IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 25, 2020, accepted March 10, 2020, date of publication March 13, 2020, date of current version March 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2980775

Spatio-Temporal Resolution of Irradiance Samples in Machine Learning Approaches for Irradiance Forecasting

ANNETTE ESCHENBACH[1], GUILLERMO YEPES[1], CHRISTIAN TENLLADO[1], JOSÉ I. GÓMEZ-PÉREZ[1], LUIS PIÑUEL[1], LUIS F. ZARZALEJO[2], AND STEFAN WILBERT[3]

[1]Computer Architecture and System Engineering Department, Universidad Complutense de Madrid, 28040 Madrid, Spain
[2]CIEMAT, Energy Department, Renewable Energy Division, 28040 Madrid, Spain
[3]DLR, Institute of Solar Research, 04200 Tabernas, Spain

Corresponding author: Christian Tenllado (tenllado@ucm.es)

This work was supported in part by the Spanish Ministry of Science and Innovation under Grant RTI2018-093684-B-I00, and in part by the Regional Government of Madrid under Grant S2018/TCS-4423.

ABSTRACT Improving short term solar irradiance forecasting is crucial to increase the market share of the solar energy production. This paper analyzes the impact of using spatially distributed irradiance sensors as inputs to four machine learning algorithms: ARX, NN, RRF and RT. We used data from two different sensor networks for our experiments, the NREL dataset that includes data from 17 sensors that cover a 1 km^2 area and the InfoRiego dataset which includes data from 50 sensors that cover an area of 94 Km^2. Several studies have been published that use these datasets individually, to the author knowledge this is the first work that evaluates the influence of the spatially distributed data across a range from 0.5 to 17 sensors per km^2. We show that all of algorithms evaluated are able to take advantage of the data from the surroundings, from the very short forecast horizons of 10s up to a few hours, and that the wind direction and intensity plays an important role in the optimal distribution of the network and its density. We show that these machine learning methods are more effective on the short horizons when data is obtained from a dense enough network to capture the cloud movements in the prediction interval, and that in those cases complex non-linear models give better results. On the other hand, if only a sparse network is available, the simpler linear models give better results. The skills obtained with the models under test range from 13% to 70%, depending on the sensor network density, time resolution and lead time.

INDEX TERMS Machine learning, forecasting, spatial resolution, solar irradiance, global horizontal irradiance.

I. INTRODUCTION

Technology for solar energy production has improved considerably over the last two decades, becoming a cost-effective alternative to the fossil energy sources. Despite this evolution, the solar energy industry is seeking to improve the spatial and temporal resolution provided by current techniques for short term solar irradiance forecasting. Operators of the distribution networks wish to handle temporal resolutions in the range of 5-10 minutes whereas intra-day auctions in the energy markets require coarser time granularities, from 0.5 to 1-2 hours.

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

This need of accurate short-term forecasting has motivated the scientific community in the search for models that incorporate both spatial and temporal information. For instance [1]–[3] used sky imagers to take measurements of cloud positions across the target area, modeling their movement and predicting their shadows in the near future. Sky imagers have also been used to obtain a velocity map of the clouds [4], that can later be used to predict cloud movements over a network of radiometric sensors to forecast the solar irradiance in the near future [5]–[7]. A similar approach uses satellite images to obtain the cloud velocity map [8]. This kind of images have also been used to infer the solar irradiance over a specific region, applying then a classic time series analysis technique to forecast future values [9]. A more recent

proposal is to combine sky-imagers with shadow cameras to improve the cloud movement detection and predict future irradiance values on a given area [10].

An alternative approach being actively explored consists in using irradiance measurements from terrestrial sensors (or even the production PV cells themselves) as inputs to some statistical model, designed to extract patterns from past samples and predict from them future values. A large amount of statistical algorithms have been explored, from simple linear models like ARX [11]–[15], VARX [16], [17], LASSO [14], [17], [18] and ARIMA [19], to more complex non-linear models like Artificial Neural Networks [14], [20]–[22], Support Vector Regressors [18], [21], LSTM [23], Boosted Regressor Trees [18] or kriging [17], [24], [25]. The literature on this topic has been extensively reviewed [15], [26]–[28]. For instance, table 1 in [15] gives a detailed comparison of the techniques proposed in the literature on this topic, including information on the spatial coverage and time resolution used on each work.

However, it is still uncertain how much advantage can the statistical models take from spatio-temporal information. Some works report evidences that the information obtained from sensors in the surrounding of the target prediction improve the accuracy for some prediction intervals. For instance [20] explored the use of Artificial Neural Networks (ANNs) for short-term prediction of Global Horizontal Irradiance (GHI), incorporating in the feature vector measurements from different neighbouring sensors within a radius of 55km. Their results improved the prediction performance for intervals up to 3 hours. However, for longer prediction intervals, the use of the spatial information available did reduce their accuracy, i.e. using information only from the location of the prediction target resulted in more accurate predictions.

Amaro and Silva [15] studied the effect of the sensor distribution in the prediction accuracy obtained by ARX models. They used the NREL dataset, that provides data from 17 sensors covering an area of 1 km^2 with samples every second, and showed that for very short intervals (1 to 5 minutes) with fast moving clouds, a high density of sensors helps to improve prediction accuracy. More precisely, their ARX model could take advantage of the correlation between measurements in the sensors positioned along the direction of the dominant winds, when they were close enough for the clouds to cover the distance between nodes in periods larger than the prediction interval. The authors worked also on a second data set, that provided data from 57 photo voltaic plants spread in an area of 10^4 km^2. However, this second data set could not provide enough data to draw trustworthy conclusions on the matter.

More recently Chao Huang et.al. [18] compared the prediction accuracy obtained with five statistical models, ARX, LASSO, ANN, SVM and BRT, on a data set with irradiation measurements taken every 30 min, for a period of two years (2014-2015), in the Solar Technology Acceleration Center in Colorado. They trained the five methods on the data set, using Jaya [29] and grid search to obtain the hyper-parameters for the models. They observed that the ARX algorithm, that considers data from sensors different from the prediction target as exogenous data, provides slightly better accuracy than the AR model, in which only local data in the target sensor is considered. The authors did not perform this analysis with the rest of the algorithms they consider. Furthermore, they did neither analyze forecasting intervals below 30 min nor the contribution of each sensor individually.

This paper extends these previous works [15], [18], [20], by analyzing the influence of spatio-temporal data on four different statistical methods: ARX, ANNs, Random Regression Forests (RRF) and Regression Trees (RT), and two different data sources 1, one with a dense grid of sensors in a small area [30], and other with a lower density but covering a larger area [31]. We study which algorithm properties and which data features are more relevant in each case, to obtain a better understanding of the problem. To the authors knowledge, this is the first work that quantitatively analyzes the impact of the spatial resolution on several machine learning algorithms with spatial resolutions from 0.5 to 17 sensors per km^2 and forecasting intervals from 10s to a few hours.

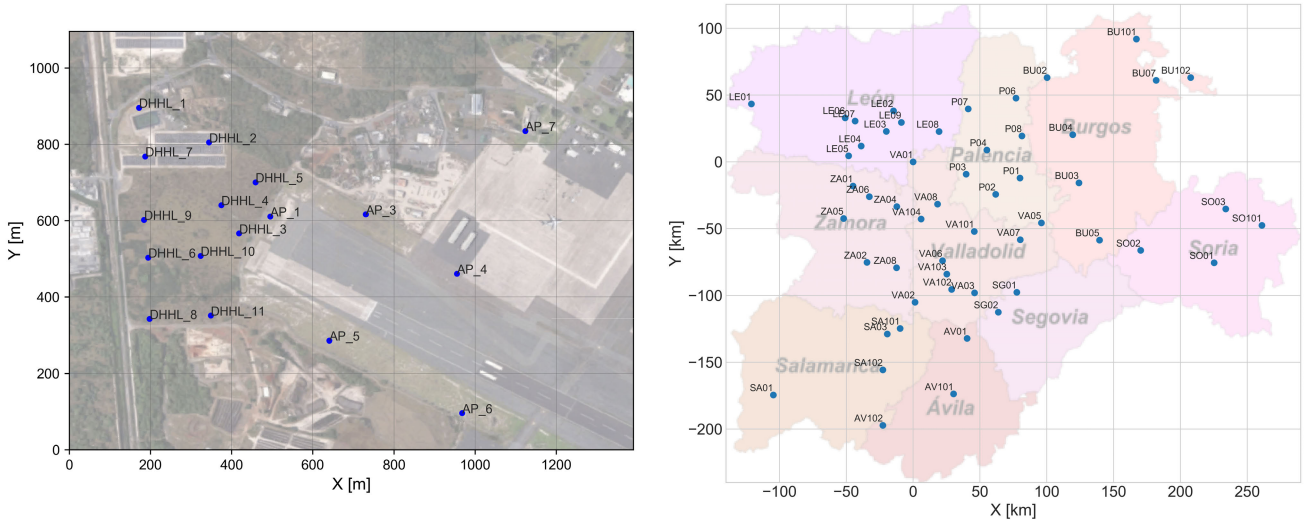
The rest of the paper is organized as follows. In Section II we describe the databases used for our work. Section III describes the statistical models we use, the feature selection and the metrics used to evaluate them. Section IV assesses the impact of the selected clear sky model on the accuracy of the statistical methods whereas Section V analyzes the impact on the prediction accuracy obtained by incorporating measurements from spatially distributed sensors. Section VI studies the influence of each feature in the prediction. Finally, Section VII summarizes the conclusions of the paper and sketches some future work.

II. DATA SOURCES

In this work we have used two data sources from very different areas, with different climates and different spatial and temporal scales.

A. NREL OAHU SOLAR MEASUREMENT GRID

This database is provided by the National Renewable Energy Laboratory ([30]), it includes GHI measurements from March 2010 to October 2011, taken at 1s intervals from 17 silicon pyranometers (LICOR LI-200) placed horizontally and distributed across an area of roughly $1km^2$ near the Kalaeloa Airport of the Oahu Island in Hawaii (USA). Figure 1a shows the irregular distribution of the sensor network. After a preliminary analysis of their data we decided to remove some days from the dataset for which the sensors were giving negative values and completely remove the AP_3 senor, for which almost all measurements where erroneous. According to [32] the dominant winds in this region come from the northeast, with an average speed of $5ms^{-1}$, which allows a cloud to traverse the whole area in about 3 minutes.



(a) NREL, $1km^2$

(b) InfoRiego, $94226km^2$

FIGURE 1. Distribution of sensors in the NREL and InfoRiego data sets.

B. InfoRiego NETWORK FROM ITACyL

This database is provided by the Instituto Tecnológico Agrario from the regional government of Castilla y León (CyL) in the north west of Spain. The network is composed of 50 weather stations irregularly distributed in an area of 94, $226km^2$, with an average distance of 25.94 km between stations. The layout of the stations is shown in Figure 1b. The dataset includes averaged measurements of GHI from silicon pyranometers (Campbell Skye SP1100), temperature and humidity on 30min intervals. According to the agricultural authorities in the region [33], the direction of the dominant winds is not homogeneous on the large area covered by these sensors.

III. METHODOLOGY

In this work we evaluate the forecasting accuracy obtained by exploiting spatial and temporal data with different machine learning regression methods. Each of them can be modeled as a parameterized function G_l:

$$y_l = G_l(\bar{f}, \bar{q}, \bar{p}), \tag{1}$$

where y_l is the forecasted value in one specific place with a lead time of l, $\bar{f}$ is the feature vector, $\bar{p}$ is a vector formed by the parameters of the model whose values will be obtained during the training (fitting) process, and $\bar{q}$ is the vector formed by the hyper-parameters, whose values are not obtained by the training process, they must be defined by the user. In the following subsections we describe in detail each of these elements and the process to obtain them.

A. FEATURE VECTORS

Figure 2 helps us to illustrate the feature selection process. First of all, time is discretized with a period appropriate for each of the data sets (10s in case of the NREL that provides samples every second, and 30min for InfoRiego, the minimum value possible in that case).

For each station, we compute the clear sky index for instant j as:

$$x[j] = \frac{\overline{GHI}[j-1, j]}{\overline{GHI_{\text{cs}}}[j-1, j]} \tag{2}$$

where $\overline{GHI}[j-1, j]$ is the mean GHI value in the time interval $[j-1, j]$, and $\overline{GHI_{\text{cs}}}[j-1, j]$ is the mean GHI value expected in that interval in the absence of clouds, as provided by a so called clear sky model. The models we have considered are described in detail in section III-B.

We refer to (2) also as the normalized irradiance. This normalization strategy is a common practice that allows two things: to eliminate the seasonality effects on irradiance and to normalize its range (convenient for machine learning models to avoid biases on different features).

The feature vector used to forecast, at instant i, the mean GHI value in the interval $[i + l - 1, i + l]$ (where l is known as the lead time) with m lagged samples, is formed by concatenating the m normalized irradiance samples (2) of each station in the interval $[i - m + 1, i]$ (highlighted in red in figure 2), including also the normalized azimuth and elevation angles of the sun from each station at instant i:

$$\begin{aligned}\bar{f}_{i+l,m,l} = [&x_0[i], \ldots, x_0[i-m+1], az_0[i], ev_0[i],\\ &x_1[i], \ldots, x_1[i-m+1], az_1[i], ev_1[i],\\ &\ldots]\end{aligned} \tag{3}$$

where the $x_j[i]$ represents the i-th sample of the normalized irradiance (2) at station j (with $j = 0$ representing the forecasting target and the remaining being the neighbouring stations) and $az_j[i]$ and $ev_j[i]$ represent the corresponding azimuth and elevation angles of the sun from station j at instant i. We use

Each sample $x[j]$ represents the mean value of the clear sky index in the interval $[j-1, j]$, i.e. the mean value of the GHI in that interval divided by the mean value of the expected GHI, for a given clear sky model.

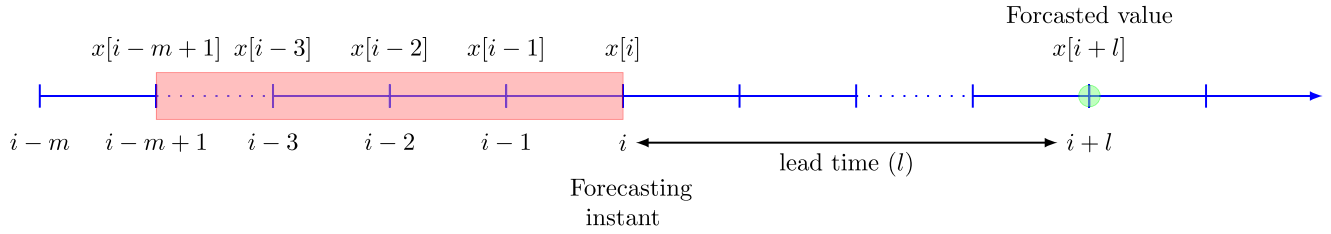


FIGURE 2. Samples selected to build the feature vector to forecast, at instant i, the mean irradiance in the interval $[i+l-1, i+l]$, where l is the lead time.

$m = 4$ in our experiments (sections V and VI). We should emphasize that each model is specifically trained for a given lead time value (l).

The training process fits the machine learning algorithms to give as output for this feature vector the normalized irradiance sample corresponding to the time instant $i+l$ ($x[i+l]$ in Figure 2).

B. CLEAR SKY MODEL

We have experimented with two models for the GHI in clear sky conditions: the McClear model proposed by [34] and the model proposed by [35]. The data for the former are generated by interpolating data obtained with a 15 min interval from the `http://www.soda-pro.com/` site, for the region of interest and the period of interest. The data for the latter is generated by the following equation:

$$GHI_{cs} = 1098 \cdot cos(z) \cdot exp\left(\frac{-0.057}{cos(z)} W/m^2\right), \quad (4)$$

where z is the zenith angle.

The McClear is an accepted reference model that incorporates the Linke turbidity factor to model the atmospheric absorption and scattering. The Haurwitz model on the other hand is a simple geometric one, that only takes into account the sun position, its coefficients were adjusted from irradiance measurements performed in the Blue Hill Observatory in Boston Massachusetts, between 1933 and 1943.

McClear however requires more computational power than Haurwitz. As we will show in section IV, the statistical models used for our work cannot exploit the accuracy differences between both models, allowing us to choose any of the two.

C. HYPER-PARAMETERS: CROSS VALIDATION

Each database is first split in two independent subsets: Training and Test. We randomly selected full days to be included in one of those sets, specifically 82% of the available days where selected for Training and the remaining 18% for Test.

The feature vectors built from the samples of the days included in the Test set are reserved to eventually evaluate and compare the accuracy of the models in the experiments shown in sections V and VI.

An exhaustive grid search is followed to obtain the best value for each of the hyper-parameters in a model ($\bar{q}$), using a traditional k-fold cross validation strategy. The Train set is split in k folds or partitions and k experiments are performed for each point in the grid. In each experiment the model is trained on $k-1$ folds and evaluated with the remaining fold. The final score assigned to the configuration of the corresponding grid point is the mean value of the scores obtained on the k experiments. The configuration of hyper-parameters that corresponds to the grid point with better score is the one selected for the model.

Once the hyper-parameters have been obtained ($\bar{q}$), we used them to train the model on the whole Training set to obtain the final values for the parameters in $\bar{p}$.

D. MACHINE LEARNING ALGORITHMS

In this work we have used four different models: ARX, ANN, RT and RRF, all of them implemented with the scikit-learn python library [36]. The simplest one is ARX, that assumes a linear relation between the forecasted irradiance and the features (as described in section III-A). In this case equation (1) can be expressed as:

$$\hat{y} = \boldsymbol{\beta}\boldsymbol{x}^T + \beta_0, \quad (5)$$

where $\hat{y}$ is the forecasted value, $\boldsymbol{x}$ is the corresponding feature vector, $\boldsymbol{\beta}$ is a row vector with the same number of coefficients as $\boldsymbol{x}$ and β_0 is the scalar bias value. A common approach to find the coefficients ($\boldsymbol{\beta}$ and β_0) is to use the ordinary least square method (as do for instance the authors in [15], [18]), which finds the values that minimize the mean square error:

$$(\boldsymbol{\beta}, \beta_0) \underset{\boldsymbol{\beta},\beta_0}{\arg\min} = \sum_{i=1}^{N} (y_i - \beta_0 - \boldsymbol{\beta}\boldsymbol{x_i}^T)^2, \quad (6)$$

where N is the number of coefficients vectors used in the training process and y_i is the measured GHI value that corresponds to the feature vector $\boldsymbol{x_i}$. A similar approach is LASSO [37], [38], which uses a regularization factor to limit the absolute value of the coefficients:

$$(\boldsymbol{\beta}, \beta_0) \underset{\boldsymbol{\beta},\beta_0}{\arg\min} = \sum_{i=1}^{N} (y_i - \beta_0 - \boldsymbol{\beta}\boldsymbol{x_i}^T)^2 + \lambda \sum_{j=1}^{M} |\beta_j|, \quad (7)$$

were λ is the regularization factor and M is the dimension of $\boldsymbol{x}$. Using this approach with small regularization factors leads to similar regression coefficients.

An ANN with Multilayer Perceptron (MLP) arquitecture is composed of several layers of neurons. The output of a neuron i on layer $j+1$ depends on the outputs of the M_j neurons of the previous layer (y_k^j) as:

$$y_i^{j+1} = f\left(\alpha_{i,0} + \sum_{k=1}^{M_j} \alpha_{i,k} y_k^j\right), \tag{8}$$

where the y_k^0 are the inputs to the network (the features). The α coefficients are obtained in a training process using the back-propagation algorithm, which has a regularization parameter that has to be specified a priori. The function f is known as the activation function and simulates the response of a human neuron, we used the most common sigmoid function. Finally, the architecture of the ANN, i.e. the number of layers and the number of neurons on each layer, can be considered as a set of additional parameters that must be explored to obtain the best results. After some experimentation we opted for an ANN with three hidden layers and 300 neurons per layer, plus an output layer with only one neuron. In the output layer we also opted for a linear activation function, which is the common approach for regression ANNs.

RTs build piece-wise constant functions for G in equation (1). The intervals in which G is constant are selected during the training process, that builds a binary tree in which each node represents a split of the space and the two sub-trees represent further sub divisions of each subspace. On the leaves of this tree the value of the function is set constant, usually selected to minimize some error criteria (usually mean square error) for all the training samples seen on that subspace.

However, RTs are known to be very unstable due to their tendency to over fit. RRF is an ensemble method based on RT that alleviates this problem. During the training process a large number of RT are build from randomly selected re-samples with replacement of the training set. The function G evaluates a given point as the average of the functions corresponding to these RTs.

E. METRICS

A common approach to quantify the accuracy of the forecasting models is to use the Root Mean Square Error (RMSE), which can be expressed as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{Y[i]} - Y[i])^2} \tag{9}$$

where $Y[i]$ is the real GHI value measured at instant i and $\hat{Y[i]}$ is the corresponding predicted GHI value. However using absolute values of RMSE makes it difficult to compare results from different locations. Therefore, a common practice is to use a normalized version of the RMSE, usually called nRMSE.

There is however no established consensus on the expression that should be used for the nRMSE. For instance, the authors in [15] used the following expression:

$$nRMSE = \frac{RMSE}{max(Y)} \tag{10}$$

where $max(Y)$ is the maximum GHI value observed. A variant of this is to divide by the maximum difference observed in the GHI (replace $max(Y)$ by $max(Y) - min(Y)$) which in the case of GHI is generally the same if large day periods are considered, as $min(Y)$ is 0 at the sunrise and nightfall.

On the other hand the authors of [18] used a different normalization strategy, dividing the RMSE instead by the mean value of the observed GHI:

$$nRMSE = \frac{RMSE}{\overline{Y}} \tag{11}$$

This approach leads to slightly larger nRMSE values, but in essence has the same properties than the expression (10). In both cases every error is divided by a large constant. A relative error (say 10%) has less influence in moments when the absolute GHI is low (early in the morning) than when the GHI is larger (around noon). This might influence the decisions when choosing the hyper-parameters if the nRMSE is used to score each configuration.

A different approach is followed in [20], where the authors compute the nRMSE from relative differences:

$$nRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{Y[i]} - Y_i}{Y[i]}\right)^2} \tag{12}$$

This is an interesting approach, using relative values does not penalize the moments of the day where the GHI is naturally lower respect to the moments of the day when that value is naturally higher. These nRMSE values are however larger than the nRMSE values obtained from (10) or (11).

We followed a more natural approach. Given that our models like most others in the literature predict the clear sky index (and then indirectly obtain the forecasted GHI value by multiplying the output by the GHI expected by the clear sky model), we score our models by computing the RMSE on clear sky index values:

$$\begin{aligned} nRMSE &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{x}[i] - x[i])^2} \\ &= \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{Y}[i] - Y[i]}{Y_{cs}[i]}\right)^2} \end{aligned} \tag{13}$$

where $Y_{cs}[i]$ is the GHI expected by the clear sky model for sample i, and $\hat{x}[i]$ and $x[i]$ are the corresponding forecast and measured clear sky indexes (as used in Figure 2). Notice that this simple approach is close to (12), it also weights the errors taking into account the moment of the day (from a model instead of from the direct measurements). The values of (13) fall in between the values of (11) and (12). Given that the goal of this work is not to obtain the best forecasting model, but

to analyze the influence of the spatial resolution; being in the exact same nRMSE scale as other authors is not critical.

As common practice, we additionally used a skill figure to evaluate each model. We used a persistence model as reference to derive the skill, which simply predicts that the clear sky index will remain constant for the lead time period:

$$x[i+l] = x[i]. \tag{14}$$

where again $x[i]$ and $x[i+l]$ are normalized GHI values (2). The skill figure is computed as the relative percentage of improvement in nRMSE respect to the normalized persistent model (14):

$$S = 100\left(1 - \frac{nRMSE_{\text{model}}}{nRMSE_{\text{persistence}}}\right). \tag{15}$$

IV. IMPACT OF THE CLEAR SKY MODEL

The purpose of the Clear Sky Model (CSM) is two-fold: eliminate seasonality effects and normalize. The latter is the most relevant role for the statistical models, specially when several magnitudes with dissimilar ranges are considered as features. To assess the impact of the CSM on the output of the statistical models we conducted a simple experiment. Table 1 shows the RMSE values obtained when training a Neural Network to predict the GHI for the DHHL_6 node of the NREL dataset, for different lead times, using the two CSMs presented in III-B. For this experiment we computed the RMSE using the raw GHI values, i.e. not normalized by the CSM.

TABLE 1. Impact of the clear sky model in the accuracy of the neural network, for the DHHL_6 station in the NREL network. The sampling period was 1s.

Lead Time	$\text{RMSE}_{\text{McClear}}$ (Wh/m^2)	$\text{RMSE}_{\text{Haurwitz}}$ (Wh/m^2)
10s	74.613	74.613
30s	136.812	136.812
1min	177.366	177.367
2min	214.255	214.258
5min	246.889	246.902
10min	266.555	266.574

As we can see from Table 1, the RMSE obtained when using the McClear model is always slightly better than the one obtained with the Haurwitz one. However these differences are negligible, affecting only the fifth significant number of RMSE, and even less in some cases. We can conclude that both CSMs can be used for the purpose of this paper, and we can take advantage of the computational simplicity of the Haurwitz model with negligible impact on the forecasting results.

V. SPATIAL RESOLUTION AND LEAD TIME

The first hypothesis of this work is that statistical models can take advantage of data supplied by spatially distributed sensors, from very short lead times up to hours. To confirm this hypothesis we conducted similar experiments on the two databases described in section II. We trained the four statistical models mentioned in section III-D for several prediction targets and several lead times. As a reference we also trained the equivalent local methods, in which we only consider the data from the prediction target sensor (no spatial information is used). Table 2 shows, for each method and different lead times the nRMSE relative to the one obtained for the corresponding local method (nRMSE_{l}), as well as the Skill value for the method and the corresponding local equivalent (Skill_{l}). Due to space constraints, we show only the data for one target station on each dataset, the VA01 station in the case of InfoRiego and the DHHL_6 for NREL. The former was selected to have a similar amount of nearby sensors in all directions, so that we can evaluate the influence of the information provided by them. For the latter, we selected one peripheral station from the opposite side of the incoming dominant winds, to maximize the impact of the information provided by the surrounding stations. Moreover, we did consider different lead times for the two databases. In the case of InfoRiego, we have samples only every half an hour, making shorter lead times unfeasible. On the other hand, the sensors in the NREL network cover a small terrain area, providing insufficient information for the prediction of large lead times, given the fast winds that dominate the covered area.

First of all, Table 2 shows that all methods considered do take advantage of the spatial information supplied for most lead times, as their nRMSE is smaller than the corresponding nRMSE_{l}, although the impact of the information provided by the surrounding stations differs significantly for the two datasets. Moreover, we can see that the statistical methods outperform the persistent model, with the exception of RT that in some cases has a negative skill. We can also see that the $\text{nRMSE}/\text{nRMSE}_{\text{l}}$ ratio for the InfoRiego dataset is generally closer to 1, meaning that the information provided by the surrounding nodes is not as useful for the statistical methods as it is in the case of the NREL dataset. This can be partially explained by the different among their geographic locations and weather characteristics. The InforRiego network covers a large area, with an average height of 750m above sea level, where winds are generally not very strong, the clouds evolve slowly, and many days have an almost clear sky [33]. In this area, the persistence model works reasonably well for lead times around 0.5h. Even so, the statistical methods reduce the forecasting error around a 10% using only local information (Skill_{l}), i.e. having measurements from the past helps to better estimate the slope of the irradiance, and probably better results would be possible by lowering the interval between samples (fixed at 1/2h for this dataset). Moreover, the skill increases up to 16.9% by using spatially distributed data (Skill), which means that some clouds can travel from the surrounding sensors in 1/2h, and the information from those sensors is exploited by the statistical methods providing around a 7% of skill improvement.

As the lead time increases the persistence model becomes less effective and the statistical methods increase the skill up to a 70% for a lead time of 4h. Although the spatial

TABLE 2. Improvement obtained from spatially distributed sensors, for the VA01 station of the InfoRiego network and the DHHL_6 station of the NREL network.

	InfoRiego (VA01)				NREL (DHH_6)			
Method	Lead Time	$\frac{\text{nRMSE}}{\text{nRMSE}_l}$	Skill_l	Skill	Lead Time	$\frac{\text{nRMSE}}{\text{nRMSE}_l}$	Skill_l	Skill
ARX	0.5 h	0.9301	10.7	16.9	10 s	0.7118	3.6	31.4
NN		0.9715	10.9	13.4		0.7229	9.4	35.0
RT		0.9181	-23.6	-13.5		0.7089	-28.3	9.1
RRF		0.9158	8.1	15.8		0.6740	2.6	34.3
ARX	1.0 h	0.9069	19.4	26.9	30 s	0.6229	4.7	40.6
NN		0.9274	18.0	24.0		0.6096	9.9	45.4
RT		0.9359	-6.1	0.7		0.6419	-27.	18.5
RRF		0.8967	18.3	26.7		0.6188	4.1	40.6
ARX	1.5 h	0.9036	28.9	35.8	1 min	0.7164	7.6	33.8
NN		0.9777	29.4	31.0		0.7316	12.7	36.5
RT		0.9730	6.4	8.9		0.7678	-23.1	5.5
RRF		0.8984	28.4	35.7		0.7324	6.6	31.6
ARX	2.0 h	0.9065	37.9	43.7	2 min	0.8008	10.9	28.7
NN		0.9390	37.2	41.0		0.8230	15.1	30.5
RT		0.9349	16.1	21.6		0.8387	-19.6	-0.3
RRF		0.9215	38.4	43.3		0.8107	9.1	26.3
ARX	3.0 h	0.9466	53.9	56.3	3 min	0.8690	12.7	24.1
NN		1.0188	54.6	53.7		0.8883	16.0	25.7
RT		1.0043	43.2	43.0		0.9006	-18.1	-6.4
RRF		0.9395	55.1	57.8		0.8737	10.3	21.6
ARX	4.0 h	0.9453	65.9	67.8	5 min	0.9346	14.2	19.8
NN		0.9519	66.6	68.2		0.9544	17.4	21.8
RT		0.9882	58.9	59.4		0.9578	-16.0	-11.1
RRF		0.9332	68.0	70.1		0.9293	11.7	17.9

information gains slightly more importance for lead times up to 2h, for larger intervals it becomes largely irrelevant, as shown by the nRMSE/nRMSE_l values approaching 1, with differences between Skill and Skill_l around 2% (being even negative in some cases). These results indicate that the area covered by the stations starts to become small for lead times above 2h, or that the relation between the local GHI in the future and the data measured further away may be too complex. This could be the case if clouds form, dissipate or change their shape during the lead time. In the next section we give some suggestions for possible improvements for such complex cases.

As we have seen, the information of the surrounding sensors, in the case of the InfoRiego data set, helps to improve the skill marginally, from 2% to 8%. On the other extreme we have the NREL network. It covers a much smaller area, in the Oahu Island (Hawaii), roughly at sea level. This area has a cloudy and windy tropical climate, where the sky is generally covered with fast moving clouds. In this scenario, the persistent model has a hard time for all the lead times considered, except for the shortest (10s). We have samples every second for each sensor in the area and, as can be seen from the nRMSE/nRMSE_l, Skill and Skill_l columns, the spatial information is in this case much more relevant, reaching its maximum for a lead time of 30s, when the Skill is roughly 10 times larger than the Skill_l. The benefits from this information are quickly reduced for lead times above 1min. This result is in consonance with the dominant northern east winds that allow a cloud to cross the covered area in about 3 minutes, which was already observed by [15]. Again here it would be necessary to extend the area covered by the network to take advantage of the spatial information for larger lead times. As can be seen, for a lead time of 5 minutes the spatial information only provides for around a 5% of improvement.

On the other hand, the trend of the Skill is very different from the trend observed in the case of the InfoRiego dataset. As the lead time increases both the persistent and the statistical models reduce their accuracy, i.e. the corresponding nRMSE increases, although the deterioration is worse for the persistent model, which translates to an increase of Skill_l. The Skill, that includes the information from the neighbouring sensors, starts to decrease strongly for lead times above 2 min, which suggests that when the area covered becomes too small (a cloud can cross the area in 3 min), the information provided by the sensors becomes less relevant and the ratio nRMSE/nRMSE_l approaches 1.

Finally, we can compare the performance of all the methods considered in this study. The RT model consistently provides the worst Skill, being even negative for some lead times, which is probably the consequence of its tendency to over fit. The rest of the methods obtain similar Skills although NN tend to be better for the short lead times used for the

NREL dataset, whereas ARX tend to have a better behaviour with the data from InfoRiego, except for the highest lead times considered, where the RRF provides a slightly better accuracy.

VI. INFLUENCE OF THE FEATURES

In this section we delve into the study of the influence of each feature, to get a better understanding of the problem. We use the ARX and the RRF methods, which make this analysis easier to interpret. The importance of each feature in the ARX model is determined by the absolute value of its coefficient in the linear regression model described by equation (5), provided that the features are normalized (recall that all the features considered by our models, are normalized as explained in Section III-A). Notice that if some features are inter-correlated they will all appear equally important, which will not help to reduce the dimensionality. On the other hand, boosting ensemble trees, like RRF, calculate the space splits iteratively, trying to find the best splits for the data. This process gives us a ranking of the features in importance, much more selective than the one obtained from the linear coefficients, that could be used for dimensionality reduction. We believe that using both methods will give us more insights on the problem we try to solve.

We start by analyzing the results on the NREL dataset, for which the models are performing better. Figure 3a represents the weight of each feature on the position of the sensor that provides the corresponding value, for the ARX model trained for the DHHL_6 station and different lead times. As can be seen, for the shortest lead time (10s) the most relevant features are the first radiation sample from the local station and the stations immediately to the north east, the direction of the dominant winds in the area. The importance of the samples on the DHHL_6 station itself indicates that the irradiance will not change significantly, and the recent past in the sensor is enough to get a good estimate for the future.

However as we increase the lead time to 30s, the local data from the DHHL_6 station looses importance and the features provided by the sensors in the north east direction become the most important ones.

Slightly worse results were obtained for lead times of 1 min. This can be explained by the presence of a building with no irradiance sensors in the north east direction.

Finally, for the largest lead time considered, we see that the sensor providing the most relevant data is located at the north west border of the sensor network. It clearly shows that we would need data from a larger area to the north east to maintain the prediction accuracy. These results are in harmony with the results obtained by [15], which showed that the stations aligned with the dominant wind directions in the target area where the most relevant for the ARX model they were using, and that larger prediction intervals would require the use of stations located further away, outside the area covered by the NREL network.

Figure 3b gives us the same representation for the the feature importance of the RRF model as obtained from scikit-learn for the same data as before and the same lead times. As can be seen the trend is very similar although the importance is heavily concentrated in fewer sensors.

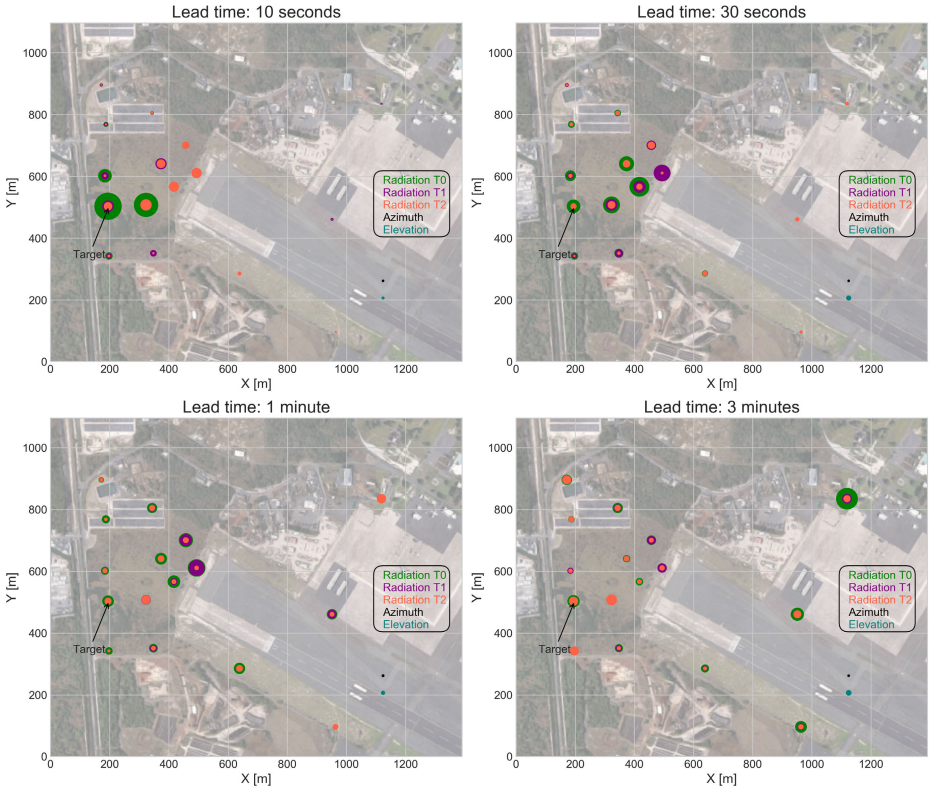
A similar analysis can be conducted with the data provided by the InfoRiego network, although in this case the spatial data has shown to be less relevant than for the models trained with the NREL dataset. Figure 4a represents the weight of each feature on the position of the sensor that provides the corresponding value, for the ARX model trained for the VA01 station and different lead times, where T1, T2 and T3 represent the samples $x[i]$, $x[i-1]$ and $x[i-2]$ from Figure 2.

Again the most important feature for the shortest lead time is the most recent irradiance measurement at the target station. This might be the consequence of the relatively slow winds in that area and the amount of clear sky days available during the year. This could also explain why the local methods are close in accuracy to the methods that include the information from other stations. Just after the local irradiance, the stations providing the most important data are the closest ones to the west, more precisely, the most recent measurements of irradiance at those points. Older irradiance measurements are relevant from further away locations (clouds need more time to travel to the target location), as are some derivative terms. As we increment the lead time the samples from further away locations gain importance in detriment of the closer stations, and the main direction is not so clearly marked (west-east direction seems slightly more important). Again the statistical methods are able to track the movement of some clouds that coming from further away locations have now time to travel to the target station.

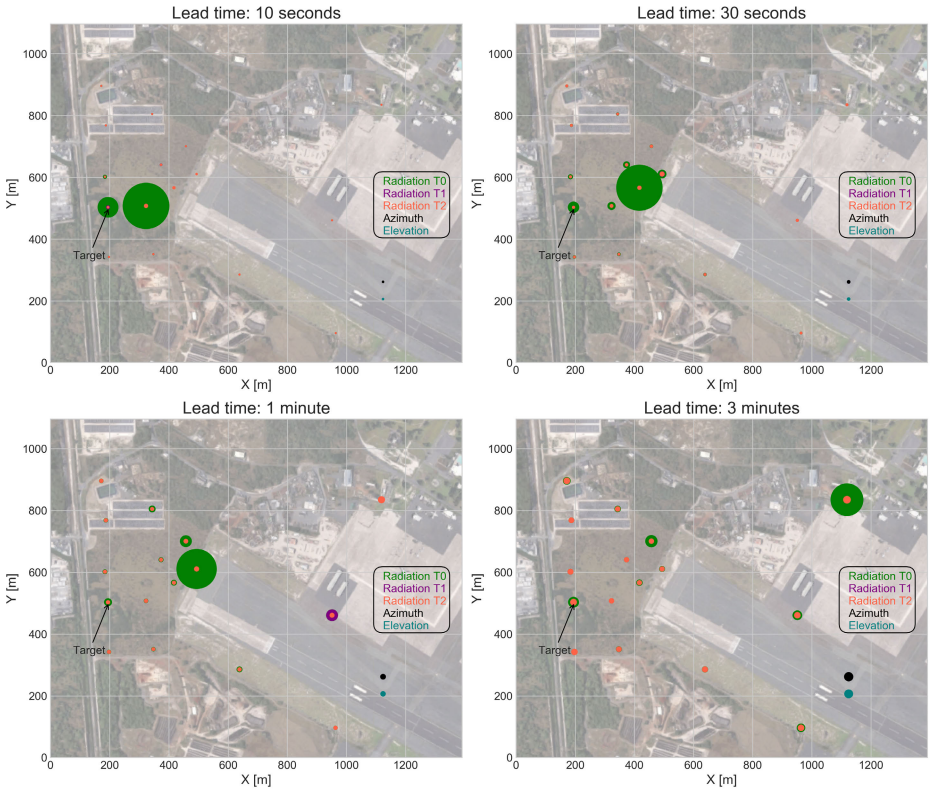
For short lead times, the direction of the relevant stations correlates well with the direction of the local dominant winds, as can be seen from the wind rose shown in Figure 5. As we increase the lead time, more distant stations become also important and the wind pattern in all the covered area affects the clouds movement, not only the local wind, making the cloud tracking problem even more complex. Having wind measurements on all stations, or incorporating expected wind fields from numerical meteorological models, might simplify the tracking process and help to improve the accuracy of the prediction.

Moreover, the azimuth feature is important for the linear model, and gains in importance as we increase the lead time. Notice that the azimuth is just a way to encode the solar time (day moment). It is very similar for all stations and the fact that it appears relevant on more than one point is the result of not removing correlated variables. This is an indication that some local seasonal effects are being captured by the model by including the azimuth as feature. These might be related to wind, in which case including wind data might turn azimuth useless, but might also be related to other meteorological phenomena, like mist or dust, that are not captured by the clear sky model.

Regarding the RRF model, Figure 4b represents the feature importance as obtained from scikit-learn, when trained

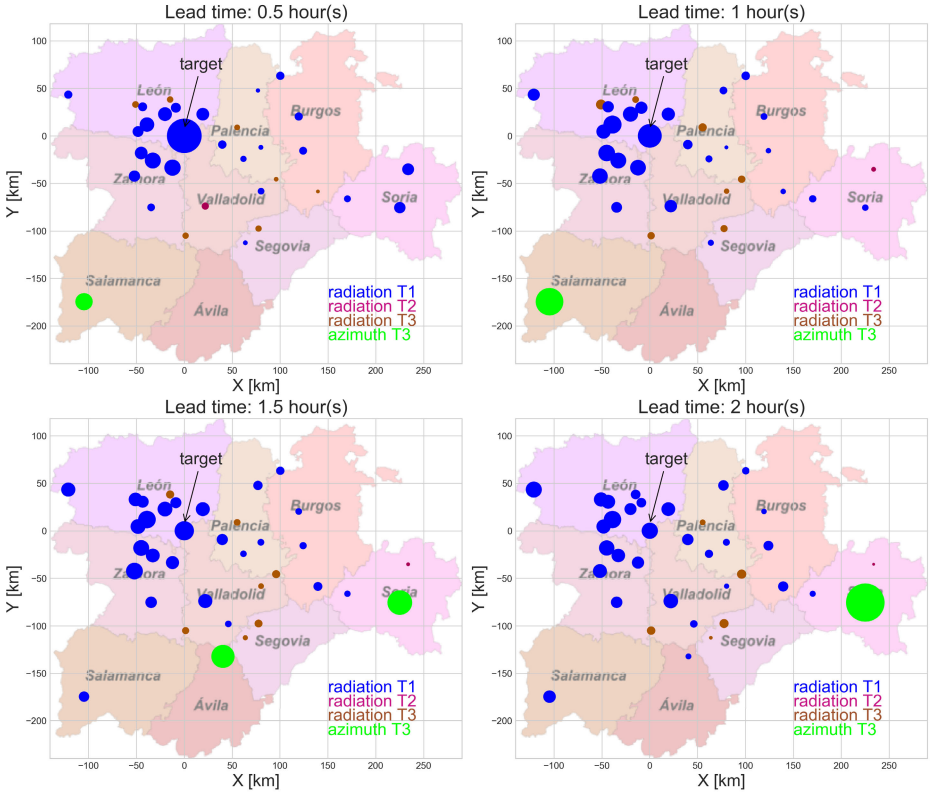


(a) Linear

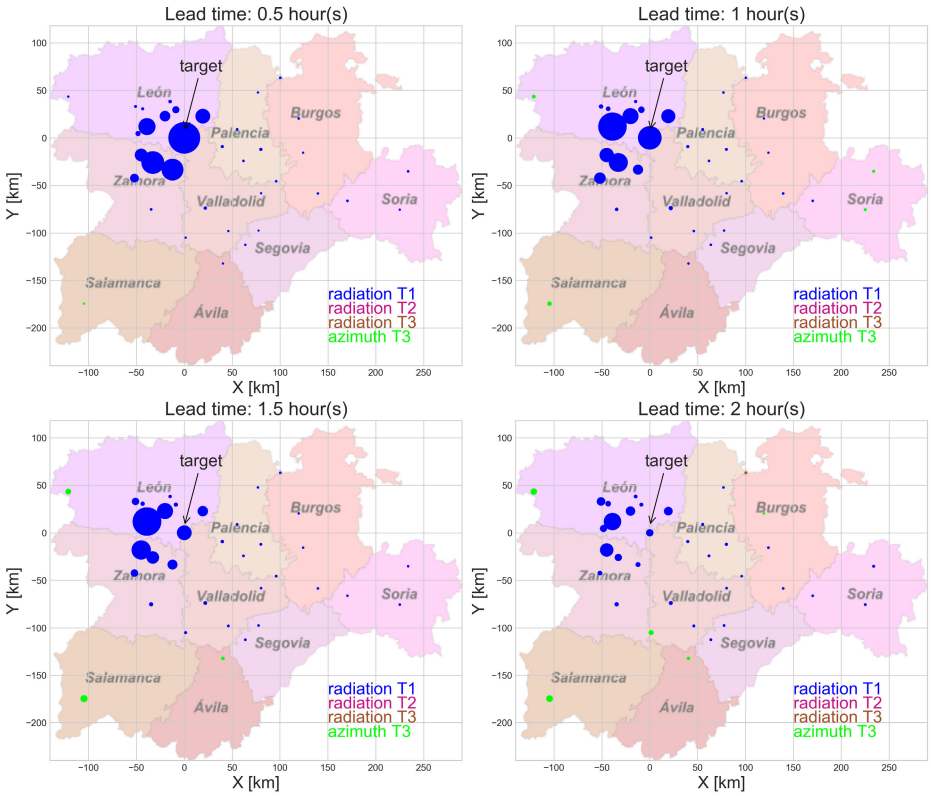


(b) RRF

FIGURE 3. Linear and RRF feature importance on the NREL database for different lead times.



(a) Linear



(b) RRF

FIGURE 4. Linear and RRF feature importance on the InfoRiego database for different lead times.

TABLE 3. Dimensionality reduction when using only a small percentage of the most important features, as ranked by RRF, on the data from InfoRiego. $nRMSE_{x\%}$ stays for the nRMSE obtained when using only the x% of the most important features.

Method	Lead Time	$nRMSE_{1.25\%}$	$nRMSE_{2.5\%}$	$nRMSE_{5\%}$	$nRMSE_{10\%}$	$nRMSE_{100\%}$
ARX	0.5 h	0.1099	0.1042	0.1028	0.1035	0.1035
NN		0.1208	0.107342	0.1067	0.1061	0.1074
RT		0.1462	0.140242	0.1446	0.1426	0.1426
RRF		0.1149	0.106542	0.1062	0.1048	0.1045
ARX	1.0 h	0.1248	0.120642	0.119	0.1192	0.1196
NN		0.1265	0.120442	0.1191	0.1204	0.1253
RT		0.1922	0.188842	0.1612	0.1641	0.1664
RRF		0.1281	0.124342	0.1209	0.1205	0.1205
ARX	1.5 h	0.1342	0.131742	0.1301	0.1285	0.1269
NN		0.1357	0.133642	0.129	0.1272	0.1299
RT		0.2248	0.171942	0.1802	0.1765	0.1796
RRF		0.1375	0.133542	0.1298	0.1288	0.1272
ARX	2.0 h	0.14	0.136942	0.1344	0.1302	0.1293
RT		0.1389	0.135642	0.136	0.1284	0.1331
NN		0.1862	0.186942	0.1842	0.1826	0.1814
RRF		0.1427	0.137342	0.135	0.132	0.1316
ARX	3.0 h	0.15	0.143442	0.141	0.1397	0.1387
RT		0.1483	0.140442	0.1419	0.1372	0.1472
NN		0.1964	0.205842	0.1931	0.1849	0.1793
RRF		0.1519	0.141942	0.1391	0.1375	0.1346
ARX	4.0 h	0.1463	0.140642	0.1407	0.1389	0.1382
RT		0.14	0.134642	0.1365	0.1349	0.1405
NN		0.2045	0.187342	0.1707	0.1746	0.18
RRF		0.1416	0.134142	0.1321	0.1298	0.1274

TABLE 4. Dimensionality reduction when using only a small percentage of the most important features, as ranked by RRF, on the data from NREL. $nRMSE_{x\%}$ stays for the nRMSE obtained when using only the x% of the most important features.

Method	Lead Time	$nRMSE_{2\%}$	$nRMSE_{4\%}$	$nRMSE_{6\%}$	$nRMSE_{10\%}$	$nRMSE_{20\%}$	$nRMSE_{100\%}$
ARX	10 s	0.1081	0.0765	0.0732	0.0712	0.0703	0.0694
NN		0.1088	0.0762	0.0685	0.0640	0.0627	0.0636
RT		0.2585	0.3314	0.2751	0.2728	0.2360	0.0890
RRF		0.2580	0.2605	0.2100	0.2042	0.1753	0.0642
ARX	30 s	0.1370	0.1267	0.1237	0.1204	0.1131	0.1104
NN		0.1347	0.1235	0.1206	0.1159	0.1008	0.0980
RT		0.2696	0.2882	0.2668	0.2483	0.2286	0.1463
RRF		0.2690	0.2253	0.2018	0.1836	0.1681	0.1065
ARX	1 min	0.1836	0.1770	0.1693	0.1666	0.1641	0.1589
NN		0.1790	0.1720	0.1656	0.1650	0.1587	0.1473
RT		0.2483	0.3125	0.2986	0.2768	0.2678	0.2191
RRF		0.2477	0.2436	0.2237	0.2030	0.1957	0.1587
ARX	2 min	0.2338	0.2142	0.2106	0.2103	0.2069	0.2049
NN		0.2258	0.2062	0.2039	0.1994	0.1957	0.1925
RT		0.2584	0.3377	0.3352	0.3276	0.3221	0.2778
RRF		0.2577	0.2631	0.2520	0.2422	0.2357	0.2042
ARX	3 min	0.2539	0.2424	0.2395	0.2385	0.2364	0.2350
NN		0.2444	0.2311	0.2277	0.2277	0.2235	0.2221
RT		0.2752	0.3616	0.3586	0.3526	0.3473	0.3178
RRF		0.2747	0.2827	0.2708	0.2626	0.2556	0.2340
ARX	5 min	0.2819	0.2697	0.2677	0.2678	0.2647	0.2636
NN		0.2687	0.2555	0.2537	0.2529	0.2486	0.2500
RT		0.2887	0.3823	0.3790	0.3774	0.3712	0.3551
RRF		0.2879	0.2992	0.2872	0.2798	0.2739	0.2623

with the same data as before and the same lead times. Again, the distance from the target sensor to the sensors providing the most important features increases with the lead time. As before, the larger it is the further away we have to sense for clouds that can move to the target area, and the statistical models seem to capture these correlations.

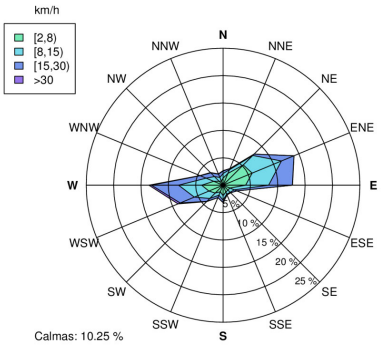


FIGURE 5. **Wind Rose from measurements between 2008 and 2011 from the station of Medina de Rioseco, a few km to the south east from the VA01 station. The 10.25% of the days were calm. Figure from the Instituto Tecnológico Agrario de Castilla y León.**

Moreover, the azimuth plays a similar role as for the ARX model, gaining relevance as the lead time increases.

A final interesting experiment is to use the dominant features provided by the RRF model to reduce the dimensionality of our problem (a similar approach as using PCA), which has interesting advantages from the computational point of view. Tables 3 and 4 show the nRMSE obtained for different lead times, training our models using only a small percentage of the most significant features as ranked by RRF. For the InfoRiego dataset the nRMSE is significantly reduced when we add features up to the 5% most significant ones, when the nRMSE remains more or less stable. This means that we could reduce the complexity of the problem by using a feature vector with only the 5% of the original features and still obtain very similar results. The data from NREL shows a similar trend, with the exception of RRF which still have a significant increment in nRMSE when using 20% of its most significant features. For NN and ARX, we could reduce the features to the 10% most important ones and obtain almost the same results.

VII. CONCLUSION

This works analyzes the accuracy improvements obtained by using spatially distributed irradiance sensors for some machine learning algorithms designed for short term GHI forecasting. We conducted similar experiments on two different datasets: NREL, that provides very dense spatio-temporal data covering a small area at the Oahu island in Hawaii, and InfoRiego, a less dense network covering a much larger area in the region of Castilla y León in Spain.

We tested four different machine learning algorithms: ARX, NN, RRF and RT. All four showed in general some improvements after using the spatially distributed data. RT provided always the worst accuracy whereas ARX showed the best results on the InfoRiego dataset, which indicates that the problem for larger lead times and less dense spatio-temporal input data is better modeled with a linear method. On the other hand, NN provided the best results for the shorter lead times and dense spatio-temporal input data from NREL, which highlights the non-linearity nature of that problem.

Moreover, the inclusion of spatially distributed inputs was more effective for the NREL dataset and lead times in which a cloud can be moved by the local wind from the neighboring sensors to the target sensor and all the area traversed by the clouds contains sensors. The spatially distributed information was less effectively exploited when sensors where missing on the NREL dataset or when using the less dense InfoRiego network and larger lead times. The feature importance analysis conducted showed that in both cases the sensors providing the most relevant data are located in the direction of the dominant winds in the area. This relation was stronger for the dense NREL network, for which we could predict with shorter lead times.

All this together suggests that including the estimated wind fields from numerical weather forecasting models could help to improve the forecasting accuracy. The inclusion of the estimated cloud maps might also help in the case of less dense networks that cover larger areas.

Finally, our analysis shows that the notion of feature importance from RRF can be used to effectively reduce the dimensionality of the problem even for the other methods, which show similar accuracy with only 5%-20% of the features. This can be very important for short lead times and dense sensor networks.

ACKNOWLEDGMENT

The authors would also like to thank the NREL and InfoRiego sites for providing us the data needed for this study. Code for all experiments can be found in https://github.com/tenllado/solarcasting

REFERENCES

[1] W. Richardson, D. Cañadillas, A. Moncada, R. Guerrero-Lemus, L. Shephard, R. Vega-Avila, and H. Krishnaswami, "Validation of all-sky imager technology and solar irradiance forecasting at three locations: NREL, San Antonio, Texas, and the Canary Islands, Spain," *Appl. Sci.*, vol. 9, no. 4, p. 684, Feb. 2019.

[2] B. Nouri, S. Wilbert, P. Kuhn, N. Hanrieder, M. Schroedter-Homscheidt, A. Kazantzidis, L. Zarzalejo, P. Blanc, S. Kumar, N. Goswami, R. Shankar, R. Affolter, and R. Pitz-Paal, "Real-time uncertainty specification of all sky imager derived irradiance nowcasts," *Remote Sens.*, vol. 11, no. 9, p. 1059, 2019.

[3] H. Yang, B. Kurtz, D. Nguyen, B. Urquhart, C. W. Chow, M. Ghonima, and J. Kleissl, "Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego," *Sol. Energy*, vol. 103, pp. 502–524, May 2014.

[4] C. W. Chow, S. Belongie, and J. Kleissl, "Cloud motion and stability estimation for intra-hour solar forecasting," *Sol. Energy*, vol. 115, pp. 645–655, May 2015.

[5] M. Lipperheide, J. L. Bosch, and J. Kleissl, "Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant," *Sol. Energy*, vol. 112, pp. 232–238, Feb. 2015.

[6] V. P. A. Lonij, A. E. Brooks, A. D. Cronin, M. Leuthold, and K. Koch, "Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors," *Sol. Energy*, vol. 97, pp. 58–66, Nov. 2013.

[7] A. T. Lorenzo, W. F. Holmgren, and A. D. Cronin, "Irradiance forecasts based on an irradiance monitoring network, cloud motion, and spatial averaging," *Sol. Energy*, vol. 122, pp. 1158–1169, Dec. 2015.

[8] Z. Wang, J. Xu, H. He, H. Zhang, and L. Zhang, "An experiment to derive cloud motion vectors from satellite images with 2-D Fourier phase analysis technique," *Plateau Meteorol.*, vol. 25, no. 1, pp. 105–109, 2006.
[9] A. Inanlouganji, T. A. Reddy, and S. Katipamula, "Evaluation of regression and neural network models for solar forecasting over different short-term horizons," *Sci. Technol. Built Environ.*, vol. 24, no. 9, pp. 1004–1013, Oct. 2018.
[10] P. Kuhn, D. Garsche, S. Wilbert, B. Nouri, N. Hanrieder, C. Prahl, L. Zarzarlejo, J. Fernández, A. Kazantzidis, T. Schmidt, D. Heinemann, P. Blanc, and R. Pitz-Paal, "Shadow-camera based solar nowcasting system for shortest-term forecasts," *Meteorologische Zeitschrift*, vol. 28, no. 3, pp. 255–270, Oct. 2019.
[11] X. G. Agoua, R. Girard and G. Kariniotakis, "Short-term spatio-temporal forecasting of photovoltaic power production," *IEEE Trans. Sustain. Energy*, vol. 9, no. 2, pp. 538–546, Apr. 2018. doi: 10.1109/TSTE.2017.2747765.
[12] E. Basha, R. Jurdak, and D. Rus, "In-network distributed solar current prediction," *ACM Trans. Sensor Netw.*, vol. 11, no. 2, pp. 1–28, Dec. 2014.
[13] J. Boland, "Spatial-temporal forecasting of solar radiation," *Renew. Energy*, vol. 75, pp. 607–616, Mar. 2015.
[14] C. Yang, A. A. Thatte, and L. Xie, "Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation," *IEEE Trans. Sustain. Energy*, vol. 6, no. 1, pp. 104–112, Jan. 2015.
[15] R. Amaro e Silva and M. C. Brito, "Impact of network layout and time resolution on spatio-temporal solar forecasting," *Sol. Energy*, vol. 163, pp. 329–337, Mar. 2018.
[16] R. J. Bessa, A. Trindade, C. S. P. Silva, and V. Miranda, "Probabilistic solar power forecasting in smart grids using distributed information," *Int. J. Electr. Power Energy Syst.*, vol. 72, pp. 16–23, Nov. 2015.
[17] D. Yang, Z. Dong, T. Reindl, P. Jirutitijaroen, and W. M. Walsh, "Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage," *Sol. Energy*, vol. 103, pp. 550–562, May 2014.
[18] C. Huang, L. Wang, and L. L. Lai, "Data-driven short-term solar irradiance forecasting based on information of neighboring sites," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9918–9927, Dec. 2019.
[19] A. M. Nobre, C. A. Severiano, S. Karthik, M. Kubis, L. Zhao, F. R. Martins, E. B. Pereira, R. Rüther, and T. Reindl, "PV power conversion and short-term forecasting in a tropical, densely-built environment in singapore," *Renew. Energy*, vol. 94, pp. 496–509, Aug. 2016.
[20] F.-V. Gutierrez-Corea, M.-A. Manso-Callejo, M.-P. Moreno-Regidor, and M.-T. Manrique-Sancho, "Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations," *Sol. Energy*, vol. 134, pp. 119–131, Sep. 2016.
[21] M. Lazzaroni, S. Ferrari, V. Piuri, A. Salman, L. Cristaldi, and M. Faifer, "Models for solar radiation prediction based on different measurement sites," *Measurement*, vol. 63, pp. 346–363, Mar. 2015.
[22] A. G. R. Vaz, B. Elsinga, W. G. J. H. M. van Sark, and M. C. Brito, "An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in Utrecht, The Netherlands," *Renew. Energy*, vol. 85, pp. 631–641, Jan. 2016.
[23] Y. Yu, J. Cao, and J. Zhu, "An LSTM short-term solar irradiance forecasting under complicated weather conditions," *IEEE Access*, vol. 7, pp. 145651–145666, 2019.
[24] D. Yang, C. Gu, Z. Dong, P. Jirutitijaroen, N. Chen, and W. M. Walsh, "Solar irradiance forecasting using spatial-temporal covariance structures and time-forward Kriging," *Renew. Energy*, vol. 60, pp. 235–245, Dec. 2013.
[25] A. W. Aryaputera, D. Yang, L. Zhao, and W. M. Walsh, "Very short-term irradiance forecasting at unobserved locations using spatio-temporal Kriging," *Sol. Energy*, vol. 122, pp. 1266–1278, Dec. 2015.
[26] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renew. Energy*, vol. 105, pp. 569–582, May 2017.
[27] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 65–76, Nov. 2013.
[28] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," *Progr. Energy Combustion Sci.*, vol. 39, no. 6, pp. 535–576, Dec. 2013.
[29] R. V. Rao, "Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems," *Int. J. Ind. Eng. Comput.*, vol. 7, no. 1, pp. 19–34, 2016.
[30] M. Sengupta and A. Andreas, "Oahu solar measurement grid (1-year archive): 1-second solar irradiance; Oahu, Hawaii (data)," NREL, Golden, CO, USA, Tech. Rep. DA-5500-56506, 2010.
[31] *Instituto Tecnológico Agrario de Castilla y León (ITACyL)*, Instituto Tecnológico Agrario de Castilla y León Consejería de Agricultura y Ganadería Ctra, Valladolid, Spain, 2015. [Online]. Available: http://www.inforiego.org
[32] L. M. Hinkelman, "Differences between along-wind and cross-wind solar irradiance variability on small spatial scales," *Sol. Energy*, vol. 88, pp. 192–203, Feb. 2013.
[33] *Atlas Agroclimático de Castilla y León*. Instituto Tecnológico Agrario de Castilla y León, Valladolid, Spain, Sep. 2013.
[34] M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola, A. Benedetti, J. W. Kaiser, and J.-J. Morcrette, "McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions," *Atmos. Meas. Techn.*, vol. 6, no. 9, pp. 2403–2418, 2013.
[35] B. Haurwitz, "Insolation in relation to cloudiness and cloud density," *J. Meteorol.*, vol. 2, no. 3, pp. 154–166, 1945.
[36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in PyThon," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[37] F. Santosa and W. W. Symes, "Linear inversion of band-limited reflection seismograms," *SIAM J. Scientific Stat. Comput.*, vol. 7, no. 4, pp. 1307–1330, Oct. 1986.
[38] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Roy. Stat. Soc. B, Methodol.*, vol. 58, no. 1, pp. 267–288, Jan. 1996.

ANNETTE ESCHENBACH was born in Werneck, Bayern, Germany, in 1980. She received the B.Sc. degree in computer science from Free University, Berlin, and the M.Eng. degree in business and engineering from the Beuth University of Applied Sciences, in 2010. From 2010 to 2014, she worked in the solar industry for juwi solar GmbH and Hanwha Q Cells Company, Ltd., Project and Product Management Departments.

GUILLERMO YEPES was born in Madrid, Spain, in 1988. He received the B.S. degree in physics from the Universidad Complutense de Madrid, in 2015, and the M.S. degree in new electronic and photonic technologies from the Universidad Complutense de Madrid, in 2018.

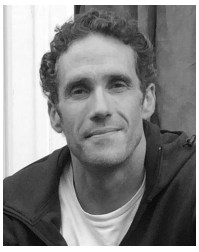
Since 2019, he has been working as a Data Analyst for solar energy in the private sector in Spain. His research interests include neural networks and renewable energy.

CHRISTIAN TENLLADO received the M.Sc. degree in electronic engineering and the Ph.D. degree in computer architecture from the Complutense University of Madrid, Spain, in 2001 and 2007, respectively.

During his Ph.D., he was a Visiting Research with imec, Leuven, working on superword level parallelism. He is currently an Associate Professor with the Department of Computer Architecture and Systems Engineering, Universidad Complutense de Madrid, Spain. Over the years, he has been working on different topics from parallel processing, with a special emphasis on emerging architectures, code generation and optimization, to memory system optimization, and low power devices and network protocols for the IoT ecosystem.

JOSÉ I. GÓMEZ-PÉREZ received the M.Sc. degree in computer science and the Ph.D. degree from the Complutense University of Madrid, Spain, in 2001 and 2007, respectively.

During his Ph.D., he was a Visiting Research with imec, Leuven, working on optimizing low-power embedded systems, both at the compiler and system level. After his Ph.D., he moved to GPGPU computing, still at the compiler level. He is currently an Associate Professor with the Department of Computer Architecture and Systems Engineering, Universidad Complutense de Madrid, Spain. He is also an Associate Professor with the ArTeCS Group, Department of Computer Architecture and Automation, UCM. His current research interest includes still in low power embedded systems but in the context of the IoT ecosystems.



LUIS PIÑUEL received the M. Sc. and Ph.D. degrees in computer science from the Universidad Complutense de Madrid (UCM), in 1996 and 2003, respectively.

From June 2010 to April 2015, he also served as an Academic Secretary with the Physics Faculty, Universidad Complutense de Madrid. He is currently an Associate Professor with the Department of Computer Architecture and Systems Engineering, Universidad Complutense de Madrid, Spain. His research interests include computer architecture, high-performance computing, low-power microarchitectures, embedded systems, and resource management for emerging computing systems. In these fields, he has coauthored more than 70 publications in prestigious journals and international conferences, several book chapters. He has advised or co-advised 5 Ph.D. dissertations. Ha has been a member of the technical program and organization committee of the some relevant conferences (e.g., HPCA). He worried about improving knowledge transfer between research institutions and industry. He has directed more than 15 research contracts with different companies, like Texas Instruments or Indra among others.

LUIS F. ZARZALEJO was born in Madrid, Spain, in 1967. He received the B.S. and M.S. degrees in physics from the Complutense University of Madrid, in 1993, the Diploma degree in advances studies and the Ph.D. degree in atomic physics and renewable energy from the Complutense University of Madrid, in 2001 and 2005, respectively. Since February 1994, he has been the Spanish National Center for Energy and Environment Research (CIEMAT), (www.ciemat.es), working in the Solar Passive Systems Group with responsibilities on Research and Development for climate characterization until 2000. In 2001, he became a Co-Founder of the Solar Characterization and Measurement Group, depend on the Plataforma Solar de Almería (www.psa.es), which leads in 2006. Since 2004, he has been involved in several activities as an Expert within the Task 36 and Task 46 of Solar Heating and Cooling (IEA) devoted to Solar Radiation Knowledge Management. From 2006 to 2010, he was responsible of Solar Resource Evaluation activities at Solar Energy Research Center (CIESOL), www.ciesol.es, research center depend on University of Almería and PSA-CIEMAT. In 2007, he was a Co-Founding Partner of technology based company IrSOLaV, of which was member of its scientific committee until 2010. He has authored more than 30 chapters in handbooks on solar energy with ISBN and more than 40 articles in scientific journals with SCI. He has been a member of the Scientific Committee of the XIII Congresso Iberico e VII Iberoamericano de Energía Solar (CIES 2006), ISES Solar World Congress 2015, and XVI Congreso Ibérico y XII Iberoamericano de Energía Solar (CIES) 2018. Since 2012, he has been a member of the board of the Spanish Association of Solar Energy (AEDES). In 2011 he was awarded with the Solar Energy Journal Best Paper Award (*Solar Energy* journal, Volume 84, Issue 10).

STEFAN WILBERT Diploma thesis in physics at the University of Bonn and his Ph.D. thesis at the University of Aachen were related to meteorological effects on solar power plants. He is currently works with the Institute of Solar Research, DLR, in the delegation at the Plataforma Solar de Almería, Spain, since 2006. He is also a Leader of the Research Group Solar Energy Meteorology. He also acts as a Sub Task and an Activity Leader in the International Energy Agency's tasks on solar resources and forecasting within the PV Power Systems and SolarPACES programs. His research interests are meteorological measurements for solar energy applications, radiative transfer modeling, nowcasting of solar radiation, and solar power plant simulation. He received the SolarPACES Technology Innovation Award, in 2014, and the Borchers Medal, in 2015.