Day-Ahead PV Power Forecasting for Control Applications

Ürkmez, Mirhan; Kallesøe, Carsten; Bendtsen, Jan Dimon; Leth, John-Josef

Published in:
2022 IECON – 48th Annual Conference of the IEEE Industrial Electronics Society

DOI (link to publication from Publisher):
10.1109/IECON49645.2022.9968709

Publication date:
2022

Document Version
Accepted author manuscript, peer reviewed version

Day-Ahead PV Power Forecasting for Control Applications

Mirhan Ürkmez, Carsten Kallesøe, Jan Dimon Bendtsen, and John Leth

Aalborg University, Fredrik Bajers Vej 7c, DK-9220 Aalborg, Denmark
(e-mail: {mu,csk,dimon,jjl}@es.aau.dk)

***Abstract*—The percentage of solar energy among all installed energy sources is increasing each year. Photo-voltaic (PV) power forecasting is key to the application of control methods in systems with PV panels. In this paper, we present a method for day-ahead PV power forecasting at each time step that is easy to train and can be applied to different power data types (e.g data from hot and cold climates, with various sampling times). Predictions made before and after sunrise are handled separately. Exponentially Weighted Moving Average (EWMA) is applied on the normalized daily power data to estimate the shape of the next-day power curve for the predictions before sunrise. Then, the multiplier value which would expectedly produce the best forecast when multiplied with the estimated shape is predicted using a time-series approach. After sunrise, the observed power data is leveraged to improve the previous forecasts. The proposed method is shown to perform well on multiple data sets with varying characteristics. Also, the method is compared with some benchmarks algorithms, and the results are presented.**

***Index Terms*—Renewable energy, PV power forecasting, Control applications**

I. Introduction

The need for renewable energy is increasing as the world is facing the climate change issue along with population growth. Solar photo-voltaic (PV) are standing out among others as it is one of the cheapest renewable energy sources and gets cheaper each year. In 2021, solar PV was responsible for 60% of the newly added renewable capacity [1]. The installment of new PV panels is expected to grow in the future as well. According to the IRENA PV report (2019), solar panels are expected to constitute 25% of the world's energy production by 2050 [2].

The projected increase in the use of PV panels brings along new problems for the management of the energy. PV power production does not follow a deterministic pattern. Various weather conditions such as solar irradiance, temperature, humidity, etc. introduce stochastic behaviour to the power production. Due to the volatile nature of solar energy, the PV power production forecast is necessary for its integration into the existing systems such as power grids. The penetration of PV power to the electrical grid damages voltage and frequency stability and power quality [3]. An accurate prediction of the produced PV power is required to build appropriate control algorithms such as Model Predictive Control (MPC) so that the power stability and quality of the electrical grid can be maintained [4]. Application of Model Predictive Control(MPC) on the systems equipped with PV panels depends on making power predictions at each time step.

This work is funded by Independent Research Fund Denmark (DFF).

Weather measurements, previous power production data and numerical weather predictions (NWP) taken from the weather stations are among the used inputs for the estimation models. NWP are effective for day-ahead predictions since power production is severely affected by rain, clouds, etc [5]. The idea of classifying days according to NWP has been applied in the literature. In [6], forecast given by an ANN is updated according to the different bias classes which are separated by their irradiance values. Forecasts of LSTM and the persistence model are used interchangeably depending on the similarity of the weather of the previous day and the forecast day [7]. Similarly, in a day-ahead prediction method, days are classified into the 3 classes: sunny, partly cloudy and overcast [8]. NWP data are also used along with physical models in the literature. Physical system properties of PVs are used to derive power production models which formulate power as a function of weather variables such as radiance, temperature, wind speed [9], [10], [11]. Then, the power estimation is made using the NWP data. Numerical weather predictions are useful for detecting cloudy days which are low points of power production. However, applicability of the methods using NWP as an input to the other PV sites is not always possible since NWP might not be available.

Recently, Neural Networks (NN) have been applied to the problem at hand. Artificial Neural Networks (ANN) methods differ in the architectures and the inputs and the outputs of the network. In[12], an ANN is trained for each of the prediction horizons targeted in the paper. NWP data and the previous power production data are fed to ANN for day-ahead prediction in some works [6], [8]. Another approach in day-ahead predictions is having an ANN that predicts the production for a smaller horizon and getting the day-ahead prediction by feeding the predicted values back to the network [13]. Long short-term memory (LSTM) networks based methods are also available in the literature as they are proven to be successful in time series forecasting [14], [15], [7]. The accuracy of the LSTM on a small data set is investigated [16]. In [17], LSTM is used with NWP inputs. They are also used for short term power forecasts [18], [19].

Most of the available methods target predicting the production at a specific time. In this paper, we are proposing a fast day-ahead PV power prediction algorithm for real-time control applications. Differing from most of the available work, we address the problem of predicting the power at each time step so that it can be used in control applications such as MPC. The method is appropriate for different PV sites with varying weather conditions and easy to apply as we do not use NWP as input to the system which are not accessible at all places. Also, the proposed method does not need an extensive training. The method is compared to some benchmark methods and shown to outperform them.

The outline of the rest of the paper is as follows. The proposed method is explained in Section II The experimental results are presented in Section III. The paper is concluded in with final remarks in Section IV.

II. Methodology

PV power data can be viewed as consisting of two parts: zero production at night and a roughly concave-shaped production curve during the day. The predictions made at night can only be based on the power data of the past days as NWP are not used in this study. The problem of predicting at night is separated into two sub-problems which are estimating the normalized shape of the next day data and the multiplier of the estimated shape to produce the power prediction.

If we do not have any information about the clouds or the solar irradiance, a common way to estimate the production data of the next day is by combining those of the preceding days [5], [20]. Moving from this assumption, we use Exponentially Weighted Moving Average (EWMA) to predict the shape of the production data. EWMA smooths out the daily fluctuations and outputs what to expect before the day starts. Let $t \in \mathbb{N}$ denote the current day. At midnight between day $t-1$ and t, the production shape of the current day given the previous data is estimated as

$$X'_{t-1} = \frac{1}{\max(X_{t-1})} X_{t-1}, \tag{1a}$$

$$Y_t = \alpha X'_{t-1} + (1-\alpha) Y_{t-1} \tag{1b}$$

where $X_{t-1} \in \mathbb{R}^N$ is the vector of production data for day $t-1$ with N the number of production data points in day $t-1$, $X'_{t-1} \in [0,1]^N$ is the normalized daily data, $\alpha \in [0,1]$, and $Y_t \in [0,1]^N$ is the moving average normalized daily production of days $1, 2, \ldots, t-1$. In this paper we also use Y_t as an estimate of the shape of the production data of day t. Daily production data X_{t-1} is normalized to equalize the contribution of low-power and high-power days to the shape. The choice of α determines the ratio between importance of newer and older observations. As α gets bigger, the contribution of the older observation to the estimation decreases. For $\alpha = 1$, the shape forecast is exactly that of the previous day.

With the method mentioned above, we get a normalized forecast of the data taking values between 0 and 1. The next step is to find the coefficient that multiplies the normalized shape to produce the final prediction. One option is to multiply the normalized shape with the estimated daily maximum power. However, it does not always produce the desired outcome, especially when there is a peak in the daily production. In Figure 1, a case is shown where multiplication with the daily maximum results in a large error. Instead, normalized forecast Y_t is multiplied with an estimation of the optimal multiplier defined as the minimizer of the cost function given by

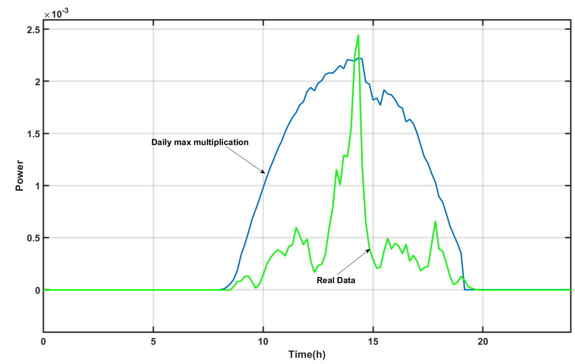

Fig. 1: A daily production curve (green) and an approximation of it which is predicted shape Y_τ multiplied with daily max power value (blue).

$$\min_{p} \quad f(p) = \sum_{i=1}^{N} (pY_{\tau i} - X_{\tau i})^2 \tag{2}$$

where $p \in \mathbb{R}$ is the multiplier, and $Y_{\tau i}$ (resp. $X_{\tau i}$) are the i^{th} coordinate of Y_τ (resp. X_τ). The aim is to minimize the squared error of the forecast. The solution of this unconstrained optimization problem can be found with the optimality condition $\nabla f(p) = 0$ and is given by

$$p_\tau = \frac{\sum_{i=1}^{N} Y_{\tau i} X_{\tau i}}{\sum_{i=1}^{N} Y_{\tau i}^2}. \tag{3}$$

The value of p_τ is calculated so that $p_\tau Y_\tau$ is the best approximation of the production data X_τ for a given shape Y_τ. It can only be calculated if all the production data X_τ of day τ is available (see (3)). That is, at day t the optimal multiplier p_t can first be calculated just before midnight between day t and $t+1$. As day t progresses the strategy is to estimate the optimal multiplier $p_{\tau=t}$.

The estimation of p_t is done using a time-series analysis method. The optimal multiplier p_τ of each day $\tau = 1, \ldots, t-1$ is calculated at the end of day using (3). The time series of optimal multipliers $p_\tau, \tau = 1, \ldots, t-1$ are modelled using an Autoregressive Moving Average model ARMA(1,1) given by

$$p_\tau = \mu + \phi p_{\tau-1} + \theta \epsilon_{\tau-1} + \epsilon_\tau \tag{4}$$

where ϵ_τ and $\epsilon_{\tau-1}$ are the forecast errors at step $\tau, \tau-1$ respectively and μ, ϕ, θ are the constants to be computed. The error terms $\epsilon_\tau = p_\tau - \hat{p}_\tau,\ \tau = 1, \ldots, t$ are assumed to be coming from independent zero-mean normal distributions. An estimation of the next value $\hat{p}_t$ is made using the calculated optimal value of the previous day p_{t-1} as

$$\hat{p}_t = \mu + \phi p_{t-1} + \theta \epsilon_{t-1}. \tag{5}$$

Moreover, because p_t is unknown when (5) is applied, it is chosen to set ϵ_t to zero based on the zero-mean assumption. The prediction made at midnight for the day t is $\hat{p}_t Y_t$.

At midnight between day $t-1$ and t, the ARMA(1,1) model (4) is retrained using $\{p_\tau\}_{\tau=1}^{t-1}$ time series to find the values of μ, ϕ, θ which minimize the sum of squared error

$$S(t) = \sum_{\tau=1}^{t-1} \epsilon_\tau^2. \tag{6}$$

The day-time forecasting differs from the night predictions in the availability of observed power data. Once we get non-zero data, those observations can be combined with the prior knowledge to improve the forecast for the future values of the current day. First, sunrise time is detected when consecutive power values are above a certain threshold. The threshold and the number of consecutive data needed may change between data sets depending on the weather conditions and the time step. The threshold should be bigger than the biggest night production value. Increasing the number of consecutive data may guarantee the correctness of the sunrise time in hot climates but it also delays the detection. On the contrary, it deteriorates the detection in cold, cloudy climates where almost zero production can be observed in some stages in the morning. In this paper, we are checking 2 consecutive power values for sunrise detection.

Then, an observation is made from the available power values at each time step after sunrise on day t by solving the problem given by

$$\hat{p}_t^m = \arg\min_p \quad \sum_{i=s_r}^{s_r+m} (pY_{ti} - X_{ti})^2 \tag{7}$$

where $\hat{p}_t^m$ is the estimation of optimal multiplier based on the observed data from sunrise time s_r and m time steps thereafter. The problem is the same as in (2), but only the values after sunrise are used in (7).

Let z_t^m denote random variable related to the observations given by (7). The observation $z_t^m = \hat{p}_t^m$ and the prior estimation $\hat{p}_t$ are combined for a better estimation of p_t using Bayes' Theorem (see (8a)). The prior is assumed to be normally distributed since $p_t = \hat{p}_t + \epsilon_t$ where the value of $\hat{p}_t$ is known and $\epsilon_t \sim N(0, \sigma^2)$. We assume the likelihood $\mathcal{L}(p_t \mid \hat{p}_t^m) = P(z_t^m = \hat{p}_t^m \mid p_t)$ can also be represented as a normal distribution using the model $z_t^m = p_t - \epsilon_t^m$ where $\epsilon_t^m \sim N(0, \sigma_m^2)$. The variances σ^2 and σ_m^2 are estimated from the sets of available ϵ_τ and $\epsilon_\tau^m, \tau = 1, \ldots, t-1$ values from the previous days .

$$P(p_t = x \mid z_t^m = \hat{p}_t^m) \propto P(p_t = x) P(z_t^m = \hat{p}_t^m \mid p_t = x) \tag{8a}$$

$$\hat{p}_{t|m} = \arg\max_x P(p_t = x \mid z_t^m = \hat{p}_t^m) \tag{8b}$$

The posterior probability distribution $P(p_t | z_t^m = \hat{p}_t^m)$ is found by multiplying the prior and the likelihood distributions, as in (8a). Then, the value maximizing the posterior probability distribution is selected as the new optimal multiplier estimation $\hat{p}_{t|m}$. as in (8b). The overall algorithm for the 1-day horizon prediction at each time stage is given in Algorithm 1.

Algorithm 1 Prediction Procedure for the day t at all time steps with 1-day time horizon beginning from midnight

```
    Input X_{t-1}, Y_{t-1}, σ, σ_m, {p_τ}_{τ=1}^{t-2}
    Output {pred_k}_{k=1}^{N}                  ▷ Predictions at time step k
procedure PV POWER PREDICTION
    Operations at midnight:
    N ← (24 × 60)/ΔT                          ▷ # of data points
    X'_{t-1} ← (1 / max(X_{t-1})) X_{t-1}
    Y_t ← αX'_{t-1} + (1 − α)Y_{t-1}
    p_{t-1} ← (Σ_{i=1}^{N} Y_{(t-1)i} X_{(t-1)i}) / (Σ_{i=1}^{N} Y_{(t-1)i}^2)
    μ, φ, θ ← arg min_{μ,φ,θ}  Σ_{τ=1}^{t-1} ε_τ^2
    p̂_t ← μ + φp_{t-1} + θε_{t-1}
    pred ← p̂_t Y_t
    night ← true
    Operations at each time step:
    for k ← 1 to N do
        if ISSUNRISE(X_{t(k-1)}, X_{tk}) then
            night ← false
            s_r ← k − 1
        if night then
            pred_k ← [pred(k : end), pred(1 : k − 1)]     ▷
1-day horizon prediction at step k
        else if !night then
            m = k − s_r
            p̂_t^m ← arg min_p  Σ_{i=s_r}^{s_r+m} (pY_{ti} − X_{ti})^2
            p̂_{t|m} ← arg max_x F(x, p̂_t, σ)F(x, p̂_t^m, σ_m)
            pred_k ← [p̂_{t|m} Y_t(k : end), pred(1 : k − 1)]
    function F(x, μ, σ)            ▷ PDF of Normal distribution
        return (1 / (σ√(2π))) exp(−(1/2)((x−μ)/σ)^2)
    function ISSUNRISE(val_1, val_2)
        if val_1 > thresh and val_2 > thresh then
            return true
        return false
```

III. Results

Two publicly available PV power data sets are used for testing and validation purposes. One is from a hot climate, Albuquerque, US [21], the other is from a colder climate, YMCA station in south of UK [22]. The Albuquerque data set includes power production data of 1 year with 1-minute sampling period, while the YMCA data set contains 160 days of data with a sampling period of 10 minutes.

Each data set is split into two parts: training and test sets. The Albuquerque training and test sets include 195 and 150 days respectively. Some days have not been used due to having incorrect information such as the date and power data. The YMCA data set is divided to include 120 and 40 days in the training and test sets. The training data sets are used for the initial training of an ARMA(1,1) model for optimal multiplier prediction and the estimation of variances for likelihood and prior distributions. The variances of the

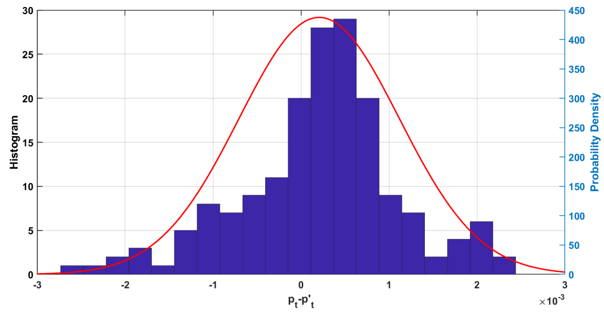

Fig. 2: Histogram and fitted probability distribution of $\epsilon_\tau^4 = p_\tau - z_\tau^4$ obtained from Albuquerque train set when z_τ^4 is observed 4 time-step (40 minutes) after the sunrise.

likelihood probability distributions are found from the time series $\epsilon_\tau^m = p_\tau - z_\tau^m, \tau = 1...t-1$ for each m. Histogram of the time series and the fitted Gaussian when z_τ^m observation is made 40 minutes after sunrise, $m = 4$, are shown in Figure 2. Our assumption that ϵ_τ^m is normally distributed seems reasonable as the histogram resembles a normal distribution. It should also be noted that the used method does not depend the underlying distribution being Gaussian. Any probability distribution approximating the behaviour of the error terms ϵ_τ^m could be used.

The tests are performed with 10-minutes sampling time and prediction horizon of 1-day for both data sets unless specified otherwise. The exponential weight parameter α in EWMA equation (1) is set to 0.9. The power threshold for the sunrise detection is 0.1 kW. The prediction performances are assessed with normalized root mean square error (nRMSE) and normalized mean absolute error (nMAE) metrics, see (9).

$$\texttt{nMAE} = \frac{100}{N y_{max}} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{9a}$$

$$\texttt{nRMSE} = \frac{100}{y_{max}} \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \tag{9b}$$

The prediction performances of night and day-time predictions are investigated on the test data sets. We expect an increase in the success of the prediction as more daily power data becomes available. This phenomenon is illustrated in Figure 3 where real data of a day and predictions at different time stages are shown. The data in the figure has a relatively smooth shape with some negative peaks. Sometimes the daily data are so volatile that the proposed method is not able to capture daily profile closely. An unusual power profile along with the predictions are given in Figure 4. In the first two hours after the sunrise, the production seems higher than expected. Accordingly, the predictions for the rest of the day are updated to the higher values as shown in the figure. However, the production falls significantly in the afternoon due to clouds or some other reason and the predictions made before the sunrise approximates the data more closely in that region. nRMSE and nMAE results of the forecasts made both before and after sunrise are given in Table I. A gradual improvement can be observed after sunrise which proves that the proposed model can capture the information from the power data arriving after sunrise on both data sets. After two hours, the estimation of the optimal multiplier $\hat{p}_\tau^m$ is generally close to the real value p_τ. On the other hand, a new value of power data from the next day is estimated with each step because the prediction horizon is fixed to 1-day. Therefore, the predictions start to become less accurate after around 2 hours.

Prediction Time	Albuquerque data set		YMCA data set	
	nRMSE (%)	nMAE (%)	nRMSE (%)	nMAE (%)
Midnight	9.66	5.47	7.10	3.40
+30 min.	9.14	4.99	6.64	3.14
+40 min.	9.07	4.97	6.65	3.15
+1 hour	8.63	4.62	6.60	3.14
+1.5 hour	7.93	4.20	6.45	3.06
+2 hours	7.80	4.12	6.19	2.93
+3 hours	7.95	4.23	6.40	3.00

TABLE I: Prediction results on 2 test sets. +X time means predictions made X time after sunrise.

Sampling Time	Proposed Method		Persistence Model (PM)	
	nRMSE (%)	nMAE (%)	nRMSE (%)	nMAE (%)
1 min	10.28	5.79	12.56	6.09
5 min.	9.88	5.59	11.35	5.86
10 min.	9.66	5.47	11.50	5.70
15 min	9.49	5.39	11.23	5.61

TABLE II: Comparison of Persistence Model (PM) and the Proposed Method with different sampling times.

The robustness of the method to the changing sampling time is tested on Albuquerque test data set which originally had 1-minute time intervals. In the tests, the time interval of the data set is increased by taking the average power values. The performance results of the proposed method and the Persistence Model (PM) which is widely used in the literature for day-ahead predictions as a baseline model are given in Table II. PM is a simple model which outputs power data of the previous day $X_{\tau-1}$ as the forecast of the next day X_τ. The results show that the proposed method is suitable for various sampling times. The effect of the parameter α is also investigated and the results are presented in Table III. Initially, $\alpha = 0.9$ is used which is determined with fine-tuning by looking at the training set results. The results show that the algorithm performs well with changing values of α once we are in an optimal neighborhood.

Relative performances of the proposed method and 3 benchmark methods which are widely used in the literature are examined. LSTM networks are used to capture the nonlinear relationships in time series predictions. Several LSTM network structures (e.g. with and without time of the day inputs, small and large number of hidden units) are trained for prediction purposes on both Albuquerque and YMCA data sets, but none of them achieved accuracy metrics worth noting here. However, when an LSTM structure is trained to only make

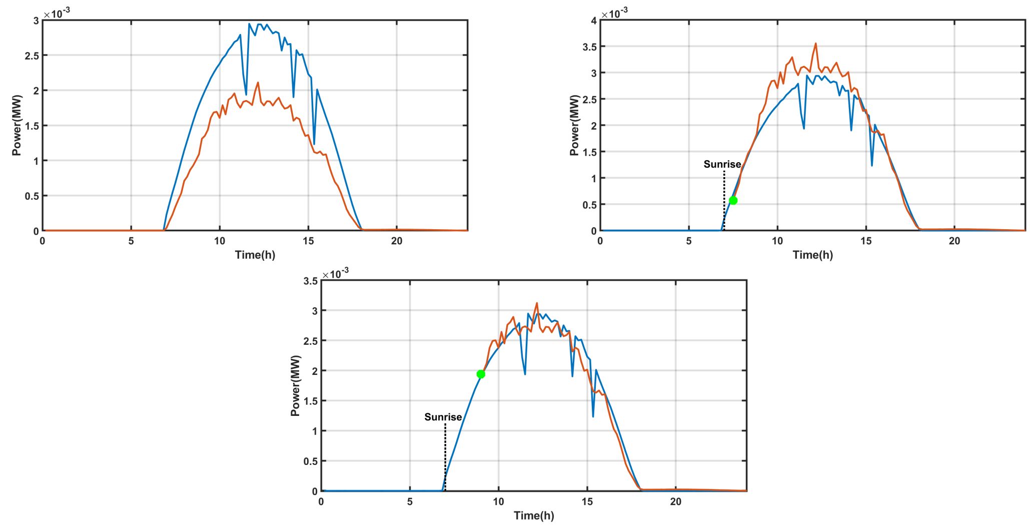

Fig. 3: Predictions (red curves) at night(left), 30 minutes after sunrise (right) and 2 hours after sunrise (bottom) and actual power data (blue). Predictions get better as more daily data becomes available.

α value	Prediction Time	nRMSE (%)	nMAE (%)
0.85	Midnight	9.74	5.49
	+ 2 hours	7.96	4.21
0.875	Midnight	9.50	5.36
	+ 2 hours	7.86	4.15
0.9	Midnight	9.66	5.47
	+ 2 hours	7.80	4.12
0.925	Midnight	9.68	5.51
	+ 2 hours	7.78	4.09
0.95	Midnight	9.79	5.62
	+ 2 hours	7.89	4.16

TABLE III: Prediction results with changing α values.

Method	Prediction Time	nRMSE (%)	nMAE (%)
LSTM	Midnight	10.31	6.01
SARIMA	Midnight	9.41	5.33
	+ 30 minutes	8.86	5.78
	+ 1 hour	8.86	6.10
	+ 2 hour	9.63	6.80
	+ 3 hour	10.26	7.34
PM	Midnight	11.50	5.70
	+ 30 minutes	11.55	5.73
	+ 1 hour	11.58	5.72
	+ 2 hour	11.63	5.72
	+ 3 hour	11.75	5.74
Proposed Method	Midnight	9.66	5.47
	+ 30 minutes	9.14	4.99
	+ 1 hour	8.63	4.62
	+ 2 hour	7.80	4.12
	+ 3 hour	7.95	4.23

TABLE IV: Prediction results on 2 test sets. +X time means predictions made X time after sunrise.

predictions at midnight with 1-day horizon, it was still worse than the proposed method but closer. The LSTM structure takes time series of daily production data $X_\tau, \tau = 1 \dots t-1$ as input and outputs X_t such that both input and output dimensions are equal to the number of data points in a day. The results of the compared methods are reported in Table IV. We used a Seasonal Autoregressive Integrated Moving Average, SARIMA $(1,1,1)(1,1,1)_{144}$ model along with the PM as our benchmarks. A detailed explanation of the SARIMA model can be found at [23]. The results show that the PM model produces more erroneous results due to daily changes in the data. To shape the expectation for the next day, one has to consider multiple days because of the high volatility among consecutive days. SARIMA is successful in its predictions made at midnight. However, it is not able to improve them as the proposed method does after sunrise with the new power data. All in all, the proposed method seems to be working better considering both the predictions made before and after sunrise.

IV. Conclusion

We have presented a PV power forecasting method for control applications. The method aims to find an exponentially weighted local mean of the normalized production curve independent of daily fluctuations using EWMA method. The optimal multiplier of the normalized curve for each day is estimated with an ARMA model. Finally, the daily estimations are updated after sunrise with incoming power data using a Bayesian learning setup. Several experiments are performed to test our assumptions and the performance of the method. The method is shown to be applicable to PV sites at both hot and cold climates. The effect of sampling time of the data set on the performance is also investigated. The method is compared with benchmark methods SARIMA, PM and LSTM. The experiments show that the proposed method achieve better results than other methods considering predictions made at different time steps.

As future work, experiments could be extended to see the effect of yearly weather changes on the results. Also, the algorithm could be improved to speculate about the daily

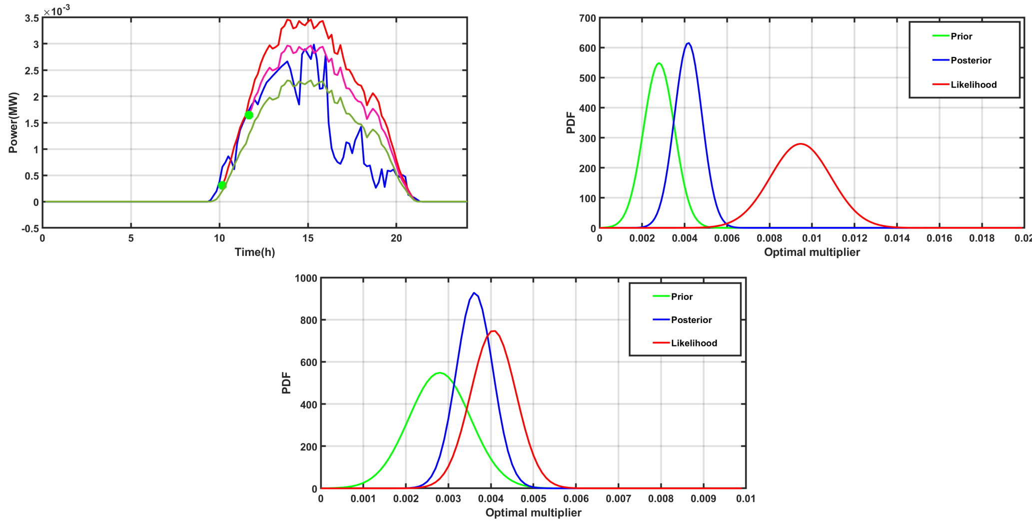

Fig. 4: Left- Predictions at night(green), 30 minutes after sunrise (red) and 2 hours after sunrise (pink) and actual power data (blue). Right- Probability distributions 30 minutes after sunrise. Bottom- Probability distributions 2 hours after sunrise.

fluctuations as well. To this end, NWP data could be taken advantage of. Another direction could be finding a better way to extract the optimal multiplier information from daily power data.

References

[1] IEA, "Renewables 2021," https://www.iea.org/reports/renewables-2021, 2021, iEA, Paris.

[2] IRENA, "Future of solar photovoltaic: Deployment, investment, technology, grid integration and socio-economic aspects (a global energy transformation: paper)," https://www.irena.org/publications/2019/Nov/Future-of-Solar-Photovoltaic, 2019, international Renewable Energy Agency, Abu Dhabi.

[3] P. K. N. Nwaigwe, P. X. Mutabilwa, and E. Dintwa, "An overview of solar power (pv systems) integration into electricity grids," *Materials Science for Energy Technologies*, vol. 2, pp. 629–633, 2019.

[4] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE Journal of Power and Energy Systems*, vol. 1, pp. 38–46, 2015.

[5] Y. Zhang, C. Jin, R. K. Sharma, and A. K. Srivastava, "Data-driven day-ahead pv estimation using hybrid deep learning," *2019 IEEE Industry Applications Society Annual Meeting*, pp. 1–6, 2019.

[6] S. Theocharides, G. Makrides, A. Livera, M. Theristis, P. Kaimakis, and G. E. Georghiou, "Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing," *Applied Energy*, vol. 268, p. 115023, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261920305353

[7] Y. Zhang, C. Jin, R. K. Sharma, and A. K. Srivastava, "Data-driven day-ahead pv estimation using hybrid deep learning," in *2019 IEEE Industry Applications Society Annual Meeting*, 2019, pp. 1–6.

[8] A. Mellit, A. M. Pavan, and V. Lughi, "Short-term forecasting of power production in a large-scale photovoltaic plant," *Solar Energy*, vol. 105, pp. 401–413, 2014.

[9] T. Huld, G. Friesen, A. Skoczek, R. P. Kenny, T. Sample, M. Field, and E. D. Dunlop, "A power-rating model for crystalline silicon pv modules," *Solar Energy Materials and Solar Cells*, vol. 95, no. 12, pp. 3359–3369, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0927024811004442

[10] F. Bizzarri, M. Bongiorno, A. Brambilla, G. Gruosso, and G. S. Gajani, "Model of photovoltaic power plants for performance analysis and production forecast," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 2, pp. 278–285, 2013.

[11] F. E. Atencio Espejo, S. Grillo, and L. Luini, "Photovoltaic power production estimation based on numerical weather predictions," in *2019 IEEE Milan PowerTech*, 2019, pp. 1–6.

[12] M. G. De Giorgi, P. Congedo, and M. Malvoni, "Photovoltaic power forecasting using statistical methods: Impact of weather data," *Science, Measurement & Technology, IET*, vol. 8, pp. 90–97, 05 2014.

[13] S. Leva, A. Dolara, F. Grimaccia, M. Mussetta, and E. Ogliari, "Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power," *Mathematics and Computers in Simulation*, vol. 131, pp. 88–100, 2017, 11th International Conference on Modeling and Simulation of Electric Machines, Converters and Systems. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378475415001238

[14] M. Park, J. Lee, W. Kang, J. Choi, and K. H. Lee, "Predictive model for pv power generation using rnn (lstm)," *Journal of Mechanical Science and Technology*, vol. 35, pp. 1–9, 01 2021.

[15] H. Chen and X. Chang, "Photovoltaic power prediction of lstm model based on pearson feature selection," *Energy Reports*, vol. 7, pp. 1047–1054, 2021, 2021 International Conference on Energy Engineering and Power Systems. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484721009768

[16] V. De, T. T. Teo, W. L. Woo, and T. Logenthiran, "Photovoltaic power forecasting using lstm on limited dataset," in *2018 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, 2018, pp. 710–715.

[17] M. S. Hossain and H. Mahmood, "Short-term photovoltaic power forecasting using an lstm neural network and synthetic weather forecast," *IEEE Access*, vol. 8, pp. 172 524–172 533, 2020.

[18] H. Ahn and N. Park, "Deep rnn-based photovoltaic power short-term forecast using power iot sensors," *Energies*, vol. 14, p. 436, 01 2021.

[19] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep lstm-rnn," *Neural Computing and Applications*, vol. 31, pp. 2727–2740, 2017.

[20] D. J. Pedregal and J. R. Trapero, "Adjusted combination of moving averages: A forecasting system for medium-term solar irradiance," *Applied Energy*, vol. 298, p. 117155, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261921005882

[21] C. W. Hansen, W. F. Holmgren, A. Tuohy, J. Sharp, A. T. Lorenzo, L. J. Boeman, and A. Golnas, "The solar forecast arbiter: An open source evaluation framework for solar forecasting," in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, 2019, pp. 2452–2457.

[22] U. P. Networks, "Validation of photovoltaic (pv) connection assessment tool closedown report)," https://www.ofgem.gov.uk/sites/default/files/docs/2015/03/pv_tool_cdr_final_0.pdf, 2015.

[23] J. Cryer and K.-S. Chan, *Time Series Analysis: With Applications in R*, 01 2008.