



Short-Term Solar Forecasting Performance of Popular Machine Learning Algorithms

Preprint

Alex Dobbs, Tarek Elgindy, Bri-Mathias Hodge,
and Anthony Florita
National Renewable Energy Laboratory

*Presented at the International Workshop on the Integration of
Solar Power into Power Systems (Solar Integration Workshop)
Berlin, Germany
October 24–26, 2017*

**NREL is a national laboratory of the U.S. Department of Energy
Office of Energy Efficiency & Renewable Energy
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy
Laboratory (NREL) at www.nrel.gov/publications.

Conference Paper
NREL/CP-5D00-70030
October 2017

Contract No. DE-AC36-08GO28308

NOTICE

The submitted manuscript has been offered by an employee of the Alliance for Sustainable Energy, LLC (Alliance), a contractor of the US Government under Contract No. DE-AC36-08GO28308. Accordingly, the US Government and Alliance retain a nonexclusive royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for US Government purposes.

This report was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or any agency thereof.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at www.nrel.gov/publications.

Available electronically at SciTech Connect <http://www.osti.gov/scitech>

Available for a processing fee to U.S. Department of Energy and its contractors, in paper, from:

U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831-0062
OSTI <http://www.osti.gov>
Phone: 865.576.8401
Fax: 865.576.5728
Email: reports@osti.gov

Available for sale to the public, in paper, from:

U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312
NTIS <http://www.ntis.gov>
Phone: 800.553.6847 or 703.605.6000
Fax: 703.605.6900
Email: orders@ntis.gov

Cover Photos by Dennis Schroeder: (left to right) NREL 26173, NREL 18302, NREL 19758, NREL 29642, NREL 19795.

NREL prints on paper that contains recycled content.

Short-Term Solar Forecasting Performance of Popular Machine Learning Algorithms

Alex Dobbs, Tarek Elgindy, Bri-Mathias Hodge, and Anthony Florita
National Renewable Energy Laboratory
Golden, CO, USA

Abstract—A framework for assessing the performance of short-term solar forecasting is presented in conjunction with a range of numerical results using global horizontal irradiation (GHI) from the open-source Surface Radiation (SURFRAD) data network. A suite of popular machine learning (ML) algorithms was compared according to a set of statistically distinct metrics and benchmarked against the persistence-of-cloudiness forecast and a cloud motion forecast. Results showed significant improvement compared to the benchmarks with trade-offs among the ML algorithms depending on the desired error metric. Training inputs included time series observations of GHI for a history of years, historical weather and atmospheric measurements, and corresponding date and time stamps such that training sensitivities could be inferred. Prediction outputs were GHI forecasts for 1, 2, 3, and 4-hours ahead of the issue time, and they were made for every month of the year for 7 locations. Future photovoltaic power and energy outputs can be produced using GHI forecasts to better understand power system impacts.

Keywords—SURFRAD; solar forecasting; numerical weather prediction; machine learning

I. INTRODUCTION

The integration of high levels of solar power into the electricity grid poses a significant challenge for grid operators because of the uncertainty and variability of solar generation. Accurately forecasting solar energy production for unit commitment can reduce this generation uncertainty, which translates to significant savings. One study found that \$5 billion in savings per year could be achieved over the Western Electricity Coordinating Council by integrating solar and wind forecasts into unit commitments [1]. A comprehensive review of state-of-the-art methods in solar forecasting [2], which primarily focused on averaged rather than instantaneous forecasts, necessitates the benchmarking and performance comparisons presented herein. Although physical models of atmospheric conditions—i.e., numerical weather prediction (NWP) methods—are extensively used for day-ahead forecasts, it has been shown that for temporal scales less than 4-hours ahead, forecasting accuracy is greatly improved by applying model output statistics to the NWP output [3]. A variety of regression approaches have been applied to improve short-term solar forecasting [4]–[8]. However, for forecasts from 15 minutes to 4-hours ahead, hybrid ML approaches have achieved significant improvements compared to the traditional NWP models [9]. Exogenous inputs such as satellite data have improved the

accuracy of short-term forecasts at several SURFRAD sites [10]. Other studies have also incorporated exogenous observations such as relative humidity and cloud cover to improve forecasting accuracy [11].

The methodology developed in this paper used irradiance and exogenous weather time series data from seven publicly available weather stations in the SURFRAD network, and then used different ML algorithms to predict solar irradiance point forecasts 1, 2, 3, and 4-hours ahead of the issue time. The following sections describe the specific data utilized, preprocessing requirements, and time-shifting techniques applied in this study. A brief overview of the ML forecasting methods is then given, followed by results and a discussion comparing the performance of the ML models to the benchmarks and against each other. Finally, concluding remarks and suggestions for future research are presented.

II. METHODOLOGY

A. Preprocessing Input Data

The routines developed in this paper were trained and tested on data from the SURFRAD observation sites in Desert Rock, NV; Fort Peck, MT; Boulder, CO; Sioux Falls, SD; Bondville, IL; Goodwin Creek, MS; and Penn State, PA. Each site has 11 years of weather measurements at 1-minute resolution from 2009 to present, and 3-minute resolution from 2004–2008. This array of sites offers climatically unique weather situations that span the U.S. Global horizontal irradiance (GHI) at the SURFRAD sites is best represented by the *global downwelling* solar measurements. The clear-sky GHI at time t is denoted by GHI_{clear}^t and represents the theoretical GHI at time t assuming zero cloud coverage; it is computed using the Bird model [12]. The clear-sky index is a metric of cloud cover that has been used extensively in forecasting literature [13]–[15]. The clear-sky index at time t denoted by $Kt_i^{(t)}$ is the ratio between the instantaneous observed GHI^t and the theoretical maximum GHI_{clear}^t . It is noted that cloud-focusing events can cause GHI^t to exceed GHI_{clear}^t . Current time, temperature, relative humidity, wind speed, wind direction, pressure, thermal infrared, GHI^t , GHI_{clear}^t , and $Kt_i^{(t)}$ were used as independent variables for the input training vectors.

Rather than training on the observed instantaneous GHI values at the 1-, 2-, 3-, or 4-hour-ahead forecast horizons (*f.h.*), which might not be representative of the *most*

probable $GHI^{f.h.}$, the ML models were trained on the averaged clear-sky index for the hour. The average hourly clear-sky index ending at time $f.h.$ is denoted by $Kt_a^{(f.h.)}$, as in:

$$Kt_a^{(f.h.)} = \frac{\sum_{s=f.h.-60}^{f.h.} Kt_i^{(s)}}{60} \quad (1)$$

$Kt_a^{(f.h.)}$ was used as the dependent variable for the training vectors when building each model, and the models were then used to predict $Kt_a^{(f.h.)}$ when given unseen test vectors. The forecasted $Kt_a^{(f.h.)}$ value was then multiplied by $GHI_{clear}^{f.h.}$ from the Bird model to predict $GHI^{f.h.}$, as in:

$$GHI_{prediction}^{f.h.} = Kt_a^{(f.h.)} \cdot GHI_{clear}^{f.h.} \quad (2)$$

This ML forecast was finally compared to the testing input's corresponding $GHI_{observed}^{f.h.}$ from the SURFRAD data to assess forecasting accuracy.

Data were partitioned by month, and any entries with missing or misreported data were removed. All nighttime entries—i.e., entries with current or future GHI readings less than 20 W/m²—were removed to improve the performance of the ML algorithms. Input data were scaled to the range [0,1] to address the large variations in the natural domain of elements. Input vectors were adjusted by finding the maximum and minimum values of each element in the training set and scaling variable x of feature j of an input data vector i to be:

$$scaled(x_i) = \frac{x_i - j_{min}}{j_{max} - j_{min}} \quad (3)$$

Each ML algorithm has many hyperparameters that can be tuned, and these internal parameters were set using a grid search method. Predictions were made for each forecasting situation at a frequency equal to the forecast horizon timescale. For example, when forecasting GHI for 3-hours ahead for a specific site and month, the ML models made predictions at 3-hour intervals every day of the month for all daylight hours.

B. Description of Forecasting Methods

1) *Persistence of Cloudiness*: Persistence forecasts use the current cloud cover to predict the future GHI. In this study, the clear-sky index at the forecast horizon was set to the current clear-sky index at time t and multiplied by $GHI_{clear}^{f.h.}$, as in:

$$Kt_i^{(f.h.)} = Kt_i^{(t)} \cdot GHI_{clear}^{f.h.} \quad (4)$$

This simple model is most effective for very short-term forecasts (e.g., minutes ahead), but it can also be used to make forecasts from 1 to 4-hours ahead. Persistence forecasts were provided as a benchmark for the forecasts made by ML methods in this study.

2) *Support Vector Machines*: Support vector machines (SVMs) have been shown to work well in conjunction with other methods in solar forecasting applications [16]–[18]. SVM regression estimates a target function based on training instances. SVMs operate by transforming a nonlinearly separable feature space into a multidimensional space in which variables can be separated by a three-

dimensional hyperplane. SVMs map the original data into this higher-dimensional space using a technique known as the “kernel trick,” which allows for different perspectives on the data. The output observations are assumed to take the form of $y_i = \omega_i \cdot \kappa(x_i) + b$, where y_i is the output observation for training instance i , x_i is the input training vector for instance i , ω_i is a weight vector which defines the functional form, b is the bias constant, and $\kappa(x, x') = \varphi(x) \cdot \varphi(x')$ is the kernel function. This study used a nonlinear radial basis function, as in:

$$\kappa(x, x') = e^{-\frac{\|x-x'\|^2}{2\sigma^2}} \quad (5)$$

where σ is a free parameter. The final objective was to minimize the deviation errors between the output observation y_i , and the linear functional form of $(\omega_i \cdot x_i) + b$ while maximizing the margin of space on either side of the hyperplane.

3) *Artificial Neural Networks*: Artificial neural networks (ANNs) are one of the most popular ML methods used in solar forecasting [19]–[21]. ANNs contain layers of nodes with connections between nodes in adjacent layers. The input layer has one node for each input signal. These included current time, temperature, and GHI. The output layer is a single output node, which in this study was the forecasted $Kt_a^{f.h.}$. One or more hidden layers contain a predetermined number of nodes and connect the input and output layers. Each node receives a weighted sum of input from the nodes in the previous layer, and it applies an activation function to the weighted sum. The weights \vec{w} are determined by training the network on inputs and known outputs with a learning function. The architecture, activation function, and learning function are three features that are predetermined before the training process. A three-layer ANN with 9-n-1 nodes in each layer was adopted in this study, expressed as:

$$y(\vec{x}, \vec{w}) = \sum_{j=1}^n w_j f\left(\sum_{k=1}^9 w_{k,j} x_k\right) \quad (6)$$

where n was the number of nodes in the hidden layer, which was calibrated by the grid search method, and $f(x)$ was the sigmoidal activation function $f(x) = \frac{1}{1+e^{-x}}$. Several back-propagation learning methods were tested to ensure a well-trained ANN model, such as the vanilla back-propagation, momentum term back-propagation, and batch version back-propagation.

4) *Random Forests*: Although SVMs and ANNs are popular for short-term solar forecasting, random forest (RF) models have been used for solar forecasting in several studies [22]–[24]. A random forest is a collection of single classification and regression trees (CART) in which each CART is trained by a bagging algorithm that avoids overfitting the RF models. To train each CART, the training set is partitioned by the bootstrap sample method. The robustness of a CART can be improved by blending the CARTs according to their performance. Although each CART might have a bias because of its structure and the specific subset of features selected, aggregating all decision trees can significantly reduce the error bias of the final

output. RFs do this by averaging all CARTs in the ensemble.

5) *Gradient Boosting Method*: The Gradient Boosting Method (GBM) [25] is a less frequently used ML approach to solar forecasting, and it is an extension of the RF method. In RF, gradient boosting uses a randomly built ensemble of decision trees to make a more accurate prediction. GBM incrementally adds trees to the ensemble during the training phase. For a training instance i with input signals \vec{x}_i and known GHI forecast of y_i , the forecast made by an ensemble of t trees is denoted by \hat{y}_i^t . The forecast with $t+1$ trees is defined as $\hat{y}_i^{t+1} = \hat{y}_i^t + f(x_i)$, i.e. $f(x_i) = y_i - \hat{y}_i^t$. Thus, to improve the model in the following iteration, a regression tree f is custom fit to the data $(\vec{x}_1, y_1 - \hat{y}_1^t), \dots, (\vec{x}_N, y_N - \hat{y}_N^t)$, where N is the number of training instances. Each new regression tree f is added to the ensemble one at a time to correct the residuals. These residuals are the negative gradients for the loss function $\phi(y, \hat{y}^t)$, which allows gradient descent methods to determine the structure of each tree before it is added to the ensemble.

C. Situation Dependent, Multi-Model Forecasting

Forecasts 1, 2, 3, and 4-hours ahead were generated for all 12 months at all 7 SURFRAD sites. The developed code was run to model each unique forecasting situation. Each run trained all four ML algorithms on preprocessed data for the desired month from the years 2004–2008 and 2010–2014. After the models were built, they were tested on unseen data from 2009, and forecasts were made for the desired forecast horizon.

III. VALIDATION METRICS

A suite of validation metrics was used to compare the forecast accuracy of different methodologies and situations in this study. A thorough discussion of different validation metrics was covered by [26] to compare N observed GHI values \vec{G} to the N forecast values \vec{H} . Root mean square error (RMSE) and mean absolute error (MAE) are commonly used metrics that measure the difference between the forecasted and actual GHI values. RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (G(i) - H(i))^2} \quad (7)$$

and MAE is defined as:

$$MAE = \frac{1}{N} \sum_{i=1}^N |G(i) - H(i)| \quad (8)$$

The RMSE metric has been commonly used to evaluate the overall accuracy of forecasts, and it penalizes large forecasting errors with its square order. The MAE metric is also appropriate for evaluating errors through the entire forecasting period, and it is widely used in regression problems and by the renewable energy industry. It does not penalize large forecast errors as much as the RMSE metric does. These two metrics can be further normalized to provide mean absolute percentage error (MAPE), defined as:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|G(i) - H(i)|}{\max(G(i))} \quad (9)$$

and normalized RMSE (nRMSE), defined as:

$$nRMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{G(i) - H(i)}{\max(G(i))} \right)^2} \quad (10)$$

These normalized validation metrics were used to compare forecasting accuracy at different spatial locations. Smaller values of these validation metrics indicate a higher forecasting accuracy.

IV. RESULTS AND DISCUSSION

A. Machine Learning Forecasts vs. Benchmark Methods

To calibrate this study against existing literature, forecasts were made for the same 1, 2, 3, and 4-hour-ahead forecast horizons as in [27]. They tested the period from August 23, 2008, through August 31, 2009. Limitations arose in this study from SURFRAD's data resolution change for the years preceding 2009, which consisted of 3-minute data, causing this study to use a slightly different testing period: from January 1, 2009–December 31, 2009. Seasons were partitioned into four 3-month periods, e.g. January 1–March 31, etc. Thus, results were compared to [27] for a slightly different time period, and to the persistence-of-cloudiness forecasts made for 2009.

Table I shows the RMSE values for the ML predictions made in this study, the persistence-of-cloudiness forecasts' RMSE values, and the RMSE values for forecasts from [27]. The values in the yellow columns were found by taking the best-performing ML algorithm per month and compiling these values into seasonal and yearly results. The ML models employed in this study outperformed the persistence-of-cloudiness forecasts in every situation, with average RMSE values of 92.36 W/m² and 122.12 W/m², respectively. This study outperformed the forecasting methodology in [27], which was tested on a different 12-month period, with average RMSE values of 92.36 W/m² and 108.29 W/m², respectively.

Performance comparisons of forecasting methods were also evaluated using the relative frequency (rounded) that a given technique had in producing the lowest error. This study outperformed the method in [27] on a seasonal basis for 1- and 4-hour-ahead forecasts in 86% and 57% of tests, respectively, based on RMSE values. However, the results in [27] outperformed the results in this study for 2- and 3-hour ahead forecasts in 57% and 61% of forecasts, respectively. The forecasts in [27] also outperformed this study in 68% of winter and 57% of spring seasonal forecasts, whereas the ML models outperformed the forecasts in [27] in 75% and 79% of all situations for the summer and fall seasons, respectively. When broken down by geographic location, this study outperformed [27] in Boulder, Fort Peck, Desert Rock, and Bondville with respective relative frequencies of 75%, 94%, 63%, and 53% in all tests. Their study outperformed this study more when forecasting across all situations in Goodwin Creek, Penn State, and Sioux Falls 75%, 56%, and 56% of the time, respectively. These relative frequencies took into account only the number of times that a method outperformed the other, and they did not consider the margin of difference, measured in W/m², between competing forecasts.

TABLE I. RMSE VALUES FOR MACHINE LEARNING (ML) FORECASTS JAN. 1 2009–DEC. 31 2009, PERSISTENCE-OF-CLOUDINESS (PC) FORECASTS JAN. 1 2009–DEC. 31 2009, AND PEREZ ET AL. (PZ) FORECASTS AUG. 23 2008–AUG. 31 2009

	Forecast Horizon	Boulder			Bondville			Goodwin Creek			Fort Peck			Desert Rock			Penn State			Sioux Falls		
		ML	PC	PZ	ML	PC	PZ	ML	PC	PZ	ML	PC	PZ	ML	PC	PZ	ML	PC	PZ	ML	PC	PZ
ALL YEAR	1-hour	74	104	120	62	83	85	71	96	80	56	79	94	52	76	80	67	96	86	52	74	68
	2-hour	108	142	139	98	118	98	103	130	101	81	110	106	72	103	88	97	132	99	81	106	84
	3-hour	123	161	154	116	135	112	125	146	114	94	126	123	83	116	96	114	151	113	96	126	102
	4-hour	125	169	166	121	143	122	120	152	127	93	130	132	82	122	104	117	157	124	103	136	115
WINTER	1-hour	55	74	64	51	66	60	58	87	48	36	53	107	45	66	46	53	72	57	41	62	48
	2-hour	81	98	71	82	104	66	98	128	59	52	74	105	63	92	48	79	102	57	65	96	58
	3-hour	96	113	81	104	117	74	122	146	66	62	84	109	75	106	59	91	122	59	82	117	69
	4-hour	87	119	85	105	123	81	111	147	70	58	84	112	84	107	70	96	127	65	89	122	78
SPRING	1-hour	97	143	125	84	114	93	94	127	92	75	108	110	71	108	86	84	117	83	66	94	69
	2-hour	137	195	141	133	154	109	125	171	122	110	149	124	106	147	95	119	161	99	103	133	90
	3-hour	170	218	157	147	178	123	159	190	144	129	174	141	120	155	111	143	183	118	124	156	107
	4-hour	162	228	170	159	189	137	145	202	164	134	186	148	115	171	115	145	190	137	131	171	126
SUMMER	1-hour	96	136	143	76	97	100	88	119	92	81	101	91	48	71	99	88	125	112	67	90	80
	2-hour	137	185	175	111	134	115	121	151	113	110	143	109	64	85	110	122	170	127	99	129	98
	3-hour	144	211	189	135	153	129	135	168	120	125	164	129	70	105	111	140	194	142	112	155	120
	4-hour	175	222	204	138	169	138	139	175	129	122	173	142	74	118	124	138	208	152	118	168	129
FALL	1-hour	46	63	85	35	56	58	44	50	55	34	52	59	43	60	55	45	71	60	35	48	49
	2-hour	78	92	97	67	80	68	67	70	66	52	73	67	57	87	62	69	96	71	57	67	54
	3-hour	81	103	110	76	90	84	83	81	81	59	81	83	65	97	69	80	104	76	65	76	64
	4-hour	75	107	120	81	143	89	87	81	94	58	78	88	56	94	72	89	102	83	74	81	80

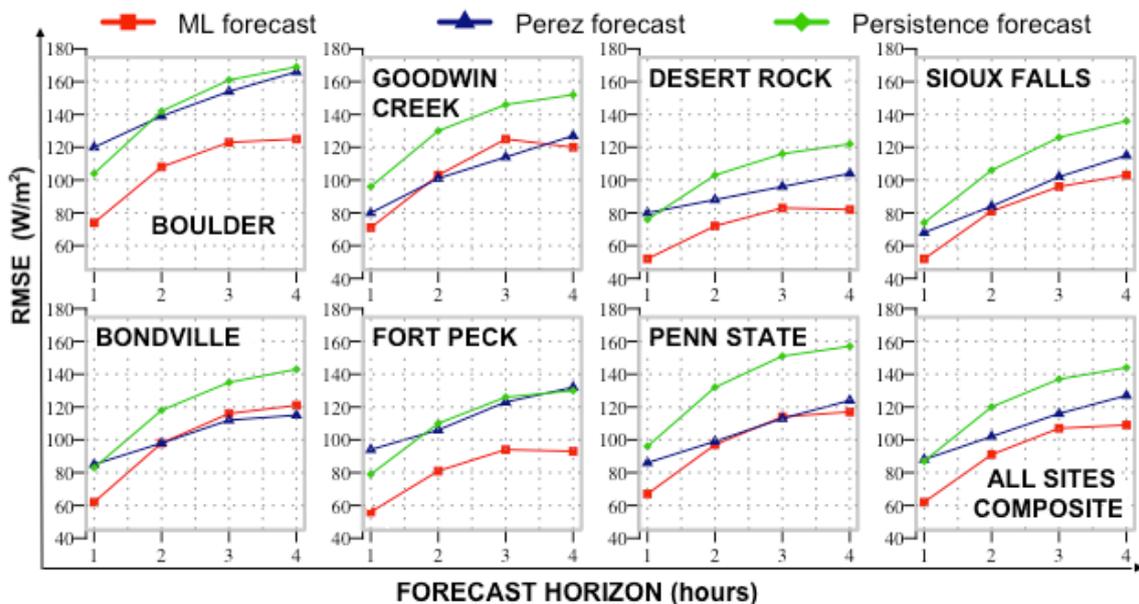


Figure 1. Annual RMSE averages for all SURFRAD sites.

Relative frequencies were useful to show which method works best in individual forecasting situations, but were not ideal for assessing a method's overall ability to minimize forecasting errors across all situations. Fig. 1 compares the RMSE values of the ML forecasts to the two benchmark forecasts, [27] and persistence, for all SURFRAD sites. The graphs indicate RMSE values as a function of the forecast horizon time span. The relative strength of the 2- and 3-hour-ahead forecasts in [27] compared to the ML models is especially apparent in the graphs for Bondville, Goodwin Creek, and Penn State. Comparisons were also made by showing the percentage improvement, defined as the difference between the RMSE of the ML forecast and the RMSE of the benchmark forecast divided by the RMSE of the benchmark. The largest improvement among all situations occurred in Fort Peck, where the suite of ML algorithms demonstrated a 28.8% improvement compared to the average RMSE values from [27], followed by a 25.7% improvement in Boulder. Improvements compared to the RMSE averages in [27] were also made in Desert Rock, Sioux Falls, Penn State, Bondville, and Goodwin Creek by 21.5%, 10.0%, 6.4%, 4.8%, and 0.7%, respectively. It is interesting to note that this study showed the largest improvements in RMSE scores for Boulder, Fort Peck, and Desert Rock. These three sites are located at the highest elevations and are the three westernmost locations in the SURFRAD network. ML forecasts outperformed the RMSE results from the persistence-of-cloudiness forecasts for all sites as well. They showed the greatest improvement in the four locations of Boulder (25.3%), Desert Rock (30.7%), Penn State (26.3%), and Fort Peck (27.2%).

B. Performance of ML Algorithms Against Each Other

Table II compares the four ML models used in this study to each other by showing the relative frequency (rounded) that each algorithm had in producing the lowest RMSE values in the listed forecasting situations. There were 84 fore

TABLE II. ML METHODS' PERFORMANCES BASED ON RMSE

Forecast Situation	RF	SVM	ANN	GBM
1-hour ahead	8%	31%	42%	19%
2-hour ahead	22%	20%	38%	20%
3-hour ahead	25%	13%	45%	16%
4-hour ahead	29%	20%	38%	13%
Winter	20%	22%	38%	20%
Spring	15%	25%	40%	20%
Summer	24%	17%	47%	12%
Fall	25%	23%	38%	14%
Boulder	27%	15%	43%	15%
Bondville	21%	21%	35%	23%
Goodwin Creek	21%	23%	39%	17%
Fort Peck	23%	31%	31%	15%
Desert Rock	15%	31%	42%	12%
Penn State	25%	15%	45%	15%
Sioux Falls	15%	12%	50%	23%
All Situations	20.8%	21.1%	41.1%	17.0%

TABLE III. ML METHODS' PERFORMANCES BASED ON MAE

Forecast Situation	RF	SVM	ANN	GBM
1-hour ahead	7%	65%	17%	11%
2-hour ahead	15%	44%	26%	15%
3-hour ahead	14%	36%	36%	14%
4-hour ahead	23%	32%	26%	19%
Winter	17%	41%	25%	17%
Spring	6%	60%	19%	15%
Summer	21%	38%	25%	16%
Fall	12%	40%	35%	13%
Boulder	10%	48%	25%	17%
Bondville	8%	52%	17%	13%
Goodwin Creek	23%	35%	19%	23%
Fort Peck	21%	42%	27%	10%
Desert Rock	21%	48%	19%	12%
Penn State	8%	42%	40%	10%
Sioux Falls	10%	44%	27%	19%
All Situations	14.6%	44.3%	26.2%	14.9%

-casting situations for each forecast horizon (12 months per 7 sites), 84 situations for each seasonal forecast (3 months per 7 sites for 4 forecast horizons), and 48 for each geographic forecast (12 months per four forecast horizons) made in this study. The ANN algorithm was the top performer in each of these situational categories and produced the lowest error value in 41.1% of the 336 total forecasting situations based on the RMSE metric. The SVM algorithm performed equally well in Fort Peck, and when the seasons were broken down by month, SVM outperformed ANN during all April forecasts by 42% to 32%. The RF and SVM algorithms performed almost equally well when considering all forecast situations. A similar table was constructed for the nRMSE metric, in which the RF, SVM, ANN, and GBM methods achieved the same overall relative frequencies as those in Table II.

Table III is similar to Table II except that it shows the relative frequency of each model's ability to produce the lowest MAE values in each forecasting situation. The SVM algorithm produced the lowest MAE values in all types of forecasting situations more often than any of the other ML models, though it tied ANN when making 3-hour-ahead forecasts. It was the top performer in more situations according to the MAE metric than the ANN was when considering the RMSE metric. The SVM performed best most often in 1-hour forecasts, and it approached smaller relative frequencies as the forecast horizon extended in time. A similar table was constructed for the MAPE metric that shows the RF, SVM, ANN, and GBM methods achieved relative frequencies of 14.6%, 44.3%, 26.2%, and 14.9%, respectively.

V. CONCLUSION AND FUTURE WORK

This paper assessed the performance of ML techniques and their validity in improving short-term solar forecasting. The ML approaches were compared to other forecasting methods, and individual ML algorithms were compared to each other. ML forecasts generated lower average RMSE values than a cloud motion forecasting method for all seven sites, with the biggest improvements for the three sites at the

highest elevations and westernmost locations in the SURFRAD network. They also outperformed the persistence-of-cloudiness forecasts at all seven sites, with the greatest improvements at the four locations of Boulder, Desert Rock, Penn State, and Fort Peck. The ML forecasts produced the lowest RMSE more often than the cloud motion method across all summer and fall seasonal forecasts as well as for 1 and 4-hour-ahead forecasts. Assessing the performance of the four algorithms against each other did not reveal any strong situation-dependent sensitivities because each algorithm was capable of occasionally making the best forecast in various forecasting situations, though some less than others. However, either SVMs or ANNs most often led to the lowest forecasting errors depending on the error metric used. An ANN was the preferred model if minimizing the largest point forecast errors was the greatest concern, according to the RMSE metric; however, the SVM was the best performer if minimizing the average absolute difference, MAE, was preferred. These results depended on the specific tuning of each algorithm's hyperparameters.

This solar irradiance forecasting methodology can be further extended by increasing the forecast horizon resolutions from hourly increments to 5-minute increments, allowing for more dynamic time series information about upcoming ramping events. Further work is needed to fine-tune each ML algorithm, and future research should also look into optimizing ML hyperparameters for each situation-dependent forecast. Improved forecasts will help facilitate higher penetrations of solar energy into the grid by providing increased grid reliability and minimizing costs associated with ramping events.

ACKNOWLEDGMENT

This work was supported by the U.S. Department of Energy under the Solar Energy Technologies Program through Contract No. DE-AC36-08-GO28308 with the National Renewable Energy Laboratory (NREL). A.D. would like to thank the NREL workforce development team and the CCI program.

REFERENCES

- [1] D. Lew, D. Piwko, N. Miller, G. Jordan, K. Clark, and L. Freeman, How Do High Levels of Wind and Solar Impact the Grid? The Western Wind and Solar Integration Study (Tech. Report). Golden, CO: National Renewable Energy Laboratory, 2010, pp. 275–300.
- [2] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," *Prog. Energy Combust. Sci.*, 6th ed., vol. 39, pp. 535–576, 2013.
- [3] E. Lorenz, D. Heinemann, H. Wickramaratne, H. G. Beyer, and S. Bofinger, "Forecast of ensemble power production by grid-connected PV systems," presented at the 20th European PV Conference, Milano, 2007.
- [4] H. T. Pedro and C. F. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Solar Energy*, 7th ed., vol. 86, pp. 2017–2028, 2012.
- [5] R. Marquez, V. G. Gueorguiev and C. F. Coimbra, "Forecasting of global horizontal irradiance using sky cover indices," *J. Solar Energy Eng.*, 1st ed., vol. 135, pp. 011017, 2013.
- [6] A. Mellit, H. Eleuch, M. Benghanem, C. Elaoun, and A. Massi Pavan, "An adaptive model for prediction of global, direct and diffuse hourly solar irradiance," *Energy Convers. Manage.*, 4th ed., vol. 5, pp. 771–782, 2010.
- [7] W. Ji and K. C. Chee, "Prediction of hourly solar radiation using a novel hybrid model of ARMA and TDNN," *Solar Energy*, 5th ed., vol. 85, pp. 808–817, 2011.

- [8] D. Yang, P. Jirutitjaroen, and W. M. Walsh, "Hourly solar irradiance time series forecasting using cloud cover index," *Solar Energy*, 12th ed., vol. 86, pp. 3531–3543, 2012.
- [9] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," *Prog. Energy and Combust. Sci.*, 6th ed., vol. 39, pp. 535–576, 2013.
- [10] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., D. Renne, and T. E. Hoff, "Validation of short and medium term operational solar radiation forecasts in the US," *Solar Energy*, 12th ed., vol. 84, pp. 2161–2172, 2010.
- [11] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," *Solar Energy*, 3rd ed., vol. 83, pp. 342–349, 2009.
- [12] R. E. Bird and R. L. Hulstrom, *Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces* (Tech. Rep.). Golden, CO: Solar Energy Research Inst., 1981.
- [13] R. Marquez, V. G. Gueorguiev, and C. F. Coimbra, "Forecasting of global horizontal irradiance using sky cover indices," *J. Solar Energy Eng.*, 1st ed., vol. 135, pp. 011017, 2013.
- [14] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., D. Renne, and T. E. Hoff, "Validation of short and medium term operational solar radiation forecasts in the US," *Solar Energy*, 12th ed., vol. 84, pp. 2161–2172, 2010.
- [15] P. Mathiesen and J. Kleissl, "Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States," *Solar Energy*, 5th ed., vol. 85, pp. 967–977, 2011.
- [16] De Giorgi, M. G., P. M. Congedo, M. Malvoni, and D. Laforgia, "Error analysis of hybrid photovoltaic power forecasting models: A case study of mediterranean climate," *Energy Convers. Manage.*, vol. 100, pp. 117–130, 2015.
- [17] Y. Chu, M. Li, H. T. C. Pedro, and C. F. Coimbra, "Real-time prediction intervals for intra-hour DNI forecasts," *Renew. Energy*, vol. 83, pp. 234–244, 2015.
- [18] Y. Zhang and Z. Hamidreza, "Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators," *IEEE Trans. Smart Grid*, 5th ed., vol. 6, pp. 2253–2262, September 2015.
- [19] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, 5th ed., vol. 2, pp. 359–366, 1989.
- [20] C. Cornaro, M. Pierro, and F. Bucci, "Master optimization process based on neural networks ensemble for 24-h solar irradiance forecast," *Solar Energy*, vol. 111, pp. 297–231, 2015.
- [21] Y. Chu, M. Li, H. T. C. Pedro, and C. F. Coimbra, "Real-time prediction intervals for intro-hour DNI forecasts," *Renew. Energy*, vol. 83, pp. 234–244, 2015.
- [22] M. Abuella and B. Chowdhury, "Random forest ensemble of support vector regression for solar power forecasting," in *Proceedings of Innovative Smart Grid Technologies, North American Conference*, 2017.
- [23] H. Zheng and A. Kusiak, "Prediction of wind farm power ramp rates: A data-mining approach," *J. Solar Energy Eng.*, 3rd ed., vol. 3, pp. 031011, 2009.
- [24] M. Zamo, O. Mestre, P. Arbogast, and O. Pannekoucke, "A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production," *Solar Energy*, vol. 105, pp. 792–803, 2014.
- [25] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Ann. Stat.*, pp. 1189–1232, 2001.
- [26] J. Zhang et al., "A suite of metrics for assessing the performance of solar power forecasting," *Solar Energy*, vol. 111, pp 157–175, 2015.
- [27] R. Perez, S. Kivalov, J. Schlemmer, K. Hemker Jr., D. Renne, and T. E. Hoff, "Validation of short and medium term operational solar radiation forecasts in the US," *Solar Energy*, 12th ed., vol. 84, pp. 2161–2172, 2010.