

Technical report 18-023

Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data*

J. Lago, K. De Brabandere, F. De Ridder, and B. De Schutter

If you want to cite this report, please use the following reference instead:

J. Lago, K. De Brabandere, F. De Ridder, and B. De Schutter, “Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data,” *Solar Energy*, vol. 173, pp. 566–577, Oct. 2018. doi:[10.1016/j.solener.2018.07.050](https://doi.org/10.1016/j.solener.2018.07.050)

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
phone: +31-15-278.24.73 (secretary)
URL: <https://www.dcsc.tudelft.nl>

* This report can also be downloaded via https://pub.bartdeschutter.org/abs/18_023.html

Short-term forecasting of solar irradiance without local telemetry: a generalized model using satellite data

Jesus Lago^{a,b,1}, Karel De Brabandere^c, Fjo De Ridder^b, Bart De Schutter^a

^a*Delft Center for Systems and Control, Delft University of Technology,
Mekelweg 2, Delft, The Netherlands*

^b*Algorithms, Modeling, and Optimization, VITO, Energyville,
ThorPark, Genk, Belgium*

^c*3E, Brussels, Belgium*

Abstract

Due to the increasing integration of solar power into the electrical grid, forecasting short-term solar irradiance has become key for many applications, e.g. operational planning, power purchases, reserve activation, etc. In this context, as solar generators are geographically dispersed and ground measurements are not always easy to obtain, it is very important to have general models that can predict solar irradiance without the need of local data. In this paper, a model that can perform short-term forecasting of solar irradiance in any general location without the need of ground measurements is proposed. To do so, the model considers satellite-based measurements and weather-based forecasts, and employs a deep neural network structure that is able to generalize across locations; particularly, the network is trained only using a small subset of sites where ground data is available, and the model is able to generalize to a much larger number of locations where ground data does not exist. As a case study, 25 locations in The Netherlands are considered and the proposed model is compared against four local models that are individually trained for each location using ground measurements. Despite the general nature of the model, it is shown that the proposed model is equal or better than the local models: when comparing the average performance across all the locations and prediction horizons, the proposed model obtains a 31.31% rRMSE (relative root mean square error) while the best local model achieves a 32.01% rRMSE.

Keywords: Solar Irradiance Forecast, Generalized Model, Deep Learning, Satellite Data

1. Introduction

With the increasing integration of renewable sources into the electrical grid, accurate forecasting of renewable source generation has become one of the most important challenges across several applications. Among them, balancing the electrical grid via activation of reserves is arguably one of the most critical ones to ensure a stable system. In particular, due to their intermittent and unpredictable nature, the more renewables are integrated, the more complex the grid management becomes [1, 2].

In this context, as solar energy is one of the most unpredictable renewable sources, the increasing use of solar power in recent years has led to an increasing interest in forecasting irradiance over short time horizons. In particular, in addition to activation of reserves to manage the grid stability, short-term forecasts of solar irradiance are paramount for operational planning, switching sources,

programming backup, short-term power trading, peak load matching, scheduling of power systems, congestion management, and cost reduction [2–4].

1.1. Solar irradiance forecasting

The forecasting of solar irradiance can be typically divided between methods for *global horizontal irradiance (GHI)* and methods for *direct normal irradiance (DNI)* [5], with the latter being a component of the GHI (together with the diffuse solar irradiance). As in this work GHI is forecasted, [5] should be used for a complete review on methods for DNI. For the case of GHI, forecasting techniques are further categorized into two subfields according to the input data and the forecast horizon [2, 6]:

1. Time series models based on satellite images, measurements on the ground level, or sky images. These methods are usually suitable for short-term forecasts up to 4-6 h. Within this field, the literature can be further divided into three groups.
 - (a) Classical statistical models like ARMA models [7], ARIMA models [4], the CARDS model [8], or the Lasso model [9].
 - (b) Artificial intelligence models such as neural networks models [10, 11], support vector machines

*Corresponding author

Email address: j.lagogarcia@tudelft.nl (Jesus Lago)

[11], decision trees-based models [12], or Gaussian models [11].

(c) Cloud-moving vector models that use satellite images [13].

2. *Numerical weather prediction (NWP)* models that simulate weather conditions. These methods are suitable for longer forecast horizons, 4-6 hours onward, time scales where they outperform the statistical models [14]. As the goal of this work are short-term forecasts, [6] should be used for more complete review of NWP methods.

While the division in accuracy between NWP and time series models is given by the predictive horizon, establishing comparisons between time series models is more complex. In particular, while some authors have reported the superiority of statistical models over artificial intelligence methods [4], others have obtained opposite results [15].

The input features typically used in the literature to predict solar irradiance vary widely, e.g. past irradiance values, satellite data, weather information, etc. In many cases, the inputs considered depend on the type of model used, e.g. cloud moving vector models require satellite images. While a detailed review on the different methods and input features is outside the scope of this paper, [6] is a good source for a more thorough analysis.

1.2. Motivation

To the best of our knowledge, due to the time series nature of the solar irradiance, the statistical and artificial intelligence methods proposed so far have considered past ground measurements of the solar irradiance as input regressors [6]. While this choice of inputs might be the most sensible selection to build time series models, it poses an important problem: local data is required at every site where a forecast is needed.

In particular, if the geographical dispersion of solar generators is considered, it becomes clear that forecasting solar irradiance is a problem that has to be resolved across multiple locations. If ground measurements of all these sites are required, the cost of forecasting irradiance can become very expensive. In addition to the cost, a second associated problem is the fact that obtaining local data is not always easy.

As a result, in order to obtain scalable solutions for solar irradiance forecasting, it is important to develop global models that can forecast without the need of local data. In this context, while current cloud-moving vectors might accomplish that, they are not always easy to deploy as they are complex forecasting techniques that involve several steps [6].

1.3. Contributions and Organization of the Paper

In this paper, a novel forecasting technique is proposed that addresses the mentioned problem by providing a prediction model that, while being accurate and easy to deploy, forecasts solar irradiance without the need of local

data. The prediction model is based on a *deep neural network (DNN)* that, using SEVIRI¹ satellite images and NWP forecasts, is as accurate as local time series models that consider ground measurements. Although the model uses satellite images just as cloud-moving vector models do, it is easier to deploy as it requires less complex computations. In addition, while obtaining satellite data might not be always easier or cheaper than installing local ground sensors, there are several locations where satellite data are available and the proposed model avoids going to the ground to install local measurements. An example of this is The Netherlands, where satellite data is provided by the national meteorological institute.

It is important to note that, to the best of our knowledge, the proposed method is the first of its class that tries to remove the dependence of local telemetry even for training. Particularly, while other methods from the literature successfully remove the local data dependence during forecasting, e.g. [16], they still require local telemetry at all sites of interest during training. While using local data in a small subsets of sites during training, the proposed model successfully predicts the irradiance in a much larger subset of locations without needing local telemetry from these sites at any stage of the estimation or the forecasting.

As a case study, 30 location in The Netherlands are considered and the model is estimated using 5 of these locations. Then, for the remaining 25 locations, the performance of the proposed estimated model is compared against individual time series models specifically trained for each site using ground data.

The remaining of the paper is organized as follows: Section 2 introduces the preliminary concepts considered in this work. Next, Section 3 presents the proposed general model for forecasting solar irradiance. Then, Section 4 introduces the case study and discusses the performance of the proposed model when compared with local models. Finally, Section 5 summarizes the main results and concludes the paper.

2. Preliminaries

In this section the concepts and algorithms that are used and/or modified in the paper are introduced.

2.1. Deep Learning and DNNs

In the last decade, the field of neural networks has experienced several innovations that have lead to what is known as *deep learning (DL)* [17]. In particular, one of the traditional issues of neural networks had always been the large computational cost of training large models. However, that changed completely when [18] showed that a deep belief network could be trained efficiently using an algorithm called greedy layer-wise pretraining. As related

¹The SEVIRI (Spinning Enhanced Visible and InfraRed Imager) is a measurement instrument of the METEOSAT satellite.

developments followed, researchers started to be able to efficiently train complex neural networks whose depth was not just limited to a single hidden layer (as in the traditional multilayer perceptron). As these new structures systemically showed better results and generalization capabilities, the field was renamed as deep learning to stress the importance of the depth in the achieved improvements [17, Section 1.2.1].

While this success of DL models initiated in computer science applications, e.g. image recognition [19], speech recognition [20], or machine translation [21], the benefits of DL have also spread in the last years to several energy-related applications [22–28]. Among these areas, wind power forecasting [22, 23] and electricity price forecasting [27, 28] are arguably the fields that have benefited the most.

While there are different DL architectures, e.g. convolutional networks or recurrent networks, in this paper a DNN is considered, i.e. a multilayer perceptron with more than a single hidden layer, in order to build the solar forecasting model. The reason for this selection is twofold: (1) DNNs are less computationally intensive than the other DL architectures [17]; (2) DNNs have empirically outperformed the other DL architectures in a similar energy-based forecasts [28], i.e. the forecast of day-ahead electricity prices.

2.1.1. Representation

Defining by $\mathbf{X} = [x_1, \dots, x_n]^\top \in \mathbb{R}^n$ the input of the network, by $\mathbf{Y} = [y_1, y_2, \dots, y_m]^\top \in \mathbb{R}^m$ the output of the network, by n_k the number of neurons of the k^{th} hidden layer, and by $\mathbf{z}_k = [z_{k1}, \dots, z_{kn_k}]^\top$ the state vector in the k^{th} hidden layer, a general DNN with two hidden layers can be represented as in Figure 1.

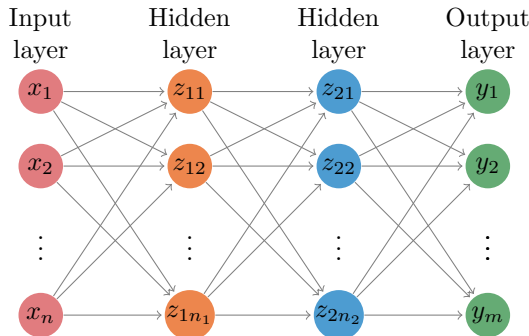


Figure 1: Example of a DNN.

In this representation, the parameters of the model are represented by the set of parameters \mathbf{W} that establish the mapping connections between the different neurons of the network [17].

2.1.2. Training

The process of estimating the model weights \mathbf{W} is usually called training. In particular, given a training set

$\mathcal{S}_{\mathcal{T}} = \{(\mathbf{X}_k, \mathbf{Y}_k)\}_{k=1}^N$ with N data points, the network training is done by solving a general optimization problem with the following structure:

$$\underset{\mathbf{W}}{\text{minimize}} \quad \sum_{k=1}^N g_k(\mathbf{Y}_k, F(\mathbf{X}_k, \mathbf{W})), \quad (1)$$

where $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the neural network map, and g_k is the problem-specific cost function, e.g. the Euclidean norm or the average cross-entropy. Traditional methods to solve (1) include the gradient descent or the Levenberg–Marquardt algorithm [29]. However, while these methods work well for small sized-networks, they display computational and scalability issues for DNNs. In particular, for DNNs better alternatives are the stochastic gradient descent algorithm and all its variants [30].

It is important to note that (1) is an approximation of the real problem one wish to solve. Particularly, in an ideal situation, the cost function w.r.t. to the underlying data distribution would be minimized; however, as the distribution is unknown, the problem has to be approximated by minimizing the cost function over the finite training set. This is especially relevant for neural networks, where a model could be overfitted and have a good performance in the training set, but perform badly in the test set, i.e. a set with a different data distribution. To avoid this situation, the network is usually trained in combination with regularization techniques, e.g. early stopping, and using out-of-sample data to evaluate the performance [17].

2.1.3. Network Hyperparameters

In addition to the weights, the network has several parameters that need to be selected before the training process. Typical parameters include the number of neurons of the hidden layers, the number of hidden layers, or the learning rate of the stochastic gradient descent method. To distinguish them from the main parameters, i.e. the network weights, they are referred to as the network hyperparameters.

2.2. Hyperparameter Optimization and Feature Selection

In this paper, to perform the hyperparameter selection, a Bayesian optimization algorithm that has been widely used for hyperparameter selection is considered: the *tree-structured Parzen estimator (TPE)* [31], an optimization algorithm within the family of sequential model-based optimization methods [32]. The basic principle of a sequential model-based optimization algorithm is to optimize a black-box function, e.g. the performance of a neural network as a function of the hyperparameters, by iteratively estimating an approximation of the function and exploring the function space using the local minimum of the approximation. At any given iteration i , the algorithm evaluates the black-box function at a new point θ_i . Next, it estimates an approximation \mathcal{M}_i of the black-box function by fitting the previously sampled points to the obtained function evaluations. Then, it selects the next sample point

θ_{i+1} by numerically optimizing \mathcal{M}_i and starts the next iteration. Finally, after a maximum number of iterations T have been performed, the algorithm selects the best configuration. Algorithm 1 represents an example of a sequential model-based optimization algorithm for hyperparameter selection.

Algorithm 1 Hyperparameter Optimization

```

1: procedure SMBO( $T, \theta_1$ )
2:    $\mathcal{H} \leftarrow \emptyset$ 
3:   for  $i = 1, \dots, T$  do
4:      $p_i \leftarrow \text{TrainNetwork}(\theta_i)$ 
5:      $\mathcal{H} \leftarrow \mathcal{H} \cup \{(p_i, \theta_i)\}$ 
6:     if  $i < T$  then
7:        $\mathcal{M}_i(\theta) \leftarrow \text{EstimateModel}(\mathcal{H})$ 
8:        $\theta_{i+1} \leftarrow \text{argmax}_{\theta} \mathcal{M}_i(\theta)$ 
9:     end if
10:  end for
11:   $\theta^* \leftarrow \text{BestHyperparameters}(\mathcal{H})$ 
12:  return  $\theta^*$ 
13: end procedure

```

In addition to optimizing the hyperparameters, the TPE algorithm is also employed for optimizing the selection of input features. In particular, the feature selection method proposed in [27] is considered, which selects the input features by first defining the input features as model hyperparameters and then using the TPE algorithm to optimally choose among them. More specifically, the method considers that each possible input feature can be either modeled as a binary hyperparameter representing its inclusion/exclusion or as an integer hyperparameter representing how many historical values of the specific input are used. In solar forecasting, an example of the former could be whether to consider the hour of the day as an input feature and an example of the latter could be the optimal number of past irradiance values.

2.3. Performance Metrics

In order to evaluate the accuracy of the proposed model, a performance metric is needed. In this paper, following the standards of the literature of solar irradiance forecasting, three different metrics are considered: the *relative root mean square error* (*rRMSE*), the *mean bias error* (*MBE*), and the forecasting skill s as defined by [33].

One of the most commonly used metrics for evaluating solar irradiance forecasting is the RMSE or rRMSE, which provide an assessment of the average spread of the forecasting errors. In particular, given a vector $\mathbf{Y} = [y_1, \dots, y_N]^\top$ of real outputs and a vector $\hat{\mathbf{Y}} = [\hat{y}_1, \dots, \hat{y}_N]^\top$ of predicted outputs, the rRMSE metric can be computed as:

$$\text{rRMSE} = \frac{\sqrt{\frac{1}{N} \sum_{k=1}^N (y_k - \hat{y}_k)^2}}{\frac{1}{N} \sum_{k=1}^N y_k} \cdot 100 \%. \quad (2)$$

A second metric that is widely used is the MBE, a measure of the overall bias of the model. Using the same definitions as before, the MBE metric can be computed as:

$$\frac{1}{N} \sum_{k=1}^N y_k - \hat{y}_k. \quad (3)$$

While both metrics can properly assess and compare models using the same dataset, they are hard to interpret when it comes to make comparisons across multiple locations, climate, and time of the year [33]. A metric that tries to solve this issue is the forecasting skill s ; particularly, S defines first a metric V that accounts for the *variability* of the solar irradiance, i.e. accounts for the specific variability due to location, climate, and time. Next, it defines a second metric U that accounts for the *uncertainty*, i.e. errors, of the forecasting model. Finally, the forecasting skill S is defined as:

$$s = 1 - \frac{U}{V}. \quad (4)$$

For the details on computing U and V as well as a detailed explanation on s , the reader is referred to [33]. The important aspect to consider for this study is that s is a normalized metric w.r.t. to a simple persistence model (see Section 4.2.1) that permits the comparison of models across different conditions. A normal forecaster should be characterized by $s \in [0, 1]$ with higher values indicating better forecasting; particularly, $s = 1$ indicates that the solar irradiance is perfectly forecasted, and $s = 0$ that the model is not better than a simple persistence model (by definition of U and V a persistence model will always have $s = 0$). Negative values would then imply the forecaster is worse than the simple persistence model.

3. Prediction Model

In this section, the proposed prediction model for solar irradiance forecasting is presented.

3.1. Model Structure

A key element to build a prediction model that can be used without the need of ground data is to employ a model whose structure is flexible enough to generalize across multiple geographical locations. As DNNs are powerful models that can generalize across tasks [17, 27], they are selected as the base model for the proposed forecaster. This concept of generalization is further explained in Section 3.6.1.

While the model is a DNN as the one illustrated in Figure 1, the number of layers, the size of the output, and the type of inputs are specifically selected according to the application. In particular, considering that 6 hours is the limit predictive horizon before NWP forecast outperform time series models [6], the model consists of 6 output neurons representing the forecasted hourly irradiance over the next 6 hours; this horizon is the standard choice for short-term irradiance forecasting [6].

In terms of hidden layers, the model is not subject to any specific depth; instead, depending on the case study, i.e. the geographical area where the forecasts are made, the number of hidden layers are optimized using hyperparameter optimization as explained in Sections 2.2. For the case study in this paper, i.e. forecasting irradiance in the Netherlands, the optimal network depth is 2 hidden layers. To select the number of neurons per layer, the same methodology applies, i.e. they need to be optimized for each geographical location.

3.2. Model Inputs

As indicated in the introduction, the aim of the model is to forecast solar irradiance without the need of ground data. As a result, to perform the selection of model inputs, it is paramount to consider the subset of inputs that, while correlating with solar irradiance, are general enough so that they can be easily obtained for any given location. Given that restriction, the proposed model considers three types of inputs: NWP forecasts of the solar irradiance, the clear-sky irradiance, and satellite images representing maps of past solar irradiance.

3.2.1. Numerical weather prediction forecast

The first type of input are NWP forecasts of the solar irradiance obtained from the *European center for medium-range weather forecasts (ECMWF)*. As indicated in the introduction, NWP forecasts of the solar irradiance are less accurate than time series models for short-term horizons. However, as they strongly correlate with the real irradiance, they are very useful regressors to build time series models.

For the proposed model, the input data consists of the 6 forecasted values for the next 6 hours given by the latest available ECMWF forecast (typically available every day around 08:00-09:00 CET).

3.2.2. Clear-sky irradiance

As second input, the model considers the clear-sky irradiance \mathcal{I}_c , i.e. the GHI under clear-sky conditions, at every hour over the next 6 hours. The clear-sky irradiance is a deterministic input that is obtained using the clear-sky model defined in [34], which computes \mathcal{I}_c using the location and time of interest.

3.2.3. Satellite images

The third input are satellite data representing the past irradiance values of a geographical area. In particular, the input data consists of images from the SEVIRI instrument of the METEOSAT satellite that are transformed to irradiance values using two different methods:

1. For data corresponding to solar elevation angles above 12° , the SEVIRI-based images are mapped to irradiance values using the *Surface insolation under clear and cloudy skies (SICSS)* algorithm [35].

2. For data corresponding to solar elevation angles below 12° , i.e. very early in the morning and late in the evening, the irradiance values are extracted by considering the interpolation method described in [36] applied to the clear sky index.

This distinction depending on the solar elevation angle is required because: (1) the SICSS method considers cloud properties; (2) at low solar elevation angles the uncertainty in the cloud properties increases strongly [36].

Once the satellite images are mapped to irradiance values, the input data simply consists of the past irradiance values in the individual pixel where the forecasting site is located. Then, to select which past irradiance values, i.e. which past images, are relevant for building the general model, the feature selection method defined in Section 2.2 is employed.

As a final remark, it is important to note that these irradiance values have a resolution that is limited by the resolution of the satellite images, which in the case of the SEVIRI instrument are pixels of 3×3 km. As a result, to represent the solar irradiance in a specific location, the accuracy of satellite-based measurements cannot be better than that of ground measurements.

3.2.4. Input selection

The three input features that the proposed model considers were selected from a larger set of input features. In particular, in order to ensure that the proposed model included the most relevant input features, a feature selection process was performed. During this feature selection process, the three considered inputs, i.e. the NWP forecasts, the clear-sky irradiance, and the satellite images were selected as the most important features. However, in addition to these three, four other features were also considered:

- Historical values of the temperature.
- Historical values of the humidity.
- Forecast of the temperature.
- Forecast of the humidity.

To perform the feature selection between these 7 input features, the feature selection method described in [27] was employed; i.e. the 7 input features were modeled as binary hyperparameters and the selection was performed together with the hyperparameter optimization described in Section 3.3. This optimization resulted in the 3 selected inputs.

3.3. Hyperparameter Optimization and Feature Selection

As briefly introduced in Section 3.1, the proposed model needs to be tuned for the specific geographical area where it is applied. In order to tune the model structure, the following four DNN hyperparameter are optimized:

1. **Number of hidden layers:** the neural network depth is a parameter that needs to be tuned in order to obtain a model that can correctly generalize across multiple geographical locations.
2. **Number of neurons per layer:** besides the number of hidden layers, the size of each layer also plays an important role in the generalization capabilities of the DNN.
3. **General learning rate:** the initial learning rate used in the stochastic gradient descent method. In particular, while the stochastic gradient descent method automatically adapts the learning rate at every iteration of the optimization process, the learning rate at the first iteration has to be selected.
4. **Dropout:** Dropout [37] is included as a possible regularization technique to reduce overfitting and to improve the training performance. To do so, at each iteration, dropout selects a fraction of the neurons and prevents them from training. This fraction of neurons is defined as a real hyperparameter between 0 and 1.

As explained in Section 2.2 and 3.2.4, in combination with the hyperparameter optimization, the proposed model also performs a feature selection. In particular, the feature selection method selects the most relevant inputs among a subset of 7 features and it also selects which past historical irradiance values are required.

3.4. Model Parameters

The parameters of the DNN are represented by the set of weights that establish the mapping connections between the several neurons of the network:

- $\mathbf{W}_{1,i}$: the vector of weights between the input \mathbf{X} and the neuron i of the first hidden layer.
- $\mathbf{W}_{k,i}$: the vector of weights between the k^{th} hidden layer and the neuron i of the $(k+1)^{\text{th}}$ hidden layer.
- $\mathbf{W}_{o,i}$: the vector of weights between the last hidden layer and the irradiance price vector $\hat{\mathcal{I}}$.
- $\mathbf{b}_k = [b_{k1}, \dots, b_{kn_k}]^\top$: the vector of bias weights in the k^{th} hidden layer, with $k = 1, 2$.
- $\mathbf{b}_o = [b_{o,1}, \dots, b_{o,6}]^\top$: the vector of bias weights in the output layer.

3.5. Model Equations

Using the above definitions, the equations of the DNN assuming two hidden layers can be defined as:

$$z_{1i} = f_{1i}(\mathbf{W}_{1,i}^\top \cdot \mathbf{X} + b_{1i}), \quad \text{for } i = 1, \dots, n_1, \quad (5a)$$

$$z_{2i} = f_{2i}(\mathbf{W}_{2,i}^\top \cdot \mathbf{z}_1 + b_{2i}), \quad \text{for } i = 1, \dots, n_2, \quad (5b)$$

$$\hat{\mathcal{I}}_{h+i} = \mathbf{W}_{o,i}^\top \cdot \mathbf{z}_2 + b_{o,i}, \quad \text{for } i = 1, \dots, 6, \quad (5c)$$

where f_{ki} represents the activation function of neuron i in the k^{th} hidden layer. In particular, for the proposed model, the *rectified linear unit (ReLU)* [38] is selected as the activation function of the two hidden layers. This choice is made because this activation function has become a standard for hidden layers of DNNs [17]. It is important to note that, as the irradiance is a real number, no activation function is used for the output layer.

3.6. Training

The DNN is trained by minimizing the mean square error². In particular, given the training set $\mathcal{S}_{\mathcal{T}} = \{(\mathbf{X}_k, \hat{\mathbf{I}}_k)\}_{k=1}^N$, the optimization problem that is solved to train the neural network is:

$$\underset{\mathbf{W}}{\text{minimize}} \quad \sum_{k=1}^N \|\hat{\mathbf{I}}_k - F(\mathbf{X}_k, \mathbf{W})\|_2^2, \quad (6)$$

where $F: \mathbb{R}^n \rightarrow \mathbb{R}^6$ is the neural network map and \mathbf{W} is the set comprising all the n weights and bias weights of the network.

3.6.1. Generalizing across geographical sites

A key element for the model to forecast without the need of ground data is to be able to generalize across locations. To do so, the proposed model is trained across a small subset of sites so that the model learns to generalize across geographical sites. It is important to note that, while ground data is required for this small subset of locations, the model generalizes across all other geographical locations where ground data is not needed. In particular, as it is shown in the case study for The Netherlands, the number of locations where ground data is required is relatively small, e.g. 3-5 sites.

3.6.2. Generalizing across predictive horizons

Enforcing generalization is not only good for obtaining a model that does not require ground data, but in general, it is also beneficial to obtain a DNN that does not overfit and that obtains more accurate predictions [17]. In particular, as it has been empirically shown in several studies [27, 28], by forcing the network to solve multiple related task, e.g. forecasting multiple sites, the network might learn to solve individual tasks better.

Therefore, to further strengthen the generalization capabilities of the network, the DNN is trained to forecast over the next 6 hours but starting at any hour of the day. As with the geographical site generalization, the goal is to build a DNN that, by performing several related tasks, it is able to learn more accurate predictions.

²Note that minimizing the mean square error is equivalent to minimizing the rRMSE metric used throughout the paper to evaluate and compare the model.

3.6.3. Implementation details

The optimization problem is solved using multi-start optimization and Adam [39], a version of stochastic gradient descent that computes adaptive learning rates for each model parameter. The use of adaptive learning rates is selected for a clear reason: as the learning rate is automatically computed, the time needed to tune the learning rate is smaller in comparison with other optimization methods. Together with Adam, the forecaster also considers early stopping [40] to avoid overfitting.

3.7. Issues

Note that the proposed model depends on another type of forecasts provided by NWP models. As a consequence, if the NWP models are performing bad, they might impact the final performance of the prediction model. For the proposed model, one of the most accurate and well-known NWP forecast models is considered: the ECMWF forecast [41]. If other NWP models are employed instead, the performance of the model might vary w.r.t. the results shown in this paper.”

3.8. Representation

Defining by h the current hour, by $\hat{\mathcal{I}}_E$ the values of the ECMWF forecast, by \mathcal{I}_S the irradiance values obtained from the satellite image, by \mathcal{I}_c the clear-sky irradiance, and by $\hat{\mathcal{I}}$ the forecasted values of the proposed model, the forecasting model can be represented as in Figure 2. In this representation, it was assumed that the optimal depth was 2 hidden layers, and that the optimal past irradiance values are lags 0, 1, and 2 w.r.t. the current hour h ; i.e. $\mathcal{I}_{S,h}$, $\mathcal{I}_{S,h-1}$, $\mathcal{I}_{S,h-2}$; and lag 24 w.r.t. the 6 prediction hours $h+1, \dots, h+6$; i.e. $\mathcal{I}_{S,h-23}, \dots, \mathcal{I}_{S,h-18}$.

4. Case study

In order to evaluate the proposed model, 30 sites in the Netherlands are considered and the accuracy of the proposed model is compared with that of specific models individually trained using local data.

4.1. Data description

The dataset spans four years, i.e. from 01/01/2014 until 31/12/2017, and comprises, for each of the 30 sites, the following four types of input data:

1. The historical ground data measured on site.
2. The satellite-based irradiance values.
3. The daily ECMWF forecasts.
4. The deterministic clear-sky irradiance.

In all four cases, these data represent hourly average values between two consecutive hours. In particular, a variable given at a time step h represents the average variable between hours h and $h+1$, e.g. the irradiance $\mathcal{I}_{S,12}$ is the average irradiance obtained from satellite images between hours 12 and 13.

4.1.1. Data Sources

For the irradiance values obtained from SEVIRI satellite images, the processed irradiance values are directly obtained from the *Royal Netherlands Meteorological Institute (KNMI)* via their Cloud Physical Properties model [42].

For the ground measurements, 30 of the meteorological stations in The Netherlands that are maintained by the KNMI [42] and that measure irradiance values using pyranometers are considered. In particular, the following 30 stations are employed: Arcen, Berkhout, Cabauw, De Kooy, De Bilt, Deelen, Eelde, Eindhoven, Ell, Gilze-Rijen, Heino, Herwijnen, Hoek van Holland, Hoogeveen, Hoorn (Terschelling), Hupsel, Lauwersoog, Leeuwarden, Lelystad, Maastricht, Marknesse, Nieuw Beerta, Rotterdam, Schiphol, Stavoren, Twente, Vlissingen, Volkel, Westdorpe, and Wijk aan Zee. The geographical location of these 30 stations is illustrated in Figure 3.

The ECMWF forecasts are directly obtained through the ECMWF website [41]. Finally, for the clear-sky irradiance, the `python PVLIB` library [43] that implements the clear-sky model [34] defined in Section 3.2 is used.

4.1.2. Data division

In order to perform the study, the data is divided into three subsets:

1. Training set (01/01/2014 to 31/12/2015): these 2 years of data are used for training and estimating the various models.
2. Validation set (01/01/2016 to 31/12/2016): a year of data is used to select the optimal hyperparameters and features, and to perform early-stopping when training the network.
3. Test set (01/01/2017 to 31/12/2017): a year of data that is not used at any step during the model estimation process, is employed as the out-of-sample data to compare the proposed model against local models.

In addition to the time separation, the data is further divided according to the location:

1. Of the 30 sites, 5 are used to train the proposed models. In particular, the following 5 were randomly selected: Herwijnen, Wijk aan Zee, Schiphol, Twente, and Lelystad.
2. The remaining 25 act as out-of-sample data to show that the model can predict irradiance at any site without the need of local data.

This separation is depicted in Figure 3, which represents the geographical distribution of the 30 sites distinguishing between training and test sites. In short, the proposed model is trained using data from 5 sites spanning three years and it is evaluated in 25 additional locations and using an additional year of data.

It is important to note that the above separation in 5+25 locations only applies for the proposed model. In particular, for the local models used as benchmark, the data

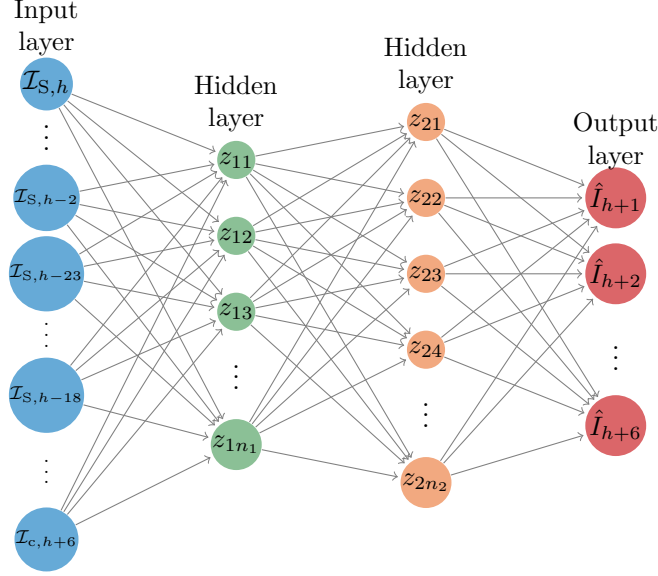


Figure 2: DNN to forecast day-ahead prices.

division is only performed as a function of time as, by definition, each local model considers only local data.

4.1.3. Data Preprocessing

To evaluate the proposed models, the hours of the day for which the irradiance is very small are disregarded. In particular, those hours that correspond with solar elevation angles below 3° are disregarded. This limitation on the solar elevation angles implies that the number of forecasts per day available to evaluate the model changes throughout the year; e.g. while in June the model makes 11-12 forecasts per day, in January that number is reduced to 3-4.

In addition to the above preprocessing step, the hourly time slots that have missing values are also disregarded.

4.2. Local models

To compare the proposed forecaster, four types of local models are considered: a persistence model [6], an *autoregressive model with exogenous inputs (ARX)* [11], a *gradient boosting tree (GBT)* algorithm [44], and a local neural network [11].

Moreover, in addition to the local models, it is also included in the benchmark the ECMWF forecast. By doing so, the accuracy between the time series models and the NWP forecast can be compared as a function of the prediction horizon, .

4.2.1. Persistence model

When evaluating a new model, a standard approach in the literature of irradiance forecasting is to check whether the new model provides better predictions than a trivial model [6]. Moreover, the trivial model normally used is a

persistence model, which assumes that the clear sky index k_c does not change from one time interval to the other [6].

In particular, given the irradiance \mathcal{I}_h at the current hour h , the clear sky index at h is defined as the ratio of \mathcal{I}_h to the clear sky irradiance $\mathcal{I}_{c,h}$, i.e.:

$$k_{c,h} = \frac{\mathcal{I}_h}{\mathcal{I}_{c,h}}. \quad (7)$$

Then, defining by $\mathcal{I}_{c,h+p}$ the clear sky irradiance at the prediction time $h+p$, the persistence model forecasts the irradiance \mathcal{I}_{h+p} at the prediction time $h+p$ as follows:

$$\hat{\mathcal{I}}_{h+p} = k_{c,h} \mathcal{I}_{c,h+p} = \frac{\mathcal{I}_h}{\mathcal{I}_{c,h}} \mathcal{I}_{c,h+p}. \quad (8)$$

4.2.2. Linear model

Another standard benchmark choice in the literature of irradiance forecasting are autoregressive linear models [6, 11]; hence, the second model considered in the comparison is a linear autoregressive model that can optimally select its exogenous inputs. As the model is local, a different model per location, per hour of the day h , and for prediction time $h+p$ is considered. Therefore, as the proposed model is evaluated in 25 locations, 6 forecasts per day are made, and each forecast is made for 6 prediction times, a total of $25 \times 6 \times 6 = 900$ models are estimated.

The exogenous inputs of these models are similar to the DNN, but instead of using the satellite irradiance maps \mathcal{I}_S , the models consider the historical irradiance ground measurements \mathcal{I}_G . In particular, the model for the prediction time $h+p$ considers the clear-sky irradiance $\mathcal{I}_{c,h+p}$ and the ECMWF forecast $\hat{\mathcal{I}}_{E,h+p}$ at the prediction time. For the historical irradiance values \mathcal{I}_G , as with the global model and the satellite-based irradiance \mathcal{I}_S , the specific

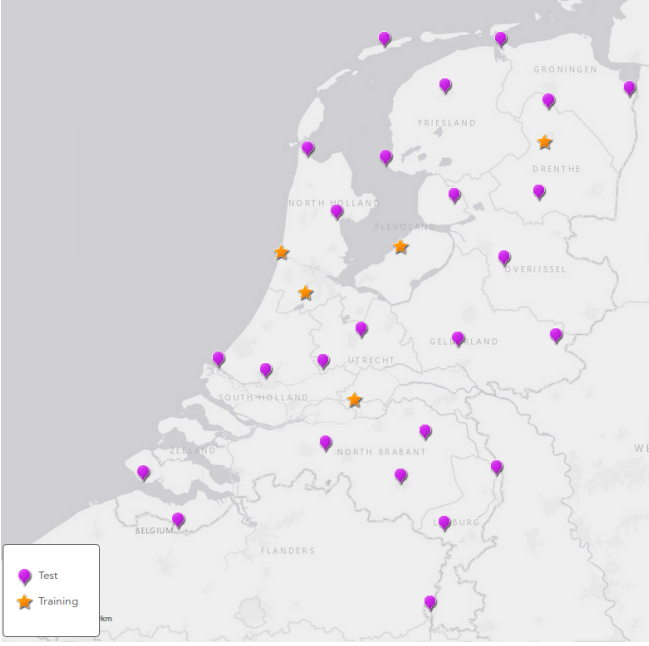


Figure 3: Geographical distribution of the 30 sites in the case study. The blue dots are the 5 sites used for estimating the model. The red dots represent the 25 out-of-sample sites to evaluate the model.

lagged values are optimally selected using the feature selection method described in Section 2.2. In addition, to ensure that the differences between models are not due to differences in input data, the model is allowed to choose satellite data through the feature selection method.

4.2.3. Gradient boosting tree

As a third model, the XGBoost algorithm [44] is considered, a GBT model that predicts data by combining several regression trees. In particular, the model is based on the principle of boosting [45, Chapter 10], i.e. combining models with high bias and low variance in order to reduce the bias while keeping a low variance.

It is important to note that, while several models based on regression trees have been proposed in the literature for forecasting solar irradiance [2], the XGBoost algorithm has, to the best of our knowledge, not yet been used. Nevertheless, including this model in the benchmark was decided for two reasons: (1) it has been shown to outperform other regression tree methods and has recently become the winner of several challenges in Kaggle, a site that hosts machine learning competitions [44]; (2) it has been successfully used in other energy-based forecasting applications, e.g. forecasting electricity prices [28].

As with the linear model, a different GBT per location, hour, and prediction time is estimated; i.e. 900 different models are estimated. Similarly, the model inputs are the same as the linear models, i.e. the clear-sky irradiance $\mathcal{I}_{c,h+p}$ and the ECMWF forecast $\hat{\mathcal{I}}_{E,h+p}$ at the prediction time, and the historical irradiance values \mathcal{I}_G optimally selected using the feature selection method. In addition, to

ensure that the differences between models are not due to differences in input data, the model is allowed to choose satellite data through the feature selection method.

It is important to note that, as done with the proposed DNN, all the GBT hyperparameters (see [44]) are optimally selected using the hyperparameter optimization algorithm defined in Section 2.2.

4.2.4. Neural network

As a fourth model, a local DNN that considers very similar inputs, outputs, structure, and training algorithm as the proposed global DNN is considered. The main difference w.r.t. to the proposed DNN is that it considers the local measurements of the irradiance \mathcal{I}_G in addition to the satellite irradiance maps \mathcal{I}_S . However, the type and number of hyperparameters that the model optimizes are the same as for the global DNN and they are also optimized using the hyperparameter optimization algorithm defined in Section 2.2.

The reason for including this model in the case study is that, similar to the linear and the persistence models, neural networks are a standard choice in the literature of solar irradiance forecasting [2, 36].

As the proposed DNN is evaluated in 25 sites and the model is local, 25 different local DNNs are estimated. Unlike the linear and GBT models, the same DNN is used for the different hours of the day; this was done because it was empirically observed that the distinction of a different DNN per hour of the day led to worse predictive accuracy.

4.3. Hyperparameter Optimization and Feature Selection

As defined in Section 3, the hyperparameters and input features of the global DNN are optimally selected according to the geographical location. In this case study, the range of the hyperparameters considered in the optimization search and their obtained optimal values are listed in Table 1.

Table 1: Optimal hyperparameters for the global DNN.

Hyperparameter	Value	Search Range
Number of hidden layers	2	{1, 2, 3, 4}
Neurons in 1 st layer	208	[100, 400]
Neurons in 2 nd layer	63	[50, 150]
Initial Learning Rate	1.16×10^{-3}	$[10^{-4}, 10^{-2}]$
Dropout	0.14	[0, 1]

In terms of the lagged satellite-based irradiance values, the optimal input features are defined by the irradiance values at lags 0, 1, 2, and 3 w.r.t. the current hour h ; i.e. $\mathcal{I}_{S,h}, \dots, \mathcal{I}_{S,h-3}$; and at lag 24 w.r.t the 6 prediction hours $h+1, \dots, h+6$; i.e. $\mathcal{I}_{S,h-23}, \dots, \mathcal{I}_{S,h-18}$.

For the local models, the hyperparameters and input features are also optimized. However, considering that 900 linear models, 900 GBT models, and 25 local DNNs are

used, displaying all their optimal hyperparameters and input features is out of the scope of this paper. However, the main results can be summarized as follows:

1. In terms of input features, all the local models (except for the persistence model) performed two types of selection:
 - (a) Use satellite data in addition to local data.
 - (b) Choose the relevant historical irradiance values.
 The addition of satellite data did not improve the performance w.r.t. using ground data only; therefore, none of the local models considered this information. In addition, in terms of ground irradiance values \mathcal{I}_G , all the local models consider the irradiance values at lags 0 and 1 w.r.t. the current hour h and at lag 24 w.r.t. the prediction hour $h + p$. In addition, most of them also consider the irradiance values at lags 2 and 3 w.r.t. the current hour h ; the exception are models that predict the solar irradiance at early hours of the day when lags of 2-3 hours represent irradiance values of 0.
2. In the case of the local DNNs, the number of hidden layers is 2 for all 25 sites. Moreover, the number of neurons in the first (second) hidden layer varies from 95 to 242 (51 to 199) neurons depending on the site. Similarly, the dropout and the learning rate respectively oscillate between 0 and 0.45, and between 5.825×10^{-4} and 5.800×10^{-2} .
3. In the case of the GBT models, the range of the hyperparameters values varies in a larger range, e.g. the number of trees per model fluctuates between 10 and 1000 and the depth of each tree varies between 1 and 20.

4.4. Overall results

After defining the setup of the case study and describing the selection of hyperparameters and features, in this section the average performance of the global DNN is compared against that of the local models. Particularly, the first metrics to take into account to compare the models are the average metrics; i.e. rRMSE, forecasting skill s , and MBE; across the 25 sites and the 6 prediction times. These average metrics are listed in Table 2, where the forecasting skill was computed using the same window length employed in [33], i.e. 200 samples³.

From Table 2, several observations can be drawn:

1. In terms of square errors, i.e. rRMSE, the predictive accuracy of the proposed global model is slightly better than all the local models and significantly better than some of them, in particular the GBT model or the persistence model. Among the local models, both the linear and local DNN perform the best and the persistence model the worst.

³As in [33], the window length for which s was stable was analyzed. Similar to [33], 200 samples were found to be a reasonable value.

Table 2: Comparison of the average predictive accuracy across sites and prediction times by means of rRMSE, forecasting skill s , and MBE.

Model	rRMSE [%]	s [%]	MBE [W/m^2]
Global DNN	31.31	22.42	-1.04
Linear	32.01	21.22	-1.07
Local DNN	32.10	19.29	-1.43
ECMWF	34.94	9.75	-2.52
GBT	35.85	9.92	1.50
Persistence	41.98	0	11.60

2. This same observation can be inferred from looking at the forecasting skill: the proposed global model performs similar to the linear model, slightly better than the local DNN, and much better than the other models. In addition, when compared across all sites and predictive horizons, all models perform better than the persistence model.
3. In terms of model bias, i.e. MBE, all models show a very small bias that indicates that the models are not biased. Particularly, considering that the average irradiance of the dataset is approximate $350 \text{ W}/\text{m}^2$, the bias of all the models is around 0.3-0.8% of the average irradiance, which represents a negligible bias. The exception to this is the persistence model, whose bias of 3% of the average irradiance is a bit larger, but still quite small.

4.5. Comparison with previously validated forecast models

While the proposed global model seems to be a good replacement of the local models considered in this paper, it is also very important to establish its quality w.r.t. previously validated forecast models from the literature. As explained in Section 2.3, while this comparison cannot fairly be done using a metric like rRMSE, it can be roughly assessed using the forecasting skill s . In particular, using the results of [33], we can establish a comparison between the proposed global model, the local NARX model proposed in [33], and the cloud motion forecast of [14]. As both models from the literature were originally only evaluated for 1-hour step ahead forecasts, we also limit the comparison of the global model to that interval. The comparison is listed in Table 3.

Table 3: Comparison of the average predictive accuracy between the global model, a NARX model from the literature, and a cloud moving forecast from the literature. The comparison is done for 1-hour ahead forecasts and by means of forecasting skill.

Model	s [%]
Global DNN	10
NARX [33]	12
Cloud moving [14]	8

What can be observed from these results is that the overall quality of the proposed global model for 1-hour ahead forecasts is very similar to those from the literature. Therefore, as initially observed when comparing the average performance of the global model w.r.t. to the local model considered in this paper, the proposed global model seems to be an excellent candidate to save the operational costs of installing local sensors and collecting ground measurements.

4.6. Comparison across prediction horizons

A third step required to analyze the performance of the proposed global model is to verify that its average performance is satisfied across all prediction times. In particular, it is important to check whether the global models can build accurate predictions at all short-term horizons. To perform this comparison, the two metrics used for comparing predictive accuracy, i.e. rRMSE and the forecasting skill s , are evaluated for each benchmark model and predictive horizon. This comparison is listed in Table 4 and illustrated in Figure 4.

Table 4: Comparison of the predictive accuracy of the various forecasters across the 6 prediction times by means of rRMSE and forecasting skill s . The best model is marked with bold font.

Horizon [h]	1	2	3	4	5	6
Model	rRMSE [%]					
Global DNN	25.07	30.18	32.36	34.19	36.10	38.71
Linear	26.67	31.36	33.11	34.63	36.44	38.35
Local DNN	26.82	30.90	32.91	34.67	36.68	39.88
GBT	30.05	34.78	36.95	39.04	40.67	43.59
Persistence	28.74	36.89	42.29	47.28	52.05	56.69
ECMWF	35.91	35.01	35.12	35.91	37.45	39.28
	s [%]					
Global DNN	9.98	18.38	23.40	27.04	28.30	27.38
Linear	7.67	15.71	21.73	26.03	27.76	28.42
Local DNN	6.34	16.98	22.13	22.64	25.13	22.51
GBT	-5.18	6.06	12.23	15.29	16.00	15.11
Persistence	0	0	0	0	0	0
ECMWF	-29.07	4.68	16.77	22.74	23.23	20.19

As can be seen from Table 4 and Figure 4, the global model seems to be the best model for the first 5 prediction horizons (both in terms of rRMSE and forecasting skill s), and the second best (very close to the best one) for the last prediction horizon. Based on these results it can be observed that not only the global model is overall equal or better than the local models, but it also performs equally well or better than them across all prediction horizons. As a result, the proposed model is a very promising candidate to replace the local models and to save operational costs without compromising the forecasting quality.

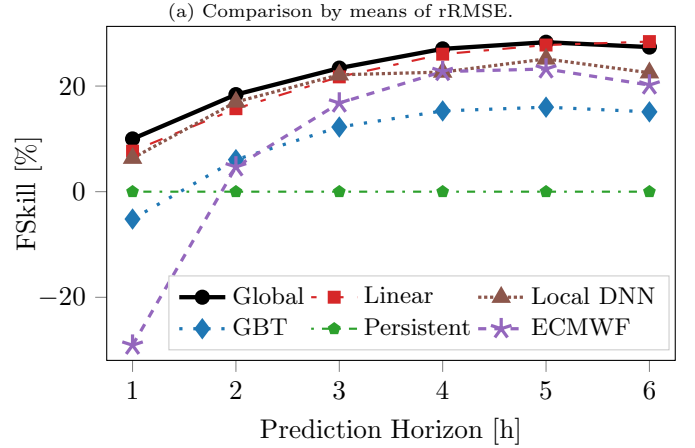
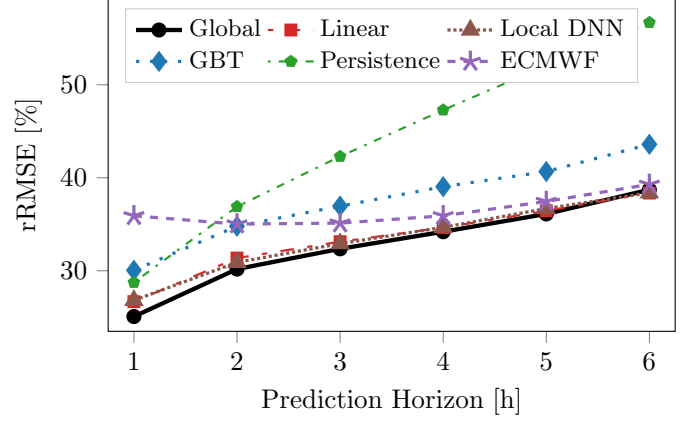


Figure 4: Comparison of the predictive accuracy of the various forecasters across the 6 prediction times.

In addition to this analysis of the global model performance, three additionally interesting observations can be made:

1. The persistence model is the worst across all prediction horizons except the first one. This result agrees with previous results from the literature [6] that stated that the persistence model only provides reasonable results for prediction horizons shorter than 1 hour.
2. Among the local models, the linear and DNN models show the best performance across all 6 prediction horizons.
3. The ECMWF forecast improves its accuracy relatively to the other models as the prediction time increases. In particular, in the case of the last prediction time, the ECMWF forecast has almost the same performance as the global DNN and the linear models. Considering previous results from the literature [6], this is highly expected as NWP models start to perform better than time series models for prediction horizons larger than 4-6 hours.
4. For 1 hour ahead predictions, the ECMWF model is

the worst; specially, considering its s value for the first prediction horizon, the weather-based model is much worse than a simple persistence model.

4.7. Comparison across geographical site

The final step to analyze the better or equal performance of the global model is to validate whether the quality of the performance is kept across the 25 different sites. In particular, it is important to check whether the global model can generalize and build accurate predictions across all geographical locations. For the same of simplicity, this comparison is only done in terms of the rRMSE metric; in particular, as it was the case with all previous results, the values of the forecasting skill s fully agree with the rRMSE across all locations, and they are a bit redundant. The comparison across the geographical locations is listed in Table 5 and illustrated in Figure 5.

As it can be seen from Table 5 and Figure 5, the global model seems to validate and maintain its performance across all geographical locations. In particular, analyzing this results, it is clear that the global model performs equal or better than the local models across all 25 sites. In particular, as listed in Table 5, the global DNN is the best model for 20 of the 25 locations, and shows an rRMSE performance that is very similar to the best model in the remaining 5 locations. Therefore, it can again be conclude that the global model is a good replacement for the local models as the performance of the former is, at least, equal to the performance of the latter.

4.7.1. Geographical dependencies

An interesting study to analyze is whether the rRMSE has any geographical dependence, i.e. it might be possible that geography or climate might have an effect on the rRMSE. To study this effect, a color map with the geographical distribution of the rRMSE can be used. Such a plot is represented in Figure 6, which depicts the geographical distribution of the rRMSE for the 6 different models. As can be observed, there is a clear difference between coastal and island sites with the latter displaying rRMSEs that are consistently higher. While this difference is not notorious, it does seem to indicate that forecasting solar irradiance at inland locations is slightly harder than at coastal sites. While analyzing the causality behind this difference is out of the scope of this paper, it is worth noting possible reasons that might cause it; particularly, differences in climate, altitude, or simple differences in irradiance ranges might explain this effect.

4.7.2. rRMSE distribution

A second interesting study is to analyze the rRMSE distribution across sites. In particular, while the variability of the rRMSE can be visually observed in Figure 5, it is interesting to analyze its empirical distribution. To perform this analysis, the histogram of the rRMSE across the 25 sites is built for each of the 6 models. This is depicted in

Figure 7, where each histogram bin represents a width of 0.5% rRMSE. As it can be observed, the rRMSE distribution across the 6 locations is very similar with an interval spanning a width of 3%-4% rRMSE where the distribution is quite homogeneous and uniform, and an outlier on the right side representing a location with a much worse rRMSE. As can be seen from Figures 5 and 6, this site representing the worst case-scenario is the same for all models: Deelen. Based on this result it can be concluded that, while the rRMSE is site-dependent, the range of variability of the rRMSE is small.

4.8. Discussion

In the previous sections, the performance of the global model has been compared to that of the local models and that of validated models from the literature. Based on the obtained results one can conclude that: (1) the global model is slightly better than the best of the local models; (2) it performs similar to other models from the literature; (3) it provides unbiased forecasts.

While based on these results it cannot be stated that the proposed model is significantly better than all other models, it is important to keep in mind that its main purpose is not to be the best, but to perform equally well as local models so that the operational costs of installing and maintaining a wide sensor network are avoided. In that respect, it can be concluded that the proposed global model is an excellent replacement for the local models: the model is overall slightly better and performs better or equally well across all individual geographical locations and prediction times.

5. Conclusion

In this paper, a general model for short-term forecasting of the global horizontal irradiance has been proposed. The main features of the model are that it replaces ground measurements by satellite-based irradiance values and that, unlike local models previously proposed in the literature, it does not need local measurements in each location where a forecast is needed.

The proposed model was shown to be equal or better than local models typically used in the literature, and in turn, to be an excellent replacement of these local models in order to save the operational costs of installing local sensors and gathering ground data.

In future research, the current work will be expanded with two further investigations. First, the model will be extended to larger regions to analyze whether it generalizes to larger geographical areas than The Netherlands. Second, the model accuracy will be improved by adding other relevant sources of input data, e.g. weather-based input data like humidity levels or ambient temperature.

Table 5: Comparison of the predictive accuracy of the various forecasters by means of rRMSE. The best model is marked with bold font.

	Site												
Model	Arcen	Berkhout	Cabauw	De Kooy	Lauwers.	Deelen	Maastric.	Eindhov.	Westdorpe	Gilze-R.	Heino	Hoek v. H.	Ell
Global	32.39	30.24	30.75	29.49	30.32	34.55	30.82	32.11	32.07	32.37	32.80	29.24	32.42
Linear	33.03	31.05	31.01	29.87	31.16	35.47	31.73	32.28	33.11	32.89	32.75	30.53	32.50
DNN	33.43	32.77	31.27	31.14	30.95	35.75	31.48	32.03	31.93	33.04	32.89	29.66	32.77
GBT	35.80	35.41	35.20	33.45	35.79	39.62	35.68	37.22	36.33	36.30	36.88	33.61	37.35
Persistence	43.63	41.04	41.51	41.18	41.14	45.47	41.20	43.20	40.28	42.86	43.80	40.59	42.65
ECMWF	35.21	34.09	33.95	32.94	33.83	38.61	34.93	34.95	36.63	35.73	35.32	33.39	36.12
Model	Hoorn	Hoogev.	Hupsel	De Bilt	Leeuward.	Eelde	Marknes.	Rotterd.	Stavoren	Vlissing.	Volkel	Nieuw B.	
Global	29.63	31.44	32.88	31.68	30.16	31.58	31.19	30.21	29.38	30.81	32.46	32.37	
Linear	30.63	32.05	32.82	32.11	30.51	32.20	31.30	31.54	30.51	32.23	33.04	33.52	
DNN	30.24	33.05	32.83	32.02	31.97	31.62	31.72	31.25	29.85	31.92	34.68	32.34	
GBT	34.46	36.44	36.99	35.94	35.20	36.19	35.30	34.53	34.14	35.50	36.50	36.98	
Persistence	40.35	42.51	42.42	43.61	40.80	42.04	41.24	40.92	40.01	41.11	42.71	43.58	
ECMWF	33.62	35.19	35.11	34.69	34.17	35.34	34.85	34.92	33.35	35.05	35.33	36.27	

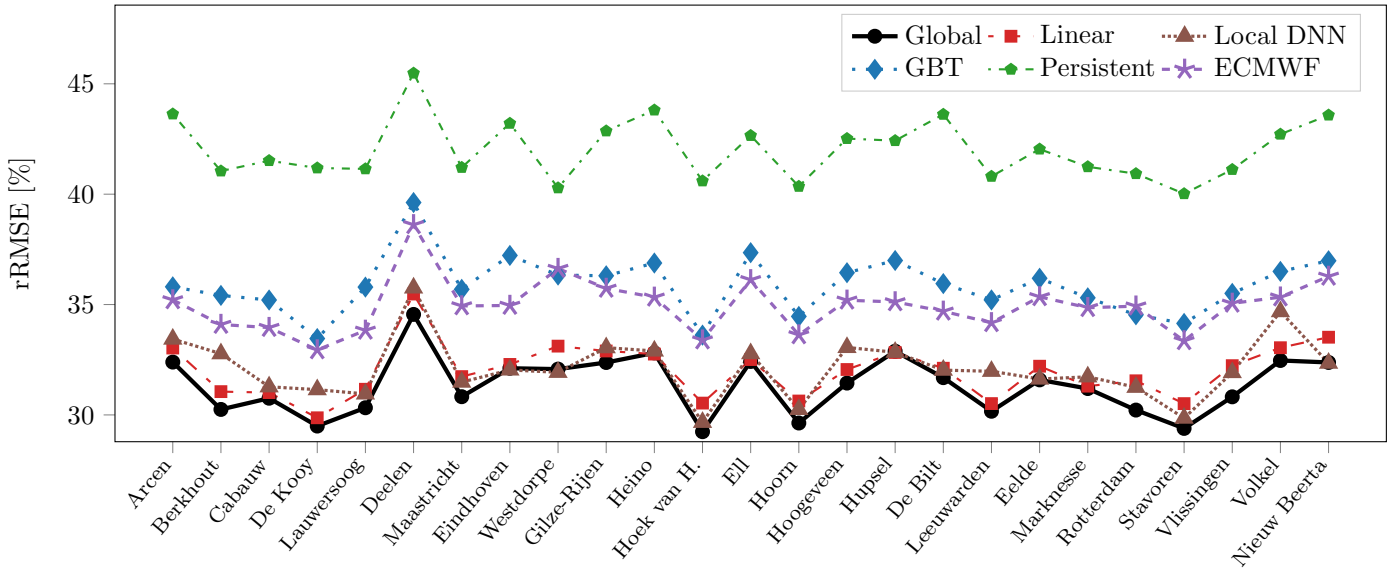


Figure 5: Comparison of the predictive accuracy of the various forecasters across the 25 locations.

Acknowledgment

This research has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 675318 (INCITE).

References

References

- [1] V. Lara-Fanego, J. A. Ruiz-Arias, D. Pozo-Vázquez, F. J. Santos-Alamillos, J. Tovar-Pescador, Evaluation of the WRF model solar irradiance forecasts in Andalusia (Southern Spain), *Solar Energy* 86 (8) (2012) 2200–2217. doi:10.1016/j.solener.2011.02.014.
- [2] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Machine learning methods for solar radiation forecasting: A review, *Renewable Energy* 105 (2017) 569–582. doi:10.1016/j.renene.2016.12.095.
- [3] A. Hammer, D. Heinemann, E. Lorenz, B. Lücke, Short-term forecasting of solar radiation: a statistical approach using satellite data, *Solar Energy* 67 (1) (1999) 139–150. doi:10.1016/S0038-092X(00)00038-4.
- [4] G. Reikard, Predicting solar radiation at high resolutions: A comparison of time series forecasts, *Solar Energy* 83 (3) (2009) 342–349. doi:10.1016/j.solener.2008.08.007.
- [5] E. W. Law, A. A. Prasad, M. Kay, R. A. Taylor, Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting – a review, *Solar Energy* 108 (2014) 287–307. doi:10.1016/j.solener.2014.07.008.
- [6] M. Diagne, M. David, P. Lauret, J. Boland, N. Schmutz, Review of solar irradiance forecasting methods and a proposition for small-scale insular grids, *Renewable and Sustainable Energy Reviews* 27 (2013) 65–76. doi:10.1016/j.rser.2013.06.042.

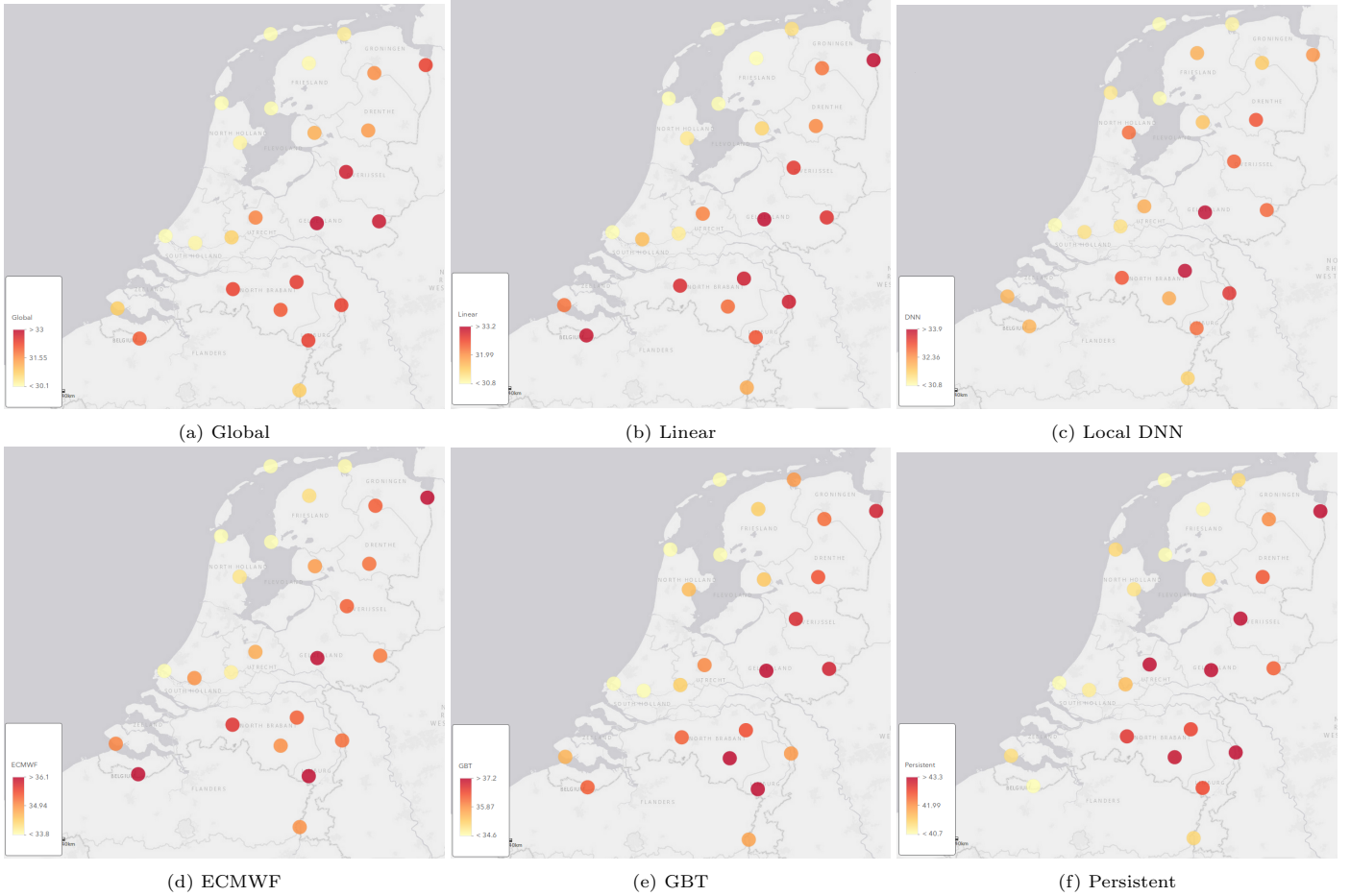


Figure 6: Geographical distribution of the rRMSE based on the 25 out-of-sample sites. Across the 6 models, it can be observed a clear difference between inland locations and coastal locations, with the latter having lower rRMSEs.

- [7] A. Ahmad, T. N. Anderson, T. T. Lie, Hourly global solar irradiation forecasting for New Zealand, *Solar Energy* 122 (2015) 1398–1408. doi:10.1016/j.solener.2015.10.055.
- [8] J. Huang, M. Korolkiewicz, M. Agrawal, J. Boland, Forecasting solar radiation on an hourly time scale using a coupled AutoRegressive and dynamical system (CARDS) model, *Solar Energy* 87 (2013) 136–149. doi:10.1016/j.solener.2012.10.012.
- [9] D. Yang, Z. Ye, L. H. I. Lim, Z. Dong, Very short term irradiance forecasting using the LASSO, *Solar Energy* 114 (2015) 314–326. doi:10.1016/j.solener.2015.01.016.
- [10] A. Mellit, A. M. Pavan, A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy, *Solar Energy* 84 (5) (2010) 807–821. doi:10.1016/j.solener.2010.02.006.
- [11] P. Lauret, C. Voyant, T. Soubdhan, M. David, P. Poggi, A benchmarking of machine learning techniques for solar radiation forecasting in an insular context, *Solar Energy* 112 (2015) 446–457. doi:10.1016/j.solener.2014.12.014.
- [12] T. C. McCandless, S. E. Haupt, G. S. Young, A model tree approach to forecasting solar irradiance variability, *Solar Energy* 120 (2015) 514–524. doi:10.1016/j.solener.2015.07.020.
- [13] E. Lorenz, D. Heinemann, Prediction of solar irradiance and photovoltaic power, in: A. Sayigh (Ed.), *Comprehensive Renewable Energy*, Elsevier, 2012, pp. 239–292. doi:10.1016/B978-0-08-087872-0.00114-1.
- [14] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker, D. Renné, T. E. Hoff, Validation of short and medium term operational solar radiation forecasts in the US, *Solar Energy* 84 (12) (2010) 2161–2172. doi:10.1016/j.solener.2010.08.014.
- [15] A. Sfetsos, A. H. Coonick, Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques, *Solar Energy* 68 (2) (200) 169–178. doi:10.1016/S0038-092X(99)00064-X.
- [16] D. P. Larson, C. F. M. Coimbra, Direct power output forecasts from remote sensing image processing, *Journal of Solar Energy Engineering* 140 (2) (2018) 021011–021011–8. doi:10.1115/1.4038983.
- [17] I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016, <http://www.deeplearningbook.org/>.
- [18] G. E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, *Neural Computation* 18 (7) (2006) 1527–1554. doi:10.1162/neco.2006.18.7.1527.
- [19] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: *Proceedings of the 25th International Conference on Neural Information Processing Systems, NIPS'12*, Curran Associates Inc., USA, 2012, pp. 1097–1105. doi:10.1145/3065386.
- [20] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, B. Kingsbury, Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups, *Signal Processing Magazine* 29 (6) (2012) 82–97. doi:10.1109/MSP.2012.2205597.
- [21] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, *arXiv eprint* (2014). arXiv:1409.0473.
- [22] H. Wang, G. Wang, G. Li, J. Peng, Y. Liu, Deep belief network based deterministic and probabilistic wind speed fore-

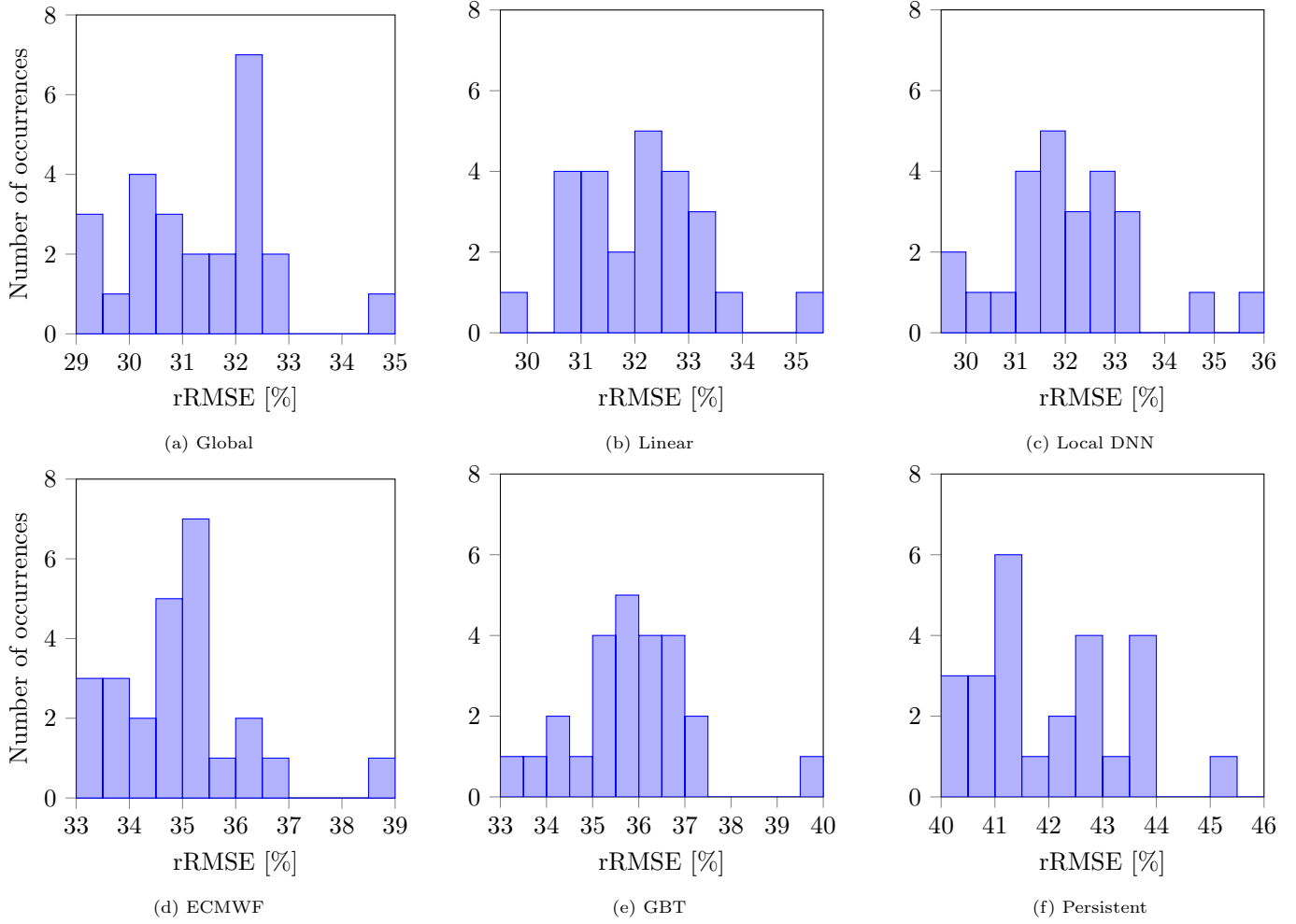


Figure 7: Distribution of the predictive accuracy of the global model across the 25 locations.

- casting approach, *Applied Energy* 182 (2016) 80–93. doi:10.1016/j.apenergy.2016.08.108.
- [23] C. Feng, M. Cui, B.-M. Hodge, J. Zhang, A data-driven multi-model methodology with deep feature selection for short-term wind forecasting, *Applied Energy* 190 (2017) 1245–1257. doi:10.1016/j.apenergy.2017.01.043.
- [24] G. Suryanarayana, J. Lago, D. Geysen, P. Aleksiejuk, C. Johansson, Thermal load forecasting in district heating networks using deep learning and advanced feature selection methods, *Energy* 157 (2018) 141–149. doi:10.1016/j.energy.2018.05.111.
- [25] I. Coelho, V. Coelho, E. Luz, L. Ochi, F. Guimarães, E. Rios, A GPU deep learning metaheuristic based model for time series forecasting, *Applied Energy* 201 (2017) 412–418. doi:10.1016/j.apenergy.2017.01.003.
- [26] C. Fan, F. Xiao, Y. Zhao, A short-term building cooling load prediction method using deep learning algorithms, *Applied Energy* 195 (2017) 222–233. doi:10.1016/j.apenergy.2017.03.064.
- [27] J. Lago, F. De Ridder, P. Vrancx, B. De Schutter, Forecasting day-ahead electricity prices in Europe: The importance of considering market integration, *Applied Energy* 211 (2018) 890–903. doi:10.1016/j.apenergy.2017.11.098.
- [28] J. Lago, F. De Ridder, B. De Schutter, Forecasting spot electricity prices: deep learning approaches and empirical comparison of traditional algorithms, *Applied Energy* 221 (2018) 386–405. doi:10.1016/j.apenergy.2018.02.069.
- [29] R. Weron, Electricity price forecasting: A review of the state-of-the-art with a look into the future, *International Journal of Forecasting* 30 (4) (2014) 1030–1081. doi:10.1016/j.ijforecast.2014.08.008.
- [30] S. Ruder, An overview of gradient descent optimization algorithms, arXiv eprint (2016). arXiv:1609.04747.
- [31] J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization, in: *Advances in Neural Information Processing Systems*, 2011, pp. 2546–2554.
- [32] F. Hutter, H. H. Hoos, K. Leyton-Brown, Sequential model-based optimization for general algorithm configuration, in: *International Conference on Learning and Intelligent Optimization*, Springer, 2011, pp. 507–523. doi:10.1007/978-3-642-25566-3_40.
- [33] R. Marquez, C. F. M. Coimbra, Proposed metric for evaluation of solar forecasting models, *Journal of Solar Energy Engineering* 135 (1) (2012) 011016–011016–9. doi:10.1115/1.4007496.
- [34] P. Ineichen, R. Perez, A new airmass independent formulation for the Linke turbidity coefficient, *Solar Energy* 73 (3) (2002) 151–157. doi:10.1016/S0038-092X(02)00045-2.
- [35] W. Greuell, J. Meirink, P. Wang, Retrieval and validation of global, direct, and diffuse irradiance derived from SEVIRI satellite observations, *Journal of Geophysical Research: Atmospheres* 118 (5) (2013) 2340–2361. doi:10.1002/jgrd.50194.
- [36] H. Deneke, A. Feijt, R. Roebeling, Estimating surface solar irradiance from METEOSAT SEVIRI-derived cloud properties, *Remote Sensing of Environment* 112 (6) (2008) 3131–3141. doi:10.1016/j.rse.2008.03.012.
- [37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever,

- R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research* 15 (2014) 1929–1958.
- [38] V. Nair, G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in: *Proceedings of the 27th international Conference on Machine Learning (ICML)*, 2010, pp. 807–814.
- [39] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv eprint (2014). [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [40] Y. Yao, L. Rosasco, A. Caponnetto, On early stopping in gradient descent learning, *Constructive Approximation* 26 (2) (2007) 289–315. doi:10.1007/s00365-006-0663-2.
- [41] European Centre for Medium-Range Weather Forecasts (ECMWF) website, <https://www.ecmwf.int/>.
- [42] Royal Netherlands Meteorological Institute (KNMI), <http://knmi.nl/>.
- [43] R. W. Andrews, J. S. Stein, C. Hansen, D. Riley, Introduction to the open source PV LIB for python photovoltaic system modelling package, in: *Photovoltaic Specialist Conference (PVSC)*, 2014 IEEE 40th, IEEE, 2014, pp. 0170–0174. doi:10.1109/pvsc.2014.6925501.
- [44] T. Chen, C. Guestrin, Xgboost: A scalable tree boosting system, in: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 785–794.
- [45] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer Series in Statistics, Springer New York Inc., New York, NY, USA, 2001. doi:10.1007/978-0-387-21606-5.