40th European Photovoltaic Solar Energy Conference and Exhibition

This presentation was selected by the Sc. Committee of the EU PVSEC 2023 for submission of a full paper to one of the EU PVSEC's collaborating peer-reviewed journals.

DETECTION OF SHADING FOR SOLAR POWER FORECASTING USING MACHINE LEARNING TECHNIQUES

Tim Kappler, Anna Sina Starosta, Nina Munzke, Bernhard Schwarz, Marc Hiller

Karlsruher Institut for Technology (KIT)

76344 Eggenstein-Leopoldshafen , Hermann-von-Helmholtz-Platz 1

ABSTRACT: This paper presents a machine learning based solar power forecast methode that can take into account partial shading related fluctuations. Generated PV power is difficult to predict because there are various fluctuations. Such fluctuations can be weather related when a cloud passes over the array. But they can also occur due to shading caused by stationary obstacles. In this work an approach is presented that improves the forecast under such fluctuations caused by partial shading. This adaptation is necessary, because partial shading is usually not detected directly. Such shading occurs after the growth of trees or later built buildings. The presented algorithm can detect such effects itself and thus works self-learning. A correction of the prediction of a forecast model could successfully reduce forecast error due to partial shading. The model is evaluated on the basis of two months of recorded shading data in which shading was caused by a tree infront a PV array. The correction uses internal inverter data and irradiance values of the previous day to perform the correction and was able to reduce the RMSE of a 10 kWp under shading and thus improve the prediction accuracy by up to 40% depending on how strong the shading is. The model can detect how intense the shading is and correct the forecast by itself.

Keywords: shading detection, solar power forecasting, machine learning

1 Introduction

The number of photovoltaic systems installed worldwide and the associated installed capacity rose by 23.6 % to more than 1100 GW in 2022. Further increase is expected to meet CO_2 emission targets in the future [1]. Since the generated power fluctuates as PV systems are subject to cloud movements, rain or changes in irradiation, problems of grid stability are more and more in focus [2]. Accurate forecasts of generated power and energy are necessary to maintain and guarantee stability and availability. Especially machine learning methods have gained popularity in this context, since their advantage is to have a good generalization capability and are therefore able to adapt quickly to new situations [3]. The disadvantage of these methods is the large amount of data that has to be collected to achieve high accuracy. Typically, exogenous data such as irradiation data, wind speed or air temperature are used for solar power forecasting [4–8]. In addition, it is difficult to predict newly occurring situations that are not considered in the recorded data set because no information about them is available. For example, incoming shade from trees that have grown taller or buildings that were constructed later can reduce the PV system's output. Other possible effects are pollution [9] by dust and leaves or degradation of the PV modules [10]. With information about irradiation, wind speed, ambient temperature or sun angle, it is not possible to represent such caused drops in power. The forecast error becomes larger when shading is present because the trained models are not able to handle the shading on their own. There is already a wide range of papers dealing in general with the prediction of the generated solar power [11,8] as well as with the effects of shading [12, 13] and how to model and analyze shading or solar systems under yield reducing effects [14,15]. Shading was already taken into account in power predictions, where the yield loss over the period of one year [16] was considered. Also, in daily forecasts as in [17], shading was included by reducing the effective irradiance and thus by preprocessing. In this paper, the focus is on a correction of the already existing prediction value with a model that has seen training data over years and thus could develop a good generalization capability. The consideration is therefore done by a postprocessing. In addition, loss effects such as shading and soiling are quantified and allow later condition monitoring approaches in an extension. To the authors' knowledge, there is no research that combines quantification of shading losses based only on PV yield data with solar power forecasting. Thus, no extra soiling or shading sensors are needed. This is investigated due to the significant impact of shading on solar cells. Long Short-Term Memory (LSTM) networks are used as a basis for the prediction model. The model is validated with four shading configurations with PV data recorded over several months. The RMSE of the shaded PV systems, which have different orientations and inclinations, is compared. In addition, soiling is taken into account. Additionally, soiling is considered. In this paper, one shading scenario is presented.

The paper is structured as follows. First, the basic method of the procedure to consider shading for PV forecasts is described. For this purpose, three individual submodels are described, which are required for the understanding of the method. Then, the presented method is validated using data from a array shaded by a tree. Finally, the results are discussed and an outlook is given.

2 Methodology

The following section explains how the whole approach works. Three sub-models are described, with a basic introduction of the models used for the forecasting part. The shading is taken into account by means of a correction value. For this purpose, the forecast values of the prediction model are corrected to take the effect of shading into account. First, the forecast model itself is described in this section. This method is first presented in a fundamentals section. Afterwards it is explained how the forecast model is trained to predict the generated PV power one day ahead. Then, a PV array model is used, which simulates the power that an unshaded PV array would provide based on irradiation and temperature values. Finally, it is discussed how the shading power is determined from the simulated PV power and the actual measured values. This refers to the power that is lost due to the shadows that occurs.

The principle of the approach is visualized in Figure 1. The forecast model provides a predicted value, which represents the solar power one day ahead. The PV array model calculates the power that an unshaded array will deliver under the same conditions. Together with the calculated and actual measured value, the shading power can be determined and the actual predicted value can be corrected afterwards so that the prediction can take the shading into account.

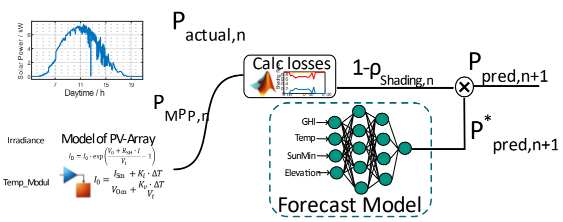

Figure 1 Overall approach to considering shading for solar performance predictions. The approach is divided into three submodels with the prediction model, the PV array model and the separation of losses.

2.1 Fundamentals

In the following, the used forecast model is introduced. The focus lies on how the forecast models work and which parameters need to be trained or optimized using collected data (which will be explained more in detail in section 2.3).

2.1.1 Long Short-Term Memory Networks (LSTM)

The basic unit of each neural network is the perceptron. It maps a weighted sum of the input data x_i with the edge weights w_{ij} and transfer function ψ to the output h_j.

Layers can be built up based on this unit. The output values of each neuron are finally calculated according to Equation 1.

$$h_j = \Psi\left(\sum_{i=1}^{n} w_{ij} \cdot x_i\right) \tag{1}$$

If several of these layers are linked together, this arrangement is called a neural network as displayed in Figure 2. These are called feed-forward networks (FFN) because the layers with p_n neurons in the k-th layer mesh in the forward direction from input to output with the transfer function Ψ [18].

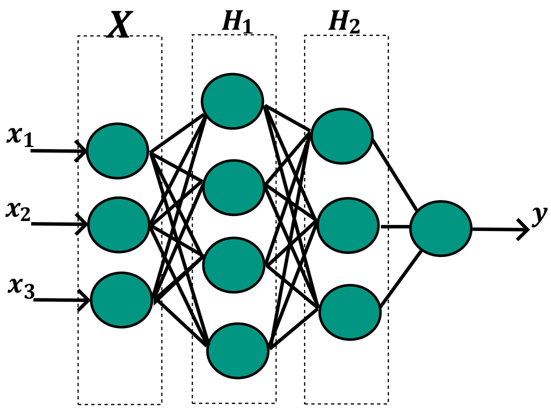

Figure 2 Structure of a two-layer neural network with three input features, two hidden layers and one output

Finally, the data are transformed from the input layer to the output y_k.

Such FFNs have been able to demonstrate good prediction capabilities in many publications in the past [19, 20]. In more recent publications, however, recurrent neural network architectures are used, which have shown an improvement in prediction accuracy several times [21,22]. In particular, LSTM networks have been able to achieve more and more popularity in scientific publications due to their memory capability [22,23] and their ability to deal with the exploding and vanishing gradient problem [24]. LSTM networks consist of a large number of gates that store knowledge about the previous state. These data are either written to, stored in or read from a cell that serves as a type of memory. When the cell reads, writes or deletes information using the input and forget gates, it makes a decision about whether to store the data. Based on the signals received, they become active and use their own weighted filters to decide whether to forward or suppress the information based on its importance and strength. These weights are similar to those that adapt during the training phase of the network to modulate the input and hidden states [25].

2.2. Data

The data are classified into endogenous and exogenous data.

Exogenous data: Since 1981 the German Weather Service (DWD) offers historical and forecast weather data from selected weather stations on its publicly accessible website [29,30]. Weather data includes data such as air temperature, irradiation, cloud cover, humidity and many others. In this work, weather data are used as input features to train and validate the forecast model. An LSTM network is used as the forecast model, as described in Section 2.1.1.

Endogenous data: Data from the solar park at KIT Campus North are used. It is located at 49.1° north and 8.44° east. There a total of 102 PV arrays with an installed capacity of around 10 kWp each. The arrays have inclinations between 2° and 60° and an orientation between 60° west and 60° east. The data of the solar park is recorded since 2014. Power data of the inverters (string voltage, string currents, string power) as well as irradiance and module temperatures of selected arrays are available. In addition, the geographical location is used to determine the solar angles with the help of the library pvLib [31] for the description of the solar trajectory.

The hourly mean values of the data are used for all investigations and calculations. Since data can be corrupted, the data are filtered beforehand. These outliers can be explained by sensor errors or communication problems during transmission of the data to the database. From the outset negative values and values that are measured significantly above the Standard Test Conditions (STC) can be considered corrupt. However, since this concerns only a small amount of data, the corrupted values can be easily compensated by interpolation of the neighboring values.

2.2.1 Shading structures

For the evaluation of the methodology, one of the 102 PV arrays is selected and is marked in Figure 3. Later, the calculated power losses are used to subsequently correct the forecast values of the model.

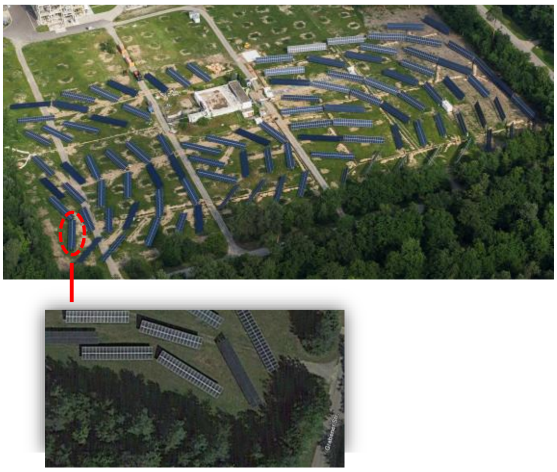

Figure 3**.** Solar park at the North Campus of the KIT with shaded array. The array has an inclination of 15° and orientation of 30° west.

2.2.1 Forecast model

An LSTM network is used to forecast the power one day ahead. As a data-driven algorithm, the endogenous and exogenous data described earlier are used to determine the parameters of the neural network. For this purpose, a total of seven years of recorded data are split into two data sets. A training data set to optimize the parameters and a validation data set to test the model on unknown data. Six years are used for training and one year for validation.

2.4 PV array model

To determine the power of an unshaded array, a model of a PV array is needed. This model uses irradiance and temperature data to determine the power of a PV array that does not face any shading. In general, it returns the power of an ideal PV array without yield degrading effects.

This model can be built from the 1-diode model [35] and thus can be parameterized completely from existing data sheets of the used modules. A perturbation and observation (PO) algorithm is used to operate the array at the maximum power point (MPP).

2.5 Separation of losses

Now the ideal power can be described using the PV array model and the predictions can be calculated using the forecast model. The recorded data and the ideal PV model can be used to estimate the shading losses similar to [36]. A ratio ρ is defined, which puts the ideal (P_{ideal}) and actual power (P_{DC}) into a direct relation, as in Equation 2.

$$\rho = \frac{P_{\text{ideal}}}{P_{\text{DC}}} \tag{2}$$

Since soiling contributes to a reduction in output in addition to shading, a soiling ratio ρ_{soiling} can be calculated as the average value of the ratio at midday hours (ρ_{Noon}), as in Equation 3.

$$\rho_{\text{Soiling}} = \overline{\rho_{\text{Noon}}} \tag{3}$$

In general, the average ρ_{Noon} should be calculated at unshaded times. The power dissipation on the soiling and shading are then calculated according to Equation 4-5.

$$P_{\text{Soiling}} = \rho_{\text{Soiling}} \cdot P_{\text{ideal}} \tag{4}$$

$$\rho_{\text{Shading}} = \rho - \rho_{\text{Soiling}} \tag{5}$$

The shading P_{Shading} losses are calculated through Equation 6.

$$P_{\text{Shading}} = \rho_{\text{Shading}} \cdot P_{\text{ideal}} \quad (6)$$

3. Results

In the following section the results are presented. In the result part, an explaination is given how shading data are recorded and evaluated and how the entire approach is validated. Furthermore, the limitations of the method are discussed.

3.1. Separation of losses

Data has been recorded over two months. This allows the manipulation of power due to shading to be recreated and observed in real terms using the inverter data. Using the strongly shaded array as an example, the power that falls on the shading can be visualized well in Figure 4.

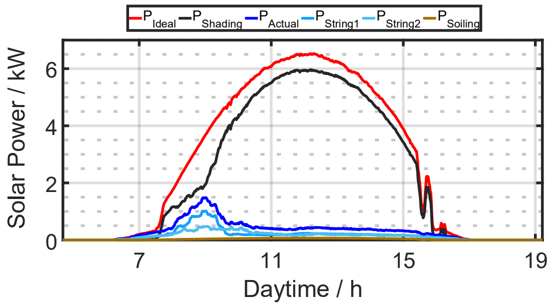

Figure 4 The calculated portion of the power due to shading is shown in black. The theoretical maximum power is reduced by this amount.

3.2. Validation

With the help of the losses from the previous day and from the shading ratio $\rho_{Shading}$ a correction can now take place by multiplication with the forecast value. It is also conceivable to subtract the power loss from the forecast value. But the correction via the shading ratio makes more sense here, since the losses proportional to the occurring shading always depend on the actual irradiation and thus the power.

It makes sense to multiply the forecast values by $1 - \rho_{Shading}$ because the shading remains constant over the days if it comes from a stationary obstacle. At the same time, the forecast model that was trained with the historical data provide forecasts in the shading period. In Figure 1 the scheme of the correction is shown as well as the change of the *RMSE* over the hours of the day when the model is applied. However, since the correction model detects the shading power at these times, it can make the correction and adjust the power values downward. So, you can also see in Figure 5 the change of the *RMSE* by the correction model. In principle, other forms of correction are also possible with the method shown. Figure 5 also shows that a correction using the shading power of the previous day ("lag correction") can already result in a significant improvement of the forecast error. A perfect correction would be if the power losses of the previous day are exactly equal to those of the current day. This illustrates the maximum potential of the presented method. In the period from February to March there was a 40% improvement in the RMSE, while in the April to May the improvement was 12.5%. The difference is in the intensity of the shading. Due to the low sun elevation in spring, there is much more shading of the array. Thus, the uncorrected LSTM model is more inaccurate than with a smaller amount of shading.

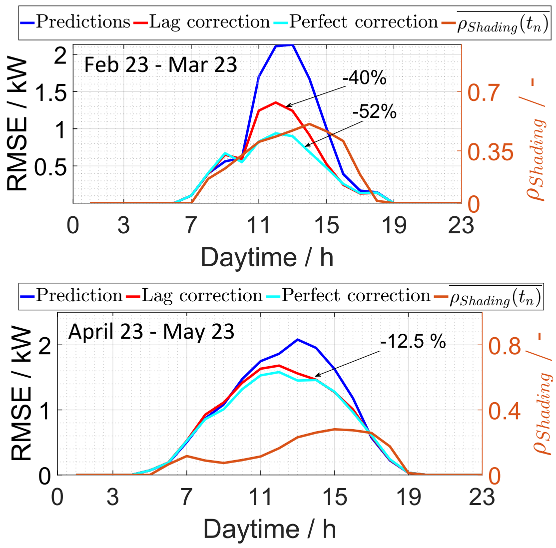

Figure 5 Decrease in RMSE due to the correction process

4. Discussion

Although the correction method in the previous validation has consistently contributed to a noticeable improvement in the prediction error under shading, the method also has limitations. It is in the nature of forecast models to over- or underestimate the true value. If the prediction values are now reduced, although the true value is underestimated, an improvement in the prediction error is not guaranteed.

In general, the power loss due to shading must be greater than the bias of the prediction model to improve the forecast error. The validation dataset showed that the true measured power value is underestimated, especially in the early morning hours and in general during the summer months. A correction in the case of very weak shading does not always lead to an improvement in the RMSE due to the negative Bias.

5. Conclusion

In this work, a forecast model was trained and a PV model was parameterized using endogenous and exogenous data. The PV model uses a PO algorithm to guarantee the MPP. Based on the power values of the PV model and the actual measured data, it was possible to calculate how much power is lost due to shading. This could finally be used to subsequently correct the prediction value of the trained prediction model and thus achieve an improvement in the prediction error under shading.

In summary the subsequent correction and therefore postprocessing of day-ahead PV power forecasts can help improve forecast error. The presented procedure was able to contribute to an improvement of the forecast error for a shaded array based on two months of recorded shading data. Particularly large shadings can be recognized and corrected in the forecast model. An improvement of the RMSE by up to 40 % could be achieved depending on the extend of shading.

As an outlook, the aproach should be extended. It would make sense to extend the forecast correction by including soiling. In principle, the following methodology could also be extended to other inverter faults. The error would then be detected at time n and the forecast would be corrected at time $n + 1$ day depending on the forecast horizon. In addition, the limitations of the methods were worked out. At the same time, the presented method is interesting for solar systems in general, since the quantified shading power can determine whether solar power optimizers can sensibly retrofit their performance. This application would lead to a reduction in the levelized cost of electricity (LCOE).

Acknowledgement

This work contributes to the research performed at KIT Battery Technology Center. The results were generated within the "Solarpark 2.0" project (funding code 03EE1135A) funded by the Federal Ministry for Economic Affairs and Climate Action (BMWK). The authors thank the project management organization Jülich (PTJ) and the BMWK.

Authors contribution

The development and evaluation of the correction model, the construction of the shading setup, and the writing of the manuscript were performed by Tim Kappler. Anna Sina Starosta contributed significantly to the development of the forecast procedures. Nina Munzke, Bernhard Schwarz, Anna Sina Starosta, and Marc Hiller also contributed significantly to the development with suggestions and intellectual input.

References

1. *Solar pv*, Renewable electricity – Renewables 2022 – Analysis - IEA: 2023-09-21
2. A.S. Saidi, *Impact of grid-tied photovoltaic systems on voltage stability of tunisian distribution networks using dynamic reactive power control* (2022), Vol. 13, p. 101537
3. H. Ye, B. Yang, Y. Han, N. Chen, *State-Of-The-Art Solar Energy Forecasting Approaches: Critical Potentials and Challenges* (2022), Vol. 10
4. J. Zeng, W. Qiao, *Short-term solar power prediction using an RBF neural network*, in *2011 IEEE Power and Energy Society General Meeting* (2011), pp. 1–8
5. A. Fentis, L. Bahatti, M. Mestari, B. Chouri, *Short-term solar power forecasting using Support Vector Regression and feed-forward NN*, in *2017 15th IEEE International New Circuits and Systems Conference (NEWCAS)* (2017), pp. 405–408
6. P. Bacher, H. Madsen, H. Nielsen, *Online Short-term Solar Power Forecasting* (2011), Vol. 83
7. F. Almonacid, P. P´erez-Higueras, E.F. Ferna´ndez, L. Hontoria, *A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator* (2014), Vol. 85, pp. 389–398
8. A. Starosta, K. Kaushik, P. Jhaveri, N. Munzke, M. Hiller,*A Comparative Analysis of Forecasting Methods for Photovoltaic Power and Energy Generation with and without Exogenous Inputs*, in *38th European Photovoltaic Solar Energy Conference and Exhibition : proceedings of the International Conference : 06 September - 10 September 2021. Ed.: J.M. Almeida Serra* (WIP-Renewable Energies (WIP), 2021), pp. 938 – 945, ISBN 3-936338-78-7, 37.12.02;
LK 01
9. M.R. Maghami, H. Hizam, C. Gomes, M.A. Radzi, M.I.Rezadad, S. Hajighorbani, *Power loss due to soiling on solar panel: A review* (2016), Vol. 59, pp. 1307–1316
10. G.G. Kim, W. Lee, B.G. Bhang, J.H. Choi, H.K. Ahn,
Fault Detection for Photovoltaic Systems Using Multivariate Analysis With Electrical and Environmental Variables (2021), Vol. 11, pp. 202–212
11. A. Alcan˜iz, D. Grzebyk, H. Ziar, O. Isabella, Energy Reports **9**, 447 (2023)
12. Y. Chaibi, M. Malvoni, A. Chouder, M. Boussetta,M. Salhi, Energy Conversion and Management **196**, 330 (2019)
13. A. Dolara, G.C. Lazaroiu, S. Leva, G. Manzolini, Energy**55**, 466 (2013)
14. R. Ahmad, A.F. Murtaza, H. Ahmed Sher, U. TabrezShami, S. Olalekan, Renewable and Sustainable Energy Reviews **74**, 721 (2017)
15. A. Babatunde, S. Abbasoglu, M. Senol, Renewable andSustainable Energy Reviews **90**, 1017 (2018)
16. S. Pareek, R. Dahiya, Energy **95**, 561 (2016)
17. M.D.L.F.L.A.P.R.L.L.G.M.F.F.R.P.M.S.R. Lucaferri, V.; Radicioni, *An indirect approach to forecast produced power on photovoltaic plants under uneven shading condition* (2022), https://iris.uniroma1.it/handle/11573/1675678
18. C.C. Aggarwal, *Neural Networks and Deep Learning* (Springer Cham, 2018)
19. M. Rana, A. Rahman, *Multiple steps ahead solar photovoltaic power forecasting based on univariate machine learning models and data re-sampling* (2020), Vol. 21, p. 100286
20. H. Malki, N. Karayiannis, M. Balasubramanian, *Shortterm electric power load forecasting using feedforward neural networks* (2004), Vol. 21, pp. 157 – 167

21. A. Yona, T. Senjyu, T. Funabashi, *Application of Recurrent Neural Network to Short-Term-Ahead Generating Power Forecasting for Photovoltaic System*, in *2007 IEEE Power Engineering Society General Meeting* (2007), pp. 1–6
22. S. Srivastava, S. Lessmann, *A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data* (2018), Vol. 162, pp. 232–247
23. M. Gao, J. Li, F. Hong, D. Long, *Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM* (2019), Vol. 187, p. 115838
24. F. Harrou, F. Kadri, Y. Sun, *Forecasting of Photovoltaic Solar Power Production Using LSTM Approach* (2020), ISBN 978-1-83880-091-8
25. C.H. Liu, J.C. Gu, M.T. Yang, *A Simplified LSTM Neural Networks for One Day-Ahead Solar Power Forecasting* (2021), Vol. 9, pp. 17174–17195
26. M. Awad, R. Khanna, *Support Vector Regression* (Apress, Berkeley, CA, 2015), pp. 67–80, ISBN 978-1-4302-5990-9, https://doi.org/10.1007/978-1-4302-5990-9$_4$
27. A. Anghel, N. Papandreou, T. Parnell, A. Palma, H. Pozidis, *Benchmarking and optimization of gradient boosted decision tree algorithms* (2018)
28. V.K. Ayyadevara, *Gradient Boosting Machine* (Apress, Berkeley, CA, 2018), pp. 117–134, ISBN 978-1-4842-3564-5, https://doi.org/10.1007/978-1-4842-3564-5$_6$
29. DWD, https://opendata.dwd.de/climate$_e$*nvironment/CDC/*
30. DWD, https://opendata.dwd.de/weather/local$_f$*orecasts/*
31. W. Holmgren, C. Hansen, M. Mikofski, *pvlib python: a python package for modeling solar energy systems* (2018), Vol. 3, p. 884
32. H. Chen, X. Chang, *Photovoltaic power prediction of LSTM model based on Pearson feature selection* (2021), Vol. 7, pp. 1047–1054, 2021 International Conference on Energy Engineering and Power Systems
33. M.F.N. Tanvir Ahmad, Sharmin Sobhan, *Comparative Analysis between Single Diode and Double Diode Model of PV Cell: Concentrate Different Parameters Effect on Its Efficiency* (2016)
34. M.H. Qais, H.M. Hasanien, S. Alghuwainem, K. Loo,M. Elgendy, R.A. Turky, *Accurate Three-Diode model estimation of Photovoltaic modules using a novel circle search algorithm* (2022), Vol. 13, p. 101824
35. E. Batzelis, G. Anagnostou, C. Chakraborty, B. Pal, *Computation of the Lambert W function in photovoltaic modeling* (2019)
36. S. Ghosh, J. Roy, C. Chakraborty, *A model to determine soiling, shading and thermal losses from PV yield data* (2022)
37. T. Selmi, M. Abdul-Niby, L. Devis, A. Davis, *PO MPPT implementation using MATLAB/Simulink*, in *2014 Ninth International Conference on Ecological Vehicles and Renewable Energies (EVER)* (2014), pp. 1–4