Lightweight Photovoltaic Forecasting Method for Agricultural Microgrids

Paul D. Brown
Department of Electrical and Electronics Engineering
Middle East Technical University
Ankara, Turkey
ORCID: 0000-0001-5753-0449

Murat Göl
Department of Electrical and Electronics Engineering
Middle East Technical University
Ankara, Turkey
ORCID: 0000-0002-2523-1169

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 compute 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.

Index Terms—Photovoltaic systems, forecasting, agriculture

I. 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.

II. 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.

A. 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.

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. Weather data utilized of this type was the global horizontal irradiance (ghi).

On the other hand, weather data 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
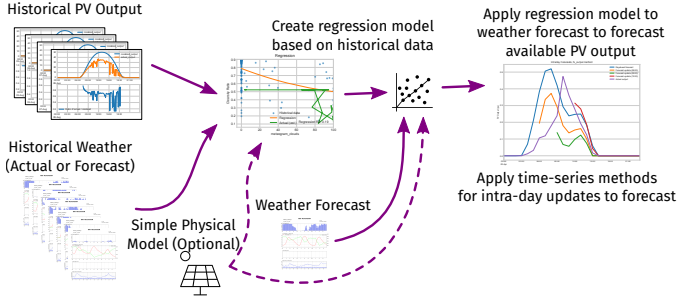
Fig. 1. Flowchart of Forecasting Method

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 (1), 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.

$$r_{cs} = \frac{P_{out} + r_\epsilon \epsilon}{P_{cs} + \epsilon} \qquad (1)$$

The regression model was then fit to the transformed data. A second-degree polynomial model was used in the case of ghi weather forecasts as well as the case of cloudiness weather forecasts. 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.

B. 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. Several approaches to intra-day updates were investigated as described in the following paragraphs.

a) 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.

b) 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 lags, order of differencing, and order of the moving average, 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(1) and MA(2), with the day-ahead PV output power forecast as an exogenous variable, and with no differencing or seasonal components included.

c) 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.

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

e) 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.

III. 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) [9]. 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 [10]. Secondly, forecasts were obtained from the SolCast PV-focused weather service [11]. 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.

IV. IMPLEMENTATION

The forecasting code was implemented using the Python programming language [12]. The Pandas [13] data analysis library and NumPy [14] were used for data manipulation. Scikit-learn [15] was used for fitting the regression model and generating predictions from it. Statsmodels [16] was used for fitting time-series auto-regressive models and generating predictions. Plots were generated using Matplotlib [17]. The

optimization and simulation implementation was previously described in [18]. Full implementation details can be obtained by examining the source code released on GitHub [19] and CodeOcean [20].

V. RESULTS

In order to test and compare the methods presented in Section II, the methods were applied to real PV time series data, described in detail in Section III. The results shown in this section are for the PV time series covering 103 days from 21 March to 29 June, 2021.

A. Day-ahead forecasts

Using data available prior to sunrise, the proposed day-ahead forecast method from Section II-A 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 [21], 28 days of historical data was used to fit the regression models. 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 [22] and [23], the methods were evaluated using mean square error (MSE), 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.

TABLE I
DAY-AHEAD FORECAST METRICS

	MSE	MAE	MBE
Persistence	0.015722	0.062034	0.000559
SolCast Cloudiness	0.006049	0.039775	-0.008772
SolCast GHI	0.006440	0.039048	-0.011762
MGM Meteogram	0.007343	0.046865	0.002728

TABLE II
DAY-AHEAD FORECAST METRICS ON DAILY SUMS

	MSE	MAE	MBE
Persistence	2.254678	1.154667	0.013278
SolCast Cloudiness	0.575034	0.583713	-0.208281
SolCast GHI	0.569468	0.564057	-0.279288
MGM Meteogram	0.792630	0.699723	0.063379

B. 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 II-B. 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.
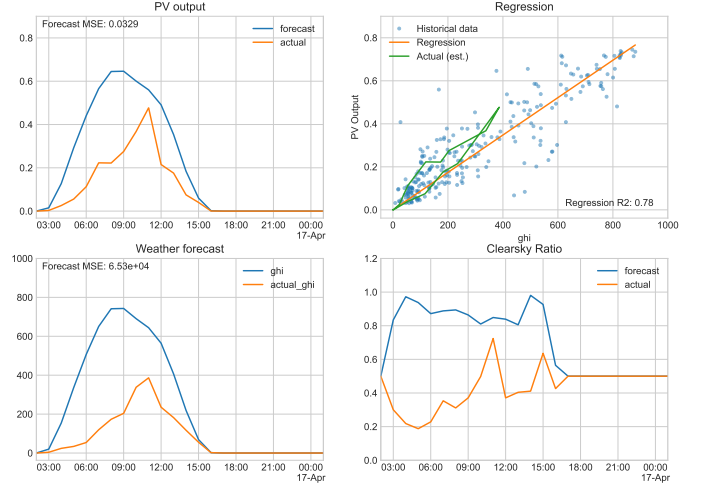

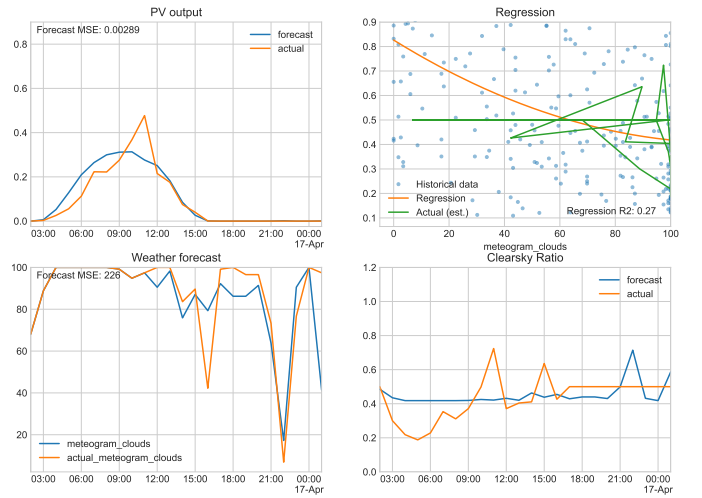Fig. 2. Example Day-ahead Forecast using Solcast GHI Forecast

Fig. 3. Example Day-ahead Forecast using MGM Meteogram Weather Forecast

For comparison, the MSE for the different intra-day forecast update methods was computed for the forecast for one hour, two hours, and six hours ahead. The result error metrics are shown in Table III.

C. 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 optimization formulation was previously presented in [18]. 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

Hours Ahead	Method	MSE	MAE	MBE
1	Persistence	0.004409	0.032320	-0.001955
	fx_output[1]	0.003743	0.031664	-0.003686
	fx_csratio[2]	0.003803	0.033566	0.006633
	exog[3]	0.003917	0.035667	0.008020
	SARIMAX	0.003726	0.032159	-0.001687
	Scaling	0.004425	0.032737	0.000284
2	Persistence	0.007663	0.041868	-0.002397
	fx_output[1]	0.005233	0.037428	-0.005418
	fx_csratio[2]	0.005684	0.041822	0.011613
	exog[3]	0.006098	0.045131	0.013845
	SARIMAX	0.005166	0.037879	-0.002296
	Scaling	0.007671	0.042636	0.001627
6	persistence	0.009497	0.047607	-0.007485
	fx_output[1]	0.005756	0.039047	-0.007912
	fx_csratio[2]	0.006253	0.042456	0.003602
	exog[3]	0.006594	0.044206	0.003111
	SARIMAX	0.005675	0.038983	-0.006740
	Scaling	0.009314	0.046933	-0.000351

[1] AR(2) on residual of PV output.
[2] AR(2) on residual of clear-sky ratio
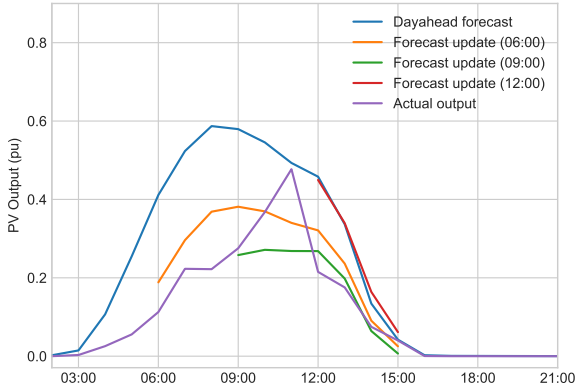[3] AR(2) on actual output with exogenous variable

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

Fig. 5. Energy Dispatch Demonstration System

Fig. 6. MPC demonstration simulation flowchart

VI. CONCLUSION

This paper demonstrated a novel light-weight forecasting method for PV generation 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 MSE compared to the persistence forecast over a test data set of 103 days. Of the weather forecast sources that were compared, SolCast cloudiness obtained the lowest MSE. Of the intra-day update methods that were compared, the SARIMAX method obtained the lowest MSE. The forecast methods with the best MSE 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 persistence forecasts were somewhat better than

actual PV data. A flowchart of the MPC simulation is shown in Fig. 6.

Table IV 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 persistence forecasts were somewhat better than the performance using the proposed method. This may potentially be explained by the relatively better bias error performance of the persistence forecasts. Additional investigation to further improve the forecasting method may be warranted.

The objective function value for optimized operation with a perfect forecast illustrates the value of improved forecasts.
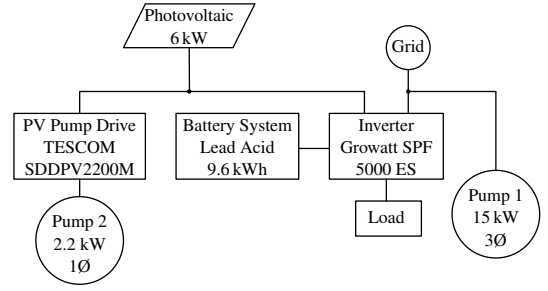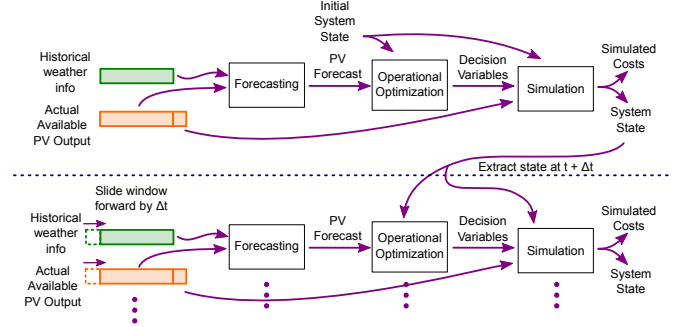
TABLE IV
MPC SIMULATION RESULTS - OBJECTIVE FUNCTION VALUE

	Persistence Forecast ($, 2020)	Proposed Method ($, 2020)	Perfect Forecast ($, 2020)
Grid energy cost	65 350	72 195	58 052
Battery use cost	8120	7979	8047
Inadequate water cost	90	99	0
Battery mode switching cost	246	244	230
Pump switching cost	94	104	80
TOTAL	82 846	90 474	66 418

the performance using the proposed method. This unexpected result may warrant additional investigation.

One direction for future work is 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.

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. [Online]. Available: http://dx.doi.org/10.1016/j.solener.2016.06.069

[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. [Online]. Available: https://www.mdpi.com/2076-3417/12/15/7529

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

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

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

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

[13] 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.

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

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

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

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

[18] 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.

[19] ——. (2023) Software repository for lightweight photovoltaic forecasting method for agricultural microgrids. [Online]. Available: https://github.com/pdb5627/ISGT2023-PV-forecasting-paper

[20] ——. (2023) Reproducible computing container for lightweight photovoltaic forecasting method for agricultural microgrids. [Online]. Available: https://codeocean.com/capsule/5295207/tree/v1

[21] M. P. Almeida, O. Perpiñán, and L. Narvarte, "Pv power forecast using a nonparametric pv model," *Solar Energy*, vol. 115, pp. 354–368, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X15001218

[22] 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. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X12001429

[23] 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. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8066360