Metrics for Evaluating the Accuracy of Solar Power Forecasting

Preprint

J. Zhang, B.-M. Hodge, and A. Florita
National Renewable Energy Laboratory

S. Lu and H. F. Hamann
IBM TJ Watson Research Center

V. Banunarayanan
U.S. Department of Energy

To be presented at 3rd International Workshop on Integration of Solar Power into Power Systems
London, England
October 21 – 22, 2013

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-5500-60142
October 2013

Contract No. DE-AC36-08GO28308

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

Available electronically at http://www.osti.gov/bridge

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
phone: 865.576.8401
fax: 865.576.5728
email: mailto:reports@adonis.osti.gov

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

U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
phone: 800.553.6847
fax: 703.605.6900
email: orders@ntis.fedworld.gov
online ordering: http://www.ntis.gov/help/ordermethods.aspx

Cover Photos: (left to right) photo by Pat Corkery, NREL 16416, photo from SunEdison, NREL 17423, photo by Pat Corkery, NREL 16560, photo by Dennis Schroeder, NREL 17613, photo by Dean Armstrong, NREL 17436, photo by Pat Corkery, NREL 17721.

Printed on paper containing at least 50% wastepaper, including 10% post consumer waste.

Metrics for Evaluating the Accuracy of Solar Power Forecasting

Jie Zhang, Bri-Mathias Hodge, Anthony Florita
National Renewable Energy Laboratory
Golden, CO, USA

Siyuan Lu, Hendrik F. Hamann
IBM TJ Watson Research Center
Yorktown Heights, NY, USA

Venkat Banunarayanan
U.S. Department of Energy
Washington, D.C., USA

***Abstract*—Forecasting solar energy generation is a challenging task due to the variety of solar power systems and weather regimes encountered. Forecast inaccuracies can result in substantial economic losses and power system reliability issues. This paper presents a suite of generally applicable and value-based metrics for solar forecasting for a comprehensive set of scenarios (i.e., different time horizons, geographic locations, applications, etc.). In addition, a comprehensive framework is developed to analyze the sensitivity of the proposed metrics to three types of solar forecasting improvements using a design of experiments methodology, in conjunction with response surface and sensitivity analysis methods. The results show that the developed metrics can efficiently evaluate the quality of solar forecasts, and assess the economic and reliability impact of improved solar forecasting.**

Keywords-grid integration; ramps; response surface; solar forecasting; sensitivity analysis; uncertainty; variability

I. INTRODUCTION

The utility solar assessment study reported that solar power could provide 10% of U.S. power needs by 2025 [1]. At these high levels of solar energy penetration, solar power forecasting will become very important for electricity system operations. Solar forecasting is a challenging task, and solar power generation presents different challenges for the transmission and distribution networks of the grid, respectively. On the transmission side, solar power takes the form of centralized solar plants, a non-dispatchable component of the generation pool. On the distribution side, solar power is generated by a large number of distributed panels installed on building rooftops, which has the effect of modulating the load. Forecast inaccuracies of solar power generation can result in substantial economic losses and power system reliability issues because electric grid operators must continuously balance supply and demand to maintain the reliability of the grid [2].

A. Overview of Solar Forecasting

Solar irradiance variations are caused primarily by cloud movement, cloud formation, and cloud dissipation. In the literature, researchers have developed a variety of methods for solar power forecasting, such as the use of numerical weather prediction (NWP) models [3-5], tracking cloud movements from satellite images [6], and tracking cloud movements from direct ground observations with sky cameras [7-9]. NWP models are the most popular method for forecasting solar irradiance several hours or days in advance. Mathiesen and Kleissl [4] analyzed the global horizontal irradiance in the continental United States forecasted by three popular NWP models: the North American Model, the Global Forecast System, and the European Centre for Medium-Range Weather Forecasts. Chen et al. [5] developed an advanced statistical method for solar power forecasting based on artificial intelligence techniques. Crispim et al. [7] used total sky imagers (TSI) to extract cloud features using a radial basis function neural network model for time horizons from 1 to 60 minutes. Chow et al. [8] also used TSI to forecast short-term global horizontal irradiance. The results suggested that TSI was useful for forecasting time horizons up to 15 to 25 minutes. Marquez and Coimbra [9] presented a method using TSI images to forecast 1-minute averaged direct normal irradiance at the ground level for time horizons between 3 and 15 minutes. As discussed above, different solar irradiance forecast methods have been developed for various timescales. Loren et al. [10] showed that cloud movement–based forecasts likely provide better results than NWP forecasts for forecast timescales of 3 to 4 hours or less. Beyond that, NWP models tend to perform better.

B. Research Motivation and Objectives

Significant work has been done to develop solar forecasting models. However, evaluation of the performance of different forecasting methodologies is still not straightforward, because different researchers use different metrics as their own criteria. In addition, solar forecasting accuracy is dependent on geographic location and timescale of the data. Conventional measures of solar forecasting accuracy include root mean square error (RMSE), mean bias error (MBE), and mean absolute error (MAE) [3, 4]. Marquez and Coimbra [11] proposed a metric for using the ratio of solar uncertainty to solar variability to compare different solar forecasting models. Espinar et al. [12] proposed several metrics based on the Kolmogorov–Smirnov test to quantify differences between the cumulative distribution functions of actual and forecast solar irradiation data. However, many of the forecast metrics developed do not take into account the types of errors that have the most impact on power system operations. Extreme forecasting errors can have disproportionate economic and reliability impacts on operations, and therefore must be emphasized to some degree by any metric that wishes to capture the true impact of the forecasts. Establishing a standard set of metrics for assessing solar forecasting accuracy is (i) critical to evaluating the success of a solar forecasting effort, and (ii) useful for decision making of power system operators under the scenario of a high penetration of solar power.

The objective of this study is to develop a suite of generally applicable, value-based metrics for solar forecasting for a comprehensive set of scenarios (different time horizons, geographic locations, applications, etc.), which can assess the economic and reliability impact of improved solar forecasting. The sensitivity of proposed metrics to improved solar forecasts is also analyzed. The next section presents the developed metrics for different types of forecasts and applications. Section III summarizes the solar

Corresponding author, email: jie.zhang@nrel.gov

power data used in the paper. The methodologies for analyzing the sensitivities of different metrics are developed in Section IV. The results and discussion of the case study are presented in Section V. Concluding remarks and ideas on areas for future exploration are given in the final section.

II. Metrics Development

The proposed solar forecasting metrics in the paper can be broadly divided into five categories: (i) statistical metrics, including *Pearson's correlation coefficient*, *(normalized) root mean squared error* (RMSE), *maximum absolute error* (MaxAE), mean absolute error (MAE), *mean absolute percentage error* (MAPE), mean bias error (MBE), *Kolmogorov–Smirnov test Integral* (KSI), skewness, and kurtosis; (ii) variability estimation metrics, including different time and geographic scales, and distributions of forecast errors; (iii) uncertainty quantification and propagation metrics, including *standard deviation* and *information entropy* of forecast errors; (iv) ramping characterization metrics, including *swinging door algorithm* signal compression and *heat maps*; and (v) economic and reliability metrics, including *non-spinning reserves service* represented by 95th percentiles of forecast errors. A detailed formulation and physical explanation of each metric is described in the following sections.

A. Statistical Metrics

1) Pearson's correlation coefficient

Pearson's correlation coefficient is a measure of the correlation between two variables (or sets of data). In this paper, the Pearson's correlation coefficient, ρ, is defined as the covariance of actual and forecast solar power variables divided by the product of their standard deviations, which is mathematically expressed as:

$$\rho = \frac{\mathrm{cov}(p, \hat{p})}{\sigma_p \sigma_{\hat{p}}} \tag{1}$$

where p and $\hat{p}$ represent the actual and forecast solar power output, respectively. Pearson's correlation coefficient is a global error measure metric; a larger value of Pearson's correlation coefficient indicates an improved solar forecasting skill.

2) Root mean squared error (RMSE) and normalized root mean squared error (NRMSE)

The RMSE also provides a global error measure during the entire forecasting period, which is given by

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{p}_i - p_i)^2} \tag{2}$$

where p_i represents the actual solar power generation at the ith time step, $\hat{p}_i$ is the corresponding solar power generation estimated by a forecasting model, and N is the number of points estimated in the forecasting period. To compare the results from different spatial and temporal scales of forecast errors, we normalized the RMSE using the capacity value of the analyzed solar plants.

3) Maximum absolute error (MaxAE), Mean absolute error (MAE), mean absolute percentage error (MAPE), and mean bias error (MBE)

The MaxAE is an indicative of local deviations of forecast errors, which is given by

$$MaxAE = \max_{i=1,2,\cdots,N} \left|\hat{p}_i - p_i\right| \tag{3}$$

The MaxAE metric is useful to evaluate the forecasting of short-term extreme events in the power system.

The MAE has been widely used in regression problems and by the renewable energy industry to evaluate forecast performance, which is given by

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{p}_i - p_i\right| \tag{4}$$

The MAE metric is also a global error measure metric, which, unlike the RMSE metric, does not excessively account for extreme forecast events.

The MAPE and MBE are expressed as

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{p}_i - p_i}{capacity}\right| \tag{5}$$

$$MBE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{p}_i - p_i\right) \tag{6}$$

The MBE metric intends to indicate average forecast bias. Understanding the overall forecast bias (over- or under-forecasting) would allow power system operators to better allocate resources for compensating forecast errors in the dispatch process.

4) Kolmogorov–Smirnov test integral (KSI) and OVER metrics

The KSI and OVER metrics were proposed by Espinar et al. [12]. The Kolmogorov–Smirnov (KS) test is a nonparametric test to determine if two data sets are significantly different. The KS statistic D is defined as the maximum value of the absolute difference between two cumulative distribution functions (CDFs), expressed as [12]

$$D = \max\left|F(p_i) - \hat{F}(p_i)\right| \tag{7}$$

where F and $\hat{F}$ represent the CDFs of actual and forecast solar power generation data sets, respectively. The associated null hypothesis is elaborated as follows: if the D statistic characterizing the difference between one distribution and the reference distribution is lower than the threshold value V_c, the two data sets have a very similar distribution and could statistically be the same. The critical value V_c depends on the number of points in the forecast time series, which is calculated for a 99% level of confidence [12].

$$V_c = \frac{1.63}{\sqrt{N}}, \qquad N \geq 35 \tag{8}$$

The difference between the CDFs of actual and forecast power is defined for each interval as [12]

$$D_j = \max\left|F(p_i) - \hat{F}(p_i)\right|, \qquad j = 1,2...,m$$
$$where \quad p_i \in \left[p_{\min} + (j-1)d, p_{\min} + jd\right] \tag{9}$$

Here the value of m is chosen as 100, and the interval distance d is defined as [12]

$$d = \frac{p_{\max} - p_{\min}}{m} \tag{10}$$

where $p_{\max}$ and $p_{\min}$ are the maximum and minimum values of the solar power generation, respectively. The KSI parameter is defined as the integrated difference between the two CDFs, expressed as [12]

$$KSI = \int_{p_{\min}}^{p_{\max}} D_n dp \tag{11}$$

A smaller value of KSI indicates a better performance of solar power forecasting. A zero KSI index means that the CDFs of two sets are equal. A relative value of KSI is calculated by normalizing the KSI value by $a_c = V_c \times (p_{\max} - p_{\min})$ [12].

$$KSIPer(\%) = \frac{KSI}{a_c} \times 100 \tag{12}$$

The OVER metric also characterizes the integrated difference between the CDFs of actual and forecast solar power. The OVER metric considers only the points at which the critical value V_c is exceeded. The OVER metric and its relative value are given by [12]

$$OVER = \int_{p_{\min}}^{p_{\max}} t dp \tag{13}$$

$$OVERPer(\%) = \frac{OVER}{a_c} \times 100 \tag{14}$$

The parameter t is defined by [12]

$$t = \begin{cases} D_j - V_c & if \quad D_j > V_c \\ 0 & if \quad D_j \le V_c \end{cases} \tag{15}$$

As with the KSIPer metric, a smaller value of OVERPer indicates a better performance of the solar power forecasting.

5) Skewness and kurtosis

Skewness is a measure of the asymmetry of the probability distribution, and is the third standardized moment, given by

$$\gamma = E\left[\left(\frac{e - \mu_e}{\sigma_e}\right)^3\right] \tag{16}$$

where γ is the skewness; e is the solar power forecast error, which is equal to the forecast minus the actual solar power value; and μ_e and σ_e are the mean and standard deviation of forecast errors, respectively. Assuming that forecast errors are equal to forecast power minus actual power, a positive skewness of the forecast errors leads to an over-forecasting tail, and a negative skewness leads to an under-forecasting tail. The tendency to over-forecast (or under-forecast) is important in that the system actions taken to correct for under-forecasting and over-forecasting events are not equal. An over-forecasting tendency could lead to a less than optimal number of large thermal units being committed, which need to be corrected through the starting of more expensive, but faster starting, units in the dispatch process [13].

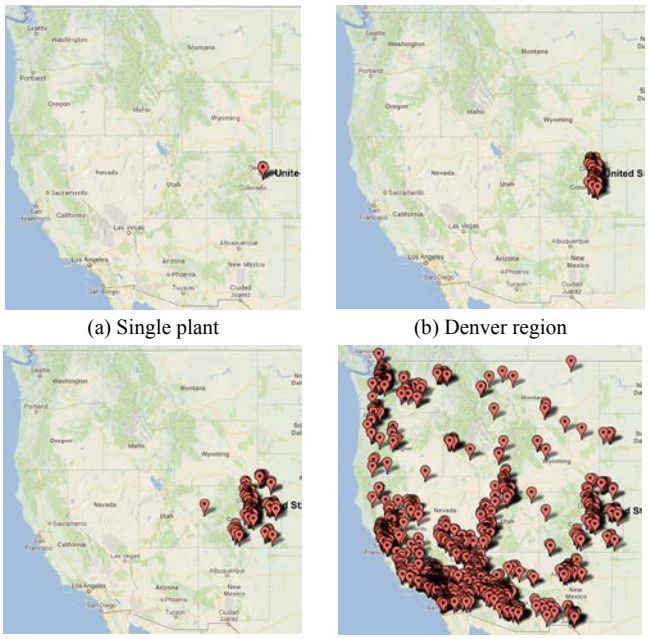
Kurtosis is a measure of the magnitude of the peak of the distribution, or, conversely, how fat-tailed the distribution is, and is the fourth standardized moment, expressed as

$$\kappa = \mu_4 / \sigma_e^{\ 4} - 3 \tag{17}$$

where κ is the kurtosis, μ_4 is the fourth moment about the mean, and σ is the standard deviation of forecast errors. The difference between the kurtosis of a sample distribution and that of the normal distribution is known as the excess kurtosis. In the subsequent analysis, the term *kurtosis* will be treated synonymously with *excess kurtosis*. A distribution with a positive kurtosis value is known as leptokurtic, which indicates a peaked distribution; whereas a negative kurtosis indicates a flat data distribution, known as platykurtic. The pronounced peaks of the leptokurtic distribution represent a large number of very small forecast errors [14].

B. Variability Estimation Metrics

Solar forecasting accuracy is dependent on geographic locations and forecast timescales. In this paper, solar plants at multiple geographic regions were analyzed to quantify the effects of geographic location on the forecasting accuracy. Distributions of forecast errors at multiple temporal and spatial scales were analyzed to investigate the variability in solar forecasting.



(a) Single plant (b) Denver region

(c) State of Colorado region (d) Western Interconnection

Figure 1. Solar plants at different geographic locations.

1) Different geographic locations

The data used in this work is obtained from the *Western Wind and Solar Integration Study Phase 2* (WWSIS-2), one of the world's largest regional integration studies to date [15]. Details of the actual and forecast solar power data are summarized in Section III. Four scenarios are analyzed based on latitude and longitude locations of solar power plants. The

first scenario analyzes the forecast of a single solar power plant with a 100-MW capacity. The second scenario analyzes the forecast of solar plants in the area of Denver, Colorado, including 46 solar plants with an aggregated 3,463-MW capacity. The third scenario investigates solar plants in the region of the state of Colorado, including 90 solar plants with an aggregated 6,088-MW capacity. The fourth scenario analyzes solar plants in the entire Western Interconnection in the United States, including 1,007 solar plants with an aggregated 64,495-MW capacity. Figure 1 shows the locations of solar plants for different scenarios.

2) Distributions of forecast errors

Reserve requirements in a power system to compensate for forecast errors of load, wind and solar power generation is determined based on interval forecasts, which is an important consideration in the commitment and dispatching of generating units. The estimation of forecast confidence intervals is generally calculated using an assumed error distribution on the point forecast. In this paper, the distribution of solar power forecast errors is estimated using the kernel density estimation (KDE) method. KDE is a nonparametric approach to estimate the probability density function of a random variable. KDE has been widely used in the renewable energy community for wind speed distribution characterization [16, 17] and wind and solar power forecasting [13, 18]. For an independent and identically distributed sample, $x_1, x_2, \cdots, x_n$, drawn from some distribution with an unknown density f, KDE is defined as [19]

$$\hat{f}(x;h) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad (18)$$

In the equation, $K(\cdot) = (1/h)K(\cdot/h)$ has a kernel function K (often taken to be a symmetric probability density) and a bandwidth h (a smoothing parameter).

C. Metrics for Uncertainty Quantification and Propagation

Two metrics are proposed to quantify the uncertainty in solar forecasting, which are: (i) standard deviation of solar power forecast errors; and (ii) Rényi entropy of solar power forecast errors.

1) Information entropy of forecast errors

An information entropy approach was proposed in the literature [20, 21] for assessing wind forecasting methods. This information entropy approach based on Rényi entropy is adopted here to quantify the uncertainty in solar forecasting. The Rényi entropy is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha}\log_2 \sum_{i=1}^{n} p_i^\alpha \quad (19)$$

where α is a parameter that allows the creation of a spectrum of Rényi entropies, and p_i is the probability density of the ith discrete section of the distribution. Large values of α favor higher probability events; whereas smaller values of α weight all of the instances more evenly [20]. A larger value of Rényi entropy indicates a high uncertainty in the forecasting.

D. Metrics for Ramps Characterization

One of the biggest concerns associated with integrating a large amount of solar power into the grid is the ability to handle large ramps in solar power output, often caused by cloud events and extreme weather events [22]. Different time and geographic scales influence solar ramps, and they can be either up-ramps or down-ramps, with varying levels of severity. The forecasting of solar power can help reduce the uncertainty involved with the power supply. In this paper, due to its flexibility and simplicity, the swinging door algorithm is applied to ramp identification over varying timeframes [23].

1) Swinging door algorithm signal compression

The swinging door algorithm extracts ramp periods in a series of power signals, by identifying the start and end points of each ramp. The algorithm allows for consideration of a threshold parameter influencing its sensitivity to ramp variations. The only tunable parameter in the algorithm is the width of a "door", represented by ε in Fig. 2. The parameter ε directly characterizes the threshold sensitivity to noise and/or insignificant fluctuations to be specified. A detailed description of the swinging door algorithm can be found in [23].

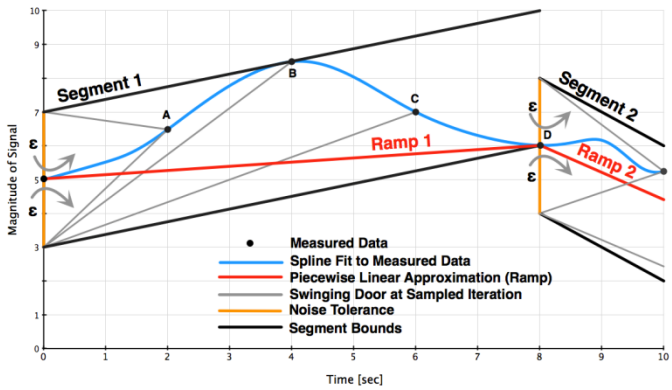


Figure 2. The swinging door algorithm for the extraction of ramps in power from the time series [23].

2) Heat maps

In addition to the ramp periods identified by the swinging door algorithm, heat maps are adopted to illustrate variations of solar power forecast errors. Heat maps allow for power system operators to observe the timing, duration, and magnitude of ramps together.

E. Economic and Reliability Metrics

Flexibility reserves have been proposed as a way to compensate for the variability and short-term uncertainty of solar output. The amount of flexibility reserves necessary is used here to evaluate the economic performance of solar forecasting. Flexibility reserves are the amount of power (in MW) needed to compensate for most hourly or intra-hourly deviations between solar forecasts and actual solar generation values. Improving solar forecasting accuracy is expected to decrease the amount of flexibility reserves that need to be procured with a high penetration of solar power in the system. Flexibility reserves are primarily determined by net load forecast error characteristics [24].

In this paper, instead of using net load forecast errors, the flexibility reserves necessary to compensate solar power

forecast errors are represented by using hourly time steps and 95th confidence intervals of solar power forecast errors.

III. Data Summary

The data used in this work is obtained from the *Western Wind and Solar Integration Study Phase 2* (WWSIS-2), which is one of the world's largest regional integration studies to date [15]. Day-ahead and 1-hour-ahead solar forecast errors are investigated in this study.

F. Day-ahead and 1-hour-ahead Solar Forecasts

The solar data is synthesized based on a 1-minute interval using satellite-derived, 10-km x 10-km gridded, hourly irradiance data. In this paper, the 60-minute solar power plant output for 2006 is used as the actual data. The solar power output data includes distributed generation rooftop photovoltaic, utility-scale photovoltaic, and concentrating solar power with thermal storage.

Day-ahead solar forecasts are taken from the WWSIS-2 solar forecasts conducted by 3TIER based on NWP simulations. The 1-hour-ahead forecasts are synthesized using a 1-hour-ahead persistence of cloudiness approach. In this method, the solar power index is first calculated, which represents the ratio between actual power and clear sky power. Next, the solar forecast power is estimated by modifying the current power output by the expected change in clear sky output.

G. Improved Solar Forecasts

To adequately study the value of improved solar power forecasts, we devised a number of scenarios that allow for the examination of different types of forecast improvements. Three types of forecasting improvements are implemented: (i) uniform improvements at all points in time, (ii) ramp forecasting magnitude improvements, and (iii) ramp forecasting threshold changes. The improvements are performed based on the day-ahead solar power forecast data sets from WWSIS-2. The uniform, ramp forecasting improvements, and threshold changes are generated through the following procedures.

- The uniform forecasting improvement is accomplished by examining the forecast error, then essentially decreasing this error by different percentages.
- The ramp extraction algorithm (swinging door algorithm) performs a piecewise linear approximation to the original signals (day-ahead forecasts). Only ramps that are identified to change greater than or equal to a threshold (e.g., 10%) of the maximum plant capacity are modified in the improved forecasts.
- The ramp forecasting threshold is changed between 10% and 30% of the solar power capacity.

To analyze the sensitivity of the proposed metrics on the three types of solar forecasting improvements, a design of experiments (DoE) methodology is proposed, in conjunction with a response surface development and sensitivity analysis. The Sobol's quasi-random sequence generator is adopted to generate training points for the development of response surfaces. Sobol's sequences [25] use a base of two to form successively finer uniform partitions of the unit interval and reorder the coordinates in each dimension. Figure 3 shows the overall structure of evaluating the sensitivity of the proposed metrics on different types of solar forecasting improvements. The methodologies used for response surface development and sensitivity analysis are discussed in Section IV.

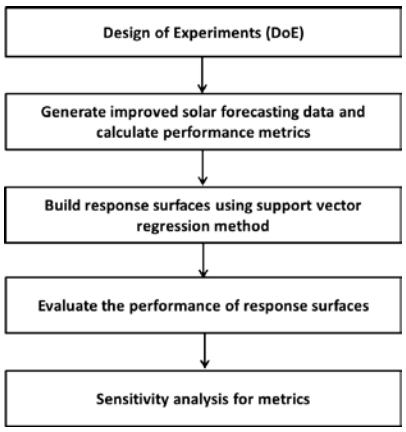


Figure 3. Overall structure of evaluating the sensitivity of metrics on solar forecasting improvements.

IV. Response Surface and Sensitivity Analysis

The response surface methodology is concerned with the construction of approximation models to estimate the system performance, and to develop relationships between specific system inputs and outputs [26]. In this paper, multiple response surfaces are constructed to represent the metric values as functions of the parameters of the three types of solar forecasting improvements. The support vector regression method is adopted for this purpose. The extended Fourier Amplitude Sensitivity Test (eFAST) is adopted for sensitivity analysis.

A. Support Vector Regression (SVR)

Support Vector Regression (SVR) has gained popularity within the statistical learning community, engineering optimization community, and renewable energy community [27] because it provides a unique way to construct smooth nonlinear regression approximations by formulating the response surface construction problem as a quadratic programming problem. SVR can be expressed as [27]

$$\tilde{f}(x) = \langle w, \Phi(x) \rangle + b \tag{20}$$

where $\langle \cdot,\cdot \rangle$ denotes the dot product; w is a set of coefficients to be determined; and $\Phi(x)$ is a map from the input space to the feature space. To solve the coefficients, we can allow a predefined maximum tolerated error ξ (with respect to the actual function value) at each data point, given by [27]

$$\left| \tilde{f}(x_i) - f(x_i) \right| \le \xi \tag{21}$$

where $f(x)$ is the actual function to be approximated. The flatness of the approximated function can be characterized by w. By including slack variables x_i to the constraint and a cost function, the coefficient w can be obtained by solving a quadratic programming problem.

B. Extended Fourier Amplitude Sensitivity Test (eFAST)

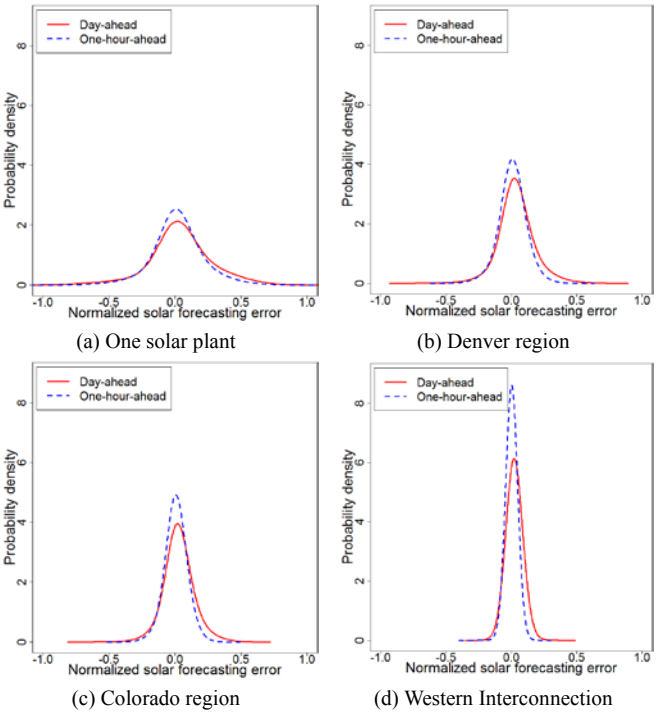
The eFAST algorithm is a variance-based sensitivity analysis method [28]. The sensitivity value is defined based on conditional variances that indicate the individual or joint effects of the uncertain inputs on the output. Two indexes are calculated: (i) the main effect index, which measures the

contribution of one type of solar forecasting improvement alone (e.g., uniform improvements) to the uncertainty (variance) in a metric value (e.g., Rényi entropy); and (ii) the total effect index, which gives the total variance in the metric value caused by the type of solar forecasting improvements (e.g., uniform improvements) and its interactions with any of the other types of forecasting improvements.

V. Results and Discussion

A. Distributions of Solar Power Forecast Errors

Distributions of day-ahead and 1-hour-ahead solar power forecast errors at the four analyzed regions are shown in Fig. 4. It is observed that the 1-hour-ahead forecasting performs better than the day-ahead forecasting. In addition, the distribution of errors at a larger geographic area has a more pronounced peak, slimmer shoulders, and longer tails.



(a) One solar plant (b) Denver region

(c) Colorado region (d) Western Interconnection

Figure 4. Distributions of forecast errors at different geographic locations.

B. Metrics Evaluation for Day-ahead and 1-hour-ahead Solar Forecasting

The values of different metrics to evaluate solar power forecasts at multiple spatial and temporal scales are shown in Table 1. As expected and inferable from the metrics of correlation coefficient, NRMSE, MaxAE, MAE, MAPE, KSIPer, OVERPer, and 95th percentile, 1-hour-ahead forecasting performs better than day-ahead forecasting. This matches the observation from the forecast error distributions in Fig. 4. The positive MBE metrics indicate an over-forecast characteristic for both day-ahead and 1-hour-ahead forecasting. Overall, all the metrics can successfully evaluate the expected performance of solar forecasting.

Table 2 shows the uncertainty metric of Renyi entropy for both day-ahead and 1-hour-ahead forecasting at the four geographic regions. Five cases are analyzed for each scenario based on forecasting time periods: (i) forecasting throughout a whole year, (ii) forecasting in January, (iii) forecasting in July, (iv) forecasting at the time of 14:00 each day, and (v) forecasting at the peak time, 10:00 to 16:00 each day. We observe that the length of forecasting period affects the uncertainty in the forecasting. The uncertainty in the forecasting of the whole year data is less than that in any of the other cases (January, July, 14:00, and 10:00 to 16:00).

C. Ramp Forecasting Results

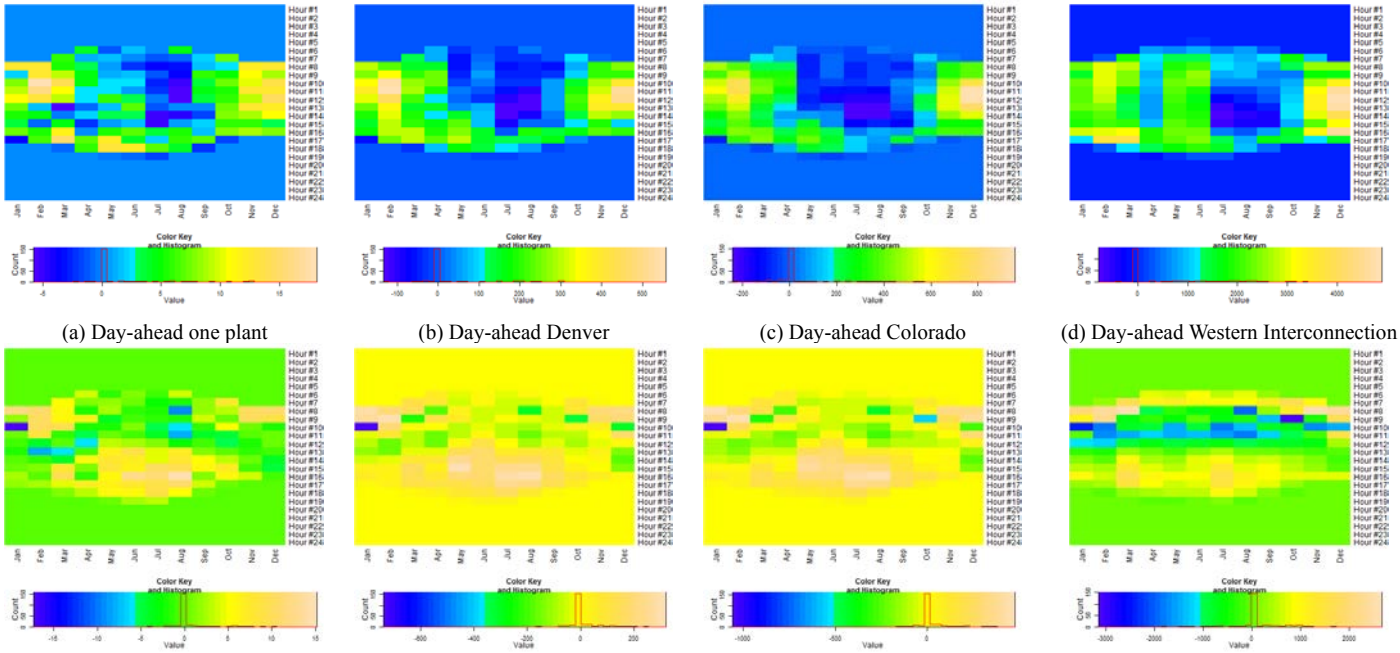
Figure 5 shows heat maps of the mean hourly day-ahead and 1-hour-ahead solar forecast errors per month at the four different geographic regions. The observations from heat maps of the regions of Denver and Colorado are similar, because the solar power capacity in the region of Denver accounts for approximately 57% of that in the region of Colorado. For the 1-hour-ahead solar forecasting, more over-forecasting events are observed than under-forecasting events for all four regions. Because reliability of a power system is more impacted by short-term forecasting, this 1-hour-ahead over-forecasting trend could be corrected through a statistical model to ensure the reliability of the power system. For day-ahead forecasting, we observe more under-forecasting events during the summer season (July and August), which significantly influences economic operations of the power system.

TABLE 1. Metrics Values Estimated by Using an Entire Year of Data

Metrics	One Plant		Denver		Colorado		Western Interconnection	
	Day-ahead	1-hour-ahead	Day-ahead	1-hour-ahead	Day-ahead	1-hour-ahead	Day-ahead	1-hour-ahead
Correlation coefficient	0.65	0.76	0.87	0.94	0.91	0.96	0.990	0.995
RMSE (MW)	22.07	17.12	438.25	284.36	624.19	378.65	2711.31	1488.28
NRMSE	0.22	0.17	0.13	0.08	0.10	0.06	0.04	0.02
MaxAE (MW)	84.10	74.33	2260.94	1304.73	3380.28	1735.24	17977.53	16127.32
MAE (MW)	14.81	11.34	286.65	191.17	413.11	256.69	1973.90	1064.52
MAPE	0.15	0.11	0.08	0.06	0.07	0.04	0.03	0.02
MBE (MW)	4.27	2.19	131.82	31.64	172.54	43.32	1497.29	132.13
KSIPer (%)	216.73	104.42	184.30	52.84	143.38	48.28	132.92	47.76
OVERPer (%)	136.36	28.16	94.43	0.77	54.65	0.37	41.43	0.00
Standard dev. (MW)	21.65	39.57	418.00	282.62	599.94	376.20	2260.09	1482.44
Skewness	-0.19	0.08	0.20	-0.20	0.18	-0.21	0.62	-0.23
Kurtosis	2.04	2.40	3.79	2.52	3.35	2.47	3.76	4.82
95th percentile (MW)	50.59	39.57	990.66	637.45	1394.85	838.27	5652.60	3079.32
Capacity (MW)	100.00	100.00	3463.00	3463.00	6088.00	6088.00	64495.00	64495.00

TABLE 2. The Uncertainty Metric of Rényi Entropy at Multiple Spatial and Temporal Scales

Cases	One solar plant		Denver region		Colorado region		Western Interconnection	
	Day-ahead	1-hour-ahead	Day-ahead	1-hour-ahead	Day-ahead	1-hour-ahead	Day-ahead	1-hour-ahead
Year	4.83	4.64	4.24	4.63	4.33	4.73	4.47	4.01
January	4.71	5.06	5.18	5.06	5.46	4.79	5.24	5.11
July	4.64	4.74	4.25	4.87	5.02	5.09	4.75	4.86
14:00	5.07	5.00	4.83	4.99	5.13	5.27	4.97	5.72
10:00 -16:00	4.95	4.73	4.60	4.79	4.82	4.94	4.90	4.45



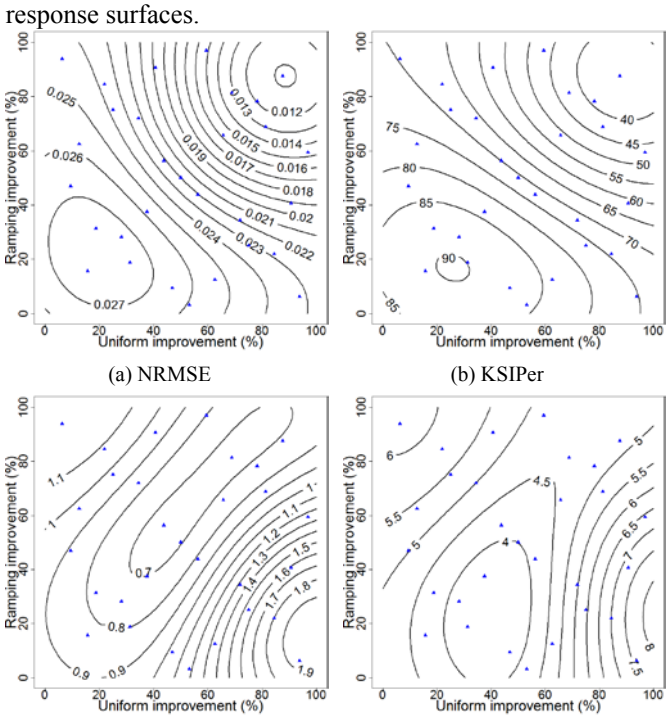
(a) Day-ahead one plant (b) Day-ahead Denver (c) Day-ahead Colorado (d) Day-ahead Western Interconnection

(e) One-hour-ahead one plant (f) One-hour-ahead Denver (g) One-hour-ahead Colorado (h) One-hour-ahead Western Interconnection

Figure 5. Heat maps of the mean solar forecast errors for each hour of the day per month of the year.

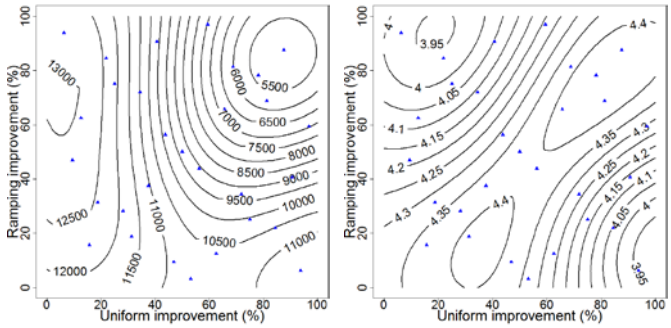
D. *Response Surfaces of Metrics*

A response surface is constructed to represent the metric value as a function of the parameters of the three types of solar forecasting improvements. The assumed forecasting improvement capability is defined as: (i) 0% to 100% for uniform improvements, (ii) 0% to 100% for ramp forecasting improvements, and (iii) 10% to 30% of solar power capacity for ramp forecasting threshold. Response surfaces of six typical metrics are shown in Fig. 6; the response surfaces are constructed based on the solar power data in the region of the Western Interconnection. A constant value of the ramp forecasting threshold is used in the plots, which is 10% of the solar power capacity. The triangle points in the figures represent the 30 training points for the development of response surfaces.



(a) NRMSE (b) KSIPer

(c) Skewness (d) Kurtosis



(e) MaxAE (f) Rényi entropy

Figure 6. Response surfaces of metrics constructed by SVR.

Figure 6(a) shows that the NRMSE value decreases with both uniform forecasting improvements and ramp forecasting improvements; a similar trend is also observed from the KSIPer metric in Fig. 6(b). Figures 6(c) through 6(f) show that the response surfaces of skewness, kurtosis, MaxAE, and Rényi entropy are highly nonlinear. For the skewness metric shown in Fig. 6(c), it is observed that: (i) a consistent positive skewness is obtained through uniform and ramp improvements, leading to an over-forecasting tail; and (ii) the minimum skewness is observed with approximately 50% uniform forecasting improvements and 60% ramp forecasting improvements. In Fig. 6(f), high uncertainty is observed in two regions; whereas low uncertainty is obtained at the top-left and bottom-right corners, resulting mainly from one type of improvement (ramp or uniform forecasting improvements).

E. *Sensitivity Analysis*

The main effects and total effects of the proposed metrics to the three types of forecasting improvements are listed in Table 3. Most metrics are highly sensitive to the uniform improvement (compared to ramp forecasting improvements and ramp threshold changes), indicating that these metrics could efficiently recognize a better solar forecasting with uniform improvements. In addition, the metrics of skewness, kurtosis, and Rényi entropy are observed to be sensitive to all

three types of forecasting improvements. These three metrics (skewness, kurtosis, and Rényi entropy) could be adopted for evaluating solar forecasts with ramp forecasting improvements and ramp threshold changes that are more important in the power system.

TABLE 3. SENSITIVITY ANALYSIS OF METRICS TO THE THREE TYPES OF FORECASTING IMPROVEMENTS

Metrics	Uniform Improvement		Ramp Improvement		Ramp Threshold	
	Main effect	Total effect	Main effect	Total effect	Main effect	Total effect
R^2	0.913	0.974	0.009	0.043	0.001	0.065
RMSE	0.883	0.958	0.016	0.060	0.000	0.088
NRMSE	0.883	0.958	0.016	0.060	0.000	0.088
MaxAE	0.899	0.977	0.011	0.063	0.002	0.057
MAE	0.896	0.955	0.016	0.056	0.003	0.077
MAPE	0.896	0.955	0.016	0.056	0.003	0.077
MBE	0.844	0.923	0.032	0.090	0.002	0.112
KSIPer	0.843	0.921	0.033	0.092	0.003	0.111
OVERPer	0.872	0.942	0.014	0.069	0.016	0.097
Standard dev.	0.897	0.969	0.012	0.052	0.001	0.075
Skewness	**0.305**	**0.686**	**0.126**	**0.473**	**0.161**	**0.343**
Kurtosis	**0.394**	**0.858**	**0.041**	**0.169**	**0.081**	**0.207**
95th percentile	0.886	0.965	0.013	0.060	0.002	0.082
Rényi entropy	**0.120**	**0.624**	**0.071**	**0.599**	**0.260**	**0.463**

VI. CONCLUSION

This paper discussed a suite of metrics for evaluating the performance of solar power forecasting. The performance of the proposed metrics was evaluated using the actual and forecast solar power data from the Western Wind and Solar Integration Study. The distribution of forecast errors indicates that relative forecast errors are smaller for a large geographic area. The results showed that the developed metrics can successfully evaluate the quality of a solar forecast. More over- than under-forecasting events were observed for the 1-hour-ahead solar forecasting; whereas the day-ahead forecasting tended to underforecast in summer.

To analyze the sensitivity of the proposed metrics to improved solar forecasts, a sensitivity analysis methodology was developed based on a design of experiments and response surfaces. The results showed that all proposed metrics were sensitive to solar forecasts with uniform forecasting improvements; and the metrics of skewness, kurtosis and Rényi entropy were also sensitive to solar forecasts with ramp forecasting improvements and ramp forecasting threshold.

Future work will (i) determine baseline values and achievable target values for the metrics for independent system operators and utilities and (ii) develop a suite of probabilistic solar forecasting metrics.

ACKNOWLEDGMENT

This work was supported by the U.S. Department of Energy under Contract No. DE-FOA-0000649 with the National Renewable Energy Laboratory.

REFERENCES

[1] R. Pernick and C. Wilder, Utility Solar Assessment Study Reaching Ten Percent Solar by 2025. Technical Report. Washington, D.C.: Clean Edge, Inc., and Co-Op America Foundation, 2008.

[2] D., Lew et al., 2013, "The Western Wind and Solar Integration Study Phase 2," Technical Report No. NREL/TP-5500-55888, National Renewable Energy Laboratory, Golden, CO.

[3] R. Marquez and C. F. M. Coimbra, "Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the NWS database," Solar Energy, vol. 85, no. 5, pp. 746–756, 2011.

[4] P. Mathiesen and J. Kleissl, "Evaluation of numerical weather prediction for intra-day solar forecasting in the continental united states," Solar Energy, vol. 85, no. 5, pp. 967–977, 2011.

[5] C. Chen, S. Duan, T. Cai, and B. Liu, "Online 24-h solar power forecasting based on weather type classification using artificial neural network," Solar Energy, vol. 85, no. 11, pp. 2856–2870, 2011.

[6] R. Perez, K. Moore, S. Wilcox, D. Renné, and A. Zelenka, "Forecasting solar radiation—Preliminary evaluation of an approach based upon the National Forecast Database," Solar Energy, vol. 81, no. 6, pp. 809–812, 2007.

[7] E. M. Crispim, P. M. Ferreira, and A. E. Ruano, "Prediction of the solar radiation evolution using computational intelligence techniques and cloudiness indices, Int. J. of Innovative Computing, Information and Control, vol. 4, no. 5, pp. 1121–1133, 2008.

[8] W. C. Chow et al., "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy test bed," Solar Energy, vol. 85, no. 11, pp. 2881–2893, 2011.

[9] R. Marquez and C. F. M. Coimbra, "Intra-hour DNI forecasting based on cloud tracking image analysis," Solar Energy, 2012, http://dx.doi.org/10.1016 /j.solener.2012.09.018.

[10] E. Lorenz, D. Heinemann, H. Wickramarathne, H. G. Beyer, and S. Bofinger, "Forecast of ensemble power production by grid-connected PV systems," Proc. 20th European PV Conference, Milano, Italy, 2007.

[11] R. Marquez and C. F. M. Coimbra, "Proposed Metric for Evaluation of Solar Forecasting Models," J. Solar Energy Eng., vol. 135, pp. 011016-1, 2013.

[12] B. Espinar et al., "Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations," Solar Energy, vol. 83, no. 1, pp. 118–125, 2009.

[13] J. Zhang, B.-M. Hodge, and A. Florita, "Investigating the correlation between wind and solar power forecast errors in the Western Interconnection." In ASME 7th Int. Conf. on Energy Sustainability, Minneapolis, MN, 2013.

[14] B.-M. Hodge and M. Milligan, "Wind power forecasting error distributions over multiple timescales," IEEE PES GM Proc., San Diego, CA, pp. 1–8, 2011.

[15] D. Lew et al., The Western Wind and Solar Integration Study Phase 2. NREL/TP-5500-55888. Golden, CO: National Renewable Energy Laboratory, 2013.

[16] J. Zhang, S. Chowdhury, A. Messac, and L. Castillo, 2013, "A multivariate and multimodal wind distribution model," Renewable Energy, vol. 51, pp. 436–447, 2013.

[17] J. Zhang, S. Chowdhury, A. Messac, and L. Castillo, "Multivariate and multimodal wind distribution model based on kernel density estimation." In ASME 5th Int. Conf. on Energy Sustainability, Washington, DC, 2011.

[18] J. Juban, N. Siebert, and G. N. Kariniotakis, "Probabilistic short-term wind power forecasting for the optimal management of wind generation," IEEE Power Eng. Society, Lausanne Power Tech Conf. Proc., pp. 683–688, 2007.

[19] M. Jones, J. Marron, and S. Sheather S., "A brief survey of bandwidth selection for density estimation," J. American Statistical Assoc., vol. 91, no. 433, pp. 401–407, 1996.

[20] B.-M. Hodge, K. Orwig, and M. Milligan, "Examining information entropy approaches as wind power forecasting performance metrics," 12th Int. Conf. on Probabilistic Methods Applied to Power Systems, Istanbul, Turkey, 2012.

[21] R. J. Bessa, V. Miranda, A. Botterud, and J. Wang, "'Good' or 'bad' wind power forecasts: A relative concept," Wind Energy, vol. 14, pp. 625–636, 2011.

[22] A. Mills and R. Wiser, Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power. LBNL-3884E. Berkeley, CA: Lawrence Berkeley National Laboratory, Environmental Energy Technologies Division: September 2010.

[23] A. Florita, B.-M. Hodge, and K. Orwig, "Identifying wind and solar ramping events," IEEE 5th Green Technologies Conf., Denver, Colorado, 2013.

[24] K. Orwig, B.-M. Hodge, G. Brinkman, E. Ela, M. Milligan, V. Banunarayanan, S. Nasir, and J. Freedman, "Economic evaluation of short-term wind power forecasts in ERCOT: Preliminary results," 11th Annual Int. Workshop on Large-Scale Integration of Wind Power into Power Systems as Well as on Transmission Networks for Offshore Wind Power Plants Conf., Lisbon, Portugal, 2012.

[25] I. M. Sobol, "Uniformly distributed sequences with an additional uniform property," USSR Computational Mathematics and Mathematical Physics, vol. 16, no. 5, pp. 236–242, 1976.

[26] J. Zhang, S. Chowdhury, and A. Messac, "An adaptive hybrid surrogate model," Structural and Multidisciplinary Optimization, vol. 46, no. 2, pp. 223–238, 2012.

[27] J. Zhang, S. Chowdhury, A. Messac, and B.-M. Hodge, "Assessing long-term wind conditions by combining different measure-correlate-predict algorithms," ASME International Design Engineering Technical Conferences, Portland, Oregon, 2013.

[28] A. Saltelli and R. Bolado, "An alternative way to compute Fourier Amplitude Sensitivity Test (FAST)," Computational Statistics & Data Analysis, vol. 26, no. 4, pp. 445–460, 1998.