EPJ Photovoltaics **15**, 17 (2024)
© T. Kappler et al., Published by EDP Sciences, 2024
https://doi.org/10.1051/epjpv/2024013

Available online at:
www.epj-pv.org

Special Issue on 'EU PVSEC 2023: State of the Art and Developments in Photovoltaics', edited by Robert Kenny and João Serra

REVIEW OPEN ACCESS

Detection of shading for short-term power forecasting of photovoltaic systems using machine learning techniques

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

Institute of Electrical Engineering (ETI), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

Received: 7 July 2023 / Accepted: 6 March 2024

Abstract. This paper presents a machine learning based solar power forecast method that can take into account shading related fluctuations. The 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, and this paper addresses this form of shading. In this work an approach is presented that improves the forecast under such fluctuations caused by shading. A correction of the prediction could successfully reduce error due to shading. The evaluation of the model is based on five sets of recorded shading data, where shading resulted from intentionally placed structures. 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 four 10 kWp systems with different orientation and tilt angle under shading and thus improve the prediction accuracy by up to 40%. The model can detect how intense the shading is and correct the forecast by itself.

Keywords: Solar power forecasting / machine learning / fault detection / shading

1 Introduction

The number of photovoltaic systems installed worldwide and the associated installed capacity rose by 22% to more than 1000 TWh in 2021. Further increase is expected to meet CO_2 emission targets in the future [1]. As the generated power of PV systems fluctuates due to factors such as cloud movements, rain, or changes in irradiation, issues related to grid stability are increasingly coming into focus [2]. Accurate forecasts of generated power and energy are necessary to maintain and guarantee stability and availability. Forecasting methods can use physical models, statistical methods or machine learning methods. 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 over years to achieve high accuracy. The trend is increasingly towards the use of new neural network architectures, which are particularly well suited to solar power forecasting. Jianwu Zeng et al. were able to show that an RBF network can represent the internal connections of the data particularly well and was able to outperform other non-linear models [4]. Furthermore, in a comprehensive study of various deep learning approaches by Dairi et al., it was investigated that variational auto-encoder networks could best represent fluctuating behavior [5]. But there are also studies that come to the conclusion that SVR can prevail over neural networks, such as Fentis et al. [6] and Starosta et al. [7]. Typically, exogenous data such as irradiation data, wind speed or air temperature are used for solar power forecasting [8]. Furthermore, in Bacher et al. it can be shown that weather forecast data as input data for the solar power prediction methods is significantly more representative for forecast periods of over two hours than for shorter periods in which the past measured PV power is significantly more meaningful [9]. Solar power forecasting is not only a subject of scientific research, but are also already being used commercially. A comprehensive comparison of such solutions has already been done. Lehmann et al. carried this out based on a test period of 6 months [10].

In addition, it is difficult to predict newly occurring situations that were 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 [11] by dust and leaves or degradation of the PV modules [12]. With information about irradiation, wind speed, ambient

* e-mail: tim.kappler@kit.edu

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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 [7,13] as well as with the effects of shading [14,15] and how to model and analyze shading or solar systems under yield reducing effects [16,17]. Shading was already taken into account in production forecasts, where the loss of yield was considered over a long period of one year [18]. Physical models can be used to compensate the shading by using geometrical information as shown in Mayer et al. [19]. Here two shading models were used to predict the solar power. Direct shading from neighboring PV arrays was taken here into account. Masa-Bote et al. [20] have also used a forecast of energy in daily frequency for an energy management system for a system of battery storage, PV with a statistical ARIMA approach used an estimation for the shading by surrounding trees. A shading factor was here determined there, which corrects the daily forecasted Energy under constant shading.

However, since complex physical or mathematical models to include dynamic shading cannot be integrated into classical machine learning algorithms, an approach to take such effects into account is necessary. All the papers mentioned so far that deal with the forecast with machine learning methods of solar power do not explicitly address the effects of shading caused by obstacles. Since such effects are subject to sesonal fluctuations due to changing sun positions in addition to the already existing fluctuation of cloud movements. Only a few papers have training data sets of more than one year and cannot take such effects into account. Since the effects of shading by obstacles can be quantified simulatively by the proposed method, a hybrid approach of data-driven methods and a physical model was chosen here to take such effects into account. 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. In addition, loss effects such as shading and soiling are quantified and allow later condition monitoring approaches in an extension. Furthermore, several PV arrays with different inclinations and orientations are considered as well as an evaluation over several months, which are located in different seasons. However, to the best of the authors' knowledge, there is no research that combines shading loss quantification with solar power prediction, which should be investigated due to the significant impact of shading on solar power generation. Long Short-Term Memory (LSTM) networks were used as a basis for the prediction model. The model is validated with four shading setups over several months of recorded PV data by comparing the RMSE of the shaded PV arrays with different orientations and inclination.

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 the shading setups. Finally, the results are discussed and an outlook is given.

The main contributions of this work are:

- Development of solar power forecasts for a day were developed based on a data set of over 7 years and a comparison of popular machine learning methods was performed.
- A simulation model based on the 1-diode model was built and validated.
- Shading effects were determined based on measurement and simulation.
- A method was developed that can explicitly consider shading effects in solar power forecasts and is validated over different array configurations.

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 subsequently corrected to take the effect of shading into account. First, the forecast model itself is described in this section. For this purpose, three classical ML methods are compared. These methods are first presented in a fundamentals section to show how the different methods work. 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 presented, 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 occur.

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.

2.1 Fundamentals

In the following, the used forecast models are briefly 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 Sect. 2.4). All models are later used for the forecast of the generated power one day ahead.

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

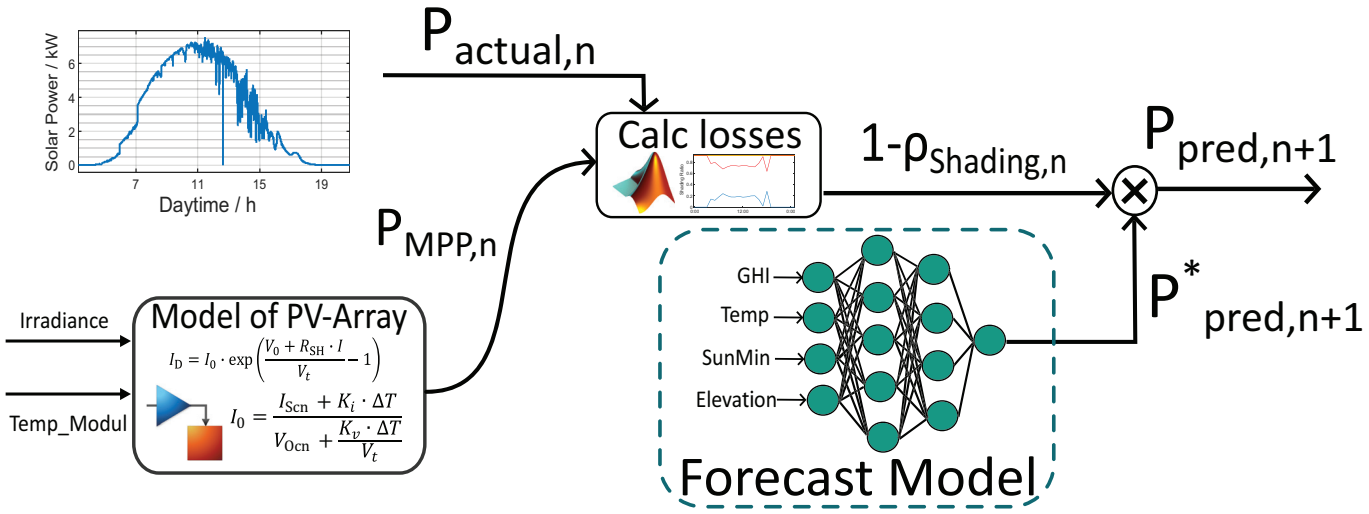


Fig. 1. Overall approach to considering shading for solar performance predictions. The approach is divided into three submodels with the prediction model, the PV array model and the separation of losses.

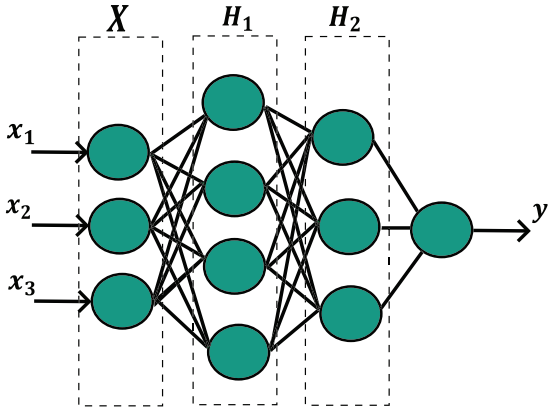


Fig. 2. Structure of a two-layer neural network with three input features, two hidden layers and one output.

values of each neuron are finally calculated according to equation (1)

$$h_{\mathrm{j}} = \psi\left(\sum_{i=1}^{n} w_{\mathrm{ij}} \cdot x_{\mathrm{i}}\right). \tag{1}$$

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 [21].

Finally, the data are transformed from the input layer to the output y_k using equations (2)–(4)

$$\overline{h_1} = \Psi\left(W_1^T X\right) \tag{2}$$

$$\overline{h_2} = \Psi\left(W_{\mathrm{p+1}}^T \overline{h_{\mathrm{p}}}\right) \tag{3}$$

$$y = \Psi\left(W_{k+1}^T \overline{h_k}\right). \tag{4}$$

Such FFNs have been able to demonstrate good prediction capabilities in many publications in the past [22,23]. In more recent publications, however, recurrent neural network architectures are used, which have shown an improvement in prediction accuracy several times [24,25]. In particular, LSTM networks have been able to achieve more and more popularity in scientific publications due to their memory capability [25,26] and their ability to deal with the exploding and vanishing gradient problem [27]. 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 [28]. The mathematical relationships between the input x_{t} and the output h_{t} can be expressed with equations (5)–(10). How the individual gates depend on each other can be visualized in Figure 3.

$$\overline{C_{\mathrm{t}}} = \tanh(W_{\mathrm{C}} \cdot [h_{\mathrm{t-1}}, x_{\mathrm{t}}] + b_{\mathrm{C}}). \tag{5}$$

Equation (5) computes the updated cell state C_{t} at time t by applying the weighted sum of the previous hidden state $h_{\mathrm{t\text{-}1}}$, the current input x_{t}, and a bias term b_{c}.The cell state reflects the memory content of the LSTM by the current input and the prior state.

$$C_{\mathrm{t}} = f_{\mathrm{t}} \cdot C_{\mathrm{t-1}} + i_{\mathrm{t}} \cdot \overline{C_{\mathrm{t}}}. \tag{6}$$

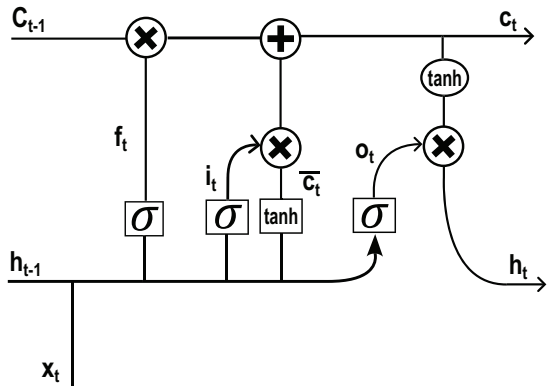


Fig. 3. Structure of a LSTM cell with input vector x_t and output vector h_t.

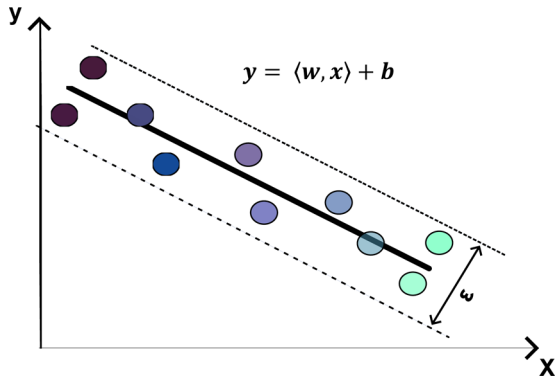


Fig. 4. Hyperplane of the SVR with epsilon strap. Values of SVR are ordered by brightness.

Equation (6) determine the change of the cell state C_t, set its update as a function of the forget gate output f_t, the input gate output i_t, and the cell state $\overline{C_t}$. The forget gate regulates the storing of information from the previous cell state, while the input gate determines the intake of the actual cell state.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f). \tag{7}$$

The forget gate f_t is determined by information from the previous cell state according to equation (7). It involves a weighted sum of the previous hidden state h_{t-1}, the current input x_t and a bias term b_f.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i). \tag{8}$$

Computed analogously to the forget gate, the input gate i_t selects information from the previous cell state h_{t-1}, the current input x_t, and a bias term b_i.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o). \tag{9}$$

The output gate o_t determines the output of the LSTM cell at time t. It is computed through the sigmoid activation function of the current cell state. The computation involves the previous hidden state h_{t-1}, the current input x_t, and a bias term b_o.

$$h_t = o_t \cdot \tanh(C_t). \tag{10}$$

The hidden state h_t at time t is computed by element-wise multiplication of the output gate o_t with the hyperbolic tangent of the current cell state C_t. This resultant hidden state serves as the output of the LSTM cell at t.

2.1.2 Support Vector Regression

Support Vector Regression (SVR) is an extension of the Support Vector Machine (SVM) for regression problems. As with classification, SVR is characterized by the number of support vectors and kernel functions. The functions that will generate the transformation into the high dimensional space are called kernel functions. This kernel trick exploits the fact that any non-linearly separable data set becomes linearly separable by transformation into a sufficiently higher dimensional space [29]. Typically, kernel functions are linear polynomials, Gaussian functions or radial basis functions. The transition to the regression method is motivated by the fact that the regression is enabled by a classification of regression errors. This is illustrated in Figure 4.

In simple terms, SVR aims to fit a hyperplane that best approximates the training data within a certain tolerance. The optimization problem penalizes deviations beyond the tolerance level, and the regularization parameter C helps control the balance between achieving a good fit and preventing overfitting. In summary, the mathematical description of Support Vector Regression involves finding the optimal hyperplane parameters (w and b) that minimize the error between predicted and actual values, considering a margin of tolerance ε. To calculate the parameters of the SVR the following optimization problem is solved:

$$min\left(\frac{1}{2}||w||^2 + C \cdot \sum_{n=1}^{N}(\zeta_i + \zeta_i^*)\right) \tag{11}$$

subject to the constraints:

$$y_i - \langle w, x_i \rangle - b \leq \varepsilon + \zeta_i \tag{12}$$

$$\langle w, x_i \rangle + b - y_i \leq \varepsilon + \zeta_i^* \tag{13}$$

$$\zeta_i^*, \zeta_i \geq 0, \tag{14}$$

where the slack variables ζ represent the allowable error due to the training samples [30]. They enable specific data points to deviate within or violate the margin constraints, creating a balance between minimizing errors and optimizing the margin.

2.1.3 Gradient boosting regression

Gradient boosting regression trees belong to the ensemble methods, which all follow the idea that an improvement in regression accuracy is associated with a combination of multiple weak regression models [31]. This idea is taken up

in Boosted Trees by building a flock of weak decision trees h_{m} with prediction F_{m} in the m-th iteration using the available data set. Then, the error to the actual values Y_{i} is calculated. In the next iteration, the resulting problem is solved:

$$F_{\mathrm{m+1}}(x_{\mathrm{i}}) = F_{\mathrm{m}}(x_{\mathrm{i}}) + h_{\mathrm{m}}(x_{\mathrm{i}}) = Y_{\mathrm{i}} \tag{15}$$

or in a different form converted

$$h_{\mathrm{m}}(x_{\mathrm{i}}) = Y_{\mathrm{i}} - F_{\mathrm{m}}(x_{\mathrm{i}}) \tag{16}$$

it can be shown that the resulting residuals minimize the squared error

$$-\frac{\partial(Y_{\mathrm{i}} - F(x_{\mathrm{i}}))^2}{\partial F x_{\mathrm{i}}} = Y_{\mathrm{i}} - F(x_{\mathrm{i}}) = h(x_{\mathrm{i}}) \tag{17}$$

With the resulting residuals another decision tree is trained, which corrects the error of the first one. In principle, this procedure can be repeated as often as necessary until a sufficiently good result is achieved [32]. To avoid overfitting, regularization methods are typically used which scale the output of each decision tree as seen in equation (18):

$$F_{\mathrm{m}}(x_{\mathrm{i}}) = F_{\mathrm{m-1}}(x_{\mathrm{i}}) + \nu h_{\mathrm{m}}, \nu \in [0, 1]. \tag{18}$$

2.1.4 Metrics

In order to be able to compare the forecast models with each other, metrics are used which form a measure of the deviation from the actual value. The Root Mean Squared (RMSE) and the Mean Absolute Error (MAE) are two frequently used metrics. The RMSE is used to give more weight to large deviations due to the squaring of the deviations as described in equation (20), while the MAE (see Eq. (19)) indicates the average error of the forecast. Metric are used for a forecast depends on the application. Often, these measures are normalized to the maximum power in order to better compare different methods. The calculation rules for RMSE and MAE, as well as their normalized metrics are as follows:

$$MAE = \sum_{\mathrm{n=0}}^{\mathrm{N-1}} \left(\frac{|Y_n - \hat{Y}_n|}{N} \right) \tag{19}$$

$$RMSE = \sqrt{\sum_{\mathrm{n=0}}^{\mathrm{N-1}} \left(\frac{Y_n - \hat{Y}_n}{N} \right)^2}. \tag{20}$$

For better comparison, these measures are normalized to the maximum observed generated power (see Eqs (21) and (22)).

$$nMAE = \sum_{\mathrm{n=0}}^{\mathrm{N-1}} \left(\frac{|Y_n - \hat{Y}_n|}{N \cdot Y_{Max}} \right) \tag{21}$$

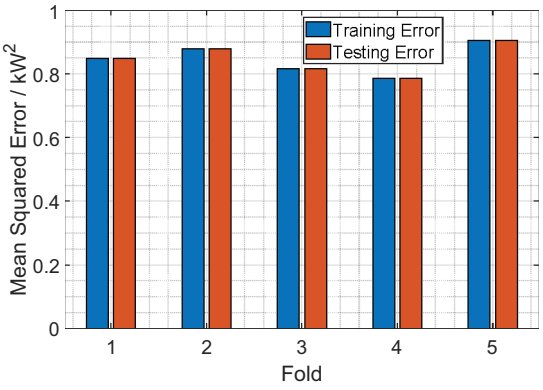


Fig. 5. Results after a 5-fold cross-validation. The training data set is divided into 5 partitions and trained on 4 partitions each. The remaining partition is then evaluated. This provides a better estimate of the forecast error.

$$nRMSE = \sqrt{\sum_{\mathrm{n=0}}^{\mathrm{N-1}} \left(\frac{Y_{\mathrm{n}} - \hat{Y}_{\mathrm{n}}}{N \cdot Y_{\mathrm{Max}}} \right)^2}. \tag{22}$$

2.2 Robustness

K-fold cross-validation serves as a means to evaluate the robustness of a model. Instead of relying on a single random division of the data into training and test sets, the data set is divided into K subsets. K-1 subsets are utilized for training for every used model, while the remaining subset is reserved for testing. This process is repeated K times, with each subset serving as the test set exactly once. By averaging the error metrics over these K iterations, a more reliable assessment of the model's performance is obtained [33]. The utilization of K-Fold cross-validation offers several advantages. Firstly, it aids in reducing variance in the performance metrics by employing multiple training and test splits, as shown in Figure 5. Additionally, it enables a comprehensive evaluation across the entire data set, as each data point is utilized once for validation. This methodology facilitates the model's ability to generalize effectively to different data sets, thereby showcasing its robustness in the face of varying conditions and data diversity.

2.3 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 [34,35]. 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 models as explained in Section 2.1.

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 are a total of 102 PV arrays with an installed capacity of around 10 kWp each. Each PV array

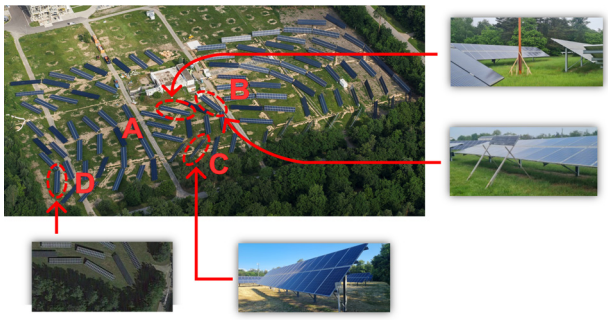


Fig. 6. Solar park at the North Campus of the KIT with array numbers. Every shading structure has its own characteristic shading pattern over time. Two self-constructed shading structures and two tree shading configurations of varying strength and orientation are considered.

uses a string inverter which can provide three MPPTs. The 10 kWp installation capacity is made up of two strings with an output of 5 kWp each. The PV tables examined each used SOLARWATT Blue modules with a peak output of 250 W according to STC conditions. This means that 20 modules are installed per string and 40 modules in total per array. 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 [36] for the description of the solar trajectory.

The hourly mean values of the data are used for all analyses and calculations, as the weather data can only be provided in hourly resolution. 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.

2.3.1 Shading structures

For the evaluation of the methodology, four of the 102 PV arrays are selected. These are marked in Figure 6. The orientation and inclination of the PV tables can be seen in Table 1. The structures were erected to study the shading effect on PV systems, which is intended to simulate shading (see Fig. 6). All shading structures apart from the wooden structure can cast a shadow over both strings.

By separating the loss effects and quantifying them, the shading times can be detected well. Later, the calculated power losses are used to subsequently correct the forecast values of the model. There are two natural shadings by trees (Fig. 6 both upper pictures) and two artificial shading structures (lower both pictures of Fig. 6). The different arrays are abbreviated in the following using the letters A–D. The different shading arrays with their characteristic shading effects are summarized in Tables 1 and 2.

Table 1. Shading structures used and their associated shading characteristics.

Array	Structure	Shading characteristic
A	Plastic pipe	Early morning hours
B	Wooden construction	Whole day
C	Tree and array	Morning and evening hours
D	Tree (protruding)	Whole day

Table 2. Corresponding array orientations and inclinations.

Array	Orientation	Inclination
A	0° South	30°
B	30° East	15°
C	60° West	30°
D	30° West	15°

The shading setup was put into operation from August 2023. After data has been recorded, the procedure is evaluated based on multiple one-month test intervals. Each structure always has its own individual shading pattern. However, the correction method should work correctly regardless of shading geometry, seasons or shading times. The shading structure at PV array D refers to the incoming shade caused by protruding trees. This shadow has a particularly strong effect in winter and spring, since the low position of the sun contributes to a particularly strong shading over both strings. With increasing time there is less and less shading, so that between May and September there is no shading on this array. In addition, there is shading at PV array C due to an adjacent tree. Here, shading occurs early in the morning and late in the evening. The shading here is relatively short and amounts to only 0.5–1 hour. Furthermore, two artificial shading setups were built up in order to be able to better assign the effects on the individual strings. One is a wooden elevation at PV array B, which creates a wide shadow throughout the day, and a tubular obstacle at array A, which contributes to shading early in the morning, but also casts a diagonal shadow instead of the rectangular one of the wooden elevation. The important aspect of the correction model presented below is that a correction can be made on the basis of the shading power. A direct detection of shadowing is realized, since an estimation of the shadow losses is always calculated in relation to the running ideal model. The procedure is thus instructed that the irradiation values accurately represent the power values. Shorter forecast intervals could solve this problem, since accuracy typically increases as the forecast interval decreases.

2.4 Forecast model

The methods presented in Section 2.1 are data-driven methods and are trained using endogenous and exogenous data so that they can predict the generated PV power. The power data are visualized in Figure 7.

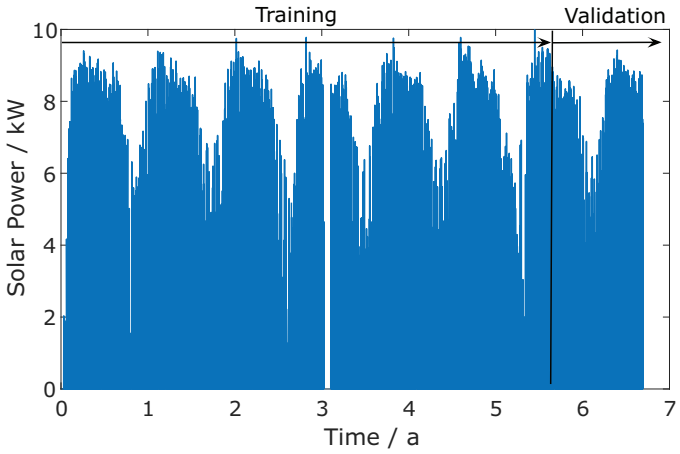


Fig. 7. Separation of the data set into training data and validation data. The validation period is one year. The split is shown on the power data. the same holds true for the input data (exogenous data).

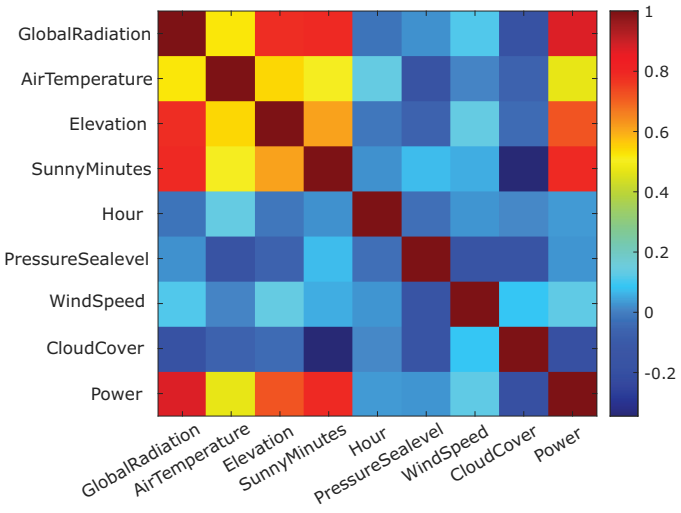


Fig. 8. Correlation matrix with Pearson coefficients.

The exogenous and endogenous data are combined in one data set. The training data set with which the forecast model is trained covers six years. Another year forms the test data set to validate it. The training data set is used to optimize the parameters with respect to the error measure and the validation data set is used to test the model on data unknown to the model.

2.4.1 Feature selection

A Pearson feature selection is necessary to create an effective and accurate model. The correlation coefficients between the input features and the output are calculated. The exogenous data is used as input features. With the forecast models the generated power is predicted one day ahead using relevant weather data. Therefore, the corresponding output feature is the power one day ahead. It is possible to determine the importance of the input characteristics for the output. For this purpose, the Pearson correlation coefficient is used, where $cov(X, Y)$ is the covariance of parameters X (Input Features) and Y (Output). σ_x and σ_y are the corresponding standard deviations of X and Y, respectively.

The results of the Pearson feature selection is summarized in Figure 8. Accordingly, Global the Horizontal Irradiation (GHI), the air temperature, the elevation angle of the sun, and the number of sunny minutes per hour are further used as input features for the forecast model. Whereas, the cloudiness degree, the current hour and the wind speed are not further used due to the low correlation ($|\rho| < 0.5$). The correlation of the features also varies over the seasons and weather conditions. This effect is comparatively low, so that considering all seasons is sufficient for feature selection, as was also shown in [37]. An alternative approach for identifying relevant features is the mutual information method, which exhibits increased sensitivity to non-linear relationships. Despite its ability to discern non-linear associations, this method converges to similar conclusions regarding the significance of features.

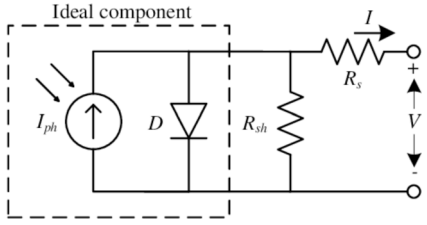


Fig. 9. 1-Diode model to describe the generated current I and applied voltage V_o of a PV array [40].

Notably, the present analysis highlights the current hour as markedly relevant. However, it is noteworthy that adding the time feature during training led to a worsening in forecast performance. Consequently, the time feature was omitted from the feature set.

2.5 PV-array model

As described in Section 2.1.4, the second step is to create a simulation model that can replicate the ideal performance of a PV array. To determine the power of an unshaded PV array, a 1-diode model is used for the PV modules. The model for the array is then consequently obtained by combining series and parallel connections of the individual module models. The structure of the 1-diode model is shown in Figure 9.

In addition to the 1-diode model, there are two- and three-diode models, whose advantage is a higher accuracy at temperature deviations [38] and additionally consider contact and optical losses [39]. The accuracy of the 1-diode model is sufficient for the following investigations and would be associated with a lower parameterization and simulation effort. The model equations are therefore as shown in equations (23)–(25):

$$I = I_L - I_D - I_p \tag{23}$$

Table 3. Characteristic values of the installed modules for all PV arrays under investigation of the solar park at KIT.

Parameter	Values
Open circuit voltage V_{oc}	37.6 V
Short circuit current I_{cs}	8.69 A
Voltage at MPP V_{MPP}	30.2 V
Maximum power P_{MPP}	250 W
Number of solar cells N_S	60
Temperature coefficient V_{oc}	−0.33%/K
Temperature coefficient I_{sc}	0.04%/K

$$I_D = I_0 \cdot \exp\left(\frac{V_0 + R_{SH} I}{V_t} - 1\right) \tag{24}$$

$$I_0 = \frac{I_{Scn} + K_i \Delta T}{\exp\left(V_{OCn} + \frac{K_v \Delta T}{V_t}\right) - 1}. \tag{25}$$

The parameters of the model equations can be obtained completely from the data sheets and compared with the current–voltage characteristics of the datasheet from the used modules. The relevant information of the data sheet for the parameterization of the model is given in Table 3.

Irradiation and temperature data are used to calculate the ideal PV power generated. For this purpose, endogenous data is used, which is recorded via sensors on the solar park.

2.6 Separation of losses

Now that the ideal power can be described using the PV array model and the forecast is calculated using the forecast model, the model for calculating the occurring shading power is still needed. The recorded data and the ideal PV model can be used to estimate the shading losses similar to [41]. For this purpose, a ratio ρ is defined, which puts the ideal and actual power into a direct relation, as in equation (26).

$$\rho(t) = \frac{P_{ideal}(t)}{P_{DC}(t)}. \tag{26}$$

However, since soiling contributes to a reduction in output in addition to shading, a soiling ratio $\rho_{\mathbf{soiling}}$ can be calculated as the average value of the ratio at midday hours, as in equation (27).

$$\rho_{soiling} = \overline{\rho_{Noon}(t)}. \tag{27}$$

In general, the average $\overline{\rho_{Noon}(t)}$ should be calculated at unshaded times. The power dissipation on the soiling and shading are then calculated according to equation (28).

$$P_{Soiling}(t) = \rho_{Soiling} \cdot P_{ideal}(t) \tag{28}$$

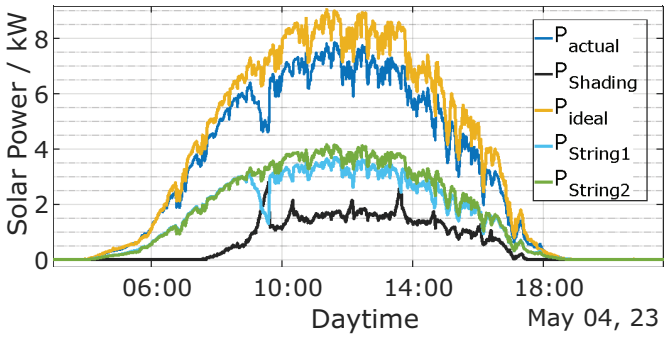


Fig. 10. Separation of the power of the heavily shaded PV array D at noon. The shadow occurs noticeably shortly before 10 a.m. by covering String 1 more than String 2. In this case, the shadows are cast by trees standing directly in front of the PV array, which cast a particularly long shadow in the early season.

with $\rho_{Shading}(t)$ as shown in equation (29).

$$\rho_{Shading}(t) = \rho(t) - \rho_{Soiling}. \tag{29}$$

The shading losses are calculated through equation (30) as follows.

$$P_{Shading}(t) = \rho_{Shading}(t) \cdot P_{ideal}(t). \tag{30}$$

Figure 10 shows an example of the separation of the power losses. By dividing the ideal and actual power, one obtains the factor ρ which is a measure for the power loss P_{Losses}. Power losses are plotted over the course of the day as shading and the resulting losses differ throughout the day (caused by structures, buildings, chimneys, trees, etc). The losses due to the soiling are proportional to the irradiation on the PV array, since the soiled modules only allow a fraction of the total irradiation power to pass through and thus cause a weakening of the irradiation.

3 Results

In the following section the results of the three submodels (forecast model, PV array model and the separation of the losses) are presented. The results are in the same order as the submodels' order in the method section. Afterwards, in the result part in Section 2.3.1, it is explained how shading data are recorded and evaluated and how the entire approach is validated. Furthermore, the limitations of the method are discussed.

3.1 Forecast model

The algorithms are trained using the filtered data set. After hyperparameter optimization was performed, predicted values were compared to actual values over a full year. Figure 11 shows three days of recorded DC power data during December 2021 and the corresponding predicted values of the three different machine learning models of array B. The errors over a whole year of the validation data

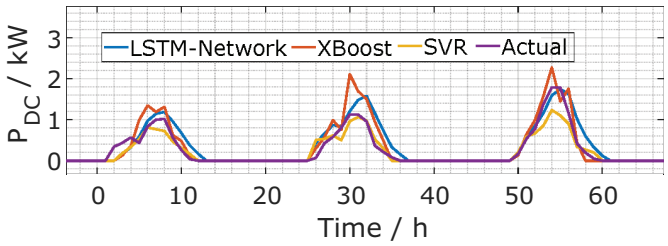


Fig. 11. Day-ahead forecasts using LSTM, gradient boosting and support vector regression over three days compared to actual measured values.

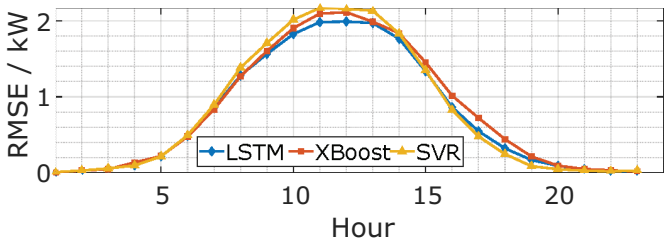


Fig. 12. The error of the forecast models is time-dependent. In particular, the error is greatest during the midday hours.

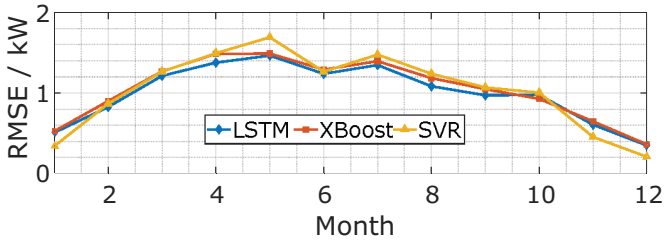


Fig. 13. The error increases during the summer months and decreases during the winter and spring months. The highest amplitude values of the generated power are to be expected in summer.

are summarized in Table 4 and displayed over the corresponding months and hours to show the distribution of the forecast error. The representation over the daytime hours (see Fig. 12) is in this case more important than the RMSE over the complete year, because the shading occurs at characteristic hours in the real field. However, the effect depends on the season, orientation and location of the obstacle that casts the shadow.

The validation over a complete year illustrates how the error is distributed over the year. Figure 13 clearly shows that the error increases in the summer months and decreases in the colder months. This is due to the high amplitude values, since more power is generated in summer than in winter or spring.

Since the LSTM network provides the best precision in this study, all further evaluations will be performed using this model in the following. However, the method works equally well regardless of the forecasting algorithm used. The investigation of the errors of the different methods should show that different methods have the same relevant error characteristics. The differences are marginal. The evaluation of the validation dataset is summarized in Table 3. For the training of the machine learning algorithms an hyperparameter tuning was performed. The parameters according to Tables 5–7 were used.

Table 4. Error measures related to the validation data set of one year for the different models LSTM, SVR and gradient boosted trees.

	RMSE (kW)	nRMSE (%)	MAE (kW)	nMAE (%)
SVR	0.97	9.7	0.47	4.7
GB	0.95	9.55	0.43	4.3
LSTM	0.90	9.0	0.40	4.0

Table 5. Parameters used for LSTM model.

Parameter	Neurons	Optimizer	Epochs	Learn rate
LSTM	20/15	Adam	500	1e-3

Table 6. Parameters used for gradient boosted trees.

Parameter	$N_{Learners}$	Learn rate	MinLeafSize
XBoost	497	57	1

Table 7. Parameter used for SVR.

Parameter	Kernel	Kernel scale	Box constrain
SVR	Rbf	57	149

3.2 PV-Array model

To validate the PV-Array model, the current-voltage characteristics of the data sheet are used. The result of the model is compared with the measured values from the data sheet (see Fig. 14). The model can thus describe the electrical behavior sufficiently well and can consequently be used to simulate the ideal behavior.

Historical irradiance and temperature data as well as real generated PV power are used to validate the model. To control the string voltages to ensure operation at maximum power point (MPP), a buck-boost converter with an Pertubation and Oberservation (PO) algorithm is used here according to [42]. Figure 15 shows that the results of the 1-diode model can reproduce the real behavior of the PV system. The model can also reproduce rapidly changing irradiation behavior as shown in Figure 16.

3.3 Separation of losses

After the shading structures have been set up, data has been recorded over several months. This allows the manipulation of power due to shading to be recreated

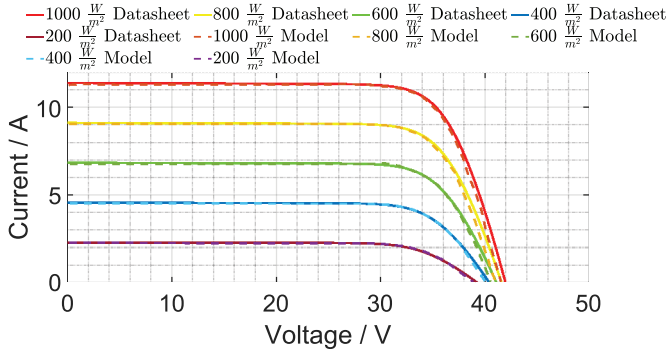


Fig. 14. Current-voltage curves of the used modules and the results of the parametric PV model (dashed).

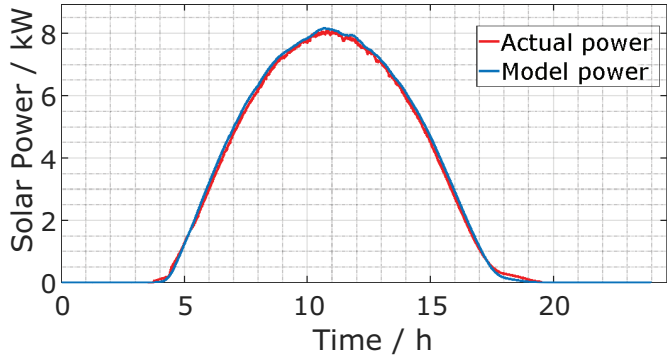


Fig. 15. Comparison of model result and measurement of the 1-diode model on a sunny day without cloud shadowing of array B.

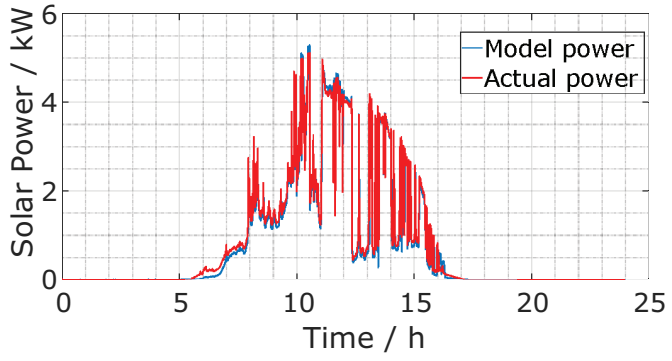


Fig. 16. Comparison of model result and measurement on a cloudy day of array B.

and observed in real terms using the inverter data (see Fig. 17). Since the shadow does not pass both strings of the PV system B at the same time, the effects of shading can be visualized particularly well by comparing the string powers. Since the simulation delivers highly accurate power values by precise irradiation measurements, it allows the exact determination of shading power throughout our measurements (see Fig. 18). Given the absence of sensor failures, the possibility of erroneously detecting shadows due to sensor errors is also precluded.

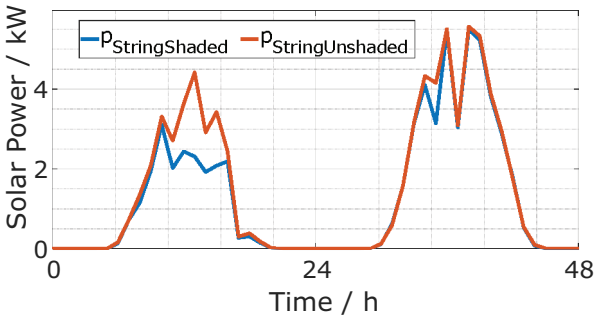


Fig. 17. Effects of shading buildup in string outputs (array B). String 2 (blue) is more affected by shadows and thus provides less power during the shading period than String 1 (red).

3.4 Validation of the approach

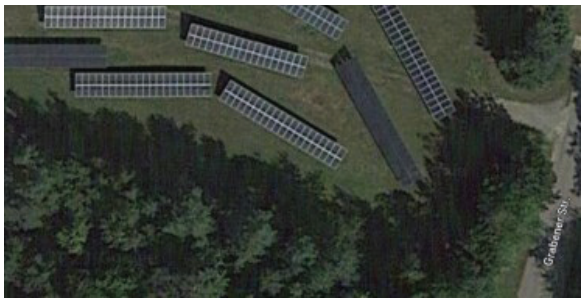
With the help of the losses from the previous day and from the shading ratio $\rho_{\text{Shading}}(t)$ a correction can now take place by multiplication with the forecast value. It is also conceivable to subtract the power loss from the forecast value. But the correction via the shading ratio makes more sense here, since the losses proportional to the occurring shading always depend on the actual irradiation and thus the power. In particular, the error is largest during the midday hours, because shading in the real environment typically occurs in the early morning and late evening hours. It makes sense to multiply the forecast values by $(1-\rho_{\text{Shading}}(t))$ because the shading remains constant over the days if it comes from a stationary obstacle related to the time points (see Fig. 19). So the new prediction values are calculated as in equation (31).

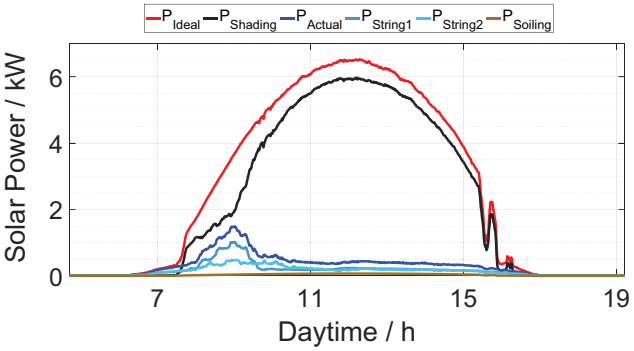
$$P_{\text{forecast}_{\text{new,n+1}}} = \left(1-\rho_{\text{Shading}_{\text{n}}}\right)\cdot P_{\text{forecast,n+1}}. \tag{31}$$

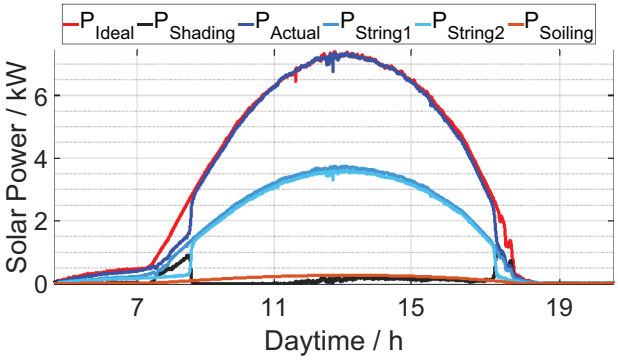
At the same time, the forecast model that was trained with the historical data and supplemented with the current weather data to provide daily forecasts in the shading period. In Figure 1 the scheme of the correction is shown as well as the change of the RMSE over the hours of the day when the model 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 1 the change of the RMSE by the correction model. In principle, other forms of correction are also possible with the method shown. Figure 20 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. The RMSE and MAE values over the entire observation period are shown in Table 8.

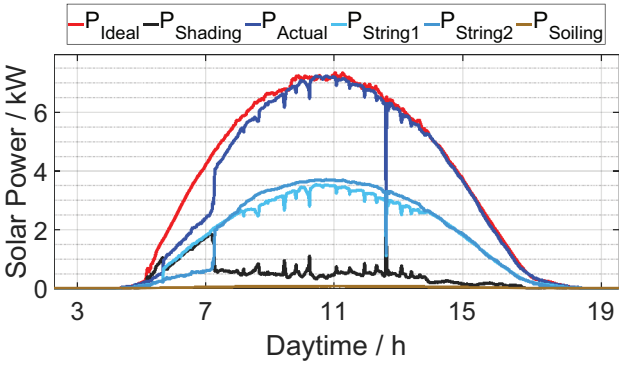
As a final comparison, other machine learning algorithms were also examined for the method, namely support vector regression, gradient boosting LSTM.

Shading data from Array A was used from July 2023 to September 2023. The weather data and measurement data, which were selected using Pearson Feature Selection, were









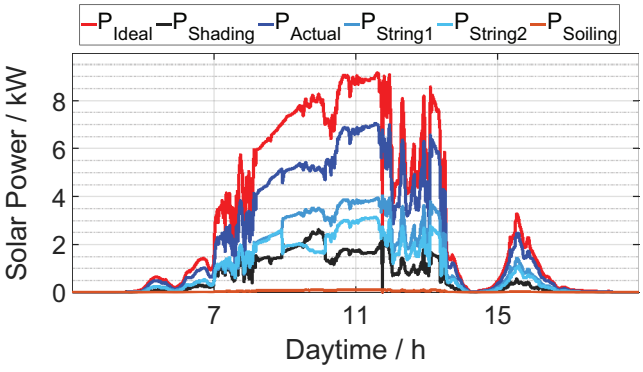


Fig. 18. Various shading structures on the arrays and their impact on the generated power. The calculated portion of the power falling on the shading is shown in black. The theoretical maximum power is reduced by this amount.

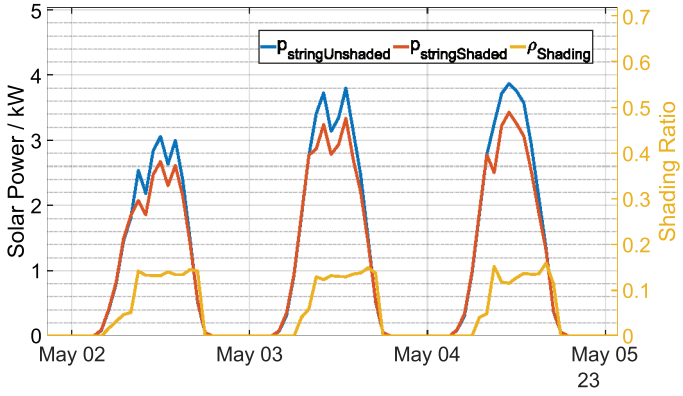


Fig. 19. Three days of recorded DC power from array B. The shaded string visibly delivers less power and the shading ratio increases during this period.

again used as input data. Since a reduction in the output power is to be expected due to the shading, an overestimation of the actual power values can be expected in the predicted values. This can also be seen in Figure 21 after the error boxplot is clearly shifted towards positive errors. The outliers can be completely traced back to individual, particularly large errors in the weather forecast.

The results of the correction are illustrated in Figure 22. It can be seen here that there are only slight differences between the models. The method can therefore contribute to an improvement in the prediction errors regardless of the algorithms considered.

3.5 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 a limitation. It is in the nature of prediction models to over- or underestimate the true value. If the prediction values are now corrected downwards, 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 shows that the true value is underestimated, especially in the early morning hours and in general during the summer months (see Fig. 23). The previous studies therefore presented good comparison months for validation, since the correction model contributes the least to an improvement in the forecast error at these times. A correction in the case of very weak shading does not always lead to an improvement in the RMSE due to the negative MAE, which is particularly visible in the summer months. Due to the fact that meterological data is constantly increasing in accuracy due to ever better satellite systems and more precise weather forecast models, it is to be expected that this will become less and less important.Even the use of endogenous features cannot eliminate shadowing, especially if no shadowing has occurred during the training process. This is also shown by Figure 24, which shows that the addition of lag features does not improve the RMSE in the presence of shadowing.

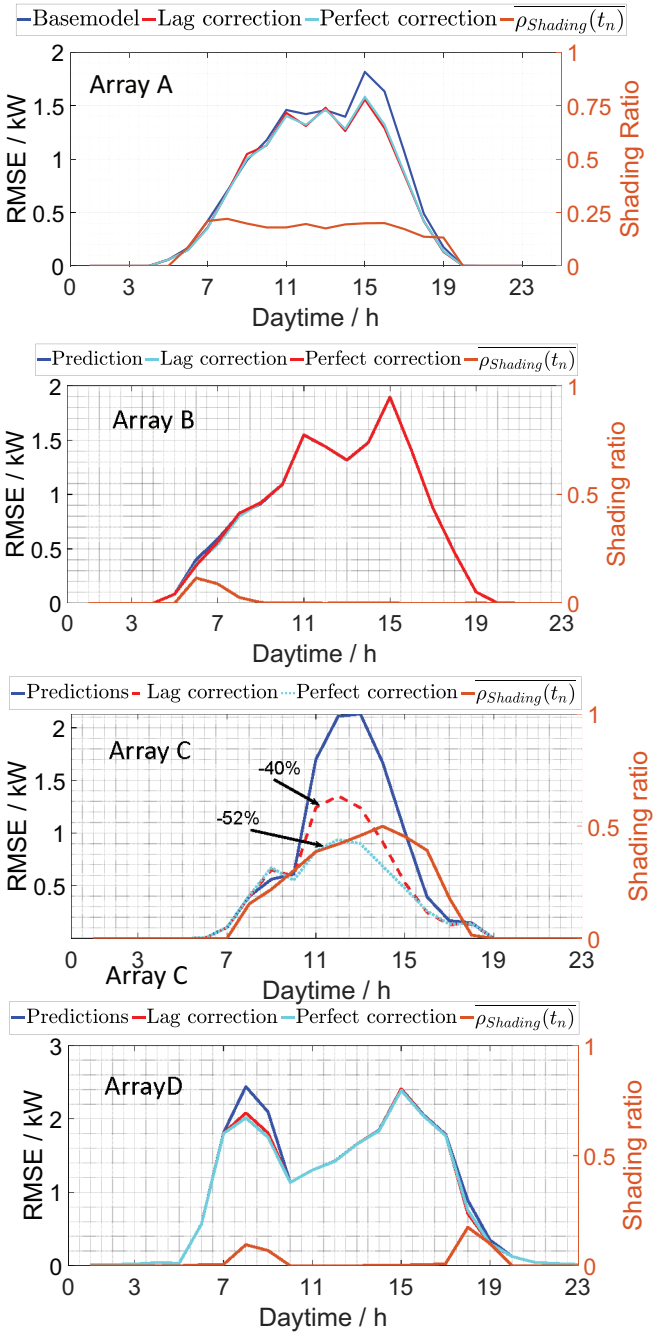


Fig. 20. Improvement of the RMSE through the correction process. The improvement is particularly clear in the case of heavy shading, while only a slight improvement can be achieved in the case of weak shading, such as with PV array A.

Furthermore, it can be concluded that the method still performs its function even if other features are added. The RMSE in the validation data set only minimally decreased from 0.9 kW to 0.88 kW with the addition of the lag feature. It would be particularly interesting to select features in the following, which could transfer the MAE into a positive bias range, but at the same time do not lead to a deterioration of the RMSE.

Table 8. Decrease in RMSE over oberservation time when correction model is applied.

	Array A	Array B	Array D	Array C
RMSE decrease/%	0.51	14.8	40.0	5.02

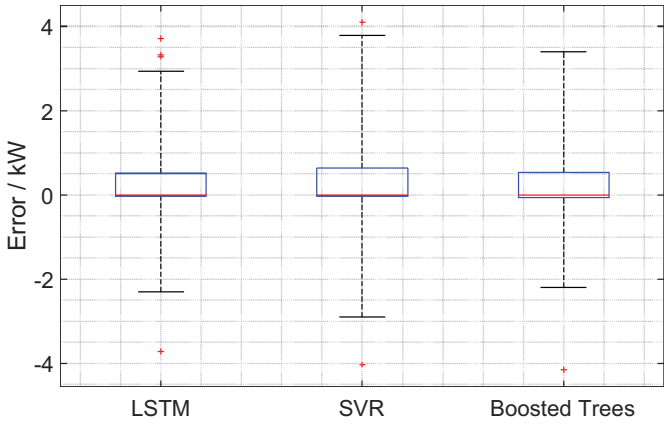


Fig. 21. Error distribution of prediction errors.

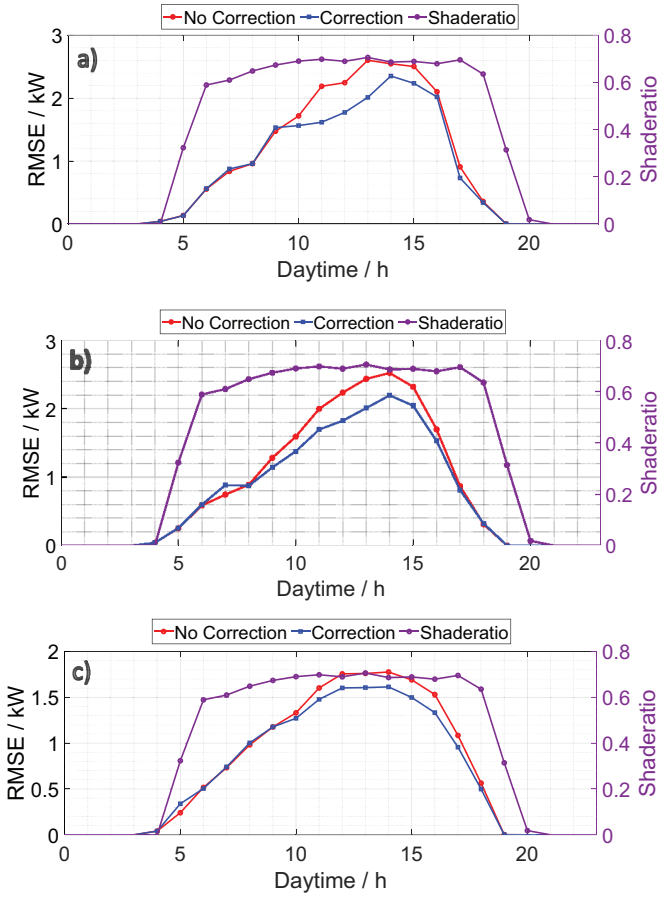


Fig. 22. Reduction in RMSE due to correction for (a) SVR, (b) Gradient Boosted Trees and (c) Neural Net.

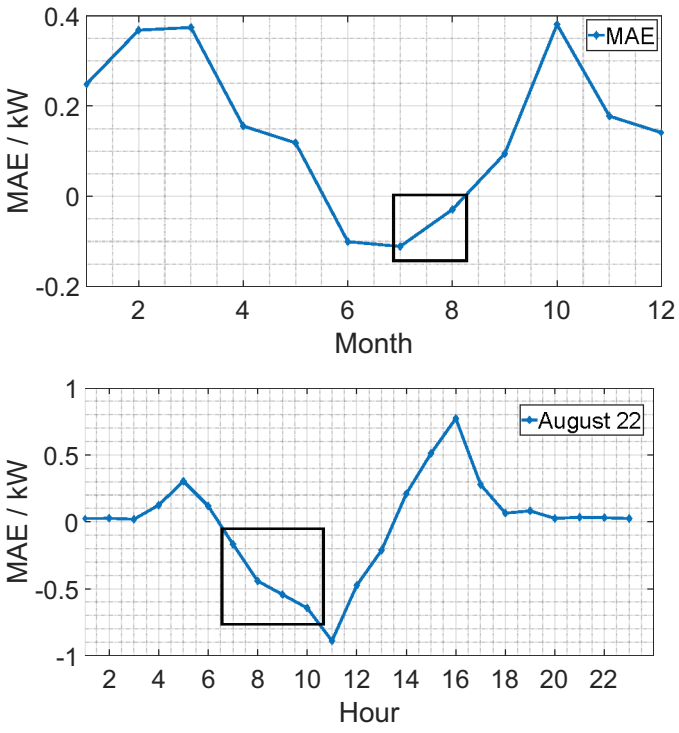


Fig. 23. The curves of the MAE show the bias of the forecast model. The curves indicate time periods in which a correction of the forecast values cannot lead to any improvement as the shading is too small.

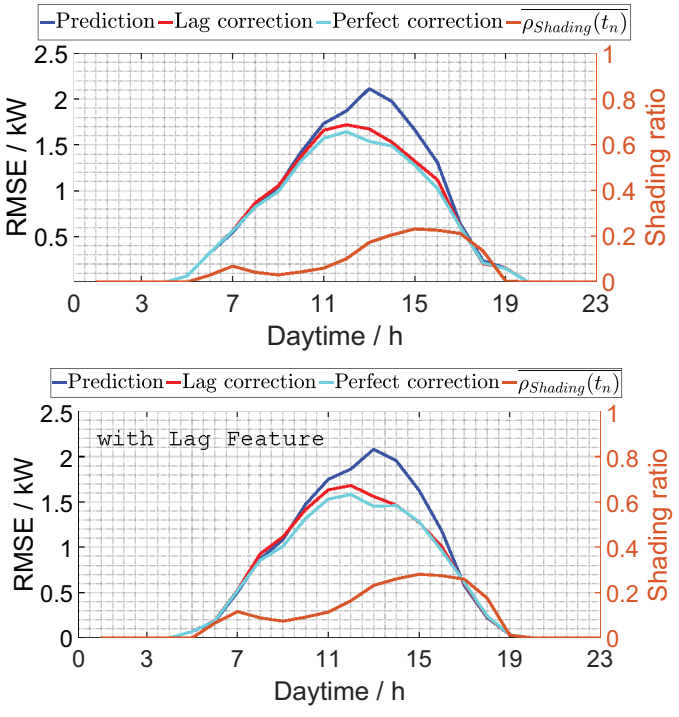


Fig. 24. Comparison of the procedure when past performance data is still taken into account in the forecast model. However, this does not lead to a reduction in the RMSE in the event of shading.

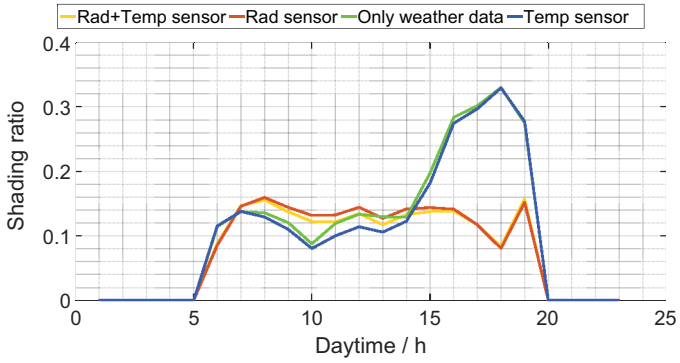


Fig. 25. Comparison of the calculated shading ratio ρ_{Shading} if both irradiation and temperature sensors, only one sensorand weather data or only weather data are used.

In addition, it was examined whether it is possible to replace the sensor data with historical weather data (GHI, temperature). This allows a complete monitoring and correction without additional sensors, but especially the data of the irradiance sensor are necessary to accurately represent the shading ratio (see Fig. 25). However, the temperature sensor could be replaced by the air temperature data without major loss of accuracy for the previous investigations.

4 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 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 post-processing 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 in all shading scenarios. The method was also validated at different orientations and seasons and was able to consistently reduce the error of the forecast models. 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. Array C achieved a 15% improvement, Array B achieved a 5% improvement, and Array A achieved a 2% improvement.

As an outlook, the correction models 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 $n+1$ depending on the forecast horizon. In addition, the limitations of the methods were worked out. Through the detection of shading and soiling, a direct quantification of the losses is possible and can therefore be used to save costs. A condition monitoring approach would be conceivable here. The energy lost from shading and soiling can thus be used to intelligently coordinate maintenance intervals in large solar storage parks. This means that they no longer have to carry out maintenance work at regular intervals, but only when it is needed. At the same time, the presented method is interesting for rooftop systems, since the quantified shading performance can determine whether solar power optimizers can sensibly retrofit their performance. Both areas of application would lead to a reduction in the levelized cost of electricity (LCOE).

Funding

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 Julich (PTJ) and the BMWK.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

The authors do not have permission to share data.

Author contribution statement

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. IEA, Solar pv power generation in the net zero scenario, 2015-2030. Available from: https://www.iea.org/energy-system/renewables/solar-pv#tracking (accessed: 2024-03-26)
2. A. Salah Saidi, Impact of grid-tied photovoltaic systems on voltage stability of Tunisian distribution networks using dynamic reactive power control, Ain Shams Eng. J. **13**, 101537 (2022). https://doi.org/10.1016/j.asej.2021.06.023
3. H. Ye, B. Yang, Y. Han, N. Chen, State-of-the-art solar energy forecasting approaches: critical potentials and challenges, Front. Energy Res. **10**, 875790 (2022). https://doi.org/10.3389/fenrg.2022.875790
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. https://doi.org/10.1109/PES.2011.6039204
5. A. Dairi, F. Harrou, Y. Sun, S. Khadraoui, Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach, Appl. Sci. **10**, 4 (2020). https://doi.org/10.3390/app10238400
6. 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. https://doi.org/10.1109/NEWCAS.2017.8010191

7. 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* (WIP-Renewable Energies (WIP), 2021), pp. 938–945. https://doi.org/10.4229/EUPVSEC20212021-5BO.7.1
8. F. Almonacid, P. Pérez-Higueras, E.F. Fernández, L. Hontoria, A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator, Energy Convers. Manag. **85**, 389 (2014). https://doi.org/10.1016/j.enconman.2014.05.090
9. P. Bacher, H. Madsen, H. Nielsen, Online short-term solar power forecasting, Sol. Energy **83**, 1772 (2009). https://doi.org/10.1016/j.solener.2009.05.016
10. J. Lehmann, C. Koessler, Benchmark of eight commercial solutions for deterministic intra-day solar forecast, EPJ Photovolt. **14**, 15 (2023). https://doi.org/10.1051/epjpv/2023006
11. 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, Renew. Sustain. Energy Rev. **59**, 1307 (2016). https://doi.org/10.1016/j.rser.2016.01.044
12. 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, IEEE J. Photovolt. **11**, 202 (2021). https://doi.org/10.1109/JPHOTOV.2020.3032974
13. A. Alcañiz, D. Grzebyk, H. Ziar, O. Isabella, Trends and gaps in photovoltaic power forecasting with machine learning, Energy Rep. **9**, 447 (2023). https://doi.org/10.1016/j.egyr.2022.11.208
14. Y. Chaibi, M. Malvoni, A. Chouder, M. Boussetta, M. Salhi, Simple and efficient approach to detect and diagnose electrical faults and partial shading in photovoltaic systems, Energy Convers. Manag. **196**, 330 (2019). https://doi.org/10.1016/j.enconman.2019.05.086
15. A. Dolara, G.C. Lazaroiu, S. Leva, G. Manzolini, Experimental investigation of partial shading scenarios on PV (photovoltaic) modules, Energy **55**, 466 (2013). https://doi.org/10.1016/j.energy.2013.04.009
16. R. Ahmad, A.F. Murtaza, H. Ahmed Sher, U. Tabrez Shami, S. Olalekan, An analytical approach to study partial shading effects on PV array supported by literature, Renew. Sustain. Energy Rev. **74**, 721 (2017). https://doi.org/10.1016/j.rser.2017.02.078
17. A. Babatunde, S. Abbasoglu, M. Senol, Analysis of the impact of dust, tilt angle and orientation on performance of PV plants, Renew. Sustain. Energy Rev. **90**, 1017 (2018). https://doi.org/10.1016/j.rser.2018.03.102
18. S. Pareek, R. Dahiya, Enhanced power generation of partial shaded photovoltaic fields by forecasting the interconnection of modules, Energy **95**, 561 (2016). https://doi.org/10.1016/j.energy.2015.12.036
19. M.J. Mayer, G. Gróf, Extensive comparison of physical models for photovoltaic power forecasting, Energy **283**, 116239 (2021). https://doi.org/10.1016/j.apenergy.2020.116239
20. D. Masa-Bote, M. Castillo-Cagigal, E. Matallanas, E. Caamaño-Martín, A. Gutiérrez, F. Monasterio-Huelín, J. Jiménez-Leube, Improving photovoltaics grid integration through short time forecasting and self-consumption, Appl. Energy **125**, 103 (2014). https://doi.org/10.1016/j.apenergy.2014.03.045
21. C.C. Aggarwal, *Neural Networks and Deep Learning* (Springer Cham, 2018). https://doi.org/10.1007/978-3-319-94463-0
22. M. Rana, A. Rahman, Multiple steps ahead solar photovoltaic power forecasting based on univariate machine learning models and data re-sampling, Sustain. Energy Grids Netw. **21**, 100286 (2020). https://doi.org/10.1016/j.segan.2019.100286
23. H. Malki, N. Karayiannis, M. Balasubramanian, *Shortterm electric power load forecasting using feedforward neural networks*, (2024) Vol. 21, pp. 157–167
24. 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. https://doi.org/10.1109/PES.2007.386072
25. S. Srivastava, S. Lessmann, A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data, Sol. Energy **162**, 232 (2018). https://doi.org/10.1016/j.solener.2018.01.005
26. M. Gao, J. Li, F. Hong, D. Long, Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM, Energy **187**, 115838 (2019). https://doi.org/10.1016/j.energy.2019.07.168
27. F. Harrou, F. Kadri, Y. Sun, Forecasting of photovoltaic solar power production using LSTM approach in *Advanced Statistical Modeling, Forecasting, and Fault Detection in Renewable Energy Systems* (IntechOpen, 2020). https://doi.org/10.5772/intechopen.91248
28. C.H. Liu, J.C. Gu, M.T. Yang, A simplified LSTM neural networks for one day-ahead solar power forecasting, IEEE Access **9**, 17174 (2021). https://doi.org/10.1109/ACCESS.2021.3053638
29. M. Awad, R. Khanna, *Support Vector Regression* (Apress, Berkeley, CA, 2015), pp. 67–80. https://doi.org/10.1007/978-1-4302-5990-9_4
30. M. Awad, R. Khanna, *Support Vector Regression* (Apress, Berkeley, CA, 2015), pp. 70–71. https://doi.org/10.1007/978-1-4302-5990-9_4
31. A. Anghel, N. Papandreou, T. Parnell, A. Palma, H. Pozidis, arXiv:1809.04559 (2018)
32. V.K. Ayyadevara, *Gradient Boosting Machine* (Apress, Berkeley, CA, 2018), pp. 117–134. https://doi.org/10.1007/978-1-4842-3564-5_6
33. H. Li, *Machine Learning Methods* (Springer Singapore, 2023). https://doi.org/10.1007/978-981-99-3917-6
34. D.W. (DWD), CDC - Climate Data Center, https://cdc.dwd.de/portal/ (Accessed: 2024-01-01)
35. D.W. (DWD), Index of weather, https://opendata.dwd.de/weather/ (Accessed: 2024-01-01)
36. W. Holmgren, C. Hansen, M. Mikofski, pvlib python: a python package for modeling solar energy systems, J. Open Source Softw. **3**, 884 (2018). https://doi.org/10.21105/joss.00884
37. H. Chen, X. Chang, Photovoltaic power prediction of LSTM model based on Pearson feature selection, in *2021 International Conference on Energy Engineering and Power Systems* (2021), Vol. **7**, pp. 1047–1054. https://doi.org/10.1016/j.egyr.2021.09.167
38. M.F.N. Tanvir Ahmad, S. Sobhan, Comparative Analysis between Single Diode and Double Diode Model of PV Cell: Concentrate Different Parameters Effect on Its Efficiency, J. Power Energy Eng. **4**, 31 (2016). https://doi.org/10.4236/jpee.2016.43004
39. 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, Ain Shams Eng. J. **13**, 101824 (2022). https://doi.org/10.1016/j.asej.2022.101824
40. E. Batzelis, G. Anagnostou, C. Chakraborty, B. Pal, Computation of the Lambert W function in photovoltaic modeling in *ELECTRIMACS 2019. Lecture Notes in Electrical Engineering, Vol. 615* (2020). https://doi.org/10.1007/978-3-030-37161-6_44
41. S. Ghosh, J. Roy, C. Chakraborty, A model to determine soiling, shading and thermal losses from PV yield data, Clean Energy **6**, 372 (2022). https://doi.org/10.1093/ce/zkac014
42. T. Selmi, M. Abdul-Niby, L. Devis, A. Davis, P&O MPPT implementation using MATLAB/Simulink, in *2014 Ninth International Conference on Ecological Vehicles and Renewable Energies (EVER)* (2014), pp. 1–4. https://doi.org/10.1109/EVER.2014.6844065

Cite this article as: Tim Kappler, Anna Sina Starosta, Nina Munzke, Bernhard Schwarz, Marc Hiller, Detection of shading for short-term power forecasting of photovoltaic systems using machine learning techniques, EPJ Photovoltaics **15**,17 (2024)