Lightweight Photovoltaic Forecasting Method for Agricultural Microgrids

Paul D. Brown[1], Murat Göl[1]

[1] Middle East Technical University, Ankara, Turkey
paul.brown@metu.edu.tr, mgol@metu.edu.tr

Abstract

A lightweight forecasting method was developed for day-ahead (12 to 72-hours) and intra-day forecasting of photovoltaic (PV) array power output for use with energy management of agricultural microgrids. The proposed method allows the PV output to be forecast without requiring extensive computing resources or high-bandwidth communication channel to the PV site. The proposed day-ahead forecasting method combines historical PV output data with historical weather data to infer the relationship between weather and PV output by fitting a simple regression model. Based on the assumption that updated weather forecasts are in general not obtained throughout the day, the current-day PV output forecast was updated using time-series techniques. The proposed forecasting methods were applied to a dataset of rooftop PV output and compared to reference persistence forecasts. The method was demonstrated by using the forecasts in the optimized operation of an agricultural microgrid and simulating with actual PV output.

1. Introduction

Over recent decades, many factors have driven significant changes in the electrical power industry, including climate change, the growing use of renewable energy resources, and ongoing rural electrification. For agricultural consumers, electrification enables increased yields and economic development. Ref. [1] reviews technology and applications for off-grid systems for rural electrification.

For agricultural applications, photovoltaic (PV) energy is generally well aligned with water needs [2]. Ref. [2] reviews numerous solar-powered water pumping applications around the world while [3] reviews the PV and water pumping modeling, design, and control approaches in the literature.

One of the barriers to efficient use of PV energy in the grid and in micro-grids is the difficulty of accurately predicting future energy availability. As reviewed in [4], many techniques have been developed for forecasting PV energy over various time horizons. The authors of this paper have been working on techniques for optimal energy planning for agricultural microgrids. For operational optimization, an effective day-ahead PV forecasting method was needed. In recent years, LSTM-based neural networks have emerged as the state-of-the art for day-ahead PV energy forecasting [5–7].

As promising as deep learning networks are for day-ahead forecasting, they have the drawback of requiring a lot of historical data [8] and significant computational resources for training. Agricultural microgrids, on the other hand, are often implemented in locations where communication infrastructure is weak and bandwidth and/or availability are limited. To cope with these conditions, the authors developed methods for day-ahead PV forecasting with a design goal to be "lightweight", that is to require minimal weather forecast data, site-specific modeling data, historical generation data, and computational resources. The target computing platform for the developed method is a single-board computer such as the Raspberry Pi or similar.

2. Proposed Method

The proposed method utilizes meteorological forecast data such as forecast cloudiness or forecast irradiance to forecast the PV system output for the upcoming 12 to 72 hours, which called the "day-ahead forecast". It is assumed that these meteorological forecasts will be updated once or at most a few times per day. Throughout the day, the current day's PV output forecast is updated by applying one of several available time-series forecasting techniques. A high-level diagram of the proposed forecast method is shown in Fig. 1.

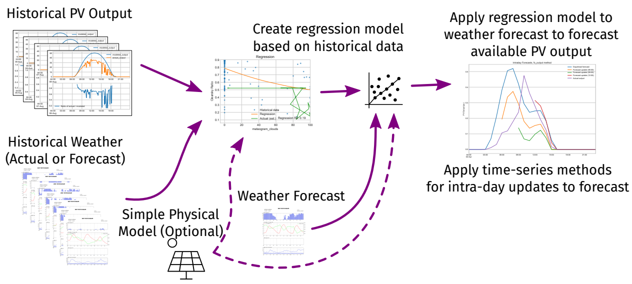

Fig. 1. Flowchart of Forecasting Method

2.1. Day-Ahead Forecasts

The proposed day-ahead forecasting method combines historical PV output data with historical weather data to infer the relationship between weather and PV output by fitting a simple regression model. The weather data used is either an estimate of solar irradiance or cloudiness. Other weather factors such as temperature are known to influence PV output values, as well as non-weather factors such as PV aging, dust, and shading. Because the proposed method uses relatively recent historical PV output data, the longer-term changes related to aging, seasonal temperatures, and dustiness are implicitly tracked as the historical data window moves forward in time. On the other hand, changes in these factors that are relatively fast compared to the historical time window used will not be accounted for. Shading may be incorporated to some extent if hourly maximum values from recent history are used to calculate the clear-sky output in cloudiness-based forecasts, however, the available PV dataset did not include any shading of the array, so the effectiveness of this method is addressing shading is not demonstrated.

Some pre-processing of the historical data was done prior to fitting the regression model. Data from periods of darkness, when output goes to small or negative values, were removed. This also removed periods when the PV array was covered with snow.

How the regression model was applied depends on the weather data that was used. If the weather data used represented the incoming solar energy, then the regression was done directly between the weather data and PV output. For this work, global horizontal irradiance (ghi) was modeled in this way. A linear model of the form shown in (1) was fit to the historical data, where P_{PV} is the actual or forecast PV output power in per-unit of the array's nominal output rating, k_1 is a linear constant to be fit, and *GHI* is actual or forecast global horizontal irradiance.

$$P_{PV} = k_1 \cdot GHI \tag{1}$$

On the other hand, weather data, such as cloudiness, that represented the attenuation of incoming solar energy was regressed against the clear-sky ratio for PV output. Two methods were developed for calculating the clear-sky ratio. The first method was to roughly model the PV array using data for its geographic position and orientation to obtain modeled clear-sky output. The second method was to estimate the maximum possible output at each time of day by calculating the maximum observed PV output for each time of day in the historical data window. This second method had the advantage of not requiring any modeling information about the PV array. It had the further advantage of naturally incorporating effects of shading of the array during parts of the day due to obstructions. Since the data set used for this analysis did not include any shading of the array, this last benefit could not be assessed in comparison to other methods.

The formula used to calculate the clear-sky ratio is shown in (2), where r_{cs} is the clear-sky ratio, P_{out} is the PV output, and P_{cs} is the clear-sky PV output. In order to provide numerical stability of the clear-sky ratio for periods when the clear-sky output is small, the parameters ϵ and r_ϵ were incorporated so that as P_{cs} goes to zero, the clear-sky ratio r_{cs} will approach r_ϵ.

$$r_{cs} = \frac{P_{out} + r_\epsilon \epsilon}{P_{cs} + \epsilon} \tag{2}$$

The regression model was then fit to the transformed data. A second-degree polynomial model of the form shown in (3) was used to model the relationship between cloudiness and the clear-sky ratio, where r_{cs} is the actual or forecast clear-sky ratio, k_2, k_1, and k_0 are coefficients to be fit, and ζ is actual or forecast cloudiness in a range from 0 (no clouds) to 100 (completely cloudy).

$$r_{cs} = k_2\zeta^2 + k_1\zeta + k_0 \tag{3}$$

The fit model was then applied to the weather forecast data available for the forecast period. If applicable, the resulting forecast was then converted from a forecast clear-sky ratio back to PV output using (2).

2.2. Intra-day Updates

Based on the assumption that updated weather forecasts are in general not obtained throughout the day, the current-day PV output forecast was updated using time-series techniques. The AR and SARIMAX time-series models are described in time-series modeling/forecasting references [9, 10]. Several approaches to intra-day updates were investigated as described in the following paragraphs.

AR(2) on actual output with exogenous variable: An auto-regressive time series model with two lags was applied to the actual clear-sky output ratio with the day-ahead clear-sky output forecast included in the model as an exogenous variable.

SARIMAX: A seasonal auto-regressive integrated moving average model was applied to the actual PV power output with day-ahead output power forecast included in the model as an exogenous variable. This model is characterized by parameters for the number of auto-regressive (AR) lags p, order of differencing (I) d, and order of the moving average (MA) q, for both the trend and seasonal components. A variety of parameter combinations was tested against the data, and the best performing model was with AR $p = 1$ and MA $q = 2$, with the day-ahead PV output power forecast as an exogenous variable, and with no differencing or seasonal components included.

AR(2) on residual of PV output: An auto-regressive time series model with two lags was applied to the residual between actual PV output and the day-ahead forecast PV output. The forecast output of the model was then added as a correction term to the day-ahead PV output forecast for the intra-day period.

Scaling: The day-ahead forecast for the rest of the day was scaled using the ratio between the actual output in the previous period and the day-ahead forecast for the previous period.

3. Data Sources

The PV time series used for this analysis was the recorded output power of the rooftop PV array on the METU EEE Department machinery building in Ankara, Turkey. Recordings were obtained by downloading them from the logger integrated with the inverter for the PV array. Logged data was recorded at approximately five-minute intervals. For the forecasting method, the logged data was aligned to exact five-minute intervals even with the hour using linear interpolation and then aggregated to hourly intervals by the mean. For the operational model, these values were scaled to the rating of the demonstration system.

The weather forecast data used for this analysis was obtained from multiple sources. Firstly, WRF Meteogram forecasts have been saved from the Turkish Meteorology Directorate (MGM) [11]. Since the MGM meteograms are available as a image rather than as structured data, code was developed to extract data from the plots in the meteogram and output it in csv format [12]. Secondly, forecasts were obtained from the SolCast PV-focused weather service [13]. SolCast weather forecasts include both a cloudiness forecast as well as an irradiance (ghi) forecast along with many other forecast quantities. SolCast data is provided using an API, so no special data extraction or conversion was needed.

4. Implementation

The forecasting code was implemented using the Python programming language [14]. The Pandas [15] data analysis library and NumPy [16] were used for data manipulation. Scikit-learn [17] was used for fitting the regression model and generating predictions from it. Statsmodels [18] was used for fitting time-series auto-regressive models and generating predictions. Plots were generated using Matplotlib [19]. The optimization and simulation implementation was previously described in [20]. Full implementation details can be obtained by examining the source code released on GitHub [21] and CodeOcean [22].

Table 1. Day-ahead forecast metrics

	RMSE	MAE	MBE
Persistence	0.108	0.045	–0.0001
SolCast Cloudiness	0.079	0.037	–0.0057
SolCast GHI	0.080	0.036	–0.0051
MGM Meteogram	0.089	0.043	–0.0076

5. Results

In order to test and compare the methods presented in Section 2, the methods were applied to real PV time series data, described in detail in Section 3. The results shown in this section are for the PV time series covering 926 days from 31 March, 2021 to 30 September, 2023.

5.1. Day-ahead forecasts

Using data available prior to sunrise, the proposed day-ahead forecast method from Section 2.1 was applied to each of the weather forecast sources (MGM meteogram cloudiness, SolCast cloudiness, and SolCast ghi) to forecast the upcoming one-day (24 hr) PV output. Following findings from [23], 28 days of historical data was used to fit the regression models. An example day-ahead forecast using the Solcast GHI forecast is shown in Fig. 2.

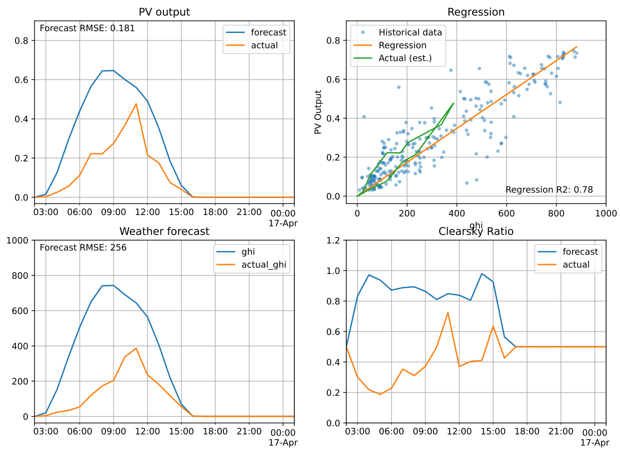

Fig. 2. Example Day-ahead Forecast using Solcast GHI Forecast

In addition to the proposed method, a reference day-ahead persistence forecast was also generated. The persistence forecast was that the PV output for the upcoming period will be the same as the PV output for the same time of day in the previous day.

Following [24] and [25], the methods were evaluated using root mean square error (RMSE), mean absolute error (MAE), and mean bias error (MBE) as the error measure. These metrics were calculated for of the whole time series as well as the for the forecast total energy output for each day. The resulting error metrics are shown in Table 1 and Table 2 respectively.

Fig. 3 shows the RMSE of the day-ahead forecasts when broken down by month of the year. The generally predictable climate pattern of clear summer skies is evident in the higher accuracy of all forecasting methods in the summer months of July and August.

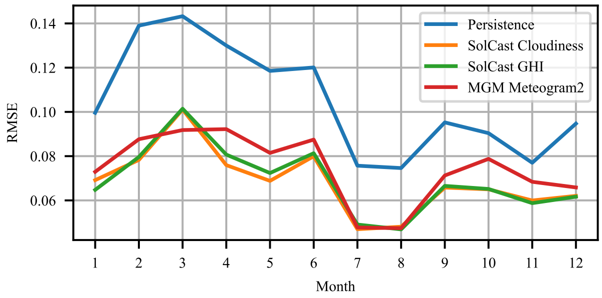

Fig. 3. Day-ahead Forecast RMSE by Month

Table 2. Day-ahead forecast metrics on daily sums

	RMSE	MAE	MBE
Persistence	1.262	0.874	–0.001
SolCast Cloudiness	0.917	0.570	–0.137
SolCast GHI	0.911	0.577	–0.122
MGM Meteogram	0.934	0.558	–0.137

5.2. Intra-day Updates

To evaluate and compare the intra-day forecast update methods, as a starting point, a day-ahead forecast was first generated using the SolCast cloudiness forecast. This day-ahead forecast was then updated for each hour of the day using each of the methods proposed in Section 2.2. In addition to the proposed methods, a reference intra-day update forecast based on persistence was produced. The persistence forecast was that the clear-sky ratio for the remainder of the day would be the same as the clear-sky ratio from the previous time period.

For comparison, the RMSE, MAE, and MBE for the different intra-day forecast update methods were computed for the forecast for the periods 0 to 1, 1 to 2, and 6 to 7 hours after of the time at which the forecast is generated. The resulting error metrics are shown in Table 3. Note that the intra-day forecast metrics exclude nighttime hours since intra-day updates are not applied to them.

Comparison of the error metrics for the various intra-day update methods to the error metrics for the original dayahead forecasts shows that the intra-day updates are most beneficial to the forecast accuracy in the upcoming few hours. When considering six hours ahead, the weather-forecast-based day-ahead forecast tends to perform best without any proposed update methods.

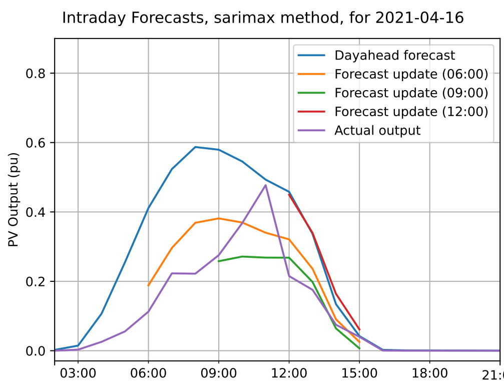

Fig. 4. Intra-day Forecast Updates - SARIMAX method

Table 3. Intraday forecast metrics

Hours Ahead	Method	RMSE	MAE	MBE
0 - 1	No update	0.1260	0.0949	–0.0064
	Persistence	0.0871	0.0575	–0.0034
	fx_output[1]	0.0876	0.0622	–0.0039
	exog[2]	0.0909	0.0682	0.0216
	sarimax	0.0901	0.0646	0.0021
	scaling	0.0917	0.0618	–0.0008
1 - 2	No update	0.1230	0.0904	–0.0042
	Persistence	0.1101	0.0723	–0.0064
	fx_output[1]	0.1059	0.0767	–0.0031
	exog[2]	0.1139	0.0882	0.0382
	sarimax	0.1090	0.0802	0.0070
	scaling	0.1127	0.0761	–0.0022
6 - 7	No update	0.0675	0.0332	–0.0070
	Persistence	0.0802	0.0379	–0.0121
	fx_output[1]	0.0685	0.0347	–0.0075
	exog[2]	0.0756	0.0387	0.0162
	sarimax	0.0698	0.0357	–0.0024
	scaling	0.0781	0.0367	–0.0075

[1] AR(2) on residual of PV output.
[2] AR(2) on actual output with exogenous variable

5.3. Optimal Energy Dispatch

To demonstrate the effectiveness of the developed methods in the target application, the forecasts were input into an agricultural microgrid operational optimization and simulation program. A system diagram of the demonstration system is shown in Fig. 5. The objective function for optimization is shown in (4), where $P_{grid,t}$ is the power drawn from the grid in each time period, $P_{BSS,ch,t}$ and $P_{BSS,disch,t}$ are the battery system (BSS) charging and discharging powers, $V_{use,desired,d}$ and $V_{use,d}$ are the desired and actual effective daily water use volumes, $s_{BSS,t}$ is a binary variable representing the BSS charging mode, and $s_{pump1,t}$ is a binary variable representing Pump 1 operating state. $C_{grid,t}$, C_{BSS}, $C_{w,short}$, $C_{BSS,switching}$, and $C_{Pump1,switching}$ are the respective cost coefficients. The optimization formulation was previously presented in full in [20].

$$\begin{aligned}\min\ & \underbrace{\sum_t C_{grid,t}\, P_{grid,t}\, \Delta t}_{\text{Cost of grid power}} \\ &+ \underbrace{\sum_t C_{BSS}\left(P_{BSS,ch,t} + P_{BSS,disch,t}\right)\Delta t}_{\text{Cost of battery usage}} \\ &+ \underbrace{C_{w,short}\sum_d \max\left(\left(V_{use,desired,d} - V_{use,d}\right),0\right)}_{\text{Penalty for inadequate water}} \\ &+ \underbrace{C_{BSS,switching}\sum_t \left|s_{BSS,t} - s_{BSS,t-1}\right|}_{\text{BSS mode-switching penalty}} \\ &+ \underbrace{C_{Pump1,switching}\sum_t \left|s_{pump1,t} - s_{pump1,t-1}\right|}_{\text{Pump 1 switching penalty}}\end{aligned} \tag{4}$$

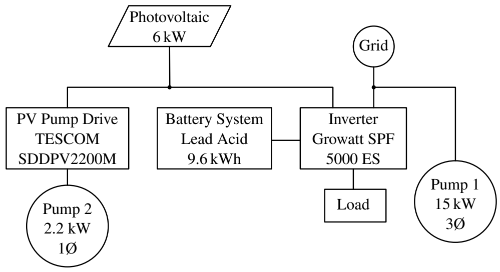

Fig. 5. Energy Dispatch Demonstration System

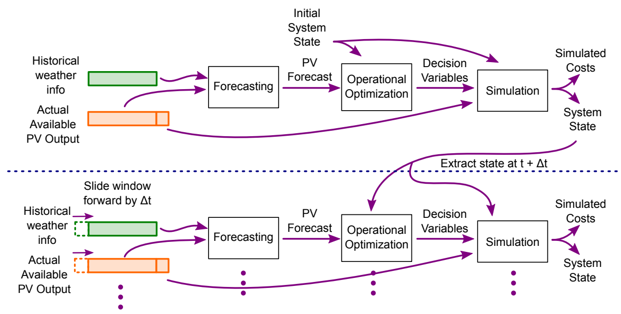

Fig. 6. MPC demonstration simulation flowchart

The use of model-predictive control (MPC) was simulated by using the optimizer to choose the optimal ratio of using excess available PV energy between energy storage in a battery energy storage system (BSS) and utilization to pump water into a reservoir. The MPC optimization was performed using the forecast PV data, including intra-day updates. The operation for the following period was then simulated using actual PV data. A flowchart of the MPC simulation is shown in Fig. 6.

The results shown in this section are for simulated operation for 64 days from 27 April to 29 June, 2021. Table 4 shows results of the MPC simulation. Simulations were performed for three forecasts:

- Persistence forecast
- SolCast cloudiness weather forecasts for day-ahead forecast with SARIMAX-based intraday updates
- Perfect "oracle" forecast where operation is optimized with the actual future PV output data

The costs for the system's simulated operation using proposed forecast method was approximately 3% lower than the performance using persistence forecasts.

The objective function value for optimized operation with a perfect forecast illustrates the value of improved forecasts. An additional 8% reduction in the objective function value would be possible with perfect forecasts.

6. Conclusion

This paper demonstrated a novel lightweight PV power forecasting method suitable for application to small agricultural microgrids. The method utilized only basic weather forecast data and did not require large computing resources. The proposed methods were compared to each other and to a reference persistence forecast. The proposed methods gave lower RMSE compared to the persistence forecast over a test data set of 926 days. Of the weather forecast sources that were compared, SolCast cloudiness obtained the lowest RMSE. Of the intra-day update methods that were compared,

Table 4. MPC Simulation Results - Objective Function Value

Cost component	Persistence Forecast ($, 2020)	Proposed Method ($, 2020)	Perfect Forecast ($, 2020)
Grid energy	65 350	63 466	58 052
Battery use	8117	7964	8047
Inadequate water	865	0	10
Battery mode switching	246	228	230
Pump switching	94	90	80
TOTAL	74 672	71 749	66 418

the SARIMAX method obtained the lowest RMSE. The forecast methods with the best RMSE metrics were integrated into a simulated agricultural microgrid with model-predictive control, and the total system costs were compared to operation with persistence forecasts. The costs for the system's simulated operation using proposed forecast method was approximately 4% lower than the performance using persistence forecasts.

One direction for future work is to verify the applicability of the method to other geographic regions and different season conditions. Another direction could be to extend the proposed lightweight method for PV power forecasting to provide probabilistic forecasts, quantifying the uncertainty in the forecast. Such forecasts may then be used in stochastic model predictive control for the microgrid controls to make more robust energy planning decisions.

7. Acknowledgments

This work is supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under grant number 119N313.

8. References

[1] S. Mandelli, J. Barbieri, R. Mereu, and E. Colombo, "Off-grid systems for rural electrification in developing countries: Definitions, classification and a comprehensive literature review," *Renewable and Sustainable Energy Reviews*, vol. 58, pp. 1621–1646, 5 2016.

[2] M. Aliyu, G. Hassan, S. A. Said, M. U. Siddiqui, A. T. Alawami, and I. M. Elamin, "A review of solar-powered water pumping systems," *Renewable and Sustainable Energy Reviews*, vol. 87, pp. 61–76, 5 2018.

[3] D. H. Muhsen, T. Khatib, and F. Nagi, "A review of photovoltaic water pumping system designing methods, control strategies and field performance," *Renewable and Sustainable Energy Reviews*, vol. 68, pp. 70–86, 2 2017.

[4] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. M. de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, 2016.

[5] W.-C. Kuo, C.-H. Chen, S.-H. Hua, and C.-C. Wang, "Assessment of different deep learning methods of power generation forecasting for solar pv system," *Applied Sciences*, vol. 12, no. 15, 2022.

[6] M. Aslam, S.-J. Lee, S.-H. Khang, and S. Hong, "Two-stage attention over lstm with bayesian optimization for day-ahead solar power forecasting," *IEEE Access*, vol. 9, pp. 107 387–107 398, 2021.

[7] C.-H. Liu, J.-C. Gu, and M.-T. Yang, "A simplified lstm neural networks for one day-ahead solar power forecasting," *IEEE Access*, vol. 9, pp. 17 174–17 195, 2021.

[8] P. Aillaud *et al.*, "Day-ahead forecasting of regional photovoltaic production using deep learning," in *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)*. IEEE, 2020, pp. 2688–2691.

[9] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[10] J. Korstanje, *Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR*. Berkeley, CA: Apress, 2021.

[11] T.C. Tarım ve Orman Bakanlığı Meteoroloji Genel Müdürlüğü. Wrf meteogram. [Online]. Available: https://www.mgm.gov.tr/tahmin/wrf-meteogram.aspx

[12] P. D. Brown and E. Canaz. (2023) Software repository for meteogram data extraction program. [Online]. Available: https://github.com/pdb5627/meteogram_extract

[13] Solcast. (2019) Global solar irradiance data and pv system power output data. [Online]. Available: https://solcast.com/

[14] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[15] W. McKinney *et al.*, "Data structures for statistical computing in python," in *Proceedings of the 9th Python in Science Conference*, vol. 445. Austin, TX, 2010, pp. 51–56.

[16] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, p. 357–362, 2020.

[17] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[18] S. Seabold and J. Perktold, "statsmodels: Econometric and statistical modeling with python," in *9th Python in Science Conference*, 2010.

[19] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in science & engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[20] P. D. Brown and M. Göl, "Operational optimization of an agricultural microgrid," in *2022 57th International Universities Power Engineering Conference (UPEC)*, 2022, pp. 1–6.

[21] P. D. Brown and M. Göl. (2023) Software repository for lightweight photovoltaic forecasting method for agricultural microgrids. [Online]. Available: https://github.com/pdb5627/ELECO2023-PV-forecasting-paper

[22] P. D. Brown and M. Göl. (2023) Reproducible computing container for lightweight photovoltaic forecasting method for agricultural microgrids. [Online]. Available: https://doi.org/10.24433/CO.5756137.v2

[23] M. P. Almeida, O. Perpiñán, and L. Narvarte, "Pv power forecast using a nonparametric pv model," *Solar Energy*, vol. 115, pp. 354–368, 2015.

[24] H. T. Pedro and C. F. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Solar Energy*, vol. 86, no. 7, pp. 2017–2028, 2012.

[25] L. Gigoni *et al.*, "Day-ahead hourly forecasting of power generation from photovoltaic plants," *IEEE Transactions on Sustainable Energy*, vol. 9, pp. 831–842, 2018.