

How Can Probabilistic Solar Power Forecasts Be Used to Lower Costs and Improve Reliability in Power Spot Markets? A Review and Application to Flexiramp Requirements

BENJAMIN F. HOBBS¹ (Life Fellow, IEEE), **VENKAT KRISHNAN²** (Member, IEEE),
JIE ZHANG³ (Senior Member, IEEE), **HENDRIK F. HAMANN⁴** (Senior Member, IEEE),
CARLO SIEBENSCHUH⁴ (Student Member, IEEE), **RUI ZHANG⁴**,
BINGHUI LI³ (Member, IEEE), **LI HE³** (Member, IEEE), **PAUL EDWARDS²**,
HAIKU SKY², **IBRAHIM KRAD²** (Member, IEEE), **EVANGELIA SPYROU²** (Member, IEEE),
XIN FANG² (Senior Member, IEEE), **YIJIAO WANG¹**, **Q. XU¹**,
AND SHU ZHANG¹ (Member, IEEE)

¹Robert O'Connor Sustainable Energy Institute, Johns Hopkins University, Baltimore, MD 21218 USA

²National Renewable Energy Laboratory, Golden, CO 80401 USA

³Department of Mechanical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA

⁴IBM Thomas J. Watson Center, Yorktown Heights, NY 10598 USA

CORRESPONDING AUTHOR: B. F. HOBBS (bhobbs@jhu.edu)

This work was supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy under Solar Forecasting II DE-FOA-0001649, Topic Area 3: Power Forecasts and Operational Integration, under Agreement EE0008215.

ABSTRACT Net load uncertainty in electricity spot markets is rapidly growing. There are five general approaches by which system operators and market participants can use probabilistic forecasts of wind, solar, and load to help manage this uncertainty. These include operator situation awareness, resource risk hedging, reserves procurement, definition of contingencies, and explicit stochastic optimization. We review these approaches, and then provide a case study in which a method for using probabilistic solar forecasts to define needs for reserves is developed and evaluated. The case study has three parts. First, we describe building blocks for enhancing the Watt-Sun solar forecasting system to produce probabilistic irradiance and power forecasts. Second, relationships between Watt-Sun forecasts for multiple sites in California and the system's need for flexible ramp capability (flexiramp) are defined by machine learning and statistical methods. Third, the performance of present methods to defining flexiramp requirements, which are not conditioned on weather and renewables forecasts, is compared with that of probabilistic solar forecast-based requirements, using a multi-timescale production costing model with an 1820-bus representation of the WECC power system. Significant potential savings in fuel and flexiramp procurement costs from using solar-informed reserve requirements are found.

INDEX TERMS Power markets, probabilistic solar forecasts, renewable energy, operating reserves, California, flexiramp.

I. INTRODUCTION

MANY jurisdictions have adopted goals of largely or completely decarbonizing the power sector by the 2030 to 2050 time frame [27]. Electrification based on renewables will play a major role in meeting these goals.

As we pursue this clean energy transformation, how can probabilistic resource and load forecasts help to cost-effectively integrate renewables [43]? To address this question, we first provide a focused survey of ways which such forecasts can inform decisions by operators and market

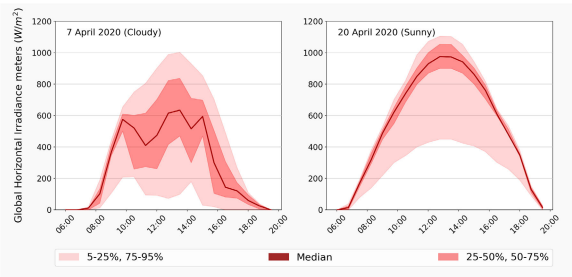


FIGURE 1. Two hour-ahead probabilistic forecasts of solar irradiance for two days for los angeles, CA from the IBM probabilistic Watt-Sun forecasting system. (Image courtesy of IBM. Reproduced with permission from [22].)

participants, and then describe a specific application to dynamic reserve procurement.

Fig. 1 illustrates an example of probabilistic forecasts, giving selected quantiles for a 2-hour ahead solar irradiance forecast. These forecasts are in the form of marginal probability distributions for particular points in time; alternatively, they can consist of collections of possible scenarios of relevant variables. Such characterizations of uncertainty have the potential to support the following five operations and market activities:

- *Situational awareness* by control room operators and energy management systems.
- *Risk characterization* for individual renewable resources and load serving entities, who can use that information to adjust spot market positions and choose hedges.
- *Procurement of operating reserves* of various types (e.g., regulation, spinning reserves, replacement reserves, and flexible ramp product, which we call FRP or “flexiramp” for brevity), along with definition of demand curves for those reserves that reflect, for instance, the likelihood of load balance constraint violations and their consequences.
- *Economic scheduling* of energy resources considering the possible time evolution of net load contingencies representing deviations from forecasts, in order to endogenously choose where and how much reserves to procure.
- *Stochastic programming*-based resource scheduling considering possible net loads and their probabilities.

All but the last activity (stochastic programming) are, to varying extents, a part of existing scheduling processes used by North American independent system operators (ISOs). Additionally, it is anticipated that the U.S. DOE ARPA-E PERFORM initiative [25] will stimulate development of additional approaches to manage risk using probabilistic forecasts.

In this paper, we first review how probabilistic forecasts can assist the above five classes of uncertainty management activities in spot markets. Previous reviews (e.g., [36], [60]) have examined the many ways in which probabilistic forecasts can inform operations, without providing details

on particular applications. In contrast, our review succinctly focuses on the present and potential roles of probabilistic forecasts in reserve determination and hedging, followed by an in-depth case study. This case describes how an enhanced probabilistic solar forecast system can be dynamically linked to reserve needs, and how the system benefits of this linkage can be quantified.

The particular application is the dynamic procurement of flexiramp, a new type of operating reserve. Traditionally, operating reserve requirements have been based on deterministic load and variable renewable forecasts augmented by a statistical analysis of historical net load variability and forecast uncertainty, area control error, and other stochastic variables [26]. These practical methods do not account for how the uncertainty that a system actually faces can vary daily or even hourly based on weather. Our case study shows how probabilistic forecasts can plug this key information gap by providing risk indices that help operators reduce reserve procurement costs when uncertainty is low, while alerting them to situations in which more reserves are needed to maintain system reliability [12].

More specifically, our case study addresses the potential use of probabilistic solar forecasts to define requirements for the California ISO’s (CAISO’s) flexible ramp product. First, we enhance an existing solar forecasting system to provide well-calibrated 2 hr-ahead probabilistic forecasts. We then relate the degree of uncertainty in those forecasts to error distributions for net load ramps for the CAISO using statistical and machine learning methods. Distributions of net load errors conditioned on solar uncertainty are translated into flexiramp requirements that consequently reflect “day-of” meteorological and solar conditions, improving on typical ISO procedures. Focusing on flexiramp, we then use a multi-settlement production simulation model to quantify how conditional ramp requirements can improve operations by (1) decreasing operating costs by reducing requirements, compared to often conservative unconditional methods, while also (2) reducing the likelihood of supply scarcity by increasing flexiramp procurement at times when unconditional requirements understate actual ramp uncertainty and the need for flexibility.

Next, in Section II, we elaborate on the above five classes of market applications of probabilistic forecasts, emphasizing the definition of reserve requirements. Section III then describes our potential application to flexiramp procurement in the CAISO and its potential benefits. Section IV offers conclusions.

II. POSSIBLE APPLICATIONS OF PROBABILISTIC SOLAR FORECASTS IN POWER MARKETS

Here we review five general uncertainty management activities in spot power markets that can benefit from incorporation of probabilistic forecasts. We point out several existing and potential frameworks for taking advantage of such forecasts, focusing on acquisition of reserves and market hedging needs. Our review is intended to provide a brief but focused

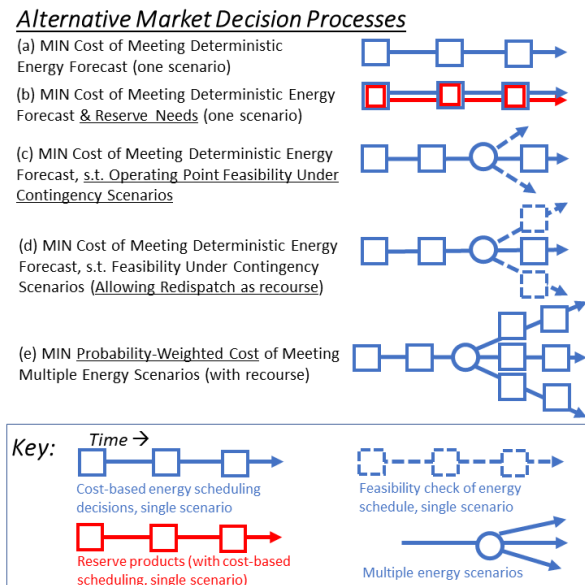


FIGURE 2. Decision tree representation of alternative spot market processes, from deterministic energy scheduling to full stochastic programming. See text.

overview of these particular uses rather than exhaustively survey the literature on probabilistic forecasts and potential uses. Broader discussions of applications are provided in [24], [36], [60], and [63].

As a framework for this discussion, Fig. 2 presents schematics of scheduling decision processes in spot markets. These range from simple energy scheduling against a deterministic forecast to full stochastic programming (see also [14]). Time proceeds from left to right; squares are decision nodes (generator commitment and dispatch), while circles are chance nodes that indicate that more than one scenario of net load can occur. Blue lines represent energy schedules, while red represents operating reserve schedules. Solid lines extending from a chance node indicate that subsequent costs are weighted by each scenario’s probability in the model’s objective function, while dashed lines imply that the model just considers feasibility and not costs of the schedule under that contingency scenario.

For instance, in Fig. 2(a), ISO scheduling decisions in day-ahead and real-time markets are simplistically represented as a sequence of decisions (one set of decisions per market interval) that schedule supply offers against either bid-in demand or a single deterministic forecast of net load at each bus in the network. There are no uncertainties or contingencies in that relatively simple process. At the other extreme, Fig. 2(e) schedules near-term generation while accounting for possible scenarios of net load that occur later, and how the system would subsequently adjust dispatch and commitment to adapt to each.

A. IMPROVING SITUATIONAL AWARENESS IN THE CONTROL ROOM [24]

Operators who are running market software to schedule energy and reserves against a deterministic net load (as in

Fig. 2(a)) of course must still recognize net load uncertainties. One way they do so is by maintaining a keen awareness of system developments over the day, so that they can manually adjust resource schedules to meet unexpected net load changes.

Control centers typically host several screens for visualizing network situational awareness, such as switch status and available reserves. Information is displayed about possible high-impact contingencies, as well as their effect on transmission line overloads, and, in some cases, voltage and transient stability [41]. The idea of situational awareness in the form of real-time visualization, event alerts, and plausible control locations/parameters is not new; however, enhancements to include probabilistic forecasts of variable renewable generation are desirable as renewable penetration and associated uncertainties grow.

With increasing uncertainty and emergent dynamic events, the awareness that is conventionally reported at aggregated levels is insufficient. Additional capabilities are needed, including:

- Innovative ways to integrate renewable forecast data—including higher spatial resolutions and distributed resources.
- Probabilistic forecast data integration that quantifies renewable resource uncertainties, which will be key to operator assessments of risk associated with system dispatch.
- Timely alerts to excessive ramping events for net load.

Fig. 3 shows an example display of such information from an open-source tool RAVIS (Resource Forecast and Ramp Visualization for Situational Awareness) [18], which we developed as part of the project described in Section III. RAVIS is intended to help forecast vendors and operators (such as utilities, ISOs, and balancing authorities) to integrate probabilistic and other advanced forecasts for variable renewables into their operations, and thereby develop situational awareness and timely mitigation strategies. A challenge is that probabilistic forecasts have much more information than median forecasts. Therefore, displays must be designed to display probabilities intuitively and simply, and to help operators quickly sort through large amounts of data to obtain the information they need.

B. MANAGING RESOURCE RISK BY OWNERS

Of course, even if the ISO uses a deterministic approach like Fig. 2(a) to schedule resources against a single net load forecast, risks remain, and market participants must reckon with them. Resource owners need to weigh the risks they face in deciding how much energy to offer, and at what price, in the day-ahead market, given uncertainties in real-time prices and the possibility of equipment outages or erroneous production forecasts.

Probabilistic forecasts offer resource owners a more sophisticated understanding of their physical supply risks. Advanced systems can also show how their output is

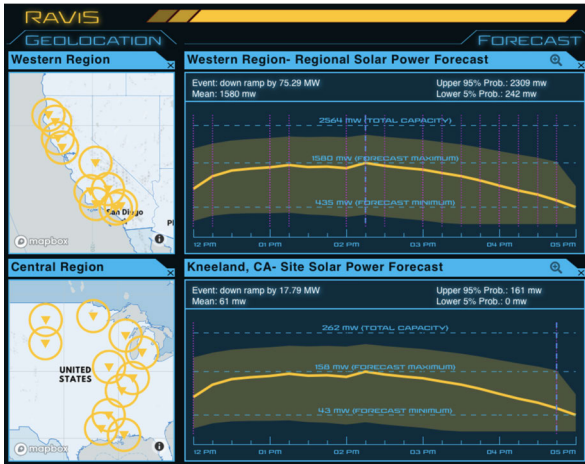


FIGURE 3. Excerpts from sample RAVIS tool screen for control room situational awareness: site-specific and power ramp alerts and time-series forecasts with confidence bands for solar production for California and U.S. [18].

correlated to overall market net load and prices [6]. Resource owners can then tailor forward schedules and virtual positions in the ISO’s day-ahead market to the conditions on that particular day. Probabilistic real-time price forecasts can also facilitate valuation of risk management instruments, such as congestion revenue rights or the novel financial products proposed under the ARPA-E PERFORM program (e.g., [51]).

Probabilistic forecasts can also inform purchases of hedges in the form of bilateral options and forward obligations. For instance, an appropriately configured hedge can put a ceiling on prices paid for imbalances when resources fall short of day-ahead schedules. Probabilistic forecasts can also contribute to better informed decisions about how to create or preserve physical hedges and options, such as charging batteries in a hybrid plant or committing a flexible generator.

Biases can arise from ISO use of deterministic rather than stochastic day-ahead unit commitment models, for example against committing flexible units with high start-up costs. However, if probabilistic forecasts of net load and prices are available, many of these biases can be corrected by individual market participants, for instance by virtual bidding together with generator self-commitment [29]. As a result, overall market efficiency resulting from a deterministic ISO scheduling process can, in theory, approach that of the sophisticated two-stage stochastic optimizations of Section II.E.

C. DEFINING REQUIREMENTS FOR OPERATING RESERVES

All ISO markets run day-ahead and real-time energy markets, and also simultaneously procure reserve products, usually considering just a single deterministic net load forecast or amount of cleared demand bids (Fig. 2(b)). There are many types of reserve products, such as up and down regulation,

up and down flexible ramp, residual unit commitment, and spin, non-spin, and replacement reserves. Present practice involves setting requirements considering explicit contingencies (such as the single largest contingency, in the case of spinning reserves), percentages of the solar and wind forecast, or a specified quantile of the probability distribution of need. “Need” might be measured, e.g., by net load ramp uncertainty (in the case of flexible ramp product, see Sect. III) or by extreme values of adjusted area control error (ACE) values expected in a five-minute real-time interval (in the case of regulation).

When basing requirements on a quantile of a distribution of need, ideally the choice of quantile (e.g., the 97.5th percentile) is justified by a careful balancing of the incremental benefits of more reserves (in terms of reducing the probability weighted consequences of being short of reserves) against the incremental costs of procuring and, if needed, deploying the reserves [42], [45]. ISOs have been moving in this direction. The concept of an operating reserve demand curve, in which higher prices are paid for reserves as the amount procured falls short of a base requirement, has been adopted in several ISO markets. Examples include overall reserves in the Electric Reliability Council of Texas (ERCOT) and for flexiramp in the Midcontinent ISO (MISO) and CAISO [57]. The willingness to pay for additional reserves is directly related to a calculation of the conditional probability of insufficient reserves and violation of the power balance in real-time. Probabilistic forecasts of reserve availability and net load would be potentially useful inputs to creating such demand curves; this has been demonstrated in the case of non-spinning reserves for the ERCOT market [21].

In addition to balancing benefits and costs when setting requirements, ISOs should condition the probability distributions of need upon dynamic information about weather and system conditions. ISOs are beginning to move away from static methods (such as the CAISO’s use of histogram of the last 40 days of weekday data to estimate weekday ramp errors [9]). In theory, conditional net load distributions could be derived by appropriately convolving separate probabilistic forecasts of wind, solar, and gross load, accounting for their correlations [37].

However, two other approaches have proven more practical. One derives a suite of possible scenarios of the evolution of net load by fitting models that relate observed net load forecasting errors to, e.g., causal weather variables and renewable forecasts [11], [35]. Proposed reserve requirements could then be compared to the set of scenarios before running the market software to see if the amount of reserves would be sufficient to meet a target level of reliability. This approach has often been proposed by researchers (e.g., [50]), and has been tested by utilities (e.g., Hawaii Electric [42]). Other examples of this approach include the following. Ref. [13] probabilistically predicts wind power ramp events by generating scenarios of wind power generation. Ref. [12] transforms probabilistic solar forecasts into scenarios which are then translated into dynamic operating reserves. Ref. [3]

applies a set of machine learning (ML) methods to generate and reduce scenarios from probabilistic solar forecasts. Compared to forecasts of marginal distributions at discrete points in time, dynamic scenarios also have the distinct advantage of also being able to evaluate whether there is enough energy available (e.g., battery charge) over multiple market intervals, if scenario construction accounts for autocorrelations of forecast errors.

The second practical approach is to use ML or quantile regression approaches to estimate distributions of net load or ACE as a function of weather conditions and forecasts of net load components. The general idea is to fit a mapping in the training stage between deployed reserves and model inputs (e.g., system states, meteorological parameters, probabilistic renewable forecasts, etc.). Typical ML algorithms used for reserve sizing include clustering (e.g., k -means [5], [15], [46]), regression methods (e.g., multiple regression [61], support vector regression, gradient boosting machines, random forests [34]), and deep learning (e.g., artificial neural networks [28], [52], convolutional neural networks [34], and extreme learning machines [62]). Other data-driven methods, such as Bayesian belief networks, can also be used [20]. These methods have been used to estimate requirements of regulation, spinning, and non-spinning reserves. For example, Ref. [15] applies k -means and k -nearest neighbors (kNN) to estimate reserve requirements in Belgium. Ref. [4] employs a clustering approach to derive dynamic reserves by convolving conditional load, wind and solar forecast errors, and plant outage distributions. Multiple regression has been used [61] to estimate regulating reserve requirements based on load and wind power forecasts. Our previous study [34] compares a set of ML methods with a kNN-based method for predicting flexiramp needs, and finds that the kNN method performs better.

The CAISO is planning to replace its histogram-based flexiramp error estimates with quantile regressions that relate, for instance, the 97.5th percentile of net load errors to deterministic forecasts of the amount of load, wind, and solar output [7]. In Section III, we show how using prediction intervals (e.g., the difference between the 25th and 75th percentiles) for solar can inform the setting of flexiramp requirements for the CAISO, potentially lowering the total amount needed, but also improving the reliability with which reserves cover the realized need.

Related to the above use of probabilistic forecasts to define *ex ante* reserve needs *before* running a market model is the idea of using chance-constrained programming to *endogenously* calculate what reserves or other decisions need to be made to ensure against low probability adverse outcomes [17]. Probabilistic forecasts of key inputs would then be required. Another approach to endogenously optimizing reserve levels is robust optimization, which uses so-called uncertainty sets rather than probability distributions as inputs [2]. Probabilistic forecasts can inform decisions about the size of uncertainty sets.

D. ASSESSING SCHEDULE FEASIBILITY UNDER MULTIPLE SCENARIOS

Using reserves to manage uncertainty poses several challenges. One is that, over time, there has been a tendency for operators to introduce additional products in order to address particular needs that existing products do not handle well, such as steep ramps. This proliferation can decrease transparency, and the job of market designers can feel like a game of whack-a-mole.

Another challenge is that procurement of reserve capacity on a system or zonal basis can result in over-procurement in some places and under-procurement elsewhere, so that congestion prevents deploying reserves when needed. There is an inherent bias towards procuring reserves where deployment would be difficult, since the opportunity cost of reserves is naturally lower in generation pockets already experiencing congestion.

One approach to dealing with these challenges is to replace—in part or entirely—explicit reserve requirements with a set of contingency constraints. Forecasting methods that generate scenarios of net load errors or equipment failures (see Sect. II.C) are crucial to implementing this approach. Each contingency represents a scenario in which the market optimization tests whether the resource schedule can still satisfy load under a particular outage or net load outcome. The most familiar version (e.g., $n - 1$ constraints) simply asks whether post-contingency flows resulting from the scheduled operating point remain feasible with respect to network constraints. We show this in Fig. 2(c), where one of the chance node's outcomes is the deterministic energy forecast used to schedule resources (solid line), while other outcomes (dashed lines) are contingency scenarios that check the feasibility of the chosen schedule.

Improvements in computation allow a second, more complex approach, in which post-contingency redispatch of resources is explicitly modelled as decision variables to be optimized. These variables are not assigned costs in the objective function; to check the feasibility of post-contingency redispatch is their only reason to exist. The price of modelling this flexibility is a larger model that includes additional dispatch variables and associated constraints for each contingency, but the result may be much lower costs of managing the contingency [8]. Fig. 2(d) thus shows a decision node (representing optimized redispatch to maintain feasibility) on each contingency's dashed line.

In theory, a “complete” set of contingencies could completely replace capacity reserve constraints, but the curse of dimensionality makes that impractical. A hybrid approach that procures a smaller amount of traditional reserves while also including just a few critical contingencies may be a practical and helpful compromise. For instance, the CAISO has proposed implementing such hybrids in two situations. One addresses certain types of network outages after which the CAISO is required to return to a safe operating point within a specified time (20-30 minutes); the CAISO proposes

to optimize the simulated redispatch rather than assume that the required up-dispatch comes from predetermined reserved capacity [8]. In the second situation, the CAISO is proposing to include two deployment scenarios for its flexiramp product: an “up” scenario in which net loads are increased across the board and the resulting redispatch of flexiramp is calculated, and a “down” scenario in which that redispatch is used to meet net loads that have been decreased from the base values [9]. A possible enhancement would be to use probabilistic methods to generate additional scenarios defining possible extreme cases that reflect correlations among net load forecast errors among all the buses in the system.

E. STOCHASTIC OPTIMIZATION TO PROCURE OPTIMAL AMOUNTS AND LOCATIONS OF RESERVES IN MARKET SOFTWARE

Stochastic multistage programming can be viewed as an extension of multi-scenario methods of Sections II.C,D in which probabilities are assigned to scenarios of, e.g., wind and solar output or net loads, and the objective is to minimize probability weighted costs across scenarios. Many researchers propose this approach as a rigorous method for endogenous determination of locations and amounts of reserves to manage uncertain net loads in unit commitment and dispatch problems (e.g., [47], [54], [56]). Indeed, it is undeniable that, in theory, probability-weighted costs would be minimized by optimizing immediate “here-and-now” decisions considering the many ways that uncertainties can unfold over the time horizon and how later “wait-and-see” decisions would optimally respond to forecast errors. Fig. 2(e) shows how these two types of decisions are related; after the initial decision, there are sets of scenarios issuing from a chance node, each with a probability and subsequent set of recourse decisions and associated costs. (We only show one chance node there; more generally, sequences of chance nodes could represent the random evolution of net load over the day.)

But as has been pointed out [16], [48], this theoretical point does not imply that practical implementations by ISOs of stochastic programming in spot markets would actually improve schedules. This is because the sheer number of uncertain variables and decision stages cause an exponential explosion in problem size, and compromises have to be made to simplify the problem so it can be solved. Although computationally clever implementations of stochastic programming have greatly improved the efficiency of hydropower operations around the world (e.g., [40]), such practical success has been elusive for high dimensional unit commitment and dispatch problems of the size typically solved by ISOs. Another challenge is the need for scenario probabilities, including for extreme events whose probabilities are subject to high sample error and might be nonstationary due to, e.g., climate change. A lack of accepted and transparent methods for calculating such probabilities is a barrier to acceptance by stakeholders, many of whom already complain about the opaqueness of market processes.

Despite these obstacles, the theoretical advantages of stochastic programming for rigorous evaluation of system flexibility and diversity are appealing, and research has shown potential for significant benefits [47]. Like Section II.D’s multi-scenario methodology, a hybrid “belt and suspenders” approach, in which a few crucial scenarios are considered, but some reserve capacity requirements are retained to cover other possibilities, may represent an optimal use of limited computing resources that may both lower system costs and increase reserve effectiveness. The fact that ISOs are already implementing post-contingency redispatch models (Sect. II.D) means that the stochastic programming camel’s nose is already under the tent—in that extending Fig. 2(d) models to a full stochastic formulation requires only inserting probabilities and costs in the objective for already existing variables.

Probabilistic forecasts are critical to implementing stochastic programming-based scheduling models in order to construct the “event trees” that describe how stochastic net load processes evolve over time. If such models are to be used, then present probabilistic forecasting methods will need to be significantly revised, since their time horizons and data outputs often are based on availability of relevant meteorological inputs, rather than the needs of system operators.

III. AN EXAMPLE OF LINKING PROBABILISTIC SOLAR FORECASTS TO RESERVE REQUIREMENTS DEFINITION

We now provide a detailed example to show the practicality and potential benefits of using probabilistic solar forecasts to create weather-informed operating requirements for inclusion in market models. Probabilistic solar forecasts for multiple sites in California from the IBM Watt-Sun probability forecasting system (see Sect. III.A) are input to statistical and ML models that predict quantiles of error distributions for forecast real-time (fifteen minute) ramp needs (Sect. III.B). We then evaluate the magnitude of resulting cost improvements for a sample of days using an 1820-bus model of the Western Electricity Coordinating Council (WECC) system, compared to using weather-independent ramp requirements based on an unconditional histogram of past forecast errors. This is done by simulating the WECC region’s scheduling processes using NREL’s *Flexible Energy Scheduling Tool for Integrating Variable Generation* (FESTIV) platform [19], yielding day-ahead and real-time schedules and costs. (Sect. III.C).

Fig. 4 gives an overview of our proposed probabilistic forecast-informed process. Numerical weather predictions and historical data are input to the Watt-Sun forecasting method, which produces probabilistic solar forecasts. Using this and other weather information, we can then project FRP and other reserve needs that are then input to the ISO’s scheduling processes. (For brevity, we emphasize the role of FRP requirements below.) Finally, historical ramp and weather data are accumulated day-to-day, and can be used to dynamically update relationships between solar forecasts and reserve needs.

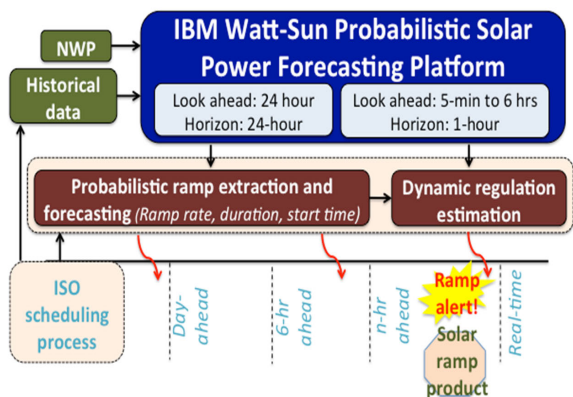


FIGURE 4. Overview of proposed process of developing and applying probabilistic solar forecast-informed flexible ramp product requirements.

The flexible ramp product that is the focus of our analysis is a new type of operating reserve that three ISO markets (CAISO, MISO, and Southwest Power Pool) use to manage uncertainties in net load changes from interval to interval. The fundamental idea is to pre-position resources in one real-time market interval so that if the net load change turns to be either much higher or lower than anticipated in the next interval, the system can still feasibly and economically satisfy load. FRP can be viewed as a type of noncontingent spinning reserve that is acquired in one interval and can be used for energy production in the next if needed. Although details of implementation differ among the three ISOs using the product, their processes attempt to characterize the degree of uncertainty in real-time net load ramps both in the up and down directions, and then ensure that there is enough capacity to meet loads in either event with a predetermined degree of confidence. We hypothesize that the cost and reliability performance of the CAISO FRP could be improved by conditioning FRP requirements on weather conditions of the day, especially as reflected in the uncertainty in solar forecasts.

A. IBM WATT-SUN PROBABILISTIC FORECASTING SYSTEM

The field of big data-driven probabilistic solar forecasting is evolving fast, driven by multiple trends including rapidly growing renewables and adoption of new decision-making processes by grid operators [53]. Solar forecasting research can be divided into (a) advancing the underlying physics in numerical weather prediction models, (b) developing non-physical but data-driven forecasting approaches that apply statistics, ML, and artificial intelligence to historical data, and (c) combinations of these physical and data-driven approaches [1]. The creation of probabilistic Watt-Sun provides an example of the third method which leverages advanced data acquisition and integration; scalable distributed computation; and validation to inform ML models for probabilistic forecasts.

The original version of Watt-Sun was deterministic, yielding forecasts of median solar insolation and power [38]. The

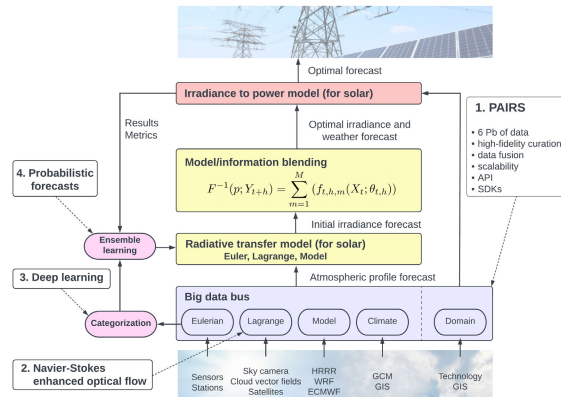


FIGURE 5. The four building blocks of the Watt-Sun Forecasting System that enable probabilistic forecasts.

flowchart within Fig. 5 summarizes the basic mechanics of the deterministic Watt-Sun forecasting method. The system is supported by a big data curation engine that integrates data from multiple numerical weather prediction models, plus shorter term models that forecast cloud movement. Then, multiple models are blended using a measurement-informed ML model, creating more accurate forecasts than any individual weather and cloud forecasting model. Predicted solar irradiance is then converted to power. The accuracy of the forecasts is then fed back to a ML approach that further optimizes model blending.

For the present project, IBM extended Watt-Sun to produce not only the median but also other quantiles (as in Fig. 1). Fig. 5 summarizes four building blocks that made this extension possible. The first building block is the replacement of the data management system of Watt-Sun with PAIRS (Physical Analytics Integrated Data Repository and Services), a completely scalable platform for geospatial-temporal data based on a Hadoop/Hbase cluster, allowing distributed and scalable processing [31], [39]. PAIRS improved speed and data processing throughput more than 50-fold compared to the previous system. PAIRS enables “automatic” fusing of satellite, weather, and sensor data; supports handling of tens of petabytes; and can inject data much faster than previously possible (eventually up to hundreds of TB/day). Crucially, PAIRS can also distribute forecast data in a scalable matter.

The second building block is a new short-term solar forecasting module, enhancing “convection-based” forecasting-based Geostationary Environmental Satellite (GOES) satellite observations with a 2-D Navier-Stokes equation. We leveraged the new GOES-R data in this project, which have significantly better spatial and temporal and spectral resolution than the previous GOES (–13/–14) satellite data used in the earlier deterministic Watt-Sun. Watt-Sun also now leverages real-time cloud information from GOES-16 [44], enabling better short-term Lagrangian-based solar forecasts.

A third building block is distinct model blendings for different weather situations. Given multimodal data input from IBM PAIRS, Watt-Sun distinguishes common weather

situations by data-driven categorization (namely volumetric convolution [59], which can identify situations based on full images or raster observations from, e.g., GOES). Then Watt-Sun assesses the performance of a wide array of prediction models for different categories. Specifically, a variational autoencoder is trained to infer the future weather situation onto a latent space [30]. That inferred state of the atmosphere and the models' forecasts are inputs to Watt-Sun's model blending (ensemble learning) to derive an ideal global horizontal irradiance (GHI) forecast. But we note that Watt-Sun's latent space of weather categories is continuous rather than discrete, as in vector quantized variational autoencoders [49]. Details are provided in [38] and [39].

The fourth building block extends Watt-Sun's capabilities to include probabilistic estimates of irradiance for points and regions. This was done by implementing linear quantile regression in the last step of the situation-dependent error analysis; this method allows capturing of non-normal distributions.

As a result of these changes, the probabilistic Watt-Sun forecasting system's accuracy increased, significantly outperforming a probabilistic baseline model, namely, a persistence estimator that is augmented with an empirical error distribution.¹ The metric used to evaluate calibration was the probability-probability (PP) metric, which is a metric of how similar the predicted probability distribution is to the actual distribution of errors, as Fig. 6 shows. The comparison showed that Watt-Sun's PP metric was better on average for 79% of the 24 sites tested over the period March-May, 2020. As an index of the reliability of these improvements for 4 hr ahead forecasts, daily values of Fig. 6's PP-based metric of forecast improvement (right panel) were <70% for at least 3 weeks straight during that period for 17 of 24 sites tested in the CAISO and MISO regions. These results are robust with respect to other probabilistic accuracy metrics and stations. Thus, Watt-Sun's improvements are statistically and economically meaningful.

To give a quantitative indication of the contribution of the four building blocks to improving Watt-Sun's probabilistic distributions, we estimate and compare the performance of 7 models of increasing complexity using an out-of-sample (spatial) location. Chicago (IL) was selected as it distant from NOAA's irradiance measurement stations in New England and California that feed into Watt-Sun. Moreover, Chicago is particularly challenging for irradiance forecasts due to persistent wind and rapidly changing cloud cover. We chose 399 days starting in Jan. 2021. Of these, 279 days serve as a training set while the remaining data were split into equally sized validation and test subsets. The latter subset was held

¹Persistence as a baseline predictor is limited in forecasting non-stationary time series data across long horizons. But as Watt-Sun is designed for intraday (GHI, wind speed) prediction, we view a probabilistic version of a persistence estimator that is corrected for the diurnal GHI pattern and augmented by an intraday GHI variability assessment to be reasonably competitive. Further, the usefulness of (slightly modified) persistence measures as a baseline has been noted for general intraday meteorological forecasts [58] and particularly GHI prediction [33], [55].

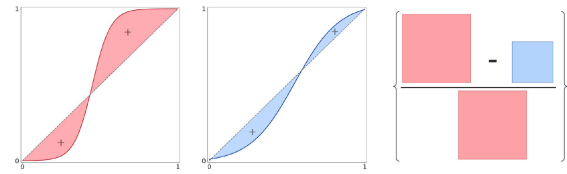


FIGURE 6. Visual interpretation of the PP metric. The size of the shaded area enclosed by the baseline method (PP-plot, left) is compared to that of Watt-Sun (PP-plot, center). The relative improvement of Watt-Sun over the baseline is assessed by the decrease in the shaded area, and visualized as a quotient (right).

out during model calibration while the former was used for hyperparameter tuning. Each predictor is distributional (i.e., yields a conditional cumulative distribution function (CDF)). However, only the most complex of the 7 models implements non-parametric CDF estimation, while the remaining models are conditional Gaussian. To compare the models' probabilistic calibration, our out-of-sample comparison used two metrics: PP and CRPS (continuously ranked probability score), the latter being a key calibration metric as it is a proper scoring rule [23].

The literature has suggested many refinements to the basic persistence approach, so we consider three versions of that method in our comparison. Specifically, we compare naïve persistence (*Persistence 1*, using the previous day's observation and fixed conditional variance estimate), a persistence estimator with adaptive conditional variance called *Persistence 2* (via lagged estimation, where the order of the lag is chosen in a data-driven fashion), and *Persistence 3*, which corrects the previous observation with diurnal and seasonal component while leveraging adaptive conditional variance. We found that *Persistence 3* shrinks CRPS by about 12% compared to both *Persistence 1* and 2, which have similar performance. Note that as we developed Watt-Sun, we consistently compared it to "smarter" persistence estimators, e.g., *Persistence 3*.

Turning to the building blocks used to create probabilistic Watt-Sun, the fourth of seven models in the comparison simulates Watt-Sun's use of PAIRS (building block 1) by adding 12 atmospheric variables to the dataset (e.g., temperature, windspeed); this yielded the most significant CRPS improvement of 26% relative to *Persistence 1* and about half that relative to the "smartest" persistence model (*Persistence 3*). In the fifth model, we add further physical data (from numerical simulations, building block 2), enhancing the forecasting accuracy in terms of CRPS by an additional ~7%. Adding deep-learning-based categorization of weather situations (building block 3) (either through variational autoencoder or unsupervised clustering via hierarchical agglomeration) decreases CRPS by another ~3.5%. Finally, switching to quantile estimation (building block 4) decreases CRPS by around another ~4.4%.

These results confirm that PAIRS is the most crucial of the four building blocks, but that the combined benefit of the last

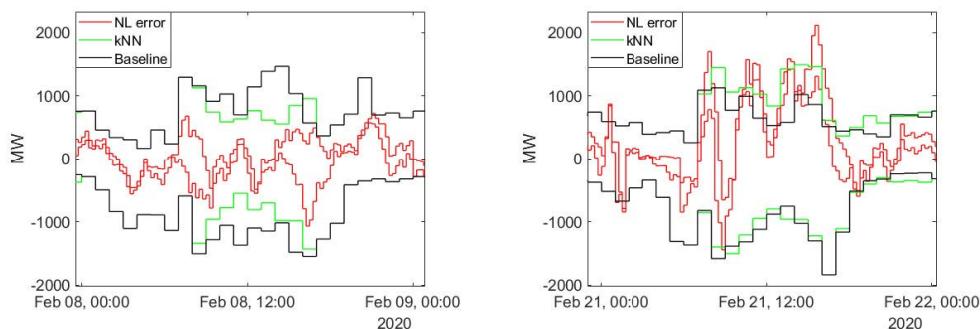


FIGURE 7. Up and down flexiramp requirements for two days by the (unconditioned) baseline requirements method used by the CAISO (black lines) versus (weather-informed) requirements based in part on probabilistic solar forecasts (kNN method, green lines). Actual net load (NL) forecast errors shown in red. Left: Less cloudy day (less forecast error); Right: Cloudier day (more forecast error).

three blocks is similar to using PAIRS. The cumulative effect of all four building blocks resulted in nearly a halving of the CRPS of *Persistence 1*. The analysis was repeated for the PP metric, and broadly similar results were obtained. Overall, this comparison provides evidence that Watt-Sun performs well across probabilistic accuracy metrics (PP, CRPS) and for data that are distributional shifted, spatially and temporarily.

In our next step, we use Watt-Sun forecasts for a geographically distributed subset of up to ten CAISO locations to calibrate models that predict the amount of flexiramp needed to meet uncertain net load ramps in the real-time market.

B. RELATING FLEXIRAMP NEEDS TO PROBABILISTIC SOLAR FORECASTS

As noted in Section II.C, the probability distribution of forecast errors for net load ramps from market interval to interval could be derived by convolution of probabilistic forecasts of the solar, wind, and gross load components of net load. Since this is not yet practical, we instead use ML and statistical methods to empirically relate quantiles of the distribution of net load forecast errors to meteorological variables, especially the width of prediction intervals from probabilistic solar forecasts. (Note that the up (down) “forecast error” for each 15-minute real-time interval being predicted by the CAISO has a specific definition. It is the difference between (a) the max (min) binding interval energy forecast from the 5-minute market over the three 5-minute intervals within the 15-minute interval of interest and (b) the first advisory 15-minute interval forecast from the previous 15 minute interval in the 15-minute real-time market.)

Our search for empirical relationships is inspired by evidence that different solar conditions are associated with different amounts of ramp forecast errors. For instance, Fig. 7 (left) shows that on an uncloudy day, those errors (shown in red) can be relatively smaller than on a somewhat cloudy day (right), when solar output is likely to be less predictable. Under the present CAISO FRP requirements process, the FRP up and down requirements (shown in black) vary from hour-to-hour based on the distribution of errors experienced in recent weeks, but are not conditioned on today’s weather

conditions. That is, whether today is cloudy or not, the black lines would stay the same on a given day. (Note that the requirements for the two days in Fig. 7 do differ because they are 13 days apart.)

Weather-conditioned requirements, by contrast, would shift the black lines on a given day depending on solar uncertainty and other variables. In Fig. 7, relationships between ramp forecast error distributions and weather variables estimated by a ML-based method (k^{th} -nearest neighbor “kNN”) have been exploited to adjust the requirements to account for the degree of cloudiness by using medians and prediction interval widths from probabilistic solar forecasts [34]. These are shown as green lines; consequently, up- and down-FRP requirements are reduced relative to baseline unconditional requirements on less cloudy days when there is less solar uncertainty (left side of figure), but are increased on cloudy days with more solar uncertainty (right side). By reducing requirements when there is less forecast error, money is saved by not procuring unneeded reserves. On the other hand, by increasing requirements when more error is anticipated, reliability of the system is improved and price spikes in the real-time energy market are avoided. Overall, the result can be a less costly and more reliable system.

Fig. 8 displays the overall performance of the kNN probabilistic solar forecast-conditioned FRP requirements (in the upwards direction) versus the CAISO unconditional baseline method. We consider two performance indices: (a) the amount of FRP procured and (b) the reliability of the requirement, in terms of the chance that actual ramps exceed the procured FRP. The target reliability of the CAISO system is that no more than 2.5% of intervals have upward ramp errors that exceed the FRP-up requirement. (The target is also 2.5% for downward ramp errors.) A system with smaller values of both (a) and (b) is preferred. Here, whisker plots (top of figure) reveal that the median FRP is reduced in 6 of 8 hourly intervals, and by as much as 50%. Also, in 6 of 8 intervals, the reliability of the kNN requirements (bottom) is as good or better than the baseline reliability, and is much better in most intervals.

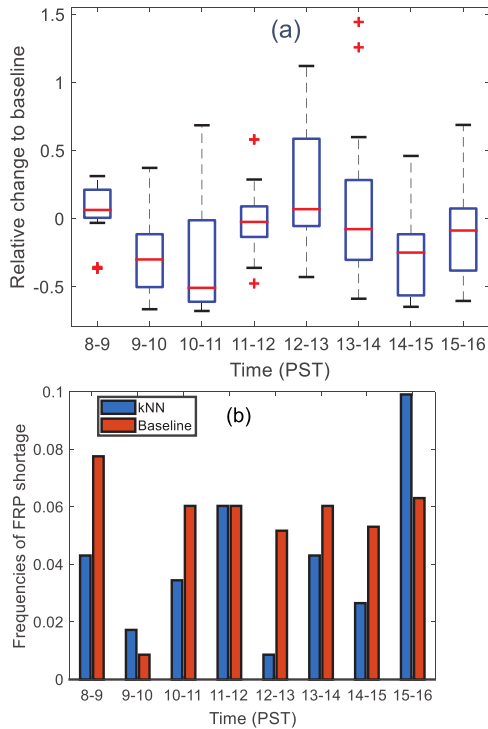


FIGURE 8. Example of performance of kNN-based estimates of upwards FRP requirements (CAISO fifteen-minute ramps occurring in eight 1-hr market intervals, Feb. 2020). Top: Distribution of changes in the portion of the FRP requirement used to meet upward uncertainty relative to the CAISO baseline (unconditional) method. 0 means no change in MW requirement, and a negative value means a decrease. Bottom: Fraction of 15-min ramps (four per hour interval for each of 29 days) that exceed the FRP requirement defined by methods.

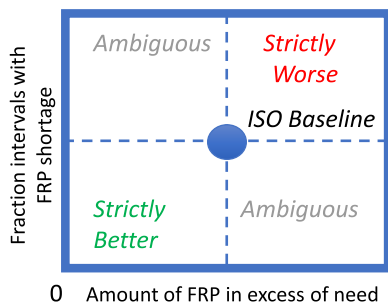


FIGURE 9. Schematic Pareto plot with two objectives: reliability (minimize frequency of FRP shortage on y-axis) and cost (minimize excess FRP, x-axis).

A Pareto plot is another way to show the relative performance of requirements estimation methods, by showing both how often ramp shortages occur and amounts of excess FRP (the amount above what is needed to meet actual ramps) for a given method as a point on an x-y plot (Figs. 9, 10). If a method's point lies southwest of the baseline, it is better in both objectives, whereas points to the northeast are worse in both.

In Fig. 10, the performance of quantile regression (QR), which is a statistical method for estimating the quantile of a

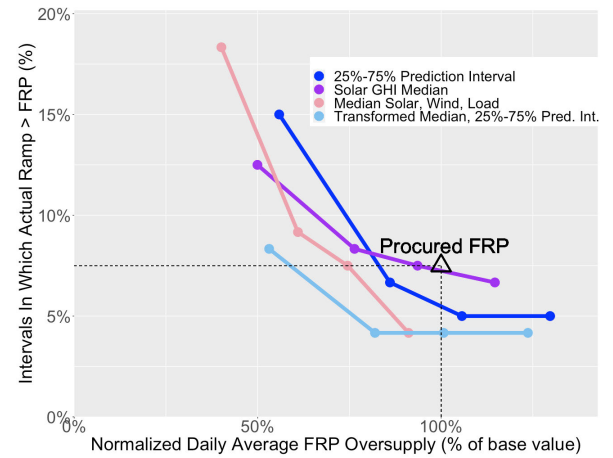


FIGURE 10. Two-objective Pareto plot (reliability (minimize frequency of FRP shortage) vs cost (minimize excess FRP)) showing performance of four quantile regression specifications (for 12:00-13:00 local time, March 2020), compared to performance of CAISO procured FRP. Specifications include: linear using width of 25th – 75th percentile prediction interval for solar global horizontal irradiance (blue); linear using median GHI (purple); linear using median GHI plus CAISO wind and gross load forecast (pink); and linear with GHI prediction interval and a nonlinear (sine) transformation of median GHI (light blue). Each set of solar variables is averaged across four CAISO solar sites. The four points on each curve (upper to lower) correspond to 10%, 5%, 3%, and 1.5% *a priori* frequency of FRP shortage.

distribution of a random dependent variable [32], is assessed by a Pareto diagram. In a three-step process, the QR method estimates what level of upward ramp forecast error corresponds to an *a priori* reliability level (e.g., 2.5% rate of shortages), and is then validated. This is done separately for each operating hour. Although QR does not obviously outperform the kNN-based ML method, it has the advantage of being simpler to implement for the CAISO because it is already using QR [7].

The steps are as follows. First, two separate QR estimations are performed for the 50th and 90th percentiles of the forecast error as a linear function of a set of independent variables related to weather and system conditions. (Four sets of such variables are considered here, as described in the figure caption.) Second, the value of error for the desired reliability (say the 97.5th percentile, which would result in a 2.5% shortage rate) is obtained by fitting a normal distribution to the 50th and 90th percentiles and extrapolating. Out-of-sample validation found that this resulted in more stable estimates of extreme percentiles rather than using QR directly to estimate that percentile, due to small sample issues with the number of observations in the tail. This is done for using 30 days of data prior to the day of interest; the desired percentile is then estimated for that day given the value of the independent variables on that day. The FRP requirement is set equal to that value. The performance of the method is then assessed by comparing the realized forecast error against the requirement. This is repeated for each of the days in the month (March 2020 in Fig. 10), and four 15-minute intervals within each

hour considered; this would give $30 \times 4 = 120$ observations to estimate the reliability and cost performance.

In Fig. 10, we repeated this for four levels of *a priori* reliability (10%, 5%, 3%, and 1.5% shortage probabilities) for each of four model specifications for the noon hour in March 2020. The best specification is based on two independent variables: the average (across four sites) of the 25th – 75th percentile prediction interval for GHI (this interval measuring the uncertain availability of solar energy; see [62] for a wind application of this measure) (Fig. 1); and a nonlinear (sine wave) transformation of median GHI, again averaged over four sites. That transformation yields values of zero if GHI is at zero or its theoretical maximum at that hour, and attains a maximum if GHI is halfway between those extremes; this reflects the fact that if solar is zero or if there is a clear sky, there is less uncertainty than if GHI is somewhere between the extremes. As Fig. 10 shows, there is one version of that model (the second point from the left, with an *a priori* reliability of 5%) that reduces oversupply by 20% (x-axis) and cuts the *ex post* frequency of FRP shortage by about half (from 7.5% to 4%), relative to the actual amount of FRP that the ISO procured for those intervals. Although that precise specification does not result in improvements in every month and time interval we considered, it often did so, and is thus worth considering as a relatively simple but effective way to condition FRP requirements on weather.

C. SIMULATION OF PRODUCTION COST SAVINGS FROM SOLAR FORECAST-INFORMED FLEXIRAMP REQUIREMENTS

The evaluation of solar probabilistic forecast-informed FRP requirements described in the above section considered only the probability of FRP shortage and the average amount of excess FRP procured. More useful to power engineers is an understanding of how different requirement estimation methods affect overall power system operating costs, including any load shedding costs associated with reliability issues.

Thus, we now compare the system cost performance of a production costing simulation that uses kNN-based estimates of FRP requirements relative to the CAISO baseline method. We use FESTIV to simulate security-constrained unit commitment and economic dispatch for an approximation of the WECC system. The approximation was created especially for this analysis, and was based on earlier NREL modeling, and models CAISO transmission in detail while omitting transmission limits outside California. Table 1 summarizes the test system.

The real-time markets simulate the operation of the real-time commitment processes used by the CAISO-run west-wide Energy Imbalance Market (EIM), which uses 15-minute settlement intervals, as well as the subsequent 5-minute dispatch processes of the EIM. The model captures the procurement of reserves, including the up- and down-flexible ramp product, and how those reserves are deployed in real-time in response to imbalances resulting from net load forecast errors. FESTIV's detailed representation of the network,

TABLE 1. Summary of WECC-scale simulation model.

Number of buses	1,820
Number of transmission lines	3,053
Number of generators	3,787
Non-renewable units/installed capacity [MW]	1,742/158,848
Hydro units/installed capacity [MW]	1,570/64,008
Wind units/installed capacity [MW]	85/20,004
Solar units/installed capacity [MW]	390/18,540
Typical day peak load [MW]	102,000
Number of regions captured	40

TABLE 2. Production cost and FRP procurement cost comparisons: Baseline vs solar-informed FRP, March 16-20, 23-25, 2020.

	Baseline (Current CAISO Method)	New kNN Solar- Probabilistic Forecast-Informed FRP	Perfect Forecast
Production Cost (M\$) (Savings, M\$) (Percent)	106.6	106.2 (0.429) (0.40%)	86.2
FRP Procurement Cost (Price*Quantity) (M\$) (Savings, M\$) (Percent)	0.1365	0.0964 (0.0401) (29.4%)	
Uncertainty Cost (Compared to Perfect) (M\$) (Savings, M\$) (Percent)	20.4	19.9 (0.429) (2.1%)	

generators, and rolling real-time market processes (including updating of solar, wind, and load forecasts) yields highly realistic representations of operations and costs, but also long execution times. To accelerate the computations, some simplifications were made, such as assuming fixed FRP requirements rather than using the CAISO FRP demand curves.

Table 2 summarizes cost results of applying FESTIV using the two sets of FRP requirements (baseline and kNN solar probabilistic forecast-informed), as well as a benchmark FESTIV simulation using perfect forecasts (i.e., day-ahead forecast = real-time forecast = actual net load). Differences between the FRP simulations' costs and the perfect forecast simulation quantify cost increases due to uncertainty in forecasts. Eight days in March 2020 are considered; because FESTIV simulations took approximately 24 hr of clock time for each simulated day, it was not possible to perform simulations for more days. Future work on FESTIV may implement faster computing procedures, allowing simulation of a more representative sample of days. In addition to production costs, we also estimated impacts on "FRP procurement cost", which equals FRP price times quantity procured.

Table 2 shows that using FRP requirements that are informed by weather, notably by probabilistic solar forecasts as described in Section III.B, would simultaneously lower FRP procurement costs by ~30% (~\$40K over the days considered) as well as production costs by 0.4% (~\$430K, over ten times the procurement cost savings). Interestingly, procurement costs decreased even though Fig. 11 shows that considering probabilistic solar forecasts did not always decrease FRP-up requirements (which tend to have higher

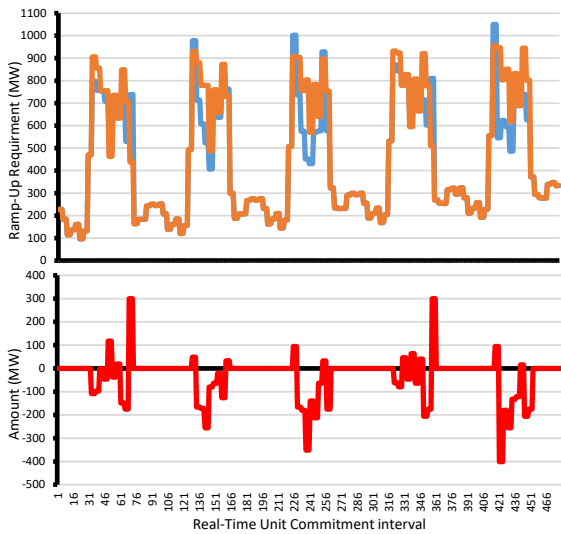


FIGURE 11. Top: Baseline (CAISO present method, orange) and new probabilistic solar forecast-informed flexible ramp-up product requirements (blue), March 16-20, 2020. Bottom: difference between new and baseline requirements. (Note: Real-Time Unit Commitment intervals each last 15 minutes.)

prices than FRP-down); thus, decreases in requirements must have occurred during periods of relatively high FRP prices. Later, we discuss possible reasons for production cost savings from solar-informed FRP requirements.

The last rows in Table 2 also reveal that, as expected, operating the system with perfect error-free net load forecasts has the lowest cost, with a large ~20% savings relative to the two runs with imperfect forecasts. This is termed “uncertainty cost” in the table, and represents the potential economic value of perfect forecasts. (Note however that those savings are likely to be much larger than what perfect forecasts would actually be expected to typically provide.) Better operating reserve requirements should reduce this uncertainty cost, and indeed the table shows that solar-informed FRP requirements do so.

To provide some context for these production and procurement cost savings, the CAISO annual market monitoring reports for 2018-20 document that annual total energy and ancillary services procurement costs for the CAISO system alone are on the order of \$8B/yr, while FRP procurement costs varied between about \$10M/yr to \$25M/yr [10]. If we extrapolate savings in production and flexiramp procurement expenses calculated in Table 2 from those 8 days to a full year, they would amount to a reduction of ~\$20M in production costs and ~\$2M in procurement costs annually. Extrapolating instead the percentage cost savings shown in the table yields somewhat higher annual savings (~\$30M/yr and \$3M-\$7M/yr, respectively). Of course, such extrapolations should be interpreted cautiously as market conditions vary over the year. But they indicate that the potential magnitude of annual savings would, if confirmed to be that large, justify efforts to use probabilistic solar forecasts to help define FRP requirements.

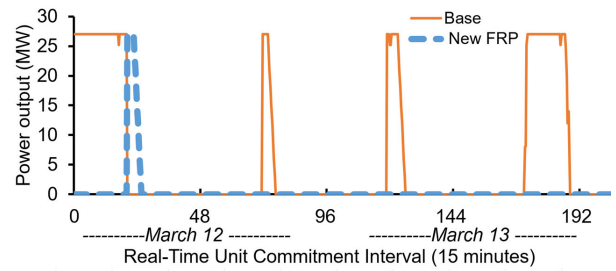


FIGURE 12. Generation (MW) from FESTIV simulations by an example peaker (combustion turbine) during March 12-13, 2020: under baseline method FRP requirements (orange) and probabilistic solar-informed requirements (blue).

We now consider some potential reasons for the production cost-savings estimated by FESTIV. Fig. 11 compares the FRP requirements (in the up direction) for five days (March 16-20, 2020) from the two methods (orange = baseline, blue = new requirements). The figure shows that the solar-informed requirements for FRP-up were increased in the early morning on some days, when net load is ramping up quickly, but were slightly reduced on other days. Meanwhile, ramp requirements were reduced during mid-day periods (the “belly” of the famous CAISO “duck curve”) as well as during evening ramps (the “neck” of the duck curve). We note that patterns were sometimes the opposite, with days earlier in March 2020 (not shown) showing higher ramp requirements by the new method in the evening, and lower requirements in the morning.

How might those changes affect production costs? Consider first a case when ramp requirements are decreased; if that occurs at times that ramp capability is actually unneeded (as in the sunny day on the left of Fig. 7, above), then fuel costs shrink by not having to commit as many flexible units. On the other hand, consider when ramp requirements are increased; if that happens when ramp forecast errors are high (as in the cloudy day on the right side of Fig. 7), then there are two possible sources of reduced production costs. One is decreased renewable curtailment (if more down-ramp was needed than previously anticipated), and the other is less severe price spikes and fuel costs because the system can avoid committing expensive short-start units to accommodate unexpectedly high ramps.

We now show an example of the latter cost savings. Because there are hundreds of generators in the western US power system, it is challenging to untangle the actual sources of cost savings from the output of a complex production costing model, but Fig. 12 provides a simple example.

Fig. 12 shows FESTIV’s simulated real-time commitment of a single short-start combustion turbine during morning and evening ramps for March 12-13, 2020 under the two sets of FRP requirements. Under the solar-informed FRP requirements, the generator needs to be committed for just one of those four peak periods (early on March 12), yielding cost savings relative to the FESTIV simulation based on the baseline FRP requirements, which committed that unit for all

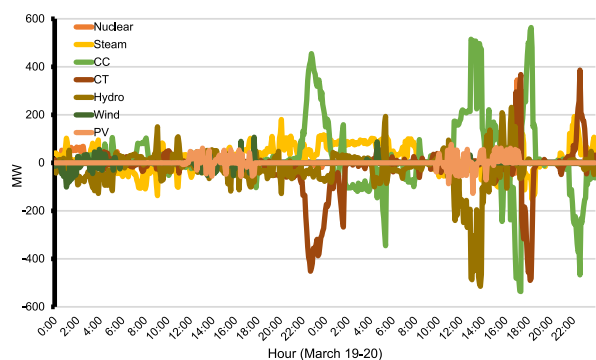


FIGURE 13. Change in total generation (MW) from FESTIV simulations (New FRP simulation minus Base simulation) for seven generation types.

four peaks. It turns out that more FRP in the up direction was procured by the new requirements at those times. As a result, we infer that it was unnecessary to commit short-start generation for three of those peaks, providing start-up and generation fuel cost savings.

Fig. 13 shows overall changes in system dispatch for two of the days considered in Table 2, illustrating possible sources of the reported \$430K production cost savings. In particular, the figure shows differences between dispatch levels (in MW) under the revised (solar-informed) FRP requirements versus the base dispatch. The green line shows that combined cycle units are dispatched more on average under the new requirements, while the dark brown line shows (on average) a decrease in more costly combustion turbine output. This change is most striking late in the evening of March 19, although there are some hours (e.g., late March 20) when instead turbine output increases somewhat and displaces combined cycle generation.

IV. CONCLUSION

Probabilistic forecasts have the potential to increase the effectiveness of several processes for managing uncertainty in spot power markets, including operator situational awareness, market participant hedging, setting reserve requirements, definition of net load contingencies for evaluating reserves, and, perhaps eventually, stochastic optimization of energy and reserve schedules. Realizing this potential poses many challenges to researchers. Methods are needed for visualizing high dimensional data sets; creating sets of scenarios and then reducing them to a manageable size for contingency constraints and event trees; estimating models for uncertainty forecasts for different net load components and locations, then aggregating them to obtain overall net load distributions; and managing the incompatibility of the timing and scale of probabilistic forecasts to the needs of system operators.

A particularly promising application of probabilistic solar, wind, load, and equipment availability forecasts is to improve the cost and reliability performance of operating reserves on various times scales, from frequency regulation to replacement reserves, even as much larger amounts of variable renewables are integrated. We have demonstrated the

potential for probabilistic solar forecasts to inform dynamic requirements for the CAISO flexible ramp product, and significantly reduce production expenses and costs to consumers. Our detailed case study illustrates the importance and effectiveness of particular methods to improve the quality of probabilistic forecasts; demonstrates statistical and machine learning methods to relate the degree of solar uncertainty to forecasts of needs for system reserves; and shows how multi-timescale production costing can quantify the value of improved reserve requirements and, by extension, the value of better probabilistic forecasts.

ACKNOWLEDGMENT

The authors thank A. Motley, R. Webb, G. Bautista, and C. Loutan of CAISO; B. Borissov and S. Rose of MISO, and J. Seuss and K. Selik of DOE for advice. Views expressed do not necessarily represent the views of DOE, the U.S. Government, or the CAISO.

REFERENCES

- [1] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez-de-Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Sol. Energy*, vol. 136, pp. 78–111, Oct. 2016.
- [2] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng, "Adaptive robust optimization for the security constrained unit commitment problem," *IEEE Trans. Power Syst.*, vol. 28, no. 1, pp. 52–63, Feb. 2013.
- [3] S. Bhavsar, R. Pitchumani, and M. A. Ortega-Vazquez, "Machine learning enabled reduced-order scenario generation for stochastic analysis of solar power forecasts," *Appl. Energy*, vol. 293, Jul. 2021, Art. no. 116964.
- [4] L. Bongers, N. Mararakanye, and B. Bekker, "Operating reserve dimensioning methodologies for renewable energy aligned power systems," in *Proc. 56th Int. Universities Power Eng. Conf. (UPEC)*, Aug. 2021, pp. 1–6.
- [5] M. Bucksteeg, L. Niesen, and C. Weber, "Impacts of dynamic probabilistic reserve sizing techniques on reserve requirements and system costs," *IEEE Trans. Sustain. Energy*, vol. 7, no. 4, pp. 1408–1420, Oct. 2016.
- [6] E. Y. Bitar, R. Rajagopal, P. P. Khargonekar, K. Poolla, and P. Varaiya, "Bringing wind energy to market," *IEEE Trans. Power Syst.*, vol. 27, no. 3, pp. 1225–1235, Aug. 2012.
- [7] *Flexible Ramping Product Refinements Initiative Appendix C—Quantile Regression Approach*, California Independent Syst. Operator, Folsom, CA, USA, May 2020.
- [8] CAISO. (2022). *Contingency Modeling Enhancements*. [Online]. Available: <http://www.stakeholdercenter-er.caiso.com>
- [9] CAISO. (2022). *Flexible Ramping Product Refinements*. [Online]. Available: <http://www.stakeholdercenter-er.caiso.com>
- [10] CAISO Dept. *Market Monitoring, Annual and Quarterly Performance Reports*. [Online]. Available: <http://www.caiso.com/market/Pages/MarketMonitoring/Default.aspx>
- [11] Y. Chen, Y. Wang, D. Kirschen, and B. Zhang, "Model-free renewable scenario generation using generative adversarial networks," *IEEE Trans. Power Syst.*, vol. 33, no. 3, pp. 3265–3275, May 2018.
- [12] N. Costilla-Enriquez, M. Ortega-Vazquez, A. Tuohy, A. Motley, and R. Webb, "Operating dynamic reserve dimensioning using probabilistic forecasts," *IEEE Trans. Power Syst.*, early access, Mar. 29, 2022, doi: 10.1109/TPWRS.2022.3163106.
- [13] M. Cui, D. Ke, Y. Sun, D. Gan, J. Zhang, and B.-M. Hodge, "Wind power ramp event forecasting using a stochastic scenario generation method," *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 422–433, Apr. 2015.
- [14] A. Daraeepour, D. Patino-Echeverri, and A. J. Conejo, "Economic and environmental implications of different approaches to Hedge against wind production uncertainty in two-settlement electricity markets: A PJM case study," *Energy Econ.*, vol. 80, pp. 336–354, May 2019.
- [15] K. De Vos, N. Stevens, O. Devolder, A. Papavasiliou, B. Hebb, and J. Matthys-Donnadieu, "Dynamic dimensioning approach for operating reserves: Proof of concept in Belgium," *Energy Policy*, vol. 124, pp. 272–285, Jan. 2019.

- [16] V. Dvorkin, S. Delikaraoglou, and J. M. Morales, "Setting reserve requirements to approximate the efficiency of the stochastic dispatch," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1524–1536, Mar. 2019.
- [17] Y. Dvorkin, "A chance-constrained stochastic electricity market," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 2993–3003, Jul. 2020.
- [18] P. Edwards, H. Sky, and V. Krishnan, "RAVIS: Resource forecast and ramp visualization for situational awareness—An introduction to the open-source tool and use cases," Nat. Renew. Energy Lab, Golden, CO, USA, Tech. Rep. NREL/TP-5D00-79746, 2021.
- [19] 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.
- [20] F. Fahiman, S. Disano, S. M. Erfani, P. Mancarella, and C. Leckie, "Data-driven dynamic probabilistic reserve sizing based on dynamic Bayesian belief networks," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 2281–2291, May 2019.
- [21] C. Feng, M. Sun, J. Zhang, K. Doubleday, B.-M. Hodge, and P. Du, "A data-driven method for adaptive reserve requirement estimation via probabilistic net load forecasting," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Aug. 2020, pp. 1–5.
- [22] J. Fox et al., "Forecasting and market design advances: Supporting an increasing share of renewable energy," *IEEE Power Energy Mag.*, vol. 19, no. 6, pp. 77–85, Nov. 2021.
- [23] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *J. Amer. Stat. Assoc.*, vol. 102, no. 477, pp. 359–378, Mar. 2007.
- [24] S. E. Haupt et al., "The use of probabilistic forecasts: Applying them in theory and practice," *IEEE Power Energy Mag.*, vol. 17, no. 6, pp. 46–57, Nov. 2019.
- [25] K. Hedman and J. King. (2020). *PERFORM Kick-off Meeting*. ARPA-E. [Online]. Available: www.arpa-e.energy.gov/sites/default/files/2021-02/ARPA-E%20Intro_PERFORM%20Kickoff_Final.pdf
- [26] H. Holttinen et al., "Methodologies to determine operating reserves due to increased wind power," *IEEE Trans. Sustain. Energy*, vol. 3, no. 4, pp. 713–723, Oct. 2012.
- [27] *Mitigation of Climate Change, 5th Assessment Report of the Intergovernmental Panel on Climate Change*, IPCC, Geneva, Switzerland, Apr. 2022.
- [28] D. Jost, A. Braun, R. Fritz, and S. Otterson, "Dynamic sizing of automatic and manual frequency restoration reserves for different product lengths," in *Proc. 13th Int. Conf. Eur. Energy Market (EEM)*, Jun. 2016, pp. 1–5.
- [29] J. Kazempour and B. F. Hobbs, "Value of flexible resources, virtual bidding, and self-scheduling in two-settlement electricity markets with wind generation—Part I: Principles and competitive model," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 749–759, Apr. 2018.
- [30] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.
- [31] L. J. Klein et al., "PAIRS: A scalable geo-spatial data analytics platform," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Oct. 2015, pp. 1290–1298.
- [32] R. Koenker and K. F. Hallock, "Quantile regression," *J. Econ. Perspect.*, vol. 15, pp. 143–156, Dec. 2001.
- [33] A. Kumlner, Y. Xie, and Y. Zhang, "A physics-based smart persistence model for intra-hour forecasting of solar radiation (PSPI) using GHI measurements and a cloud retrieval technique," *Sol. Energy*, vol. 177, pp. 494–500, Jan. 2019.
- [34] B. Li et al., "Sizing ramping reserve using probabilistic solar forecasts: A data-driven method," *Appl. Energy*, vol. 313, May 2022, Art. no. 118812.
- [35] B. Li, K. Sedzro, X. Fang, B.-M. Hodge, and J. Zhang, "A clustering-based scenario generation framework for power market simulation with wind integration," *J. Renew. Sustain. Energy*, vol. 12, no. 3, May 2020, Art. no. 036301.
- [36] B. Li and J. Zhang, "A review on the integration of probabilistic solar forecasting in power systems," *Sol. Energy*, vol. 210, pp. 68–86, Nov. 2020.
- [37] B. Li, J. Zhang, and B. F. Hobbs, "A copula enhanced convolution for uncertainty aggregation," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Feb. 2020, pp. 1–5.
- [38] 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)*, Jul. 2015, pp. 283–290.
- [39] S. Lu et al., "IBM PAIRS curated big data service for accelerated geospatial data analytics and discovery," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2016, pp. 2672–2675.
- [40] M. P. Maceiral et al., "Twenty years of application of stochastic dual dynamic programming in official and agent studies in Brazil," in *Proc. PSCC*, 2018, pp. 1–7.
- [41] M. Marinelli et al., "Demonstration of visualization techniques for the control room engineer in 2030," Future Control Room Functionality, Danish Tech. Univ., Kongens Lyngby, Denmark, Tech. Rep. ELECTRA Deliverable D8.1, WP8, 2017.
- [42] C. B. Martinez-Anido and B. Botor, "The value of day-ahead solar power forecasting improvement," *Sol. Energy*, vol. 129, pp. 192–203, May 2016.
- [43] A. D. Mills et al., "Solar-to-grid: Trends in system impacts, reliability, and market value in the United States with data through 2019," Lawrence Berkeley Natl. Lab., Berkeley, CA, USA, Tech. Rep., 2021. [Online]. Available: <https://escholarship.org/uc/item/3jn8b0qq>
- [44] NASA. *Images: GOES-16/GOES East Data and Imagery*. Accessed: Nov. 4, 2022. [Online]. Available: www.goes-r.gov/multimedia/dataAndImageryImagesGoes-16.html
- [45] M. A. Ortega-Vazquez and D. S. Kirschen, "Optimizing the spinning reserve requirements using a cost/benefit analysis," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 24–33, Feb. 2007.
- [46] A. Papavasiliou, "An overview of probabilistic dimensioning of frequency restoration reserves with a focus on the Greek electricity market," *Energies*, vol. 14, no. 18, p. 5719, Sep. 2021.
- [47] A. Papavasiliou, S. S. Oren, and B. Rountree, "Applying high performance computing to transmission-constrained stochastic unit commitment for renewable energy integration," *IEEE Trans. Power Syst.*, vol. 30, no. 3, pp. 1109–1120, May 2015.
- [48] W. B. Powell and S. Meisel, "Tutorial on stochastic optimization in energy—Part I," *IEEE Trans. Power Syst.*, vol. 31, pp. 1459–1475, 2015.
- [49] A. Roy, A. Vaswani, A. Neelakantan, and N. Parmar, "Theory and experiments on vector quantized autoencoders," 2018, *arXiv:1805.11063*.
- [50] J. D. Sprey, P. Schultheis, and A. Moser, "Dynamic dimensioning of balancing reserve," in *Proc. 14th Int. Conf. Eur. Energy Market (EEM)*, Jun. 2017, pp. 1–5.
- [51] E. Spyrou et al., "Flexibility options: A proposed ISO product for managing energy imbalance risk," Nat. Renew. Energy Lab., Golden, CO, USA, Tech. Rep. NREL/PR-5D00-80535, 2021.
- [52] Y. Sun et al., "Machine learning derived dynamic operating reserve requirements in high-renewable power systems," *J. Renew. Sustain. Energy*, vol. 14, no. 3, May 2022, Art. no. 036303.
- [53] A. Tuohy et al., "Solar forecasting: Methods, challenges, and performance," *IEEE Power Energy Mag.*, vol. 13, no. 6, pp. 50–59, Nov. 2015.
- [54] C. Uckun, A. Botterud, and J. R. Birge, "An improved stochastic unit commitment formulation to accommodate wind uncertainty," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2507–2517, Jul. 2016.
- [55] C. Voyant and G. Notton, "Solar irradiation nowcasting by stochastic persistence: A new parsimonious, simple and efficient forecasting tool," *Renew. Sustain. Energy Rev.*, vol. 92, pp. 343–352, Sep. 2018.
- [56] J. Wang et al., "Wind power forecasting uncertainty and unit commitment," *Appl. Energy*, vol. 88, no. 11, pp. 4014–4023, Nov. 2011.
- [57] Q. Wang and B.-M. Hodge, "Enhancing power system operational flexibility with flexible ramping products: A review," *IEEE Trans. Ind. Inform.*, vol. 13, no. 4, pp. 1652–1664, Aug. 2017.
- [58] D. Yang, "Making reference solar forecasts with climatology, persistence, and their optimal convex combination," *Sol. Energy*, vol. 193, pp. 981–985, Nov. 2019.
- [59] J. Yang et al., "Reinventing 2D convolutions for 3D images," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 8, pp. 3009–3018, Aug. 2021.
- [60] S. Yin, J. Wang, Z. Li, and X. Fang, "State-of-the-art short-term electricity market operation with solar generation: A review," *Renew. Sustain. Energy Rev.*, vol. 138, Mar. 2021, Art. no. 110647.
- [61] G. Zhang and J. D. McCalley, "Estimation of regulation reserve requirement based on control performance standard," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1173–1183, Mar. 2018.
- [62] C. Zhao, C. Wan, and Y. Song, "Operating reserve quantification using prediction intervals of wind power: An integrated probabilistic forecasting and decision methodology," *IEEE Trans. Power Syst.*, vol. 36, no. 4, pp. 3701–3714, Jul. 2021.
- [63] Z. Zhou et al., "Application of probabilistic wind power forecasting in electricity markets," *Wind Energy*, vol. 16, no. 3, pp. 321–338, Apr. 2013.

...