trained with the historical dataset.

However, a well-known drawback with neural networks non-linear approximation ability is the possibility to occur in bad generalization. In fact, an ANN properly sized, in terms of number of neurons in the hidden layers, can be correctly trained with an arbitrary training data set; nevertheless, it could learn both investigated dependencies and noise (worsening the predictive ability of the network) giving rise to the well-known issue of *overtraining*, that is when the network learns most likely the gross existing links among parameters first, and then the fine structure that is generated also by noise [32]. This issue can be mitigated by control techniques on the convergence of the ANN model, as described in the following section.

A. CASE STUDY

In order to check the performance and the accuracy of the novel approaches proposed in this work, the study was conducted on experimental data collected from the SolarTech^{LAB} at Politecnico di Milano, Italy, located on the rooftop of Department of Energy [33]. In this Laboratory, different technologies are simultaneously tested and in order to test the modules under different meteorological conditions, each of them is singularly controlled by a dedicated micro-inverter. All PV modules are oriented with an azimuth γ equal to $-6^\circ 30'$, assuming 0° is the South positive West, and a tilt θ of 30° . The tilt of the modules can be modified together with the distance among arrays.

In the order to properly assess the capability of the proposed model, the production from a single monocrystalline module of nominal power of 245 W_p is considered, by adopting its DC power recordings of the year 2017. The PV module and micro-inverter datasheets are reported in Table 1. A publicly available dataset with all these measurements has been previously provided by the authors in [34] as described by [35].

Meteorological forecasts for the following days were collected from a commercial weather service provider every

TABLE 1. PV module and inverter datasheets.

PV module	
P_{MPP}	245 W
Classification Range	-0/+4.99 W
Measurement Accuracy (P_{MPP})	-3/+3 %
U_{MPP}	31.30 V
I_{MPP}	7.84 A
U_{OC}	37.10 V
I_{SC}	8.48 A
U_{max}	1000 V
Inverter	
$V_{DC, max}$	65 V
$V_{DC, MPP}$	12- 60 V
$V_{DC, Full Power}$	30-50V
$I_{DC, max}$	10.5 A
$I_{SC, max}$	12.5 A
$V_{AC, nom}$	230 V
f_{nom}	50 Hz
$P_{AC, nom} \cos \phi = 1$	250 W @ 65 °C amb.
$I_{AC, max}$	1.3 A

day at 11 p.m. The weather parameters considered by the proposed model are summarized in table 2, where they are listed together with the theoretical irradiance in clear-sky conditions, computed according to the deterministic CSRSM, the day of the year (DOY), the hour of the day (h) and the desired target P , i.e. the measured power production in the considered PV plant. All these variables are provided to the network both in the training and the test phases.

Finally, the here considered dataset covers 269 days for the year 2017; however, the described procedure has general validity, and it can be easily extended to different time range, PV modules, power plants and also load forecasting, according to the suitable data set availability.

In order to properly set the forecasting model avoiding overtraining issues, the data were divided into two sub sets, called training and validation sets. They comprises the parameters reported in Table 2, together with the measured power of the PV module, in order to compute errors and evaluate accuracy, as proposed in [36].

In particular, the dataset is divided randomly assigning 90% of data to the training set and 10% to the validation

set. These specific choices and shares have been previously defined through sensitivity analysis conducted in [37].

TABLE 2. Dataset composition for year 2017.

Acronym	Description	UoM
P	DC Power produced by the module (output)	(W)
GHI	Global Horizontal irradiance (input)	(W/m ²)
GPOA	irradiance on the plane of the array (input)	(W/m ²)
T _{amb}	Ambient temperature (input)	(° C)
DOY	Day of the Year (input)	-
h	Hour of the day (input)	-
GHI _{CS}	GHI in clear sky condition (CSRM) (input)	(W/m ²)

B. ACCURACY MEASUREMENT

The effectiveness and the accuracy of the forecasting techniques must be proven by means of evaluation indicators. A wide definition's variety can be found in literature [38], [39] and the most common will be here proposed.

Since the considered model is for day-ahead forecasting with hourly resolution, the root for all the other metrics is the hourly error e_h , defined as:

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

where $P_{m,h}$ is the measured power corresponding to the hour h and $P_{p,h}$ is the prediction provided by one of the forecasting methods. For a fair accuracy assessment and easy identifiable indicators, normalized error have been preferred. Hence, starting from this basic definition, the *normalized mean absolute error* (NMAE), which is the *Mean Absolute Error* (MAE) normalized by the nominal power P_n of the plant and the *normalized root mean square error* (nRMSE) which is defined as the *Root Mean Square Error* normalized by the maximum power recorded in the PV system have been adopted.

Besides, the *weighted mean absolute error* (WMAE) is defined as follows:

$$\text{WMAE} = \frac{\sum_{h=1}^N |e_h|}{\sum_{h=1}^N P_{m,h}} \cdot 100 \quad (2)$$

and it is based on the total energy production in the considered period: when this last data is significantly lower than the predicted one, large values can occur (often significantly above 100%).

Additionally, two novel error metrics have been introduced by the authors in [40]; these are specifically tailored on PV power production in order to provide a more significant evaluation of the overall error behavior.

The *enveloped mean absolute error* (EMAE) is similar to the WMAE, but an upper bound limited to 100% is set by its definition, to avoid the previously described drawback of the standard WMAE indicator:

$$\text{EMAE} = \frac{\sum_{h=1}^N |e_h|}{\sum_{h=1}^N \max(P_{m,h}, P_{p,h})} \cdot 100 \quad (3)$$

Finally, the *objective mean absolute error* (OMAE) whose definition has been specifically designed to introduce an adaptive normalization factor, more representative of the maximum level of irradiance theoretically available in a specific time:

$$\begin{aligned} \text{OMAE} &= \frac{\sum_{h=1}^N |e_h|}{\sum_{h=1}^N G_{POA,h}^{CS}} \cdot \frac{G_{STC}}{P_n} \cdot 100 = \\ &= \text{NMAE} \cdot \frac{N \cdot G_{STC}}{\sum_{h=1}^N G_{POA,h}^{CS}} \end{aligned} \quad (4)$$

where $G_{POA,h}^{CS}$ is the theoretical irradiance on the plane of the array provided by the CSRM for a specific time and G_{STC} is the reference irradiance in standard test condition. This correction factor makes this indicator more related to the theoretical available power in any time of the day and the year, with respect to just the nominal capacity P_n used in NMAE. Night measurements are hence included in the definition of these metrics to properly compare errors in different periods of the year.

III. ENSEMBLE FORECAST

The forecasting technique adopted in this work is based on a ML method, such as ANN. During the network initialization, weights among neurons are firstly randomly assigned and later optimized during training. As highlighted in [41], this process presents many local minima, hence the obtained values for weights can greatly differ from one run to the other, resulting in a differentiation of the errors committed by the networks on different subsets of the input space. For this reason, an "ensemble method" is usually adopted, as reported in [42]. This method consists in performing the forecast of power profile with various models and finally averaging their results to obtain the desired single day profile.

In our work, the considered models for the ensemble are parallel runs of the MLP defined in Section II: in particular, for each run, commonly referred to as "trial", the network's weights and biases are re-initialized and then trained; thus, every considered single MLP is trained on different training and validation datasets, assuring a proper diversity among all the parallel trials. For obvious reasons, the number of computed trials and their respective accuracy are parameters that have a great influence both on the overall prediction reliability and the computational burden, thus, in the following subsection 3.A we will describe how the minimum number of parallel MLPs can be set, while in subsection 3.B we will define the acceptance criteria in order to select a single trial for the considered ensemble.

A. NUMBER OF TRIALS

To evaluate how many MLPs should be trained in parallel in order to have a reliable prediction, several ensembles are considered for every day of the year, each composed of an increasing number of trials, with a maximum of 250. For every single MLP added to the ensemble, the corresponding predicted power profile is then obtained averaging the

current number of trials; the error committed throughout the whole year range is then evaluated for any trial added to the ensemble, and the corresponding metrics are computed as described above. An additional parameter can be determined with respect to each indicator and the number of considered trials, *i.e.* the marginal benefit $M_{\%}$ of increasing the number of trials N_t , *e.g.* for the EMAE indicator it can be defined as:

$$M_{\%} = -\frac{\Delta \text{EMAE}}{\Delta N_t} \cdot 100 \quad (5)$$

It indicates the performance increase (error reduction) obtained for any trial added to an ensemble.

For the sake of generalization, this process has been repeated 8 times, considering 8 independent ensembles of maximum 250 trials each, for a total of 2000 independent trials.

In Figure 2, two graphs are shown: the first one is representative of the error committed in terms of EMAE indicator, with respect to the number of adopted trials. The second graph, on the other hand, shows the $M_{\%}$ as defined in equation (5). In both graphs, the grey lines represents the eight aforementioned independent ensembles, while the orange line their average.

As it is possible to see for each model, accuracy increases (reduction of EMAE) with the number of trials, approaching an asymptotic value:

$$\lim_{N_t \rightarrow \infty} M_{\%} = 0 \quad (6)$$

Nevertheless, as seen in the figure below, the marginal benefit, despite being generally positive, quickly tends toward zero in less than 40 trials. Since the computational load linearly increases with the number of parallel trials (at each repetition, a new network is created, trained and the forecast is computed), a trade off between the time required to predict the power profile and the increase of performance is required. After the preliminary study described above, in the following a number of trials equal to 40 is adopted. Finally, the purpose of the marginal benefit is a criterion to properly define the number of trials to be adopted in the ensemble.

B. TRIALS SELECTION

One of the main advantages of forecasting PV power is that almost half of the desired outputs are known *a priori*. For instance, power production is known to be zero during night hours. Additionally, power production in daytime is expected not to exceed nominal power, modulated by the hourly profile of maximum irradiation, which can be daily determined by the CSRSM. Thus, a mask can be defined daily by setting the maximum and minimum admitted profiles $P_{h,top}$ and $P_{h,bot}$, respectively:

$$P_{h,top} = \frac{G_{POA,h}^{CS}}{G_{STC}} \cdot P_n \quad (7)$$

$$P_{h,bot} = 0 \quad \forall h \quad (8)$$

where $P_{h,top}$ has been defined following the same methodology adopted in (4) for OMAE.

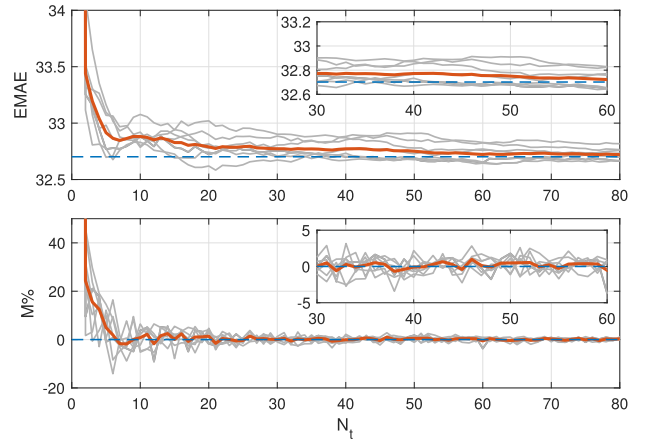


FIGURE 2. Increase of EMAE performance (top) and the marginal benefit $M_{\%}$ with increasing number of trials N_t (bottom).

This valuable information can be used for admitting or rejecting values produced by the forecasting model: this helps validating the accuracy of single outputs obtained by the ANN, which allows an improvement with respect to the previously introduced ensemble logic.

In fact, analyzing the result of each single trial, it is possible to detect the worst performing ones, which can be eliminated in advance without affecting the overall average prediction. For instance, in Figure 3 (top) the forecast of a sample day is presented: the gray lines represent the output of each of the 40 considered trials, the orange line is their average (*i.e.*, the output of the ensemble), the blue line is the actual measurements and finally the yellow shaded area represents the mask defined by the CSRSM. As it is possible to see in the first plot, a few trials present a power profile exceeding the admission mask: these are not consistent, since they strongly differ from the expected value. These unfortunate cases can be explained by the nature of ANN training, based on the Error Back-Propagation (EBP) algorithm, which can end up in sub-optimal solutions. Consequently, those single trials are not providing a good generalization, hence being not suited for the scope.

Starting from these considerations, an accuracy measure can be defined for each trial, defining an exclusion criterion of a single trial based on the error committed with respect to the mask defined by (7) and (8), respectively:

$$P_{err,top} = \sum_h P_{p,h} - P_{h,top}, \quad \forall h | (P_{p,h} > P_{h,top}) \quad (9)$$

$$P_{err,bot} = \sum_h P_{h,bot} - P_{p,h}, \quad \forall h | (P_{p,h} < P_{h,bot}) \quad (10)$$

For each forecasted day, this cumulative error has to be lower than or equal to a control threshold P_{thr} :

$$P_{err,top} + P_{err,bot} \leq P_{thr} \quad (11)$$

It is here worth to underline that this methodology aims at identifying the properly trained network excluding from the ensemble the results when (11) is not verified.

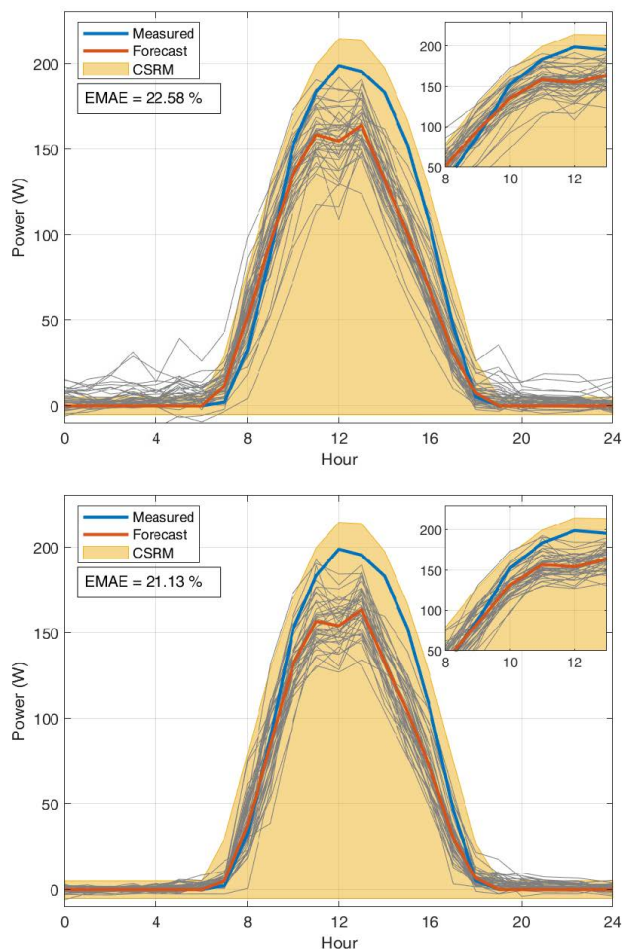


FIGURE 3. Power forecast for a sample day: without (top) and with control threshold (bottom).

Fig. 3 (bottom) presents the power forecast of the same day, after the introduction of the control threshold. As it can be noticed, both the power profiles and the observed EMAE error indicator are significantly improved for that particular day. Thus, with specific reference to the sample day reported in Fig. 3, it is worth highlighting that precision of the forecast also at nighttime is particularly important in order to train the forecasting model to the best precision at sunrise and sunset: indeed, these two conditions are those when the highest relative errors usually occur during the day, due to the small value of produced power, which is usually at the denominator of error indicators, as shown in (2), for instance. Moreover, since PV power forecasting should take care of accuracy not only on average but also in peak hours, a report of the percentage error behavior in different hours of the day (average over the whole year) is reported in Figure 4, with respect to the plant capacity. As shown in this figure, while absolute errors at peak hours are the most relevant in percentage, the selective ensemble approach is effective in reducing them, as reported in the overall distribution during the year, shown in the boxplots: this improvement is clearly highlighted by the bottom histogram, where the average values from the

previous boxplots are compared, showing the higher accuracy due to ensemble selection, in particular during peak hours.

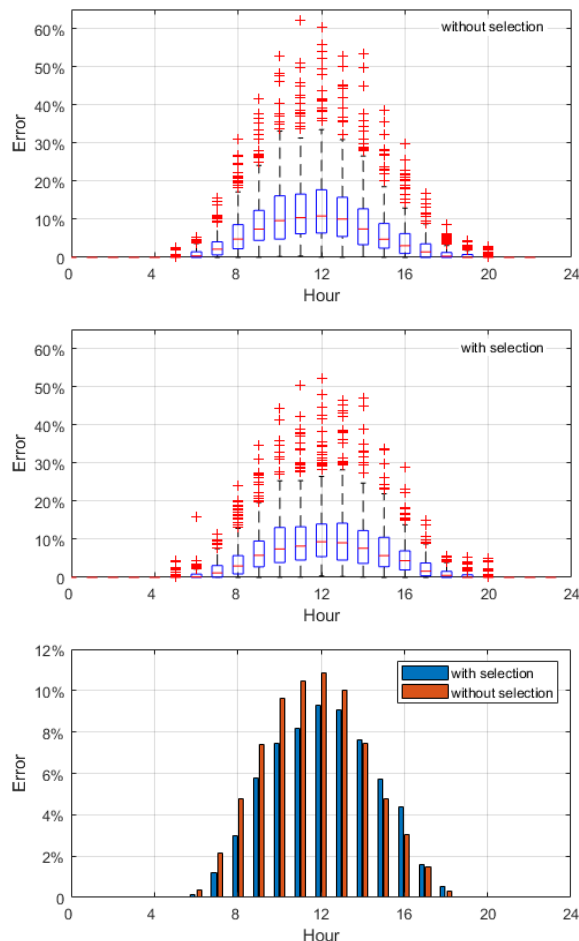


FIGURE 4. Distribution of the hourly error during one year: without ensemble selection (top); with selective ensemble (middle); comparison of the hourly average (bottom).

The overall procedure is synthesized in the flowchart of Figure 5: in order to compose the ensemble of trained MLPs, the output of each trial is compared to P_{thr} to evaluate its accuracy: when the trial satisfies the control threshold, its model is included in the ensemble, otherwise it is discarded and a new MLP is trained and evaluated, until the completion of the ensemble. The ensemble forecast is hence computed *a posteriori*, once the well trained networks have been selected.

While a preliminary implementation of this method was introduced and its overall validity demonstrated in [28] focusing mainly on night hours, in the current paper the control threshold is extended to full day analysis and an additional study is proposed, in the following subsection, to further analyze the performance improvement and mitigate the additional load associated with the computation of several trials.

C. COMPUTATIONAL LOAD

The main disadvantage coming from the implementation of the described method is given by the increased computational

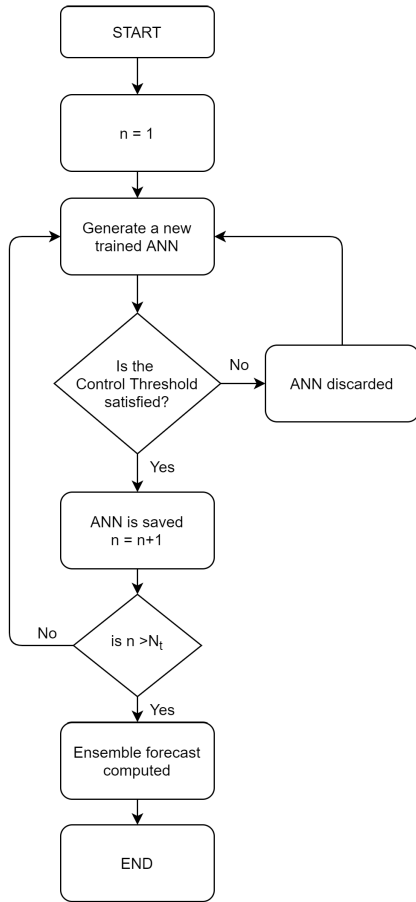


FIGURE 5. Scheme of the proposed selective ensemble approach to reject trials which do not satisfy the control threshold.

burden required to provide the desired number of valid trials: to achieve it, respecting the control threshold, some networks must be trained and later discarded.

In order to quantify the added computational burden, power profile predictions are simulated day-by-day for the whole year 2017, through 250 independent trials; these are later reassembled respecting the logic described above, to obtain a valid ensemble of 40 trials; this study is conducted by varying the value of the threshold P_{thr} in equation (11).

In Figure 6, the blue bars represent the number of trials that was possible to select, on average, with respect to the control threshold P_{thr} , on a yearly basis; the red stacks, on the other hand, represent the number of runs that had to be discarded. Hence, the overall number of networks that, on average, must be trained, is given by the sum of the two. As it is possible to see, reducing P_{thr} , the vast majority of the trials are not compliant and are discarded.

In the lower part of Figure 6, the blue line shows the corresponding computational load, which, as expected, tends to infinity as P_{thr} approaches zero. In fact, during training ANNs learn to generalize trends but not the specific behaviour of the training dataset, as it occurs in case of overtraining.

From the aforementioned considerations it is evident that a trade-off between the accuracy and increased computational

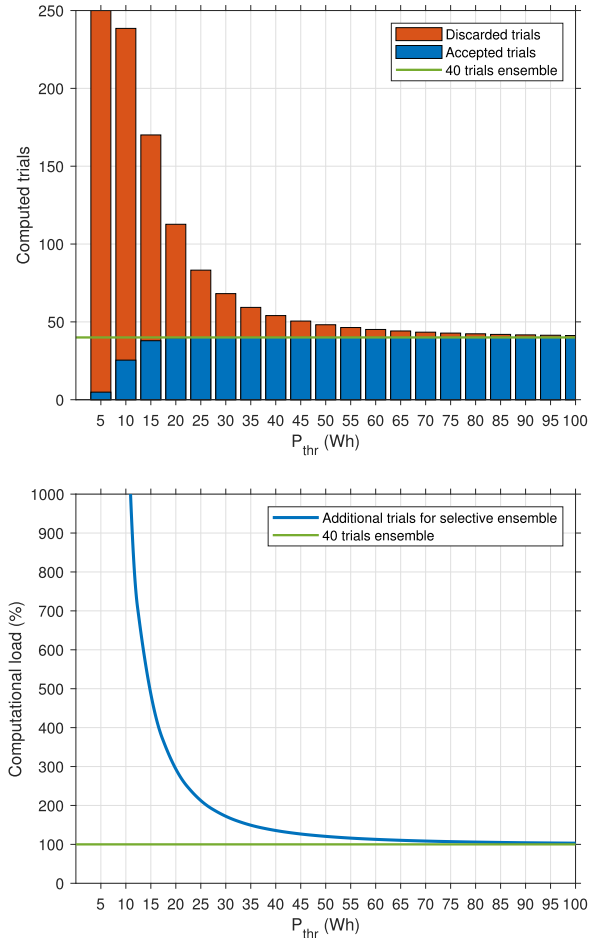


FIGURE 6. Number of accepted and discarded trials with respect to control threshold (top) and corresponding computational load (bottom).

burden is fundamental. For this reason, based also on preliminary studies conducted in [28], a threshold $P_{thr} = 30$ Wh is considered a good compromise and will be used in the followings.

IV. DATA PRE-PROCESSING AND POST-PROCESSING

In the previous section a technique aiming at reducing the overall prediction error was presented by means of selective ensemble approach. Since the required computational load is greatly increased, in the following a few approaches are proposed and analyzed to enhance convergence speed of ANN training, thus reducing the total computational effort needed is now presented.

When dealing with ANN, and adopting a gradient descent method to perform the training process (for example EBP [43]), it is particularly important to pre-process and post-process the inputs and the outputs.

According to [44], convergence is faster when every input variable of the training set has a mean value that is close to zero. In [45], the importance of normalizing data is further explained, in fact it is shown how removing covariance shift from internal activation of the network may aid in the training process.

In the current section, a novel normalization procedure is proposed: as a starting point is always beneficial to shift the mean value according to [44]. Moreover convergence is faster and works better if inputs are scaled so that they have approximately the same covariance C_i . Scaling speeds the learning process because it helps to balance out the rate at which the weights connected to the input nodes learn, as shown in [46].

Data normalization is a fundamental pre-processing step for mining and learning from data; nevertheless, finding the proper approach to deal with time series normalization is not obvious: several normalization methods proposed in literature are valid only on specific time series [47].

In this section, we propose different approaches for normalizing non-stationary time series to be used with ANNs in forecast problems, by means of properly defined data pre- and post-processing, as shown in Fig. 1.

Generally speaking, the covariance value should be chosen with respect to the adopted neuron's activation function. Here, the hyperbolic tangent is used and a covariance equal to 1 is demonstrated to be a good choice. An exception to this guideline can be made when the importance of every input to the output (i.e. for example, the correlation) is known *a priori* [46].

In the following, a step-by-step analysis is conducted on how different input and output processing influence the overall prediction performances, both in terms of committed error and time required to compute the forecast, in order to define suitable data processing to improve ANN efficiency.

A. CORRELATION

The importance of every input to the output is studied by means of the Bravais-Pearson correlation factor [48]:

$$\rho_{X,Y} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}} \quad (12)$$

where X and Y are two generic variables, \bar{x} and \bar{y} their mean value and x_i and y_i a single observation belonging to X and Y respectively.

B. NORMALIZATION

One of the most common way of pre- and post-processing the output is normalizing the variable in a predefined range, usually $[-1; +1]$. Generally speaking, the normalization process is performed according to equation (13),

$$Z' = \frac{Z - Z_{\min}}{Z_{\max} - Z_{\min}} \cdot (Z_{\max}^* - Z_{\min}^*) + Z_{\min}^* \quad (13)$$

where Z is a generic vector of data, whose minimum observation is Z_{\min} and maximum observation is Z_{\max} . Z_{\max}^* and Z_{\min}^* represent the upper and lower bound of the desired range.

C. COVARIANCE

As said, the covariance plays a central role in the ANNs training process. It is defined for generic

vectors X and Y as follows:

$$\begin{aligned} \text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] = \\ &= E[XY] - E[X]E[Y] \end{aligned} \quad (14)$$

where $E[\cdot]$ is the expected value operator.

When the normalization expressed in equation (13) is performed, the covariance coefficient changes consequently following equation (15).

$$\begin{aligned} \text{Cov}(X', Y') &= E \left[\left(\left(\frac{X - X_{\min}}{X_{\max} - X_{\min}} \right) \cdot (X_{\max}^* - X_{\min}^*) + X_{\min}^* + \right. \right. \\ &\quad \left. \left. - E \left[\left(\frac{X - X_{\min}}{X_{\max} - X_{\min}} \right) \cdot (X_{\max}^* - X_{\min}^*) + X_{\min}^* \right] \right) \right. \\ &\quad \cdot \left(\left(\frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} \right) \cdot (Y_{\max}^* - Y_{\min}^*) + Y_{\min}^* + \right. \\ &\quad \left. \left. - E \left[\left(\frac{Y - Y_{\min}}{Y_{\max} - Y_{\min}} \right) \cdot (Y_{\max}^* - Y_{\min}^*) + Y_{\min}^* \right] \right) \right] \end{aligned} \quad (15)$$

and, given the linearity of the expectation operator $E[\cdot]$:

$$\begin{aligned} \text{Cov}(X', Y') &= \\ &= \frac{(X_{\max}^* - X_{\min}^*) \cdot (Y_{\max}^* - Y_{\min}^*)}{(X_{\max} - X_{\min}) \cdot (Y_{\max} - Y_{\min})} \cdot \text{Cov}(X, Y) \end{aligned} \quad (16)$$

As it is possible to notice, the new covariance coefficient is both dependent on the observed and desired upper and lower bounds.

D. ADJUSTED RANGE

When the information about the correlation of the inputs with respect to the output is known, instead of normalizing all the variables in a range of $[-1, +1]$ as it is commonly done, a smarter normalization can be applied: each parameter's range could be adjusted, in fact, in order to obtain the covariance equal to the correlation factor.

As far as the power P is concerned, being the correlation with itself equal to 1, it is possible to rearrange equation (16) and, being the desired range symmetric with respect to the origin, it is possible to write:

$$\begin{cases} (P_{\max}^* - P_{\min}^*) = \sqrt{\frac{\rho_{P,P}}{\text{Cov}_{P,P}}} \cdot (P_{\max} - P_{\min}) \\ (P_{\max}^* + P_{\min}^*) = 0 \end{cases} \quad (17)$$

which allows to find the optimal power range.

As for each of the other i -th selected inputs of the ANN model:

$$\begin{cases} (X_{i,\max}^* - X_{i,\min}^*) = \\ = \frac{\rho_{X_i,P}}{\text{Cov}_{X_i,P}} \cdot \frac{(P_{\max} - P_{\min}) \cdot (X_{i,\max} - X_{i,\min})}{(P_{\max}^* - P_{\min}^*)} \\ (X_{i,\max}^* + X_{i,\min}^*) = 0 \end{cases} \quad (18)$$

E. CONSIDERED NORMALIZATION APPROACHES

Finally, four different approaches can be proposed for data normalization: a detailed description is presented in the following, while their comparison and analysis of numerical results is reported in the next Section.

1) Ap1: BASELINE APPROACH

The first approach (Ap1) can be considered as the baseline, since it does not include any pre- and post-processing of the dataset. When the available data are not processed, they are directly provided to the ANN in order to compute the forecast. In table 3, the above described input and output parameters are given.

TABLE 3. Correlation and Covariance referred to power.

Parameter	Correlation	Covariance	Range
Power	1	5640	[0 , 267.76]
DOY	-0.04	-319	[4 , 365]
Hour	0.15	79	[0 , 23]
T _{amb}	0.37	245	[-1.5 , 37.2]
GHI	0.92	17566	[0 , 957]
GPOA	0.93	22240	[0 , 1037]
CSRM	0.90	24513	[0 , 1036]

2) Ap2: TRADITIONAL NORMALIZATION APPROACH

The second approach (Ap2) normalizes all the data in a range of [-1, +1]. This normalization is the traditionally adopted one, as it is built-in in common ML libraries, and it can be used as a benchmark for the others; Normalizing the available variables between [-1, +1], the consequent covariance coefficients can be computed as in the following equation

$$\text{Cov}(X', Y') = \frac{4 \cdot \text{Cov}(X, Y)}{(X_{\max} - X_{\min}) \cdot (Y_{\max} - Y_{\min})} \quad (19)$$

The obtained values are listed in table 4.

TABLE 4. Covariance and ranges for Ap2.

Parameter	Covariance	Range
Power	0.31	[-1 , +1]
DOY	-0.01	[-1 , +1]
Hour	0.05	[-1 , +1]
T _{amb}	0.09	[-1 , +1]
GHI	0.27	[-1 , +1]
GPOA	0.32	[-1 , +1]
CSRM	0.35	[-1 , +1]

3) Ap3: ADAPTIVE NORMALIZATION APPROACH

The third approach here considered (Ap3) adjusts the normalization range for every available variable to reflect the importance of each input determining the output.

In table 5, the obtained covariances and ranges deriving from equations (17) and (18) are reported.

TABLE 5. Covariance and ranges for Ap3.

Parameter	Covariance	Range
Power	1	[-1.78,+1.78]
DOY	-0.04	[-1.71,+1.71]
Hour	0.15	[-1.66,+1.66]
T _{amb}	0.37	[-2.21,+2.21]
GHI	0.92	[-1.89,+1.89]
GPOA	0.93	[-1.64,+1.64]
CSRM	0.90	[-1.43,+1.43]

4) Ap4: ENHANCED NORMALIZATION APPROACH

The fourth approach here considered (Ap4) is similar to the previous one, but it divides all the obtained ranges by two, to properly match the activation window of the chosen transfer function (hyperbolic tangent). The active range of a transfer function is defined as the range of the input which produces a value for the first derivative significantly different from zero. For the tansigmoid function, the active input range is usually set as [-2, +2]. For this reason the input range of every available parameter is halved with respect to Ap3, resulting in a covariance equal to a fourth of the previous.

It is worth highlighting that the information about the relative importance of each input to the output is retained. Results are shown in table 6.

TABLE 6. Covariance and ranges for Ap4.

Parameter	Covariance	Range
Power	0.25	[-0.89 , +0.89]
DOY	-0.01	[-0.86 , +0.86]
Hour	0.04	[-0.83 , +0.83]
T _{amb}	0.09	[-1.11 , +1.11]
GHI	0.23	[-0.95 , +0.95]
GPOA	0.24	[-0.82 , +0.82]
CSRM	0.23	[-0.72 , +0.72]

V. COMPARISON AND NUMERICAL RESULTS

Following the aforementioned procedures, it was possible to compute the optimal ranges for data processing. In particular, as described in Section II, a simple Multi-Layer Perceptron (MLP) is here adopted, with 6 input neurons (parameters described in Table 2), 1 output neuron (predicted power), 12 and 5 neurons in the first and second hidden layers, respectively. In order to analyse the performance of the four approaches presented above, they will be compared from two different point of view. Firstly, the committed error will be evaluated, secondly, the time required to perform the simulations of the whole considered year is reported, used as a proxy of the computational burden.

A. ACCURACY

The overall results of the combination of the presented normalization approaches with the ensemble forecast, are shown

in table 7, where the computed outcome are filtered with a control threshold P_{thr} equal to 30 Wh/day. In particular, in table 7 the combination of the four considered normalization methodologies to the ensemble selection is compared with the most widely adopted normalization approach when the ensemble selection is disabled (Ap2*); for the sake of clarity, the lowest attained values for each error indicator are highlighted in bold.

TABLE 7. Error metrics comparison among the four approaches, compared to the approach Ap2* when the selective ensemble is not considered (the best results are highlighted in bold).

	Ap2*	Selective Ensemble			
		Ap1	Ap2	Ap3	Ap4
NMAE (%)	3.64	7.47	3.37	3.39	3.36
WMAE (%)	43.67	65.01	42.95	42.99	42.96
EMAE (%)	24.34	35.49	23.20	23.30	23.17
nRMSE (%)	18.88	106.02	17.60	17.55	17.62
OMAE (%)	13.68	27.85	12.36	12.33	12.32

From the performed comparison, it is evident to what extent a correct pre- and post-process of inputs and outputs affects the prediction accuracy: indeed, Ap2, Ap3 and Ap4 show a significant improvement with respect to Ap1 (when pre- and post-processing are avoided). Moreover, the proposed data processing approaches Ap2, Ap3 and Ap4 combined with the selective ensemble present a similar improvement on the forecasting error with respect to the benchmark case Ap2* (when the selective ensemble is not applied).

In order to validate the proposed model results, a comparison with other architectures commonly used to address the PV power forecasting has been conducted, considering in particular CNN [14] and LSTM [15].

For what concerns the adopted CNN approach, the combined use of a convolutional layer and a pooling layer is the commonly considered structure, here implemented with a 1D convolutional layer (32 filters), max pooling layer, flattening layer and dense layer (fully connected, 20 neurons); simulations were performed under different input sequences, considering T_{amb} (a), GHI (b), GPOA (c) and wind speed (d), as reported in Table 8.

For what concerns the adopted LSTM approach, a sliding window forecast was assumed, considering different units, hidden layers and dropout: in Table 8, the considered models are indicated as (*units* \times *hidden layers* \times *dropout*).

After this comparison, it is possible to conclude in general terms that the CNN and LSTM are not well suited for the data distribution available in the considered case study, due, for instance, to the overall size, probably too small for CNN approach, or to some missing days in the dataset, which break the continuity of data flow needed by LSTM: these are common issues when using real data from PV plant, which usually are not easily available, and often subject to interruptions. Indeed, as reported in Table 8, the forecasting computed by

these two approaches have a similar behavior in terms of accuracy, but slightly above the error level reached by the presented PHANN with the selective ensemble: apparently, the here proposed approach is simple and lightweight, better suited to address the day-ahead power forecast based on numerical weather predictions, and more robust with respect to the dataset limitations presented above.

TABLE 8. Comparison with other forecasting architectures (the best results are highlighted in bold).

Approach	NMAE (%)	nRMSE (%)	EMAE (%)
CNN (a,b,c,d)	4.05	24.21	30.08
CNN (a,b,d,c)	4.31	24.98	31.44
CNN (a,c,b,d)	4.70	27.34	34.38
CNN (a,c,d,b)	4.25	25.45	31.60
LSTM (24 \times 1 \times 0.1)	5.19	30.61	38.24
LSTM (24 \times 2 \times 0.1)	4.91	29.93	36.86
LSTM (36 \times 1 \times 0.1)	5.20	31.16	38.66
LSTM (36 \times 2 \times 0.1)	5.02	31.92	38.60
LSTM (48 \times 1 \times 0.1)	5.11	34.33	40.58
LSTM (48 \times 2 \times 0.1)	5.10	31.53	38.59
LSTM (72 \times 1 \times 0.1)	5.11	30.01	37.57
LSTM (72 \times 2 \times 0.1)	4.90	30.73	37.38
sel. ensemble (Ap2)	3.37	17.60	23.20
sel. ensemble (Ap3)	3.39	17.55	23.30
sel. ensemble (Ap4)	3.36	17.62	23.17

B. COMPUTATIONAL BURDEN

Additionally, it is possible to analyze the overall computational time taken by each approach for a whole year forecasting of daily power profiles (with 40 effective trials): Table 9 reports the results of simulations performed on an Intel(R) i7-7700 CPU @3.60 GHz - 64 GB ram.

TABLE 9. Computational burden for the Selective Ensemble with different normalization approaches.

	Ap1	Ap2	Ap3	Ap4
Computed trials	68	68	68	68
Accepted trials	40	40	40	40
Discarded trials	28	28	28	28
Total time (min)	160.3	83.9	73.9	69.3

These results show how pre-processing the available dataset is relevant to reduce the computational burden. In fact, the two newly proposed adaptive and enhanced normalization approaches (Ap3 and Ap4, respectively) allow to significantly reduce the time required to train the network with respect to the most commonly used normalization procedure Ap2.

In particular, the newly proposed enhanced normalization approach (Ap4) guarantees a reduction in the computational effort required because, during the first iteration, when weights are randomly picked, it allows the transfer function

to work in a region of the domain where its derivative is greater and the optimization algorithm can proceed faster.

Moreover, in table 10, a comparison of the computational load of the different normalization procedures is reported, considering the combination of the here introduced selective ensemble approach with the control threshold P_{thr} set to 30 Wh/day.

TABLE 10. Comparison of computational burden of each approach (column) with respect to the others (rows).

	Ap1	Ap2	Ap3	Ap4
Ap1	–	-48%	-54%	-57%
Ap2	+91%	–	-12%	-17%
Ap3	+117%	+14%	–	-6%
Ap4	+131%	+21%	+7%	–

As it was possible to see from Table 7, all the error metrics present a significant reduction, when the ensemble selection is enabled, with respect to the scenario considered as a benchmark (Ap2*). Nevertheless, the newly proposed enhanced normalization procedure (Ap4) is significantly the fastest in terms of convergence speed of the ANN model, with respect to the other traditional (Ap2) and adaptive (Ap3) normalization approaches, although having a comparable accuracy.

VI. CONCLUSION

In this paper, the combination of two novel approaches aimed at improving forecasting accuracy and computational efficiency has been analyzed, applied to the day-ahead PV power prediction.

First, a selective ensemble methodology has been proposed, in order to select the most promising trained MLPs for the ensemble forecast. This approach led to a significant reduction of the overall error committed, with respect to several considered error evaluation metrics (i.e. about 1 percentage point, on average).

Additionally, through the correlation and covariance analysis of the available dataset, it was possible to obtain the optimal normalization ranges of ANN. By applying the newly proposed adaptive and enhanced normalization approaches, it was possible to get a relevant mitigation (i.e. -17%) of additional computational load associated to ensemble selection, with respect to the traditionally adopted normalization procedure.

Future perspective of this study envisage the possibility to conduct additional adaptive data analysis to further increase the computational efficiency of the proposed combined approach and to apply it to the forecasting of power production and load consumption in challenging micro-grid management applications.

REFERENCES

- [1] (2015). *United Nation Development Programme: Sustainable Development Goals. Goal 7: Affordable and Clean Energy*. [Online]. Available: <http://www.un.org/sustainabledevelopment/energy/>
- [2] United Nations Development Programme. *Goal 7 Targets*. Accessed: Jan. 15, 2022. [Online]. Available: <http://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-7-affordable-and-clean-energy/targets/>
- [3] L. Bird, M. Milligan, and D. Lew, "Integrating variable renewable energy: Challenges and solutions," Nat. Renew. Energy Lab., Golden, CO, USA, Tech. Rep. NREL/TP-6A20-60451, Sep. 2013. [Online]. Available: <http://www.nrel.gov/docs/fy13osti/60451.pdf>
- [4] G. Carcangiu, C. Dainese, R. Faranda, S. Leva, and M. Sardo, "New network topologies for large scale photovoltaic systems," in *Proc. IEEE Bucharest PowerTech*, Jun. 2009, pp. 1–7.
- [5] Y. Zhang, M. Beaudin, R. Taheri, H. Zareipour, and D. Wood, "Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators," *IEEE Trans. Smart Grid*, vol. 6, no. 5, pp. 2253–2262, Sep. 2015.
- [6] T. Carriere, C. Vernay, S. Pitaval, and G. Kariniotakis, "A novel approach for seamless probabilistic photovoltaic power forecasting covering multiple time frames," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2281–2292, May 2020.
- [7] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Sol. Energy*, vol. 136, pp. 78–111, Oct. 2016, doi: [10.1016/j.solener.2016.06.069](https://doi.org/10.1016/j.solener.2016.06.069)
- [8] Y. Fujimoto, H. Kikusato, S. Yoshizawa, S. Kawano, A. Yoshida, S. Wakao, N. Murata, Y. Amano, S.-I. Tanabe, and Y. Hayashi, "Distributed energy management for comprehensive utilization of residential photovoltaic outputs," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1216–1227, Mar. 2018.
- [9] R. Ulbricht, U. Fischer, W. Lehner, and H. Donker, "First steps towards a systematical optimized strategy for solar energy supply forecasting," in *Proc. 1st Int. Workshop Data Anal. Renew. Energy Integr. (DARE)*, 2013, pp. 14–25.
- [10] D. Yang, J. Kleissl, C. A. Gueymard, H. T. C. Pedro, and C. F. M. Coimbra, "History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining," *Sol. Energy*, vol. 168, pp. 60–101, Jul. 2018, doi: [10.1016/j.solener.2017.11.023](https://doi.org/10.1016/j.solener.2017.11.023)
- [11] R. Blaga, A. Sabadus, N. Stefu, C. Dughir, M. Paulescu, and V. Badescu, "A current perspective on the accuracy of incoming solar energy forecasting," *Progr. Energy Combustion Sci.*, vol. 70, pp. 119–144, Jan. 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0360128518300303>
- [12] S. Leva, A. Dolara, F. Grimaccia, M. Mussetta, and E. Ogliari, "Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power," *Math. Comput. Simul.*, vol. 131, pp. 88–100, Jan. 2015, doi: [10.1016/j.matcom.2015.05.010](https://doi.org/10.1016/j.matcom.2015.05.010)
- [13] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.
- [14] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network," *Appl. Energy*, vol. 251, Oct. 2019, Art. no. 113315. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0306261919309894>
- [15] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," *Energy*, vol. 148, pp. 461–468, Apr. 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0360544218302056>
- [16] E. Ogliari, A. Dolara, G. Manzolini, and S. Leva, "Physical and hybrid methods comparison for the day ahead PV output power forecast," *Renew. Energy*, vol. 113, pp. 11–21, Dec. 2017, doi: [10.1016/j.renene.2017.05.063](https://doi.org/10.1016/j.renene.2017.05.063)
- [17] J. Shi, Z. Ding, W.-J. Lee, Y. Yang, Y. Liu, and M. Zhang, "Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 521–526, Jan. 2014.
- [18] S. Leva, M. Mussetta, and E. Ogliari, "PV module fault diagnosis based on microconverters and day-ahead forecast," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3928–3937, May 2019.
- [19] J. Wang, H. Zhong, X. Lai, Q. Xia, Y. Wang, and C. Kang, "Exploring key weather factors from analytical modeling toward improved solar power forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1417–1427, Mar. 2019.
- [20] A. Telikani, A. Tahmassebi, W. Banzhaf, and A. H. Gandomi, "Evolutionary machine learning: A survey," *ACM Comput. Surv.*, vol. 54, no. 8, pp. 1–35, Nov. 2022, doi: [10.1145/3467477](https://doi.org/10.1145/3467477)
- [21] D. Opitz and R. Maclin, "Popular ensemble methods: An empirical study," *J. Artif. Intell. Res.*, vol. 11, pp. 169–198, Aug. 1999.

- [22] N. Dvornik, J. Mairal, and C. Schmid, "Diversity with cooperation: Ensemble methods for few-shot classification," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3723–3731.
- [23] Y. Frayman, B. F. Rolfe, and G. I. Webb, "Solving regression problems using competitive ensemble models," in *Proc. Austral. Joint Conf. Artif. Intell.* Berlin, Germany: Springer, 2002, pp. 511–522.
- [24] A. Ahmadi, Z. Zargarán, A. Mohebi, and F. Taghavi, "Hybrid model for weather forecasting using ensemble of neural networks and mutual information," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 3774–3777.
- [25] R. Jovanovic, L. M. Pomares, Y. E. Mohieldeen, D. Perez-Astudillo, and D. Bachour, "An evolutionary method for creating ensembles with adaptive size neural networks for predicting hourly solar irradiance," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 1962–1967.
- [26] M. Abdel-Nasser, K. Mahmoud, and M. Lehtonen, "Reliable solar irradiance forecasting approach based on choquet integral and deep LSTMs," *IEEE Trans. Ind. Informat.*, vol. 17, pp. 1873–1881, 2021.
- [27] M. Abdel-Nasser, K. Mahmoud, and M. Lehtonen, "HIFA: Promising heterogeneous solar irradiance forecasting approach based on kernel mapping," *IEEE Access*, vol. 9, pp. 144906–144915, 2021.
- [28] S. Leva, M. Mussetta, A. Nespoli, and E. Ogliari, "PV power forecasting improvement by means of a selective ensemble approach," in *Proc. IEEE Milan PowerTech*, Jun. 2019, pp. 1–5.
- [29] R. E. Bird, "A simple, solar spectral model for direct-normal and diffuse horizontal irradiance," *Sol. Energy*, vol. 32, no. 4, pp. 461–471, 1984.
- [30] A. Dolara, F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, "A physical hybrid artificial neural network for short term forecasting of PV plant power output," *Energies*, vol. 8, no. 2, pp. 1138–1153, Feb. 2015.
- [31] A. Nespoli, M. Mussetta, E. Ogliari, S. Leva, L. Fernández-Ramírez, and P. García-Triviño, "Robust 24 hours ahead forecast in a microgrid: A real case study," *Electronics*, vol. 8, no. 12, pp. 1–13, 2019.
- [32] I. V. Tetko, D. J. Livingstone, and A. I. Luik, "Neural network studies. 1. Comparison of overfitting and overtraining," *J. Chem. Inf. Comput. Sci.*, vol. 35, no. 5, pp. 826–833, 1995.
- [33] *SolarTech Lab: Forecasting*. Accessed: Sep. 1, 2021. [Online]. Available: <http://www.solartech.polimi.it/activities/forecasting/>
- [34] S. Leva, A. Nespoli, S. Pretto, M. Mussetta, and E. Ogliari, "Photovoltaic power and weather parameters," *IEEE Dataport*, Sep. 2020, doi: [10.21227/42v0-jz14](https://doi.org/10.21227/42v0-jz14).
- [35] S. Leva, A. Nespoli, S. Pretto, M. Mussetta, and E. G. C. Ogliari, "PV plant power nowcasting: A real case comparative study with an open access dataset," *IEEE Access*, vol. 8, pp. 194428–194440, 2020.
- [36] F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, "ANN sizing procedure for the day-ahead output power forecast of a PV plant," *Appl. Sci.*, vol. 7, no. 6, p. 622, Jun. 2017.
- [37] A. Dolara, F. Grimaccia, S. Leva, M. Mussetta, and E. Ogliari, "Comparison of training approaches for photovoltaic forecasts by means of machine learning," *Appl. Sci.*, vol. 8, no. 2, p. 228, Feb. 2018. [Online]. Available: <http://www.mdpi.com/2076-3417/8/2/228>
- [38] C. Monteiro, T. Santos, L. Fernandez-Jimenez, I. Ramirez-Rosado, and M. Terreros-Olarte, "Short-term power forecasting model for photovoltaic plants based on historical similarity," *Energies*, vol. 6, no. 5, pp. 2624–2643, May 2013.
- [39] J. Kleissl, C. Coimbra, and R. Marquez, "Overview of solar forecasting methods and a metric for accuracy evaluation," 2013, doi: [10.1016/B978-0-12-397117-7.00008-5](https://doi.org/10.1016/B978-0-12-397117-7.00008-5).
- [40] A. Nespoli, E. Ogliari, A. Dolara, F. Grimaccia, S. Leva, and M. Mussetta, "Validation of ANN training approaches for day-ahead photovoltaic forecasts," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–6.
- [41] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 10, pp. 993–1001, Oct. 1990.
- [42] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, vol. 1. Cambridge, MA, USA: MIT Press, 1986.
- [43] G. O. Stone, "An analysis of the delta rule and the learning of statistical associations," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1. Cambridge, MA, USA: MIT Press, 1986, pp. 444–459.
- [44] R. Rojas, *Neural Networks: A Systematic Introduction*. Springer, 2013.
- [45] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 1, Jul. 2015, pp. 448–456.
- [46] Y. A. LeCun, L. Bottou, G. B. Orr, and K. R. Müller, "Efficient BackProp," in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 9–48.
- [47] E. Ogasawara, L. C. Martinez, D. de Oliveira, G. Zimbrao, G. L. Pap, and M. Mattoso, "Adaptive normalization: A novel data normalization approach for non-stationary time series," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2010, pp. 1–8.
- [48] P. Courriou, "Three algorithms for estimating the domain of validity of feedforward neural networks," *Neural Netw.*, vol. 7, no. 1, pp. 169–174, Jan. 1994.



ALFREDO NESPOLI (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in energy engineering from the Politecnico di Milano, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. Since then, he has been working as a Temporary Researcher with the Electrical Division, Energy Department, Politecnico di Milano. His research interests include AI and renewable energy sources.



SONIA LEVA (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Politecnico di Milano, Italy, in 2001. She is currently a Full Professor of electrical engineering with the Department of Energy, Politecnico di Milano. She is also the Director of the Solar Tech Laboratory and the Laboratory of Microgrids (MG2LAB), Politecnico di Milano. She is a member of the IEEE Working Group "Distributed Resources: Modeling and Analysis" and a Senior Member of the IEEE Power and Energy Society.



MARCO MUSSETTA (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Politecnico di Milano, Italy, in 2007. He is currently an Associate Professor of electrical engineering with the Department of Energy, Politecnico di Milano. He is the Chair of the IEEE Computational Intelligence Society (CIS) Task Force on "Fuzzy Systems in Renewable Energy and Smart Grid." He is a Senior Member of the IEEE CIS, PES, and IES.



EMANUELE GIOVANNI CARLO OGLIARI (Member, IEEE) received the M.Sc. degree in electrical engineering and the Ph.D. degree in electrical engineering from the Politecnico di Milano, Italy, in 2016, with a thesis titled: "Novel methods for photovoltaic energy forecasting of the day-ahead hourly power curve." He is currently an Assistant Professor with the Department of Energy, Politecnico di Milano, where he teaches electrical engineering. Since 2010, he has been working on photovoltaic power plant design and their optimization. Since 2020, he has been working on RES expected power by means of computational intelligence techniques.

• • •