

Comparative Study of Photovoltaic Power Forecasting Methods

Angelo A. Pelisson¹, Thiago F. Covões²,
Anderson W. Spengler¹, Pablo A. Jaskowiak¹

¹Joinville Technological Center
Federal University of Santa Catarina (UFSC)
Joinville, Santa Catarina, Brazil

²Center of Mathematics, Computing and Cognition
Federal University of ABC (UFABC)
Santo André, São Paulo, Brazil

angelo.a.p@posgrad.ufsc.br, thiago.covoes@ufabc.edu.br
anderson.spengler@ufsc.br, pablo.andretta@ufsc.br

Abstract. *Electricity consumption is growing rapidly worldwide. Renewable energy resources, such as solar energy, play a crucial role in this scenario, contributing to satisfy demand sustainability. Although the share of Photovoltaic (PV) power generation has increased in the past years, PV systems are quite sensitive to climatic and meteorological conditions, leading to undesirable power production variability. In order to improve energy grid stability, reliability, and management, accurate forecasting models that relate operational conditions to power output are needed. In this work we evaluate the performance of regression methods applied to forecast short term (next day) energy production of a PV Plant. Specifically, we consider five regression methods and different configurations of feature sets. Our results suggest that MLP and SVR provide the best forecasting results, in general. Also, although features based on different solar irradiance levels play a key role in predicting power generation, the use of additional features can improve prediction results.*

1. Introduction

Electricity demand is growing rapidly. Industries account for a large fraction of such demand, with consumption figures close to 35.6% of global energy production in 2017 [IEA 2017]. Technological innovations and improving living standards also mean more household appliances at homes. In Brazil, specifically, revenues related to household appliances grew 1.58% from 2018 to 2019 and an increase of 1.56% is expected in the year 2020 [Statista 2019]. Estimates suggest that, on average, global electricity demand will grow 2.1% a year until 2040 [IEA 2019]. The growing demand combined with the necessity to reduce the environmental impact of power generation has led to the development and adoption of clean energy resources.

The share of solar Photovoltaic (PV) systems in power generation has increased over the past years worldwide. Specifically, it grew 34.75% from 2016 to 2017 [IEA 2017]. While solar energy is a great source of renewable energy, its widespread adoption is hampered by its sensitivity to climatic and meteorological

conditions [Pedro and Coimbra 2012]. In fact, critical challenges in the transition to clean and renewable energy sources are often related to their capacity, availability, and reliability [Gielen et al. 2019]. These challenges are faced not only by Photovoltaic systems but other energy sources, such as, wind and tidal power [Inman et al. 2013].

Prior to the deployment of a PV Plant the environmental and climatic characteristics of the candidate locations should, ideally, be investigated. Ground and satellite data can be employed in order to quantify factors that directly affect the performance of a PV System, such as, wind speed, cloud coverage, ambient temperature and solar radiation [Das et al. 2018, Kaplani and Kaplanis 2014, Malvoni et al. 2017, Souza-Echer et al. 2006]. These factors will not only help to determine the best candidate location, but also support design and optimization at a given site during operation [Nonnenmacher et al. 2014]. Indeed, the variable and sometimes irregular nature of PV Plants poses substantial challenges for the current operation methods of energy producers, public companies, independent service operators, and mainly the concessionaires responsible for energy distribution [Pedro and Coimbra 2012].

In order to optimize grid response, the energy production of a PV Plant must be predicted as accurately as possible, providing grid managers with information on how much energy will be available at a given time [Kudo et al. 2009, Lorenz et al. 2009, Mahtta et al. 2014]. Thus, the accurate forecast of the power output of a PV Plant can provide security for the entire energy system, improving stability and reliability, and allowing for a better planning of energy generation, distribution and economic return [Das et al. 2018, Lo Brano et al. 2014, Muhammad Ehsan et al. 2017, Raza et al. 2016].

In this work, we evaluate the impact of different regression methods and feature sets when applied to short term forecasting of power output from a PV Plant. We consider ground collected data from a solarimetric station spanning the whole year of 2018. Our goal is to identify configurations that will allow the development of tools to assist management of PV generation systems and grids.

The remainder of the work is organized as follows. In Section 2, we review related work on power output forecasting for PV Plants. Experimental setup is presented in Section 3. First we describe the data, its features, and the pre-processing steps performed before its actual analysis. We then proceed to present methods and evaluation metrics. Results from our evaluation are discussed in Section 4. Conclusions and future work directions are given in Section 5.

2. Related Work

Several studies have focused on the prediction of power output for PV systems. Comprehensive reviews on the topic are provided by Raza et al. [Raza et al. 2016] and, more recently, by Das et al. [Das et al. 2018]. Both works discuss the main aspects related to the problem, such as forecasting horizon, features (variables), and models. In the following we briefly discuss findings related to these particular topics.

In terms of forecasting horizon, models fall within three major categories, namely: short, medium, and long term [Das et al. 2018]. Short term models aim to predict power

output for a few hours ahead, up to seven days. Medium term models consider a broader horizon, ranging from a week to a month. Models that provide predictions for periods greater than a month ahead are deemed long term. Some authors [Raza et al. 2016] also include a fourth category, named very short term forecasting, for predictions made almost in a real time fashion, considering minutes ahead.

Although solar irradiance plays a central role in PV systems, it is not the only feature affecting power output. Among the other features with moderate to high influence on the power output of PV systems, temperature, wind speed/direction, cloud coverage, humidity, and aerosol index are usually considered [Das et al. 2018]. Different feature subsets were evaluated by De Giorgi et al. [De Giorgi et al. 2014] and Huang et al. [Huang et al. 2010]. Both studies obtained better results when combined solar irradiance data with other features, such as ambient/module temperature, and humidity.

Raza et al. [Raza et al. 2016] observes that better predictions may be obtained with specialized models built for different weather conditions, e.g., sunny, rainy, or cloudy. However, results seem to depend on more factors than solely stratification by weather. Ding et al. [Ding et al. 2011] and Leva et al. [Leva et al. 2017] obtained better forecasting results for clear days (in which there are no clouds) than for rainy or cloudy days. With a similar procedure, however, Mellit et al. [Mellit et al. 2014] observed greater improvements for the model built for cloudy days in comparison to no stratification.

Besides weather, the geographical location of a PV Plant is often regarded as a key factor in the observed output of the system. Given that the correlations between meteorological parameters may change according to location, there is recommendation for their investigation in different application scenarios [Das et al. 2018].

In terms of prediction techniques, Artificial Neural Network (ANN) based models are amongst the most usual choices in the literature. Albeit a wide variety of ANN types have been used, Multilayer Perceptrons (MLPs) seem to prevail [Chow et al. 2012, Lo Brano et al. 2014, Raza et al. 2016, Muhammad Ehsan et al. 2017, Das et al. 2018]. Other methods, such as k-Nearest Neighbors (kNN) and Support Vectors Regression (SVR) have also been widely adopted [Fernandez-Jimenez et al. 2012, Arize and Nogueira Rios 2019]. Das et al. [Das et al. 2018] suggest, based on their literature review, that ANN models and Support Vector Regression (SVR) show the best results for the task, in general.

3. Material and Methods

In this section we detail the dataset employed in our study, its features and main characteristics. Then, we describe the methods used in our experimental section and the evaluation metrics adopted.

3.1. Data

The dataset considered in this work consists of ground collected data from a solarimetric station located at the Fotovoltaica/UFSC Solar Energy Research Laboratory, Florianópolis - SC - Brazil. The PV Plant responsible for generating power output is located at the same site. The dataset comprehends the energy production for the 2018 year, with a one minute resolution. Initially, raw data comprehended a total of 545,806 measurements (pre-processing is described in the following) obtained with a Campbell CR6 Datalogger.

The features collected by the solarimetric station are detailed in Table 1. The target feature, which corresponds to the power output at a given instant is also shown (this corresponds to the inverter output).

Table 1. Features from the solarimetric station and the PV Plant output.

Parameter	Abbreviation	Sensor	Unit
Direct Normal Irradiance	DNI	SHP1	W/m ²
Global Horizontal Irradiance	GHI	SMP22	W/m ²
Diffuse Horizontal Irradiance	DHI	SMP11	W/m ²
Ultraviolet Index	UV	CUV 5	W/m ²
Ambient Temperature	AT	HMP155	°C
Wind Speed	WS	WindSonic1	m/s
Relative Humidity	RH	HMP155	%
Power Output	PO	ABB Uno 2.0 Inverter	W

Before performing any analysis we pre-processed the raw data. In order to have a standard daily schedule we kept for further analysis measurements taken from 07:30 to 17:00. This interval is also employed in the work and analysis of Muhammad et al. [Muhammad Ehsan et al. 2017] and comprises most of the power generated by the PV Plant. With a regular daily schedule we proceeded to verify eventual data quality problems. In this step a few duplicated timestamps were removed. Furthermore, we identified that in some days the datalogger responsible for data collection went *stuck at a constant value*¹. Values of these spurious readings were discarded.

The solarimetric station and PV Plant from which the data came from are located at a laboratory. When maintenance or personal training is carried out the power inverter is turned off, for safety reasons. Such events result in misleading readings, given that data from the solarimetric station is correctly captured, while power output is fixed at zero, since the power inverter is not operating. Such events were also identified and removed from the dataset.

In order to reduce the size of the dataset we aggregated its timestamps considering 30 minutes intervals, thus resulting in a total of 20 observations per day with half an hour intervals (07:30, 08:00, ..., 16:30, 17:00). Each one of the 20 aggregated values was obtained by averaging the values within ± 15 minutes of its specified time. Such transformation was performed in order to allow us to evaluate the performance of the methods through the whole year period. After such pre-processing steps, we obtained a total of 5,375 observations (timestamps).

To give a glimpse on how daily profiles may vary through the year, we depict in Figure 1 illustrative examples of average daily profiles for the months of January and June (20 points per line). One can see that the daily power output is, on average, higher in January than in June (Figure 1a). The average daily profile considering irradiance components (GHI, DNI, and DHI) for these two months are depicted in Figure 1b and Figure 1c, considering the aggregates previously described. As expected, irradiance values rise sharply in the morning and decay at noon.

¹The person responsible for data collection confirmed these as spurious readings.

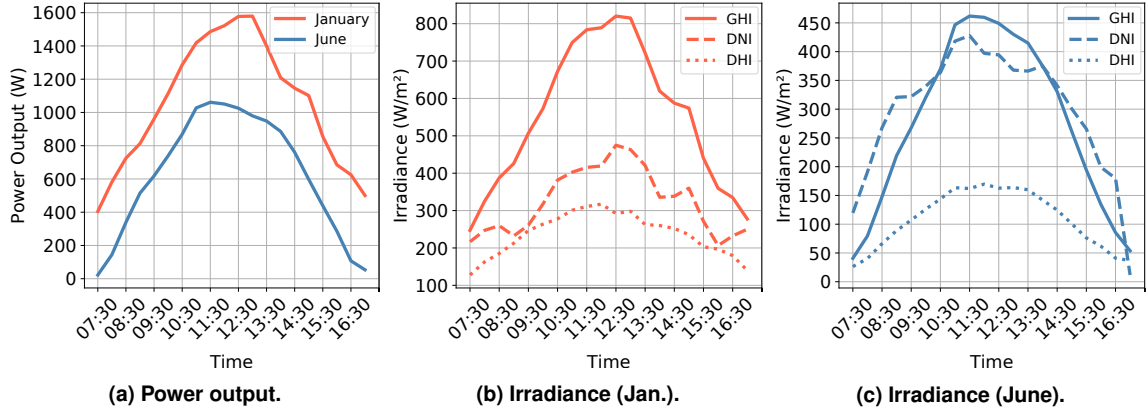


Figure 1. Daily average profile curves for January and June.

We consider five different Feature Sets (FS) in our evaluation, namely: FS 1 (GHI, DNI, DHI, UV, RH, AT and WS), FS 2 (GHI, DNI, DHI, UV and RH), FS 3 (UV, RH, AT and WS), FS 4 (GHI, DNI and DHI) and FS 5 (GHI). The first set is composed by all the measurements available from the solarimetric station. The second set is composed by all the irradiance components alongside the two features with highest correlation with Power Output. The third and fourth Feature Sets are built considering solely climatic and irradiance information, respectively. The fifth set, which we adopt as a baseline, is based solely on GHI data, which has the highest overall correlation with Power Output.

3.2. Methods

In this section we review the regression methods employed in the study. All methods, except for MLP, were trained and evaluated using scikit-learn [Pedregosa et al. 2011].

MLP is a well-known feed-forward neural network [Haykin 1998]. It has the ability to map non-linear relationships between input features and target. We employ MLPs with a single hidden layer and different activation functions for the neurons on this layer. Specifically, we considered $\{3, 5, 7, 9, 11\}$ neurons in the hidden layer and Rectified Linear Unit (ReLU) and Hyperbolic Tangent (tanh) as activation functions. Thus, we obtained 10 configurations of MLPs for evaluation. Models were trained with the Adam optimizer [Kingma and Ba 2014] with default parameters on Keras [Chollet et al. 2015].

The Support Vector Regression (SVR) method [Vapnik 1998] is a statistical Machine Learning technique, which can be used to solve regression problems. We considered different settings for the parameters C and $kernel$. The former is the regularization parameter, for which we considered the values $\{1, 10, 50, 100, 1000\}$. For the latter, we used the Radial Basis Function, Polynomial, and Linear kernels. Both parameters have substantial impact on the model's accuracy. For ϵ we considered its default value (0.01). In total, 15 SVR configurations were evaluated.

k-Nearest Neighbors (kNN) is one of the simplest Machine Learning methods. We used kNN with Euclidean distance, considered features normalized to have mean zero and standard deviation one (z-score normalization). We considered different values of k , i.e., the number of neighbors: $\{3, 5, 7, 9, 11\}$, thus resulting in five different configurations for evaluation. In order to obtain predictions, we consider the unweighted arithmetic average of the neighbors target value.

Decision Trees (DTs) can be applied for both regression and classification problems [Breiman et al. 1984]. In our experiments we adopted the Mean Squared Error (MSE) as splitting criterion. Regarding its parameters, we assessed for the minimum number of samples in a leaf node, for which we considered the values 5 and 10. This parameter indirectly controls the depth of the final tree.

Linear Regression (LR) is the most conventional regression algorithm and we adopt it here as a baseline for comparison. We employ LR in its most basic form, i.e., without regularization or basis expansion. Since our dataset is composed of a small number of features most of the common problems encountered with vanilla linear regression are avoided.

3.3. Methodology

In this work we take into account short term forecasting, more specifically, we are interested in next day forecasting, which is a common goal in the literature [Das et al. 2018, Raza et al. 2016]. In order to estimate the generalization capability of each model, we adopt a non-anchored walk-forward validation procedure [Kirkpatrick and Dahlquist 2006], that is, we perform training on a fixed size window, then predict power output of the next day. The fixed size window then moves one day ahead and the procedure is repeated until the end of the year (data) is reached. With this approach models are built solely on past observations, which is a realistic scenario. Moreover, given that the training window is not anchored (i.e., it has always the same size) models can adapt and reflect seasonal changes. In our experiments, we consider roughly 30 days for training (a fixed size window of 600 timestamps — 30 days \times 20 timestamps per day) then predicted the power output of the following day which, in turn, is comprised of another 20 timestamps.

To assess the quality of the methods we used two well known regression error metrics. Specifically, the Root Mean Square Error (RMSE) — Eq. (1) — and the Coefficient of Determination also known as R^2 — Eq. (2). In the aforementioned equations y_i accounts for the observed value of power production (power output) of the i -th observation, \hat{y}_i is the predicted power production of the i -th observation, and \bar{y} is the average of power production considering all observations.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1) \quad R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

4. Results and Discussion

As we discussed in the previous section, different parameter configurations were considered for each one of the methods under evaluation. In total 33 different configurations (10 MLPs, 15 SVRs, 5 kNNs, 2 DT, and 1 LR) were evaluated within each one of the 5 Feature Sets (FSs). Given the number of configurations under evaluation we report the best results of each method in a particular Feature Set (according to its overall RMSE). These results are presented in Table 2.

In the case of MLP, across all FSs, the best results were observed with 9 neurons in the hidden layer and the tanh activation function. For SVR, in terms of kernel there was

no consensus among FSs, whereas for the regularization parameter, $C = 1000$, led to the best results in all cases. Regarding kNN, the best results were achieved with k equal 7, 5, 9, 5, 11 for FSs 1 through 5, in that order. Finally, for DTs the best results were achieved with minimum number of samples equal to 5 in FSs 1, 2, and 4 and 10 in the remaining.

Table 2. Best results obtained for each model within Feature Sets.

Model	FS 1		FS 2		FS 3		FS 4		FS 5	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
MLP	45.92	0.99	49.43	0.99	107.68	0.96	53.19	0.99	66.17	0.99
SVR	52.20	0.99	52.63	0.99	112.69	0.96	56.78	0.99	88.61	0.98
kNN	94.95	0.97	76.65	0.98	155.42	0.92	71.68	0.98	87.33	0.98
DT	65.47	0.99	64.81	0.99	122.47	0.95	63.31	0.99	89.93	0.98
LR	59.03	0.99	60.42	0.99	112.35	0.96	63.33	0.99	87.82	0.98

We highlight in bold the best overall results in terms of RMSE (which allows for a better discrimination of the methods than R^2) within each one of the FSs. In general, the best results were obtained with MLP, followed by SVR. These results are in accordance with Das et al. [Das et al. 2018]. Indeed, the best overall result was achieved with MLP, when all features are taken into account (FS 1). In most of the cases, LR showed better results than both kNN and DT. Regarding these two particular models, the former has the worst results across most of the FSs.

In terms of Feature Sets, specifically, one can observe that the worst results were obtained with FS 3, which is composed by UV, RH, AT, and WS. Although UV is clearly correlated with output, this is the only Feature Set that does not include any irradiance feature in its composition (i.e., GHI, DHI, and DNI). Given that solar irradiance is one of the main factors in photovoltaic generation, this seems to be a plausible/possible cause for such a behavior. Indeed, for all the remaining FSs the errors are quite lower than for FS 3. When considering the results obtained with MLP, SVR, and LR, the best results were obtained with FS 1. Although FS 5 is composed solely by GHI (Global Horizontal Irradiance), the highest correlated feature with output, improvements are observed when additional features are taken into account in other Feature Sets.

In Figure 2 we depict individual errors obtained for each one of the best methods within each Feature Set (these are the individual errors — per timestamp — summarized in Table 2). One can observe that the final RMSE value is affected by outliers in all cases. A careful analysis of the outliers revealed that some of them occur within days related with reading problems (days that the datalogger went stuck at a particular value). Although the spurious stuck at readings were removed (as discussed in Section 3.1), we believe that a few adjacent readings may also be inaccurate, thus leading to these particular errors.

Overall, kNN was the method that produced the largest number of outliers. In terms of variance, the best results were obtained with MLP and SVR in all Feature Sets, but FS 3. Linear regression (LR) and Decision Tree (DT) produced fewer outliers than the remaining methods, however, their 75th percentile was greater than those observed with MLP and SVR in all cases.

In the sequel, we analyze the behavior of the different methods in more detail,

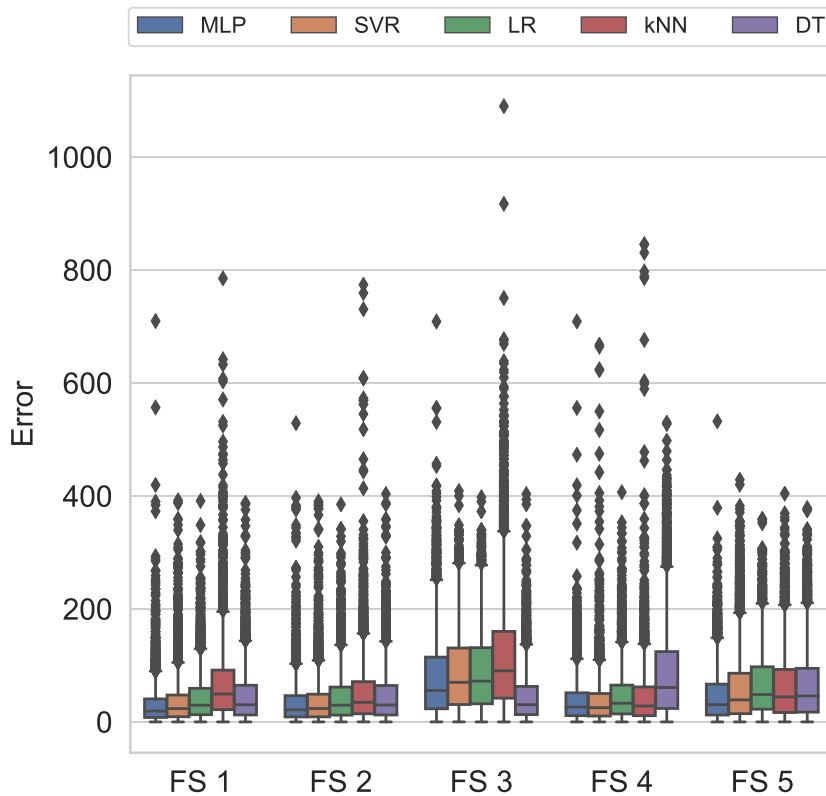


Figure 2. Boxplots depicting the individual errors for each method and FS.

considering the best Feature Set in general, FS 1. First, we select the predictions from four months: April, July, August, and December. For each month, we present expected values and predictions for selected days in Figure 3. One can note that the lowest values of Power Output occur at morning and evening. In December these values are higher than April, given that in the Summer the sun rises earlier than in the Fall.

For April and December, all models obtained predictions close to the expected values. For July and August the same does not hold, as most models had difficulties predicting the power generated around noon. Considering July, LR was able to predict correct values at this time, but suffered from bad predictions in the remaining of the afternoon. These considerably bad predictions can be explained by the presence of clouds and rain, which affect the measurements obtained by some of the sensors. In August, we observe a similar phenomenon. The quantity of power generated around noon was the hardest period to predict. As in the case of July, kNN was the method that had the worst prediction errors at this period. This is frequent for kNN, the worst method overall.

To better understand the difficulties of all algorithms in predicting the power output around noon on July 11 (upper right plot in Figure 3), we present in Figure 4 the values of each feature at that day (for all timestamps). One can note that, close to 11:30, DNI values were zeroed due to a sensor measurement problem, which is only reestablished after 15:30. Thus, all models that have their predictions influenced by DNI had worst performance at this period, with kNN and SVR being the most affected ones. Such an example shows the importance of robust methods for power output forecasting in PV systems.

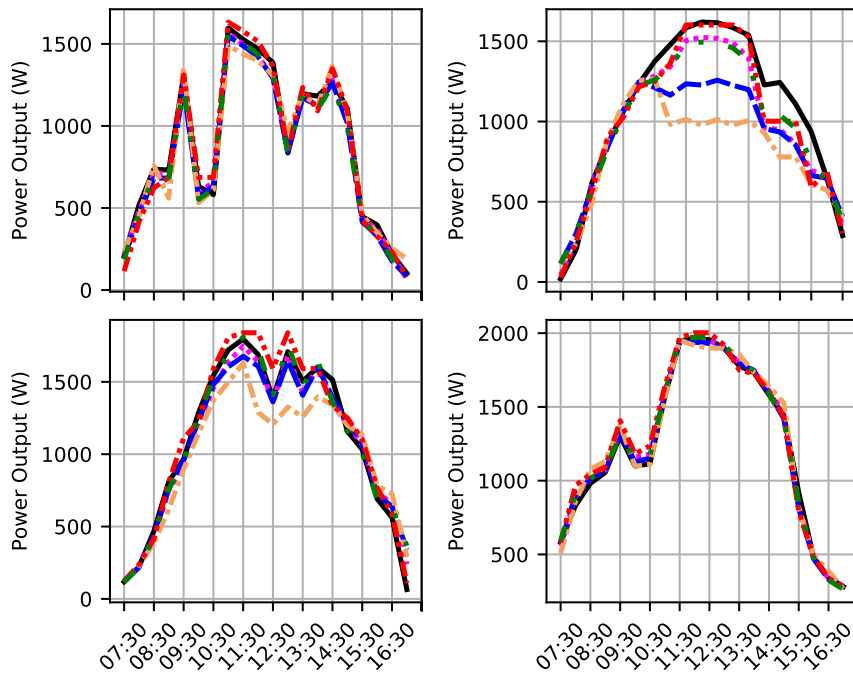


Figure 3. Expected and predicted values for power output on different days. Clockwise from top left corner: 18-04-27, 18-07-11, 18-12-12, and 18-08-28.

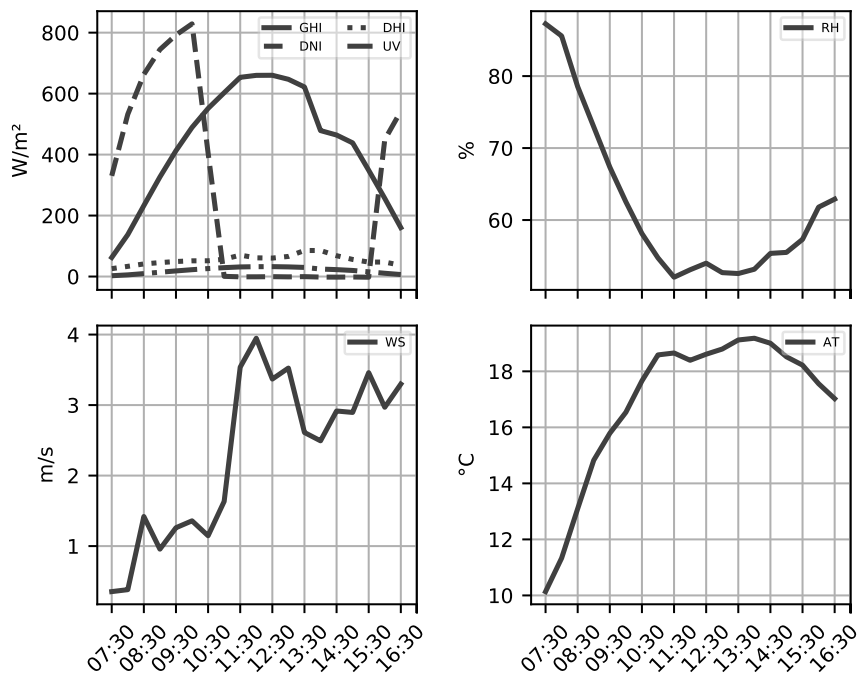


Figure 4. Feature values (FS 1) for all timestamps of 18-07-11.

5. Conclusions

We presented a comparison of different regression methods for predicting the power output of a Photovoltaic plant, considering only climatic and meteorological features. All features were obtained by a solarimetric station located at the same site of the PV plant. Given that we considered raw collected data, in a first step we performed pre-processing to remove records with clear spurious measurements due to problems with the datalogger. Our analysis considered daily data with aggregated in 30 minute intervals. Models were built considering windows of past data spanning through 30 days and evaluated for next day prediction.

Regarding methods, we assessed MLP, SVR, kNN, DT, and LR. The most important parameters of each method were tuned, and the best result for each one was selected for further evaluation. Our empirical analysis suggests that MLP and SVR are the best methods for this prediction task. Despite the significant correlation observed between a couple of features and Power Output, considerable improvements were obtained by MLP and SVR over LR. Meanwhile, kNN and DT suffered most with the noisy nature of the data from the sensors.

Considering the different feature sets, even though Global Horizontal Irradiance (GHI) is highly correlated with power output, the use of additional features demonstrated its value. Specifically, by using MLP with all features we obtained a reduction of 30% on RMSE. This result indicates that a potential future line of work is the aggregation of more meteorological/climate data to improve prediction. We also intend to employ/investigate different models according to particular weather conditions (clean, rainy, or cloudy) and seasons of the year.

Acknowledgments

This work was partially funded by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). The authors would like to thank Fotovoltaica/UFSC Solar Energy Research Laboratory for providing the dataset employed in this study.

References

- Arize, D. and Nogueira Rios, T. (2019). A comparison study on time series forecasting given smart grid load uncertainties. *Proceedings - 2019 Brazilian Conference on Intelligent Systems, BRACIS 2019*, pages 257–262.
- Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.
- Chollet, F. et al. (2015). Keras. <https://keras.io>.
- Chow, S. K., Lee, E. W., and Li, D. H. (2012). Short-term prediction of photovoltaic energy generation by intelligent approach. *Energy and Buildings*, 55:660–667.
- Das, U. K., Tey, K. S., Seyedmahmoudian, M., Mekhilef, S., Idris, M. Y. I., Van Deventer, W., Horan, B., and Stojcevski, A. (2018). Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, 81(April 2017):912–928.

- De Giorgi, M. G., Congedo, P. M., and Malvoni, M. (2014). Photovoltaic power forecasting using statistical methods: Impact of weather data. *IET Science, Measurement and Technology*, 8(3):90–97.
- Ding, M., Wang, L., and Bi, R. (2011). An ANN-based approach for forecasting the power output of photovoltaic system. *Procedia Environmental Sciences*, 11(PART C):1308–1315.
- Fernandez-Jimenez, L. A., Muñoz-Jimenez, A., Falces, A., Mendoza-Villena, M., Garcia-Garrido, E., Lara-Santillan, P. M., Zorzano-Alba, E., and Zorzano-Santamaria, P. J. (2012). Short-term power forecasting system for photovoltaic plants. *Renewable Energy*, 44:311–317.
- Gielen, D., Boshell, F., Saygin, D., Bazilian, M. D., Wagner, N., and Gorini, R. (2019). The role of renewable energy in the global energy transformation. *Energy Strategy Reviews*, 24(January):38–50.
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, USA, 2nd edition.
- Huang, Y., Lu, J., Liu, C., Xu, X., Wang, W., and Zhou, X. (2010). Comparative study of power forecasting methods for PV stations. *2010 International Conference on Power System Technology: Technological Innovations Making Power Grid Smarter, POWERCON2010*, pages 1–6.
- IEA (2017). International energy agency, data and statistics 2017. <https://www.iea.org/data-and-statistics/data-tables/?country=WORLD&energy=Electricity&year=2017>. Accessed: 2020-03-17.
- IEA (2019). International energy agency, world energy outlook 2019. <https://www.iea.org/reports/world-energy-outlook-2019/electricity>. Accessed: 2020-03-17.
- Inman, R. H., Pedro, H. T., and Coimbra, C. F. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39(6):535–576.
- Kaplani, E. and Kaplanis, S. (2014). Thermal modelling and experimental assessment of the dependence of PV module temperature on wind velocity and direction, module orientation and inclination. *Solar Energy*, 107:443–460.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization.
- Kirkpatrick, C. and Dahlquist, J. (2006). *Technical Analysis: The Complete Resource for Financial Market Technicians*. FT Press, first edition.
- Kudo, M., Takeuchi, A., Nozaki, Y., Endo, H., and Jiro, S. (2009). Forecasting electric power generation in a photovoltaic power system for an energy network. *Electrical Engineering in Japan (English translation of Denki Gakkai Ronbunshi)*, 167(4):16–23.
- Leva, S., Dolara, A., Grimaccia, F., Mussetta, M., and Ogliari, E. (2017). Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. *Mathematics and Computers in Simulation*, 131:88–100.

- Lo Brano, V., Ciulla, G., and Di Falco, M. (2014). Artificial neural networks to predict the power output of a PV panel. *International Journal of Photoenergy*, 2014.
- Lorenz, E., Hurka, J., Heinemann, D., and Beyer, H. G. (2009). Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2(1):2–10.
- Mahtta, R., Joshi, P. K., and Jindal, A. K. (2014). Solar power potential mapping in India using remote sensing inputs and environmental parameters. *Renewable Energy*, 71:255–262.
- Malvoni, M., De Giorgi, M. G., and Congedo, P. M. (2017). Forecasting of PV Power Generation using weather input data-preprocessing techniques. *Energy Procedia*, 126:651–658.
- Mellit, A., Massi Pavan, A., and Lughi, V. (2014). Short-term forecasting of power production in a large-scale photovoltaic plant. *Solar Energy*, 105:401–413.
- Muhammad Ehsan, R., Simon, S. P., and Venkateswaran, P. R. (2017). Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron. *Neural Computing and Applications*, 28(12):3981–3992.
- Nonnenmacher, L., Kaur, A., and Coimbra, C. F. (2014). Verification of the SUNY direct normal irradiance model with ground measurements. *Solar Energy*, 99:246–258.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pedro, H. T. C. and Coimbra, C. F. M. (2012). Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86(7):2017–2028.
- Raza, M. Q., Nadarajah, M., and Ekanayake, C. (2016). On recent advances in PV output power forecast. *Solar Energy*, 136:125–144.
- Souza-Echer, M. P., Pereira, E. B., Bins, L. S., and Andrade, M. A. (2006). A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera. *Journal of Atmospheric and Oceanic Technology*, 23(3):437–447.
- Statista (2019). Statista - global business platform, world energy outlook 2019. <https://www.statista.com/outlook/256/115/household-appliances/brazil>. Accessed: 2020-03-17.
- Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley-Interscience.