Importance Based Selection Method for Day-ahead Photovoltaic Power Forecast Using Random Forests

Ali Lahouar*, Amal Mejri and Jaleleddine Ben Hadj Slama
Laboratory of Advanced Technology and Intelligent Systems
National Engineering School of Sousse, BP 264, Sousse Erriadh 4023, University of Sousse, Tunisia
Email: ali.lahouar@gmail.com, elmejri.amal@gmail.com, bhslama@yahoo.fr

***Abstract*—With the great recent moves towards green energy exploitation worldwide, the solar photovoltaic (PV) power has gained much attention. Thanks to PV panels' cost drop and recent improvements in energy conversion systems, the PV installations are getting more and more integrated into power plants. Because of high correlation with weather conditions, accurate short-term PV output forecast is highly recommended. An accurate prediction is needed to assess the effective contribution of solar energy in the grid, and to overcome the problems of intermittence. This paper proposes a day-ahead prediction method of PV output, which estimates the power generated by solar panels with and without prior knowledge of solar irradiance. The proposed model is the random forest using bagging algorithm, characterized by built-in cross validation and immunity to irrelevant inputs. A special attention is paid to the choice of most influential weather conditions on future power. The proposed approach is validated through tests on real data from PV sites in Australia.**

***Index Terms*—Photovoltaic power forecast, random forest, importance measure, solar irradiance, feature selection.**

I. Introduction

The PV energy becomes the center of interest in many countries, especially in deserts that receive huge amounts of solar irradiance. It is among best alternatives to conventional thermal and nuclear generation. Being environmentally friendly and having low production costs, it is now getting more and more integrated into power plants. The global installed capacity all over the world reached 227 GW in 2016. China is the world leader with 15.3 GW installed, followed by United States, Europe and India. The PV technology becomes easier to manipulate thanks to the decreasing costs of solar panels and recent advances in power converters and batteries. Nowadays, the smart grid concept includes necessarily renewable energy resources with different levels of penetration. The novel technologies ensure better management and exploitation of these intermittent resources, and provide sophisticated tools to deal with any problems that may occur.

However, the PV energy has to overcome various problems. The intermittence of power output is among most important issues, since it is highly dependent on weather conditions. Forecasting the power that may be provided by the solar panel is then mandatory for dispatch. This forecast is performed for a period of time in the future called forecast horizon, through one or various time steps. It is possible to predict power, energy or solar irradiance. According to the forecast horizon, the forecast ranges from very-short to long term. In general, researchers focus on short-term, precisely hour-ahead and day-ahead forecasts. Indeed, they are extremely important for planning, feasibility and scheduling. Long-term predictions are also useful, especially for maintenance. Some possible definitions of forecast horizons as well as forecast interests for each horizon can be easily found in literature [1]. The accurate forecast becomes more critical with high penetration levels of solar installation in the grid. In fact, the effect of neighboring solar plants on power forecast is quantified [2], as well as the economic benefits of accurate PV prediction [3]. In addition to traditional forecast, recent research advances are moving towards probabilistic forecast, which provides confidence intervals instead of single points. The probabilistic forecast handles the uncertainty more efficiently, and offers a margin within which the dispatcher can act in case of major error.

One of the most important steps to achieve an accurate forecast is the feature selection, which means choosing the appropriate factors that drive the predicted output. In the case of PV forecast, the solar irradiance is obviously the most important feature. However, several other weather conditions have also their impact, such as air temperature, humidity, wind speed, and wind direction. Various approaches are used to determine the appropriate weather factors, generally by regression and correlation analyses. Autocorrelation and cross correlation functions are also tools to select the suitable features across time. In addition to feature selection, some intrinsic characteristics are also responsible for reducing the forecast error, such as the area of the solar farm. In fact, enlarging the panels' surface makes the overall production smoother and minimizes the disturbing phenomena such as cloud shading.

Forecast methods may be direct or indirect. In case of indirect methods, the solar irradiance is predicted, and then converted into power using the characteristics of panels. In direct methods, power is predicted directly. Indirect methods may be physical or statistical. Physical approaches rely on meteorological equations that are solved by calculators in order to estimate future weather states. The numerical weather

978-1-5090-6287-4/17/$31.00 **© 2017 IEEE**

prediction (NWP) is among the well-known physical models. Statistical approaches make use of time series, regression and artificial intelligence.

Several reviews summarize the commonly used methods by researchers [4]. Time series predictors are the simplest models, such as nonlinear autoregressive model with exogenous inputs [5], or autoregressive integrated moving average [6]. The artificial neural network (ANN) is one of the most elaborated approaches for PV forecast. It may be utilized as standalone forecaster [7] or combined with particle swarm optimization [8], Markov model [9], wavelet transform [10], or even enhanced by confidence intervals [11]. The support vector machine (SVM) is likewise a well-developed method. It may be combined with ANN [12] or wavelet decomposition [13]. The random forest (RF) is also used, where it may be combined with confidence intervals [14]. Probabilistic forecast has gained much attention on his part [15], with possible combination with Markov chain [16] or elaboration of stochastic processes [17]. Physical methods are also developed, such as tracking clouds' motion [18], NWP optimization [19], and electric circuit modeling [20]. Hybridizing physical and artificial intelligence models is also possible [21] in addition to comparison between them [22].

This paper proposes a random forest (RF) predictor for day-ahead PV output forecast, hour-by-hour. The RF is chosen since it is immune to irrelevant inputs and has good generalization capacity. The most important aspect is the RF ability to measure the importance of its inputs. The importance measure is then exploited in the feature selection process. In addition, correlation and profile analysis are elaborated in order to select the suitable features. The model has 15 variants for 15 prediction hours in the day, between 5 am and 7 pm. Tests are carried out on real data from the Australian PV sites, with and without prior knowledge of solar irradiance. The remainder of the paper is organized as follows: section II gives necessary mathematical tools, section III depicts the design of forecasters and the feature selection process, section IV analyzes the obtained results, and section V concludes the paper.

II. Mathematical development

The random forest is an ensemble method that aggregates the output of several uncorrelated decision trees. The decision tree will then presented first.

A. Decision tree

A decision tree, or classification and regression tree, is a statistical model introduced in 1984. It depicts the different classes or values that an output may take in terms of a set of input features. A tree in general is an organized set of branches and nodes with no loops. The nodes of a decision tree store a test function that should be applied to incoming data. Terminal nodes are called leaves of the tree. Each leaf stores the final test result. The tree is binary if each internal node has only two outgoing branches, called right child and left child. The decision tree may be used for classification or for regression. In this paper, only the regression variant will be depicted, since forecast is a kind of regression. Let X be the input vector containing m features, Y the output scalar and S_n a training set containing n observations (X_i, Y_i):

$$S_n = \{(X_1, Y_1), ..., (X_n, Y_n)\},\ X \in \mathbb{R}^m,\ Y \in \mathbb{R} \quad (1)$$

During training, an algorithm drives the inputs' split at each node, in a way to optimize the split function parameters so they fit with S_n. The concept is to split recursively the input space X by searching the optimal sub-partitions according to:

$$\{X^j < d\} \cup \{X^j > d\} \quad (2)$$

Where $j \in \{1, ..., m\}$ and $d \in \mathbb{R}$. In order to determine the best split, the couple (j, d) has to minimize a cost function, which is most of the time the variance of children nodes. The variance VAR of a child node p is defined by:

$$VAR(p) = \sum_{i:X_i \in p} (Y_i - \bar{Y}_p)^2 \quad (3)$$

Where $\bar{Y}_p$ is the average of observations Y_i present at node p. Then, children nodes are also split according to the same procedure. The tree development is stopped by a termination criterion. Generally, the tree is stopped when a maximum number of levels is reached, or when a node contains less than a minimum number of observations. At the end of the training process, a prediction function $\hat{h}(X, S_n)$ is built over S_n.

The test process consists in determining an estimation $\hat{Y}$ of the output Y that matches a new input X. A new input means not contained in S_n:

$$\hat{Y} = \hat{h}(X, S_n) \quad (4)$$

Starting from root, each node applies its own split function to X. According to the result of the binary test, data are sent either to right child or to left child. This procedure continues until X reaches a leaf (a terminal node).

B. Random forest

The random forest (RF) is a combination of several weak predictors. The base principle is called bagging (or bootstrap aggregation); where a sample of size n from the training set S_n is selected randomly and fitted to a regression tree. This sample, called bootstrap, is chosen by replacement. Replacement means that the same observations (X_i, Y_i) may appear several times.

A bootstrap sample is built by selecting randomly n observations with replacement from S_n, where each observation has a probability of $1/n$ to be selected. The independent identically distributed random variables Θ_l represent this random selection. The bagging algorithm selects various bootstrap samples $(S_n^{\Theta_1}, \ldots, S_n^{\Theta_q})$, applies the previous regression tree algorithm to these samples in order to get a collection of q prediction trees $(\hat{h}(X, S_n^{\Theta_1}), \ldots, \hat{h}(X, S_n^{\Theta_q}))$, and finally aggregates the outputs of all these predictors.

In addition to bagging, the random forest chooses a predefined number $mtry$ of features among the m available features

for the split at each node. The RF algorithm tries to find the best split according the $mtry$ selected features only. The selection is uniform, thus each feature has a probability of $1/m$ to be chosen. The number $mtry$ is the same for all prediction trees, and it is recommend to be the square root or the one third of m:

$$mtry = \lfloor \sqrt{m} \rfloor \tag{5}$$

Or:

$$mtry = \lceil m/3 \rceil \tag{6}$$

Were $\lfloor x \rfloor$ and $\lceil x \rceil$ stand for floor and ceiling functions of x, respectively. The remainder of the algorithm is similar to that of regression trees: the best couple for split (j, d) is obtained by minimizing a cost function, and the procedure continues until full development of all trees. The aggregation is performed by averaging the outputs of these trees. Then, the estimation $\hat{Y}$ of output matching any new input X is computed by:

$$\hat{Y} = \frac{1}{q} \sum_{l=1}^{q} \hat{h}(X, S_n^{\Theta_l}) \tag{7}$$

The major advantage of bootstrap aggregation is immunity to noise, since it combines the prediction of several uncorrelated trees. In addition, two other characteristics distinguish the RF, which are the out-of-bag error, and the measure of importance. These two special functions are not detailed here in order to lighten the paper. The number of trees q will be denoted $ntree$ in the following sections.

III. Design of the Predictor

A. Profile analysis

The University of Queensland in Australia has several PV sites with a total installed capacity of 5796 kW. The data used in this paper are relative to one of these sites, where power, cumulative energy and weather conditions are recorded every minute.

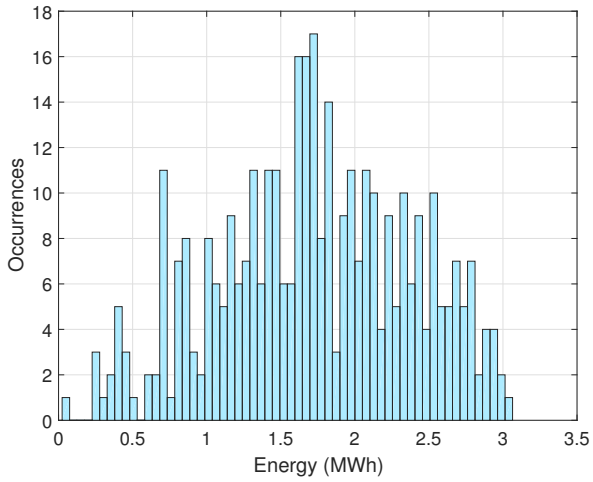

Fig. 1. Distribution of daily cumulative energy over 2013

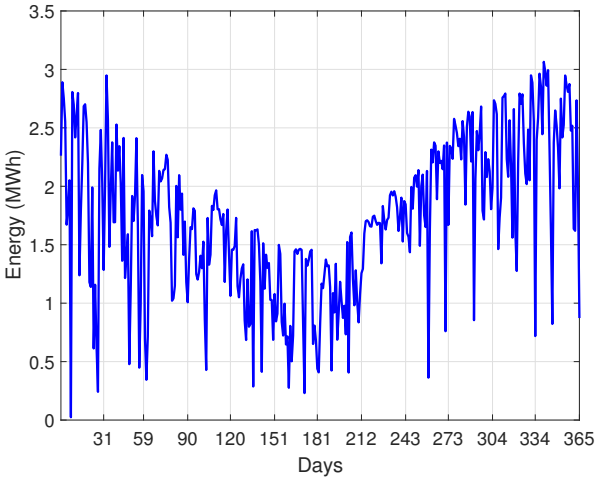

Fig. 2. Evolution of daily cumulative energy throughout the year 2013

Fig. 1 shows the daily distribution of cumulative energy over 2013. The cumulative energy is the overall produced energy by solar panels at the end of the day. As it can be seen, most occurrences are between 0.5 and 3 MWh, while only few days have no production at all. To have such a distribution is beneficial for the predictor, since it makes the learning process more efficient. Fig. 2 tracks the evolution of cumulative energy throughout the year. The seasonal effect is obvious, where winter months (June and July) have the minimum production. Dealing with seasonal effect is relatively easy. It is enough either to add a season index to the predictor's inputs, or to perform an online training.

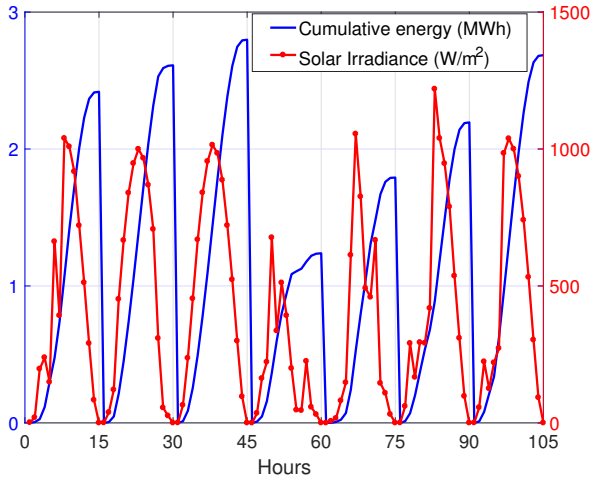

Fig. 3. Hourly evolution of cumulative energy and solar irradiance from 11 to 17 January 2013

Fig. 3 shows the hourly evolution of cumulative energy across 7 days in January. Each day consists of 15 hours only, from 5 am to 7 pm, when the sunlight is available. The energy is correlated with solar irradiance to some extent. However,

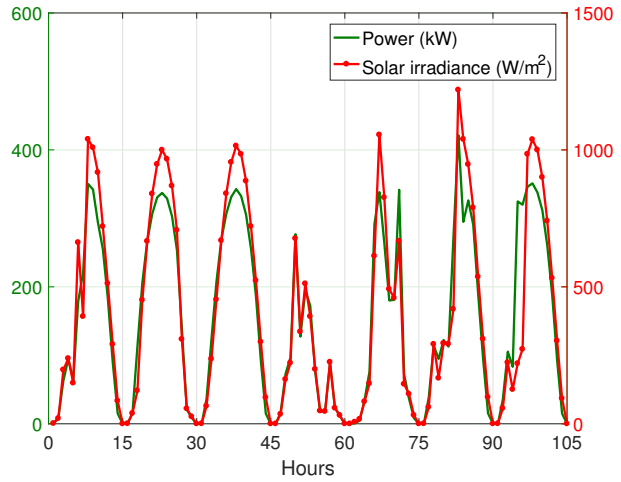

Fig. 4. Hourly evolution of power and solar irradiance from 11 to 17 January 2013

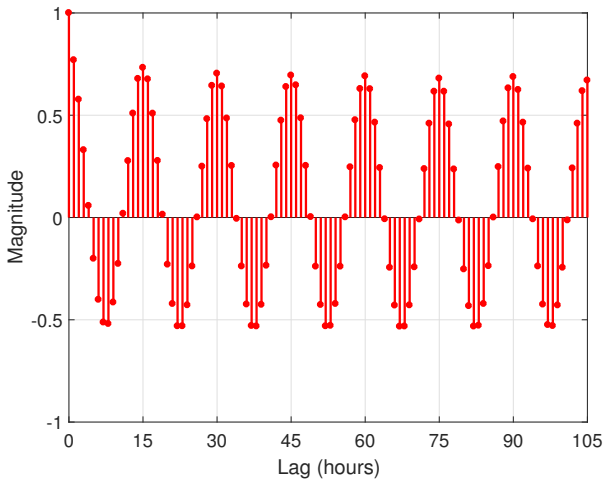

Fig. 5. Autocorrelation plot of hourly power across 7 days in January 2013

correlation with irradiance is much more visible in case of instantaneous power, as shown in Fig. 4. The two curves overlap most of the time. Indeed, the solar irradiance is the most important factor that drives the power. The curves' shape is sometimes smooth, and sometimes rough and spiky. The roughness is a consequence of cloudy sky, and it hardens the forecast process.

B. Feature selection

The feature selection is an important step. It is the art of choosing the most appropriate factors driving the power output. Statistical methods in general make use of past values to predict future values. For this purpose, the autocorrelation function is plotted in Fig. 5. The autocorrelation determines similar past samples to the current power value. Obviously, lags of 15 hours have the maximum magnitude, hence the maximum similarity. There is then a strong daily seasonality, since each day contains 15 hours. It can be concluded that power at hour h of a given day is highly correlated with power at hour h of previous days.

TABLE I
IMPORTANCE MEASURES OF THE SIX CHOSEN INPUTS

Importance at noon ($\times 10^{10}$)	**7 October**	**7 January**	**7 April**	**7 July**
Month	7.6964	1.2006	4.7092	1.9354
Temperature	9.2263	16.4887	26.2148	4.5168
Humidity	20.3370	22.4457	20.6520	9.0860
Solar irradiance	32.3992	48.0994	38.0906	28.7998
Wind speed	3.6926	5.1924	11.0516	4.0792
Wind direction	5.1447	7.1231	11.8031	5.6302

The measure of importance is another way to determine the suitable features. The importance measures of six variables are then calculated. These variables are: month index Mo, temperature T, humidity Hu, solar irradiance Ir, wind speed Ws and wind direction Wd. The measures, which appear in Table I, are taken on four different days at noon, after 90 days of training. The considered output is obviously the produced power P at noon, expressed in watts. The solar irradiance has unsurprisingly the maximum importance measure. However, the second most important variable is humidity, and not temperature as expected. Indeed, some PV sites have their special intrinsic characteristics that may differ from usual sites. Also, the wind direction is a little bit more important than wind speed.

C. Model construction

Because of high daily seasonality, it is judicious to predict each hour of the day apart through a dedicated forecaster. Hence, 15 variants of the RF predictor are called to forecast the 15 hours of the day. Each variant has the six previously selected inputs, and one output which is the predicted power at that hour. In other words, the model is depicted by these equations for each hour h of the day:

$$X(h) = [Mo(h), T(h), Hu(h), Ir(h), Ws(h), Wd(h)]^T \quad (8)$$

$$\hat{Y}(h) = \hat{P}(h) \quad (9)$$

Where $\hat{Y}(h)$ is computed according to equation (7), and h is the hour index ($h = 1, \ldots, 15$). The quantity $\hat{P}(h)$ is the predicted power at hour h. The constructed model assumes that all input variables are known in advance. Building such a model is useful to assess the RF power forecast ability, assuming perfect weather predictions.

The RF machine is set to default parameters, $ntree = 500$ and $mtry$ follows equation (6). Four tests are performed over four weeks in October, January, April and July. The training period for each week is set to 90 days. The learning process is online, which means new training for each new day. The 15 forecasters predict the first day of the testing week. Then the real power values of the first day are appended to the training period and the farthest values from it are removed, so that the

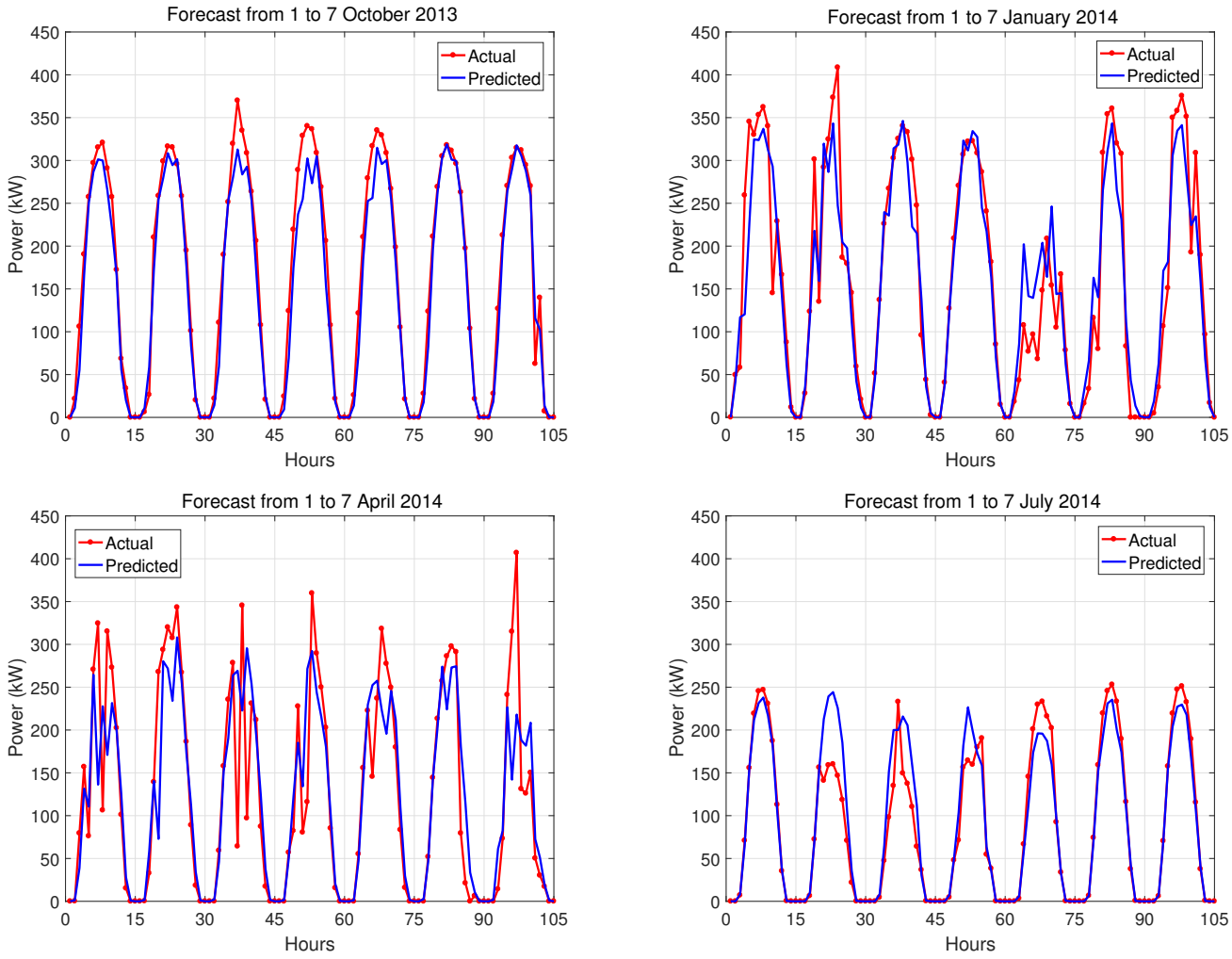

Fig. 6. Predicted and actual power curves throughout four test weeks

second day can be predicted. This procedure continues until the last day of the test period. This looks like a sliding train period, used in order to teach RF predictors the new trends of power pattern. The online learning helps to overcome slow seasonal effects, such as transit from winter to summer.

IV. Case Study

A. Forecast with prior knowledge of solar irradiance

This type of forecast assumes that future weather conditions, namely temperature, humidity, irradiance, wind speed and wind direction are known in advance. They are determined for each hour of the day, either by physical/NWP models or by dedicated statistical models. These future values should be available at predictors' inputs before forecasting. In this paper, exact historical values are utilized. It is true this looks like conversion rather than a proper forecast, but it is useful to assess the efficiency of the predicting machine. The four tests are then applied using the online learning depicted in the previous section. The weeks of test range from the 1st to the 7th day of October 2013, January, April and July 2014, respectively. The predicted power curves are given in Fig. 6. In October, the forecast is almost perfect. Fortunately, there was no major shading. In January, some little errors occurred. The fifth day is a bit special; the produced power was lesser than usual. April has the worse results; the curve is pretty spiky. The curve's roughness is expected in autumn, since it is the season with maximum cloud movements. The results in July are somehow acceptable, even if some difficulties are expected.

The qualitative evaluation through curves is generally not enough. Quantitative assessment through error metrics is also required. The evaluation metrics are criteria that measure the distance between actual and predicted quantities. This paper uses, among well-known error metrics, the mean absolute error (MAE), the root mean squared error (RMSE) and the mean absolute percentage error (MAPE), defined by:

$$MAE = \frac{1}{s}\sum_{h=1}^{s}|\hat{P}_h - P_h| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{s}\sum_{h=1}^{s}(\hat{P}_h - P_h)^2} \quad (11)$$

$$MAPE = 100 \times \frac{1}{s}\sum_{h=1}^{s}\frac{|\hat{P}_h - P_h|}{P_s} \quad (12)$$

Where h is the hour index, s is the total number of hours in the test period, $\hat{P}_h$ and P_h are predicted and actual power at hour h, respectively, and P_s is the average power over the test period:

$$P_s = \frac{1}{s}\sum_{h=1}^{s}P_h \quad (13)$$

The MAE is the most natural criterion; it is simply the average of all errors, through absolute value to avoid offset. The RMSE does the same thing, except avoiding offset by squaring instead of absolute value. The MAE and RMSE are raw quantification, i.e. they are expressed in terms of the original signal unit, which is kW. This is not very useful. For instance, the same error of 50 kW has not the same impact on a little PV site as it may have on large sites. The MAPE is the consequence of the need for relative error metric. As it is expressed in percentage, it is more significant. However, it is also affected by the PV site capacity.

TABLE II
FORECAST RESULTS IN TERMS OF PROPOSED EVALUATION CRITERIA

Period	1 to 7 October	1 to 7 January	1 to 7 April	1 to 7 July
Max P (kW)	369.9900	408.6350	406.9250	253.0700
MAE (kW)	16.6341	30.1435	35.5628	15.1879
RMSE (kW)	24.6179	44.3428	61.7241	26.5178
MAPE (%)	10.2558	19.1602	28.9766	17.4471

Table II shows the obtained results for all test weeks through the three proposed criteria, in addition to the maximum recorded power measure P. The RMSE is a bit higher than MAE in most cases. In accordance with qualitative evaluation, the least accurate results are those achieved in April. The MAPE is affected by the average measured power P_s, so it has higher values for lower P_s. For this reason, the MAPE has lower values in large sites, where P_s is pretty high. This is the major inconvenient of this type of criteria.

B. Forecast without prior knowledge of solar irradiance

In this section, the future solar irradiance will be removed from inputs. Indeed, future weather conditions, especially irradiance, are not known in advance. The role of the predictor is to estimate the future power without prior knowledge of weather conditions. However, only future humidity and temperature will be kept, since they are relatively easy to predict. The model will then be modified to take only three inputs; future temperature, future humidity, and current power. The output should be the future power.

Two tests are then carried out in order to assess the efficiency of the new model. Naturally, a decrease of accuracy is expected, since the most influential input is absent. The first weeks of January and August 2014 are chosen for test. Predicted and actual curves have similar general trends and magnitudes, but they differ in sharp and brief variations. The error criteria are also computed, including MAE, RMSE and MAPE. In addition, persistence and neural network predictions are also performed in order to compare results. The persistence (PER) assumes simply that future power (of day d) is equal to current power (of day $d-1$). The artificial neural network (ANN) is a simple feedforward network with one hidden layer of ten neurons. Results are given by Table III, which proves the superiority of RF in both cases. It is true that ANN could provide much better results once optimized, but the aim here is to prove the advantage of nonparametric methods such as RF. Nonparametric methods do not need optimization or parameters' tuning. It is important to notice also that persistence could provide very accurate results in case of consecutive similar days. Consecutive similar days is a phenomenon that occurs often in spring and summer, when the solar irradiance remains the same for several days, along with complete absence of clouds.

TABLE III
RESULTS OF FORECAST WITHOUT PRIOR KNOWLEDGE OF SOLAR IRRADIANCE

Period	1 to 7 January 2014			1 to 7 August 2014		
Method	**PER**	**ANN**	**RF**	**PER**	**ANN**	**RF**
Max P (kW)	408.6350			301.9750		
MAE (kW)	66.4783	66.5198	44.2708	25.6051	25.6964	20.9176
RMSE (kW)	104.4875	101.4455	59.3910	47.0946	42.0095	35.0088
MAPE (%)	42.2558	42.2822	28.1400	25.9903	26.0830	21.2323

V. CONCLUSION

Photovoltaic power is becoming more and more integrated in modern grids, thanks to panels' cost decrease and power conversion advances. Hence, forecasting future power provided by PV sites becomes a necessity, to deal with intermittence and volatility induced by this type of resources. This paper proposed a predictor based on random forests, for day-ahead power forecast. The random forest is characterized by built-in cross validation, immunity to irrelevant inputs and importance measures. This last feature is utilized to select the most appropriate inputs of the predictor, in addition to conventional correlation plots and profile analyses. The forecaster was then built using 15 variants for 15 hours of the day. Qualitative and quantitative analyses are performed in order to assess the efficiency of the proposed method, with and without prior knowledge of solar irradiance. Tests show accurate and satisfactory results in most cases, and a good ability to handle seasonal effects.

REFERENCES

[1] F. Barbieri, S. Rajakaruna, and A. Ghosh, "Very short-term photovoltaic power forecasting with cloud modeling: A review," *Renewable and Sustainable Energy Reviews*, pp. –, 2016.

[2] A. Vaz, B. Elsinga, W. van Sark, and M. Brito, "An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in utrecht, the netherlands," *Renewable Energy*, vol. 85, pp. 631 – 641, 2016.

[3] J. Zhang, B.-M. Hodge, S. Lu, H. F. Hamann, B. Lehman, J. Simmons, E. Campos, V. Banunarayanan, J. Black, and J. Tedesco, "Baseline and target values for regional and point {PV} power forecasts: Toward improved solar forecasting," *Solar Energy*, vol. 122, pp. 804 – 819, 2015.

[4] M. Q. Raza, M. Nadarajah, and C. Ekanayake, "On recent advances in {PV} output power forecast," *Solar Energy*, vol. 136, pp. 125 – 144, 2016.

[5] A. E. Hendouzi and A. Bourouhou, "Forecasting of pv power application to pv power penetration in a microgrid," in *2016 International Conference on Electrical and Information Technologies (ICEIT)*, May 2016, pp. 468–473.

[6] D. van der Meer, G. R. C. Mouli, G. Morales-Espana, L. R. Elizondo, and P. Bauer, "Energy management system with pv power forecast to optimally charge evs at the workplace," *IEEE Transactions on Industrial Informatics*, vol. PP, no. 99, pp. 1–1, 2016.

[7] F. Almonacid, P. Prez-Higueras, E. F. Fernndez, and L. Hontoria, "A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a {PV} generator," *Energy Conversion and Management*, vol. 85, pp. 389 – 398, 2014.

[8] Z. Zhong, J. Tan, T. Zhang, and L. Zhu, "Pv power short-term forecasting model based on the data gathered from monitoring network," *China Communications*, vol. 11, no. 14, pp. 61–69, Supplement 2014.

[9] K. Ding, L. Feng, X. Wang, S. Qin, and J. Mao, "Forecast of pv power generation based on residual correction of markov chain," in *2015 International Conference on Control, Automation and Information Sciences (ICCAIS)*, Oct 2015, pp. 355–359.

[10] M. Q. Raza, M. Nadarajah, and C. Ekanayake, "An improved neural ensemble framework for accurate pv output power forecast," in *2016 Australasian Universities Power Engineering Conference (AUPEC)*, Sept 2016, pp. 1–6.

[11] D. AlHakeem, P. Mandal, A. U. Haque, A. Yona, T. Senjyu, and T. L. Tseng, "A new strategy to quantify uncertainties of wavelet-grnn-pso based solar pv power forecasts using bootstrap confidence intervals," in *2015 IEEE Power Energy Society General Meeting*, July 2015, pp. 1–5.

[12] M. D. Giorgi, M. Malvoni, and P. Congedo, "Comparison of strategies for multi-step ahead photovoltaic power forecasting models based on hybrid group method of data handling networks and least square support vector machine," *Energy*, vol. 107, pp. 360 – 373, 2016.

[13] M. G. D. Giorgi, P. M. Congedo, M. Malvoni, and D. Laforgia, "Error analysis of hybrid photovoltaic power forecasting models: A case study of mediterranean climate," *Energy Conversion and Management*, vol. 100, pp. 117 – 130, 2015.

[14] M. P. Almeida, O. Perpin, and L. Narvarte, "{PV} power forecast using a nonparametric {PV} model," *Solar Energy*, vol. 115, pp. 354 – 368, 2015.

[15] K. E. Hagan, O. O. Oyebanjo, T. M. Masaud, and R. Challoo, "A probabilistic forecasting model for accurate estimation of pv solar and wind power generation," in *2016 IEEE Power and Energy Conference at Illinois (PECI)*, Feb 2016, pp. 1–5.

[16] 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. PP, no. 99, pp. 1–1, 2016.

[17] J. Vasilj, P. Sarajcev, and D. Jakus, "Pv power forecast error simulation model," in *2015 12th International Conference on the European Energy Market (EEM)*, May 2015, pp. 1–5.

[18] Z. Zhen, F. Wang, Z. Mi, Y. Sun, and H. Sun, "Cloud tracking and forecasting method based on optimization model for pv power forecasting," in *2015 Australasian Universities Power Engineering Conference (AUPEC)*, Sept 2015, pp. 1–4.

[19] D. P. Larson, L. Nonnenmacher, and C. F. Coimbra, "Day-ahead forecasting of solar power output from photovoltaic plants in the american southwest," *Renewable Energy*, vol. 91, pp. 11 – 20, 2016.

[20] A. Dolara, S. Leva, and G. Manzolini, "Comparison of different physical models for {PV} power output prediction," *Solar Energy*, vol. 119, pp. 83 – 99, 2015.

[21] A. Dolara, S. Leva, M. Mussetta, and E. Ogliari, "Pv hourly day-ahead power forecasting in a micro grid context," in *2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*, June 2016, pp. 1–5.

[22] B. Wolff, J. Khnert, E. Lorenz, O. Kramer, and D. Heinemann, "Comparing support vector regression for {PV} power forecasting to a physical modeling approach using measurement, numerical weather prediction, and cloud motion data," *Solar Energy*, vol. 135, pp. 197 – 208, 2016.