Electric Power Systems Research 212 (2022) 108415

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

A cross-learning approach for cold-start forecasting of residential photovoltaic generation

J. Bottieau [a,*], Z. De Grève [a], T. Piraux [b], A. Dubois [c], F. Vallée [a], J.-F. Toubeau [a]

[a] Power Systems & Markets Research Group, University of Mons, Mons, Belgium
[b] Wesmart Entreprise, Brussels, Belgium
[c] Regional Development Agency Ideta, Tournai, Belgium

ARTICLE INFO

Keywords:
Cold-start forecasting
Machine learning
Multi-horizon forecasting
Photovoltaic generation

ABSTRACT

This paper addresses the problem of forecasting, over a daily horizon, quarter hourly profiles of residential photovoltaic (PV) power production for sites with no historical data available. Typically, such forecasts are required for improving the local operation of low-voltage systems, where observability is still a practical challenge. In this context, we develop a cross-learning forecasting approach to predict unobserved PV sites, which exploits common patterns learned from neighboring monitored PV production profiles. Concretely, the proposed approach fits a single, generic forecasting function across the entire panel of monitored PV time series based only on series-specific features – i.e., the peak power installed, geographical position, orientation and inclination – and local numerical weather predictions. This allows to enlarge the dataset for training more complex data-driven techniques, while ensuring scalability for predicting each PV site. The proposed approach is evaluated using a k-nearest neighbors algorithm, different variants of neural networks and gradient boosted trees on five new residential PV sites. Outcomes highlight the ability of the cross-learning forecasting models to better generalize on new PV sites in comparison with a clear sky-based physical approach, without needing any adjustment of the models.

1. Introduction

The prediction of local renewable generation is gaining an increasing attention in the power system community [1]. This interest is fostered by the emergence of local market mechanisms, such as Renewable Energy Communities, which are constituted by organized groups of consumers/prosumers who can exchange electricity locally without resorting to the traditional wholesale/retail market structure [2]. In such communities, the coordination between production and generation is achieved locally, and is usually informed by predictions of local renewable energy sources and end-users consumption.

Numerous community projects involving residential end-users are currently launched in Europe—see e.g., [3,4]. In most of these projects, historical data of photovoltaic (PV) generation (e.g., quarter hourly profiles) at the forecasting location is not readily available, which hinders the forecasting task. This paper aims therefore at proposing a machine learning (ML)-based solution to predict, over a daily horizon, quarter hourly profiles of PV generation for sites where no historical data is available. This challenging task is generally referred to as 'cold-start problem', which typically occurs for the forecasting of new variables with no history [5,6]. To the best of our knowledge, such a topic has been barely addressed for local renewable generation sites, where most of the methods are fitted based on past data pertaining to the target site [4,7].

Overall, PV generation is highly related to the total amount of solar irradiance incident on the panels, which fluctuates over time. The extraterrestrial component of irradiance can be precisely calculated from solar constant and Sun–Earth geometry [8]. However, the changing atmospheric conditions and cloud movements introduce stochasticity in the amount of rays of sunshine collected by the panels, which consequently jeopardizes the PV forecasting task [7]. In this line, several mathematical models were developed for forecasting the surface solar irradiance such as (i) clear-sky models [9], which estimate the irradiance in the absence of clouds, or (ii) time series-based models, which incorporate the temporal information on cloudiness [10]. Typically, physical-based models, i.e., methods estimating the PV output power from a set of analytical equations, heavily rely on such inputs [11]. While physical-based models do not need historical data to be calibrated, their performance is directly impacted by the accuracy of irradiance forecasts and the availability of technical specifications of the PV system. On the other hand, statistical models mostly combine

* Corresponding author.
E-mail address: jeremie.bottieau@umons.ac.be (J. Bottieau).

https://doi.org/10.1016/j.epsr.2022.108415
Received 3 October 2021; Received in revised form 17 April 2022; Accepted 2 July 2022
Available online 14 July 2022

linear inferences of lagged PV power measurements and predicted solar irradiance, e.g., in [12,13]. Although these methods have proved to be successful, they cannot be easily deployed for sites where no local telemetry is available. Recently, ML techniques have emerged as promising candidates for predicting the output power of PV systems [14]. The main techniques include regression trees [15], support vector machine [16] and different variations of neural networks [17–19]. Their success rely on their ability to model highly non-linear relationships, without requiring any prior assumptions about the PV generation data. In particular, deeper architectures of neural networks have been developed in [20,21] to extract more information from raw data alleviating the need of hand-crafted features. However, their performance can be limited by an insufficient amount of training data when fitted on individual PV time series, which prevents them from reaching their full potential [11,21].

Generally, few research works investigate a generic, scalable forecasting approach, which can be applied for any individual PV site without recurring to a complete retraining of the forecaster. In such a framework, the forecasting approach has to be robust against the lack of ground measurements of the target sites. Indeed, installing and operating local telemetric data for all the forecasting locations may become very expensive, while leading to privacy concerns. A fist approach consists of approximating these non-available local telemetric data using interpolation methods, which then allows training different statistical methods [22,23]. This two-step approach has shown accurate forecasts for very short forecast horizons (e.g., sub-5-min), but their extension towards longer forecasting horizons is still subject to research [24]. In contrast, we propose a cross-learning forecasting approach based on ML techniques, which inherently tackles both insufficient data and scalability issues. A cross-learning approach consists in fitting a single, generic function to all time series available in the dataset [25]. The obtained generic model can then be applied on new related time series. Such a model can compete with individual models, even for dataset containing heterogeneous time series [26]. Refs. [27,28] adopt a similar approach with traditional feed-forward neural networks for predicting the solar irradiance at various locations. Here, we differ from these works by focusing on the local PV prediction task, while combining the cross-learning approach with a representative panel of state-of-the-art machine learning techniques including the k-nearest neighbor algorithm, feed-forward and recurrent neural networks, and gradient boosted regression trees. The contributions can be summarized as follows:

- We propose a cross-learning approach for predicting unobserved PV sites. This approach offers the benefit to enlarge the training data set, while providing a scalable forecasting model that can be applied for any local PV site.
- The cross-learning approach is combined with a representative panel of state-of-the-art machine learning techniques. This allows to evaluate the pros and cons of each technique considering our real-world data on local PV generation.

We evaluate the performance of the machine learning techniques in a deterministic framework, which permits a direct comparison with a clear sky-based physical approach. However, it should be noted that techniques such as neural networks and gradient boosted regression trees can be readily extended toward a probabilistic framework [29]. The obtained outcomes show the ability of the cross-learning forecasting models to leverage common patterns of neighboring PV sites for predicting unobserved PV sites. In particular, the gradient boosted trees model provides a superior performance in our case study. In complement, its performance is further analyzed by varying the number of available monitored PV sites and input features, which allows to provide insights on its sensitivity on these both elements.

The paper is outlined as follows. Section 2 details the cross-learning forecasting approach. Section 3 presents the clear sky-based physical approach and the evaluation metric used in the paper. Section 4 conducts the case study, while providing numerical results. Finally, Section 5 concludes the paper.

2. Methodology

Section 2.1 introduces the cross-learning forecasting approach, which is applied for unobserved PV sites. Then, the available input features are described in Section 2.2. Finally, the machine learning techniques are presented in Section 2.3.

2.1. Cross-learning forecasting approach

Overall, at the beginning of each day, we consider the following multi-step time series regression problem:

$$\mathbb{E}(p^{\mathrm{PV}}_{j,t_0},\dots,p^{\mathrm{PV}}_{j,t_T}|\mathbf{x}^s_j,\mathbf{x}^f_{j,t_0},\dots,\mathbf{x}^f_{j,t_T}),\quad \forall j\in\mathcal{J} \tag{1}$$

where $p^{\mathrm{PV}}_{j,t_0:t_T}$ is the power production (to predict) of a new PV site j over steps t_0 to t_T of the prediction horizon. Two types of explanatory variables (inputs) are considered: (i) series-specific features $\mathbf{x}^s_j$, e.g., the latitude and longitude of the PV site, and (ii) available temporal covariates $\mathbf{x}^f_{j,t_0:t_T}$ such as numerical weather predictions (NWP) obtained from different locations. Note that, in contrast with traditional approaches, no lagged PV power measurements are leveraged as their availability is not guaranteed for all PV sites $j\in\mathcal{J}$. Finally, we consider the quarter hours $t\in[t_0,t_T]$ occurring between 5.30 a.m. and 10.30 p.m., thereby disregarding the time steps where the PV output power is always null.

The objective in Eq. (1) is achieved by relying on a cross-learning forecasting approach, which results in a generic forecasting function f_Θ suitable for all the forecasted PV sites $\mathcal{J}$:

$$\mathbb{E}(p^{\mathrm{PV}}_{j,t_0:t_T}|\mathbf{x}^s_j,\mathbf{x}^f_{j,t_0:t_T}) = f_\Theta(\mathbf{x}^s_j,\mathbf{x}^f_{j,t_0:t_T}),\ \ \forall j\in\mathcal{J} \tag{2}$$

The parameters Θ defining f_Θ are learned jointly using all the available samples $\mathcal{I}$ of the monitored PV time series. In particular, they are fitted by optimizing an error function $\mathcal{L}$ measuring the compatibility of the forecasting function f_Θ with respect to all the dataset $(\mathcal{P}_i,\mathcal{X}_i),\ \forall i\in\mathcal{I}$, where $\mathcal{P}_i=(p^{\mathrm{PV}}_{i,t_0:t_T})$ and $\mathcal{X}_i=\left(\mathbf{x}^s_i,\mathbf{x}^f_{i,t_0:t_T}\right)$. The optimization problem for the forecasting function f_Θ is:

$$\Theta^* = \underset{\Theta}{\operatorname{argmin}} \sum_{i\in\mathcal{I}} \mathcal{L}(\mathcal{P}_i, f_\Theta(\mathcal{X}_i)) \tag{3}$$

where $\mathcal{L}$ is typically the mean square error function in a deterministic framework.

The cross-learning approach allows to build a forecasting function f_Θ that extracts the common patterns from all samples of the monitored PV sites. One practical advantage of this approach is its potential for dealing with time series limited in terms of data points. Indeed, when fitting a data-driven model on an individual short time series, the model is not likely to capture the underlying data generation process of the series, which will undermine its performance. However, when the model is trained on multiple time series, the odds of approximating adequately the data generating process is increased as it automatically augments the data availability. The local PV forecasting task is well-suited for adopting the cross-learning approach as the dataset is composed of multiple time series that are governed by similar physical processes for a defined geographical area. Of course, the overall performance of this approach is dependent on the availability of a sufficient number of monitored PV sites. This relation between the performance of the model and the number of available monitored PV sites is studied in Section 4.2.

Practically, Fig. 1 depicts the geolocations of the monitored PV sites (eleven black circles) for which historical production profiles are available, and the set $\mathcal{J}$ (five red triangles), i.e., new sites to be forecasted. Note that, at the forecast creation time, the production profiles of the monitored PV sites have not yet been measured over the prediction horizon $t\in[t_0,t_T]$, and these values cannot thus be leveraged for directly predicting the generation of the new PV sites. In addition, the geolocations of the NWP are denoted by four blue diamonds in Fig. 1.

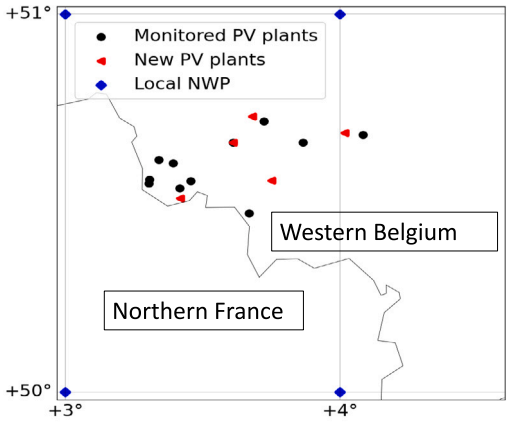


Fig. 1. Geolocations of the monitored (black circles) and new (red triangles) residential PV plants, while the blue diamonds denote the geolocations of the local numerical weather predictions (NWP).

2.2. Input features

The series-specific features $\mathbf{x}_i^s$ are the peak power installed c_i, the geolocation of the PV site, i.e., latitude ϕ_i and longitude λ_i, the tilt angle β_i and orientation γ_i of the considered PV system in sample i. On the other hand, the time-varying inputs defined by $\mathbf{x}_{i,t_0:t_T}^f$ are the tilted global clear-sky irradiance and the four NWP. The tilted global clear-sky irradiance $I_{i,t_0:t_T}^{\text{PV}}$ [W/m^2] is the estimated irradiance received on the tilted PV surface β_i with orientation γ_i based on the isotropic transposition model in [30]. This transposition is performed on the global clear-sky irradiance obtained via the Ineichen-Perez model defined in [31]. Note that both models are available in the pvlib package [32]. Regarding the NWP, we consider the ambient air temperature $T_{k,t_0:t_T}^A$ [°C] and cloud coverage $\text{CC}_{k,t_0:t_T}$ [%], i.e., the fraction of sky covered, for the four geolocations $k \in \mathcal{K}$ surrounding the PV sites. These forecasts are provided by the Global Forecast System (GFS) publicly available in [33], which yields a temporal resolution of three hours and a spatial resolution of 1°.

In the following, all the inputs are min–max normalized between $[0, 1]$ for alleviating underlying optimization issues, especially for the training of neural networks. Note that Section 4.2 also investigates the loss of performance when a type of inputs, e.g., series-specific features or NWP, is removed.

2.3. Machine learning techniques

We combine six different machine learning techniques with the cross-learning approach for assessing their respective performance on our local PV generation forecasting task. Let $\hat{\mathcal{P}}_j = \left(\hat{p}_{j,t_0:t_T}^{\text{PV}}\right)^\intercal$ and $\mathcal{X}_j = \left(\mathbf{x}_j^s, \mathbf{x}_{j,t_0:t_T}^f\right)^\intercal$ be the PV generation predictions of the site j and its associated inputs.

The first machine learning technique is the feed-forward neural network (FFNN), which is the traditional neural network architecture used in the literature [27]. The FFNN consists in two linear transformations, with a non-linear activation in between:

$$\hat{\mathcal{P}}_j = \mathbf{W}_2 \; f^{\text{Relu}}(\mathbf{W}_1 \mathcal{X}_j + \mathbf{b}_1) + \mathbf{b}_2 \tag{4}$$

where $\mathbf{W}_1 \in \mathbb{R}^{d \times |\mathcal{X}_j|}$, $\mathbf{W}_2 \in \mathbb{R}^{|\mathcal{P}_j| \times d}$, $\mathbf{b}_1 \in \mathbb{R}^{\in d}$ and $\mathbf{b}_2 \in \mathbb{R}^{|\mathcal{P}_j|}$ are parameters to be trained, f^{Relu} is the element-wise application of the rectified linear activation function, and $|.|$ stands for the cardinality of the associated vector. Note that the dimension d, defining the number of neurons in the hidden layer, is a task-dependent hyperparameter to be tuned.

The computational power of FFNN can be increased by simply stacking successive non-linear transformations [28]. This 'deep' version of FFNN is denoted as S-FFNN. For this variant, two types of hyperparameters can be tuned: the number of hidden layers, i.e., the non-linear transformation $f^{\text{Relu}}(.)$ in Eq. (4), and the number of neurons per hidden layer.

An alternative for modeling dynamic processes such as time series are recurrent neural networks (RNN). RNN are self-connected units whose recurrent connection h_t allows the network to have a memory of previous time steps. Given $\mathcal{X}_{j,t} = \left(\mathbf{x}_j^s, \mathbf{x}_{j,t}^f\right)^\intercal$, the predictions are obtained by iterating the following equations from $t = t_0$ to t_T:

$$h_t = \mathcal{H}\left(\mathbf{W}_3 \mathcal{X}_{j,t} + \mathbf{W}_4 h_{t-1} + \mathbf{b}_3\right) \tag{5a}$$

$$\hat{p}_{j,t}^{\text{PV}} = \mathbf{W}_5 h_t + \mathbf{b}_4 \tag{5b}$$

where $\mathbf{W}_3 \in \mathbb{R}^{d \times |\mathcal{X}_{j,t}|}$, $\mathbf{W}_4 \in \mathbb{R}^{d \times d}$, $\mathbf{W}_5 \in \mathbb{R}^{1 \times d}$, $\mathbf{b}_3 \in \mathbb{R}^{\in d}$ and $\mathbf{b}_4 \in \mathbb{R}^1$ are parameters to be trained. $\mathcal{H}$ is the hidden layer function, which was traditionally an element-wise application of a sigmoid function. The hidden dimension d is also a hyperparameter to be tuned.

However, when using an element-wise sigmoid function as $\mathcal{H}$, RNN can suffer from exploding or vanishing gradients issues for modeling long term dependencies [34]. To mitigate the latter issue, Hochreiter and Schmidhuber in [35] suggested the long short term memory (LSTM) architecture. The $\mathcal{H}^{\text{LSTM}}$ is composed of a memory cell, whose interactions with the inputs are controlled via three multiplicative gates: the input, forget and output gates. These multiplicative gates allow the memory cell to store and access information as time elapses, while preserving the gradient flow. For this reason, we use the LSTM architecture $\mathcal{H}^{\text{LSTM}}$, whose mathematical expressions is completely detailed in [36].

While the LSTM allows an access to previously fed information, no connection with (known) later inputs are permitted in its original form. The bidirectional LSTM (denoted BLSTM) typically enable this access on available future information by processing the input sequence in both directions [37]:

$$\overrightarrow{h_t} = \mathcal{H}^{\text{LSTM}}\left(\mathcal{X}_{j,t}, \overrightarrow{h}_{t-1}\right) \tag{6a}$$

$$\overleftarrow{h_t} = \mathcal{H}^{\text{LSTM}}\left(\mathcal{X}_{j,t}, \overleftarrow{h}_{t+1}\right) \tag{6b}$$

$$\hat{p}_{j,t}^{\text{PV}} = \mathbf{W}_6 \overrightarrow{h}_t + \mathbf{W}_7 \overleftarrow{h}_t + \mathbf{b}_5 \tag{6c}$$

where $\mathbf{W}_6 \in \mathbb{R}^{1 \times d}$, $\mathbf{W}_7 \in \mathbb{R}^{1 \times d}$ and $\mathbf{b}_5 \in \mathbb{R}^1$ are the parameters of the BLSTM output layer.

Another efficient machine learning technique is the gradient boosted regression trees (GBRT), which was applied successfully on various regression tasks; see e.g., [38]. GBRT is an ensemble method, where each regression tree is estimated in a stage-wise manner to produce collectively a strong learner. Each boosting iteration trains a new regression tree against the residual of the previously trained ones by minimizing a loss function. In a deterministic regressive framework, the loss function is typically the mean squared error function. Following [38], a separate GBRT model is fitted for each step of the prediction horizon using all the available inputs. Four hyperparameters have to be adjusted: the number of boosting iterations M_{GBRT}, the depth of each tree J_{GBRT}, the learning rate λ_{GBRT}, and the subset of inputs used at each split m_{GBRT}.

Finally, we also investigate the k-nearest neighbor technique, which is a simple yet competitive machine learning technique. Similarly to the GBRT, we instantiate a separate k-NN model for each step of the prediction horizon. Hence, when receiving a newly input, the k-NN algorithm is specified by two operations. First, it ranks the training samples of the associated dataset based on the similarity of their inputs with the new one. Then, the prediction is formed by aggregating the k-nearest target values using a simple arithmetic mean with equal weight. Let $\{i, j\}$ be respectively indexes of the training and new samples, the

Table 1
Series-specific features of all the PV sites.

Series-specific features	Sites															
	Dataset											Test set				
	# 1	#2	# 3	# 4	# 5	# 6	# 7	# 8	# 9	# 10	# 11	# 12	# 13	# 14	# 15	# 16
Orientation γ from south [°]	10	10	−40	35	70	0	−40	20	35	85	−45	55	0	50	50	−40
Tilt angle β [°]	20	15	55	15	25	25	35	40	21	90	35	45	20	45	35	15
Peak power installed c [kW]	11.28	9.95	3.76	5.1	11.6	9.95	9.95	5.1	9.84	9.5	24	9.6	9.84	3.333	5.1	8.25

similarity function between both samples are computed based on the following Euclidean distance:

$$\sqrt{(I^{\mathrm{PV}}_{j,t} - I^{\mathrm{PV}}_{i,t})^2 + (\overline{T}^{A}_{j,t} - \overline{T}^{A}_{i,t})^2 + (\overline{\mathrm{CC}}_{j,t} - \overline{\mathrm{CC}}_{i,t})^2} \tag{7}$$

where $I^{\mathrm{PV}}_{\{i,j\},t}$, $\overline{T}^{A}_{\{i,j\},t}$ and $\overline{\mathrm{CC}}_{\{i,j\},t}$ are respectively the tilted global irradiance, the averaged ambient air temperature and averaged cloud cover obtained from the NWP for samples $\{i, j\}$ at the prediction step t.

All the neural networks are optimized using Adam [39], an adaptive learning rate version of stochastic gradient descent, by minimizing the mean square error function. An early stopping criteria is used to prevent overfitting. Early stopping is also applied for GBRT, where the maximum number of boosting iterations M_{GBRT} is fixed to 150. A random search is employed to find the best hyper-parameters associated with each ML technique [40]. The search range of the number of neurons is $\{10, 25, 50, 100, 250\}$, while the number of hidden layer can vary between $\{2, 3, 4, 5\}$ for the S-FFNN. The initial learning rate of Adam is also tested between $\{0.1, 0.01, 0.001, 0.0001\}$. Concerning the GBRT, the search ranges are $J_{\mathrm{GBRT}} = \{3, 5, 8, 12\}$, $m_{\mathrm{GBRT}} = \{1/2, 1/3, 1/4, 1/5\}$ and $\lambda_{\mathrm{GBRT}} = \{0.3, 0.2, 0.1, 0.05\}$. Finally, the search range of the number of neighbors k for k-NN is discretized between $\{1, 3, 5, 10, 25, 50, 75, 100, 250, 500\}$.

3. Baseline and metric

Section 3.1 presents the baseline model, i.e., the clear sky-based physical model, while Section 3.2 introduces the evaluation metric for comparing the forecasting methods.

3.1. Baseline

For each prediction step t, The baseline model (denoted Phys.) approximates the PV output power $\hat{p}^{\mathrm{PV}}_{t}$ of site j as follows [41]:

$$\hat{p}^{\mathrm{PV}}_{j,t} \approx c_j \frac{I^{\mathrm{PV}}_{j,t}}{1000}\left(1 + \mathrm{TC}(T^{\mathrm{PV}}_{j,t} - 25)\right) \tag{8}$$

where c_j, $I^{\mathrm{PV}}_{j,t}$ are the peak power installed and the transposed solar radiation defined in Section 2.2. $T^{\mathrm{PV}}_{j,t}$ and TC denote respectively the temperature [°C] and temperature coefficient [%/°C] of the PV panels.

We set TC at −0.45, which is a common value for the widely used c-SI PV module [42]. The panels' temperature $T^{\mathrm{PV}}_{j,t}$ can be estimated as [41]:

$$T^{\mathrm{PV}}_{j,t} = \overline{T}^{\mathrm{A}}_{j,t} + \frac{I^{\mathrm{PV}}_{j,t}}{800}(N^{\mathrm{oct}} - 20) \tag{9}$$

where $\overline{T}^{\mathrm{A}}_{j,t}$ is the averaged ambient air temperature [°C] obtained via the NWP, and N^{oct} denotes the nominal operating cell temperature, which is typically around 48 °C.

3.2. Evaluation metric

We assess the performance of the different models by computing a normalized root mean square (NRMSE) [%] for each considered unobserved PV site. Over one day, the NRMSE of site j is calculated as follows:

$$\mathrm{NRMSE} = \sqrt{\frac{1}{T}\sum_{t=t_0}^{t_T}(p^{\mathrm{PV}}_{j,t} - \hat{p}^{\mathrm{PV}}_{j,t})^2} \cdot \frac{100\%}{c_j} \tag{10}$$

Table 2
Optimal hyperparameters for the neural networks.

	Learning rate	No. units	No. hidden layer
FFNN	0.001	100	1
S-FFNN	0.01	{10,100,10,25}	4
LSTM	0.01	50	1
BLSTM	0.01	25	1

Table 3
Training and inference times of forecasting methods.

Models	Training time [s]	Inference time [s]
FFNN	4	0.02
S-FFNN	9	0.02
LSTM	20	0.04
BLSTM	125	0.2
GBRT	18	0.03

where T is the length of the prediction horizon.

This metric facilitates the comparison between the different PV sites, which varies in terms of peak power installed.

4. Case study

The case study is performed on an Intel® Core™ i7-3770 CPU @ 3.4 GHz with 16 Gb of RAM. The forecasting methods are implemented using scikit-learn, statsmodels and TensorFlow packages in Python 3.6. The data is a collection of sixteen local PV sites from January to mid-August 2021. Their respective orientation, tilt angle and peak power installed are shown in Table 1. Eleven PV sites are used for training the models, with their data spanning from January to mid-July 2021. This set of data is divided into two subsets: (i) a training set (from 01/01/2021 to 14/06/2021), which is used for estimating the parameters of each ML technique. (ii) a validation set (from 15/06/2021 to 14/07/2021) allowing to select optimal hyperparameters and to perform early stopping. The set of selected hyperparameters for the neural networks are listed in Table 2. In the case of the GBRT and k-NN models, the optimal hyperparameters vary according to the prediction step. Overall, the GBRT models adopts a moderate depth of tree $J_{\mathrm{GBRT}} = \{5, 8\}$ with the smallest learning rate 0.05, while the fraction of inputs used at each split are below 1/3. The k parameter is around 100 on average. In addition, Table 3 provides a brief overview concerning the training and inference times of each forecasting model. For neural models, the training time increases along with the complexity of the neural network variation. For GBRT, a different model is needed per prediction step, which consequently augments the training time along with the forecasting horizon. Overall, the training time of all forecasting models are below 2 min in our experiments. Finally, we can observe that the inference time for generating new predictions is lower than one second for all prediction models, which renders them operational for our local PV forecasting task.

The remaining five PV sites are used as a test set with data spanning from mid-July to mid-August. This allows to evaluate the models on out-of-sample data distinct in time and location. We perform the multi-horizon deterministic forecasts at 00.00 a.m. of each day, with a prediction horizon of 68 quarter hours, i.e., between 5.30 a.m. and 22.30 p.m. each day. Globally, after cleaning the data, 1096 daily samples are used in the training set, 297 for the validation set and 126 for the test set.

Table 4
Averaged NRMSE [%] for all forecasting models over the entire test set.

Model	Total	Sites				
		# 12	# 13	# 14	# 1 5	# 16
GBRT	**10.67**	**10.57**	10.13	11.3	10.6	8.95
k-NN	11.16	11,43	10.46	11.85	11.25	9.16
BLSTM	11.48	10.77	11.48	14.08	10.63	9.78
LSTM	11.79	11.14	11.73	13.84	11.41	9.56
S-FFNN	10.7	11.12	**10.03**	**11.13**	10.43	9.21
FFNN	10.68	10.88	10.54	11.56	**10.42**	**8.84**
Phys.	18.51	16.28	18.1	17.85	17.39	18.72

4.1. Forecast evaluation

Table 4 lists the averaged NRMSE performance of each forecasting model over the entire test set. The NRMSE is computed over all sites in the Total column, while it is evaluated per site in the remaining columns. We can observe that the machine learning techniques clearly outperform the Phys. model on all PV sites. In total, the outcomes reveal a relative increase of 73.5% in terms of NRMSE when using the Phys. model compared to the GBRT model. This gap arises from the fact that the Phys. model ignores the impact of the clouds in its data generating process, which heavily penalizes it. Overall, none of the machine learning techniques clearly establishes a superior performance compared to the others. However, the results suggest that the GBRT, FFNN and S-FFNN models are slightly more precise for our application. Surprisingly, the LSTM and BLSTM architectures provide the worst performance amongst the machine learning techniques. This observation appears, at first sight, counter-intuitive compared with that observed in the literature; see e.g., [20,43]. This may arise from the presence of different sampling frequencies in the data. Indeed, the input data is composed of the series-specific features $\mathbf{x}_j^s$ and the NWP, which are respectively time-invariant and signals covering a 3-h period, while the output is sampled on a quarter-hourly basis. The performances of the LSTM and BLSTM may be altered when processing such mixed frequency data [44]. In contrast, this issue does not occur with the FFNN and S-FFNN as they use a static fixed size window encompassing all the inputs. In the same vein, increasing further the depth of the FFNN leads to a limited increase of performance only for particular sites. The tree-based ensemble method GBRT provides the best performance in our case study. This is aligned with observations in [45], which considers them a natural choice for competition settings as they provide an excellent compromise between efficiency and their straightforward application. Finally, the *k*-NN provides an interesting alternative for practical use. First, the *k*-NN is characterized by a single hyperparameter, i.e., the number of neighbors, which has a direct physical meaning. Besides, this technique can be continuously updated by simply storing the newly data of monitored sites in the dataset. The major downside of this technique concerns its computational cost when performing the prediction, which necessitates to search over the whole training sample for obtaining the *k*-nearest neighbors.

These global results are complemented with Fig. 2, which depicts the averaged NRMSE performance over the entire prediction horizon. Similarly to Table 4, Fig. 2 shows a significant difference between the Phys. model and the machine learning techniques. Overall, all the machine learning techniques perform equally well across the prediction horizon between 10.00 a.m. and 8.00 p.m.. Interestingly, the FFNN, LSTM and BLSTM models have difficulties to predict the beginning and final parts of the day. Naturally, the prediction steps with the highest errors are the quarter hours between 12.00 a.m. and 5.00 p.m., for which the rays of sunshine are the most impacted by the cloud movements.

In that sense, Fig. 3 depicts the difference of behaviors between the GBRT and Phys. model for three days characterized by different atmospheric conditions. Typically, the Phys. model (the black dotted line) follows a half-sine wave over the day. In addition, as site 12 is characterized by an orientation $\gamma_j = 55°$ and tilt angle $\beta_j = 45°$, the effect of the transposition model on the clear-sky irradiance is noticeable at the beginning of the day. Yet, Figs. 3(b) and 3(c) clearly demonstrate the limits of the Phys. model on cloudy days. In that case, the GBRT technique approximates the PV output power more precisely, although it is not able to capture all the quick variations of the PV output power. One reason could be the low temporal resolution of the obtained NWP. In this respect, the importance of NWP and the other input features on the performance of the GBRT is further investigated in the following section.

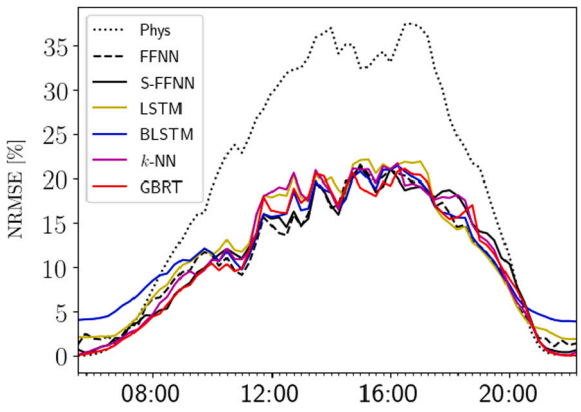


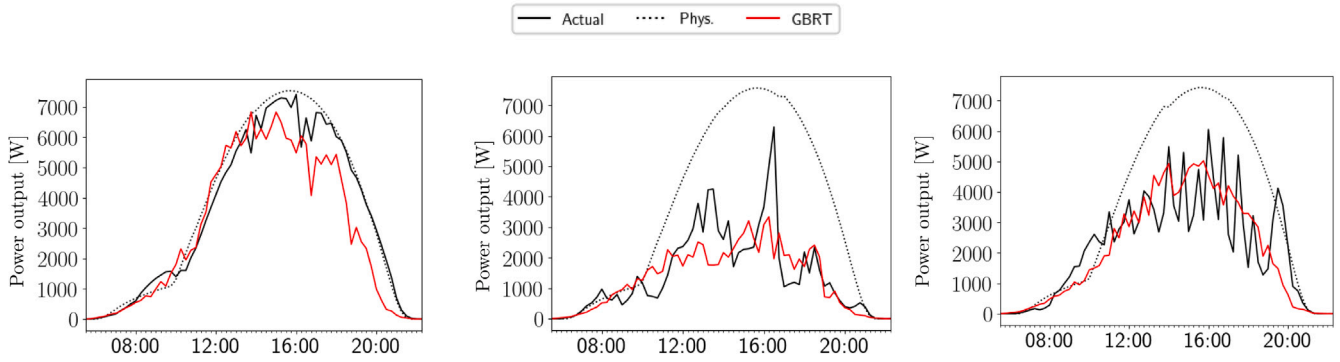
Fig. 2. Averaged NRMSE across all sites for each prediction step and each forecasting model.

4.2. Sensitivity analysis

In this section, we perform a sensitivity analysis on the number of available monitored PV sites and input features when predicting the unobserved PV sites. More specifically, we investigate the decrease in the GBRT performance when a more limited number, i.e., $\{8, 5, 3, 1\}$, of monitored PV sites is available. In this case, the monitored PV sites are selected based on their order of appearance in the dataset. In complement, a similar analysis is conducted on the input features, where we remove distinctively the tilted solar irradiance $I_{j,t}^{\text{PV}}$, the NWP and the series-specific features $\mathbf{x}_j^s$. Note that the removal of $\mathbf{x}_j^s$ also affects the input $I_{j,t}^{PV}$ as we do not have information on the tilt angle β_j or orientation γ_j of the PV site, which is necessary for the transposition model. Besides, as the latitude ϕ_j and longitude λ_j is also unknown, the model is fed for all PV sites with the clear-sky irradiance of the Ineichen–Perez model for $\phi = 50.5$ and $\lambda = 3.5$, which corresponds to the center of the area delimited by the geolocations of the NWP. Of course, we assume that the peak power installed c_j is still known $\forall j \in \mathcal{J}$.

From Table 5, it can be observed that the number of available PV sites has a significant impact on the GBRT performance, with a relative increase in the NRMSE of around 25% for the Total column in case of one available PV site. Note that the NRMSE with one site is still above the Phys. model. Besides, we can see that the NRMSE starts to decline substantially below five available sites. This is in concordance with the cross-learning approach, which boosts the generalization ability of the machine learning models by mixing different time series such that the models have a greater number of samples for adequately approximating the data generating process.

The lower half of Table 5 concerns the performance of the GBRT model without certain inputs. It can be observed that the removal of NWP has a slightly worse impact than the removal of $I_{j,t}^{\text{PV}}$, while the withdrawal of $\mathbf{x}_j^s$ is the most critical in terms of averaged NRMSE. The obtained results confirm the importance of NWP and irradiance data in the forecasting task of local PV sites. In this line, the public availability



(a) Forecasts of a clear-sky day (b) Forecasts of a cloudy day (c) Forecasts of a partial cloudy day

Fig. 3. Multi-horizon deterministic forecasts of the PV site 12 on 17th July 2021 (a clear-sky day, Fig. 3(a)), 27th July 2021 (a cloudy day, Fig. 3(b)) and 3rd August 2021 (a partial cloudy day, Fig. 3(c)).

Table 5
Sensitivity analysis of the GBRT model on the number of available sites and input features.

Model	Total	Sites				
		# 12	# 13	# 14	# 15	# 16
GBRT	10.67	10.57	10.13	11.3	10.6	8.95
GBRT (8 sites)	10.83	11.13	10.24	11.6	10.74	8.88
GBRT (5 sites)	10.91	11.19	10.32	11.83	10.88	8.92
GBRT (3 sites)	12.74	14.61	10.37	13.52	13.33	9.16
GBRT (1 site)	13.34	14.83	10.76	14.08	13.55	10.54
GBRT (w/o $I_{j,t}^{PV}$)	11.82	12.53	10.24	13.44	12.38	8,83
GBRT (w/o NWP)	12,1	12,69	10,14	11,59	12,34	9,83
GBRT (w/o $\mathbf{x}_j^s$)	12.34	14.07	9.83	13.29	12.77	9.05

of satellite images, such as cloud-moving vector models, or higher frequency NWP could be a great asset for fostering the performance of the cross-learning forecasting models.

5. Conclusion

This paper proposes a new framework for developing a cross-learning forecasting approach applied for unobserved PV time series. The benefit of this approach is twofold: this allows increasing the data availability for approximating the data generating process, while providing a scalable method for predicting each individual PV site.

Six machine learning techniques, i.e., the k-NN, four variants of neural networks and gradient boosted trees, are combined with this approach, and are assessed on five new PV sites. The outcomes reveal that the obtained forecasting models yield more accurate predictions than a clear sky-based physical approach, which shows a relative increase of 73.5% in NRMSE compared with the best models. Regarding the specificities of our dataset, the best models are GBRT, FFNN and S-FFNN. The LSTM and BLSTM models provide a slightly worse performance compared to the top three with a relative increase above 8.8% in NRMSE, which can be due to the presence of mixed data frequency in the data. The k-NN method provides a good trade-off between efficiency, showing a limited increase of 5.6% in NRMSE compared to the best models, and simplicity, where it necessitates the tuning of the single (interpretable) hyperparameter k.

A first perspective might be to increase the spatio-temporal resolution of the NWP inputs for better capturing the ramping trajectories of local PV generation. Another interesting perspective could be to investigate the application of a similar approach on a dataset composed of both solar PV and wind generation. Hence, the resulting final model could be used for predicting each individual renewable production in low-voltage systems.

CRediT authorship contribution statement

J. Bottieau: Conceptualization, Methodology, Software, Validation, Visualization, Writing – original draft. **Z. De Grève:** Conceptualization, Methodology, Validation, Writing – review & editing. **T. Piraux:** Conceptualization, Validation. **A. Dubois:** Data curation, Conceptualization. **F. Vallée:** Methodology, Validation, Writing – review & editing. **J.-F. Toubeau:** Methodology, Validation, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, Z. Hu, Photovoltaic and solar power forecasting for smart grid energy management, CSEE J. Power Energy Syst. 1 (4) (2015) 38–46.
[2] M. Hupez, J.-F. Toubeau, Z. De Grève, F. Vallée, A new cooperative framework for a fair and cost-optimal allocation of resources within a low voltage electricity community, IEEE Trans. Smart Grid 12 (3) (2021) 2201–2211.
[3] cVPP—Community-Based Virtual Power Plant: A Novel Model of Radical Decarbonisation Based on Empowerment of Low-Carbon Community Driven Energy Initiatives, Interreg NW Europe, 2017–2020, [Online]. Available: https://www.nweurope.eu/projects/project-search/cvpp-community-based-virtual-power-plant/.
[4] Z. De Grève, et al., Machine learning techniques for improving self-consumption in renewable energy communities, Energies (18) (2020).
[5] J. Moon, J. Kim, P. Kang, E. Hwang, Solving the cold-start problem in short-term load forecasting using tree-based methods, Energies 13 (4) (2020).
[6] C. Aguilar-Palacios, S. Muñoz-Romero, J.l. Rojo-Álvarez, Cold-start promotional sales forecasting through gradient boosted-based contrastive explanations, IEEE Access 8 (2020) 137574–137586.
[7] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.M. de Pison, F. Antonanzas-Torres, Review of photovoltaic power forecasting, Sol. Energy 136 (2016) 78–111.
[8] D. Yang, Choice of clear-sky model in solar forecasting, J. Renew. Sustain. Energy 12 (2020) 026101.
[9] X. Sun, J.M. Bright, C.A. Gueymard, B. Acord, P. Wang, N.A. Engerer, Worldwide performance assessment of 75 global clear-sky irradiance models using principal component analysis, Renew. Sustain. Energy Rev. 111 (2019) 550–570.
[10] M. Diagne, M. David, P. Lauret, J. Boland, N. Schmutz, Review of solar irradiance forecasting methods and a proposition for small-scale insular grids, Renew. Sustain. Energy Rev. 27 (2013) 65–76.
[11] M.J. Mayer, G. Gróf, Extensive comparison of physical models for photovoltaic power forecasting, Appl. Energy 283 (2021) 116239.
[12] C. Yang, A.A. Thatte, L. Xie, Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation, IEEE Trans. Sustain. Energy 6 (1) (2015) 104–112.
[13] R. Amaro e Silva, M. C. Brito, Impact of network layout and time resolution on spatio-temporal solar forecasting, Sol. Energy 163 (2018) 329–337.
[14] M.N. Akhter, S. Mekhilef, H. Mokhlis, N. Mohamed Shah, Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques, IET Renew. Power Gener. 13 (7) (2019) 1009–1023.

[15] C. Persson, P. Bacher, T. Shiga, H. Madsen, Multi-site solar power forecasting using gradient boosted regression trees, Sol. Energy 150 (2017) 423–436.
[16] B. Wolff, J. Kuehnert, E. Lorenz, O. Kramer, D. Heinemann, Comparing support vector regression for PV power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data, Sol. Energy 135 (2016) 197–208.
[17] F. Golestaneh, P. Pinson, H.B. Gooi, Very short-term nonparametric probabilistic forecasting of renewable energy generation— with application to solar energy, IEEE Trans. Power Syst. 31 (5) (2016) 3850–3863.
[18] J.-F. Toubeau, J. Bottieau, F. Vallée, Z. De Grève, Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets, IEEE Trans. Power Syst. 34 (2) (2019) 1203–1215.
[19] Y. Lu, I. Murzakhanov, S. Chatzivasileiadis, Neural network interpretability for forecasting of aggregated renewable generation, 2021, arXiv preprint, 2106.10476.
[20] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, Y. Du, Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism, IEEE Access 7 (2019) 78063–78074.
[21] J. Dumas, C. Cointe, X. Fettweis, B. Cornélusse, Deep learning-based multi-output quantile forecasting of pv generation, in: 2021 IEEE Madrid PowerTech, 2021, pp. 1–6.
[22] D.J. Gagne, A. McGovern, S.E. Haupt, J.K. Williams, Evaluation of statistical learning configurations for gridded solar irradiance forecasting, Sol. Energy 150 (2017) 383–393.
[23] M. Jamaly, J. Kleissl, Spatiotemporal interpolation and forecast of irradiance data using Kriging, Sol. Energy 158 (2017) 407–423.
[24] A.W. Aryaputera, D. Yang, L. Zhao, W.M. Walsh, Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging, Sol. Energy 122 (2015) 1266–1278.
[25] A.-A. Semenoglou, E. Spiliotis, S. Makridakis, V. Assimakopoulos, Investigating the accuracy of cross-learning time series forecasting methods, Int. J. Forecast. 37 (3) (2021) 1072–1084.
[26] P. Montero-Manso, R.J. Hyndman, Principles and algorithms for forecasting groups of time series: Locality and globality, Int. J. Forecast. 37 (4) (2021) 1632–1653.
[27] A.F. Zambrano, L.F. Giraldo, Solar irradiance forecasting models without on-site training measurements, Renew. Energy 152 (2020) 557–566.
[28] J. Lago, K.D. Brabandere, F.D. Ridder, B.D. Schutter, Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data, Sol. Energy 173 (2018) 566–577.
[29] J.-F. Toubeau, J. Bottieau, Y. Wang, F. Vallee, Interpretable probabilistic forecasting of imbalances in renewable-dominated electricity systems, IEEE Trans. Sust. Energy (2021) 1.
[30] B. Liu, R. Jordan, Daily insolation on surfaces tilted towards equator, ASHRAE J. 10 (1961) 53–59.
[31] P. Ineichen, R. Perez, A new airmass independent formulation for the linke turbidity coefficient, Sol. Energy 73 (2002) 151–157.
[32] R.W. Andrews, J.S. Stein, C. Hansen, D. Riley, Introduction to the open source PV LIB for python photovoltaic system modelling package, in: 2014 IEEE 40th PVSC, 2014, pp. 0170–0174.
[33] Global Forecast System, National Centers for Environmental Information, 2021, [Online]. Available: https://www.ncei.noaa.gov/thredds/dodsC/model-gfs-004-files/.
[34] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, J. Schmidhuber, A novel connectionist system for unconstrained handwriting recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (5) (2009) 855–868.
[35] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780.
[36] J.-F. Toubeau, T. Morstyn, J. Bottieau, K. Zheng, D. Apostolopoulou, Z. De Grève, Y. Wang, F. Vallée, Capturing spatio-temporal dependencies in the probabilistic forecasting of distribution locational marginal prices, IEEE Trans. Smart Grid 12 (3) (2021) 2663–2674.
[37] A. Graves, J. Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures, Neural Netw. 18 (5–6) (2005) 602–610.
[38] S. Ben Taieb, R.J. Hyndman, A gradient boosting approach to the Kaggle load forecasting competition, Int. J. Forecast. 30 (2) (2014) 382–394.
[39] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2017, arXiv preprint, 1412.6980.
[40] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, J. Mach. Learn. Res. 13 (10) (2012) 281–305.
[41] Y. Wang, N. Zhang, Q. Chen, D. Kirschen, P. Li, Q. Xia, Data-driven probabilistic net load forecasting with high penetration of behind-the-meter PV, IEEE Trans. Power Syst. 33 (2018) 3255–3264.
[42] B. Mihaylov, T.R. Betts, A. Pozza, H. Müllejans, R. Gottschalg, Uncertainty estimation of temperature coefficient measurements of PV modules, IEEE J. Photovolt. 6 (6) (2016) 1554–1563.
[43] J. Bottieau, L. Hubert, Z. De Grève, F. Vallée, J.-F. Toubeau, Very-short-term probabilistic forecasting for a risk-aware participation in the single price imbalance settlement, IEEE Trans. Power Syst. 35 (2) (2020) 1218–1230.
[44] K. Ma, H. Leung, A novel LSTM approach for asynchronous multivariate time series prediction, in: 2019 IJCNN, 2019, pp. 1–7.
[45] M. Landry, T. Erlinger, D. Patschke, C. Varrichio, Probabilistic gradient boosting machines for GEFCom2014 wind forecasting, Int. J. Forecast. 32 (2016).