Proc. of the 2nd International Conference on Electrical, Communication and Computer Engineering (ICECCE)
12-13 June 2020, Istanbul, Turkey

Day-Ahead Prediction of PV Generation Using Weather Forecast Data: a Case Study in the UK

Mohammad Monfared
Department of Electrical and Electronic Engineering
Swansea University
Swansea, U.K.
Department of Electrical Engineering
Ferdowsi University of Mashhad
Mashhad, Iran
mohammad.monfared@swansea.ac.uk

Meghdad Fazeli
Department of Electrical and Electronic Engineering
Swansea University
Swansea, U.K.
m.fazeli@swansea.ac.uk

Richard Lewis
SPECIFIC, College of Engineering
Swansea University
Swansea, U.K.
r.p.lewis@swansea.ac.uk

Justin Searle
SPECIFIC, College of Engineering
Swansea University
Swansea, U.K.
j.r.searle@swansea.ac.uk

***Abstract*—With the ever-increasing capacity of PV installations, the requirement for accurate and computationally efficient day-ahead forecasts is becoming more evident, while the solution remains a real challenge. Different techniques are being employed to transform a combination of weather forecasts and historical measurements into PV generation predictions. In this paper, the weather forecast data, provided by the UK Met Office, and the historical measurements are used to construct three different prediction models, based on linear least square regression (LSR), artificial neural network (ANN), and fuzzy. All the models can learn from new available data while running the forecasts. The results of an almost one-year study show that the best permutation (for the under-study case) is achieved by averaging the forecasts from LSR and ANN.**

Keywords—Day-ahead, forecast, prediction, PV

I. Introduction

Day-ahead generation prediction is of crucial importance for management of storage systems and the next day trade and load planning for both Photovoltaic (PV) plant owners and operators [1]. Day-ahead prediction of PV production is predominantly performed by transforming a combination of weather forecasts and historical measurements into power or energy data. To do so, there are several statistical approaches that can effectively employ the numerical weather prediction (NWP) and historical generation data to provide accurate and reliable PV energy predictions. Among different statistical techniques, the ones based on least square regression (LSR) and artificial intelligence (AI) have already shown success [2-4]. The LSR technique is a linear model and much easier to implement in contrast to the more advanced AI-based techniques, such as artificial neural networks (ANNs) and fuzzy systems, which are highly capable to model very nonlinear systems at the price of more complexity and computational effort. All these solutions use the most relevant information from weather forecasts (provided by a specialized meteorological agency) in conjunction with the site's historical power/energy measurements to build a prediction model estimating various kinds of PV related outputs, such as solar irradiance and power/energy generation. Evidently, the results obtained with these approaches are highly dependent on the quality of the weather forecasts. On the other hand, these solutions are efficient in terms of resources and costs, because they benefit from the expertise and immense computing capabilities of organizations specialized in the weather forecasts [5].

In this work, the NWP data provided by the UK Met Office and the historical measurements are used to develop three different prediction models for the day-ahead total PV energy generation in kWh. The best features that can be used for prediction purpose are first obtained through a correlation analysis and then the three predictors, based on LSR, ANN and fuzzy, are implemented and initially trained with just one-week data. Then these models are used to provide daily prediction of (almost) one-year energy production. All the techniques are designed such that they are able to learn from new data while running the forecasts. The results of one-year study show that the best permutation is achieved by averaging the forecasts from LSR and ANN, which also avoids the risk of overprediction of energy generation.

II. Modeling the Prediction Problem

In this study, the day-ahead generation forecasts for a 22.3kWp PV system, located in the "Active Office Building" in Swansea University Bay Campus, using LSR, ANN and fuzzy, are developed and compared. The NWP information and the site historical recorded energy generations are used as the inputs.

A. NWP Data

The detailed forecasts from NWP model of the UK Met Office, known as the Unified Model (UM), for Swansea City are used as the inputs to the energy prediction model [6]. The UK Met office is recognized as a world-leader in climate and weather science and a trusted provider of public weather information across the globe. The weather predictions include temperature, UV index, weather description (as a code), probability of precipitation, wind speed, visibility, etc. [7] for periods of 3-hours, up to 5 days ahead. According to the sunrise and sunset hours of different months in the UK, the relevant data periods are used for statistical analysis and prediction purposes, listed in Table I. Also, for this analysis and the benchmark test, the real PV power data measured at

This work was made possible by the support given to the SPECIFIC Innovation and Knowledge Centre by the Engineering and Physical Science Research Council (EP/N020863/1); Innovate UK (920036) and by the European Regional Development Fund (c80892) through the Welsh Government.

978-1-7281-7116-6/20/$31.00 ©2020 Crown

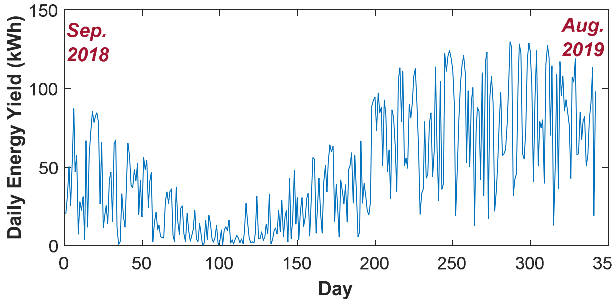

Fig. 1. Measured PV energy generation for the Active Office building in 2018-2019.

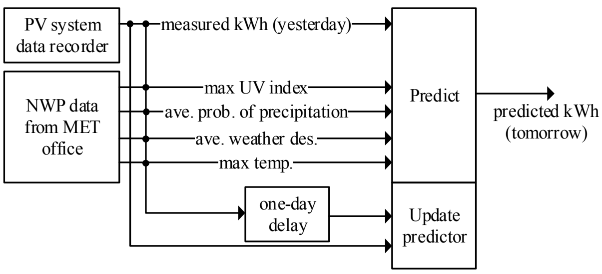

Fig. 2. One-day ahead prediction from NWP data.

Swansea University (Active Building Centre) in the UK during 2018-2019 are used. The daily energy generation data for almost one year, illustrated in Fig. 1, indicate the highly volatile weather conditions with high fluctuations and rapid changes over all seasons.

An essential part for design and implementation of a reliable time series prediction algorithm is the input selection. The set of inputs chosen for the day-ahead prediction of PV energy is decided based on the correlation analysis, detailed as follows. Among different meteorological data, eight daily indices are considered as candidate features for prediction purposes. These parameters are calculated over each day from the data bins detailed in Table I. The total energy generation of yesterday is also considered as another possible input to the prediction model. To find the best choices, a linear regression analysis using the well-known Pearson correlation coefficient was conducted between each of these parameters and the total energy generation on the same day. The results are summarized in Table II in a descending order. According to the Pearson correlation coefficient scale, the closer the absolute value to unity, the stronger the correlation between that parameter and the energy generation. As it can be seen from Table II, the maximum daily UV index and the maximum daily wind speed have the most and the least correlation to the energy generation, respectively. To decide about the proper features among this set of parameters, the possibility of correlation among these variables must also be studied. In other words, the parameters that are least dependent to each other with a high correlation factor (to the generated energy) are the best candidates. So, the matrix of correlation is calculated, and the results are presented in Table III that show strong correlations among some parameters. For example, the maximum and average daily temperatures and UV indices have a correlation of 0.98, which means one can be used for predictions while the other one is practically redundant. Following the correlation analysis, the set of proper features for prediction is decided as:

W = Function (the maximum UV index, the maximum temperature, the average weather description, the average probability of precipitation, yesterday kWh). (1)

TABLE I. PERIODS OF WEATHER PREDICTION DATA USED FOR PREDICTION PURPOSE.

Month	daytime	Bins to be included in analysis
1 Jan	8:00-16:00	9AM, 12PM
2 Feb	7:00-17:00	6AM, 9AM, 12AM, 3PM
3 Mar	6:30-19:30	6AM, 9AM, 12AM, 3PM
4 Apr	5:30-20:00	6AM, 9AM, 12AM, 3PM
5 May	5:00-21:00	6AM, 9AM, 12AM, 3PM, 6PM
6 June	5:00-21:00	6AM, 9AM, 12AM, 3PM, 6PM
7 July	5:00-21:00	6AM, 9AM, 12AM, 3PM, 6PM
8 Aug	5:30-20:30	6AM, 9AM, 12AM, 3PM, 6PM
9 Sep	6:00-19:30	6AM, 9AM, 12AM, 3PM
10 Oct	7:00-18:30	6AM, 9AM, 12AM, 3PM
11 Nov	7:00-16:30	6AM, 9AM, 12AM, 3PM
12 Dec	8:00-15:30	9AM, 12PM

TABLE II. CORRELATION ANALYSIS OF PV GENERATION STATISTIC CHARACTERISTICS (KWH TO DIFFERENT METEOROLOGICAL DATA OF THE SAME DAY AND KWH OF YESTERDAY).

Meteorological Parameter	Correlation
Max. UV index	0.86
Ave. UV index	0.83
Max. temperature	0.60
Ave. weather description	-0.53
Ave. temperature	0.51
Ave. probability of precipitation	-0.46
Ave. wind speed	-0.20
Max. wind speed	-0.09
Historical data	
Yesterday kWh	0.65

TABLE III. CORRELATION AMONG DIFFERENT DATA AVAILABLE FOR PREDICTION PURPOSE.

Max. temp.	Ave. temp.	Max. UV index	Ave. UV index	Ave. weather description	Ave. probability of precipitation	Max. wind speed	Ave. wind speed	Yesterday kWh	
1	0.98	0.74	0.71	-0.16	-0.17	0.16	0.06	0.58	**Max. temp.**
	1	0.69	0.68	-0.06	-0.09	0.22	0.14	0.51	**Ave. temp.**
		1	0.98	-0.28	-0.30	0.02	-0.08	0.70	**Max. UV index**
			1	-0.25	-0.28	0.01	-0.08	0.70	**Ave. UV index**
				1	0.84	0.40	0.42	-0.23	**Ave. weather description**
					1	0.48	0.49	-0.17	**Ave. probability of precipitation**
						1	0.92	-0.02	**Max. wind speed**
							1	-0.14	**Ave. wind speed**
								1	**Yesterday kWh**

B. *Prediction Models*

The total daily energy generation by the Active Office system (rated at 22.3 kWp) during September 2018 to August 2019 is predicted using the three prediction methods: LSR, ANN and fuzzy models. The first is a well-known linear approximation and the two others are successful non-linear statistical models. The first week data were used for initial learning of the models and then the prediction for daily energy generation in next day ($i + 1$) was performed on day before (i) using the latest weather forecast information for the next day ($i + 1$) and the measured energy generation for that day (i).

It is noted that all the above techniques have a training stage using the historical energy and weather data and can derive a function that computes day-ahead energy from the set of forecasted weather parameters and the measured energy generation. As mentioned above, the training data set for all models is just one week (first week of September 2018) and the predictions start from the second week. However, with an active learning strategy, detailed in [8], all models are continuously learning from the new data while the predictions are running based on the available information up to that instance of time. Indeed, the prediction models will be updated at the end of each day. In other words, they will be trained again by including the newly available data for that day. Therefore, as the time goes on the accuracy of prediction improves. The overall block diagram of the prediction system is shown in Fig. 2.

1) Linear Least Square Regression (LSR)

LSR is almost the most common predictor that establishes a linear relationship between a dependent variable, e.g. $[Y]_{n\times1}$ = (PV energy generation), and a set of independent variables, here $[X]_{n\times6} = [1, W]$ that $[W]_{n\times5}$ is defined in (1). The trained model $[Z]_{6\times1}$ is then obtained as (2), which contains the constant gains of the prediction model of (3).

$$Z = (X^T X)^{-1} X^T Y \tag{2}$$

$$\text{Predicted kWh} = \sum_{j=1}^{6} Z_i X_i \tag{3}$$

2) Generalized Regression Neural Network (GRNN)

The GRNN is a new member of radial basis artificial neural networks that offers a faster training time, the generalization ability and less parameters to set (no need for backpropagation). Also, as a very interesting and unique feature, the GRNN training result is deterministic and not sensitive to the initial weight values and consequently the local minima is no longer a problem. For these reasons, the GRNN has already found more successful applications than the feedforward back-propagation method.

The architecture of the GRNN has two layers, the radial basis at first layer followed by an especial linear layer. The output of GRNN can be calculated using

$$\text{Predicted kWh} = \sum_{j=1}^{n} Y_i e^{\left(-d_i/_{2\sigma^2}\right)} \Big/ \sum_{j=1}^{n} e^{\left(-d_i/_{2\sigma^2}\right)} \tag{4}$$

where d_i, defined in (5), is the Euclidean distance and σ is the smoothing parameter of the GRNN [9], [10].

$$d_i = (X - X_i)^T (X - X_i) \tag{5}$$

A serious problem in the training process of most ANNs is the possibility of overfitting, i.e. if the training process is not stopped at the right point, the ANN begins to "memorize" the training dataset rather than performing "generalization" from it. So, a supervisory training is almost inevitable. However, the GRNN has already shown a superior generalization performance and the training stage is not iterative. Consequently, if one decides to add new input-output data pairs to an existing GRNN, while running the predictions, then the whole training process can be simply repeated without the need for any kind of supervision or concerns about overfitting or local minima problems.

In this work, the biases of radial basis functions are all set to 0.8326.

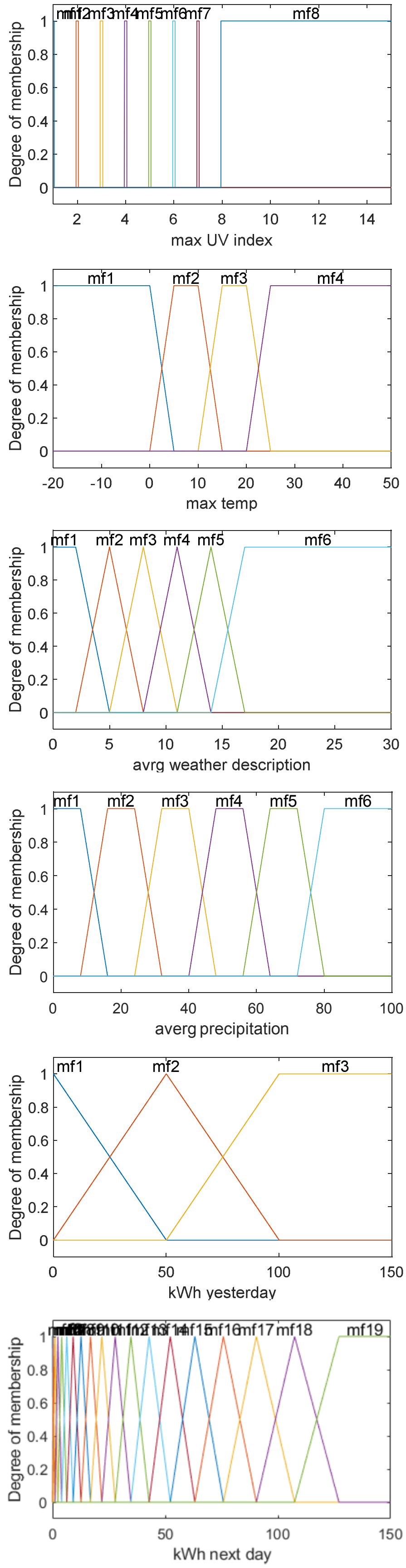

Fig. 3. Fuzzy membership functions.

3) Fuzzy

As another AI technique, fuzzy logic can provide a nonlinear mapping of input-output data pairs using soft linguistic variables. The fuzzy logic system is composed of the

Knowledge Base (KB) and the Inference System (IS). The KB has the available information about the problem in the form of linguistic IF-THEN rules and is used by the IS to decide on the system outputs from inputs. Based on the successful work in [8], a generalized fuzzy predictor that uses the Wang-Mendel training algorithm to directly extract the fuzzy rules from the training data pairs is also employed here. This strategy exploits the limited training data more efficiently. In addition, an online additive learning routine, which is also proposed in [8] and enables the predictor to learn from new data while running the predictions, is used. As a result, the KB will be adaptively updated while the predictions are running. The type and shape of membership functions (mfs) for input and output variables are shown in Fig. 3.

III. Performance Comparison

In order to investigate the effectiveness and compare the accuracy of different prediction strategies for the Swansea University's Active Office PV system, root mean square error (RMSE), coefficient of determination (COD) and mean bias error (MBE), defined in (6), (7) and (8), respectively, are employed. In these equations, N is the total number of samples, kWh_i and kWh_{ip} are the actual and the predicted values of generated PV energy and kWh_m is the mean of kWh_{is}. The ideal values for the RMSE and the COD are 0 and 1, respectively.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(kWh_{ip} - kWh_i\right)^2} \quad (6)$$

$$\mathrm{COD} = 1 - \left(\frac{\sum_{i=1}^{N}\left(kWh_i - kWh_{ip}\right)^2}{N-2} \bigg/ \frac{\sum_{i=1}^{N}\left(kWh_i - kWh_m\right)^2}{N-1}\right) \quad (7)$$

$$\mathrm{MBE} = \frac{1}{N}\sum_{i=1}^{N}\left(kWh_{ip} - kWh_i\right) \quad (8)$$

It is worth emphasizing again that the data of the first week of September 2018 are used for initial training of all predictors and as the time series progresses the predictors are updated with the new data available. The results for approximately one year running of different prediction algorithms are summarized in Table IV. Primarily from two main error metrices of RMSE and COD one can conclude that the LSR outperforms the two artificial intelligence techniques. The superior performance of the LSR somehow shows that a simple linear function can adequately model the relationship between the weather parameters and PV energy and this function is not highly affected by the size of the data. The fuzzy is in the second place while the GRNN shows the poorest performance. These results are mainly due to the fact that with a limited number of historical data pairs available for training, both the GRNN and the fuzzy cannot learn enough about the process to be able to provide accurate predictions. This problem is especially more evident with the GRNN predictor. Indeed, a proper AI predictor inevitably requires a large set of training data to build an adequate knowledge. It was especially seen during the predictions by the fuzzy that for many of input data sets no rules could be fired. In this situation the precited kWh was just simply assumed to be equal to the previous day measured kWh, as what is conventionally done in the persistence method. Another important practical issue with the prediction algorithm is the bias, which can be quantitively investigated with MBE. The ideal value for MBE is zero and a positive/negative MBE means overprediction/underprediction. In practice and from the operator point of view, overprediction of the PV energy generation causes more technical and financial problems.

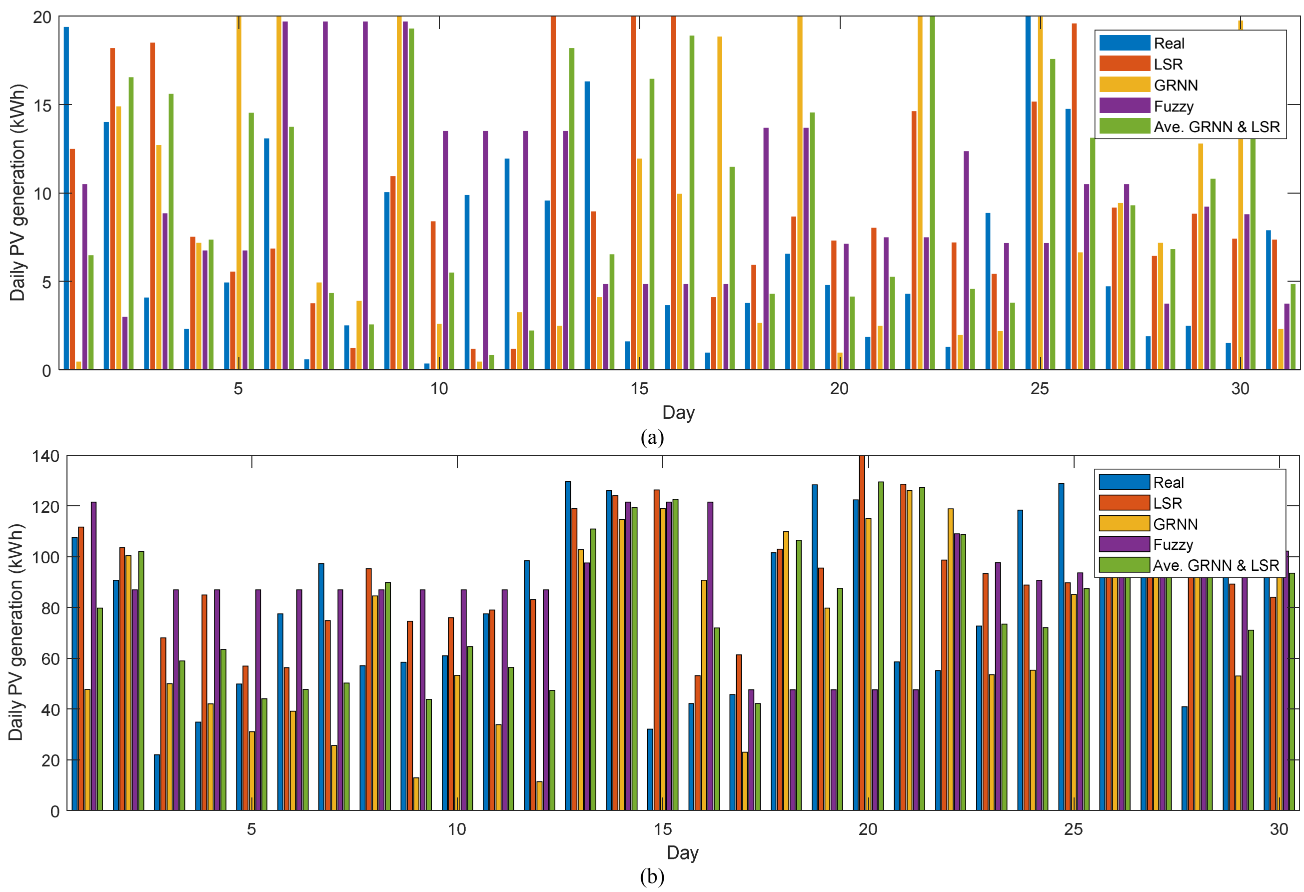

Fig. 4. Prediction results for (a) December 2018, and (b) June 2019.

TABLE IV. STATISTICAL PERFORMANCE ANALYSIS.

Strategy	RMSE (kWh)	COD (%)	MBE (kWh)
LSR	16.2	0.83	+3.2
GRNN	25.9	0.57	-2.9
Fuzzy	24.4	0.64	+4.1
Averaged GRNN & LSR	17.8	0.78	+0.2

Indeed, it is more convenient and safer to curtail the excess PV power (as with the underpredictions) than shed the loads or bring in the backup sources (as with the overpredictions). Table IV shows that both LSR and fuzzy overestimate the energy generation while GRNN underestimates it that can be considered as an advantage for the GRNN. Based on the above discussions, one may decide to have more accurate predictions with the LSR or an underprediction dominated error with the GRNN. The other compromise is to run both techniques simultaneously and use the averaged value of both outputs as the final prediction. The results are also shown in the last row of Table IV. This way, the accuracy is much better than the GRNN and relatively comparable to the LSR and at the same time the overprediction problem with the LSR is effectively solved. Finally, Fig. 4 shows bar charts of real and predicted values by different techniques for two seasons on winter (December 2018) and summer (June 2019). It can be said that the accuracy of predictions during the wintertime is less than the summertime. However, the amount of power generated each day in winter is much less than summer.

IV. CONCLUSION

The need for solar power prediction from weather forecast data has become a serious challenge with the increasing deployment of PV systems in the UK. This paper used the data provided by the UK Met Office for Swansea city to predict the day-ahead total PV power generation using LSR, GRNN and Fuzzy. All models had the ability of online learning while doing the predictions. Final results show that the averaged forecasts by LSR and GRNN have a good level of accuracy while the risk of overprediction of energy generation is also avoided.

REFERENCES

[1] P. J. Alet et al., "Assessing the need for better forecasting and observability of pv," 2017 [Online]. Available: https://etip-pv.eu/publications/etip-pv-publications

[2] A. Nespoli et al., "Day-ahead photovoltaic forecasting: A comparison of the most effective techniques," in Energies, vol. 12, no. 9, pp. 1–15, 2019.

[3] J. G. Kim, D. H. Kim, W. S. Yoo, J. Y. Lee, and Y. B. Kim, "Daily prediction of solar power generation based on weather forecast information in Korea," in IET Renew. Power Gener., vol. 11, no. 10, pp. 1268–1273, 2017.

[4] D. P. Larson, L. Nonnenmacher, and C. F. M. Coimbra, "Day-ahead forecasting of solar power output from photovoltaic plants in the American Southwest," in Renew. Energy, vol. 91, pp. 11–20, 2016.

[5] L. H. Duy, S. Lespinats, and Y. M. Bourien, "D3.5–Summary of PV Production Forecasting Tool Design and Implementation," 2015 [Online]. Available: www.greendatanet-project.eu

[6] "Numerical weather prediction models" [Online]. Available: www.metoffice.gov.uk/research/approach/modelling-systems/unified-model/weather-forecasting

[7] "Weather and climate data" [Online]. Available: www.metoffice.gov.uk/services/data/datapoint/code-definitions

[8] M. Monfared, M. Fazeli, R. Lewis, and J. Searle, "Fuzzy Predictor with Additive Learning for very Short-Term PV Power Generation," in IEEE Access, vol. 7, pp. 91183-91192, 2019.

[9] D. F. Specht, "A general regression neural network," in IEEE Transactions on Neural Networks, vol. 2, no. 6, pp. 568-576, 1991.

[10] A. J. Al-Mahasneh, S. G. Anavatti, and M. A. Garratt, "Review of applications of generalized regression neural networks in identification and control of dynamic systems," International Conference on Aeronautics, Astronautics and Aviation 2018.