Progress In Electromagnetics Research C, Vol. 125, 35–49, 2022

A Two-Step Learning-by-Examples Method for Photovoltaic Power Forecasting

Alessandro Polo*

Abstract—In this paper, an innovative machine learning (ML) approach for the prediction of the output power generated by photovoltaic (PV) plants is presented. Toward this end, a two-step learning-by-examples (LBE) strategy based on support vector regression (SVR) is proposed to learn the complex relation among the heterogeneous parameters affecting the energy production of the power plant. More specifically, the first step is aimed at down-scaling the weather forecasts from the standard air temperature and the solar irradiance to the local module temperature and the plane-of-array (POA) irradiance. Then, the second step predicts the output power profile given the down-scaled forecasts estimated at the previous step. The advantages and limitations of the proposed two-step approach have been experimentally analyzed exploiting a set of measurements acquired in a real PV plant. The obtained results are presented and discussed to point out the capabilities of the proposed LBE method to provide robust and reliable power predictions starting from simple weather forecasts.

1. INTRODUCTION

Nowadays, the increasing cost of energy production through fossil fuels and a renewed attention towards environmental issues have led worldwide Governments to introduce new laws in order to increase energy efficiency and decrease the environmental impact of energy production. In this framework, the United Nations (UN) has recently identified within its "2030 Agenda for Sustainable Development" goals urgently addressing paramount issues such as (i) affordable and clean energy (i.e., access to renewable, safe and widely available energy sources), (ii) sustainable cities and communities (i.e., making cities safe, inclusive, resilient, and sustainable), (iii) responsible consumption and production (i.e., reversing current consumption trends and promoting a more sustainable future), and (iv) climate action (i.e., regulating and reducing emissions and promoting renewable energy) [1]. Consequently, the diffusion of renewable energy plants exploiting different energy sources (e.g., wind, sunlight, tides, geothermal heat) has been consistently increasing in the last years. For instance, the European Member States have recently developed different strategies in order to fully exploit their territorial characteristics and maximize the energy production from renewable sources. As an example, Denmark is producing about 30% of the energy demand using wind-farms. The Netherlands have concentrated their production on off-shore wind plants. Concerning the Mediterranean countries such as Italy, Spain, and Greece, the photovoltaic (PV) renewable energy source has been largely adopted to exploit the high solar radiation on the ground, which is almost twice compared to the north European countries. As an example of legislation in favour of the renewable energies, the Italian Authority for Electrical Energy and Gas ($AEEG$) introduced in 2007 an economic incentive proportional to the PV plant size and power in order to stimulate the diffusion of PV and solar thermodynamic plants, also at the residential level. Moreover, the legislation introduced the possibility to sell the produced energy at the market price. As

Received 10 June 2022, Accepted 9 September 2022, Scheduled 4 October 2022
* Corresponding author: Alessandro Polo (alessandro.polo.1@unitn.it).
The authors is with the ELEDIA Research Center (ELEDIA@UniTN, University of Trento), DICAM — Department of Civil, Environmental, and Mechanical Engineering, University of Trento, Via Mesiano 77, Trento 38123, Italy.

a direct consequence, the number of PV plants rises significantly, and such an increase of PV energy production has caused deep changes in the Italian grid management. Since the PV systems are sensitive to the fluctuation of the weather conditions, the uncertainty of the power generation may impact on the stability and reliability of the distribution grid. In order to avoid such a drawback, the energy generation of conventional power plants (e.g., coal, hydro-electric, nuclear) has to be well managed for compensation purposes. Regarding the Italian example, such a complex balancing is addressed in the so-called *day-before market*, where the energy stakeholders buy and sell the "energy slots" one day in advance. This approach enables a balanced energy supply to the grid, the reduction of the energy waste, and the maximum exploitation of the green energy sources. Accordingly, the prediction of the output power generated by weather-dependent plants is requested by the $AEEG$ at least one day before and with a error lower than 10% [2]. Moreover, the output power forecasting is mandatory for energy plants with nominal power higher than 200 [kW].

Artificial intelligence (AI) and machine learning (ML) are promising candidates to effectively address the PV forecasting problem [3]. More specifically, the learning-by-examples (LBE) paradigm encompasses different supervised learning methodologies [e.g., support vector machines (SVMs), artificial neural networks (ANNs), Gaussian processes (GPs), and deep neural networks (DNNs)] enabling the creation of accurate predictors starting from the information embedded within an off-line generated database of training examples/observations [4–17]. In such a framework, different strategies have been proposed in the state-of-the-art to predict the output power of PV plants [3]. On the one hand, the so-called *indirect* methods [18, 19] are based on the prediction of the solar irradiance exploiting historical irradiance data in combination with weather forecasts. Successively, the irradiance is adopted for the estimation of the output power. As an example, Lorenz et al. [20] derived the power predictions from a refined irradiance forecast computed with the physical modeling and simulation of PV plants. The regional power production is then obtained by means of up-scaling strategies. However, up-scaling procedures are suitable for regions characterized by morphological homogeneity, like the German test fields considered in [20] characterized by uniform and stable weather conditions. It should also be noticed that the behavior of the solar irradiance is very complex, and its prediction is challenging. The existing numerical models like the clear-day radiation model or the half-sine hour-by-hour solar model [21] are often customized and difficult to be widely adopted. LBE methods such as ANNs [22] or support vector machines SVMs [23] are powerful tools to exploit the statistical properties of the irradiance time series.

On the other hand, *direct* methods [24–27] are aimed at estimating the power production considering the relation between the status of the considered PV plant and the available weather information. In this context, Shi et al. [28, 29] proposed a power forecast strategy considering a set of predefined weather models (i.e., sunny, foggy, cloudy, and rainy) and using a support vector regression (SVR) method trained with the measured solar irradiance, air temperature, and PV output power. The main drawback of such a predefined weather model is the reduced flexibility in dealing with unpredictable changes of weather conditions during the day. Similar results have been achieved by Fei et al. [25] using an ANN trained with multiple time series of the output power, average air temperature, and clear sky index. However, sub-optimal predictions are achieved using the air temperature, since the power production is more closely related to the PV module temperature. Approximated equations exist to compute the PV module temperature from solar irradiance and air temperature, but they are usually adopted only for quantitative analyses during the performance assessment of the PV modules.

Following this line of reasoning, the objective of this work is to develop a ML methodology for predicting the output power profile of a PV plant exploiting only the standard weather forecast of the considered area. Towards this end, an innovative two-step LBE approach based on SVR is proposed in order to (i) accurately downscale the available weather forecast information to the plane-of-array (POA) irradiance and PV module temperature, and to (ii) predict the output power profile starting from the estimated weather-related PV status and considering the complex non-idealities of the power conversion process, such as the nonlinear behavior of the inverters, cable connection losses, impedance mismatch, and module orientation angles.

Therefore, the main novelty of this work over the existing literature lies in the development of an innovative two-step SVR-based methodology capable of predicting the output PV power using common weather forecasts made available by commercial services. To the best of the author's knowledge, this is

the first attempt to apply *SVR*s to the prediction of the PV power in any weather condition. As a matter of fact, different from available approaches where daily weather conditions are commonly grouped (e.g., sunny, cloudy, etc.) in order to select the customized trained model to improve the performance of the statistical-based forecasts [23, 24, 35–38], thanks to the adopted downscaling step the proposed method is particularly flexible and usable when a weather forecast service is available.

The paper is organized as follows. After the definition of the output power as a function of the weather conditions, the proposed two-steps *ML* strategy is formulated in Section 2. The experimental validation performed on a real PV power plant is presented and discussed in Section 3 to point out the prediction performance of the proposed strategy. Finally, the conclusions are summarized in Section 4.

2. TWO-STEPS *LBE* STRATEGY

Let us be given a PV power plant composed by K modules, with each one providing the nominal power $p_k(t)$, $k = 1, \ldots, K$, at the time instant t. The electrical performance of a PV module can be modeled as a function of the environmental conditions, and consequently the power profile prediction can be performed exploiting the weather forecast. The commonly forecasted weather data are the air temperature $\tau^{(air)}(t)$ and solar irradiance $\iota^{(sun)}(t)$ computed with meteorological models. The spatial resolution of such predictions is often at regional or district level and could partially represent the real environmental conditions at the PV plant site. Moreover, the power generation performance is more dependent on the PV module temperature $\tau^{(mod)}(t)$ and the POA irradiance $\iota^{(POA)}(t)$.

The first step of the proposed strategy aims at down-scaling the weather forecast through the unknown functions Υ_τ and Υ_ι

$$\tau^{(mod)}(t) = \Upsilon_\tau \left\{ \tau^{(air)}(t), \iota^{(sun)}(t) \right\} \tag{1}$$

$$\iota^{(POA)}(t) = \Upsilon_\iota \left\{ \tau^{(air)}(t), \iota^{(sun)}(t) \right\} \tag{2}$$

which model the complex effects of the territorial morphology and of the PV plant installation (e.g., modules orientation and inclination). Once the down-scaled environmental information is computed, the total output power $P(t)$ produced by the plant can be estimated as

$$P(t) = \Phi \left\{ \pi(t), \tau^{(mod)}(t), \iota^{(POA)}(t) \right\} \tag{3}$$

where $\pi(t)$ is the theoretical output power numerically computed as follows [30, 31]

$$\begin{aligned} \pi(t) &= \eta \times \iota^{(POA)}(t) \\ &\quad \times \sum_{k=1}^{K} \left\{ p_k(t) \times \left[1 - \gamma_k(t) \left(\tau^{(mod)}(t) - \widehat{\tau} \right) \right] \right\} \end{aligned} \tag{4}$$

where η is the efficiency of the DC/AC inverter; $\gamma_k(t)$, $k = 1, \ldots, K$, are the modules power losses related to the temperature variations; and $\widehat{\tau}$ is the reference temperature. The function $\Phi\ \cdot\}$ is introduced to take into consideration all the unknown non-idealities of the hardware affecting the ideal output power $\pi(t)$, like the power losses of the cables, the non-linearity of the string boxes, and the inaccuracy in the calibration of the modules tilt angle. Summarizing, the uncertainty and time variability of the weather conditions, the ground morphology, and the complexity of the internal plant components make the closed-form computation of the functions Υ_τ, Υ_ι, and Φ very complex. The proposed two-steps *LBE* strategy is aimed at estimating such functions exploiting a set of known input-output parameters.

2.1. *SVR*-Based Machine Learning Strategy

Let us suppose the availability of weather forecast concerning the geographical area where the considered PV plant is deployed and with prediction time $\Delta t^{(pred)}$. In particular, the forecasted solar irradiance $\iota^{(sun)}(t + \Delta t^{(pred)})$ and air temperature $\tau^{(air)}(t + \Delta t^{(pred)})$ are provided at the time instant t. Starting from such input information, the final objective of the learning strategy is to accurately estimate

the forecasted output power $P(t + \Delta t^{(pred)})$. Toward this end, the problem is formulated as a two-step regression problem since the input-output data samples are observable and exploitable for the estimation of the unknown system dynamic and for the approximation of the functions $\widetilde{\Upsilon}_\tau$, $\widetilde{\Upsilon}_\iota$, and $\widetilde{\Phi}$. The basic principle of nonlinear *SVR* [32] is to find the relation from the input space to the output space, mapping the input data into a higher dimensional feature space, where a linear regression can be performed. According to the *SVM* theory, the *SVR* has the advantage of self-determining its structure and avoiding initialization problems of the training process [4, 5]. However, the *SVR* parameters such as the hyper-parameter $C^{(SVR)}$ and parameter $\gamma^{(SVR)}$ of the widely used Radial Basis Function (*RBF*) kernel have to be determined in advance [13]. Such parameter calibration is a powerful mean for the regularization, for adaptation to the noise of the training data, and for the control of the network complexity and generalization capability [34].

The proposed two-step learning process, which is graphically summarized in Fig. 1, is detailed as follows.

- **First step — Weather forecast down-scaling.** The first step aims at finding the surrogate functions $\widetilde{\Upsilon}_\tau$ and $\widetilde{\Upsilon}_\iota$ on the basis of the training sets $\{(\underline{\upsilon}(t_m), \tau^{(mod)}(t_m))\}_{m=1,\dots,M}$ and $\{(\underline{\upsilon}(t_m), \iota^{(POA)}(t_m))\}_{m=1,\dots,M}$, where $\underline{\upsilon}(t_m) = [\tau^{(air)}(t_m), \iota^{(sun)}(t_m)]$, $m = 1, \dots, M$, are the input training patterns, while $\tau^{(mod)}(t_m)$ and $\iota^{(POA)}(t_m)$, $m = 1, ldots, M$, are the corresponding outputs, such that $\tau^{(mod)}(t_m) = \Upsilon_\tau\{\underline{\upsilon}(t_m)\}$ and $\iota^{(POA)}(t_m) = \Upsilon_\iota\{\underline{\upsilon}(t_m)\}$, respectively. Then, once the regression functions $\widetilde{\Upsilon}_\tau$ and $\widetilde{\Upsilon}_\iota$ have been estimated off-line exploiting the available training data, the online test phase is aimed at down-scaling new weather forecasts given as input the test set $\underline{\upsilon}(t_n)$, $n = 1, \dots, N$, where $t_n > t_M + \Delta t^{(pred)}$. The output of the first step is the estimation of

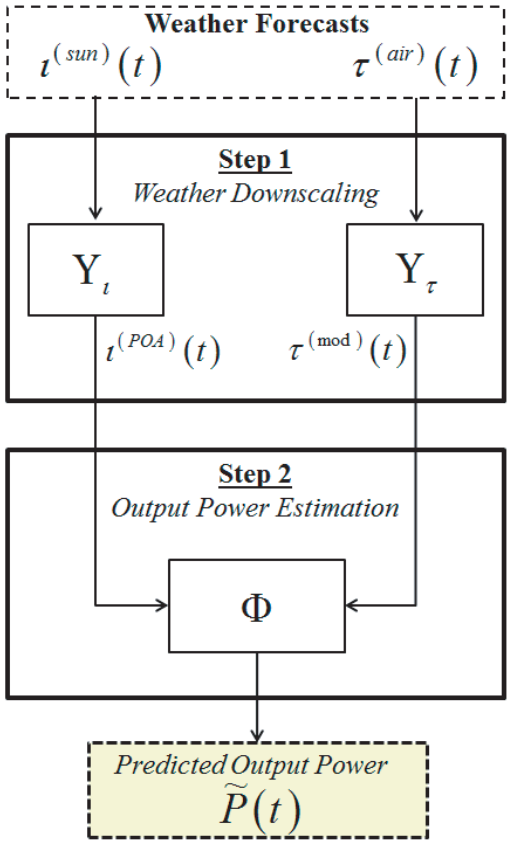


Figure 1. Two-steps *SVR*-based *ML* strategy.

the predicted parameters according to

$$\widetilde{\tau}^{(mod)}(t_n) = \widetilde{\Upsilon}_\tau \{\underline{\upsilon}(t_n)\} \tag{5}$$

$$\widetilde{\iota}^{(POA)}(t_n) = \widetilde{\Upsilon}_\iota \{\underline{\upsilon}(t_n)\} \tag{6}$$

- **Second step — Output power estimation.** The second step addresses the prediction of the total output power $\widetilde{P}(t)$ exploiting the down-scaled weather forecast computed at the previous step. Toward this end, the unknown operator Φ in (3) is estimated using the set of input-output training pairs $\{\underline{\delta}(t_m), P(t_m)\}_{m=1,\ldots,M}$, where the input vector $\underline{\delta}(t_m) = [\tau^{(mod)}(t_m), \iota^{(POA)}(t_m), \pi(t_m)]$, $m = 1, \ldots, M$. After the offline generation of the estimated function $\widetilde{\Phi}$, the predicted output power obtained using the output of the first learning step is

$$\widetilde{P}(t_n) = \widetilde{\Phi}\{\underline{\delta}(t_n)\}, \; n = 1, \ldots, N \tag{7}$$

where $t_n > t_M + \Delta t^{(pred)}$, $n = 1, \ldots, N$, are the forecasting time instants.

Combining (5), (6), and (7), the final output obtained by the proposed two-steps *LBE* strategy turns out to be

$$\widetilde{P}(t_n) = \widetilde{\Phi}\left\{\widetilde{\Upsilon}_\tau\left\{\tau^{(air)}(t_n), \iota^{(sun)}(t_n)\right\}, \widetilde{\Upsilon}_\iota\left\{\tau^{(air)}(t_n), \iota^{(sun)}(t_n)\right\}\right\}. \tag{8}$$

According to (8), the predicted output power profile of the PV plant has been obtained starting from the commonly available forecasts of air temperature and solar irradiance.

3. EXPERIMENTAL VALIDATION

The performance of the proposed approach has been assessed using an experimental dataset measured in a real PV plant with nominal power $\pi \simeq 700\,[\text{kW}]$ located in a rural area in the central part of Italy. The output power $P(t)$, irradiance $\iota^{(POA)}(t)$, and temperature $\tau^{(mod)}(t)$ have been collected every $\Delta t = 300\,[\text{sec}]$ in different seasons to take into consideration the changing environmental conditions. The forecasted solar irradiance $\iota^{(sun)}(t + \Delta t^{(pred)})$ and air temperature $\tau^{(air)}(t + \Delta t^{(pred)})$ have been acquired using a commercial service for regional weather forecast, with $\Delta t^{(pred)} = 24\,[\text{h}]$.

The selected experiments aim to validate the learning and forecasting capabilities of both the first and second steps of the proposed method. In particular, the first experiment described in Section 3.1 deals with the validation of the SVR capability to estimate the output power assuming the knowledge of the weather parameters at the PV plant, while Section 3.2 is aimed at showing the calibration of the SVR method. The down-scaling of the weather forecast is validated in Section 3.3, and the results are successively exploited to predict the PV output power. The performance of the two-steps procedure is presented in Section 3.4. The last experiment addresses the power prediction with training and test sets collected in different seasons to point out the robustness of the proposed strategy in dealing with different weather conditions.

The performance of the proposed strategy can be analytically quantified by computing the *absolute error* ε between the estimated and real output powers

$$\varepsilon_n = \left|\widetilde{P}(t_n) - P(t_n)\right|; \; n = 1, \ldots, N \tag{9}$$

and the *relative error* $\langle\varepsilon\rangle$ normalized with respect to the maximum output power

$$\langle\varepsilon\rangle_n = \frac{\varepsilon_n}{\max_n\{P(t_n)\}}; \; n = 1, \ldots, N. \tag{10}$$

Moreover, the *mean absolute error* $\bar{\varepsilon}$ and *mean relative error* $\langle\bar{\varepsilon}\rangle$ have been calculated as follows

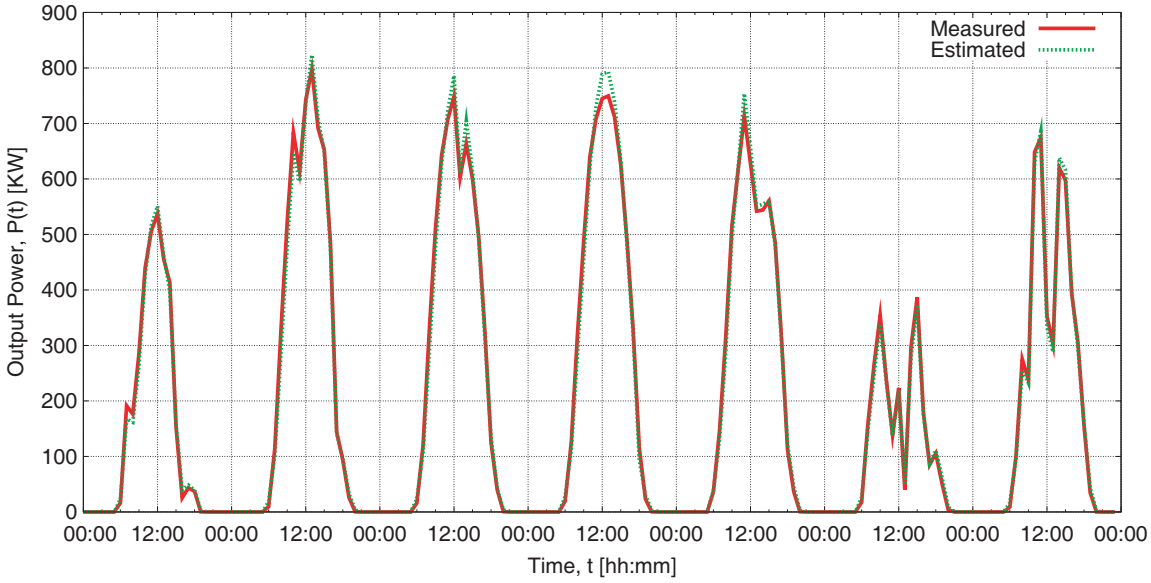
$$\bar{\varepsilon} = \frac{1}{N}\sum_{n=1}^{N}\varepsilon_n \tag{11}$$

$$\langle\bar{\varepsilon}\rangle = \frac{1}{N}\sum_{n=1}^{N}\langle\varepsilon\rangle_n \tag{12}$$

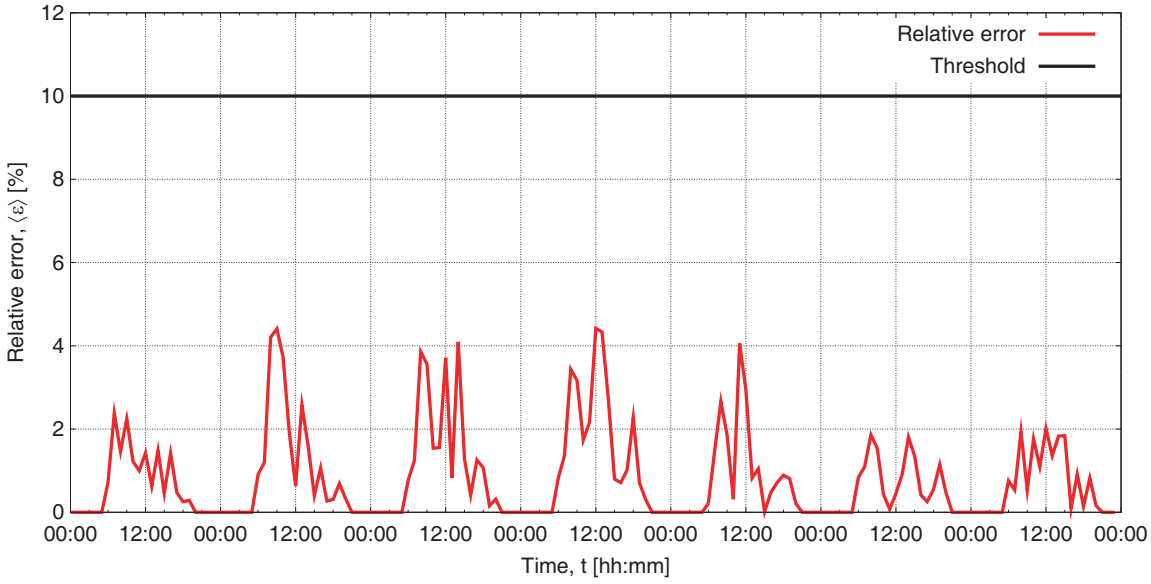
in order to provide a simple indication of the performance during the whole duration of the prediction experiments. Similar performance indicators $\overline{\varepsilon}_\tau = \frac{1}{N}\sum_{n=1}^{N}\varepsilon_\tau^{(n)}$, $\overline{\varepsilon}_\iota = \frac{1}{N}\sum_{n=1}^{N}\varepsilon_\iota^{(n)}$, where $\varepsilon_\tau^{(n)} = |\widetilde{\tau}^{(mod)}(t_n) - \tau^{(mod)}(t_n)|$ and $\varepsilon_\iota^{(n)} = |\widetilde{\iota}^{(POA)}(t_n) - \iota^{(POA)}(t_n)|$ have also been defined for the estimation of the down-scaled weather parameters.

3.1. Output Power Estimation with Known Weather-related Parameters

The first experiment is aimed at estimating the output power $\widetilde{P}(t)$ assuming the knowledge of the irradiance $\iota^{(POA)}(t)$ and of the module temperature $\tau^{(mod)}(t)$. Toward this end, the operator



(a)



(b)

Figure 2. (a) Estimated output power $\widetilde{P}(t)$ [kW] with known irradiance $\iota^{(POA)}(t)$ and module temperature $\tau^{(mod)}(t)$, and (b) the corresponding *relative error* $\langle\varepsilon\rangle$ [%].

Φ has been estimated applying the second step of the proposed *SVR*-based strategy. A set of $M = 672$ input-output training pairs $\{\underline{\delta}(t_m), P(t_m)\}_{m=1,\ldots,M}$, corresponding to a 4-week measurement campaign has been created, whereas the test data set is composed by $N = 168$ input vectors $\underline{\delta}(t_n) = [\tau^{(mod)}(t_n), \iota^{(POA)}(t_n), \pi(t_n)]$, $n = 1, \ldots, N$, measured one week after the training period (Table 1). Such a test period has been selected because varying weather conditions happened, and the generalization capabilities of the learning approach can be better verified. More in detail, cloudy and rainy weather conditions occurred during the first and last two days of the considered week, while the remaining days were sunny. The *SVR* parameters have been calibrated with the cross-validation procedure and have been set to $C^{(SVR)} = 64$ and $\gamma^{(SVR)} = 6.25 \times 10^{-2}$, respectively. The estimated output power compared with the actual measurement is shown in Fig. 2(a). As can be noticed, the power profile has been correctly estimated also during the rainy days. The corresponding *relative error* $\langle\varepsilon\rangle$ shown in Fig. 2(b) is always below the threshold $\langle\varepsilon\rangle_{th} = 10\,[\%]$ defined by the *AEEG* regulation. The *mean absolute error* is $\bar{\varepsilon} = 17.77\,[\mathrm{kW}]$ and *mean relative error* $\langle\bar{\varepsilon}\rangle = 1.56\,[\%]$.

Table 1. Size of the training and test sets for the different prediction scenarios.

Scenario	Training Set Size, M	Test Set Size, N
Short-Term Forecasting	672	168
Long-Term Forecasting	1273	1376

The outcome of this first experiment has pointed out that the *SVR*-based learning strategy is suitable to estimate the total output power when the *POA* irradiance and the module temperature are known.

3.2. *SVR* Parameters Calibration

In order to evaluate the stability of the obtained performance respect to the *SVR* calibration, an extensive assessment of the training phase has been performed changing the control parameters $C^{(SVR)}$ and $\gamma^{(SVR)}$. It is well known that the generalization capabilities of the *SVR* depend on a good selection of such meta-parameters, which are usually related to the application-domain knowledge and to the input data type [4]. The constant $C^{(SVR)} > 0$ determines the trade-off between the flatness of the

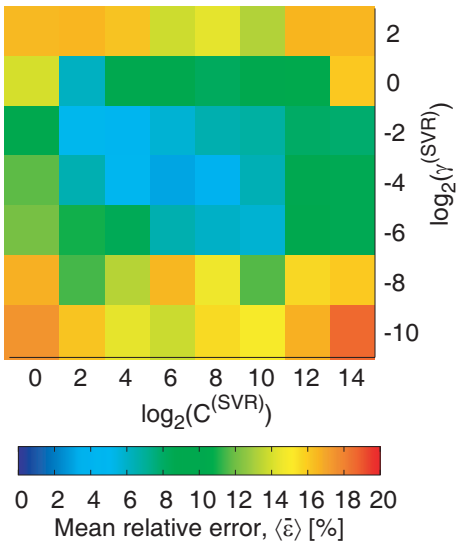
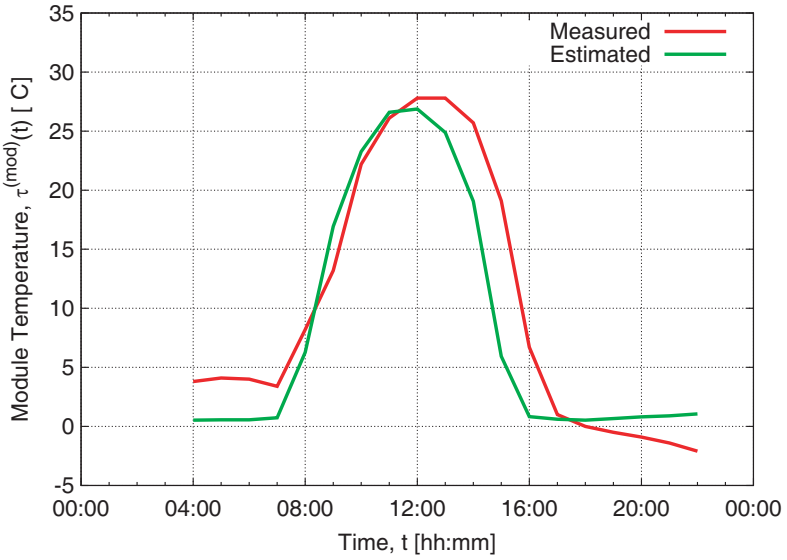
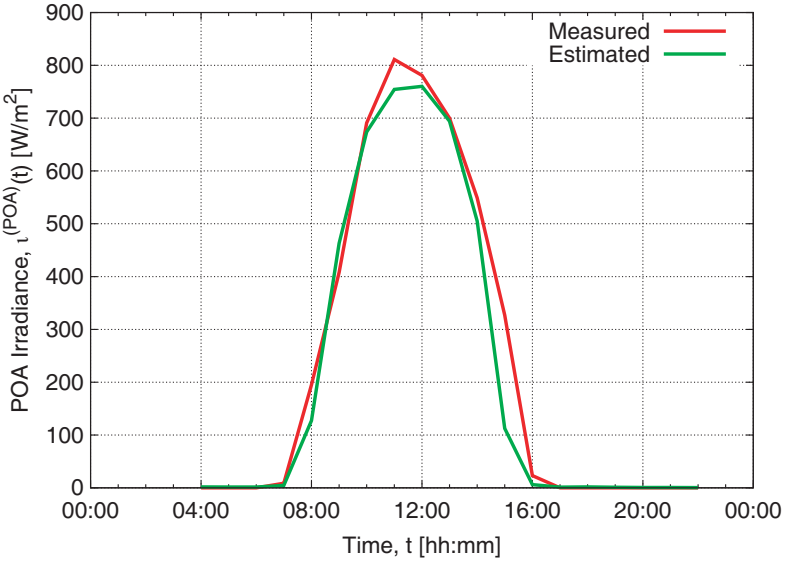


Figure 3. *Mean relative error* $\langle\bar{\varepsilon}\rangle$ [%] versus the *SVR* parameters $C^{(SVR)}$ and $\gamma^{(SVR)}$.

SVR function and the deviation from the constraints of the optimization problem performed during the training phase. If $C^{(SVR)}$ is too large, then the main objective of the optimization is to consider the empirical risk only, with less regard to the model complexity. Concerning the parameter $\gamma^{(SVR)}$ of the radial basis kernel function, it defines the influence of a single training example selected by the model as support vector. With large values of $\gamma^{(SVR)}$, the area of influence of the support vectors only includes the support vector itself, and even increasing the C parameter for regularization, the over-fitting problem is more probable. On the contrary, when $\gamma^{(SVR)}$ is very small the model cannot "capture" the complexity of the data at hand, behaving similarly to a linear model. Summarizing, the optimal parameters combination is not trivial, and it cannot be a-priori identified. In order to practically understand the impact of the calibration on the final performance of the learning method, the a-posteriori analysis of the *mean relative error* $\langle\bar{\varepsilon}\rangle$ has been performed. Fig. 3 shows the behavior of $\langle\bar{\varepsilon}\rangle$ changing the combination



(a)



(b)

Figure 4. Weather forecast down-scaling. Estimation (a) of the module temperature $\widetilde{\tau}^{(mod)}(t)$ and (b) of the *POA* irradiance $\widetilde{\iota}^{(POA)}(t)$.

of parameters in the wide range $0 \leq \log_2(C^{(SVR)}) \leq 14$ and $-10 \leq \log_2(\gamma^{(SVR)}) \leq 2$. The obtained performances satisfy the *AEEG* boundary $\langle\varepsilon\rangle_{th} = 10\,[\%]$ for a high number of parameter combinations and in particular with intermediate values of $\gamma^{(SVR)}$, where less sensitivity to the variations of $C^{(SVR)}$ is pointed out. It can be observed that a diagonal region with smaller errors exists, and the minimum error is obtained with $C^{(SVR)} = 2^6 = 64$ and $\gamma^{(SVR)} = 2^{-4} = 6.25 \times 10^{-2}$ (Fig. 3).

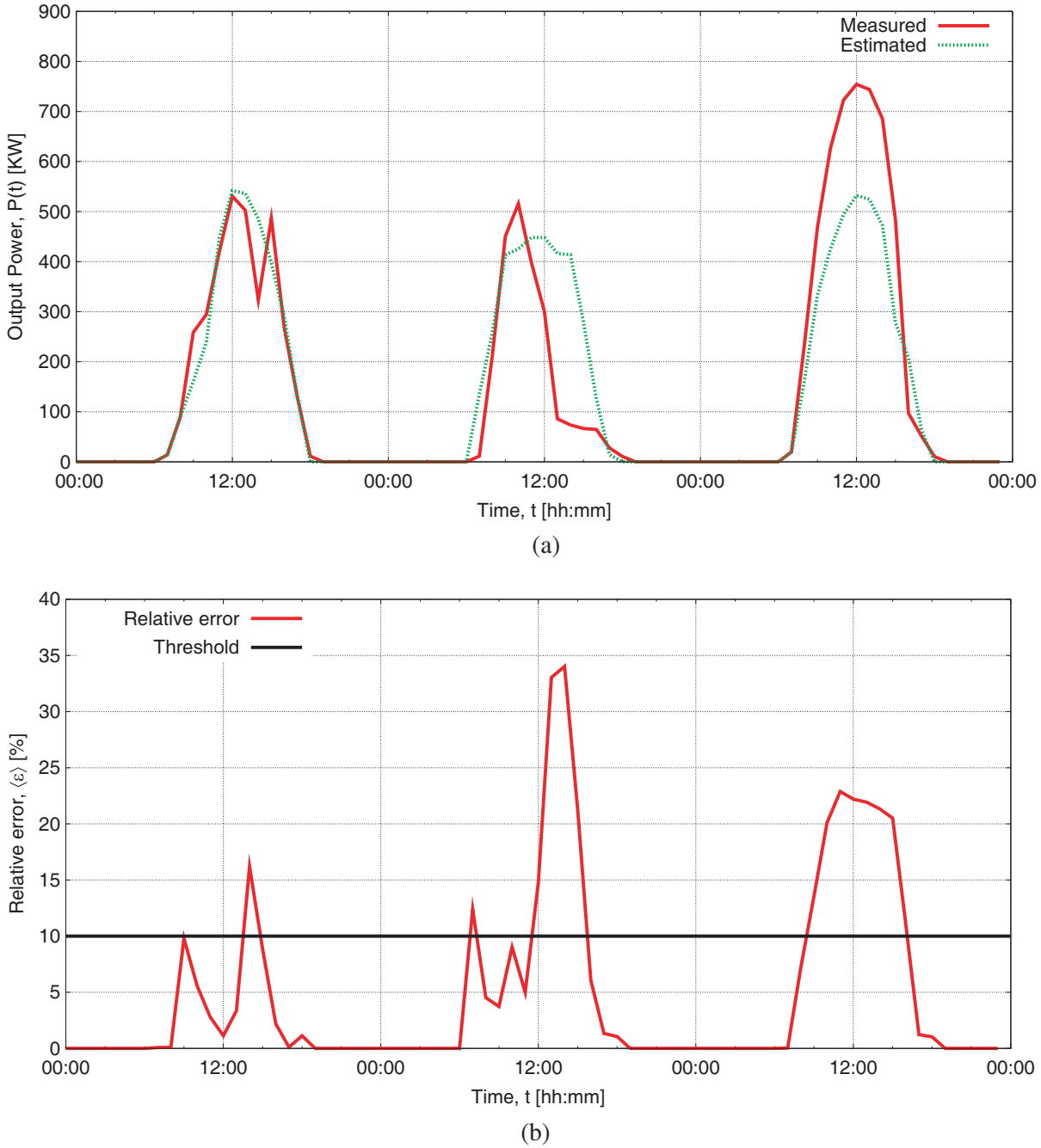
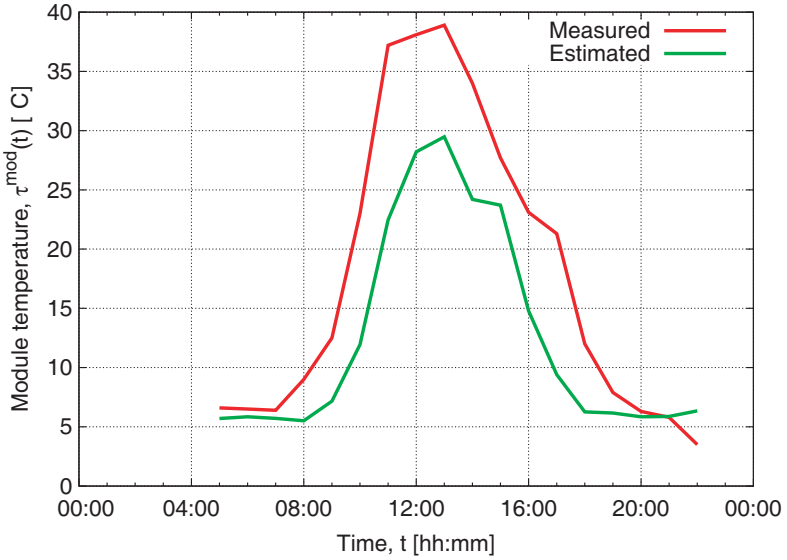


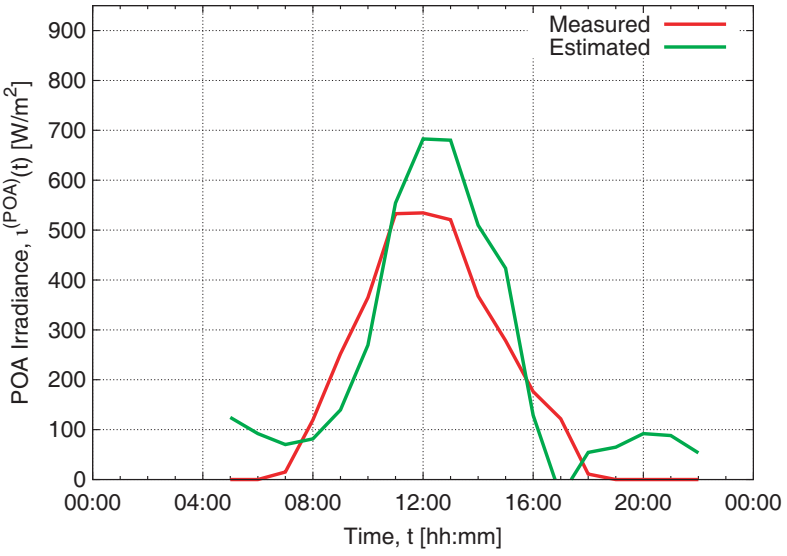
Figure 5. Output power prediction using weather forecasts — cloudy/rainy weather. (a) Predicted vs measured output power and (b) *relative error* $\langle\varepsilon\rangle$.

3.3. Weather Forecasts Down-Scaling

The results of the experiment described in Section 3.1 pointed out good performance in the estimation of the output PV power when the measurement of the module temperature $\tau^{(mod)}(t)$ and POA irradiance $\iota^{(POA)}(t)$ are available. In this section, the first step of the proposed learning strategy, which is aimed at the weather forecasts down-scaling, is validated. The capability to correctly estimate the operators $\widetilde{\Upsilon}_\tau$ and $\widetilde{\Upsilon}_\iota$ has been verified considering the experimental data of the 4-week measurement campaign described in Section 3.1, together with the air temperature $\tau^{(air)}(t)$ and solar irradiance $\iota^{(sun)}(t)$ given by the commercial weather forecast service, which provides a maximum spatial resolution of 1.5 [km^2] and a forecast prediction time $\Delta t^{(pred)} = 24$ [h]. It has to be clarified that the training input-output pairs have been synchronized in time in order to have temporal match between the weather forecasted



(a)

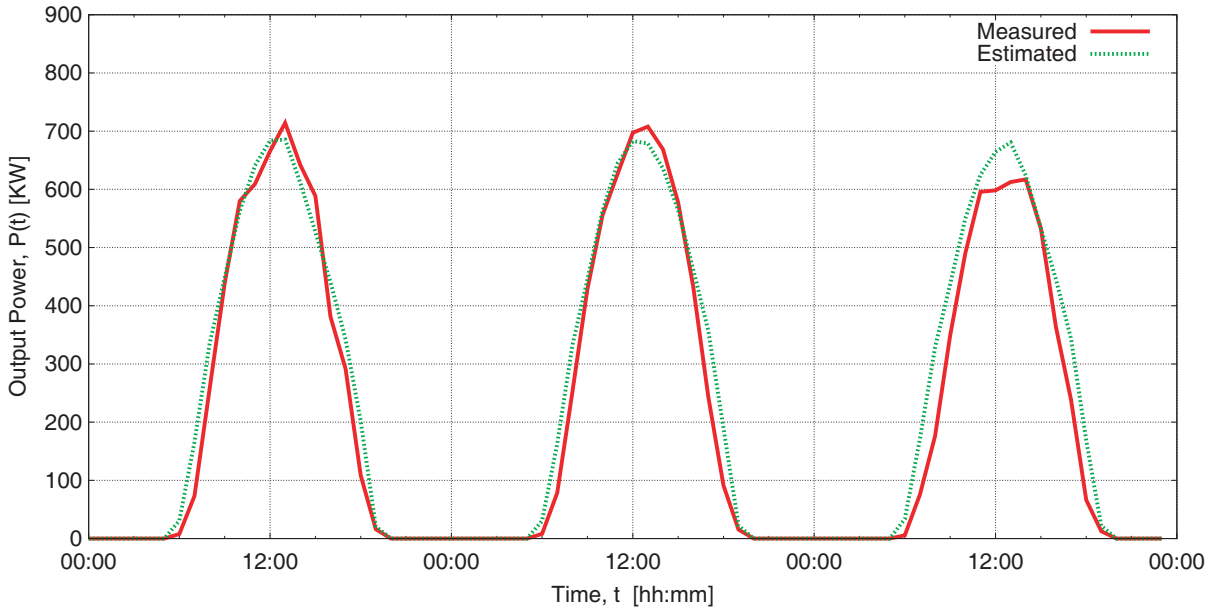


(b)

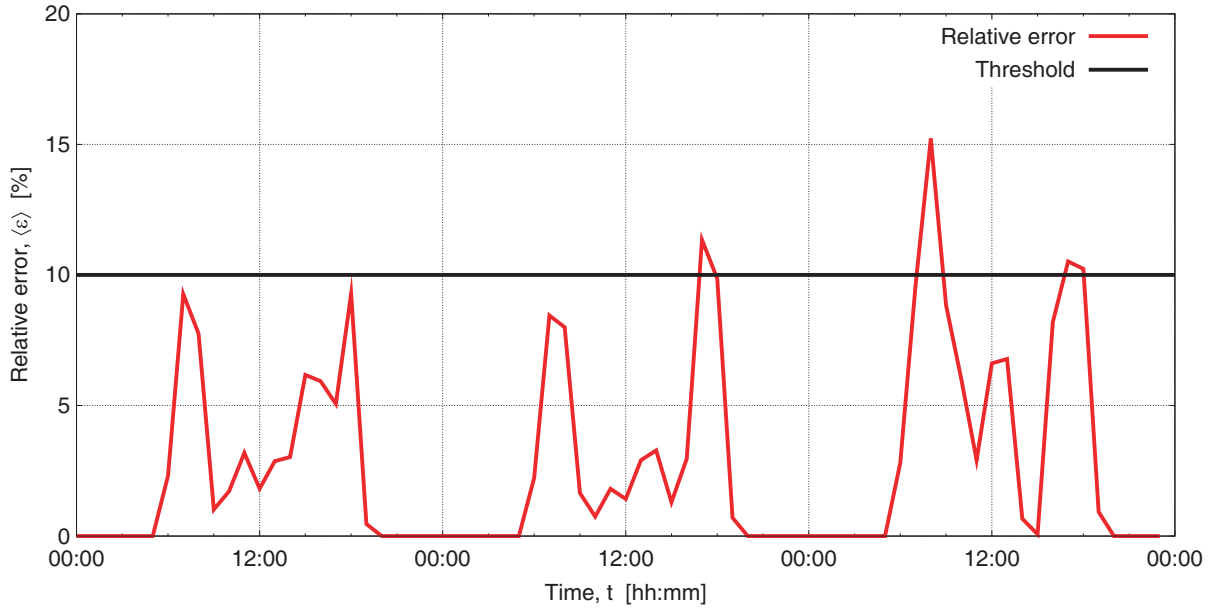
Figure 6. Cloudy weather forecast down-scaling. Estimation (a) of the module temperature $\widetilde{\tau}^{(mod)}(t)$ and (b) of the POA irradiance $\widetilde{\iota}^{(POA)}(t)$.

values (input) and the local measurements on the plant (output). The *SVR* functions have been trained using the optimal parameter set described in Section 3.2.

The validation has been performed with a test set collected during the week after the training period, in order to downscale unknown weather forecast. The estimated module temperature $\widetilde{\tau}^{(mod)}(t)$ and *POA* irradiance $\widetilde{\iota}^{(POA)}(t)$ of a representative day are reported in Fig. 4(a) and Fig. 4(b), respectively. The comparisons point out a good reconstruction of both the temperature and irradiance during the whole considered day, with *mean relative errors* $\bar{\varepsilon}_\tau = 7.33\,[°\mathrm{C}]$ and $\bar{\varepsilon}_\iota = 35.61\,[\mathrm{W/m^2}]$. As can be noticed,



(a)



(b)

Figure 7. Output power prediction using weather forecasts — sunny weather. (a) Predicted vs measured output power and (b) *relative error* $\langle\varepsilon\rangle$.

a time delay exists between the measured and estimated module temperatures due to the common misalignment of the weather forecasts. However, the similarity between the two patterns points out the feasibility of the down-scaling process.

3.4. Power Prediction with Standard Weather Forecasts

The validation reported in this section is aimed at assessing the performance of the whole two-step *LBE* process. The operators $\widetilde{\Upsilon}_\tau$, $\widetilde{\Upsilon}_\iota$, and $\widetilde{\Phi}$ have been trained according to the experiments described in the previous sections, and successively tested with the available regional weather forecast. More in detail, the forecasted solar irradiance $\iota^{(sun)}(t + \Delta t^{(pred)})$ and air temperature $\tau^{(air)}(t + \Delta t^{(pred)})$, with $\Delta t^{(pred)} = 24\,[\text{h}]$, have been acquired during (i) three cloudy/rainy days characterized by unstable forecasts, and during (ii) three sunny days.

The results of the first cloudy/rainy test case are reported in Fig. 5(a), and the corresponding *relative error* $\langle\varepsilon\rangle$ is shown in Fig. 5(b). The prediction retrieved by the proposed approach shows inaccuracies mainly related to the unreliable weather forecasts. As an example, during the second day of the considered test period, the method underestimated in the morning and overestimated in the afternoon the output power with *relative error* $\langle\varepsilon\rangle > 30\,[\%]$ [Fig. 5(b)]. Fig. 6 points out the inaccuracy of the weather down-scaling during the second test day, which is the main cause of the inaccurate power prediction. For instance, in the early afternoon the *POA* irradiance [Fig. 6(b)] has been forecasted with values lower than the measured one ($\bar{\varepsilon}_\iota = 132.45\,[\text{W/m}^2]$), while the module temperature has been overestimated [Fig. 6(a)].

The second test case points out the performance of the proposed method when dealing with a stable sunny period. The measured and predicted output powers and the *relative error* are shown in Fig. 7(a) and Fig. 7(b), respectively. As a matter of fact, the proposed approach provided a satisfactory energy forecast except for three short-term error peaks with maximum value $\max_n\{\langle\varepsilon\rangle_n\} = 15.25\,[\%]$, which exceeded the *AEEG* threshold $\langle\varepsilon\rangle_{th} = 10\,[\%]$. As can be noticed, the prediction performance degrades in the morning and in the evening when the weather forecasts are less accurate [23]. However, the obtained error metrics pointed out satisfying values of *mean absolute error* $\bar{\varepsilon} = 61.85\,[\text{kW}]$ and *mean relative error* $\langle\bar{\varepsilon}\rangle = 3.72\,[\%]$.

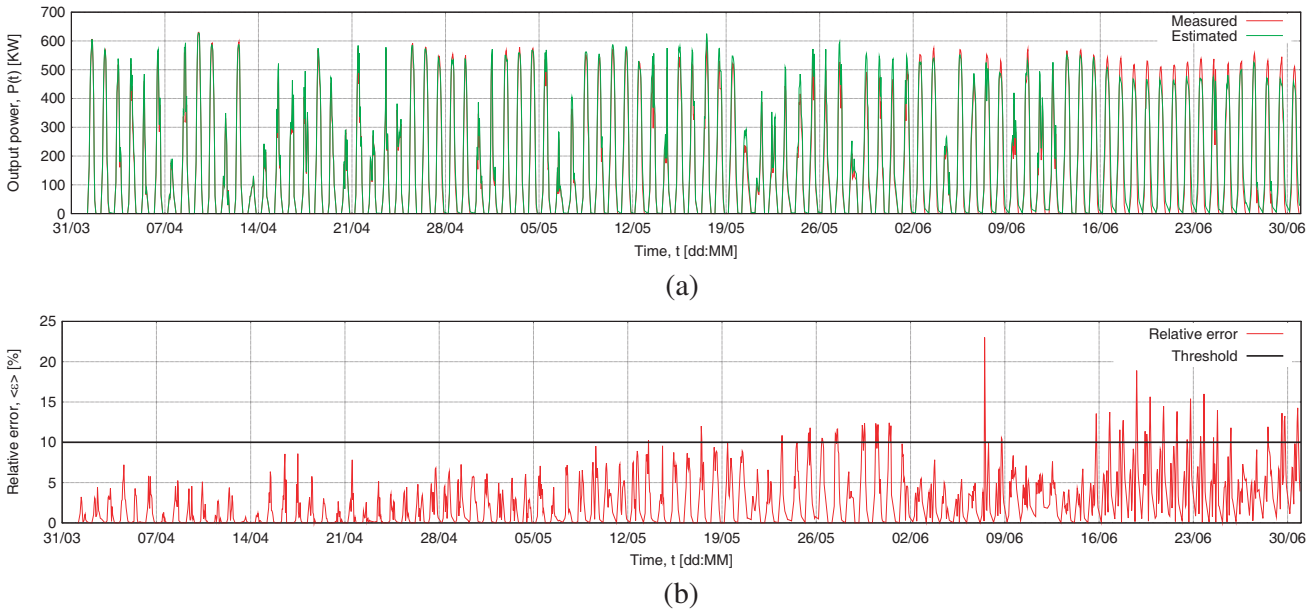


Figure 8. Long-term output power prediction with training and test sets from different seasons. (a) Predicted vs measured output power and (b) *relative error* $\langle\varepsilon\rangle$.

3.5. Long-Term Output Power Forecasting

The last experiment is aimed at testing the prediction capabilities using long-term training and testing data acquired in different seasons. In particular, $M = 1273$ training samples have been acquired during the winter period (i.e., January, February, and March), while the spring weather forecasts of April, May, and June have been considered for the test set ($N = 1376$ — Table 1). The two-step *LBE* strategy has been applied using the *SVR* parameters calibrated in Section 3.2, and the results of the prediction are shown in Fig. 8. As expected, the *relative error* reported [Fig. 8(b)] points out higher prediction inaccuracies at the end of the test, when the weather conditions are very different with respect to the winter season. However, only 4.65 [%] of the test samples presented a *relative error* $\langle\varepsilon\rangle > 10$ [%].

4. CONCLUSION

In this paper, an innovative two-step *LBE* strategy has been proposed to predict the daily output power profile of a *PV* plant. The first step is devoted to the *SVR*-based prediction of the *POA* irradiance and the module temperature starting from the available weather forecasts. Successively, the obtained down-scaled weather parameters are exploited as input of the second-step *SVR* for the estimation of the next-day output power profile. The proposed method has demonstrated the following capabilities:

- The *estimation* of the output power profile (with *mean relative error* $\langle\bar{\varepsilon}\rangle < 10\%$) when the local measurement of the *POA* irradiance and module temperature are known. This outcome pointed out that the *SVR* is a suitable methodology to learn the relation between the weather-related *PV* features and the output power.
- The *prediction* of the output power profile when the forecasted solar irradiance and air temperature are used as input data. As expected, the experimental validation of the prediction performance has pointed out a strict relation with the quality of the weather forecasts.

The proposed *ML* approach has shown satisfactory performance also dealing with the long-term prediction of the *PV* output power with training samples acquired in different seasons. The experimental validation has pointed out that the main limitations are related to the quality of the weather forecast. Future studies will be focused on (a) the integration of additional weather indicators, such as the wind speed and direction, as features of the *LBE*-based strategy, (b) the exploitation of heterogeneous data fusion models, leveraging on the spreading of environmental wireless sensors based on the Internet of Things paradigm, as well as (c) the introduction of progressive/reinforcement learning strategies to dynamically update the training data set and self-adapt the learning method according to the measurement data.

ACKNOWLEDGMENT

This work benefited from the networking activities carried out within the Project "MITIGO — Mitigazione dei rischi naturali per la sicurezza e la mobilita' nelle aree montane del Mezzogiorno" (Grant No. ARS01_00964) funded by the Italian Ministry of Education, University, and Research within the PON R&I 2014-2020 Program (CUP: B64I20000450005) and the Project "SMARTOUR — Piattaforma Intelligente per il Turismo" (Grant No. SCN_00166) funded by the Italian Ministry of Education, University, and Research within the Program "Smart cities and communities and Social Innovation" (CUP: E44G14000040008).

REFERENCES

1. UN General Assembly, "Transforming our world: The 2030 Agenda for Sustainable Development," Oct. 21, 2015, A/RES/70/1, available at: https://www.refworld.org/docid/57b6e3e44.html [accessed Jun. 7, 2022].
2. GSE, "ARG/elt 4/10 — Procedure to improve the predictability of the electricity input from implants using renewable energy sources that are not programmable about not relevant production units," Jan. 2010.

3. Nespoli, A., et al., "Day-ahead photovoltaic forecasting: A comparison of the most effective techniques," *Energies*, Vol. 12, No. 9, 1621, 2019.
4. Massa, A., G. Oliveri, M. Salucci, N. Anselmi, and P. Rocca, "Learning-by-examples techniques as applied to electromagnetics," *Journal of Electromagnetic Waves and Applications*, Vol. 32, No. 4, 516–541, 2018.
5. Salucci, M., M. Arrebola, T. Shan, and M. Li, "Artificial intelligence: New frontiers in real-time inverse scattering and electromagnetic imaging," *IEEE Trans. Antennas Propag.*, DOI: 10.1109/TAP.2022.3177556.
6. Massa, A., D. Marcantonio, X. Chen, M. Li, and M. Salucci, "DNNs as applied to electromagnetics, antennas, and propagation — A review," *IEEE Antennas Wireless Propag. Lett.*, Vol. 18, No. 11, 2225–2229, Nov. 2019.
7. Elshennawy, W., "Large intelligent surface-assisted wireless communication and path loss prediction model based on electromagnetics and machine learning algorithms," *Progress In Electromagnetics Research C*, Vol. 119, 65–79, 2022.
8. Salucci, M., N. Anselmi, G. Oliveri, P. Calmon, R. Miorelli, C. Reboud, and A. Massa, "Real-time NDT-NDE through an innovative adaptive partial least squares SVR inversion approach," *IEEE Trans. Geosci. Remote Sens.*, Vol. 54, No. 11, 6818–6832, Nov. 2016.
9. Liu, F., Y. Wu, H. Duan, and R. Du, "SVR-CMT Algorithm for null broadening and sidelobe control," *Progress In Electromagnetics Research*, Vol. 163, 39–50, 2018.
10. Salucci, M., L. Tenuti, G. Oliveri, and A. Massa, "Efficient prediction of the EM response of reflectarray antenna elements by an advanced statistical learning method," *IEEE Trans. Antennas Propag.*, Vol. 66, No. 8, 3995–4007, Aug. 2018.
11. Salucci, M., J. Vrba, I. Merunka, and A. Massa, "Real-time brain stroke detection through a learning-by-examples technique — An experimental assessment," *Microw. Opt. Technol. Lett.*, Vol. 59, No. 11, 2796–2799, Aug. 2017.
12. Hosseinzadeh, S. and M. Shaghaghi, "GPR data regression and clustering by the fuzzy support vector machine and regression," *Progress In Electromagnetics Research M*, Vol. 93, 175–184, 2020.
13. Salucci, M., G. Oliveri, and A. Massa, "Real-time electrical impedance tomography of the human chest by means of a learning-by-examples method," *IEEE J. Electromagn., RF, Microw. Med. Biol.*, Vol. 3, No. 2, 88–96, Jun. 2019.
14. Li, H., B. Zhu, and J. Chen, "Optimal design of photonic band-gap structure based on Kriging surrogate model," *Progress In Electromagnetics Research M*, Vol. 52, 1–8, 2016.
15. Salucci, M., N. Anselmi, G. Oliveri, P. Rocca, S. Ahmed, P. Calmon, R. Miorelli, C. Reboud, and A. Massa, "A nonlinear kernel-based adaptive learning-by-examples method for robust NDE-NDT of conductive tubes," *Journal of Electromagnetic Waves and Applications*, Vol. 33, No. 6, 669–696, Feb. 2019.
16. Rayala, R. and S. Raghavan, "Hexagon shape SIW bandpass filter with CSRRS using artificial neural networks optimization," *Progress In Electromagnetics Research Letters*, Vol. 104, 47–55, 2022.
17. Oliveri, G., M. Salucci, and A. Massa, "Towards efficient reflectarray digital twins — An EM-driven machine learning perspective," *IEEE Trans. Antennas Propag.*, DOI: 10.1109/TAP.2022.3155204.
18. Tanaka, K., K. Uchida, K. Ogimi, T. Goya, A. Yona, and T. Senjyu, "Optimal operation by controllable loads based on smart grid topology considering insolation forecasted error," *IEEE Trans. Smart Grid*, Vol. 2, No. 3, 438–444, Sep. 2011.
19. Capizzi, G., C. Napoli, and F. Bonanno, "Innovative second-generation wavelets construction with recurrent neural networks for solar radiation forecasting," *IEEE Trans. Neural Netw. Learn. Syst.*, Vol. 23, No. 11, 1805–1815, Nov. 2012.
20. Lorenz, E., T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz, "Regional PV power prediction for improved grid integration," *Progress in Photovoltaic: Research and Applications*, Vol. 19, 757–771, 2011.

21. Knight, K. M., S. A. Klein, and J. A. Duffie, "A methodology for the synthesis of hourly weather data," *Solar Energy*, Vol. 46, No. 2, 109–120, 1991.
22. Liu, J., W. Fang, X. Zhang, and C. Yang, "An improved photovoltaic power forecasting model with the assistance of aerosol index data," *IEEE Trans. Sustain. Energy*, Vol. 6, No. 2, 434–442, Apr. 2015.
23. Shi, J., W. J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Trans. Ind. Appl.*, Vol. 48, No. 3, 1064–1069, May 2012.
24. Yona, A., T. Senijyu, A. Y. Seber, T. Funabashi, H. Sekine, and C. H. Kim, "Application of neural network to 24-hour-ahead generating power forecasting for PV system," *Proceedings of Power and Energy Society General Meeting — Conversion and Delivery of Electrical Energy in 21st Century*, 1–6, 2008.
25. Wang, F., Z. Mi, S. Su, M. Chen, and C. Zhang, "A practical model for single-step power prediction of grid-connected PV plant using artificial neural network," *IEEE Innovative Smart Grid Technologies Asia (ISGT)*, 1–4, 2011.
26. Fonseca, J. G., T. OOzeki, T. Takashima, G. Koshimizu, Y. Uchida, and K. Ogimoto, "Photovoltaic power production forecasts with support vector regression: A study on the forecast horizon," *37th IEEE Photovoltaic Specialists Conference (PVSC) 2011*, 1–6, Seattle, WA, USA, Jun. 2011.
27. Mellit, A. and A. M. 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*, Vol. 84, No. 5, 807–821, 2010.
28. Shi, J., W.-J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic system based on weather classification and support vector machine," *Industry Applications Society Annual Meeting (IAS)*, 1–6, Orlando, FL, USA, Oct. 2011.
29. Shi, J., W.-J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machine," *IEEE Trans. on Industry App.*, Vol. 46, No. 3, 1064–1069, Jun. 2012.
30. Luque, A. and S. Hegedus, *Handbook of Photovoltaic Science and Engineering*, 2nd Edition, Wiley, New York, 2011.
31. Colli, A. and W. J. Zaaiman, "Maximum-power-based PV performance validation method: Application to single-axis tracking and fixed-tilt c-Si systems in the Italian Alpine region," *IEEE J. Photovolt.*, Vol. 2, No. 4, 555–563, Oct. 2012.
32. Cristianini, N. and J. S. Taylor, *An Introduction to Support Vector Machine*, Cambridge University Press, Cambridge, U.K., 2000.
33. Ito, K. and R. Nakano, "Optimizing support vector regression hyperparameters based on cross-validation," *Proceedings of the International Joint Conference on Neural Networks*, Vol. 3, 2077–2082, Jul. 2003.
34. Smola, A. J. and B. Scholkopf, "From regularization operators to support vector kernels," *Neural Information Processing Systems*, MIT Press, Cambridge, MA, 1997.
35. Mellit, A., A. Massi Pavan, and V. Lughi, "Short-term forecasting of power production in a large-scale photovoltaic plant," *Sol. Energy J.*, Vol. 105, 401–413, Jul. 2014.
36. Yang, C., A. A. Thatte, and L. Xie, "Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation," *IEEE Trans. Sustain. Energy*, Vol. 6, 104–112, Jan. 2015.
37. Yang, H. T., C. Huang, Y. C. Huang, and Y. Pai, "A weather-based hybrid method for one-day ahead hourly forecasting of PV power output," *IEEE Trans. Sustain. Energy*, Vol. 5, 917–926, Jul. 2014.
38. Chen, C., S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," *Sol. Energy J.*, Vol. 85, 2856–2870, Nov. 2011.