PV Generation Forecasting Model for Energy Management in Buildings

Brígida Teixeira[1,2](✉), Tiago Pinto[1,2], Pedro Faria[1,2], and Zita Vale[2]

[1] GECAD - Research Group on Intelligent Engineering and Computing for Advanced Innovation and Development, Porto, Portugal
[2] Institute of Engineering, Polytechnic of Porto (ISEP/IPP), Porto, Portugal
{bccta,tcp,pnf,zav}@isep.ipp.pt

Abstract. The increasing penetration of renewable energy sources and the need to adjust to the future demand requires adopting measures to improve energy resources management, especially in buildings. In this context, PV generation forecast has an essential role in the energy management entities by preventing problems related to intermittent weather conditions and allowing participation in incentive programs to reduce energy consumption. This paper proposes an automatic model for the day-ahead PV generation forecast, combining several forecasting algorithms with the expected weather conditions. To this end, this model communicates with a SCADA system, which is responsible for the cyber-physical energy management of an actual building.

Keywords: Energy management system · Forecast · PV generation

1 Introduction

Currently, sustainability is one of the biggest challenges in the energy sector. In an environment where the use of electronic devices and the internet is increasingly significant in the daily routine, studies show that the demand will more than double in the coming years, conducting to the need to produce more energy [4]. European Commission also says that the energy consumed in buildings corresponds to 40% of the total energy demand [2].

In order to deal with current and future demand requirements and considering the urgency of significantly reducing the environmental impact of fossil fuels, there is an investment in the penetration of renewable energy sources worldwide through the creation of new technologies, models, and legislation [3]. However, the intermittency of energy sources based on weather conditions raises several

This work was supported by the MAS-Society Project co-funded by Portugal 2020 Fundo Europeu de Desenvolvimento Regional (FEDER) through PO CI, and under Grant UIDB/00760/2020. Brígida Teixeira was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) PhD studentship with reference 2020.08174.BD.

G. Marreiros et al. (Eds.): EPIA 2021, LNAI 12981, pp. 176–182, 2021.
https://doi.org/10.1007/978-3-030-86230-5_14

challenges due to their uncertainty to satisfy the demand. In this way, forecasting has an essential role in providing energy management systems with information to better use energy, allowing them to take full advantage of renewable energy sources while decreasing cost and waste of energy.

This paper presents an automatic PV generation forecast model for Anonymous building energy management. The proposed methodology aims to improve the solar forecast results, supported by different data models and four forecast methods, namely Neural Artificial Networks (ANN), Support Vector Machines (SVM), Random Forest, and ARIMA. It is connected to a SCADA system, which ensures the cyber-physical energy management of the building. The Forecasting process executes in five main phases: data selection and transformation, creation of data models, model hyperparameters, forecast, and evaluation.

2 SCADA System

In order to monitor and manage GECAD research group energy resources, this section presents the implemented SCADA system [1] in Building N facilities located in Porto, Portugal. This building has twelve offices, a meeting room, a server room, two laboratories, two toilets, a kitchen, and three corridors. These rooms are equipped with several sensors and energy meters, which communicate with Programmable Logic Controllers (PLCs) connected to a central PLC to monitor and control the building. The communications are via TCP/IP protocol. The visualization and control of these resources are available from a touch panel installed in the hall or an internal web page. The sensors collect real-time data of the building's environment, allowing the observation of several indicators, such as external temperature, internal temperature, brightness, humidity, CO_2 levels, air quality, generation, and consumption.

The building has three-phase meters, which allows to analysis separately the consumption by type of resource. Phase 1 allows monitoring the loads in the area where the energy meter is installed; phase 2 observes the consumption of the air conditioning systems; phase 3 reads the consumption of the lights. Besides, this building has a PV system with a maximum 7.5 KW capacity installed on the rooftop. The energy generated by the PV satisfies part of the energy demand, and if there is a surplus of energy, it is injected into the network.

Furthermore, GECAD also has access to the Institute of Engineering of Polytechnic of Porto meteorological station (meteo@isep)[1], which enables to access real-time weather information (i.e., feel temperature, real temperature, wind speed, radiation, atmospheric pressure, humidity, and rain), and consult forecasts for the next three days.

3 Solar Forecasting Model

In building energy management and smart grids, having information on the estimated consumption and generation is essential for the optimization of energy

[1] meteo@isep - https://meteo.isep.ipp.pt/gauges.

resources, taking advantage of grid incentives to reduce electricity consumption, namely by participating in demand response programs. However, the weather conditions variation represents a significant challenge in forecasting renewable energy since they can harm the results, especially when the frequency is less than an hour. For example, the passage of a cloud causes a decrease in radiation, originating an error that will have more impact in a 15-min time interval forecast than a 1-h frequency.

This paper proposes a methodology to forecast the day-ahead PV generation, with a 15-min time interval, which corresponds to a total of 96 periods. The model is implemented in Python, and it contains four different artificial intelligence techniques to forecast PV generation, namely ANN, SVM, and Random Forests from scikit-learn library, and ARIMA from pmdarima library. The model's architecture is based on five phases. The selection of these techniques is based on previous works presented in the current literature.

The first phase is related to selecting and preparing the necessary historical data for the learning process of forecasting algorithms. This data is imported from SCADA's database and consists of the last 20 days of PV generation. Rain, radiation, and outside temperature are imported from the meteo@isep API as additional features for the learning process. After this, cleaning data occurs by detecting and replacing missing data, incorrect data, and outliers. The strategy for replacing corrupt or inaccurate data consists of an estimation based on the average between the last recorded value and the next.

The second phase is the generation of data models (or scenarios). For this purpose, the historical PV generation is used as a basis of the data model. Next, it will be added new information to this dataset, according the combination of several data transformations. The use of different data models helps to identify which type of information helps the most in the algorithms' training process. This are the considered transformations:

- Separation of the timestamp into the four columns (month, day, hour and minute);
- Insert a new column with the distance between the entry and the first value of forecast $t(0)$. For example, if the forecast starts at 00:00h, the train entry that corresponds to the previous period 23:45 h $t(-1)$ will have the value of -1, 23:30 h $t(-2)$ will have the value of -2, and so on. This strategy helps the algorithm to identify how old the entry is compared to the forecast;
- Insert three new columns with the values of the three periods of the previous day, at the same time $t(-96)$, $t(-97)$, $t(-98)$;
- Insert five new columns with the values of the five periods of the previous day, at the same time $t(-96)$, $t(-97)$, $t(-98)$, $t(-99)$, $t(-100)$;
- Insert seven new columns with the values of the seven periods of the previous day, at the same time $t(-96)$, $t(-97)$, $t(-98)$, $t(-99)$, $t(-100)$, $t(-101)$, $t(-102)$;
- Exclusion of night time periods, from 23:00 h to 05:00 h of the next day;
- Insert three new columns with the information of additional features (radiation, rain, and outside temperature).

Then, the data is split, where 80% is for training the model and 20% is for testing, following the normalization process.

The third phase is the tuning of the hyperparameters and training of the forecasting algorithms. For each algorithm, several configurations of the hyperparameters are tested to identify which one achieves more precise results. In the case of ANN, different solvers, activation functions, layers, and the number of nodes are tested. For the SVM, different kernels and gammas are tested. For the random forest, the studied parameters are the number of estimators and the criterions. In Arima, since it is a timeseries algorithm, its configuration is adjusted to the data frequency. Each hyperparameters setup is combined with all generated models to train the forecast algorithms.

The fourth phase is forecast execution. Once a week, this process runs in parallel with the previous phase, as the parameters are tested together with the forecast. In the remaining days, it is used only the algorithm that had the best performance in the training process. Instead of historical data, to perform the forecast, the algorithms require the forecast of additional features for the next day, obtained through access to the meteo@isep platform. This data is transformed according to the model with the best results (if used).

The fifth phase is the evaluation phase. The scenario that presented the best results is selected, namely the forecasting algorithm, the hyperparameters, and the model that most improved the learning process. The results obtained by combining these three factors are evaluated by calculating several error metrics: Minimum Error (MinE), Maximum Error (MaxE), Mean Absolute Error (MAE) (Eq. 1), Mean Absolute Percentage Error (MAPE) (Eq. 2), Root Mean Squared Error (RMSE) (Eq. 3) and Root Mean Squared Percentage Error (RMSPE) (Eq. 4). The scenario with the lowest MAPE value is selected.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \tag{1}$$

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(F_t - A_t)^2}{n}} \tag{3}$$

$$RMSPE = \frac{\sqrt{\frac{\sum_{t=1}^{n}(F_t - A_t)^2}{n}}}{\frac{\sum A_t}{n}} \tag{4}$$

where n is the number of error entries, A_t is the real value and F_t is the forecasted value, for the PV generation at period t.

4 Case Study

This section presents a practical case study in order to demonstrate the proposed model. This case study aims to forecast the PV generation for 21 April 2020 and analyze the behavior of the different forecast methods. In this way, the selected historical dataset is from 01 April 2020 to 20 April 2020, and it is composed by PV generation, radiation, rain, and outside temperature. Table 1 presents the best results obtained for the April 21 2020 forecast.

Table 1. Top 10 best results for solar forecast

	Study	Model	MinE	MaxE	MAE	MSE	RMSE	MAPE	RMSPE	Feat.	Alg.
1	1308	30	0.0000	0.3099	0.0333	0.0048	0.0694	0.1868	0.3453	True	ann
2	2718	56	0.0000	0.2293	0.0290	0.0035	0.0590	0.1893	0.3500	True	ann
3	1391	31	0.0000	0.3114	0.0321	0.0048	0.0690	0.1909	0.353	True	ann
4	1254	29	0.0000	0.3251	0.0336	0.0056	0.0746	0.1935	0.3607	True	ann
5	1385	31	0.0004	0.2578	0.0307	0.0038	0.0614	0.2006	0.3624	True	ann
6	1315	30	0.0000	0.3607	0.0392	0.0067	0.0820	0.2008	0.3641	True	ann
7	1411	32	0.0000	0.3637	0.0378	0.0064	0.0800	0.2050	0.3714	True	ann
8	1389	31	0.0001	0.3422	0.0347	0.0055	0.0740	0.2059	0.3621	True	ann
9	1252	29	0.0000	0.3436	0.0360	0.0060	0.0776	0.2062	0.3786	True	ann
10	1393	31	0.0000	0.3664	0.0389	0.0068	0.0825	0.2073	0.3726	True	ann

Regarding the table above, it is possible to observe that the best results were obtained using the ANN method (indicated in the Alg. column), using the additional features (indicated in the Feat. column). The other algorithms had a worse performance and are below the tenth position. SVM presents the best results after the ANN, in the 31st position. This variation uses the transformations of the timestamp division, count of the distance from the entries to the initial instant of the forecast, and uses the data from the periods $t(-96)$, $t(-97)$ and $t(-98)$ of the previous day. In this case, the MAPE value is 18.68%. It is also possible to see that this model appears more than once (studies/scenarios 1308 and 1315). The difference between them relies on the used hyperparameters to configure the forecasting algorithm. Study 1308 uses the 'adam' solver, the activation function 'relu', and three layers with 100, 50 and 25 nodes. Study 1315 uses the 'lbfgs' solver, 'logistic' activation function three layers with two nodes each (2, 2, 2), as parameters. The results of study 1308 can be seen in Fig. 1.

The graph in Fig. 1 shows that in the days before the forecast, the values of PV generation were higher than those that occurred on the forecast day (represented by the green line). However, the forecast line (shown in red) can detect the current weather changes and adjust to the actual values, analyzing the forecast of the additional features. Model 5, which corresponds to the same study but does not include features, has a MAPE value of 35.59%, representing an increase in the error of 16.91%.

Moreover, another critical factor that influences the error is the forecast of values near sunrise or sunset, where the sensitivity of the error is higher than in other periods. In other words, when the real value is 0, if the forecast value is greater than 0, the error is calculated with a large penalty. This penalty can be greater than 100% if the forecasted value is much higher or much lower than the

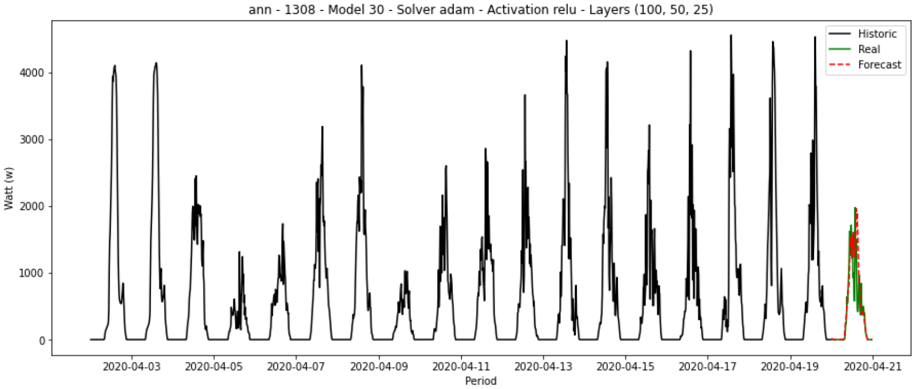

Fig. 1. Solar forecast results. (Color figure online)

real value, punishing the analysis of the model's error. In this specific case, the adjustment of these cases was made to consider a 100% error. However, the most appropriate solution in the future will be to use more suitable metrics to deal with these situations, as is the case of Mean Absolute Scaled Error (MASE).

5 Conclusions

This paper proposes an automatic PV generation forecasting model for building energy management. Thus, the model performs the forecast for the next day, with a 15-min interval, having at its disposal four different forecasting methods: ANN, SVM, Random Forest, and ARIMA. The one with the best results is selected after an exhaustive study of the intersection of the several forecasting methods, their hyperparameters, and data models that include features that help in the learning process. A case study was presented of the application of this model by using data collected in a real building. The results show that the model achieves promising results despite the intermittency of the weather.

As future work, it is suggested to use the error metric MASE for a more precise analyzes. Furthermore, a longer period of history may also be included, and other dataset transformations may be considered.

References

1. Abrishambaf, O., Faria, P., Vale, Z.: SCADA office building implementation in the context of an aggregator. In: Proceedings - IEEE 16th International Conference on Industrial Informatics, INDIN 2018, pp. 984–989 (2018). https://doi.org/10.1109/INDIN.2018.8471957
2. European Commission: Energy efficiency in buildings. Technical report (2020). https://ec.europa.eu/info/news/focus-energy-efficiency-buildings-2020-feb-17_en

3. European Commission: Directive (EU) 2019/944 of the European Parliament and of the Council. Official Journal of the European Union 125 (2019). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L0944&from=en
4. International Energy Agency: World energy outlook 2014 factsheet how will global energy markets evolve to 2040? p. 75739 (2015). www.worldenergyoutlook.orgwww.iea.org, www.worldenergyoutlook.org