

Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution

Lennard Visser ^{a,*}, Tarek AlSkaif ^b, Wilfried van Sark ^a

^a Copernicus Institute of Sustainable Development, Utrecht University, Princetonlaan 8a, 3584, CB, Utrecht, the Netherlands

^b Information Technology Group, Wageningen University and Research, Droevendaalsesteeg 4, 6708, PB, Wageningen, the Netherlands

ARTICLE INFO

Article history:

Received 26 July 2021

Received in revised form

4 October 2021

Accepted 31 October 2021

Available online 4 November 2021

Keywords:

Solar forecast

Photovoltaics

Regional solar forecasting

Machine learning

LSTM

Day-ahead markets

PV aggregation

ABSTRACT

Accurate forecasts of the power production of distributed photovoltaic (PV) systems are essential to support grid operation and enable a high PV penetration rate in the electricity grid. In this study, we analyse the performance of 12 different models that forecast the day-ahead power production in agreement with market conditions. These models include regression, support vector regression, ensemble learning, deep learning and physical based techniques. In addition, we examine the effect of aggregating multiple PV systems with a varying inter-system distance on the forecast model performance. The models are evaluated both on their technical and economic performance. From a technical perspective, the results show a positive effect from both an increasing inter-system distance and a larger sized PV fleet on the model performance, which was not the case for the economic assessment. Furthermore, the ensemble and deep learning models perform better than the alternatives from a technical point of view. For the economic assessment, the results indicate the superiority of the physical based model, followed by the deep learning models. Lastly, our findings show the importance of considering the user's objective when assessing solar power forecast models.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Vast decreasing costs associated to solar photovoltaic (PV) systems have increased the competitiveness of PV systems to other power generation technologies. This has resulted in a surge of the global installed capacity of PV systems in recent years. The global installed capacity is estimated at 739 GWp in 2020 and is expected to grow up to 1,800 GWp in 2025, making it the most rapidly growing source for power generation [1]. Moreover, the installation of distributed PV systems is expected to account for about half of the projected capacity growth [2]. This strong growth of (distributed) PV capacity will affect the electricity supply and subsequently the power dispatch.

The integration of variable renewable energy sources, i.e. solar PV, in the electricity grid poses challenges to grid operators in maintaining grid stability [3]. Moreover, the power output of PV systems may drop or increase by almost 70% within 1 min and can differ by more than 75% from 1 h to the following [4]. These

fluctuations in the power output cause imbalances in the grid, which affect the operating frequency and (local) voltage if not managed properly by balancing reserves [5]. Adequate scheduling of supply and demand can limit the occurrence of imbalances and lower the need for these often carbon intensive, inefficient and costly reserves. As a result, accurate forecasts of the PV power output are required for scheduling the dispatch of power production by other technologies. This includes distributed PV systems, as the power output of large amounts of distributed PV will affect the dispatch. Reliable forecasts of PV power output therefore have the potential to support stable grid operation when the PV penetration rate increases, while limiting the need for balancing reserves. Subsequently, these forecasts are in the first place valuable for grid operators as it reduces the integration costs associated with PV systems [6]. Besides, accurate forecasts of power production of distributed PV can also support flexibility services of aggregated systems in the grid if managed by, e.g., an aggregator [7]. Consequently, accurate forecasts are identified as a requirement to facilitate a high PV penetration level [8].

Fluctuations in PV power output are caused by variations in irradiance that is received at the plane of the PV array [3]. This observed irradiance can be described by a deterministic and a

* Corresponding author.

E-mail addresses: l.r.visser@uu.nl (L. Visser), tarek.alskaif@wur.nl (T. AlSkaif), W.G.J.H.M.vanSark@uu.nl (W. van Sark).

stochastic component [9]. Here, the deterministic is described by the movement of the earth with respect to the sun and can therefore be estimated very accurately at every location and at any time of interest. Moreover, this component can be expressed by the Clear Sky Irradiance (CSI), while considering the panel orientation, i.e. tilt and azimuth angle. On the other hand, the stochastic component is rather difficult to predict as it is affected by the chaotic nature of the atmosphere, where the formation, evolution and motion of clouds affect the Global Horizontal Irradiance (GHI) received at the surface. In addition, the power output of a PV system is also directly affected by the ambient temperature and wind speed [3].

In recent years, the development and application of solar forecasting methods have received an increasing attention from researchers, grid operators and other users involved in electricity markets [10]. In solar forecasting, the objective of the user who is applying the results of the forecast defines the forecast requirements. These requirements determine what forecast methods generates the most accurate results [11]. In many electricity markets around the world, the majority of electricity is traded in the day-ahead market (DAM) and consequently, the dispatch of generators is largely decided in this DAM. Hence, operators of solar PV systems, in particular traders such as utilities and aggregators that operate many and/or large-sized systems have a high interest in PV power forecast models that can predict the day-ahead production in order to optimize their market bids.

As a result, the market requirements related to the DAM should be considered when developing day-ahead PV power forecast models. In many countries, including the Netherlands, single value electricity volume and price bids are accepted until noon on day T , thereafter the market is cleared and a day-ahead price is set for each trading block on day $T + 1$ [12]. Acknowledging this 12 h lead time for the purpose of PV power forecasting, Numerical Weather Prediction (NWP) is preferred over other methods including satellite and all-sky imaging. Moreover, statistical models are often applied to improve the forecast performance by post-processing NWPs [6].

Previous studies have already shown the success of statistical models that post-process NWPs to forecast the day-ahead PV power output. Different studies test a selection of models that include a variety of predictor variables [6,13]. In more recent years, the attention of researchers appears to be shifting towards the development and application of more advanced models. Most of these approaches include machine learning and deep learning models [14,15], ensemble models [16,17] and/or hybrid models [18,19]. However, these more advanced forecast models come at a cost of the difficulty in interpretation, require extensive hyperparameter tuning and are therefore more complex to adopt. In addition, they are often not compared to one another. Besides, proposed novel methods often gain a marginal performance improvement and are commonly executed using a single case study so that it remains uncertain if the found improvements are universal. This argument is supported by the fact that the forecast accuracy is greatly affected by the local climate conditions and therefore the location of interest [3]. The lack of an extensive comparison raises the question if these advanced models are worth developing and adopting for different users.

Furthermore, due to the limited information that is disclosed, the applicability of these models by interested parties remains in many cases not clear. For example, users that are interested to adopt a forecast model that can support bidding into the DAM are bounded to gate closure time and trading blocks. This sets several requirements to the operationalization of forecast models, including the temporal resolution, lead time, time horizon and

available information, which is essential to interested parties as discussed in Ref. [20]. For example, in Ref. [21] day-ahead refers to a 0–24 h ahead forecast, which does not meet the operational conditions of the actual DAM [12]. In other studies it is unclear if the application of these models meet the market conditions [14,16] or it is unknown to what extent the obtained results can be achieved when adopted in real-time [15]. Exceptions to this are found in few studies, where the evaluated models meet the operational conditions of the DAM [22–24] or intraday market [25]. In addition, these models are commonly examined by means of technical performance metrics including the Mean Bias Error (MBE), Mean Absolute Error (MAE), Root Mean Square Error and/or the Skill Score [26]. Nevertheless, users willing to adopt the forecast models may be more interested in the economic performance of the proposed forecast models, as also outlined by Antonanzas et al. [27]. Although the selection of adequate error metrics to evaluate the performance of models in the field of solar forecasting is already an element of discussion [28], in literature there is limited attention to the potential discrepancy between technical and economic metrics. Moreover, an economic examination is only found in Ref. [24], where three different day-ahead forecast models are compared on several performance indicators including the potential profits in the DAM for a case study in Spain.

Another limitation of the above-mentioned studies is related to the testing conditions, which is in many cases either a single site [14,15,18,19] or a single aggregation of a number of systems [16,17,22]. In practise the performance of solar forecast models is found to be correlated with the spatial distribution and/or the number of systems included, since this smoothens out the PV power production profiles [29]. The effect of increasing the spatial area on the performance of solar irradiance forecast models was studied by Lorenz et al. [30], with a coarse spatial resolution within an area of 200×120 km. This raises the question how an increased spatial area relates to the performance of different PV power forecast models and how this is effected by the number of the systems, i.e. system density. To the best of the authors' knowledge, similar studies that consider the effect of a varying spatial resolution on the performance of PV power forecast models are missing in current literature. A summary of the literature is presented in Table 1.

The goal of this paper is to investigate the performance of a wide variety of solar power forecast models. As motivated above, grid operators require reliable forecasts to guarantee adequate grid management. On the other side, electricity traders, such as utilities and aggregators, have a high interest in these models to maximize the economic revenue of the PV systems they operate. Besides, utilities and aggregators may also deploy the models to offer flexibility services to grid operators. In this study the model performance is evaluated with both technical and economic metrics to provide insights into the effect that the user's objective has on the best performing models, and subsequently to uncover potential contradictory interests between different users. Lastly, the effect of aggregating PV systems with varying inter-system distances is assessed on the model performance for both metrics. As a result, with this study we aim to lower potential barriers for grid integration of PV systems by empowering PV system operators, electricity traders, aggregators, utilities, grid operators and other users in the development and implementation of operational PV power forecast models.

The contributions of the paper can be summarized as follows:

1. First of all, this study provides an extensive comparison of 12 models that forecast the day-ahead PV power production while considering the DAM requirements. These models include

Table 1
Overview of relevant literature.

Reference	Forecast target (Power/GHI)	Forecast site	Evaluation metrics (Technical/economic)	Meet market requirements	Considers impact of spatial area
[14]	Power	Single site	Technical	No	No
[15]	Power	Single site	Technical	No	No
[16]	Power	Single aggregation	Technical	No	No
[17]	Power	Single aggregation	Technical	DAM	No
[18]	Power	Single site	Technical	No	No
[19]	Power	Single site	Technical	No	No
[21]	Power	Single site	Technical	No	No
[22]	Power	Single aggregation	Technical	DAM	No
[23]	Power	Single site and multiple aggregation levels	Technical	DAM	No
[24]	Power	Single site	Both	DAM	No
[25]	GHI	Single site	Technical	Intraday	No
[30]	GHI	Single site and multiple aggregation levels	Technical	No	Yes

indirect and direct approaches, including a PV model and several regression and machine learning based models that rely on a varying degree of complexity.

- Subsequently, we evaluate the performance of these models on both technical and economic metrics. As a result, this study shows the effect of the user's objective on the best performing model as well as the trade-off between the objectives of different users, i.e. minimizing the forecast error from a technical perspective or maximizing the economic value.
- Furthermore, this study evaluates the impact of aggregating PV systems on the technical and economic performance of solar power forecast models and hence bridges the current gap in literature.
- Lastly, we decouple the value of aggregation in the context of the model performance into a spatial and density effect by considering different numbers of PV systems within varying spatial areas.

The paper is further organized as follows. In the next section the methods including the performance metrics are discussed (Section 2). In Section 3 the data input is described and assessed. The results of this study are discussed in Section 4. Finally, the conclusions are presented in Section 5.

2. Methods

This section explains the methods adopted in this study, starting with a description of the 12 forecast models. Next, additional steps in the methods including pre- and post-processing steps are discussed. Finally, the metrics adopted to examine the performance of the forecast models are presented. An overview of the methods is provided in Fig. 1. The data used in this study is discussed in more detail in Section 3 (see Table 2).

2.1. Forecast models

Most forecast models considered in this study are supervised learning models. This means that the model is trained on a subset of the dataset holding observations of predictor and target variables. An overview of the predictor variables is presented in Table 2, whereas the target variable presents the PV power output. Based on the observed relationship, the model identifies the parameter settings [13]. In addition to the supervised learning models, a PV model is considered that simulates the power output of a PV system using its characteristics and few weather variables. Finally, two benchmark models are included, where forecasted values are based on historic observations [31]. Furthermore, all simulations in this

study are conducted in Python 3.6.5 [32]. The experiments are supported by the PVlib [33], Scikit-learn [34] and Tensorflow [35] libraries.

2.1.1. Linear regression

The first model that is considered in this study is a Multivariate Linear Regression (MLR) model. This is a simple and widely applied model in solar forecasting. The MLR model forecasts the PV power output by considering a linear relationship between a matrix (\mathbf{X}) of (n) predictor and (m) timestamps and the power output (\hat{y}^{mlr}), which is described by a vector of regression coefficients β :

$$\hat{y}^{mlr} = \beta \mathbf{X} + \varepsilon, \quad (1)$$

where ε presents the uncertainty. The coefficients are found by minimizing differences between the actual (y) and predicted (\hat{y}^{mlr}) power output:

$$\min (y - \hat{y}^{mlr}). \quad (2)$$

The Least Absolute Shrinkage and Selection Operator (LASSO) model is a variation of the linear regression model in Eq. (1), where the number of included input variables is reduced. This is done by penalizing the regression coefficients using a L_1 norm function, which is the sum of the absolute coefficients. The penalty forces some of the coefficients that have a minor contribution to the model to reduce to zero (see Eq. (3)).

$$\min (y - \hat{y}^{lasso}), s.t. \|\beta\|_1 \leq t, \quad (3)$$

where t is a hyperparameter to be defined by the user and sets the upper bound for the sum of the coefficients $\|\beta\|_1$. To boost the model performance, the value of this and other hyperparameters is optimized by means of a grid search (see Section 2.2).

Another popular regression model for forecasting is Seasonal Auto-Regressive Integrated Moving Average with exogenous input variables (SARIMAX). Different from the models discussed above, SARIMAX is a time-series model that requires a temporal ordering in the data. The SARIMAX model is based on an Auto-Regressive Integrated Moving Average (ARIMA) model that integrates an auto-regressive (AR) and moving average (MA) model to forecast time-series. In the SARIMAX model, external variables are considered and a periodicity factor is added to the ARIMA model. This periodicity factor enables the model to observe a natural seasonal cycle, e.g. the solar diurnal cycle. The SARIMAX model is mathematically expressed as [36]:

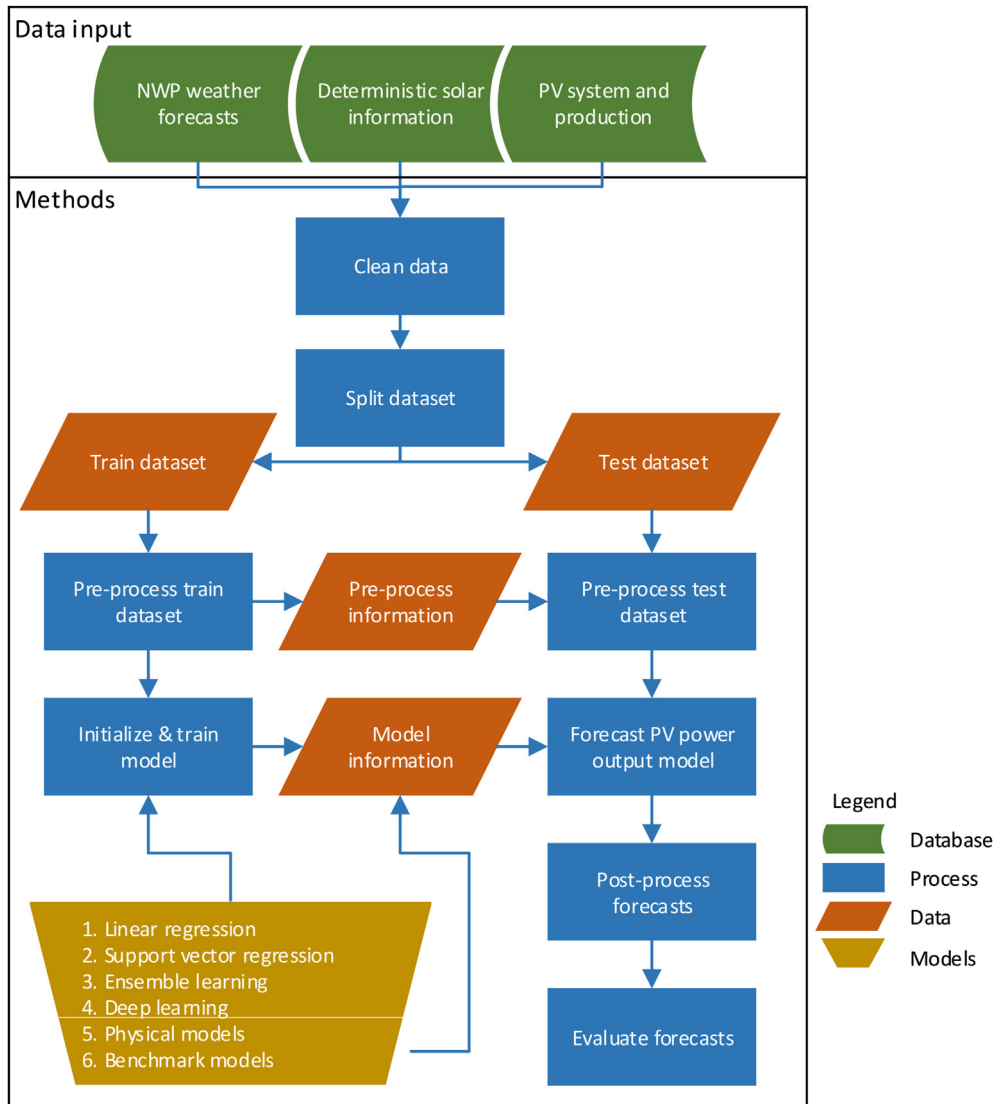


Fig. 1. Overview of the methods. Details can be found in the text. Note that the six types of models lead to 12 forecast models used in total.

$$\varphi_p(B) \Phi_p(B^s) \nabla^d \nabla_s^D \hat{y}^{sarimax} = \theta_q(B) \Theta_Q(B^s) \beta X \varepsilon, \quad (4)$$

where $\hat{y}^{sarimax}$ is the target variable (i.e. PV power output). The ARIMA operators are characterized by $\varphi_p(B)$, ∇^d and $\theta_q(B)$ and respectively present the AR polynomial, difference operator and MA polynomial. The seasonal AR polynomial, difference operator and MA polynomial are defined by $\Phi_p(B^s)$, ∇_s^D and $\Theta_Q(B^s)$.

2.1.2. Support vector regression

Support Vector Regression (SVR) is a kernel based forecast technique that evolved from the Support Vector Machine (SVM), which is typically used for classification problems. Similar to SVM, SVR constructs a set of hyperplanes in a multidimensional space in order to describe the relationship between predictor and target variables [37]. In this study we first consider an SVR model with a linear kernel (LSVR), which is described by Eq. (5).

$$\hat{y}^{svr} = \sum_{i=1}^m (\alpha_i - \alpha_i^*) K(x_i, x_j) + b, \quad (5)$$

where the target variable (i.e. PV power output) is \hat{y}^{svr} , $\alpha_i - \alpha_i^*$ is the

Table 2

Overview of the predictor variables.

	Variable	Unit
1.	Air Pressure	Pa
2.	Mean Sea Level Pressure	Pa
3.	Ambient Temperature	°C
4.	Dewpoint Temperature	°C
5.	Total Precipitation	M
6.	Surface Solar Radiation Downwards	W/m ²
7.	Zonal Wind Speed at 10m height	m/s
8.	Meridional Wind Speed at 10m height	m/s
9.	Total Cloud Cover	—
10.	Low Cloud Cover	—
11.	Medium Cloud Cover	—
12.	High Cloud Cover	—
13.	Sine of Hour of Day	—
14.	Cosine of Hour of Day	—
15.	Clear Sky Irradiance	W/m ²
16.	Solar Zenith Angle	°
17.	Solar Azimuth Angle	°

difference between the Lagrange multipliers and b the bias. The kernel function is denoted with $K(x, x_i)$, in case of a linear SVR:

$$K(x_i, x_j) = x_i^T x_j + c, \quad (6)$$

where c is a constant.

In addition to the linear SVR model, we consider a nonlinear kernel variant of SVR denoted as KSVR. By replacing the linear kernel function in Eq. (5) with a nonlinear kernel, a more complex relation between the predictor and target variables can be considered as the problem can now be described in a higher dimensional space. In this study we consider a Gaussian Radial Basis Function, which is mathematically expressed as [13]:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad (7)$$

where σ is a hyperparameter set by the user.

2.1.3. Ensemble learning

In this study we include two different ensemble learning based forecast algorithms: a Random Forest regression (RF), and a Gradient Boosting regression (GB). Both RF and GB consist of a number of trees each made up of t_n layers and 2^{t_n} decision nodes. At each of these decision nodes, a test to any of the input variables is applied. The test outcome determines what sub-branch of the tree is selected until a leaf node is reached, where a prediction is made. Each decision node in a tree is constructed by randomly sampling a few variables. From these variables the most valuable split variable and value is chosen by optimizing on a loss function, i.e. the Mean Squared Error (MSE) [13]. In RF the decision trees are created independently from one another based on a bootstrap sample of the training dataset. Consequently, each tree is trained on the bootstrap sample, while considering the loss function. The mean of the output of all constructed trees forms the forecasted value [38].

Contrary to RF, in GB the individual decision trees depend on each other and are built considering the entire dataset. In GB each subsequent tree is built and stacked upon its predecessor, in order to diminish the obtained forecast or estimation error (i.e., boosting). Moreover, each new built tree takes the residual of the previous learner to decrease the errors of the model. The first built tree considers the deviation of the mean as input residual. The generated model output is equal to the sum of the output of all trees [13].

2.1.4. Deep learning

Furthermore, this study includes two different types of Neural Networks (NNs). The architectural design of NNs allows to build complex nonlinear relationships between the predictor and target variables, without assuming any form of relationship between these variables. A NN consists of an input layer that receives the input data, an output layer that yields the predictions and a pre-defined number of hidden layers in between that transform the input data. Each of these layers are made up of a number of nodes, in which the data transformation takes place [39].

The first NN architecture we consider is a feed-forward Artificial Neural Network (ANN). In this ANN, data is transmitted in a uni-directional manner, i.e. information is passed from each node in one layer to every node in the consecutive layer. Within every node the data is transformed according to the activation function, weight and bias [39]. This process is described by equation (8).

$$L_{i+1} = \alpha(W_i * L_i + \mathbf{b}_i), \quad (8)$$

where \mathbf{W}_i presents the weights of each input node per processing node \mathbf{b}_i holds a vector of the bias per processing node present for layer i . The activation function α determines if the fed data triggers the operations in the node, and L indicates the layer that consists of

multiple nodes. In this study several activation functions are tested including Sigmoid, Rectified Linear Unit, Softmax and Hyperbolic Tangent [39]. The model output, in this study \hat{y}^{ann} , is calculated in the output layer as in equation (8). The output node in the output layer is fed with information of all nodes in the last hidden layer.

Secondly, we consider a Recurrent Neural Network (RNN) architecture in the form of a Long Short-Term Memory (LSTM). In contrast to ANN, where samples are fed to the model separately, in an RNN a sequence of samples is given to the model at once. The RNN model then processes each sample individually, while in every layer the output of the preceding sample is passed on as additional input [40].

LSTM is a special type of RNN where multiple preceding outputs of nodes in all layers including the output layer can be remembered. To guide the conservation of information, three steps are added in the form of gates to the data transformation steps taken in the nodes. These include a forget, input and output gate. These gates respectively determine what information is removed, stored and provided as an input to the following sample. Similar to the operations in the node, these gates utilize an activation function, weight and bias to generate its output [40].

Apart from the difference in the node operation and the subsequent ability to remember information, an LSTM model operates similarly to an ANN. Consequently, the forecast \hat{y}^{lstm} can be described similarly to ANN by equation (8), where the inserted information from the previous layer (L_i) is accompanied by information from the previous samples.

2.1.5. Physical model

We include a PV model that obtains a power forecast by considering the NWP's of the GHI, wind speed, surface pressure, and ambient and dew point temperature. This PV model forecast is also referred to as an indirect forecast approach as it is computed in three consecutive steps. Firstly, the GHI is decomposed in the Direct Normal Irradiance and Diffuse Horizontal Irradiance with the DIRINT model [41]. Secondly, the Perez model is used to transpose the irradiance components in their in-plane counterparts [42]. And thirdly, the power output of the PV system is simulated through a PV model [33]. An advantage of the PV model to the models discussed above is that the former does not require any training data. This advantage comes at a cost of the additional information that is required regarding the characteristics of the PV system.

2.1.6. Benchmark models

Finally, in order to provide context with respect to the accuracy obtained by the forecast models presented above, we consider two benchmark models. Firstly, we include a Diurnal Persistence (DP) model where the PV forecast equals the observations for the most recent available daily series. As PV production values are only available up to noon at day T , we consider the production values for $T - 1$, see Eq. (9). In addition, we include a Clear Sky Persistence (CSP) model that presumes that the current condition, which is defined by the clearness index (k^*), will last in the future. As a result, the model accounts for solar irradiance variations due to the diurnal and seasonal cycles (see Eq. (10)). The k^* is calculated for the most recent observed PV power output value, which is noon (see Eq. (11)).

$$\hat{y}^{dp}(t) = y(t - 48), \quad (9)$$

where the time t ranges from $h = 0$ to $h = 24$.

$$\hat{y}^{csp}(t) = y_{cs}(t)k_t^*, \quad (10)$$

where y_{cs} is the expected PV production in case of a clear sky and:

$$k_t^* = \frac{y(t_{h=12})}{y_{cs}(t_{h=12})}. \quad (11)$$

2.2. Pre-processing

The variables considered in this study (see section 3) are collected in a matrix \mathbf{X} with dimension $m \times n$. Here, m presents the number of hours considered in this study and n is the number of predictor variables. Next, the dataset is cleaned, which includes: (i) the power output during timestamps that fall in the night are set to 0, (ii) if present, negative values are labeled as NaN and lastly (iii) NaN values are filled by linear interpolation of the clear-sky index for a maximum of 3 consecutive hours. In case a day holds more than 3 NaN values, the entire day is removed from the dataset. Furthermore, in case of time-series based models like SARIMAX and LSTM, nighttime values are kept. This is required as these models consider a dataset as an equally spaced sequence of observations and therefore demand a successive equally spaced dataset. For the remaining models, all nighttime values are excluded from the dataset, as this is found to improve the model performance and simultaneously reduces the computational requirements. To create a fair comparison between all the models, nighttime values are excluded for evaluation.

Subsequently, the dataset is split in a training and testing set, respectively $\mathbf{X}_{\text{train}}$ and \mathbf{X}_{test} . The training set concerns 2 years, from February 2014 until February 2016. The test set runs for one year, until February 2017. The training set is firstly used for hyperparameter tuning, where the best settings according to a numerical score are found by means of a grid search. A k-fold cross-validation method is used, where the training set is divided in $k = 8$ sequential folds. An 8-fold grid search considering a two year training period (see section 3) results in 8 iterations where the model is trained on 21 months and evaluated on one quarter of a year. The performance of each model is for all hyperparameter configurations tested on the MSE, for all 8-folds. Due to computational constraints the grid search is executed for 10 PV systems, which are, based on the system characteristics (i.e. location, size, orientation, tilt), deemed to be representative for the entire PV fleet. Next, the optimal settings for each system and model are weighted to obtain the best overall hyperparameter setting configuration. This configuration is then used in the execution of all PV power output forecasts.

Finally, the predictor and target variables are all pre-processed before being used in the forecast models. Moreover, the predictor variables are scaled according to the maximum observed value during the training period. Similar to the predictor variables, the target variable PV power output is normalized according to the PV system capacity. In order to produce model performance results that are independent of the PV system size, the normalized values are also considered in the evaluation.

2.3. Post-processing

After the forecast models are executed, the predicted values are post-processed (see Fig. 1). In this post-processing step, all forecasted values are compared and limited to the theoretical maximum production. Since at an hourly time resolution potential cloud enhancement effects are undetectable [43], the actual yield should not exceed the maximum production of the PV system that is obtained under clear sky conditions. Moreover, the production of a PV system under these circumstances can be estimated by using a PV model that is fed with CSI values and depends on the PV system

characteristics, location and time (see section 2.1.5).

2.4. Aggregating PV systems

To evaluate the impact of aggregating PV systems on the technical and economic performance of the forecast models, a multitude of sets of aggregated systems is considered by the forecast models. The power output of these aggregated systems is described:

$$y_t = \frac{1}{N} \sum_{s=1}^N y_{t,s}, \quad (12)$$

where s is an index for each system in the number of systems considered (N). Note, since the constructed aggregated production dataset requires all systems to have data available, the performance results obtained for single and multi-site forecasting cannot be compared directly.

Subsequently, in the analysis of the results the impact of aggregating PV systems is categorized according to the spatial distribution of the PV systems, which includes the spatial area as well as the density of the aggregated systems. The latter is considered by labelling each set with the number of PV systems included N in the aggregation, namely $N = 1, 2, 3, 5, 10, 25, 50$ or 100 systems. Next, we label each of these sets using the inter-system distance D , with $D = 0.5, 1, 2, 5, 10, 15, 20, 25$ or 30 km. Here, we denote a combination of a number of systems and an observed inter-system distance as a category. For instance, $N = 5$ PV systems with an inter-system distance of $D = 10$ km is considered as one category. Subsequently, within each category ten different sets are included in order to identify trends and minimize the impact of potential outliers. The sets within each group are random.

The inter-system distance D is defined as the diameter of the smallest circle that encloses the area holding all PV systems. We construct the smallest enclosing circle with the randomized incremental construction algorithm [44]. Here, let l_1, \dots, l_N be the location of the PV systems in a certain set N and C_i the smallest enclosing circle of i points (i.e. PV systems) l_1, \dots, l_N . To establish this smallest enclosing circle we first create a circle where the center is exactly in the middle of l_1 and l_2 (see Fig. 2a). If N exceeds 2, we want to add an additional location l_3 , which will be located either within the current circle C_i such that $C_3 = C_2$ or outside (see Fig. 2b). In the latter case, a new larger circle C_3 is constructed, where l_3 is considered a boundary point (see Fig. 2c). Next, the radius of the circle is reduced until another location meets the boundary of the circle. Hereafter, the origin of the circle is moved towards the centre of these two locations until either any third location hits the boundary or to the point where the circle diameter is equal to the distance between these two known locations (see Fig. 2d). This process is repeated until all points in N are included in the circle C_{l_1, \dots, l_N} .

2.5. Performance metrics

The last step in the methods considers evaluation of the forecasts (see Fig. 1). To examine the performance of the proposed forecast models we firstly consider the technical evaluation metrics the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Bias Error (MBE) [6]. These can be referred to as a technical evaluation of the model performance as it pronounces the magnitude of the forecast error in terms of the quantity.

In addition to the technical performance metrics described above, we include an economic metric that measures the performance of the forecast models. This indicator considers the net revenues made when the generated forecast of a model was used

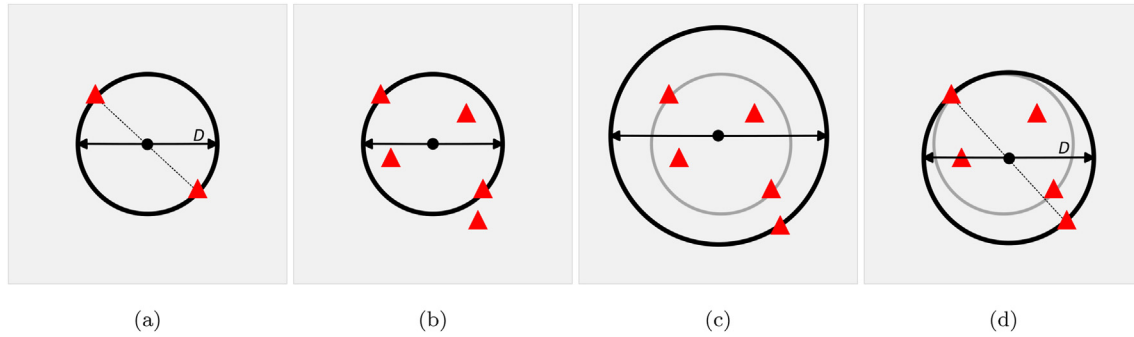


Fig. 2. A visual example of the process followed to define the inter-system distance (D), where the red triangles present the spatial distribution of the selected PV systems in case of $N = 5$.

for bidding in the DAM, where the production of electricity per trading block is compensated with the DAM Price (DAMP). To include the economic consequences of a forecast error, the economic indicator includes the financial penalty that a bidding party is required to pay when the day-ahead bid deviates from the actual power production. This penalty, i.e. Settlement Price (SP), is raised in case of a production deficit or surplus and must be paid by the responsible party to the local Transmission System Operator (TSO), which is TenneT in the Netherlands [45]. The sum of the penalties depends on the grid imbalance at the respective moment in time and therefore depends on all participants in the DAM [46]. The distribution of the DAMP [47] and SPs [45] during the study period can be found in Fig. 3, where the SP is split into a SP in case of a surplus (SP_s) and deficit (SP_d). During most of the time the bidding party is paid when the actual production exceeds the forecast and vice versa. Moreover, the compensation of the overshoot is usually lower than the DAMP and the costs of the shortage higher than the revenues made on the DAM (see Fig. 3). This translates as a price incentive to the bidding party to forecast the production as accurately as possible. However, due to major imbalances, the mentioned logic does not always persist, which may lead to unexpected revenues from forecast errors [45]. In extreme cases the logic can even be reversed such that the bidding party is paid by the TSO for their production shortage and vice versa [45]. This occurs in case of a negative SP, see Fig. 3.

In this study, the economic indicator is represented by the Economic Revenues (ER). The ER is calculated by Eq. (13): the first term describes the initial revenues made on the DAM and the second term presents the net imbalance costs due to the observed forecast error. In contrast with the technical performance metric, the ER evaluates the economic consequences of forecast errors. Moreover, a large forecast error could have little economic

consequences at times the SPs approaches the DAMP.

$$ER = \sum_{t=0}^{m-1} DAMP * \hat{y} + SP_s * (y - \hat{y}) \quad \text{if } \hat{y} < y, \quad (13)$$

$$ER = \sum_{t=0}^{m-1} DAMP * \hat{y} + SP_d * (y - \hat{y}) \quad \text{if } \hat{y} > y.$$

3. Data input and analysis

3.1. PV data

In the present study the PV power output of a total of 152 PV systems is considered. All these systems are located in the province of Utrecht, the Netherlands and cover an area of 38 by 54 km (see Fig. 4). The PV systems are all rooftop mounted and differ in their characteristics, with an installed capacity ranging from 0.5 to 6.8 kWp. The system orientation varies between 86° and 285° (south is 180°), with one exception of a single system oriented towards the north (5°). The system tilt ranges from 2° to 60° . More details on the distribution of the PV systems according to these characteristics and the observed average annual system yield for three years from January 2014 are depicted in Fig. 5. Among others, this figure shows that the capacity of most systems is lower than 3 kWp. Besides, most systems are tilted and oriented towards the south. Systems due south are found to have an average energy yield of 1050 kWh/kWp, whereas yield of all systems varies between 660 and

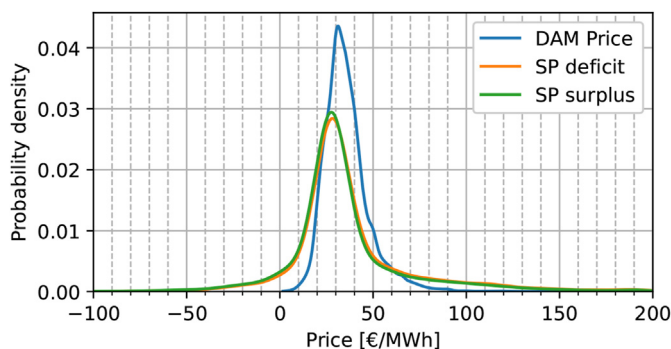


Fig. 3. Distribution of the observed DAMP, and SP in case of a production deficit or surplus for 2016 and 2017.

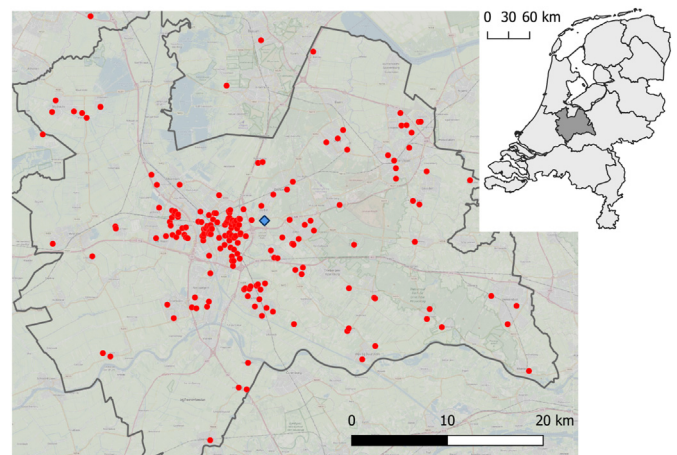


Fig. 4. Geographical distribution of the PV-systems (red circles) in the province Utrecht, the Netherlands.

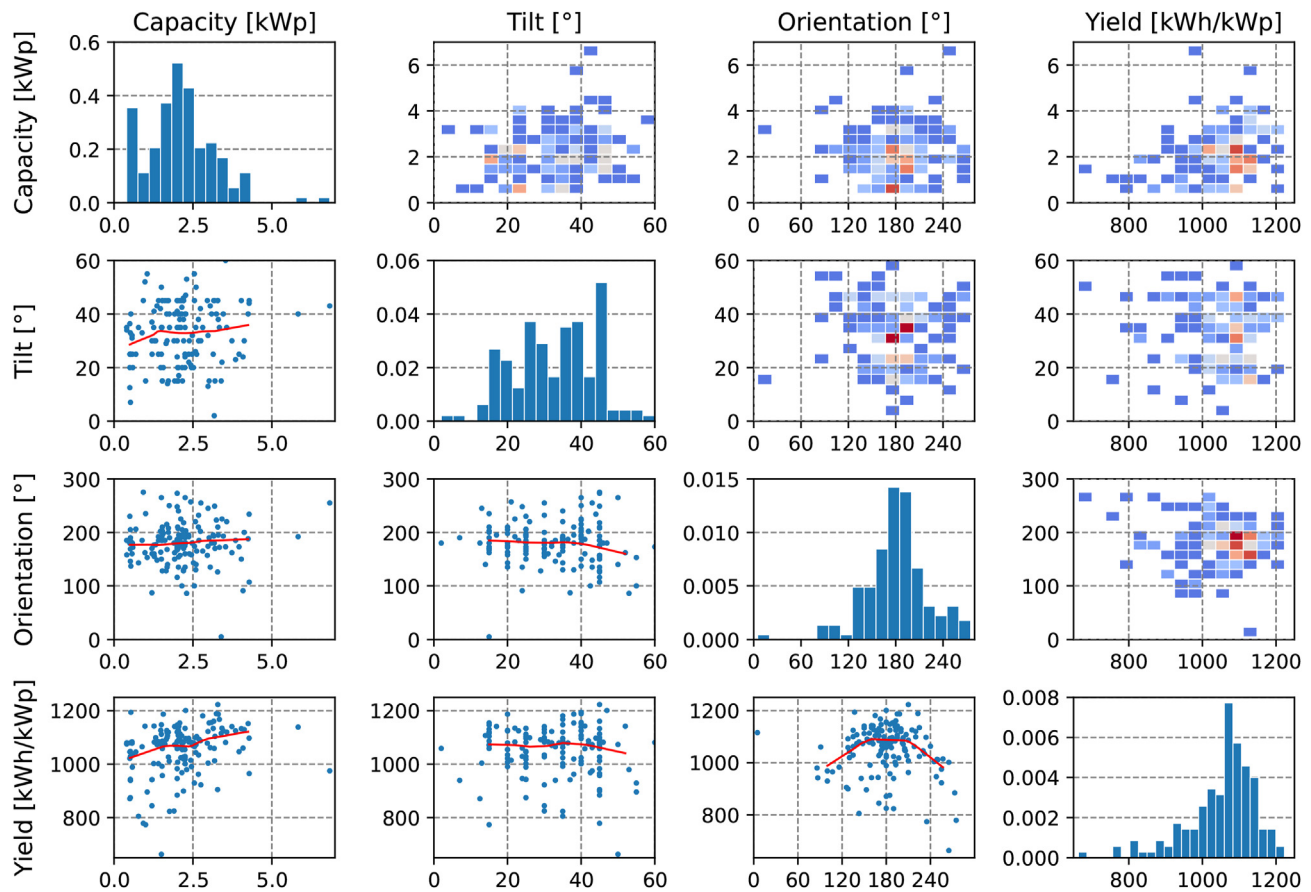


Fig. 5. An overview of the characteristics and observed annual yield of the PV systems considered in this study. The figure should be read as a matrix, where the sub figures in the diagonal are probability density functions that present the probability distribution of the characteristics among the PV fleet. The sub figures to the left are scatter plots, whereas the trend, i.e. a best fit line, is represented by the red line. The figures in the top right corner present a heat map, where the color marks the density of systems found from low (blue) to high (red).

1225 kWh/kWp.

The power output of these PV systems has been collected with a 1-min resolution from January 2014 until February 2017 and are averaged to obtain mean hourly production values. The considered PV systems have a data availability of at least 95% for diurnal hours. The distribution of the monthly summed PV power generation of the considered systems is shown in Fig. 6. The figure indicates annual variations in the PV power output per month as well as a significant distribution among the PV systems in terms of electricity generation per unit of installed capacity.

3.2. Predictor variables

For the same period, historic weather predictions of several variables are collected from the European Centre for Medium-Range Weather Forecasts (ECMWF) weather archive [48]. These NWP are generated by the High Resolution Forecast Configuration (HRES) of the Integrated Forecast System (IFS) developed by ECMWF. The historic archive holds the weather predictions for two separate simulations per day, one at noon and one at midnight. These have a time horizon of 10 days, an ascending time step from 1 to 6 h depending on the time horizon, and a lead time of 0 h. Since in this study we focus on day-ahead PV power output forecasts, we only collect day-ahead weather predictions. This corresponds to weather predictions with a lead time of 12 h and a time horizon of 36 h. Moreover, for these hours, predictions are available with an hourly time resolution. In total 12 predictor variables are collected from these NWP forecasts, i.e. variables 1–12 in Table 2.

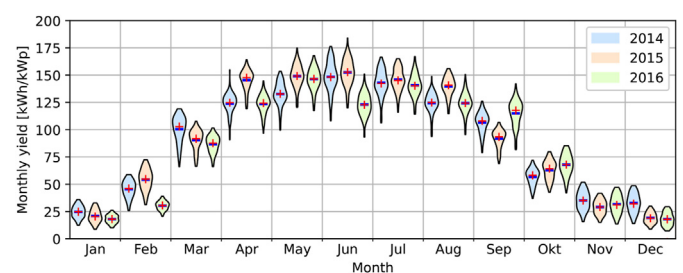


Fig. 6. A distribution plot presenting the observed monthly yield per year of all the PV systems considered in this study. The mean yield of the 152 PV systems per month is indicated by a blue line, the red marker presents the median.

These variables are complemented with the hour of the day (HoD), which is presented by a sine and cosine component in order to deal with its cyclic nature [9]. In addition, several variables are included that describe the time dependent position of the sun in the sky. These variables include the CSI, Solar Zenith Angle (SZA) and the Solar Azimuth Angle (AZI).

4. Results

This section firstly presents the technical and economic performance of the forecast models when a single PV system is considered. Next, the impact of aggregating PV systems on the performance of the models is discussed.

4.1. Single PV system

4.1.1. Technical comparison

Fig. 7 presents the performance of each forecast model for the test period (i.e. February 2016 until February 2017) by showing the distribution of the MAE observed per PV system. Similar results are found in terms of the RMSE. The MAE is depicted, as this metric is easier to interpret with respect to under- and overestimation of the forecasted PV power output. From Fig. 7 it becomes clear that all forecast models outperform the benchmark models, i.e. DP and CSP. Moreover, the competitiveness amongst the top performing models, with respect to the average observed MAE among all considered PV systems, is obvious from Fig. 7. The RF model is the best performing model with an average MAE of 6.13%. This is followed by the LSTM and GB models, for which MAE values of 6.22% and 6.29% are found. Since the ANN model achieves a similar performance, it can be concluded that in terms of the MAE ensemble and deep learning based models are superior to the alternatives. Consequently, these models are found to best filter and capture the relevant information from the available predictor variables.

Besides, Fig. 7 indicates that the PV model performs slightly better than the regression-based models, including MLR, SARIMAX and LASSO, and LSVR. This is remarkable as the PV model requires the input of less predictor variables nor does it need any training before it can be applied. However, the PV model does rely on the availability of information concerning the characteristics of the PV system, e.g. azimuth and tilt angle, which information is not included in any other model. The results prove the advantage of utilizing a PV model to describe the relation between predictor variables that affect the PV power output directly and exclude the remainder variables. Subsequently, the inclusion of variables that indirectly affect the PV power output within the linear regression models may cloud the learning process, causing a poorer performance for the linear regressors.

Furthermore, the difference in the performance between the two benchmark models should be pointed out as the simpler DP model shows a slightly lower MAE than the CSP with an average MAE of 11.3% and 11.4%, respectively. This can be explained as the DP model can better describe similar weather patterns that occur on consecutive days. Hence, the DP model performs better during a series of typical summer days where cloud formation occurs in the afternoon. Since the CSP forecasts the day-ahead PV power production based on the current situation only, it lacks the capability to include such daily patterns.

Lastly, a MBE of less than 1% is observed for the two benchmark models and under 0.5% for all others.

4.1.2. Economic comparison

The economic performance in this study is assessed based on

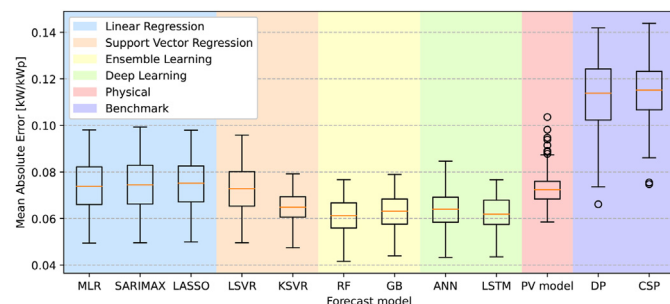


Fig. 7. The Mean Absolute Error per forecast model for each individual PV system. Color reflects ordering per model type.

the ER, see Section 2.5. Fig. 8 presents the average economic performance per model over all 152 PV systems for the entire test period. Moreover, this figure presents the revenues made on the DAM as well as those made on the imbalance markets during the studied period. The different revenue streams are separated, whereas the first bar per model presents the revenues from trading on the DAM. The second bar shows the net revenues due to a deficit in the delivery of electricity. Since these are negative, the deficit comes at a net costs for the PV operator and results in lower revenues made per installed capacity. The third bar presents the net revenues from a surplus in electricity delivery, increasing the average annual revenues per kWp.

Although in Fig. 8 it can be observed that the deficit in electricity production comes at a net cost for the PV operator during the studied period, it should be noted that it can lead to additional revenues for the PV operator from time to time in case of a (strong) negative SP. In practice, this would be the case when the electricity grid is highly congested, such that involved parties are rewarded for reducing their production [45]. Similarly, a surplus of electricity production could at times be met with net costs. Moreover, a varying SP results in less and higher penalized forecast errors. This observation is important to note because it explains the fact that a (high) prediction error may increase the benefits, promoting a less accurate forecast model. Nevertheless, imbalance market prices are unpredictable and in general the revenues on the imbalance market are less profitable for traders, which stimulates them to minimize the forecast error (see Section 2.5).

The overall economic performance of all models as well as a perfect forecast case (Perfect) is presented in Fig. 8. The results show that over the test period, the revenues made in case of a perfect forecast of the PV power output amount to 27.18 € per kWp installed. This could also be referred to as the theoretical potential of a perfect forecast model. We find that adopting the forecast models lead to revenues varying between 24.77 and 26.39 € per kWp on an annual basis. Remarkably, the best performing model is the PV model. Fig. 8 shows that this is closely followed by the LSTM, KSVR and ANN models, respectively. Similar to the MAE, the benchmark models also show worse performance according to the economic metric. However, in this case the CSP is found to outperform the DP model.

Furthermore, Fig. 8 shows a clear relationship between high revenues made at the DAM and high costs related to production deficits. This relationship subsist due to an overestimation of the electricity production, i.e. a positive bias, and the SP raised on the deficit in production that emerges from this overestimation. This is the case for MLR, SARIMAX, LASSO, CSP and the PV model, where a high production deficit occurs due to an overestimation of the PV production in the day-ahead bid (on the DAM). The deficits are met by significant amounts of imbalance penalties that must be paid.

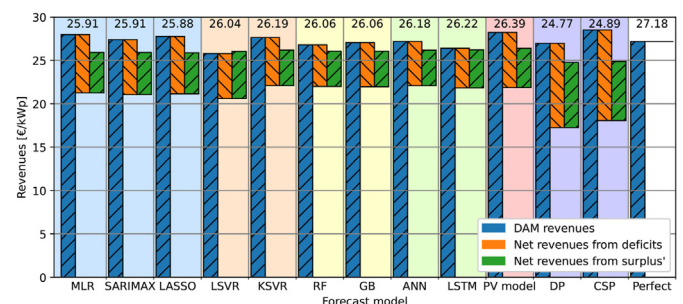


Fig. 8. Overview of the observed economic performance in terms of costs and revenues per kWp installed capacity of solar PV for each forecast model when used for trading on the DAM, including raised imbalance penalties.

Another interesting result is found for the LSVR model, which is the only model found to generate more income from the surplus in electricity production than costs associated with production deficits. These net profits made on the imbalance market are explained by a structural underestimation of the PV power output by the LSVR model, i.e. a significant negative bias. This production underestimation leads to relative low quantity bids resulting in lower revenues made at the DAM and a higher surplus of electricity when compared to other models, which leads to the observed net profits.

Finally, it should be noted that the smallest costs and revenues components associated with the imbalance market are present for the LSTM and RF models. This is closely followed by the ANN and GB models, see Fig. 8. The limited revenues of the ensemble and deep learning models on the imbalance market is in line with the low MAE found for these models in Section 4.1.1, since a low prediction error results in less deviations that are to be settled.

4.1.3. Overall comparison

The order of best performing models with the purpose of forecasting the day-ahead power output of a single PV system differs depending on the performance metric considered, e.g. MAE or ER. While the results show some overlap, the differences are more striking. Firstly, the PV model outperforms all other forecast models when examined on the ER metric, whereas it is rated the sixth model in terms of the MAE. The opposite applies to the RF model, which was rated as the best model to the MAE while in terms of the economic metric it is outperformed by the PV, LSTM, KSVR and ANN models. On the other hand, similarities are found between the performance of the forecast models according to the performance metrics for the MLR, SARIMAX and LASSO models. In addition, the benchmark models considered in this study are the worst performing alternative according to both performance metrics. Nevertheless, whereas CSP was the worst model when evaluated on the average MAE, the DP model performs the worst in case of considering the economic metric. These results are in line with one other study that was found to evaluate operational solar forecasts on both technical and economic metrics [24]. Although the order of best performing models did not significantly change in this case study on the Spanish DAM, the results show differences in the best model based on the technical and economic metric used.

The preference for the PV model in terms of the ER metric is explained by the interaction between the observed forecast errors, the DAMPs and the varying SPs. Consequently, no single reason can be indicated that explains the better performance of the PV model. Nevertheless, one unique feature of the PV model that should be highlighted in this context is the ability to forecast outliers. The LSTM model proves a valuable option to forecast the PV power production as in this study it is found to perform as the second best model for both performance metrics. However, the decision to adopt a specific model to forecast the day-ahead power production of a PV system is in this study shown to depend on the objective of the user. If the objective is to forecast the exact power production as accurate as possible, those models that perform well on the MAE (RF and LSTM) are preferred. On the other hand, if the objective is to maximize the revenues from electricity trading, the ER is a more important metric leading to a preference for the PV model.

4.2. Multiple PV systems

4.2.1. Technical comparison

The performance of all forecast models in terms of the MAE are summarized in Figs. 9 and 10. Fig. 9 presents the MAE and shows how the MAE per forecast model depends on the number of PV systems considered in the forecast and the spatial distribution covered by these systems. Firstly, Fig. 9 clearly points out the

benefit of considering PV systems that are spread over a larger spatial area as the MAE decreases with an increasing inter-system distance. Moreover, this trend is observed for all forecast models, except for DP and CSP, and is independent of the number of systems considered. As no clear stagnation of this trend is found in this study and the improvement are observed to be constant with an increasing distance, the results raise the question at what inter-system distance the positive effect will reduce. Lorenz et al. [30] found that this effect was exponential and stabilized at an area of approximately 200×200 km for a solar irradiance forecast model.

Secondly, the results depicted in Fig. 9 show how the MAE per forecast model depends on the density of the PV systems considered in the forecast. From these results it is found that the performance of all forecast models increases with an increasing number of PV systems that is considered. This is explained as in addition to spatial smoothing also an increasing number of systems scattered throughout a region of interest smoothens the PV production profile, enhancing the forecast performance. Moreover, the MAE of each forecast model improves significantly as the number of PV systems considered in the forecast increases from e.g. 2 to 5 PV systems. However, the rate of improvement becomes limited after a total number of 10 PV systems are included in the forecast. Also, no significant difference between the MAE in case of 25, 50 and 100 PV systems is observed. Subsequently, for the spatial area considered in this study, the advantage found of increasing the number of PV systems appears to stagnate around 10 PV systems and is not present anymore at a level of 25. The advantage of adding a PV system is therefore the greatest when the size of the initial set is limited.

Overall, when we compare the separate effects of the density and spatial distribution of the aggregated systems, it is found that the number of systems is a more significant factor in improving the model forecast performance, if the initial set holds a limited number of PV systems. In this study, this tipping point lays around 10 PV systems. When the initial set of PV systems exceeds 10, the inter-system distance becomes a more significant factor to improve the forecast accuracy. Consequently, this implies that there is a saturation level with respect to the spatial density of the number of systems included and the observed forecast performance improvement.

Next, Fig. 10 presents the average MAE per forecast model observed for the different kind of aggregations sets. This figure therefore enables to compare the performance of the forecast models among each other. Firstly, the results show the dominance of the RF and LSTM models where the MAE decreases from 0.062 to 0.048 kW per kWp with an increasing number of systems and inter-system distance. These models, which alternate one another as the top performing forecast model for all categories, are followed by the ANN and GB models. Fig. 10 also indicate the large difference in the superiority of all forecast models compared to the benchmark models, DP and CSP. In addition, the DP model outperforms the CSP models in terms of the MAE as for most distance and number of systems combinations 4.1.1. These results are largely inline with the results observed for single PV system forecasting, except that LSTM may be preferred over the RF model in case of forecasting the power production of multiple PV systems. Finally, the PV model appears to profit less from an increasing number of systems spread over a larger distance, as the linear-based regression models are found to outperform the PV model more and more with an increasing number of systems and distance. Lastly, similar observations with respect to the best performing models and the observed trends are found when considering the RMSE instead of the MAE.

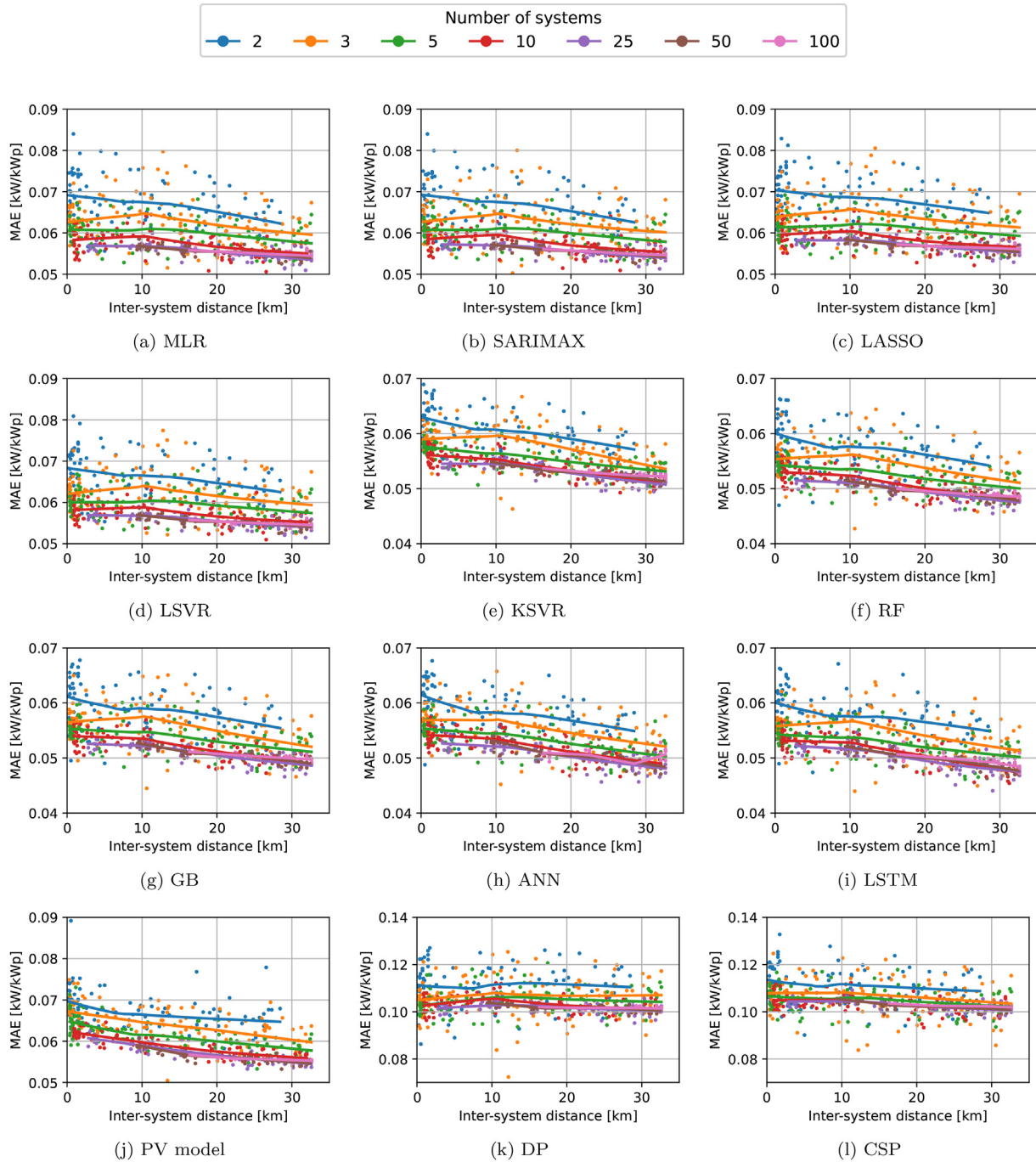


Fig. 9. The MAE per forecast model where each point considers a set of N PV systems. The trend lines show the average MAE for an increasing inter-system distance. Note, the y-axis scales are different (best viewed in color).

4.2.2. Economic comparison

Fig. 11 presents the performance of the forecast models in terms of the ER in a similar fashion as Section 4.2.1 described the technical performance. Moreover, Fig. 11 shows the influence of the number of PV systems and inter-system distance on the observed ER per forecast model. Firstly, from the results the revenues are found to increment slightly as the inter-system distance increases up to 10 km. Hereafter the revenues remain approximately constant per kWp with an increasing spatial distribution. This relation is found for all forecast models and is also observed in case of a perfect forecast model. Consequently, the gain in ER cannot be credited to

an improved forecast performance as was found in Section 4.2.1. Moreover, this observation is explained by the smoothing effect that a larger considered spatial distribution of PV systems has on the power production profile, resulting in a more balanced electricity supply to the DAM. Furthermore, Fig. 11 does not show any consistent relationship between the number of PV systems that is considered by the model and the revenues made nor the potential revenues that could have been made, i.e. in case of a perfect forecast.

Fig. 12 depicts the average revenues per aggregation level as a fraction of the maximum revenues that could have been generated

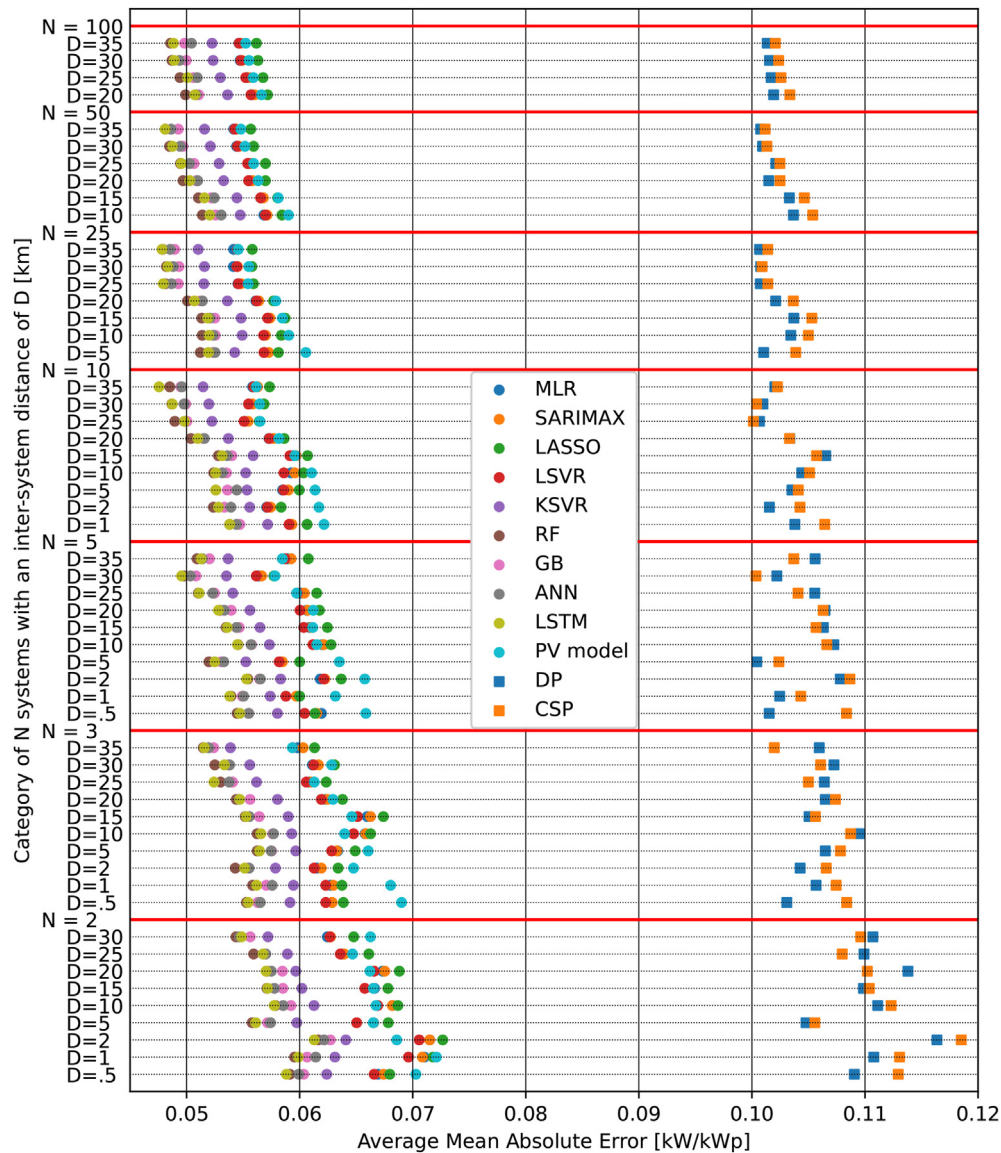


Fig. 10. An overview of the average MAE found for a varying number of systems (N) and inter-system distances (D) (best viewed in color).

in case of a perfect forecast. Consequently, this figure enables to examine the model forecast performance to its potential revenues, giving insights into the economic forecast improvements that can still be gained. Fig. 12 shows that the performance of the forecast models, as a fraction of the maximum revenues, remain more or less constant with an increasing inter-system distance and number of systems. This proves the absence of a clear trend showing a gain in the performance of the forecast models as a result of aggregating PV systems over an increasing spatial area. Nevertheless, Fig. 12 does show a more constant economic performance of the forecast models when the number of systems included grow. The absence of an economic gain is remarkable as it implies that the improvement that was observed for the forecast models in terms of the MAE in section 4.2.1 does not translate into increased ER.

The results in Fig. 12 also enable for comparing the economic performance of the forecast models amongst one another. The results clearly indicate the superiority of the PV model, which is found to outperform the other models in all categories except one. The ER for the PV model vary between 26.7 and 30.3 € per kWp.

Next, with a tested ER in the range of 26.6 and 30.1 € per kWp the deep learning models tested in this study prove as valuable alternatives, where the LSTM model outperforms the ANN model. Subsequently, a relative consequent order of performance for the forecast models can be observed over all categories in Fig. 12, where the deep learning models are in order followed by the ensemble learning, SVRs and linear regressor models. Finally, all forecast models are found to outperform the CSP and DP model that obtain an ER value between 25.2 and 28.5 € per kWp. Moreover, in terms of the ER, the CSP and DP forecast models are identified to be competitive to one another.

4.2.3. Overall comparison

Similar to section 4.1, the order of the best performing models that forecast the PV power output of multiple sites depends on the performance metric of interest. Again, the RF and LSTM models showed their superiority when assessed on the MAE. This is followed by the ANN and GB models. If we consider the model performance from an economic perspective, these models lose their

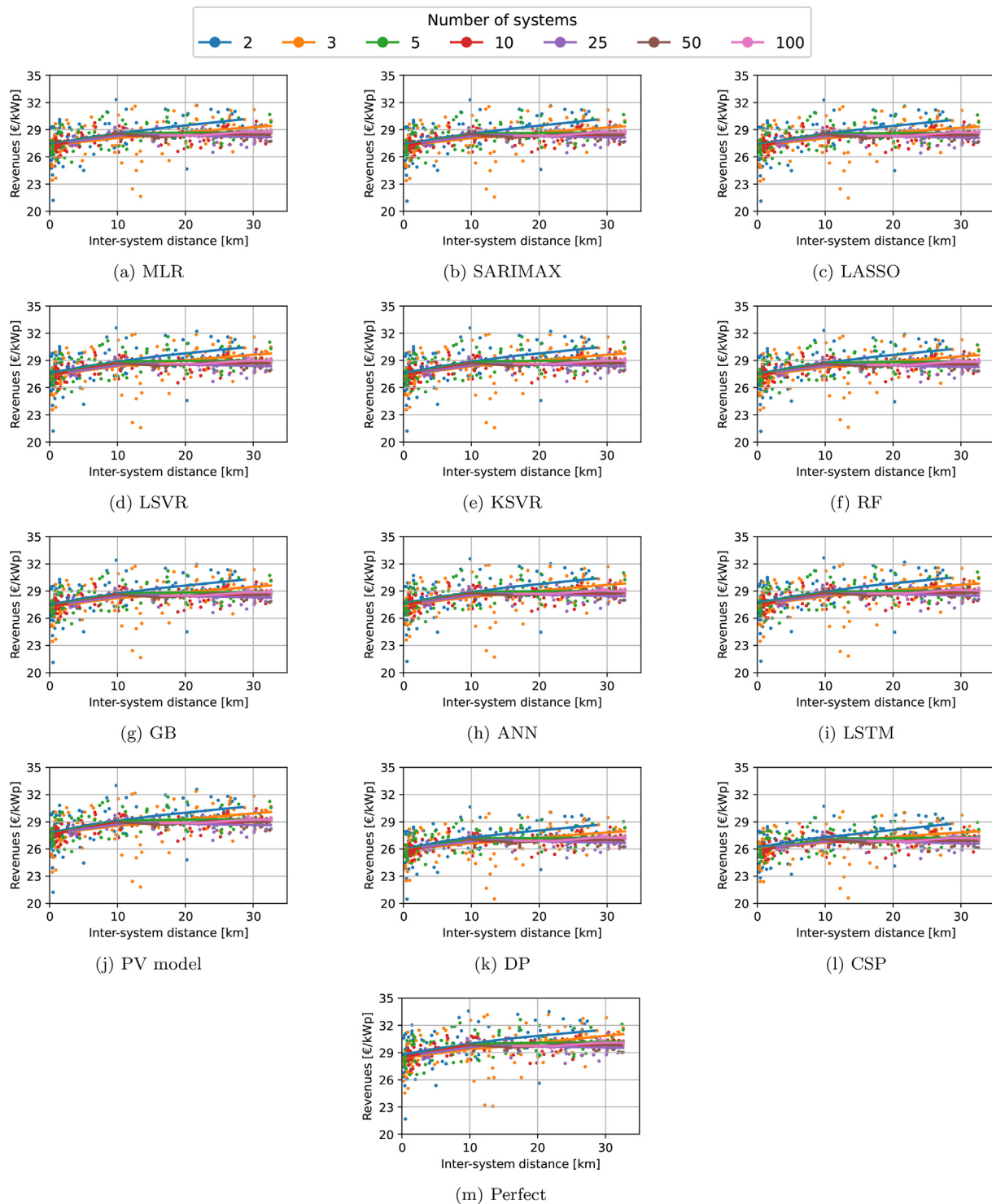


Fig. 11. The revenues per forecast model where each point considers a set of N PV systems. The trend lines show the average revenues over an increasing inter-system distance (*best viewed in color*).

superiority to the PV model. Along with the ANN model, the LSTM model proves to be a valuable alternative to the PV model in terms of the ER metric. Therefore, completely in line with the results found for single site forecasting, the objective of the user determines what model is preferred.

Furthermore, the results show that a decreasing MAE can be expected as the inter-system distance or the number of aggregated

systems grow. Similar results were found for a case study in Italy, where the performance of day-ahead PV power output forecast models improved when several regions were aggregated [3]. On the other hand, in this study the improvement is not observed for the model performance in terms of the ER with respect to the maximum ER. Still, the maximum ER, i.e. Perfect, is found to increase when a multitude of systems is considered with inter-

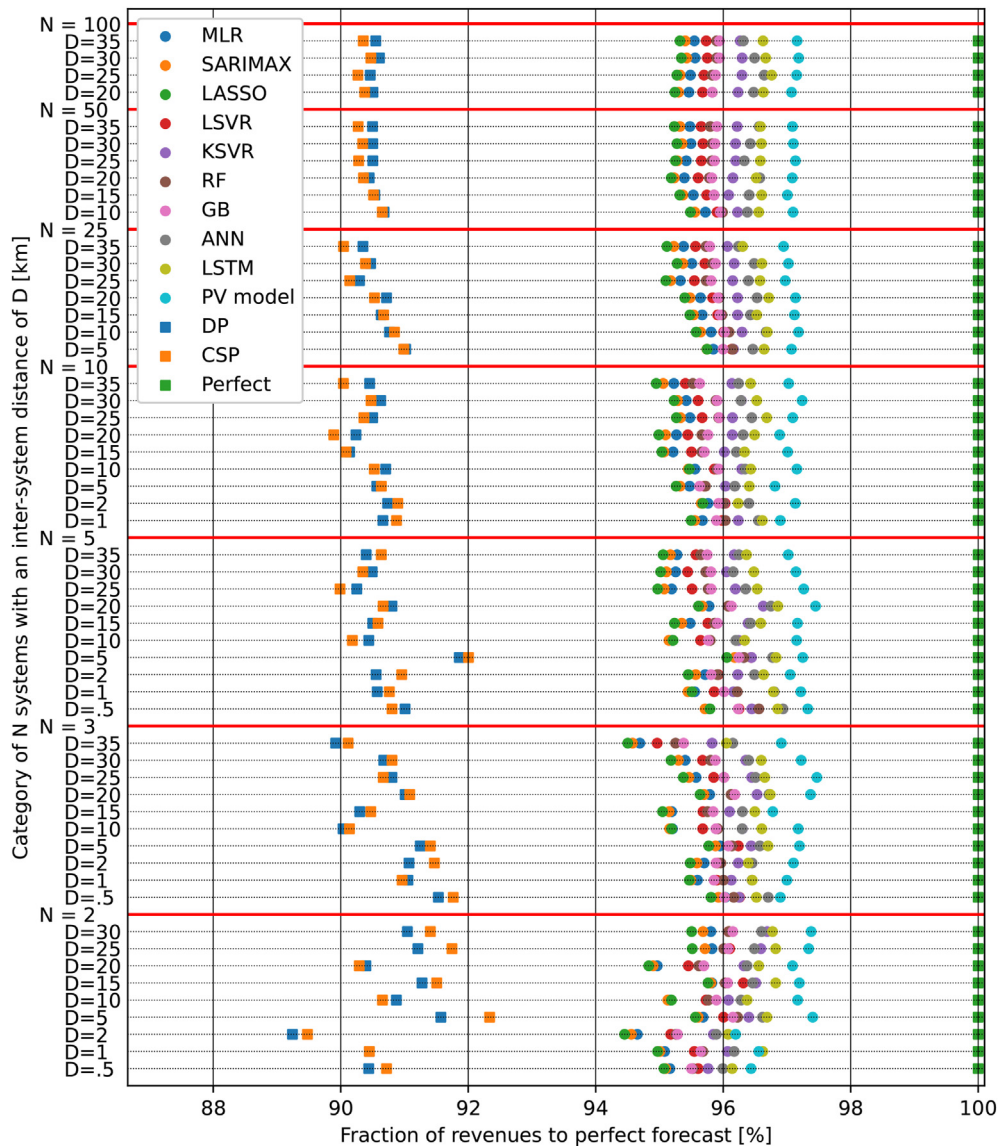


Fig. 12. An overview of the average revenues per forecast model found as a fraction of the maximum trading revenues in case of a perfect forecast. The results show the average revenues per category, i.e. the number of systems (N) and their inter-system distance (D) (best viewed in color).

system distance of up to 10 km.

Consequently, users who are interested in implementing an operational PV power forecast model should carefully consider the objective of the forecast. Moreover, for parties involved in trading on the DAM the ER may be considered as more important, resulting in a preference for adopting a PV model. On the other hand, grid operators, responsible for balancing the electricity grid and scheduling balancing reserves, may prefer the forecast model with the best technical performance, i.e. LSTM and RF.

5. Conclusions

In this study we have developed and compared the performance of 12 models that forecast the PV power production on a day-ahead basis. The models operate in agreement with the DAM and are tested when applied to forecast the production of individual and aggregated PV systems. The performance of the forecast models is examined to their technical (MAE) and economic (ER) performance, where the latter concerns a novel application to evaluating solar

forecasting methods.

Firstly, the results show that from a technical perspective and in case of forecasting the PV power production for a single PV system, the best performance is obtained by the RF and subsequently the LSTM model. Moreover, as we aggregate a number of PV systems, the technical performance of the RF and LSTM model is tied and superior over all other models. Consequently, these models are recommended to users that aim to minimize grid imbalances, e.g. system operators. Furthermore, from an economic viewpoint the results in this study show that the PV model outperforms all alternatives as it delivers the highest ER for both single sites and aggregations. The PV model is therefore recommended to users that participate in the DAM, e.g. utilities and aggregators. Following the PV model, the deep learning models (i.e. LSTM and ANN) obtain the best economic performance. Hence, based on the consistent good performance achieved by the LSTM model for both error metrics and a varying spatial distribution, it could claim superiority over the investigated alternatives.

Furthermore, since the preferred forecast model was found to

depend on the interest of the user, this study uncovers a conflict of interest as pursuing maximal economic revenues leads to increased grid imbalances. This adds to the ongoing discussion on the selection of error metrics and confirms that the importance of considering users' objective while developing a forecast model.

Next, this study shows the dependence of the performance of each model on the spatial distribution of the considered PV systems, which was decomposed in two factors: density and spatial area. From a technical perspective, the empirical results show that both factors have a positive effect on the model performance, where the effect of the number of aggregated PV systems was more significant in improving the model performance up to an aggregation of approximately 10 PV systems. Thereafter, the positive impact of an increasing inter-system distance was found to prevail. Consequently, system operators or portfolio managers that aim to minimize grid imbalances are found to benefit from aggregating PV systems, especially over a larger spatial area. System operators may therefore prefer a larger spatial spread of installed PV systems. Since the performance improvement in terms of the MAE did not translate into a monetary benefit, participants in the DAM can be indifferent to the number of systems in their portfolio as well as the spatial area.

From this study a number of directions for future work can be derived. Firstly, the application of probabilistic forecasts along with a strategy to extract an optimal single value for bidding in the DAM deserves attention. Secondly, we did not observe a stagnation of the model performance improvement in the MAE with respect to the inter-system distance. Subsequently, additional research should consider a larger area in order to quantify the improvement of the MAE considering larger spatial distributions. Furthermore, future research into the economic value of PV power forecast models should consider additional markets, e.g. the intraday market, as this may increase revenues by reducing the raised imbalance penalty. Finally, as we identified that the forecast objective sets the best forecast model, a topic that deserves attention is the interaction between the objective, the selection of a model and the effect on the required balancing capacity.

CRediT authorship contribution statement

Lennard Visser: Conceptualization, Data curation, Methodology, Software, Writing – original draft, Writing – review & editing. **Tarek AlSkaif:** Writing – review & editing, Supervision. **Wilfried van Sark:** Writing – review & editing, Supervision.

Declaration of competing 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.

Acknowledgements

This work is part of the Energy Intranets (NEAT: ESI-BiDa 647.003.002) project, which is funded by the Dutch Research Council NWO in the framework of the Energy Systems Integration & Big Data programme.

References

- [1] IEA, World energy outlook 2020. Tech. Rep., IEA, Paris, 2020. URL, <https://www.iea.org/reports/world-energy-outlook-2020>.
- [2] IEA, Renewables 2019, Tech. rep., IEA, Paris, 2020. URL, <https://www.iea.org/reports/renewables-2019>.
- [3] A. Betti, M. Pierro, C. Cornaro, D. Moser, M. Moschella, E. Collino, D. Ronzio, D. van der Meer, L. Visser, J. Widen, T. AlSkaif, W. van Sark, Regional solar power forecasting 2020, Tech. Rep. July (2020), <https://doi.org/10.13140/RG.2.2.32196.24964>.
- [4] F.P. Kreuwel, W.H. Knap, L.R. Visser, W.G. van Sark, J.V.-G. de Arellano, C.C. van Heerwaarden, Analysis of high frequency photovoltaic solar energy fluctuations, *Sol. Energy* 206 (2020) 381–389.
- [5] T.O. Olowu, A. Sundararajan, M. Moghaddami, A.I. Sarwat, Future challenges and mitigation methods for high photovoltaic penetration: a survey, *Energies* 11 (7) (2018) 1782.
- [6] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. Martinez-de Pison, F. Antonanzas-Torres, Review of photovoltaic power forecasting, *Sol. Energy* 136 (2016) 78–111.
- [7] I. Lampropoulos, T. AlSkaif, J. Blom, W. van Sark, A framework for the provision of flexibility services at the transmission and distribution levels through aggregator companies, *Sustainable Energy, Grids and Networks* 17 (2019) 100187.
- [8] M.Q. Raza, M. Nadarajah, C. Ekanayake, On recent advances in PV output power forecast, *Sol. Energy* 136 (2016) 125–144.
- [9] T. AlSkaif, S. Dev, L. Visser, M. Hossari, W. van Sark, A systematic analysis of meteorological variables for pv output power estimation, *Renew. Energy* 153 (2020) 12–22.
- [10] D. Yang, J. Kleissl, C.A. Gueymard, H.T. Pedro, C.F. Coimbra, History and trends in solar irradiance and pv power forecasting: a preliminary assessment and review using text mining, *Sol. Energy* 168 (2018) 60–101.
- [11] R. Ahmed, V. Sreram, Y. Mishra, M. Arif, A review and evaluation of the state-of-the-art in pv solar power forecasting: techniques and optimization, *Renew. Sustain. Energy Rev.* 124 (2020) 109792.
- [12] I. Lampropoulos, M. van den Broek, E. van der Hoofd, K. Hommes, W. van Sark, A system perspective to the deployment of flexibility through aggregator companies in The Netherlands, *Energy Pol.* 118 (2018) 534–551.
- [13] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Machine learning methods for solar radiation forecasting: a review, *Renew. Energy* 105 (2017) 569–582.
- [14] K. Wang, X. Qi, H. Liu, A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network, *Appl. Energy* 251 (2019) 113315.
- [15] 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 (2019) 115838.
- [16] X. Zhang, Y. Li, S. Lu, H.F. Hamann, B.-M. Hodge, B. Lehman, A solar time based analog ensemble method for regional solar power forecasting, *IEEE Trans. Sustainable Energy* 10 (1) (2018) 268–279.
- [17] D. Yang, Z. Dong, Operational photovoltaics power forecasting using seasonal time series ensemble, *Sol. Energy* 166 (2018) 529–541.
- [18] F. Wang, Z. Xuan, Z. Zhen, K. Li, T. Wang, M. Shi, A day-ahead pv power forecasting method based on lstm-rnn model and time correlation modification under partial daily pattern prediction framework, *Energy Convers. Manag.* 212 (2020) 112766.
- [19] A.T. Eseye, J. Zhang, D. Zheng, Short-term photovoltaic solar power forecasting using a hybrid wavelet-pso-svm model based on scada and meteorological information, *Renew. Energy* 118 (2018) 357–367.
- [20] D. Yang, A guideline to solar forecasting research practice: reproducible, operational, probabilistic or physically-based, ensemble, and skill (ropes), *J. Renew. Sustain. Energy* 11 (2) (2019), 022701.
- [21] C.L. Dewangan, S. Singh, S. Chakrabarti, Combining Forecasts of Day-Ahead Solar Power, *Energy* (2020) 117743.
- [22] P. Bacher, H. Madsen, H.A. Nielsen, Online short-term solar power forecasting, *Sol. Energy* 83 (10) (2009) 1772–1783.
- [23] L. Visser, T. AlSkaif, W. van Sark, Benchmark analysis of day-ahead solar power forecasting techniques using weather predictions, in: 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), IEEE, 2019, pp. 2111–2116.
- [24] J. Antonanzas, D. Pozo-Vázquez, L. Fernandez-Jimenez, F. Martinez-de Pison, The value of day-ahead forecasting for photovoltaics in the Spanish electricity market, *Sol. Energy* 158 (2017) 140–146.
- [25] D. Yang, E. Wu, J. Kleissl, Operational solar forecasting for the real-time market, *Int. J. Forecast.* 35 (4) (2019) 1499–1519.
- [26] D. Yang, S. Alessandrini, J. Antonanzas, F. Antonanzas-Torres, V. Badescu, H.G. Beyer, R. Blaga, J. Boland, J.M. Bright, C.F. Coimbra, et al., Verification of deterministic solar forecasts, *Sol. Energy* 210 (2020) 20–37.
- [27] J. Antonanzas, O. Perpinan-Lamigueiro, R. Urraca, F. Antonanzas-Torres, Influence of electricity market structures on deterministic solar forecasting verification, *Sol. Energy* 210 (2020) 44–46.
- [28] J. Zhang, A. Florita, B.-M. Hodge, S. Lu, H.F. Hamann, V. Banunarayanan, A.M. Brockway, A suite of metrics for assessing the performance of solar power forecasting, *Sol. Energy* 111 (2015) 157–175.
- [29] J. Marcos, L. Marroyo, E. Lorenzo, M. García, Smoothing of pv power fluctuations by geographical dispersion, *Prog. Photovoltaics Res. Appl.* 20 (2) (2012) 226–237.
- [30] E. Lorenz, J. Hurka, D. Heinemann, H.G. Beyer, Irradiance forecasting for the power prediction of grid-connected photovoltaic systems, *IEEE J. Sel. Top. Appl. Earth. Obs. Remote Sens.* 2 (1) (2009) 2–10.
- [31] J. Kleissl, Solar Energy Forecasting and Resource Assessment, Academic Press, 2013.
- [32] G. Van Rossum, F.L. Drake Jr., Python Reference Manual, Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [33] W.F. Holmgren, C.W. Hansen, M.A. Mikofski, Pvlb python: a python package

- for modeling solar energy systems, *Journal of Open Source Software* 3 (29) (2018) 884.
- [34] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, G. Varoquaux, API design for machine learning software: experiences from the scikit-learn project, in: *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013, pp. 108–122.
- [35] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G.S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: large-scale machine learning on heterogeneous systems, Software available from, <https://www.tensorflow.org>, 2015.
- [36] S.I. Vagropoulos, G. Choularas, E.G. Kardakos, C.K. Simoglou, A.G. Bakirtzis, Comparison of sarimax, sarima, modified sarima and ann-based models for short-term pv generation forecasting, in: *2016 IEEE International Energy Conference (ENERGYCON)*, IEEE, 2016, pp. 1–6.
- [37] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer science & business media, 2013.
- [38] L. Breiman, Random forests, *Mach. Learn.* 45 (1) (2001) 5–32.
- [39] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, *Deep Learning*, vol. 1, MIT press Cambridge, 2016.
- [40] H. Sak, A. W. Senior, F. Beaufays, Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling.
- [41] P. Ineichen, R. Perez, R. Seal, E. Maxwell, A. Zalenka, Dynamic global-to-direct irradiance conversion models, *Build. Eng.* 98 (1) (1992) 354–369.
- [42] R. Perez, P. Ineichen, R. Seals, J. Michalsky, R. Stewart, Modeling daylight availability and irradiance components from direct and global irradiance, *Sol. Energy* 44 (5) (1990) 271–289.
- [43] C.A. Gueymard, Cloud and albedo enhancement impacts on solar irradiance using high-frequency measurements from thermopile and photodiode radiometers. part 1: impacts on global horizontal irradiance, *Sol. Energy* 153 (2017) 755–765.
- [44] E. Welzl, Smallest enclosing disks (balls and ellipsoids), in: *New Results and New Trends in Computer Science*, Springer, 1991, pp. 359–370.
- [45] TenneT, Settlement prices (2020). URL <https://www.tennet.org/english/operationalmanagement/exportdata.aspx?exporttype=Onbalansprijs>.
- [46] S. Clò, E. Fumagalli, The effect of price regulation on energy imbalances: a difference in differences design, *Energy Econ.* 81 (2019) 754–764.
- [47] Entso-E, European network of transmission system operators for electricity, ENTSO-E, URL, <https://transparency.entsoe.eu/>, 2020.
- [48] ECMWF, European centre for medium-range weather forecasts, ECMWF, URL, <https://www.ecmwf.int/en/forecasts/datasets/archive-datasets>, 2020.