A Methodology for Quantifying Reliability Benefits From Improved Solar Power Forecasting in Multi-Timescale Power System Operations

Mingjian Cui, *Member, IEEE*, Jie Zhang, *Senior Member, IEEE*, Bri-Mathias Hodge, *Senior Member, IEEE*, Siyuan Lu, and Hendrik F. Hamann, *Member, IEEE*

***Abstract*—Solar power forecasting improvements are mainly evaluated by statistical and economic metrics, and the practical reliability benefits of these forecasting enhancements have not yet been well quantified. This paper aims to quantify reliability benefits from solar power forecasting improvements. To systematically analyze the relationship between solar power forecasting improvements and reliability performance in power system operations, an expected synthetic reliability (ESR) metric is proposed to integrate multiple state-of-the-art independent reliability metrics. The absolute value and standard deviation of area control errors (ACEs), and the North American Electric Reliability Corporation Control Performance Standard 2 (CPS2) score are calculated through a multi-timescale scheduling simulation, including the day-ahead unit commitment, real-time unit commitment, real-time economic dispatch, and automatic generation control sub-models. The absolute ACE in energy, CPS2 violations, CPS2 score, and standard deviation of the raw ACE are all calculated and combined as the ESR metric. Numerical simulations show that the reliability benefits of multi-timescale power system operations are significantly increased due to the improved solar power forecasts.**

***Index Terms*—Area control error, multi-timescale power system operation, photovoltaic, reliability benefit, forecast.**

NOMENCLATURE

Acronyms (Alphabetically)

AACEE	Absolute area control error in energy.
ACE	Area control error.
AGC	Automatic generation control.
BSF	Baseline solar forecasting.
CPS2	Control Performance Standard 2.
DA, HA	Day-ahead, hour-ahead.
DU	Day-ahead security-costrained unit commitment.
ESR	Expected synthetic reliability.
FESTIV	Flexible Energy Scheduling Tool for Integration of Variable Generation.
MILP	Mixed-integer linear programming.
RE	Real-time security-constrained economic dispatch.
RU	Real-time security-constrained unit commitment.
TSF	Target solar forecasting.

Indices/Sets (Alphabetically)

b, NB	Index and set for buses, $b = 1, 2, \ldots, NB$.
i, NI	Index and set for thermal units, $i = 1, 2, \ldots, NI$.
r, NR	Index and set for reserve types. $r = 1$, spinning reserve; $r = 2$, non-spinning reserve; $r = 3$, regulation reserve; and $r = 4$, replacement reserve.
$j, \mathscr{J}$	Index for forecasting improvement percentage and the set of combination $(x_j\%, y_j\%)$.
s, NS	Index and set for solar units, $b = 1, 2, \ldots, NB$.
$t^{(\cdot)}$	Index for time periods. $t^{\mathrm{DU}} = 1, 2, \ldots, T^{\mathrm{DU}}$ in DU model; $t^{\mathrm{RU}} = 1, 2, \ldots, T^{\mathrm{RU}}$ in RU model, $t^{\mathrm{RU}} \in t^{\mathrm{DU}}$; $t^{\mathrm{RE}} = 1, 2, \ldots, T^{\mathrm{RE}}$ in RE model, $t^{\mathrm{RE}} \in t^{\mathrm{RU}}$; and $t^{\mathrm{AGC}} = 1, 2, \ldots, T^{\mathrm{AGC}}$ in AGC model, $t^{\mathrm{AGC}} \in t^{\mathrm{RE}}$.

Parameters (Alphabetically)

$d_b^{t^{(\cdot)}}$	Expected load of bus b at time $t^{(\cdot)}$, in MW.
$e_{\mathrm{B}}^t, e_{\mathrm{T},j}^t$	Forecasting error of BSF and TSF with the jth forecasting improvement at time t.
$p_{\mathrm{act}}^t, p_{\mathrm{B}}^t, p_{\mathrm{CS}}^t$	Actual, baseline, clear-sky solar power at time t.
$p_s^{t^{(\cdot)}}$	Solar power of unit s at time $t^{(\cdot)}$, in MW.
t_{AGC}	Interval in AGC model.
ACE_{CPS2}	Sum of instantaneous ACE until the 10-minute CPS2 interval ends, in MW-10min.
$C_i^{t^{(\cdot)}}$	Operation cost of thermal unit i at time $t^{(\cdot)}$.
$E_{\mathrm{B}}^{\mathscr{D}}, E_{\mathrm{T},j}^{\mathscr{D}}$	Day-ahead forecast error time series of BSF and the jth possible TSF.

Manuscript received February 17, 2017; revised June 5, 2017; accepted July 14, 2017. Date of publication July 18, 2017; date of current version October 19, 2018. This work was supported by the National Renewable Energy Laboratory (under the U.S. Department of Energy Prime under Contract DE-AC36-08GO28308) under Contract XHQ-6-62546-01. Paper no. TSG-00238-2017. *(Corresponding author: Jie Zhang.)*

M. Cui and J. Zhang are with the Department of Mechanical Engineering, The University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: mingjian.cui@utdallas.edu; jiezhang@utdallas.edu).

B.-M. Hodge is with the National Renewable Energy Laboratory, Golden, CO 80401 USA (e-mail: bri.mathias.hodge@nrel.gov).

S. Lu and H. F. Hamann are with IBM T.J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: lus@us.ibm.com; hendrikh@us.ibm.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2017.2728480

$E_{\mathrm{B}}^{\mathscr{H}}$, $E_{\mathrm{T},j}^{\mathscr{H}}$ Hour-ahead forecast error time series of BSF and the jth possible TSF.
$ENS_{t^{(\cdot)}}$ Energy not served at time $t^{(\cdot)}$, in MWh.
$\mathscr{H}_{\mathrm{DU}}$ Scheduling horizon of the DU model.
$IR_{t^{(\cdot)}}^{r}$ Insufficient reserve r at time $t^{(\cdot)}$, in MWh.
L_C A sufficient large constant.
$LS^{t^{\mathrm{AGC}}}$ Losses on the transmission lines at time t^{AGC}.
P_i^{max}, P_i^{min} Maximum/minimum capacity of unit i, in MW.
$RR_{t^{(\cdot)}}^{r}$ Reserve r requirements of thermal units at time $t^{(\cdot)}$, in MW.
R_i^{up}, R_i^{dn} Maximum up/down ramping rate of thermal unit i, in MW/min.
$S_i^{t^{(\cdot)}}$ Start-up cost of thermal unit i at time $t^{(\cdot)}$, in \$.
SPI^t Solar power index at time t.
$T^{(\cdot)}$ Number of time periods. $T^{\mathrm{DU}} = 24$ with 1-hour time resolution in DU model; $T^{\mathrm{RU}} = 4$ with 15-minute time resolution in RU model; and $T^{\mathrm{RE}} = 3$ with 5-minute time resolution in RE model.
$T_{i,\mathrm{on}}$, $T_{i,\mathrm{off}}$ Minimum ON/OFF time limits of unit i.
T_{CPS2} Time interval of CPS2, 10 minutes.
$VOIR^r$ Value of insufficient reserve r (\$/MWh).
$VOLL$ Value of loss load (\$/MWh).
$X_{i,\mathrm{on}}^{t^{(\cdot)}}$, $X_{i,\mathrm{off}}^{t^{(\cdot)}}$ ON/OFF time of thermal unit i at time $t^{(\cdot)}$.
ξ, $\overline{\xi}$ Sets of ramping and non-ramping time intervals.
γ_{sp}, γ_{ns} Bidding price of spinning and non-spinning reserves of thermal generators, in \$/MWh.
$\gamma_{i,t^{(\cdot)}}^{r}$ Bidding price of reserve r of thermal unit i at time $t^{(\cdot)}$, in \$/MWh.
$\mathbf{D}$ Vector of expected load or demand.
$\mathbf{K_P}$, $\mathbf{K_S}$, $\mathbf{K_D}$ Bus-thermal unit, bus-solar unit, and bus-load incidence matrices.
$\mathbf{P}$, $\mathbf{P_S}$ Vector of thermal dispatch and PV generation.
$\mathbf{PL}^{\max}$ Vector of power limit for transmission lines.
$\mathbf{SF}$ Shift factor matrix.

Variables (Alphabetically)

$p_i^{t^{(\cdot)}}$ Dispatch of thermal unit i at time $t^{(\cdot)}$, in MW.
$u_i^{t^{(\cdot)}}$ 1, unit i scheduled at time $t^{(\cdot)}$; and 0, otherwise.
$x_j\%$, $y_j\%$ jth forecasting improvement percentage by which the forecast errors of no-ramping and ramping time series are uniformly decreased.
$R_{i,t^{(\cdot)}}^{r}$ Schedule of reserve r of thermal unit i at time $t^{(\cdot)}$, in MW.

I. Introduction

STEADILY rising solar power penetrations in power system operations are now catching the attention of independent system operators (ISOs). The variable and uncertain characteristics of solar power can have profound impacts on the economics and reliability of power system operations [1].

Currently, there are various solar power forecasting methods proposed to improve forecasting accuracy. These methods can be broadly divided into two categories. The first are physical methods that model the physics of the atmosphere to obtain accurate irradiance forecasts [2]. The second are statistical methods that identify relationships based on historical data to predict the future behavior of irradiance or power. These are considered to be data-driven approaches and are commonly divided into regressive methods and machine learning methods. Moreover, the forecasting methods can also be classified into direct and indirect forecasting models, i.e., directly calculating the solar power generation using the physical or statistical method, or first predicting solar irradiation and then obtaining the solar power generation using an irradiance-to-power model.

To quantify the benefits of various solar power forecasting methods, there are different types of metrics that are generally composed of statistical, economic, and reliability metrics. Statistical metrics are the most widely used metrics [3]. Economic metrics have recently been used to assess the benefits of the improved solar forecasts. Zhang *et al*. [4] suggested that improving solar forecasting accuracy could decrease the amount of additional operating reserves. Martinez-Anido *et al*. [5] found that solar power forecasting improvements could reduce the annual operational electricity costs with annual economic values. However, few studies considered the power system reliability benefits of solar power forecasting improvements. Though Lew *et al*. [6] pointed out that solar power forecasting improvements could affect the reliable balance between supply and demand, it is still challenging to quantify these reliability benefits. This is because the reliability standards, defined by the North American Electric Reliability Corporation (NERC) [7], are estimated via the area control error (ACE), which is very strongly influenced by the performance of the automatic generation control (AGC). Yet few optimization models (i.e., unit commitment or economic dispatch) consider the impacts of AGC. Therefore, the reliability benefits from solar power forecasting improvements have not been well studied in the literature.

The contribution of this work is to propose a methodology for quantifying reliability benefits from improved solar power forecasts. Though a number of papers have focused on improving the accuracy of solar power forecasts, the impacts of the improved solar power forecasts on power system reliability are still not well studied. Based on state-of-the-art solar power forecasts and a multi-timescale scheduling model, this paper aims to bridge the gap by measuring and analyzing the reliability benefits from solar power forecasting improvements, through a large number of scenarios that cover multiple solar forecasting methods at different locations, time ahead forecasts, and solar penetration levels. Though several individual reliability metrics have been used in the literature, the trade-off of different reliability metrics is still not well understood. A multi-objective optimization could better explore the trade-off of these reliability metrics (or together with economic metrics). An alternative is to integrate these metrics and utilize a single metric to represent the reliability performance of

power system operations. The weights of individual reliability metrics could be adaptively chosen by considering the status and risk preferences of the particular balancing authority. In this paper, an expected synthetic reliability (ESR) metric is proposed to integrate individual reliability metrics that are widely used in the literature, including the absolute area control error in energy (AACEE), control performance standard 2 (CPS2) violations, CPS2 score, and standard deviation of the raw ACE.

The organization of this paper is as follows. In Section II, three solar power forecasting improvements are briefly introduced. In Section III, a multi-timescale power system operation model and reliability metrics are presented. The case studies and results performed on the IEEE 118-bus system are described in Section IV. Concluding remarks and future work are provided in Section V.

II. Solar Power Forecasting Improvements: Baseline, Target, and Watt-Sun

Most existing renewable integration studies that attempt to gauge the impact of forecasts assume a uniform forecasting improvement using a percentage-based methodology (e.g., 25%, 50%, 75%, and 100% [8], [9]). However, these assumptions are simplifications that ignore the fact that improvements may also have large reliability impacts at certain specific times in power system operations. In practice, forecasting improvements (no matter the method) do not tend to create uniform improvements, but rather improvements during certain situations. In this paper, the actual measured solar power data is used as the forecast in the perfect case. Three practical forecasting methodologies are applied to obtain solar power forecasts for analyzing the possible enhancement of reliability performance. These methodologies consist of baseline, target, and Watt-Sun [10], [11] solar power forecasts that are briefly described as follows.

As part of the project work performed under the SunShot Initiative's Improving the Accuracy of Solar Forecasting program, a system for improving solar forecasting, namely Watt-Sun, has been developed by this team [10], [11]. To truly measure the improvements that any new solar forecasting methods provide, the baseline and target solar forecasts at different spatial and temporal scales were also developed in close collaboration with utility and independent system operator partners. Baseline and target solar forecasts mainly focus on quantifying the difference between ramping and non-ramping error periods [12].

A. Baseline Solar Forecasting (BSF)

1) Persistence of Cloudiness Approach for 1-Hour-Ahead (1HA) and 4-Hour-Ahead (4HA) BSF: For 1HA and 4HA forecasts, a persistence of cloudiness approach is adopted due to its superiority in the shorter forecasting period. Solar power is predicted by the power output in clear-sky and the current measured solar power generation, as shown in Fig. 1.

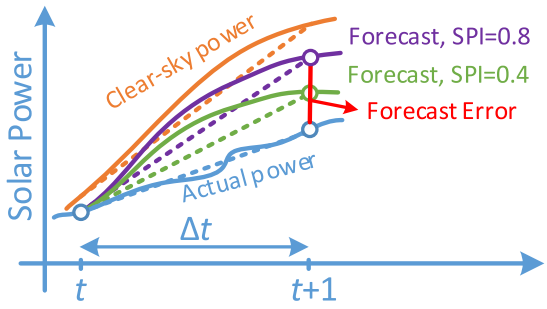


Fig. 1. Persistence of cloudiness approach for BSF [6].

The expected power change in clear-sky is calculated by multiplying the current solar power index (SPI^t) and adding the estimated forecast error, which is the power difference between time t and time $t+1$ [6]. Thus, the forecasted solar power can be formulated as:

$$p_{\mathrm{B}}^{t+1} = p_{\mathrm{act}}^{t} + \mathrm{SPI}^t \times \left(p_{\mathrm{CS}}^{t+1} - p_{\mathrm{CS}}^{t}\right). \tag{1}$$

2) 1-Day-Ahead (1DA) and 2-Day-Ahead (2DA) Numerical Weather Prediction (NWP) Model Approach: For 1DA and 2DA forecasts, a model approach is combined with a streamer radiative transfer model (RTM) and an irradiance-to-power model. This method is based on a state-of-the-art NWP model, specifically the North American Mesoscale Forecast System (NAM) [13]. The global horizontal irradiance (GHI) and the direct normal irradiance (DNI) are first calculated by the streamer RTM. Then the California Energy Commission (CEC) model implemented in PV-Lib [14] is used to convert GHI and DNI to AC solar power.

B. Target Solar Forecasting (TSF)

Based on the aforementioned BSF values, the TSF is derived by the non-ramping and ramping periods in BSF. These ramping periods are detected by the swinging door algorithm [15], [16]. For the non-ramping period, uniform forecasting improvements by $x\%$ are applied to the BSF values. For the ramping period, forecasting improvements by $y\%$ are applied to the BSF values. The objective of determining TSF is to reduce reserves to a level that is both feasible and economically impactful based on the consensus of balancing authorities. A spinning/non-spinning reserve-based model is used to calculate TSF values. The forecasting error of BSF at time t is defined as e_{B}^{t} and formulated as:

$$e_{\mathrm{B}}^{t} = p_{\mathrm{B}}^{t} - p_{\mathrm{act}}^{t} \tag{2}$$

Then, the forecasting error of TSF at time t is defined as $e_{\mathrm{T},j}^{t}$ and formulated as:

$$e_{\mathrm{T},j}^{t} = \underbrace{\left(p_{\mathrm{B}}^{t} - p_{\mathrm{act}}^{t}\right)_{t\in\bar{\xi}}}_{\text{Non-ramping}} \times x_j\% + \underbrace{\left(p_{\mathrm{B}}^{t} - p_{\mathrm{act}}^{t}\right)_{t\in\xi}}_{\text{Ramping}} \times y_j\% \tag{3}$$

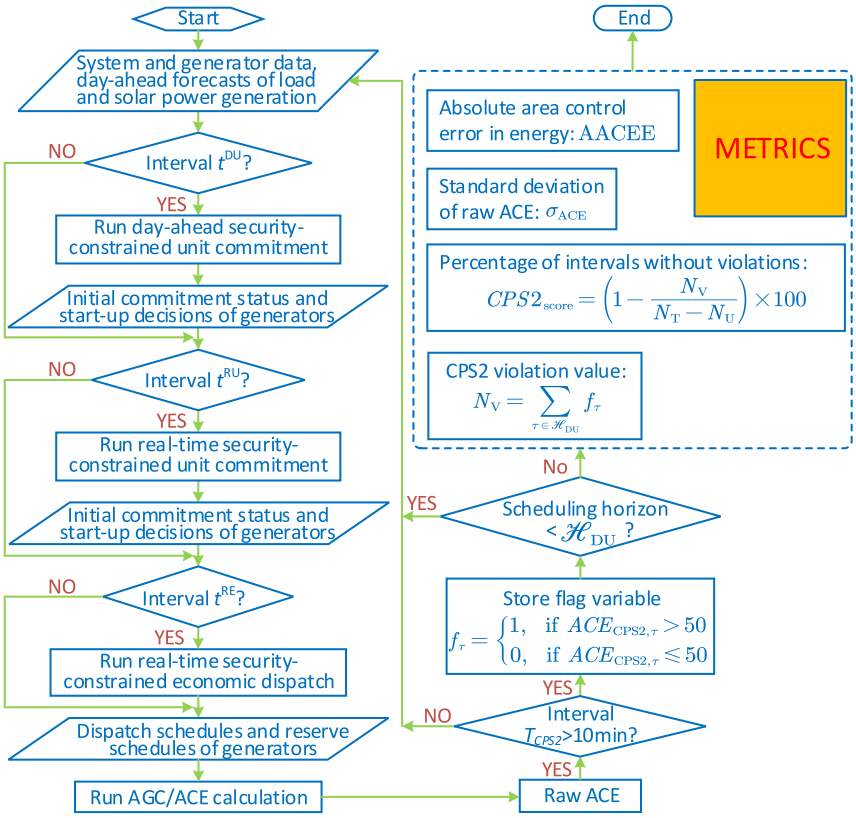


Fig. 2. Schematic for calculating reliability metrics by using multi-timescale scheduling models.

Based on (2) and (3), the objective function for 1HA and 4HA solar power forecasts is formulated as:

$$\min_{e^t_{\mathrm{T},j} \in E^{\mathscr{H}}_{\mathrm{T},j}, e^t_{\mathrm{B}} \in E^{\mathscr{H}}_{\mathrm{B}}, j \in \mathscr{J}} \underbrace{\left| \theta_{95\%}\left(e^t_{\mathrm{T},j}\right) - \rho\theta_{95\%}\left(e^t_{\mathrm{B}}\right) \right| \gamma_{\mathrm{sp}}}_{\text{Spinning Reserve}} \tag{4}$$

As the total number of forecasting improvement percentage pairs $(x\%, y\%)$ approaches infinity, the error would approach zero. The combination $(x_j\%, y_j\%)$ is generated by a widely used method in design of experiments, i.e., Sobol's quasi-random sequence generator [17]. For 1DA and 2DA solar power forecasting, the objective function is formulated as:

$$\begin{aligned} \min_{e^t_{\mathrm{T},j} \in E^{\mathscr{D}}_{\mathrm{T},j}, e^t_{\mathrm{B}} \in E^{\mathscr{D}}_{\mathrm{B}}, j \in \mathscr{J}} \Bigg| & \underbrace{\left[\theta_{70\%}\left(e^t_{\mathrm{T},j}\right) - \rho\theta_{70\%}\left(e^t_{\mathrm{B}}\right)\right] \gamma_{\mathrm{sp}}}_{\text{Spinning Reserve}} \\ & + \underbrace{\left[\theta_{95\%}\left(e^t_{\mathrm{T},j}\right) - \theta_{70\%}\left(e^t_{\mathrm{T},j}\right)\right] \gamma_{\mathrm{ns}}}_{\text{Non-spinning Reserve of TSF}} \\ & - \rho \underbrace{\left[\theta_{95\%}\left(e^t_{\mathrm{B}}\right) - \theta_{70\%}\left(e^t_{\mathrm{B}}\right)\right] \gamma_{\mathrm{ns}}}_{\text{Non-spinning Reserve of BSF}} \Bigg| \end{aligned} \tag{5}$$

where ρ is the target reserve cost requirement percentage of the baseline reserve cost. In this paper, a reduction of 25% in reserve levels is assumed, which is based on a project partner utility consensus. Thus, ρ is set as 75%. This advanced reserve calculation algorithm was originally developed in the Western Wind and Solar Integration Study Phase 2 (WWSIS-2) study [6]. The spinning reserve for the DA forecasting is defined as the 70% confidence interval ($\theta_{70\%}$) of the 1DA (or 2DA) solar power forecast errors. The non-spinning reserve is defined by the difference between the 95% confidence interval ($\theta_{95\%}$) and the 70% confidence interval ($\theta_{70\%}$) of the 1DA (or 2DA) solar power forecast errors [18]. The objective of (5) is to minimize the difference between the target reserve cost and the 75% baseline reserve costs. Ideally, the objective function value should be equal to 0. The target spinning, baseline spinning, target non-spinning, and baseline non-spinning reserve costs are expressed as: $\theta_{70\%}(e^t_{\mathrm{T},j})\gamma_{\mathrm{sp}}$, $\theta_{70\%}(e^t_{\mathrm{B}})\gamma_{\mathrm{sp}}$, $[\theta_{95\%}(e^t_{\mathrm{T},j}) - \theta_{70\%}(e^t_{\mathrm{T},j})]\gamma_{\mathrm{ns}}$, and $[\theta_{95\%}(e^t_{\mathrm{B}}) - \theta_{70\%}(e^t_{\mathrm{B}})]\gamma_{\mathrm{ns}}$, respectively. The parameters of γ_{sp} and γ_{ns} represent bidding prices of spinning and non-spinning reserves of thermal generators in \$/MWh, respectively.

C. Watt-Sun Solar Power Forecasting – A Multi-Expert Machine Learning Method

Watt-sun solar forecasting is based on a situation-dependent multi-expert machine learning method, which combines the linear model, random forest, and support vector machine methods to enhance the forecast accuracy. A dozen single machine-learning models are set-up and ingest the NWP models for the situation dependent learning. The algorithm that provides the best accuracy for the last two days is selected for future solar power forecasts. In comparison to the Dynamically Integrated ForeCast (DICast) method developed

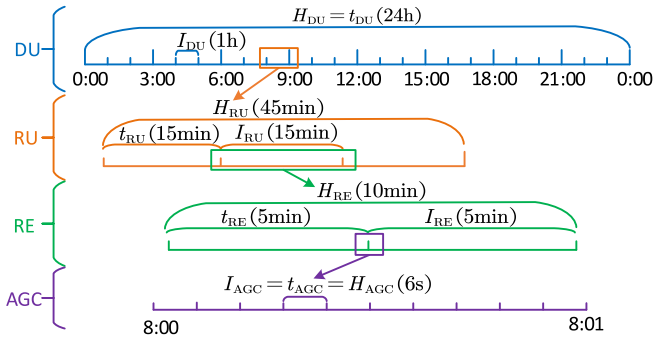


Fig. 3. Timeframes of FESTIV with four sub-models.

by the National Center for Atmospheric Research (NCAR), the main difference of the Watt-sun system is that it learns on historical forecasts [19]. Numerical results have shown a 30% improvement in solar irradiance/power forecast accuracy compared to forecasts based on the best individual method, and 10% improvement compared to model forecasts processed by machine learning methods without situation categorization. Detailed information about the Watt-sun forecast method can be found in [10] and [11].

III. Descriptions of Multi-Timescale Scheduling Models and Metrics

A. Multi-Timescale Scheduling Models

To study the reliability impacts of solar power forecasts on multi-timescale power system operations, we use an in-house tool, the Flexible Energy Scheduling Tool for Integration of Variable Generation model (FESTIV) [20], [21]. FESTIV is a multi-timescale steady-state power system operations simulation tool and uses several scheduling sub-models, as shown in Fig. 2. Four power system operation sub-models are included in FESTIV, including day-ahead security-constrained unit commitment (DU), real-time security-constrained unit commitment (RU), real-time security-constrained economic dispatch (RE), and AGC model. The solar power forecasts are updated in each scheduling sub-model during the simulation. Fig. 3 shows how the four sub-models are coupled at different timeframes. In this figure, t represents the time between updates in each scheduling sub-model, I represents the interval length, and H represents the scheduling horizon. For the DU sub-model, the interval resolution is I_{DU} (1 h). DU is simulated every t_{DU} (24 h) which is usually updated once per day in the current system operation. Thus, the solar power forecast in DU is also updated once per day. The optimization horizon (H_{DU}) in DU is 24 hours (one day). The RU sub-model is updated every t_{RU} (15 min) at an interval resolution of I_{RU} (15 min) and an optimization horizon of H_{RU} (45 min). Thus, the solar power forecast in RU is also updated every t_{RU} (15 min). The RE sub-model is updated every t_{RE} (5 min) at an interval resolution of I_{RE} (5 min) and an optimization horizon of H_{RE} (10 min). Thus, the solar power forecast in RE is also updated every t_{RE} (5 min). The AGC sub-model is updated every t_{AGC} (6 s). H_{AGC} and I_{AGC} are equal to t_{AGC}.

1) Objective Functions: The DU sub-model allows for variable startup costs that depend on the offline time of the unit starting up. A piecewise linear approximation of the cost curves of generators is utilized to retain a mixed-integer linear programming (MILP) formulation. The DU sub-model is run for a day-ahead time horizon (such as 24-hour commitment) and makes the initial commitment status and start-up decisions for all generators. With the initial status as input, the RU sub-model is repeated throughout the day to continuously update the commitment status of all generators. Both the DU and RU objectives are to minimize the total power system production costs including the operation costs, start-up costs, conventional reserve costs, loss load costs, and insufficient reserve costs. The objective function of DU and RU sub-models is given in (6). The RE sub-model aims to minimize the production costs that consist of the same costs as above without the start-up costs. The objective function of the RE sub-model is stated in (7). Note that the binary variables u_i^t and u_i^{t-1} have been determined in the RU sub-model and are fixed as constant in the RE sub-model. Thus, the start-up costs are not considered in the objective function of the RE sub-model.

2) Equality and Inequality Constraints: The objective functions comply with a number of prevailing constraints, including the system load balance (8), DC transmission capacity limits (9), minimum ON/OFF time limits (10)–(11), max/min active power limits of thermal units (12)–(13), up- and down-ramping rate limits (14)–(15), up- and down-ramping capacity limits (16)–(17), and all reserves' requirements (18) [including the spinning ($r = 1$), non-spinning ($r = 2$), regulation ($r = 3$), and replacement ($r = 4$) reserves]. The regulation reserve is supplied in both the up and down operating directions with the same attributes whereas the spinning, non-spinning, and replacement reserves are supplied only in the up operating direction. Prevailing RU and RE models share similar non-integer constraints. The binary variable $u_i^{t(\cdot)}$ is solved in the DU and RU models and used as a constant in the RE model.

Specific descriptions of equality and inequality constraints are provided as follows. Constraint (8) enforces the sum of total thermal generation and total solar power generation to strictly equal the total expected load at any time interval. Constraint (9) ensure that the power flowing on transmission lines does not exceed the maximum transmission capacity. Constraints (10)–(11) enforce the feasibility of thermal generators in terms of minimum ON/OFF time limits. Constraints (12)–(17) are all unit specific for thermal generators. Constraint (12) ensures that the power generation and its scheduled reserves are below the maximum capacity. Constraint (13) ensures that the power generation and its scheduled down-regulation reserve are not below the minimum capacity. Constraints (14)–(15) ensures that the power generation difference between successive time intervals is below the maximum up- and down-ramping rate limits. Constraint (16) ensures that the scheduled upward reserves are below the maximum up-ramping rate limit. Constraint (17) ensures that the scheduled downward reserves (regulation reserves) are below the maximum down-ramping rate limit. Constraint (18) ensures that the total amount of the scheduled

reserves provided by all the thermal units meets the rth reserve requirement.

$$\min \sum_{t^{(\cdot)}=1}^{T^{(\cdot)}} \left\{ \underbrace{VOLL \times ENS_{t^{(\cdot)}}}_{\text{Loss Load Cost}} + \underbrace{\sum_{r=1}^{NR} VOIR^r \times IR^r_{t^{(\cdot)}}}_{\text{Insufficent Reserve Cost}} + \sum_{i=1}^{NI} \left[\underbrace{C_i^{t^{(\cdot)}}\left(p_i^{t^{(\cdot)}}, u_i^{t^{(\cdot)}}\right) + S_i^{t^{(\cdot)}}\left(u_i^{t^{(\cdot)}}, u_i^{t^{(\cdot)}-1}\right)}_{\text{Operation and Start-up Cost}} + \underbrace{\sum_{r=1}^{NR} \gamma^r_{i,t^{(\cdot)}} R^r_{i,t^{(\cdot)}}}_{\text{Reserve Cost}} \right] \right\} \tag{6}$$

$$\min \sum_{t^{\mathrm{RE}}=1}^{T^{\mathrm{RE}}} \left\{ \sum_{i=1}^{NI} \left[\underbrace{C_i^{t^{\mathrm{RE}}}\left(p_i^{t^{\mathrm{RE}}}\right)}_{\text{Operation Cost}} + \underbrace{\sum_{r=1}^{NR} \gamma^r_{i,t^{\mathrm{RE}}} R^r_{i,t^{\mathrm{RE}}}}_{\text{Reserve Cost}} \right] + \underbrace{VOLL \times ENS_{t^{\mathrm{RE}}}}_{\text{Loss Load Cost}} + \underbrace{\sum_{r=1}^{NR} VOIR^r \times IR^r_{t^{\mathrm{RE}}}}_{\text{Insufficent Reserve Cost}} \right\} \tag{7}$$

$$\sum_{i=1}^{NI} p_i^{t^{(\cdot)}} + \sum_{s=1}^{NS} p_s^{t^{(\cdot)}} = \sum_{b=1}^{NB} d_b^{t^{(\cdot)}} \tag{8}$$

$$-\mathbf{PL}^{\max} \le \mathbf{SF} \times [\mathbf{K_P} \times \mathbf{P} + \mathbf{K_S} \times \mathbf{P_S} - \mathbf{K_D} \times \mathbf{D}] \le \mathbf{PL}^{\max} \tag{9}$$

$$\left[X_{i,\mathrm{on}}^{t^{(\cdot)}-1} - T_{i,\mathrm{on}}\right] \cdot \left[u_i^{t^{(\cdot)}-1} - u_i^{t^{(\cdot)}}\right] \ge 0 \tag{10}$$

$$\left[X_{i,\mathrm{off}}^{t^{(\cdot)}-1} - T_{i,\mathrm{off}}\right] \cdot \left[u_i^{t^{(\cdot)}} - u_i^{t^{(\cdot)}-1}\right] \ge 0 \tag{11}$$

$$p_i^{t^{(\cdot)}} + \sum_{r=1}^{NR} R^r_{i,t^{(\cdot)}} \le P_i^{\max} \times u_i^{t^{(\cdot)}}, \ \forall i, \forall r, \forall t^{(\cdot)} \tag{12}$$

$$p_i^{t^{(\cdot)}} - R^3_{i,t^{(\cdot)}} \ge P_i^{\min} \times u_i^{t^{(\cdot)}}, \ \forall i, \forall t^{(\cdot)} \tag{13}$$

$$p_i^{t^{(\cdot)}} - p_i^{t^{(\cdot)}-1} \le R_i^{\mathrm{up}} \times \triangle t^{(\cdot)} + L_C \times \left(2 - u_i^{t^{(\cdot)}} - u_i^{t^{(\cdot)}-1}\right) \tag{14}$$

$$p_i^{t^{(\cdot)}-1} - p_i^{t^{(\cdot)}} \le R_i^{\mathrm{dn}} \times \triangle t^{(\cdot)} + L_C \times \left(2 - u_i^{t^{(\cdot)}} - u_i^{t^{(\cdot)}-1}\right) \tag{15}$$

$$\sum_{r=1}^{NR} R^r_{i,t^{(\cdot)}} \le R_i^{\mathrm{up}} \times \triangle t^{(\cdot)}, \quad \forall i, \forall r, \forall t^{(\cdot)} \tag{16}$$

$$R^3_{i,t^{(\cdot)}} \le R_i^{\mathrm{dn}} \times \triangle t^{(\cdot)}, \quad \forall i, \forall t^{(\cdot)} \tag{17}$$

$$\sum_{i=1}^{NI} R^r_{i,t^{(\cdot)}} \ge RR^r_{t^{(\cdot)}}, \quad \forall i, \forall r, \forall t^{(\cdot)}. \tag{18}$$

3) Automatic Generation Control Model: The AGC model is a rule-based algorithm that is the final line of scheduling defense to control resources to reduce the imbalance error [22]. Instead of optimizing the scheduling of units based on costs, the AGC model utilizes all units that are providing regulation reserve as scheduled by the RE model to assist in correcting the area control error (ACE). AGC schedules the ACE correction via the proportion of the units regulating schedules and ramp rates provided by the last RE model. As seen in Fig. 2, AGC uses the dispatch and reserve schedules from the RE model as its input and provides AGC schedules and realized generation as output. The realized generation output is determined by the prior AGC schedule and the resources behavior rate. The reliability benefits are calculated with the AGC model as a true steady-state simulation.

B. Reliability Metrics

Area Control Error (ACE) is the difference between the sum of total generation and the total load at any given time period. It is the main driver of all imbalance metrics and formulated by the smoothed AGC mode in actual operations [23] as shown below:

$$ACE_{t,sm} = K_1 ACE_{t,raw} + \frac{K_2}{T_n} \int_{t-T_n}^{t} ACE_{x,raw} \mathrm{d}x \tag{19}$$

$$ACE_{t,raw} = \sum_{i=1}^{NI} p_i^{t^{\mathrm{AGC}}} + \sum_{s=1}^{NS} p_s^{t^{\mathrm{AGC}}} - \sum_{b=1}^{NB} d_b^{t^{\mathrm{AGC}}} - LS^{t^{\mathrm{AGC}}} \tag{20}$$

$$K_1 = \frac{T_{\mathrm{CPS2}} - mod(t, T_{\mathrm{CPS2}})}{T_{\mathrm{CPS2}}} \tag{21}$$

$$K_2 = \frac{mod(t, T_{\mathrm{CPS2}})}{T_{\mathrm{CPS2}}} \tag{22}$$

$$T_n = mod(t, T_{\mathrm{CPS2}}) \tag{23}$$

where K_1 and K_2 are parameters used to produce the low frequency ACE signal by passing the raw ACE signal through a PI filter. K_1 and K_2 are selected as 1 and 2, respectively. T_n is the integral length of the PI filter to smoothen the raw ACE over the past 3 minutes, i.e., T_n=3×60=180 [23].

Absolute area control error in energy (AACEE) is the absolute value of ACE at every t_{AGC} interval, summed over the study period in units of MWh. This gives an indication of the total imbalance occurring for the study period in either direction. AACEE is a function of the time resolution of the RE model (I_{RE}), the time horizon of the RE model (H_{RE}), the amount of PV on the system (P_{PV}), the load (P_{LOAD}), and the amount of total ramp available for managing the variability (P_{RAMP}) [23], given by:

$$AACEE = \alpha_1 I_{\mathrm{RE}} + \alpha_2 H_{\mathrm{RE}} + \alpha_3 P_{\mathrm{PV}} + \alpha_4 P_{\mathrm{LOAD}} + \alpha_5 P_{\mathrm{RAMP}} \tag{24}$$

$$[\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5] = \left[\frac{\partial AACEE}{\partial I_{\mathrm{RE}}}, \frac{\partial AACEE}{\partial H_{\mathrm{RE}}}, \frac{\partial AACEE}{\partial P_{\mathrm{PV}}}, \frac{\partial AACEE}{\partial P_{\mathrm{LOAD}}}, \frac{\partial AACEE}{\partial P_{\mathrm{RAMP}}} \right] \tag{25}$$

where $\alpha_{(\cdot)}$ is a sensitivity coefficient and calculated by the standard deviation of output changes at different timescales.

Control performance standard 2 (CPS2) is a reliability standard (defined by the North American Electric Reliability Corporation (NERC) [7]) that measures the amount of intervals where the absolute value of ACE exceeds a predefined threshold. The reliability indicator ACE_{CPS2} measures the sum of instantaneous ACE until the 10-minute CPS2 interval (L10)

ends for assessing the $CPS2_{\text{score}}$. The unit of ACE_{CPS2} is MW-10min and the τth value of ACE_{CPS2} is given by:

$$ACE_{\text{CPS2},\tau} = \sum_{t=(\tau-1)\times T_{\text{CPS2}}\times 60}^{\tau\times T_{\text{CPS2}}\times 60-t_{\text{AGC}}} ACE_{t,raw}\times\frac{t_{\text{AGC}}}{T_{\text{CPS2}}\times 60} \tag{26}$$

where τ is the time period index of the ACE_{CPS2} value.

$CPS2_{\text{score}}$ measures the percentage of intervals without violations. A violation occurs when an interval exceeds the 50 MW-10 min ACE limit in a 10-minute CPS2 interval, given by:

$$CPS2_{\text{score}} = \left(1-\frac{N_{\text{V}}}{N_{\text{T}}-N_{\text{U}}}\right)\times 100\% \tag{27}$$

where N_{V}, N_{T}, and N_{U} represent the number of violation periods, total periods, and unavailable periods, respectively. $CPS2_{\text{score}}$ gives an indication of how often the system encounters severe imbalance errors. Normally, the $CPS2_{\text{score}}$ is greater than or equal to 90%. The number of violation periods N_{V} is calculated by:

$$N_{\text{V}} = \sum_{\tau\in\mathcal{H}_{\text{DU}}} f_\tau \tag{28}$$

where f_τ is a flag variable defined to indicate whether the $ACE_{\text{CPS2},\tau}$ value of an interval exceeds the ACE limit (i.e., 50 MW-10 min in this paper) in a CPS2 interval (10 minutes as defined by NERC). If one violation occurs, f_τ equals 1; otherwise, f_τ equals 0. The mathematical formulation of f_τ is given by:

$$f_\tau = \begin{cases} 1, & \text{if } ACE_{\text{CPS2},\tau} > 50 \\ 0, & \text{if } ACE_{\text{CPS2},\tau} \le 50 \end{cases} \tag{29}$$

The σ_{ACE} metric is the standard deviation of raw ACE values and gives an indication of the distribution of ACE values for the study period. Generally, smaller AACEE, σ_{ACE}, CPS2 violation values, and larger CPS2 scores indicate better reliability performance in power system operations. Hence, an integrated metric of the expected synthetic reliability (ESR) is proposed by normalizing the reliability metrics, and is given by:

$$\begin{aligned} ESR &= \boldsymbol{\lambda}\cdot\mathbf{DIAG}\cdot\mathbf{Rel} \\ &= \lambda_1 CPS2^{\text{N}}_{\text{score}} - \lambda_2 AACEE^{\text{N}} - \lambda_3\sigma^{\text{N}}_{\text{ACE}} - \lambda_4 N^{\text{N}}_{\text{V}} \end{aligned} \tag{30}$$

$$\begin{cases} \boldsymbol{\lambda} = [\lambda_1\ \lambda_2\ \lambda_3\ \lambda_4] \\ \mathbf{DIAG} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \\ \mathbf{Rel} = \left[CPS2^{\text{N}}_{\text{score}}\ AACEE^{\text{N}}\ \sigma^{\text{N}}_{\text{ACE}}\ N^{\text{N}}_{\text{V}}\right]^{\text{T}} \\ N^{\text{N}}_{\text{V}} = \left(N_{\text{V}}-\mu_{N_{\text{V}}}\right)/\sigma_{N_{\text{V}}} \\ \sigma^{N}_{\text{ACE}} = \left(\sigma_{\text{ACE}}-\mu_{\sigma_{\text{ACE}}}\right)/\sigma_{\sigma_{\text{ACE}}} \\ AACEE^{\text{N}} = (AACEE-\mu_{AACEE})/\sigma_{AACEE} \\ CPS2^{\text{N}}_{\text{score}} = \left(CPS2_{\text{score}}-\mu_{CPS2_{\text{score}}}\right)/\sigma_{CPS2\text{score}} \end{cases} \tag{31}$$

where $\mu_{(\cdot)}$ and $\sigma_{(\cdot)}$ represent the mean value and standard deviation of each of the reliability metrics, respectively. **Rel** is the reliability index vector. In this paper, four reliability metrics widely used in the literature are deployed as the elements of the reliability index vector, i.e., $CPS2_{\text{score}}$, $AACEE$, σ_{ACE}, and N_{V}. Note that multi-objective optimization could also be leveraged to find the Pareto front of these four reliability metrics. This reliability index vector can also allow more reliability metrics to be added. This paper does not aim to develop new reliability metrics, but rather to examine the trade-off of an integrated reliability metric instead of an individual metric value. $\boldsymbol{\lambda}$ is the weight coefficient vector of the reliability index vector. The sum of weight coefficients must equal unity. The proposed ESR metric is flexible, which allows balancing authorities naturally place more weight on certain metrics than other balancing authorities, according to their systems status and particular risk preferences, such as renewable penetrations, peak load, flexibility, etc. Here in this paper we have equally weighted the four metrics (i.e., the weight coefficient is 1/4), but each user or balancing authority can choose weightings that match their particular risk preferences. **DIAG** is the diagonal matrix of the attribute of reliability metrics. If a larger reliability metric indicates a better reliability performance, this metric is defined as a positive index. If a smaller reliability metric indicates a better reliability performance, this metric is defined as a negative index. If the ith reliability metric is a positive index, the element diag_{ii} equals 1. If the ith reliability metric is a negative index, the element diag_{ii} equals -1. In this case, $CPS2_{\text{score}}$ is a positive index and $\text{diag}_{11} = 1$. The other three reliability metrics are negative indexes and $\text{diag}_{22} = \text{diag}_{33} = \text{diag}_{44} = -1$.

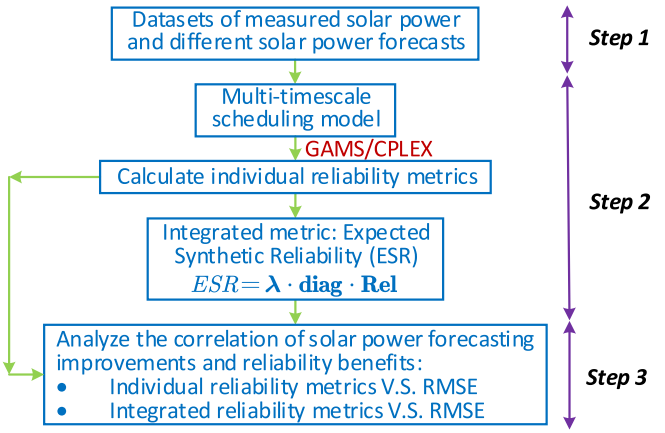
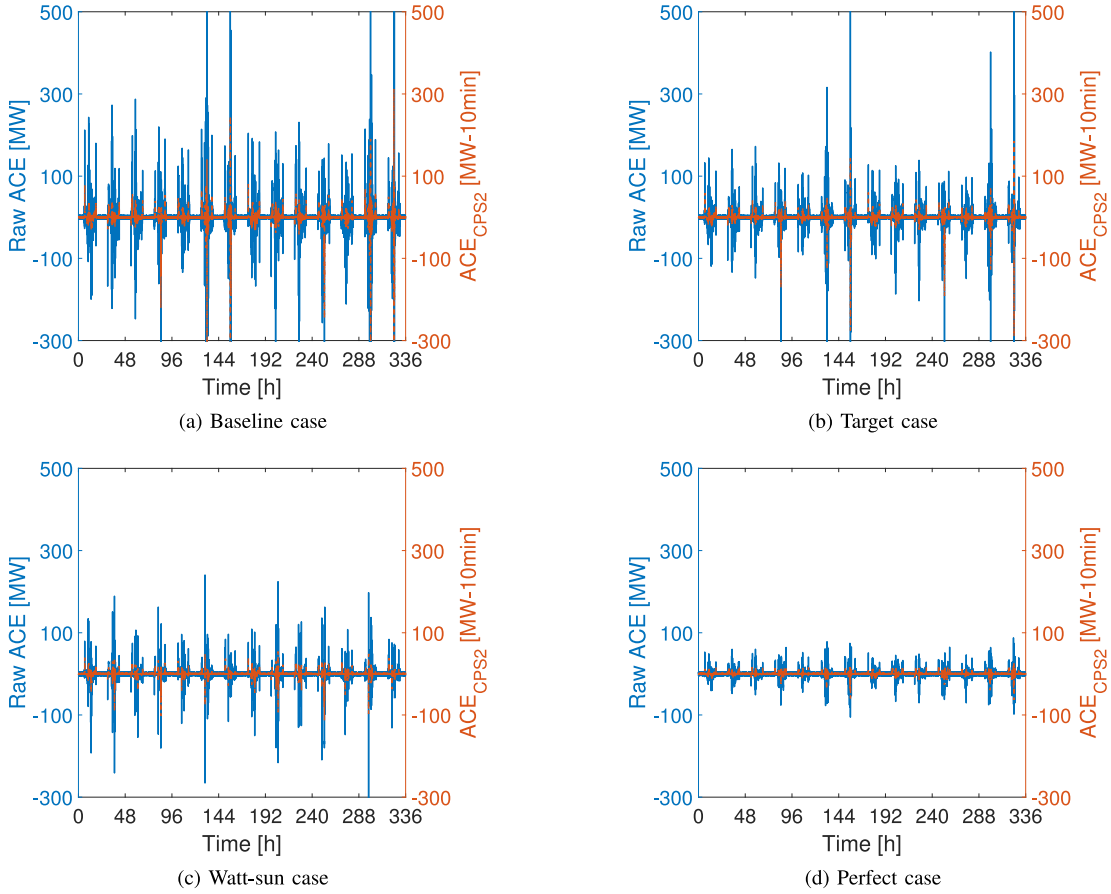


Fig. 4. The flow chart for quantifying and analyzing reliability benefits from improved solar power forecasts.

C. Procedure for Quantifying and Analyzing Reliability Benefits

The flow chart for quantifying and analyzing reliability benefits from improved solar power forecasts is shown in Fig. 4, which consists of three major steps:

- ***Step 1***: Prepare the datasets of measured solar power and different solar power forecasts (as described in Section II).
- ***Step 2***: The datasets are input into the multi-timescale scheduling model to calculate different individual reliability metrics and the integrated ESR metric (as described in Sections III-A and III-B).



(a) Baseline case (b) Target case

(c) Watt-sun case (d) Perfect case

Fig. 5. Comparison of ACE in the baseline, target, Watt-sun, and perfect forecast cases at 1HA forecast with a 5.08% solar penetration.

- ***Step 3***: The impacts of solar power forecasting improvements on the reliability of power system operations are analyzed (as described in the following Section IV).

IV. Case Studies and Results

Numerical simulations are performed by using multi-timescale scheduling models based on FESTIV and the representative IEEE 118-bus system [24]. This system has 54 thermal units, 186 branches, and 91 load buses. The parameters of generators, transmission network, and load profiles are given in [25] and [26] with the same timeframe. All tests are carried out using the General Algebraic Modeling System (GAMS) Distribution 24.7 [27], and solved using ILOG CPLEX 12.6 [28] on two Intel-e5-2603 1.6-GHz workstations with 32 GB of RAM memory. We simulate 400 different scenarios by using five solar power penetration levels (5.08%, 10.16%, 15.24%, 20.32%, and 25.40%), five locations (two utilities: Green Mountain Power (GMP) and Tucson Electric Power (TEP); two ISOs: ISO-New England (ISO-NE) and California Independent System Operator (CAISO); and one individual solar plant: Smyrna), four time ahead forecasts (2DA, 1DA, 4HA, and 1HA), and four solar forecasting methods (baseline, target, perfect, and Watt-sun). Representative findings are statistically analyzed and discussed as follows.

A. Overview of ACE Profiles

The ACE profiles in the baseline, target, Watt-sun, and perfect cases at the 1HA forecast for TEP with the 5.08% penetration level are compared in Fig. 5. It is seen that the ACE magnitudes in the target, Watt-sun, and perfect cases are significantly smaller than in the baseline case. The perfect case shows much more smooth and flat profiles for both raw ACE and CPS2 ACE. Spikes of raw ACE and CPS2 ACE are gradually reduced by the solar power forecasting improvement in multi-timescale power system operations. This is because solar power forecasting improvement can help accurately commit thermal units. Consequently, the reliability performance of the multi-timescale power system operations is improved.

As shown in Fig. 5, the difference of ACE during nighttime hours in the four cases is negligible. The variation of ACE is very small during nighttime hours. However, the difference of ACE is relatively more significant in the four cases during daylight hours. Table II compares the average and standard deviation of ACE values during daylight and nighttime hours. Due to the variation of solar power during daylight hours, both the average and standard deviation of ACE values are larger than those during nighttime hours. This is because solar power is zero during nighttime hours, and the slight variation of ACE values is caused by the imbalance between conventional generation and load.

Another interesting finding is that the variation of ACE in weekend is slightly higher than that in weekdays. This is because the average load demand in weekend (1.12×10^5 MWh) is lower than that in weekdays (1.34×10^5 MWh). The solar power variation on the weekends impacts the imbalance between power generation and load demand relatively more distinctly, comparing to the solar power variation

TABLE I
OVERALL RESULTS OF AACEE UNDER THREE SOLAR POWER PENETRATION LEVELS

		5.08% penetration				15.24% penetration				25.40% penetration			
		Baseline	Target	Perfect	Watt	Baseline	Target	Perfect	Watt	Baseline	Target	Perfect	Watt
GMP	Value [MWh]	1785	926	739	908	1861	1387	1113	1339	2569	2034	1379	1832
2DA	Improvement [%]	-	**48.12**	58.61	49.13	-	**25.47**	40.19	28.05	-	20.83	46.32	28.69
GMP	Value [MWh]	1632	925	739	762	1771	1381	1113	1339	2476	1903	1379	1526
1DA	Improvement [%]	-	43.32	54.72	53.31	-	22.02	37.15	24.39	-	23.14	44.31	38.37
GMP	Value [MWh]	1957	1356	739	1225	2024	1507	1113	1473	2674	2437	1379	2229
4HA	Improvement [%]	-	30.71	62.24	37.40	-	25.54	45.01	27.22	-	8.86	48.43	16.64
GMP	Value [MWh]	851	756	739	926	1339	1179	1113	1429	1586	1477	1379	2141
1HA	Improvement [%]	-	11.16	13.16	-8.81	-	11.95	16.88	-6.72	-	6.87	13.05	-34.99
TEP	Value [MWh]	1446	1029	577	689	1541	1256	851	992	2317	1982	1113	1381
2DA	Improvement [%]	-	28.84	60.11	52.35	-	18.49	44.78	35.63	-	14.46	51.96	40.41
TEP	Value [MWh]	1563	1117	577	589	1662	1341	851	871	2404	2012	1113	1345
1DA	Improvement [%]	-	28.53	63.08	**62.32**	-	19.31	**48.81**	**47.59**	-	16.31	**53.71**	**44.05**
TEP	Value [MWh]	1306	849	577	879	1342	1153	851	1161	2164	1682	1113	1962
4HA	Improvement [%]	-	34.99	55.82	32.71	-	14.08	36.59	13.49	-	22.27	48.57	9.33
TEP	Value [MWh]	826	607	577	699	1069	975	851	1055	1614	1369	1113	1612
1HA	Improvement [%]	-	26.51	30.15	15.38	-	8.79	20.39	1.31	-	15.18	31.04	0.12
Smyrna	Value [MWh]	1804	1543	869	1427	2043	1857	1241	1832	2696	2212	1576	1984
2DA	Improvement [%]	-	14.47	51.83	20.91	-	9.11	39.26	10.33	-	17.95	41.54	26.41
Smyrna	Value [MWh]	1765	1499	869	1188	1949	1852	1241	1668	2603	2025	1576	1946
1DA	Improvement [%]	-	15.07	50.76	32.69	-	4.98	36.33	14.42	-	22.21	39.45	25.24
Smyrna	Value [MWh]	1916	1677	869	1488	2093	1931	1241	1832	2727	2393	1576	1988
4HA	Improvement [%]	-	12.47	54.65	22.34	-	7.74	40.71	12.47	-	12.25	42.21	27.11
Smyrna	Value [MWh]	1402	943	869	1254	1821	1378	1241	1751	1982	1691	1576	1959
1HA	Improvement [%]	-	32.74	38.02	10.56	-	24.33	31.85	3.84	-	14.68	20.48	1.16
ISO-NE	Value [MWh]	2199	1483	948	1774	2209	1683	1268	1784	2673	1841	1405	2353
2DA	Improvement [%]	-	32.56	56.89	19.33	-	23.81	42.61	19.24	-	**31.13**	47.44	11.97
ISO-NE	Value [MWh]	1693	1186	948	1286	1758	1526	1268	1665	2193	1794	1405	1807
1DA	Improvement [%]	-	29.95	44.01	24.04	-	13.21	27.87	5.29	-	18.19	35.93	17.61
CAISO	Value [MWh]	1985	1258	655	1061	2048	1901	1144	1787	2625	2216	1292	2079
2DA	Improvement [%]	-	36.62	**67.01**	46.55	-	7.18	44.14	12.74	-	15.58	50.78	20.81
CAISO	Value [MWh]	1599	991	655	691	1941	1767	1144	1334	2343	1885	1292	1569
1DA	Improvement [%]	-	38.02	59.04	56.79	-	8.96	41.06	31.27	-	19.55	44.86	33.03

Note: (i) Watt-sun performs better than both target and baseline (green); (ii) Watt-sun performs better than baseline but worse than target (blue); and (iii) Watt-sun performs worse than both baseline and target (red).

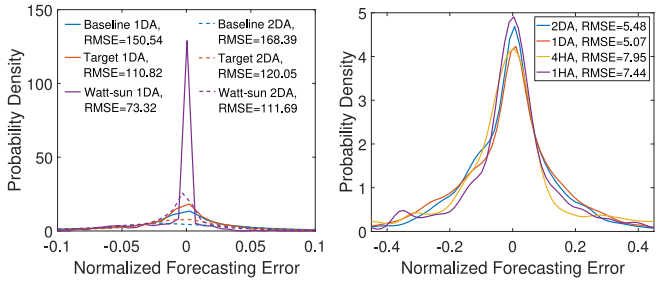
TABLE II
COMPARISON OF ACE VALUES DURING DAYLIGHT AND NIGHTTIME HOURS IN THE WATT-SUN CASE

Time	Average ACE [MWh]	Standard Deviation of ACE [MWh]
Daylight	937	23.45
Nighttime	95	4.32

in weekdays due to its being a larger component of the generation mix.

B. AACEE Enhancement

The overall AACEE results with three solar power penetration levels (5.08%, 15.24%, and 25.40%) for weekly simulations are shown in Table I. Smaller AACEE values represent a better reliability performance. For all cases, perfect forecasts perform the best among the four forecasting methods while baseline forecasts perform the worst except for the GMP 1HA case. At the 5.08% penetration, the highest improvement for perfect forecasts is the CAISO 2DA case that has a 67.01% AACEE reduction and 655 MWh AACEE value. At the 15.24% penetration, the highest improvement for perfect forecasts is the TEP 1DA case with a 48.81% AACEE reduction and 851 MWh AACEE value. At the 25.40% penetration, the highest improvement for perfect forecasts is also the TEP 1DA case with a 53.71% AACEE reduction and 1,113 MWh AACEE value. Target forecasts perform better than baseline forecasts and worse than perfect forecasts among different forecast scenarios. At the 5.08% penetration, the highest improvement for target forecasts is the GMP 2DA case with a 48.12% AACEE reduction and 926 MWh AACEE value. At the 15.24% penetration, the highest improvement for target forecasts is also the GMP 2DA case with a 25.47% AACEE reduction and 1387 MWh AACEE value. At the 25.40% penetration, the highest improvement for target forecasts is the ISO-NE 2DA case with a 31.13% AACEE reduction and 1,841 MWh AACEE value. Watt-sun forecasts perform worse than the baseline only in the GMP 1HA case, as shown in red in Table I. This is because that Watt-sun improved the baseline forecasts using the machine learning method instead of applying uniform improvements in target forecasts. Though Watt-sun is more accurate than the baseline forecasts over the entire study period, it is possible that Watt-sun has larger forecast errors at certain time periods. In TEP 4HA, TEP 1HA, Smyrna 1HA, ISO-NE 2DA, and ISO-NE 1DA cases, Watt-sun forecasts perform better than the baseline forecasts but worse than the target forecasts, as shown in blue in Table I. However, Watt-sun forecasts perform better than both baseline and target forecasts in most cases (ten cases), as shown in green in Table I. Taking the CAISO 1DA case as an example, the root mean square error (RMSE) of the Watt-sun case is 73.32 MW but RMSEs of baseline and target forecasts are 150.54 MW and 110.82 MW, respectively. Distribution of forecast errors



(a) CAISO forecasts at 2DA and 1DA (b) Watt-sun forecasts for GMP

Fig. 6. Distributions of forecasting errors. (a) CAISO site at 2DA and 1DA forecasts with baseline, target, and Watt-sun forecasts; (b) GMP site at different time ahead forecasts with Watt-sun forecasts.

TABLE III
COMPARISON OF σ_{ACE} VALUES FOR GMP, TEP, AND SMYRNA AT 5.08% PENETRATION [MWh]

Scenarios		Baseline	Target	Watt-sun	Perfect
GMP	2DA	16.55	12.73	10.99	8.43
	1DA	14.43	11.24	9.45	8.43
	4HA	19.01	13.62	13.42	8.43
	1HA	9.85	9.37	12.76	8.43
TEP	2DA	16.43	12.72	9.49	6.13
	1DA	18.87	13.91	13.19	6.13
	4HA	15.21	12.58	12.66	6.13
	1HA	11.61	6.69	11.54	6.13
Smyrna	2DA	20.93	18.27	15.53	9.42
	1DA	19.29	17.72	13.73	9.42
	4HA	21.32	19.09	17.71	9.42
	1HA	15.05	12.68	14.47	9.42

of Watt-sun CAISO 1DA case is shown in Fig. 6a. AACEE values are increased with the solar power penetration, due to more stable but expensive thermal generation is replaced by the increasing solar power, which is much economic but more intermittent.

C. Standard Deviation of ACE (σ_{ACE}) Analysis

Table III compares the σ_{ACE} values of GMP, TEP, and Smyrna at the 5.08% penetration level. Smaller σ_{ACE} values represent better reliability performance. The results are similar to those shown in the AACEE analysis. Target forecasts perform better than baseline forecasts and worse than perfect forecasts among all forecast scenarios. In most cases, Watt-sun forecasts perform better than baseline forecasts and sometimes even better than target forecasts. In addition, the 1HA forecast shows the smallest σ_{ACE} values for all three sites, comparing to 2DA, 1DA, and 4HA forecasts. This is because shorter time ahead forecast can usually obtain more accurate forecast results and consequently reduce the σ_{ACE} values. However, a more accurate solar forecast may not necessarily result in a small σ_{ACE} value at all cases, as shown in Table III. For instance, the 4HA forecast shows the largest σ_{ACE} values for the GMP and Smyrna cases, whereas the 1DA forecast shows the largest σ_{ACE} values for the TEP case. The distribution of forecast errors of the GMP case by Watt-sun forecasts is shown in Fig. 6b.

D. CPS2 Score and Violations

Table IV compares the CPS2 violations and CPS2 score for CAISO, GMP, TEP, and Smyrna at the 2DA forecast and 15.24% penetration. Lower CPS2 violations and higher CPS2 score represent better reliability performance. Perfect forecasts outperform the other three forecasts while baseline forecasts perform the worst. Target forecasts perform better than baseline forecasts and worse than perfect forecasts in all forecast scenarios. Watt-sun forecasts perform better than both baseline and target forecasts. As shown in the CAISO case, Watt-sun forecasts reduce the CPS2 violations from 26 to 20, and enhance the CPS2 score from 91.61% to 93.12%. The CPS2 violations improvement is 23.08% and the CPS2 score improvement is 1.65%. In this case, both CPS2 score and violations metrics decrease with the improvement of solar power forecasting. The average enhancement of Watt-sun forecasts is 13.37% for the CPS2 violations metric and 1.59% for the CPS2 score metric.

TABLE IV
COMPARISON OF CPS2 VIOLATION AND SCORE RESULTS FOR CAISO, GMP, TEP, AND SMYRNA AT 15.24% PENETRATION AND 2DA FORECAST

Scenarios		CPS2 Violations		CPS2 Score	
		Value	Improvement [%]	Value [%]	Improvement [%]
CAISO	Baseline	26	-	91.61	-
	Target	22	15.38	92.99	1.51
	Watt-sun	20	23.08	93.12	1.65
	Perfect	15	42.31	97.92	6.89
GMP	Baseline	44	-	92.26	-
	Target	34	22.73	96.31	4.39
	Watt-sun	26	40.91	97.42	5.59
	Perfect	19	56.82	98.12	6.34
TEP	Baseline	43	-	92.76	-
	Target	36	16.28	95.39	2.85
	Watt-sun	27	37.21	97.13	4.71
	Perfect	21	51.16	98.91	6.63
Smyrna	Baseline	45	-	91.35	-
	Target	35	22.22	92.43	1.18
	Watt-sun	32	28.89	95.31	4.33
	Perfect	21	53.33	97.62	6.86

E. Correlation Between Solar Power Forecasting Improvements and Reliability Benefits of Power System Operations

Fig. 7 shows the statistical correlations between forecasting improvements and individual reliability metrics, i.e., $CPS2^{\text{N}}_{\text{score}}$, $AACEE^{\text{N}}$, $\sigma^{\text{N}}_{\text{ACE}}$, and N^{N}_{V}, for the TEP case with a 10.16% solar penetration level. It is seen that the $CPS2^{\text{N}}_{\text{score}}$ metrics increase with the improvement of solar power forecasts (i.e., the decrease of RMSE). The $AACEE^{\text{N}}$, $\sigma^{\text{N}}_{\text{ACE}}$, and N^{N}_{V} metrics decrease with improvement of solar power forecasts. Since smaller $AACEE^{\text{N}}$, $\sigma^{\text{N}}_{\text{ACE}}$, N^{N}_{V}, and larger $CPS2^{\text{N}}_{\text{score}}$ indicate better reliability performance, the improvement of solar power forecasts could increase the reliability benefits of multi-timescale power system operations. While 1DA forecasts result in better reliability performance than 2DA with the same RMSE value, the 2DA forecasts present a larger improving rate (i.e., the rate of the linear curve). Note that there are large enough errors that might cause a step change in reliability performance, which would not continue the approximate linear relationship between RMSE and the expected (or individual) reliability metrics as shown in Fig. 7.

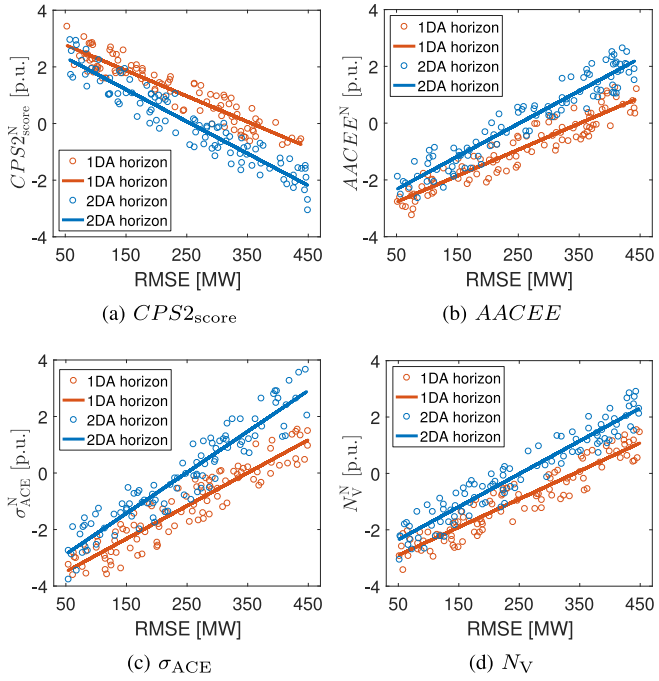


Fig. 7. Statistical correlation between RMSE and individual reliability metrics for the TEP case with a 10.16% solar penetration level.

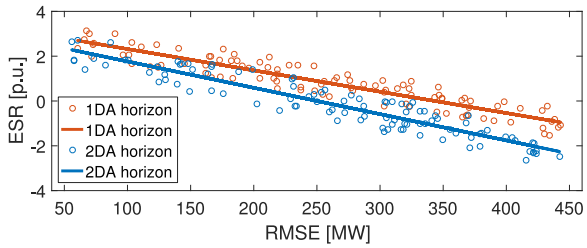


Fig. 8. Statistical correlation between RMSE and the expected reliability metrics of power system operations.

Fig. 8 shows the statistical correlation between forecasting improvements and the expected reliability benefits. It is seen that the ESR metrics increase with the improvement of solar power forecasts (i.e., the decrease of RMSE). According to the empirical criteria of correlationship in [29], both correlation coefficients of forecasting improvements and reliability benefits in 1DA and 2DA forecasts, (i.e., 0.7836 and 0.7966, respectively), are greater than 0.5 and less than 0.8. There exists a *clear correlation* between the solar power forecasting improvement and the power system reliability benefit. In the 1DA forecast, a 10% forecasting improvement increases the reliability benefit by approximately 8.72%. In the 2DA forecast, a 10% forecasting improvement increases the reliability benefit by approximately 12.88%.

V. Discussion

The solar forecast performance can improve the reliability benefits of power system operations. Our previous study [11] has shown that the forecast error has a strong dependence on the global horizontal irradiance (GHI), solar zenith angle, and cloud column integrated liquid water content. Hence, advanced forecast techniques of these three parameters in NWP models would decrease the forecast error and improve the solar forecasting performance. Another way to improve the solar forecasting performance is to blend different individual NWP models. A blending model like Watt-sun has been shown to reduce the localized forecast error of individual NWP models by ∼30% [11], compared to the best individual model.

VI. Conclusion

This paper developed a methodology for quantifying and analyzing the reliability benefits of improved solar power forecasts in the multi-timescale power system operations. An expected synthetic reliability (ESR) metric was proposed to quantify the enhanced power system reliability benefits from improved solar power forecasts. The absolute area control error in energy (AACEE), CPS2 violations, CPS2 score, and standard deviation of the raw ACE were calculated and combined as the ESR metric. Numerical simulation results based on the IEEE 118-bus system showed that the expected synthetic reliability (ESR) metrics increased with the improvement of solar power forecasts and a 10% solar power forecasting improvement could increase the reliability benefit by approximately 8.72%–12.88%.

In the future work, this research can be further extended by quantifying the reliability benefits of solar power ramping forecast improvements. In addition, the reliability quantification method developed in this paper can also be extended to estimate the reliability impacts of other energy efficiency and renewable energy technologies (e.g., wind forecasting and electric vehicle grid integration) on multi-timescale power system operations.

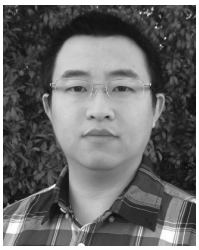
Acknowledgment

The authors would also like to thank the anonymous reviewers for their constructive suggestions to this research.

References

[1] T. Mai *et al.*, "Renewable electricity futures for the United States," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 372–378, Apr. 2014.
[2] Y. Chu *et al.*, "Short-term reforecasting of power output from a 48 MWe solar PV plant," *Solar Energy*, vol. 112, pp. 68–77, Feb. 2015.
[3] J. Antonanzas *et al.*, "Review of photovoltaic power forecasting," *Solar Energy*, vol. 136, pp. 78–111, Oct. 2016.
[4] J. Zhang *et al.*, "A suite of metrics for assessing the performance of solar power forecasting," *Solar Energy*, vol. 111, pp. 157–175, Jan. 2015.
[5] C. B. Martinez-Anido *et al.*, "The value of day-ahead solar power forecasting improvement," *Solar Energy*, vol. 129, pp. 192–203, May 2016.
[6] D. Lew *et al.*, "The western wind and solar integration study phase 2," Nat. Renewable Energy Lab., Golden, CO, USA, Tech. Rep. NREL/TP-5500-55588, Sep. 2013.
[7] *Reliability Standards for the Bulk Electric Systems of North America*, North Amer. Elect. Rel. Corporat., Atlanta, GA, USA, 2017. [Online]. Available: http://www.nerc.com/pa/Stand/Reliability%20Standards%20Complete%20Set/RSCompleteSet.pdf
[8] C. B. Martinez-Anido, A. R. Florita, and B.-M. Hodge, "The impact of improved solar forecasts on bulk power system operations in ISO-NE," Nat. Renewable Energy Lab., Golden, CO, USA, Tech. Rep. NREL/CP-5D00-62817, Sep. 2014.
[9] Q. Wang, C. B. Martinez-Anido, H. Wu, A. R. Florita, and B.-M. Hodge, "Quantifying the economic and grid reliability impacts of improved wind power forecasting," *IEEE Trans. Sustain. Energy*, vol. 7, no. 4, pp. 1525–1537, Oct. 2016.

[10] *Watt-Sun: A Multi-Scale, Multi-Model, Machine-Learning Solar Forecasting Technology*, IBM, Armonk, NY, USA, 2012. [Online]. Available: http://energy.gov/eere/sunshot/watt-sun-multi-scale-multi-model-machine-learning-solar-forecasting-technology
[11] S. Lu *et al.*, "Machine learning based multi-physical-model blending for enhancing renewable energy forecast-improvement via situation dependent error correction," in *Proc. Eur. Control Conf. (ECC)*, Linz, Austria, 2015, pp. 283–290.
[12] J. Zhang *et al.*, "Baseline and target values for regional and point PV power forecasts: Toward improved solar forecasting," *Solar Energy*, vol. 122, pp. 804–819, Dec. 2015.
[13] F. Mesinger *et al.*, "North American regional reanalysis," *Bull. Amer. Meteorol. Soc.*, vol. 87, no. 3, pp. 343–360, Mar. 2006.
[14] J. S. Stein, "The photovoltaic performance modeling collaborative (PVPMC)," in *Proc. IEEE 38th Photovoltaic Specialists Conf. (PVSC)*, Austin, TX, USA, 2012, pp. 003048–003052.
[15] A. Florita, B.-M. Hodge, and K. Orwig, "Identifying wind and solar ramping events," in *Proc. IEEE 5th Green Technol. Conf.*, Denver, CO, USA, 2013, pp. 147–152.
[16] M. Cui *et al.*, "An optimized swinging door algorithm for identifying wind ramping events," *IEEE Trans. Sustain. Energy*, vol. 7, no. 1, pp. 150–162, Jan. 2016.
[17] A. Forrester, A. Sobester, and A. Keane, *Engineering Design via Surrogate Modelling: A Practical Guide*. New York, NY, USA: Wiley, 2008.
[18] B.-M. Hodge, A. Florita, J. Sharp, M. Margulis, and D. Mcreavy, "The value of improved short-term wind power forecasting," Nat. Renewable Energy Lab., Golden, CO, USA, Tech. Rep. NREL/TP-5D00-63175, Feb. 2015.
[19] J. K. Williams, P. P. Neilley, J. P. Koval, and J. McDonald, "Adaptable regression method for ensemble consensus forecasting," in *Proc. 30th AAAI Conf. Artif. Intell.*, Phoenix, AZ, USA, 2016, pp. 3915–3921. [Online]. Available: http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/12492
[20] *Flexible Energy Scheduling Tool for Integrating Variable Generation (FESTIV) Model*. Accessed on Aug. 2016. [Online]. Available: http://www.nrel.gov/ electricity/transmission/festiv.html
[21] M. Cui, J. Zhang, H. Wu, and B.-M. Hodge, "Wind-friendly flexible ramping product design in multi-timescale power system operations," *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 1064–1075, Jul. 2017.
[22] N. Jaleeli, L. S. VanSlyck, D. N. Ewart, L. H. Fink, and A. G. Hoffmann, "Understanding automatic generation control," *IEEE Trans. Power Syst.*, vol. 7, no. 3, pp. 1106–1122, Aug. 1992.
[23] E. Ela and M. O'Malley, "Studying the variability and uncertainty impacts of variable generation at multiple timescales," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1324–1333, Aug. 2012.
[24] M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems*. New York, NY, USA: Wiley, 2002.
[25] *The IEEE 118-Bus 54-Unit 24-Hour System*. Accessed on Aug. 2016. [Online]. Available: http://motor.ece.iit.edu/data/JEASIEEE118.doc
[26] H. Wu, M. Shahidehpour, Z. Li, and W. Tian, "Chance-constrained day-ahead scheduling in stochastic power system operation," *IEEE Trans. Power Syst.*, vol. 29, no. 4, pp. 1583–1591, Jul. 2014.
[27] *GAMS/SCENRED Documentation*. [Online]. Available: http://www.gams.com/help/index.jsp
[28] ILOG CPLEX. (2014). *ILOG CPLEX Homepage*. [Online]. Available: http://www.ilog.com
[29] W. Lin, J. Wen, S. Cheng, and W.-J. Lee, "An investigation on the active-power variations of wind farms," *IEEE Trans. Ind. Appl.*, vol. 48, no. 3, pp. 1087–1094, May/Jun. 2011.



Mingjian Cui (S'12–M'16) received the B.S. and Ph.D. degrees in electrical engineering and automation from Wuhan University, Wuhan, China, in 2010 and 2015, respectively.

He is currently a Research Associate as a Post-Doctoral with the University of Texas at Dallas. He was also a Visiting Scholar with the Transmission and Grid Integration Group, National Renewable Energy Laboratory, Golden, CO, USA, from 2014 to 2015. His research interests include renewable energy forecasting, power system operation and control, unit commitment, economic dispatch, optimization modeling, electricity market, data analytics, and statistical analysis.

Jie Zhang (M'13–SM'15) received the B.S. and M.S. degrees in mechanical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2006 and 2008, respectively, and the Ph.D. degree in mechanical engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 2012.

He is currently an Assistant Professor of the Department of Mechanical Engineering at The University of Texas at Dallas. His research interests include multidisciplinary design optimization, complex engineered systems, big data analytics, wind and solar forecasting, renewable integration, and energy systems modeling and simulation.

Bri-Mathias Hodge (M'10–SM'17) received the B.S. degree in chemical engineering from Carnegie Mellon University in 2004, the M.S. degree from the Process Design and Systems Engineering Laboratory of Åbo Akademi, Turku, Finland, in 2005, and the Ph.D. degree in chemical engineering from Purdue University in 2010.

He is currently the Manager of the Power System Design and Studies Group, National Renewable Energy Laboratory, Golden, CO, USA. His current research interests include energy systems modeling, simulation, optimization, and wind power forecasting.

Siyuan Lu received the B.S. degree in physics from Fudan University, Shanghai, China, and the Ph.D. degree in physics from the University of Southern California, Los Angeles, CA, USA, in 2001 and 2006, respectively.

He is currently a Manager and a Research Staff Member leading the Data Intensive Physical Analytics Research Group, IBM T.J. Watson Research Center, Yorktown Heights, NY, USA. His research interests include architectures of big geospatial data services and the union of physics and data-driven approaches for modeling complex systems with applications in renewable energy forecasting, climate forecasting, and remote-sensing based land surveying and environmental monitoring.

Hendrik F. Hamann (M'01) received the Ph.D. degree from the University of Göttingen, Göttingen, Germany, in 1995.

He is currently a Senior Manager and a Distinguished Research Staff Member with IBM T.J. Watson Research Center, Yorktown Heights, NY, USA. In 1999, he joined the IBM T.J. Watson Research Center, where is leading the Physical Analytics and cognitive IoT program. He has authored and co-authored over 90 peer-reviewed scientific papers and holds over 90 patents and has over 100 pending patent applications. His current research interest includes sensor networks, sensor-based physical modeling, renewable energy, energy management, precision agriculture, system physics, and big data technologies. He was a recipient of several awards, including the 2016 AIP Prize for Industrial Applications of Physics. He is an IBM Master Inventor, a member to the IBM Academy of Technology and has served on governmental committees such as the National Academy of Sciences, the National Science Foundation and as an industrial advisor to Universities. He is a member of the American Physical Society, Optical Society of America, and the NY Academy of Sciences.