



HAL
open science

A Novel Approach for Seamless Probabilistic Photovoltaic Power Forecasting Covering Multiple Time Frames

Thomas Carriere, Christophe Vernay, Sébastien Pitaval, Georges Kariniotakis

► **To cite this version:**

Thomas Carriere, Christophe Vernay, Sébastien Pitaval, Georges Kariniotakis. A Novel Approach for Seamless Probabilistic Photovoltaic Power Forecasting Covering Multiple Time Frames. *IEEE Transactions on Smart Grid*, 2020, 11 (3), pp.2281 - 2292. 10.1109/TSG.2019.2951288 . hal-02369413

HAL Id: hal-02369413

<https://hal.science/hal-02369413v1>

Submitted on 19 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Novel Approach for Seamless Probabilistic Photovoltaic Power Forecasting Covering Multiple Time Frames

Thomas Carriere, Christophe Vernay, Sebastien Pitaval, George Kariniotakis *Senior Member, IEEE*

Abstract—Uncertainty in the upcoming production of photovoltaic (PV) plants is a challenge for grid operations and also a source of revenue loss for PV plant operators participating in electricity markets, since they have to pay penalties for the mismatch between contracted and actual productions. Improving PV predictability is an area of intense research. In real-world applications, forecasts are often needed for different time frames (horizon, update frequency, etc.) and are derived by dedicated models for each time frame (i.e. for day ahead and for intraday trading). This can result in both different forecasted values corresponding to the same horizon and discontinuities among time-frames. In this paper we address this problem by proposing a novel seamless probabilistic forecasting approach able to cover multiple time frames. It is based on the Analog Ensemble (AnEn) model, however it is adapted to consider the most appropriate input for each horizon from a pool of available input data. It is designed to be able to start at any time of day, for any forecast horizon, making it well-suited for applications like continuous trading. It is easy to maintain as it adapts to the latest data and does not need regular retraining. We enhance short-term predictability by considering data from satellite images and in situ measurements. The proposed model has low complexity compared to benchmark models and is trivially parallelizable. It achieves performance comparable to state-of-the-art models developed specifically for the short term (i.e. up to 6 hours) and the day ahead. The evaluation was carried out on a real-world case comprising three PV plants in France, over a period of one year.

Index Terms—Analog-Ensemble Model, Photovoltaics, Probabilistic Forecasting, Satellite Imagery

I. INTRODUCTION

IN recent years, continuous reduction of the cost of photovoltaic (PV) panels together with favorable policies have been the main drivers behind increased PV installation in electricity grids. However, weather-dependent PV production brings several challenges for grid operators. Accurate forecasts of the power output of PV plants are necessary to maintain economic and secure operation of the power system. Equally, PV producers and aggregators participating directly in electricity markets attempt to avoid financial penalties for any

imbalances due to a mismatch between what they propose to the market and what they actually produce.

In practice, decisions are often made over different cycles, i.e. trading on the day-ahead market requires models able to provide forecasts for the medium term (up to 48h ahead), while improving positions on the intraday market or managing coupled storage devices calls for short- or very short-term forecasting models (i.e. up to 6h ahead or minutes ahead respectively). The decisions related to the participation in an electricity market must take into account the uncertainty of the PV production [1], which is why we chose to have a probabilistic framework in this paper. Although the work proposed in this paper is applicable for the case of other weather-dependent renewables, we focus here on PV prediction, since the state of the art comprises much more distinct approaches developed for short and medium time frames.

The state of the art in solar power forecasting has developed rapidly in recent years. Reference [2] provides a fairly complete literature review of research in the field. Generally, the features used as inputs are largely dependent on the forecast horizon. Short-term forecasting (0-6 hours) mostly employs endogenous data, although input from Numerical Weather Predictions (NWP) and meteorological records can be used [3]. Works considering satellite imagery appear, i.e. [4] and [5], while data from neighboring PV plants can be employed in spatio-temporal models [6], [7]. Data from sky imagers are also useful for the very short term (up to a few minutes) [8], but harder to apply as they require significant preprocessing work. For the medium-term (up to few days ahead), forecasts mostly rely on NWPs.

Regarding forecasting techniques, linear auto-regressive models are popular, as they are light and can issue forecasts at any time of the day using the same model [9]. Numerous machine learning models are also used, such as Artificial Neural Networks (ANN) [9] and Support Vector Machines (SVM) [10]. Recently, several new methods have emerged. The Extreme Learning Machine (ELM) is a fairly popular variant of ANN [11], [12], [13]. Gaussian Process Regression (GPR) [14] and Markov Chain (MC) [15] models are also becoming more frequent in the literature. The global forecasting competition GEFCOM 2014 [16] showed that the most efficient algorithms were often non-parametric, such as Quantile Regression Forests (QRF) [17] and Gradient Boosting (GB) [18]. However, [19] performed a comparison of several non-parametric models and found that the performance difference between the tested models was low.

T. Carriere and G. Kariniotakis are with MINES ParisTech, PSL University, Centre PERSEE, CS 10207, 1 rue Claude Daunesse, 06904 Sophia Antipolis Cedex, France (e-mail: thomas.carriere, georges.kariniotakis each with @mines-paristech.fr)

S. Pitaval and T. Carriere are also with Third Step Energy, 55 allée Pierre Ziller, 06560 Sophia Antipolis Cedex, France (e-mail: thomas.carriere, sebastien.pitaval each with @thirdstep.energy)

C. Vernay and S. Pitaval are also with SOLAÏS, 55 allée Pierre Ziller, 06560 Sophia Antipolis Cedex, France (e-mail: cvernay@solais.fr)

For each time frame, dedicated models are employed that use as input the data sources that best contribute to performance during that particular frame. The operators or decision-support tools are presented with probabilistic forecasts from different models that are not always compatible, due to either discontinuities between time frames, or differences in the forecast values that correspond to the same horizon. In this paper we address this technical issue by proposing a forecasting approach independent of the time frame. This requires, however, that its performance must not be significantly lower than that of the state-of-the-art models specifically developed for each time frame.

The motivation goes beyond the above-mentioned technical issue, since the proposed approach simplifies the overall prediction model chain and thus makes easier the deployment of predictive analytics in a context of large-scale penetration of renewable energy sources. It is much easier for Renewable Energy Sources (RES) producers to employ just one model for all of their forecasting needs.

To achieve our goal of attaining good performance for any time frame, the proposed model must be able to use all of the data mentioned above, depending on the forecast horizon. Regarding the choice of model, we based our approach on models from the Analog Ensemble (AnEn) family, which have the dual advantage of being lightweight compared to other models and of naturally providing non-parametric probabilistic forecasts. The base AnEn model that is extended in this paper was described in [20], where past NWP are used to forecast 10-m wind speed and 2-m temperature. In [21], the authors implemented a model in which they also looked for analogs using NWPs, but applying a different metric than that of [20]. In [22], the authors used the AnEn model to forecast probabilistic PV power for three large power plants. They emphasized the very low computational time needed to produce the forecasts compared to other models.

Several papers have also proposed corrections or adaptations of the AnEn model to obtain better forecasts. Reference [23] proposed a modification of the Euclidian distance frequently used as a metric to evaluate analogs and used measures along with NWP to search for analogs. This reduced the forecast horizon, making it necessary to run the model multiple times to obtain a complete forecast. In [24] a 20% improvement over the standard AnEn model with a brute-force optimization of the parameters was demonstrated. Several algorithmic variations of the AnEn ensemble, along with a dynamic way of selecting the number of analogs to retain in the ensemble, and a wrapper method to dynamically optimize the model parameters, were proposed in [25]. Reference [26] extended the AnEn model with an Artificial Neural Network (ANN) which significantly improved the forecasts. Finally, [5] proposed an Analog method to produce short-term forecasts of solar irradiance using geostationary satellite images only.

Overall, the AnEn model still has some drawbacks that we address in this paper. In previous implementations of the AnEn, most of the time only NWPs are used, and thus forecasts are always the same until a new NWP run is available. It is thus not possible to use state-of-the-art model settings in an intra-day configuration. Besides, although some papers have

used data other than NWP in an AnEn setting, no model has used both NWP and other sources of data. Moreover, previous works do not present how they generate probabilistic forecasts from analogs. The key contributions of this paper to overcome these drawbacks are:

- 1) A methodology to extend the data that can be used in AnEn models, which includes a pre-processing step for using satellite images with a low computational cost;
- 2) A filter approach to compute the model parameters based on historical data without the need for a costly wrapper or offline optimization procedure. This results in a model that is "plant-free", because the model can be used as is on any new PV plant and will directly compute its parameters without a learning process;
- 3) A formal way to produce probabilistic forecasts from the Analog Ensemble using a Kernel Density Estimation (KDE);

These contributions ultimately result in a model that is dependent on the forecast horizon and the starting time of the forecast, so that the number of parameters are not multiplied when starting the model at different times. This is a unique seamless model for short and medium time frames that allows us to simplify the overall model chain in practical applications without compromising accuracy. This model is also able to have state-of-the-art performance in the 5-minute to 36-hours horizon range, and thus can be used in all energy applications such as control of a PV/Storage system combined plants, or day-ahead bidding in an electricity market.

The paper is structured as follows: in Section II, the modifications made to the AnEn model are presented. Then Section IV presents the case study on which we tested the extended model, comparing it to a state-of-the-art forecasting model and Section V presents the results of the study. Finally, Section VI draws the conclusions of the paper.

II. ANALOG ENSEMBLE MODEL

As mentioned in the introduction, the modeling approach we based our work on is the Analog Ensemble model [22], which presents similarities to a k-Nearest-Neighbor (kNN) approach. In this section we show the main improvements we made on the existing model, which are the modification of the AnEn metric used to find the analogs, the derivation of the PDF from the analogs and the weighting of the features.

A. The AnEn Metric

The aim of the AnEn model is to generate a set of past observations considered similar to the situation we want to forecast, and use this set to build the forecast density. Initially we generate an N -member ensemble for a given lead-time by computing a metric between the situation to forecast and all of the past situations. Then we select the N most similar to past situations and look for the PV power measured at the time of these similar situations. These N measures constitute the analog ensemble, and each analog can be thought of as a sample from the probability density function (PDF) of the PV power.

In the most general formulation of the metric taken from [20], only NWP were used as inputs. The metric used in this paper is based on the one defined in [20] but it is adapted to allow different sources of data to be taken into account. The similarity between an instant t for which a forecast is requested and another instant t' in the past is written as follows:

$$D(\mathbf{X}_t, \mathbf{H}_{t'}^h) = \sum_{i=1}^{N_v} w_i^h \sqrt{\sum_{j=-k}^0 (\mathbf{X}_{i,t+j} - \mathbf{H}_{i,t'+j}^h)^2} \quad (1)$$

where N_v is the number of features used as input, h is the forecast horizon (that is, the difference between the time t and the time when the forecasting model is run) and k is a parameter that indicates the length of the time window over which the metric is computed.

The sets \mathbf{H}^h and \mathbf{X} are two sets containing the features and input of the model. The set \mathbf{H}^h contains features from NWP and clear-sky data, along with the measures and the satellite data lagged h times. This set is dependent on the horizon of the forecast. When we make a forecast with horizon h , the latest measurements and satellite data were observed h time steps ago, and this should be reflected in the historical data set. Following through, \mathbf{X} contains the NWP for time t and the latest measures and satellite data, observed at time $t - h$.

The sets \mathbf{H}^h and \mathbf{X} are scaled and centered, so that each variable contributes roughly to the metric with the same proportion. Thus, only the weights w_i^h can control the relevance of each feature. The weight calculation takes into account the forecast horizon and is presented in the next sub-section.

The different sources of data that the paper aims to integrate are: variables from NWP, in situ measurements, clear-sky profile, and spatial data derived from satellite imagery. The clear-sky profile is an estimation of the solar irradiance on the ground under the assumption that there are no clouds. Following equation (1) and using the notations introduced with this equation, the integration of a clear-sky profile and ground measurements is pretty straightforward.

However, satellite data is more complex to use for two main reasons. The data we use in this work are time series of Global Horizontal Irradiance (GHI) estimations derived for each pixel of the satellite picture using the method proposed in [27]. Thus, there is a high number of features available (more details are provided in subsection II-D), which contain redundant information. Nearest-Neighbors methods such as the AnEn are very sensitive to both the number of features and their redundancy. As the number of redundant features increases, all observations become near equidistant in the feature space. Ultimately, the analogs identified by a Nearest-Neighbors are not representative of the conditional distribution of the PV power on the features. The additional processing work required to use satellite data is detailed in Section II-C.

B. Derivation of the Probability Density Function

To our knowledge, previous papers have not presented how they generated the PDF from the analogs. In this paper, the PDF is built with a weighted Kernel Density Estimation (KDE), using the metric value of each member as its weight in

the distribution. Given a set of N_{An} AnEn members P_i , which are power observations made at a time when forecasts from the past were similar to the current observations, according to the AnEn distance, and their distance value D_{An} , the PDF \hat{f} is estimated by:

$$\hat{f}(P) = \frac{1}{\sum_{i=1}^{N_{An}} s_i} \sum_{i=1}^{N_{An}} \frac{s_i}{bw} \left[K\left(\frac{P - P_i}{bw}\right) + K\left(\frac{P + P_i}{bw}\right) + K\left(\frac{P + P_i - 2Pn}{bw}\right) \right] \quad (2)$$

where Pn is the installed power of the given plant, the s_i terms are similarity measures inversely proportional to the distances, K is the Epanechnikov kernel, and bw is the bandwidth of the kernels. The first Epanechnikov evaluation corresponds to the kernel centered around the i -th AnEn member value, and the two others are there to ensure that the integral of the distribution is 1 between 0 and Pn . Outside this range, the PDF is set to 0.

Regarding the bandwidth selection, it is difficult to define guidelines on which bandwidth to choose, as the optimal bandwidth is very dependent on the shape of the true density. When approximating a Gaussian density, the optimal bandwidth can be computed as in [28], but this optimal bandwidth can yield poor results on densities that are not close to Gaussian. A typical example is the case of a mixture Gaussian density composed of the sum of two Gaussian distributions, where the Silverman's optimal bandwidth from [28] fails to see the two components of the distribution.

In any case, it seems not reasonable to make assumptions on the shape of the solar power density, as shown in [13]. Therefore, we implemented the binning method from [29] for deriving the optimal bandwidth, as it yielded better results on the learning set without making such assumptions

C. Dynamic Weights Computation

The calculation of the feature-weights is critical. In previous works [24], [30], [25], these were obtained from an off-line optimization for each power plant and remained the same throughout the testing period. In these cases, a measure of the probabilistic performance of the models was used as the optimization objective. In this paper, we do not use the final performance as a criterion, but we propose a dynamic way to estimate weights based on the most recent data, since the model operates in a sliding window scheme and the weight of the latest measurements will not be the same for a forecast started at noon as for one started at midnight.

The criterion used to quantify the weights is Mutual Information (MI), which is often used in machine learning for feature selection [31]. This is a measure of how much the fact of knowing a variable reduces the uncertainty of another variable. The MI between two random variables X and Y , knowing their respective marginal density distributions p_X and p_Y and their joint density distribution $p_{X,Y}$, is:

$$MI(X, Y) = \int \int p_{X,Y}(x, y) \log \left(\frac{p_{X,Y}(x, y)}{p_X(x)p_Y(y)} \right) dx dy \quad (3)$$

The main reason we chose this criterion rather than simpler ones, e.g. Pearson's correlation, is because it can identify non-linear relationships between random variables. As we are dealing with both features that are strongly and almost linearly correlated to the production (clear-sky profile, last production measure), and features that are not (temperature from NWP), using a linear correlation criterion would overestimate the weight of the former over the latter. On the other hand, the MI calculation identifies the non linear information contained in the variables and thus avoids giving significant weights only to variables that are linearly correlated with the PV production.

The weight calculation is carried out in two steps. The first step is to evaluate the relevance of each variable individually by calculating its MI with the measured power according to (3). Computing the MI is not trivial, since it requires formulation of both the marginal and joint distributions of the random variables. We implemented a simple method involving a discretization of the random variables prior to calculating the MI. The discretization algorithm is described in [32]. The MI is finally computed as follows:

- First, discretize the random variables X and Y in k bins, so that each bin has the same number of observations
- Then, compute the discrete probabilities of the couple of events (Ai, Bj), with the events being defined as Ai: "x is in the i-th bin" and Bj: "Y is in the j-th bin", for all possible values of i and j. These probabilities are simply computed by counting the occurrence of the events.
- Apply the discrete formula for computing the MI:

$$MI(X, Y) = \sum_{i=1}^k \sum_{j=1}^k p(A_i, B_j) \log \left(\frac{p(A_i, B_j)}{p(A_i)p(B_j)} \right) \quad (4)$$

The second step is to set a limit on the total cumulative weight of each source of data. If such a limit is not set and all the weights are computed independently for each feature, some features are overrepresented, because they are highly colinear. In particular, the features from the satellite data are highly redundant. If we compute their weights independently, they all have significant weights. This strongly overestimates the relevance of these features. By reorganizing the features into groups, we can first compute a global weight for each group that will represent the total allowed contribution from this group. Then, the weights computed independently for each feature are normalized so that the cumulated contribution of each group of features does not exceed the global weight for this group.

In a more formal way, the weights are computed as follows. In the first step, intermediate weights $w'_{i,h}$ are obtained by (5), for each feature i and forecast horizon h . E_{PV} is the time series of energy generation measurements.

Then, the weights are normalized. In practice, the features variables are organized into N_s subsets depending on their source (NWP, measures, satellite data, clear-sky profile). The sets $S_v, v \in [1, \dots, N_s]$ contain the indexes of the features

contained in each of the N_s sets, corresponding to each data source. A global weight $W_{S_v,h}$ is attributed to each of these sets, and for each horizon. We then obtain the final weights with (6).

$$w'_{i,h} = MI(\mathbf{H}_i^h, E_{PV}) \quad (5)$$

$$w_{i,h} = \left(\frac{w'_{i,h}}{\sum_{v=1}^{N_s} \mathbb{1}(i \in S_v) w'_{v,h}} \right) \sum_{k=1}^{N_s} \mathbb{1}(i \in S_k) W_{S_k,h} \quad (6)$$

The computation of $W_{S_v,h}$ should be done taking into account redundancy between variables from the same source. This is especially important for satellite data where the variables are strongly colinear. We propose to use:

$$W_{S_v,h} = \max_{i \in S_v} \{w'_{i,h}\} \quad (7)$$

This means that each source of data contributes to the metric as much as the weight of its most informative feature does. As a result, we may underestimate the information conveyed by variables that are not strongly correlated to the PV production, but we will not overestimate the global contribution of a source of data that contains numerous colinear variables, since the individual weights of the redundant features will not add up to more than the weight $W_{S_v,h}$ of their source.

Fig. 1 shows the average weights obtained for each type of data and for different time horizons. For every forecast, all features that have non-zero weights for the given forecast horizon are used. The weights represented on Fig. 1a are averaged over all starting times but, when observing forecasts always made at the same time, they are different. For example, when looking at forecasts made at 3AM only 1b, the weight given to the latest measurement is near zero, as it is not informative. For night hours, no weights are computed. They correspond to the white spaces in Fig. 1b, 1c and 1d.

The values of the weights of the different sources correspond well with results reported in the literature concerning which source of data is informative for which horizon. The most recent power measurement is very informative for the first few time steps, but its weight decreases quickly. NWPs are always relevant, but even more so for day-ahead purposes. For horizons shorter than 4 h, they carry less information than the latest measurements and the satellite data. Satellite data are very useful up to 6 h, even though their value decreases steadily. Equally of interest, the clear-sky profile is useful almost only for the beginning and end of the day, when it becomes the most important feature. It is not surprising to see that for these instants with very low incoming irradiance even in clear-sky conditions, the amount of solar power is not dominated by the presence of clouds, but by the Sun's path.

D. Integration of Satellite Data

In the proposed model, estimated GHI time series are derived from the satellite images for each pixel (see Section IV-A). We propose to use these time series as conditioning features for the AnEn model when it looks for analogs. Since the model looks for analogs by matching features on several time steps as described by the parameter k from (1), and the

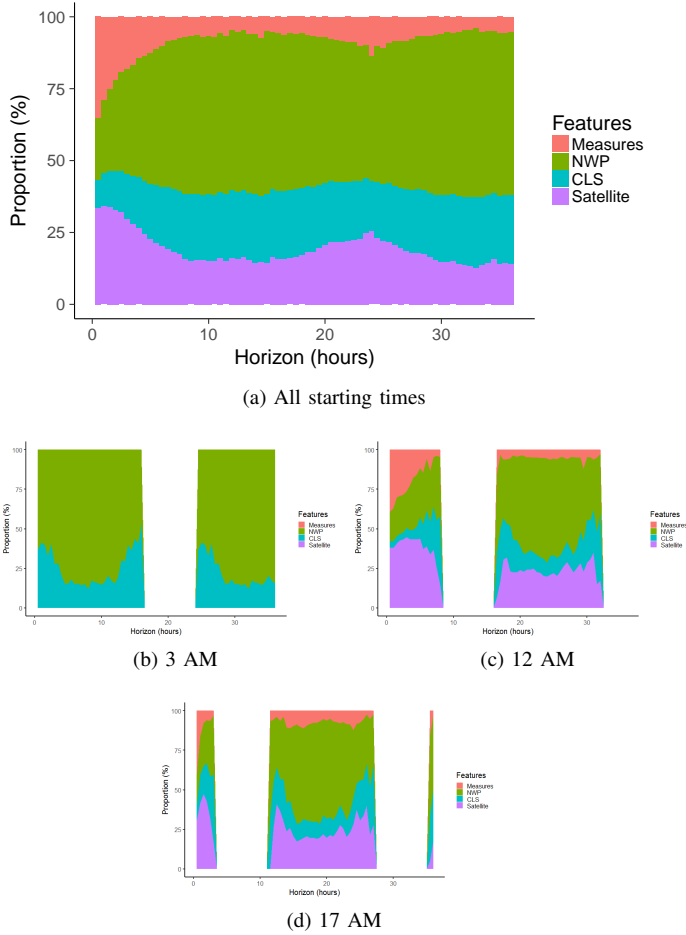


Fig. 1. Weights of the different sources of data depending on the forecast horizon and for different start times. "CLS" refers to the clear-sky profile.

pixels used to condition the forecasts are in the neighborhood of the plant, we use both the spatial and temporal information from the images. Still, this method is simpler than standard ones, i.e. Cloud Motion Vector [33], as it does not try to anticipate the future state of the cloud layer.

The choice of taking a purely "data-driven" approach rather than including a preprocessing step for the images to derive cloud motion information is one of the design requirements set here in order to maintain simplicity in the proposed model chain. The mechanism of the AnEn model, where a series of past images is linked to future situations through the analogs, is a process that is expected to reflect the mechanism according to which the temporal variations in the images that reflect cloud motion impact PV production. Section V demonstrates that the inclusion of satellite images as conditioning features is beneficial to the model. This section describes the selection process for the pixels we use to compute the model.

Using the whole satellite picture is not possible with the AnEn approach because of the dimensionality issues explicated in section II. Thus, two parameters must be estimated before including the satellite-estimated GHI.

The first one is the maximal distance D_{max} between the power plant and the points for which we use the GHI estimation in the model. Theoretically, the greater this distance, the

longer the time horizons for which the estimated GHI time series can be useful. However, increasing D_{max} quadratically increases the number of pixels to be considered. To avoid computational and dimensionality issues, we have to set a limit on this distance D_{max} .

To define D_{max} , we first obtained features estimated from pixels within a 150 km radius as an upper-bound. Then, for different time horizons h , we computed the MI between the GHI time series derived from each pixel lagged h time and the PV production over one year. We could then visualize the location of the most informative time series for each horizon, as shown in Fig. 2. From visualizing the data, we found that for time horizons below 90 min, the global level of the information of the estimated GHI time series is significant compared to other horizons. There is also a significant difference between the most and the least informative time series, the most informative ones being located within a 50 km radius from the plant. On the contrary, for greater horizons, the information is scant. Thus, it does not seem necessary to have a large D_{max} for longer time horizons, or to use satellite data at all. In the end we kept a 50 km value for D_{max} .

The second parameter is the number N_{pix} of pixels we select to derive the features, within the area defined by D_{max} . The selection of a specific pixel from the satellite image results in a GHI time-series from consecutive images. We propose to keep the most informative N_{pix} time-series according to their MI with PV production.

We defined the number of features selected within a D_{max} radius by fitting LASSO (Least Absolute Shrinkage and Selection Operator) models with a 10-fold cross validation, using all the estimated GHI time series as features to predict the production for horizons ranging from +30 minutes to +36 h with a half-hourly time step, that is, for 72 different horizons. For a given horizon h , we obtained the LASSO models with:

$$\hat{E}_{PV,t+h} = \hat{\beta}_0^h + \sum_i \hat{\beta}_i^h X_{t,i} \quad (8)$$

$$\hat{\beta}^h = \underset{\beta}{\operatorname{argmin}} \left((E_{PV} - \hat{E}_{PV})^2 + \sum |\beta_i^h| \right) \quad (9)$$

where $X_{t,i}$ is the estimated GHI for the i -th pixel for time t , and β^h are the parameters of the model.

LASSO models have a built-in tendency to produce a sparse feature selection, and to randomly drop features when they are strongly correlated. As our estimated GHI time series are indeed correlated, we took the feature selection performed by the LASSO models as a measure of the redundancy in the time series, instead of actually selecting the most relevant features. We then obtained N_{pix} by averaging the number of features kept by the LASSO models over all horizons:

$$N_{pix} = \frac{1}{72} \sum_{i=1}^{72} \frac{\#j : \hat{\beta}_j^i = 0}{\#\hat{\beta}^i} \quad (10)$$

In the end, we obtain $N_{pix} = 37$ features, which represent around 12% of the total number of pixels considered. This value is retained for the forecasts in Section IV. Fig. 3 shows the average selection probability for each feature within the area defined by D_{max} , for all forecasts started at 12 p.m. For

the 30-minute horizon, the selected pixels are concentrated in an area east of the plant. This suggests that the weather conditions propagate from east to west, which could be explained by the Sun's path, but also by local weather conditions (e.g. a systematic east wind). For the 90-minute forecast, a slight concentration persists to the east of the plant. After this horizon, no recognizable pattern can be found.

III. BENCHMARK MODELS

To compare the AnEn model with the literature, several benchmark models were also implemented depending on the forecast horizon and resolution.

A. Persistence model

The persistence model is often used as a simple benchmark for PV power forecasting models. In this paper, we use two variants of the persistence, depending on the forecast horizon. We denote them as "Persistence 1" and "Persistence 2" in the remainder of the paper.

The Persistence 1 model simply states that the power forecast \hat{y} for all future times is the power y observed at the time of the forecast, that is:

$$\hat{y}_{t+h} = y_t, \forall h \in \mathbb{N} \quad (11)$$

The Persistence 2 model states that the power forecast for a given time is the power observed the day before at the same time:

$$\hat{y}_t = y_{t-24h} \quad (12)$$

The Persistence 1 model is a benchmark more suited to short-term forecasts, while the Persistence 2 model is a benchmark for day-ahead forecasts.

B. ARIMA model for short-term forecast

For the 5-minute resolution forecasts, we used an Auto Regressive Integrated Moving Average (ARIMA) model [9], which is best suited to short-term forecasts, using only the production data as input.

For an ARIMA model fitted with order (p, d, q) , where p is the number of auto-regressive terms, d the order of differentiation, and q the number of moving average components, a deterministic forecast of the solar output y is readily computed as follows:

$$\hat{y}_t^d = a_0 + \sum_{i=1}^p a_i y_{t-i}^d + \sum_{j=1}^q b_j (y_{t-j}^d - \hat{y}_{t-j}^d) \quad (13)$$

where y^d is the time series y differentiated d times:

$$y_t^d = y_t^{d-1} - y_{t-1}^{d-1} \text{ with } y_t^0 = y_t \quad (14)$$

The vectors of parameters a and b are estimated by maximizing the conditional sum of squares on the most recent data for each forecast, following a sliding window scheme as for the AnEn.

C. QRF model

The AnEn approach was compared with two state-of-the-art models. The first is the QRF model, because it is widely used and featured several times in the leaderboard of the GEFCOM 2014 [16].

This is a modification of the random forest algorithm [34] that can provide quantile forecasts, and was first proposed by reference [35]. In the original random forest model, a large number of regression trees are grown over the training set, and the conditional mean of the distribution is obtained by a weighted average of the output of the trees. More specifically, each tree is grown on a random sample with replacement ("bagged version") of the training set, and each split of the trees is done on a random subset of the predictor variables. This prevents the trees from being correlated, and finally avoids overfitting on the training set. For quantile regression, a random forest is grown over the training set, but instead of the conditional mean, the full distribution is estimated from the observations in the output of the trees.

Here, it is trained using the same variables as the AnEn except for the satellite data that caused computational time of the QRF to be too long, along with their one-time-step lagged values. Note that what we call the QRF model is actually a collection of 72 models, each trained to forecast a specific horizon, because the models have to treat the relative importance of the last measure differently regarding the horizon. This process is automated in the AnEn model, so that there is a single model for all of the horizons. This argument supports the fact that the AnEn model "unifies" time frames, as a single model gives consistent forecasts from +5-minute to +36-hour horizons.

D. Bayesian ARD model

The second model we used is a Bayesian regression with an Automatic Relevance Determination (ARD) prior [36]. This prior is known to introduce sparsity into the feature selection.

The Bayesian ARD approach models the output y as a normal distribution, with the mean being a linear combination of the inputs X , and the precision being a parameter α :

$$y \sim \mathcal{N}(Xw, \alpha^{-1}) \quad (15)$$

The ARD prior on the weights, which introduces sparsity to the approach, models them as centered standard deviations with precisions λ :

$$w_i \sim \mathcal{N}(0, \lambda_i^{-1}) \quad (16)$$

Then, the parameters α and λ are obtained by maximizing the likelihood of the data with respect to them. Once they are obtained, the density of the output is perfectly defined, and can be used to predict new inputs.

Since this model also provides an automatic derivation of the relevance of each feature similar to the AnEn, and naturally provides a probabilistic output, it is a good comparison model for the AnEn.

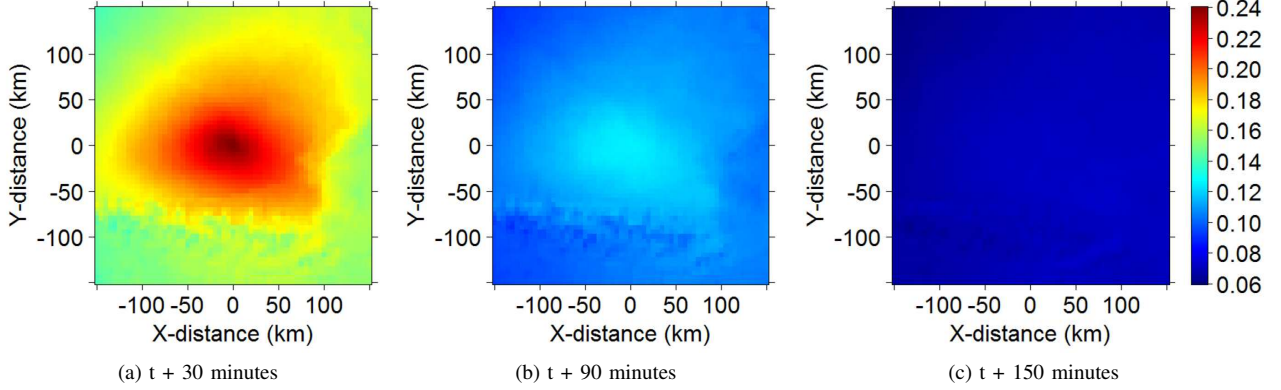


Fig. 2. Mutual information between each estimated GHI time series and the lagged production for different time lags

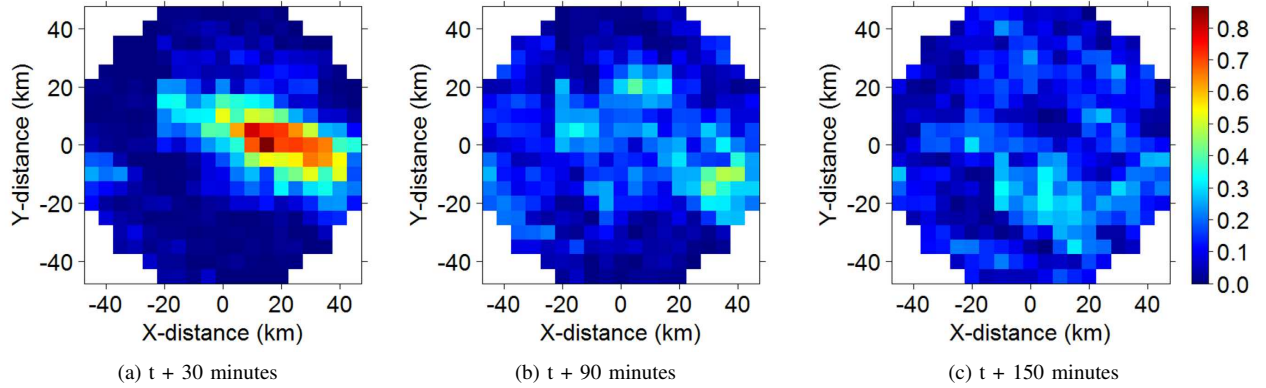


Fig. 3. Probability for each pixel to be selected for different forecast horizons

TABLE I
COMPUTATION TIME REQUIRED FOR PROVIDING THE FORECASTS FOR ONE GIVEN HORIZON, IN SECONDS

	30-minute resolution		5-minute resolution	
	Training	Forecasting	Training	Forecasting
AnEn	-	1.87	-	8.77
Persistence 1	-	5e-3	-	6e-3
Persistence 2	-	5e-3	-	6e-3
ARIMA	9.2e-2	2.5e-7	10e-2	3.7e-3
QRF	4.26	1.3e-2	68.0	4e-2
ARD	10.75	10e-3	154	1e-3

E. Computation Time

The computing times required by the different models, including the AnEn, are reported on Table I. These times include the preprocessing of the data. It is possible to derive a worst-case complexity for some of the algorithms tested.

For the QRF model, during the training we must build N_t decision trees by performing consecutive splits of the features depending on the information gain. Each decision tree has a maximum depth of n , the number of training vectors (after n splits, there is only one training vector remaining in each leaf of the tree and thus no further split can be performed). To perform a split, the information gain has to be evaluated when splitting for each of the p features, and each of the p

evaluations is $O(n)$. In the end, the training of the QRF is thus $O(n^2 p N_t)$. During the testing, we must drop the input along all trees. At most, there is a split for each of the features, and thus evaluating the model is $O(p N_t)$

For the AnEn model, there is no training. The critical step is to compute and select the k closest analogs. Computing the distance for one training vector costs at most p operations for Euclidian distances. For the whole training set, it is $O(np)$. Then, we must select the k closest analogs by looping through the training set k times. In the end, performing the forecast is $O(np + nk)$.

For the ARD model, it is more difficult to assess the complexity. However, computation times from Table III suggest that it is $O(n^2)$ since computation time increased roughly with the same ratio as the QRF.

Thus, it seems natural that the AnEn model is faster, since it does not require training, and its forecasting complexity is lower than the other models training. Since we retrain the models at each timestep to have a fair comparison with the AnEn, it is easily the fastest of the benchmark models (except for persistence and ARIMA models that are extremely fast for short-term forecasting). This is especially visible when we go from 30-minutes to 5-minutes forecast, effectively multiplying the training set size by 6. In that case, the AnEn remains the only tractable model. Also, the algorithms are linear with

respect to the number of features p , which explains why it is difficult to use satellite data without preprocessing.

IV. CASE STUDY

A. Data

The proposed AnEn model was used to forecast the power output of three PV plants located in southwest France. The plants are denoted as P1, P2 and P3 respectively and have a nominal power of 11994, 2694 and 6876 kWp. The available measurements cover the period from January 2014 to September 2018.

NWPs are obtained from the European Center for Medium-range Weather Forecasts (ECMWF). The forecasts are made on a $0.1^\circ \times 0.1^\circ$ latitude/longitude grid every 12 hours. The NWP variables used as features are the Surface Solar Radiation Downwards (SSRD), 10-m U- and V-wind speed (10U and 10V), 2-m temperature (2T), Total Cloud Cover (TCC) and Total Precipitations (TP).

In situ measurements come from the power plants' monitoring systems. The measurements taken into account are PV power, temperature, and GTI. The clear-sky profile is computed using the model McClear [37].

Lastly, features obtained from the MSG satellite imagery are computed using the HelioClim-3 database with the HelioSat-2 method [38], [27]. The images are converted into a time series of estimated GHI for each pixel. At the location of the plants, each pixel corresponds to a $5 \text{ km} \times 5 \text{ km}$ surface.

B. Simulation

Taking the French energy market EPEX SPOT as reference for the case study, the forecast horizons required for trading on both the day-ahead and the intra-day markets are from +30 minutes to +36 hours. Since the measurements were available from January 2014 to September 2018, all of the data necessary to perform the simulation were obtained for the same period. All of the data were then converted to 30-minute time series to obtain the same time step as the market time unit. The clear-sky profile and in situ measurements have a native 5-minute resolution and the satellite images have a 15-minute resolution. These variables were summed over 30-minute intervals to obtain the 30-minute time series. The NWP have a native 1-hour resolution. All NWP fields were linearly interpolated to obtain the 30-minute time series.

However, numerous applications, such as real-time control of a combined PV and storage power plant, also require short-term forecasts with horizons lower than 30 minutes. Using the exact same model but feeding it with the native 5-minute data, and with 5-minute interpolated NWPs and satellite data, we could provide forecasts with a 5-minute resolution.

The period from May 2016 to April 2017 was used to estimate the structural parameters of the AnEn, QRF and ARD models through trial and error. The ARIMA order was also obtained by minimizing the Akaike Information Criterion (AIC) on the testing set. Then, to assess the performance of the models, PV power was forecast from May 2017 to April 2018 with the AnEn and QRF models, from 30 minutes to 36 hours ahead with a 30-minute resolution. The forecasts were updated

every 30 minutes following a sliding window scheme: for each new forecast, the set \mathbf{X} was updated using the latest in situ measurements, NWP, and satellite-derived GHI estimations. In addition, the set \mathbf{H} was updated with the most recent data available at the time to identify the analogs. We also performed the same evaluation for the ARIMA and AnEn models with the 5-minute resolution forecasts, forecasting from +5-minute to +60-minute horizons, and updating the forecasts every 5 minutes.

V. EVALUATION RESULTS

Probabilistic forecasts are more complex to evaluate than deterministic ones. Numerous properties are required for predictive densities, while identifying some aspects of the forecasts may fail when using only proper scoring rules. The main required properties are reliability and sharpness. The results are presented for plant P1 in the following parts, as the behavior of the models was similar for the different plants. Deterministic criteria are also presented for comparison purposes with results from standard deterministic models in the state of the art.

The averaged results are summarized in Table II. All the evaluation criteria from Table II are described in detail in the following parts. Except for the reliability score (R), which is naturally a percentage, all other scores are given in %, relative to the installed power of the plant.

A. Reliability

The standard way to evaluate a forecasting system is first to assess the reliability of the model. The forecasts are reliable if the level of each quantile matches the frequency of the observations. For a perfectly reliable model, the empirical quantile level should be the same as the nominal one, and thus the reliability diagram should be a diagonal line. Fig. 4 shows the reliability diagram of the two models. Consistency bars are also added following [39] to indicate a range within which even a perfectly reliable model could be situated due to the finite size of the testing set. The AnEn and QRF models seem reliable, since they both fall within the acceptable range. However, the ARD model seems to lack reliability. We explain this by the parametric representation of the uncertainty of PV production. It is difficult to make any assumption on the shape of the distribution of PV power, because it might be skewed and varies over time, thus making the Gaussian assumption from the ARD model highly detrimental to the reliability. In Table II, Reliability (R) reports the mean absolute reliability deviations from the diagonal over several forecast horizons.

B. Sharpness

Sharpness measures the spread of the forecast distribution. As an example, a Dirac distribution would have a perfect sharpness, while a uniform distribution would have a very low sharpness. The goal of any probabilistic forecasting system is to be as sharp as possible, while maintaining its reliability. In this paper, sharpness is assessed using the Prediction Interval Normalized Averaged Width (PINAW) metric. Noting $I_{i,\alpha}$ the width of the PI with coverage rate α , it writes as follows:

TABLE II
EVALUATION RESULTS FOR 30-MINUTE RESOLUTION FORECASTS

Plant	AnEn				QRF				ARD				Persistence 1	Persistence 2
	R	S	CRPS	RMSE	R	S	CRPS	RMSE	R	S	CRPS	RMSE	RMSE	RMSE
	%	% of Pn			%	% of Pn			%	% of Pn			% of Pn	% of Pn
Forecast horizon 0-2 h														
P1	1.43	12.4	4.18	7.13	0.90	13.4	4.39	7.43	3.76	13.75	4.79	7.26	20.9	16.5
P2	1.97	12.8	4.20	7.22	2.02	14.1	4.41	7.37	3.22	14.0	4.93	7.25	23.4	16.6
P3	1.90	12.3	4.10	6.92	1.74	13.3	4.34	7.25	3.71	13.8	4.75	6.99	20.7	15.9
Forecast horizon 2-36 h														
P1	2.60	14.4	5.17	8.67	2.18	14.3	5.00	8.53	2.66	17.0	5.97	8.78	20.9	36.5
P2	1.71	16.1	5.47	9.07	2.02	15.1	5.29	8.95	2.54	18.3	6.57	9.34	23.4	37.6
P3	1.00	14.8	5.24	8.68	1.80	14.2	5.12	8.61	3.01	17.6	6.17	8.83	20.7	35.6

TABLE III
RMSE RESULTS FOR PLANT P3 AND 5-MINUTE RESOLUTION FORECASTS

	AnEn	ARIMA	Persistence 2
	% of Pn	% of Pn	% of Pn
Forecast horizon 0-30 min	10.1	10.2	11.6
Forecast horizon 30 min-4 h	12.8	16.8	25.0

$$PINAW(\alpha, h) = \frac{\sum_{i=1}^N I_{i,\alpha}}{P_n} \quad (17)$$

Fig. 5 shows the PINAW for different forecast horizons and nominal coverage rates α . Similar to reliability, the PINAW related to the representation of uncertainty is worse for the ARD model. The PINAW are very similar for the AnEn and QRF models. In Table II, Sharpness (S) reports the mean PINAW over several horizons and all nominal coverage values.

Regarding the parameter k from equation (1), we found that when using $k \lesssim 1$, the forecast lost considerable sharpness, moving increasingly closer to a climatological forecast. Using $k = 0$ also led to significant errors for short-term forecasts where the model relies heavily on the last observation. Since the generation is the same in the morning or in the evening, past observations in the evening were deemed similar to upcoming situations in the morning and vice-versa. With $k = 1$, the pattern of increasing (for the morning) or decreasing (for the evening) generation is included, and the problematic behavior is avoided, without losing too much sharpness.

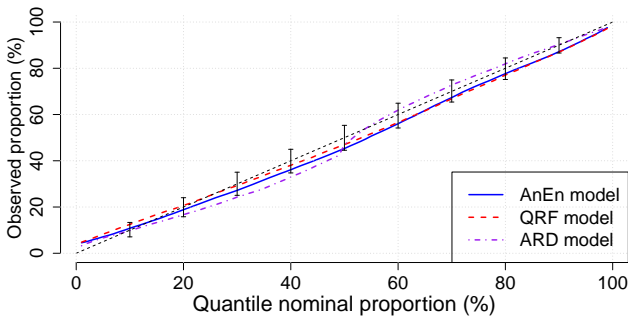


Fig. 4. Reliability diagram of the two models including consistency bars

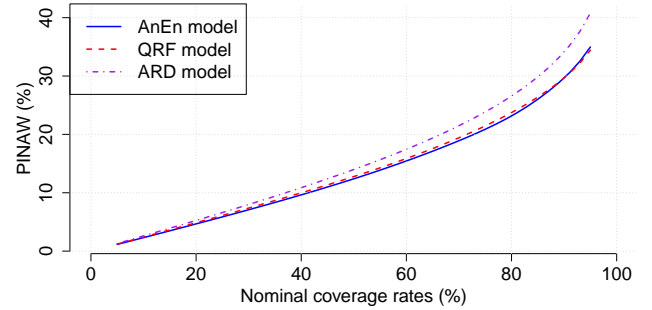


Fig. 5. PINAW for different forecast horizons

C. CRPS score

At this point, it is difficult to tell which model performs better, since they show very similar results for both reliability and sharpness. The overall performance of the models is evaluated using the CRPS, which is a score system dedicated to probabilistic forecasts. It takes into account both reliability and sharpness. The CRPS is defined for a given distribution f and its correspondent cumulative distribution function F , along with an observation y , by:

$$CRPS(F, y) = \int_{-\infty}^{+\infty} (F(u) - \mathbb{1}(y \leq u))^2 du \quad (18)$$

Fig. 6 presents the CRPS of the two models depending on the horizon, normalized by the nominal power of the plant. The difference between the models is clearer when the score is plotted conditionally to the forecast horizon. As expected from the reliability and sharpness results, the ARD model performs worse than the two other models for all horizons. The QRF model outperforms for forecast horizons longer than 3 hours. For shorter horizons, the AnEn model performs much better. However, as can be seen from Table II, the overall CRPS differences between the AnEn and QRF models are very low, and they both show state-of-the-art performance.

As the CRPS difference is very low between the QRF and AnEn models, we implemented a Diebold-Mariano (DM) test [40] using the CRPS as the measure of performance, to see if the difference in the forecasts between the models was statistically significant.

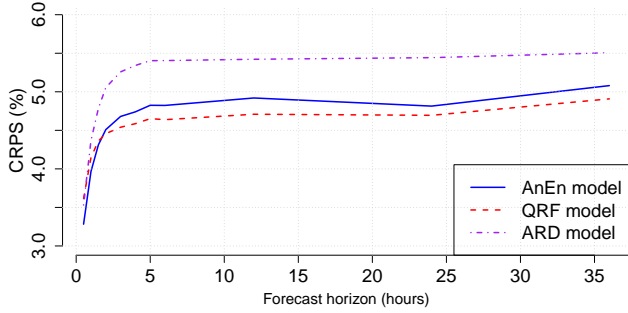


Fig. 6. CRPS performance of the AnEn and the QRF models

Given two sets of forecast CDFs and their matching observations $\{\hat{F}_{1,i}, y_i\}$, and $\{\hat{F}_{2,i}, y_i\}$, we note the loss differential d :

$$d_i = CRPS(\hat{F}_{2,i}, y_i) - CRPS(\hat{F}_{1,i}, y_i) \quad (19)$$

the DM tests the null hypothesis H_0 versus the adverse hypothesis H_1 :

$$H_0 : \mathbf{E}(d) = 0 \quad (20)$$

$$H_1 : \mathbf{E}(d) \neq 0 \quad (21)$$

To do so, it can be shown that under the null hypothesis the DM statistic follows the standard normal distribution:

$$\sqrt{(1/T)} \frac{\sum_{i=1}^T d_i}{\sqrt{2\pi f_d(0)}} \sim \mathcal{N}(0, 1) \quad (22)$$

Where $f_d(0)$ is the spectral density of d at frequency 0 which can be approximated from the data, and T is the number of forecast/verification pairs available. On Fig. 7 we plot the DM statistic between the AnEn and QRF models for different forecast horizons. We choose a standard 5 % significance level, and thus values of the DM statistic lower than the 2.5 % quantile or higher than the 97.5 % quantile of the normal distribution (respectively -1.96 and +1.96) allow us to reject the null hypothesis.

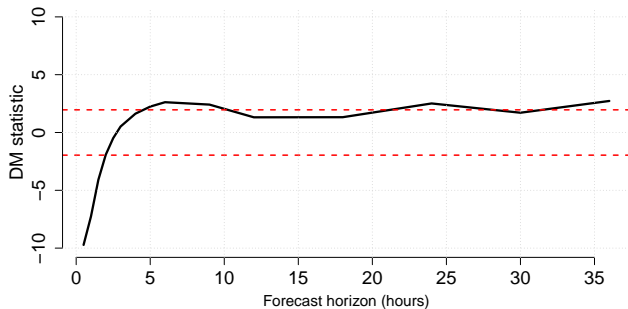


Fig. 7. DM statistic between the QRF and AnEn model

From Fig. 7 we can see that the difference in the AnEn and QRF forecasts is not significant for horizons between 2 and

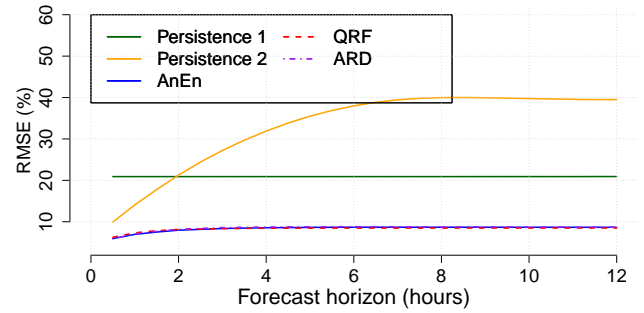


Fig. 8. RMSE of the models depending on the horizon

5 hours. For horizons lower than two hours, the difference is clearly significant, while for longer horizons, it is less clear. The QRF seems slightly better, but the null hypothesis is on the verge of being rejected for many horizons.

This supports the facts that the AnEn performs better for short forecast horizons, and that the difference for long forecast horizons is not significant.

D. RMSE

The Root Mean Square Error (RMSE) is also computed, taking the densities medians as a deterministic forecasts. This allows us to compare the model with standard deterministic ones. We compared the AnEn and QRF with two variants of the classic persistence models. The first variant, noted Persistence 1, gives the power measurement of the day before at the same time of the day as the forecast. The second, noted Persistence 2, gives the power observed at the starting time of the model as the prediction for all horizons.

The RMSE of the models depending on the forecast horizon is presented on Fig. 8. All models consistently outperform Persistence 1 and 2 for all horizons. For the ARD model, although the uncertainty representation is not as good as the AnEn and QRF, as can be seen from the reliability, PINAW and CRPS scores, the deterministic forecasts are good, with RMSE scores similar to the AnEn and QRF.

E. Intra-hourly forecasts

Figure 9 shows the average RMSE conditioned to the forecast horizon for the AnEn, ARIMA and Persistence 2 models for the 5-minute resolution forecasts. The AnEn model is consistently more accurate than the two other models for intra-hourly forecasting. Besides, even though it is not shown with the RMSE criterion, the AnEn provides uncertainty information since it gives a probabilistic estimate of the production.

F. Contribution of Each Source of Data

Figure 10 shows how CRPS performance increases when adding incrementally different sources of data. The addition of the last measure significantly increases performance for time steps up to 5 h. This is self-explanatory, as the last measure

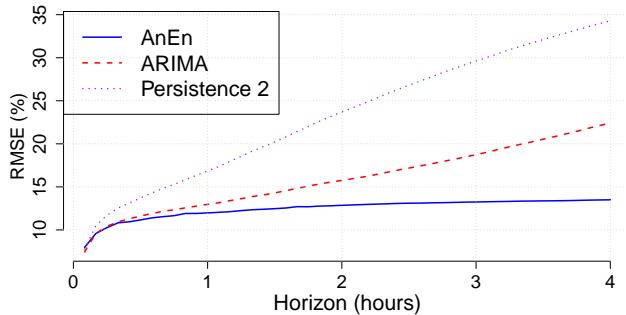


Fig. 9. Comparison of AnEn, ARIMA and Persistence 2 models for intra-hourly forecasts

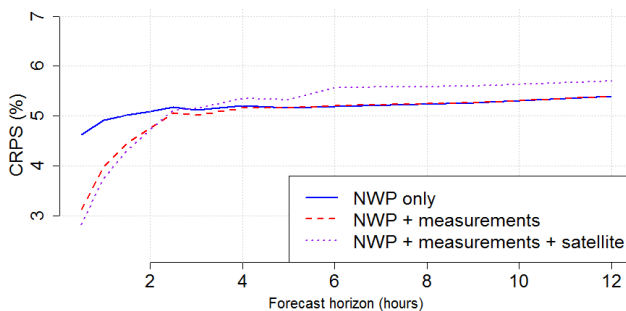


Fig. 10. Comparison of the performance of the AnEn model depending on the inputs

is very informative about the current meteorological situation, but carries little predictive information.

The addition of satellite data increases further the forecasts accuracy for the first forecast horizon. There is also a slight improvement up to 3 hours ahead. It is surprising that the satellite data added value mostly in the same horizon range as the ground measurements (that is, in the first timesteps). We expected that it would improve forecasts in the typical range of satellite data, that is in the 2-6 hours horizon range. This can be explained by the use of satellite data as "snapshots" to condition the future forecasts, instead of explanatory variables that we could derive using an intermediate forecasting system from the satellite pictures e.g. a Cloud Motion Vector model. In that case, satellite data would be similar to NWP's rather than measurements. However, satellite data had little effect for day-ahead forecasts. This is probably caused by the residual weights for day-ahead that can be seen in Fig. 1.

In the end, the model is able to process both temporal and spatial information from very different sources of data. It could be extended by other features, that are known to improve solar power forecasting, such as measurements from neighboring PV plants or weather stations. However, this would require further work when assigning global weights to each source of data. In our case, using the maximal feature weight as the global weight of the whole source was efficient as the information between each source of data was not redundant.

When increasing the sources of data, the chances are higher that two sources of data will carry correlated information. This should be considered for the global weight assignment.

VI. CONCLUSIONS

In this paper, a probabilistic PV power forecasting model is proposed that is well suited for use by a power plant operator trading in day-ahead and intra-day electricity markets. The model was found to (1) have state-of-the-art performance in the +5-minute to +36-hour forecast range; (2) be able to dynamically adjust to the most recent data, so that any change in the PV plant (surrounding environment, partial outage, soiling, etc.) can be automatically dealt with; (3) be able to start at any time of the day for any forecast horizon without multiplying the number of models.

The good performance of the model for both intra-day and day-ahead forecasts was achieved by including in situ measurements and data from satellite imagery along with NWP's. Specifically, efforts were made to deal with the high dimensionality of satellite data. Further work could include other sources of data in the model. The proposed model was proven to be an efficient alternative, both in terms of computational cost and accuracy, to the state-of-the-art approach where different forecasting models are employed as a function of the available data types and the application time frame.

REFERENCES

- [1] R. Bessa, C. Möhrlen, V. Fundel, M. Siefert, J. Browell, S. Haglund El Gaidi, B.-M. Hodge, U. Cali, and G. Kariniotakis, "Towards Improved Understanding of the Applicability of Uncertainty Forecasts in the Electric Power Industry," *Energies*, vol. 10, no. 9, p. 1402, 2017. [Online]. Available: <http://www.mdpi.com/1996-1073/10/9/1402>
- [2] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Conversion and Management*, vol. 156, no. May 2017, pp. 459–497, 2018.
- [3] P. Lauret, M. David, and H. Pedro, "Probabilistic Solar Forecasting Using Quantile Regression Models," *Energies*, vol. 10, no. 10, p. 1591, 2017.
- [4] L. M. Aguiar *et al.*, "Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intraday solar forecasting," *Renewable Energy*, vol. 97, pp. 599–610, 2016.
- [5] A. Ayet and P. Tandeo, "Nowcasting solar irradiance using an analog method and geostationary satellite images," *Solar Energy*, vol. 164, pp. 301–315, 2018.
- [6] X. G. Agoua, R. Girard, and G. Kariniotakis, "Short-Term Spatio-Temporal Forecasting of Photovoltaic Power Production," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 538–546, 2018.
- [7] X. G. Agoua, R. Girard, and G. Kariniotakis, "Probabilistic models for spatio-temporal photovoltaic power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 2, pp. 780–789, April 2019.
- [8] H. T. C. Pedro, C. F. M. Coimbra, M. David, and P. Lauret, "Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts," *Renewable Energy*, vol. 123, pp. 191–203, 2018.
- [9] H. T. C. Pedro and C. F. M. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Solar Energy*, vol. 86, no. 7, pp. 2017–2028, 2017.
- [10] J. G. da Silva Fonseca *et al.*, "On the use of maximum likelihood and input data similarity to obtain prediction intervals for forecasts of photovoltaic power generation," *Journal of Electrical Engineering and Technology*, vol. 10, pp. 1342–1348, 2015.
- [11] C. Wan, J. Lin, Y. Song, Z. Xu, and G. Yang, "Probabilistic Forecasting of Photovoltaic Generation: An Efficient Statistical Approach," *IEEE Transactions on Power Systems*, vol. 32, no. 3, pp. 2471–2472, 2017.
- [12] B. Jing, Z. Qian, Y. Pei, and J. Wang, "Ultra short-term PV power forecasting based on ELM segmentation model," *The Journal of Engineering*, vol. 2017, no. 13, pp. 2564–2568, 2017.

- [13] F. Golestaneh, P. Pinson, and H. B. Gooi, "Very Short-Term Nonparametric Probabilistic Forecasting of Renewable Energy Generation With Application to Solar Energy," *IEEE Transactions on Power Systems*, vol. 31, no. 5, 2016.
- [14] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, and S. Wang, "Short-Term Solar Power Forecasting Based on Weighted Gaussian Process Regression," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 1, pp. 1–1, 2017.
- [15] M. J. Sanjari and H. B. Gooi, "Probabilistic Forecast of PV Power Generation Based on Higher Order Markov Chain," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2942–2952, 2017.
- [16] T. Hong, P. Pinson, S. Fan, H. Zareipour, A. Troccoli, and R. J. Hyndman, "Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond," *International Journal of Forecasting*, vol. 32, pp. 896–913, 2016.
- [17] G. I. Nagy, G. Barta, S. Kazi, G. Borbély, and G. Simon, "GEF-Com2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach," *International Journal of Forecasting*, vol. 32, no. 3, pp. 1087–1093, 2016.
- [18] J. Huang and M. Perry, "A semi-empirical approach using gradient boosting and k-nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting," *International Journal of Forecasting*, vol. 32, pp. 1081–1086, 2016.
- [19] L. Gigoni, A. Betti, E. Crisostomi, A. Franco, M. Tucci, F. Bizzarri, and D. Mucci, "Day-Ahead Hourly Forecasting of Power Generation from Photovoltaic Plants," *IEEE Transactions on Sustainable Energy*, vol. 9, no. 2, pp. 831–842, 2018.
- [20] L. Delle Monache, F. A. Eckel, D. L. Rife, B. Nagarajan, and K. Searight, "Probabilistic Weather Prediction with an Analog Ensemble," *Monthly Weather Review*, vol. 141, no. 10, pp. 3498–3516, 2013.
- [21] C. Monteiro, T. Santos, L. A. Fernandez-Jimenez, I. J. Ramirez-Rosado, and M. S. Terreros-Olarte, "Short-term power forecasting model for photovoltaic plants based on historical similarity," *Energies*, vol. 6, pp. 2624–2643, 2013.
- [22] S. Alessandrini, L. Delle Monache, S. Sperati, and G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast," *Applied Energy*, vol. 157, pp. 95–110, 2015.
- [23] B. O. Akyurek, A. S. Akyurek, J. Kleissl, and T. i. Rosing, "TESLA: Taylor expanded solar analog forecasting," in *2014 IEEE International Conference on Smart Grid Communications, SmartGridComm 2014*, Venice, Italy, 2015, pp. 127–132.
- [24] C. Junk, L. Delle Monache, S. Alessandrini, G. Cervone, and L. von Bremen, "Predictor-weighting strategies for probabilistic wind power forecasting with an analog ensemble," *Meteorologische Zeitschrift*, vol. 24, no. 4, pp. 361–379, 2015.
- [25] A. Gensler, B. Sick, and V. Pankraz, "An analog ensemble-based similarity search technique for solar power forecasting," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC16)*, Budapest, Hungary, 2016, pp. 2850–2857.
- [26] G. Cervone, L. Clemente-Harding, S. Alessandrini, and L. Delle Monache, "Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble," *Renewable Energy*, vol. 108, pp. 274–286, 2017.
- [27] P. Blanc, B. Gschwind, M. Lefèvre, and L. Wald, "The HelioClim project: Surface solar irradiance data for climate applications," *Remote Sensing*, vol. 3, no. 2, pp. 343–361, 2011.
- [28] B. Silverman, "Density estimation for statistics and data analysis," *Chapman and Hall*, vol. 37, no. 1, pp. 1–22, 1986.
- [29] S. Sheather and M. Jones, "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 53, no. 3, pp. 683–690, 1991.
- [30] S. Alessandrini, L. Delle Monache, S. Sperati, and J. N. Nissen, "A novel application of an analog ensemble for short-term wind power forecasting," *Renewable Energy*, vol. 76, pp. 768–781, 2015.
- [31] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [32] J. Dougherty, R. Kohavi, and M. Sahami, "Supervised and Unsupervised Discretization of Continuous Features," in *Machine Learning: Proceedings of the Twelfth International Conference*, San Francisco, USA, 1995.
- [33] J. L. Bosch and J. Kleissl, "Cloud motion vectors from a network of ground sensors in a solar power plant," *Solar Energy*, vol. 95, pp. 13–20, 2013.
- [34] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [35] N. Meinshausen, "Quantile Regression Forests," *Journal of Machine Learning Research*, vol. 7, pp. 983–999, 2006.
- [36] M. E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–245, 2001.
- [37] Lefèvre *et al.*, "McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions," *Atmospheric Measurement Techniques*, vol. 6, no. 9, pp. 2403–2418, 2013.
- [38] L. F. Zarzalejo, J. Polo, L. Martín, L. Ramírez, and B. Espinar, "A new statistical approach for deriving global solar radiation from satellite images," *Solar Energy*, vol. 83, no. 4, pp. 480–484, 2009.
- [39] J. Bröcker and L. A. Smith, "Increasing the Reliability of Reliability Diagrams," *Weather and Forecasting*, vol. 22, no. 3, pp. 651–661, 2007.
- [40] F. Diebold and R. Mariano, "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, vol. 13, pp. 253–263, 1995.

Thomas Carriere graduated from Ecole Supélec (2016) with a specialization in Energetics, and received his MSc. degree in Energy Processes from Paris-Saclay University (2016). He is pursuing a Ph.D. at MINES ParisTech, PSL - Research University, PERSEE - Centre for Processes, Renewable Energies and Energy Systems. His research interests are forecasting from statistical modelling, storage systems, and renewable energies integration to electricity markets.

Christophe Vernay graduated from Ecole Supélec (1997), a general engineering school in France. He started his career working in the space industry, for Alcatel Space Industries (1998–2001), as a research engineer in Attitude and Orbit Control Systems (AOCS). He then worked in the telecoms industry for Nortel Networks then Alcatel-Lucent (2001–2010) in several positions ranging from integration to engineering of 3G and 4G systems. His interest in renewable energies and his convictions led him to study photovoltaics and wind power in 2010 at the Conservatoire national des arts et métiers (CNAM), Paris. In 2011, he received a specialized Master's degree in renewable energy from Ecole Nationale Supérieure d'Arts et Métiers (ENSAM) and then pursued a professional thesis at SOLAÏS, a French consulting company dedicated to photovoltaics, in collaboration with MINES ParisTech/Armines, on characterizing measurement campaigns of global irradiation. He has since occupied the position of Technical Director at SOLAÏS, in charge of R&D, engineering, technical due diligence and commissioning.

Sebastien Pitaval graduated from Ecole Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, d'Hydraulique et de Télécommunications (2000, Toulouse, France), with a specialization in fluid mechanics and energy. He started his career working in the space industry, for Alcatel Space Industries then Thalès Alenia Space (2000–2008), occupying successive management positions for Herschel and Planck satellites: propelling systems, Attitude and Orbit Control Systems (AOCS) then systems engineering. His interest in renewable energies, conviction and entrepreneurial spirit led him to co-create SOLAÏS in 2008. He developed R&D, Engineering and Export activities while setting up Corporate Social Responsibility (CSR) within the company. He co-created THIRD STEP in 2012 for the development of ground-mounted PV plants in France and Asia. He is also a legal expert at several French Courts of Appeal.

Georges Kariniotakis (S'95-M'02-SM'11) was born in Athens, Greece. He received his Eng. and M.Sc. degrees from Greece in 1990 and 1992 respectively, and his PhD from Ecole des Mines de Paris in 1996. He currently works at the MINES ParisTech PERSEE Centre as a senior scientist and head of the Renewable Energies and Smart Grids Group. He has authored more than 240 scientific publications in journals and conferences. He has been involved as participant or coordinator in more than 45 R&D projects in the fields of renewable energies and distributed generation. Among them, he was the coordinator of several major EU projects in the field of renewable power forecasting such as the Anemos, Anemos.plus and SafeWind projects, as well as the Smart4RES H2020 project starting in 2019. His scientific interests include among others timeseries forecasting, decision-making under uncertainty, modeling, management and planning of power systems.