# Machine learning derived dynamic operating reserve requirements in high-renewable power systems © SOM

Cite as: J. Renewable Sustainable Energy 14, 036303 (2022); https://doi.org/10.1063/5.0087144 Submitted: 01 February 2022 • Accepted: 17 May 2022 • Accepted Manuscript Online: 18 May 2022 • Published Online: 24 June 2022



#### ARTICLES YOU MAY BE INTERESTED IN

Best practices in renewable energy resourcing and integration Journal of Renewable and Sustainable Energy 14, 030402 (2022); https:// doi.org/10.1063/5.0098482

Quantifying and mitigating soiling and abrasion in solar power Journal of Renewable and Sustainable Energy 14, 030401 (2022); https:// doi.org/10.1063/5.0097947

Techno-economic assessment of application of particle-based concentrated solar thermal systems in Australian industry Journal of Renewable and Sustainable Energy 14, 033702 (2022); https://doi.org/10.1063/5.0086655



## **APL Machine Learning**

Machine Learning for Applied Physics Applied Physics for Machine Learning First Articles Now Online!



J. Renewable Sustainable Energy 14, 036303 (2022); https://doi.org/10.1063/5.0087144

## Machine learning derived dynamic operating reserve requirements in high-renewable power systems **6 6**

Cite as: J. Renewable Sustainable Energy **14**, 036303 (2022); doi: 10.1063/5.0087144 Submitted: 1 February 2022 · Accepted: 17 May 2022 · Published Online: 24 June 2022



Yuchi Sun, James H. Nelson,<sup>a)</sup> (b) John C. Stevens, (b) Adrian H. Au, Vignesh Venugopal, Charles Gulian, Saamrat Kasina, Patrick O'Neill, Mengyao Yuan, (b) and Arne Olson

#### AFFILIATIONS

Energy and Environmental Economics, Inc., 44 Montgomery Street, Suite 1500, San Francisco, California 94104, USA

<sup>a)</sup>Author to whom correspondence should be addressed: jimmy.nelson@ethree.com

#### ABSTRACT

Accurately forecasting wind and solar power output poses challenges for deeply decarbonized electricity systems. Grid operators must commit resources to provide reserves to ensure reliable operations in the face of forecast errors, a process which can increase fuel consumption and emissions. We apply neural network-based machine learning to expand the usefulness of median point forecast data by creating probabilistic distributions of short-term uncertainty in demand, wind, and solar forecasts that adapt to prevailing grid conditions. Machine learning derived estimates of forecast errors compare favorably to estimates based on incumbent methods. Reserves derived from machine learning are usually smaller than values derived using incumbent methods, which enables fuel savings during most hours. Machine learning reserves are generally larger than incumbent reserves during times of higher forecast error, potentially improving system reliability. Performance is tested using multistage production simulation modeling of the California Independent System Operator system. Machine learning reserves provide production cost and greenhouse gas (GHG) emission reductions of approximately 0.3% relative to historical 2019 requirements. Savings in the 2030 timeframe are highly dependent on battery storage capacity. At lower levels of battery capacity, savings of 0.4% from machine learning reserves are shown. Significant quantities of battery storage are expected to be added to meet California's resource adequacy needs and GHG reduction targets. The addition of these batteries saturates reserve needs and results in minimal withinhour balancing costs in 2030.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0087144

#### I. INTRODUCTION

Efforts to decarbonize energy systems will rely heavily on solar and wind power to displace emissions from fossil fuel combustion.<sup>1,2</sup> Power production from weather-driven resources such as solar and wind poses challenges to electric system operators on many timescales. It has been shown that forecasting solar and wind power output on a short-term basis is difficult.<sup>3,4</sup> More difficult still is to estimate the forecast error of "net demand," defined here as demand minus wind and solar production potential, due to non-trivial weather-driven correlations between solar and wind output and demand. In this paper, we focus on the uncertainty of net demand within the operating hour.

Grid operators must be prepared to quickly rebalance power supply in response to a range of possible forecast errors. Therefore, grid operators hold operating reserves to ensure that system resources can balance fluctuations in net demand. Setting requirements for operational flexibility involves a trade-off between economics and reliability: excessively large requirements increase electricity production costs and customer bills for little benefit, but the inability to address a large forecast error can jeopardize reliable operations for an entire interconnection.

In this paper, we develop an open-source machine learning model, RESERVE, which is capable of characterizing the distribution of net demand forecast errors under a wide variety of power system conditions. Our goal in this work is not to improve the accuracy of point forecasts for demand, wind, or solar, but to determine the potential size of demand, wind, and solar forecast error, while also capturing covariance between these forecasts. We use machine learning to mine the historical record to provide grid operators with the best estimate of net demand uncertainty at different levels of likelihood of occurrence. Grid operators could use these data to balance cost and reliability by specifying the confidence intervals of forecast error to hold for operating reserves.

Published under an exclusive license by AIP Publishing

We use RESERVE to derive California Independent System Operator (CAISO) 15-minute flexible ramping product (FRP) requirements. FRP is a type of operating reserve that allows CAISO's 15-minute real time market to prepare for net load uncertainty in the subsequent 5-minute real time market. To perform this function, capacity held for 15-minute FRP is economically dispatched in the 5-minute energy market.

We then use a commercial production cost model, PLEXOS,<sup>5</sup> to simulate how the outputs from RESERVE impact the cost, greenhouse gas (GHG) emissions, and reliability of the CAISO balancing area relative to incumbent CAISO methods for deriving ramping requirements. We test three systems: a base case in 2019 and a low and high battery case representative of CAISO in 2030. Finally, we explain the mechanisms by which energy storage affects the benefits of using RESERVE to derive ramping requirements.

The rest of this article is arranged as follows: Sec. II provides a literature review, Sec. III outlines the methodology for creating the neural network-based machine learning model, Sec. IV provides results from the machine learning model, Sec. V outlines the methodology for creating the production simulation model, Sec. VI provides results from the production simulation model, and Sec. VII draws conclusions from this study. An Appendix provides technical details on machine learning and production cost modeling.

#### II. LITERATURE REVIEW

To ensure real-time supply demand balance on electric grids, system operators need to make informed judgments about how much within-hour flexibility to hold such that they can manage forecast error of net demand on the sub-hourly timescale. In practice, system operators frequently set operating reserve requirements for future time intervals based on forecast errors observed in the historical record.6-Looking backward to set future requirements does not allow operators to dynamically increase or decrease requirements based on current conditions. The distribution of forecast errors-both of net demand forecasts and of other types of forecasts-is generally related to the underlying physical phenomena such as temperature, wind speed, and solar irradiance. Some methods, therefore, quantify uncertainty by grouping historical forecast error observations from different periods based on known predictive factors of uncertainty such as the level of cloudiness or the wind speed.9 However, extreme forecast error events in the historical record are sparse, which implies that grouping historical events to capture underlying correlations between errors and system conditions increases the risk of having statistically insignificant sample sizes.

Various researchers have sought to use probabilistic models to quantify distributions of demand, wind, and solar forecast error. For instance, previous work has used the method of moments to fit beta and two-sided power distributions to solar irradiance data.<sup>10</sup> However, the choice of parametric distribution family for the conditional probability distribution functions (PDFs) can lead to overfitting of the training data and poor testing performance, whereas too narrow or specific a choice of distribution family introduces bias into the model's predictions.

Quantile regression, by contrast, is a non-parametric probabilistic forecasting approach that directly estimates quantiles of the conditional forecast error distributions as a function of grid conditions.<sup>11,12</sup> Non-parametric probabilistic forecasting methods have been used

within the electric power industry, with some notable examples being lower–upper bound estimation (LUBE) and variants thereof,<sup>13,14</sup> extreme learning machines (ELM),<sup>15</sup> and ensemble-based approaches such as quantile regression forests.<sup>16</sup>

Non-parametric quantile regression models may be constructed by fitting a model to data using the pinball loss function as the cost function. Example classes of models are polynomial functions of the explanatory variables or highly nonlinear functions such as multi-layer perceptrons (a type of artificial neural network). The pinball loss function described mathematically in Appendix A1 is theoretically minimized by the true conditional quantiles of the response variable (conditional on the explanatory variables) and can be adjusted to estimate any conditional quantile of the response variable.<sup>17</sup> Multi-layer perceptron-based quantile regression models can be trained via minimization of the pinball loss function over historical data because of the method's simplicity and documented success in a variety of applications.<sup>11,18</sup> Multi-layer perceptrons are a highly flexible class of model that has been widely used in both regression and classification problems in machine learning. They are capable of modeling arbitrary complex and nonlinear relationships between large numbers of explanatory variables and response variables. Therefore, we adopt a multi-layer perceptron machine learning formulation and use a pinball loss function.

In the context of power systems, researchers have recently improved upon the simpler methods employed by system operators by using different methods to determine dynamic operating reserve needs that adapt to real-time grid conditions. For example, researchers have used machine learning models coupled to a mixed-integer optimization simulation and reserve under- and over-procurement cost penalties to determine optimal prediction intervals of day-ahead wind forecast error.<sup>19</sup> Other researchers have applied probabilistic methods to co-optimize the provision of operating reserve and energy dispatch.<sup>20</sup> De Vos *et al.* used information on day-ahead net demand forecast error and forced outage rates to train a machine learning model to predict hourly contingency reserve requirements for the Belgian grid.<sup>21</sup> Finally, the Electric Power Research Institute has developed a machine learning module in its DynADOR model that can be used to calculate dynamic operating reserves.<sup>18</sup>

Optimization models, such as production simulation and capacity expansion models, are an industry-standard method for assessing the production cost, emissions, curtailment, and other qualities of changing parameters in deeply renewable electricity grids.<sup>22,23</sup> Prior research has demonstrated the production cost benefits of reducing overall operating reserve procurement needs.<sup>24</sup> Additional work has described and shown the benefit of procuring operating reserves from flexible solar power plants.<sup>25–27</sup>

This study builds upon prior work in this field in several ways. Similar to other researchers, we develop a machine learning model that determines dynamic operating reserves for wind- and/or solarheavy grids. However, we specifically configured RESERVE so it could be integrated into the existing sub-hourly operations of an ISO by training our model with the data that would be available to system operators in real-time operations and adhering to their conventions for calculating net demand forecast error. We also evaluate the benefits of the machine learning model using a multistage PLEXOS production simulation model that was benchmarked to historical ISO operation to ensure its reasonableness. This model's multistage architecture

Published under an exclusive license by AIP Publishing

J. Renewable Sustainable Energy 14, 036303 (2022); doi: 10.1063/5.0087144

captures unit commitment and market execution processes to quantify production costs, emissions, and renewable curtailment savings from applying the machine learning model. This investigation also examines the benefits of the machine learning model in both the recent past (2019) and the future (2030). This allows us to examine how and why deriving dynamic operating reserves with machine learning provides benefits, and how and why these benefits are affected by planned short-duration battery energy storage additions. Finally, we document and release our open-source machine learning model to the public via a GitHub repository for further use by the modeling and system operator community.

#### **III. MACHINE LEARNING METHODOLOGY**

Our multi-layer perceptron neural network model, which we call the RESERVE model, is trained to produce simultaneous quantile forecasts of net demand, demand, solar production, and wind production. The model employs a pinball loss function using an array of calendar data, weather data, and lag terms containing information about weather and forecast errors in the preceding forecast periods. In practical application, utilities would hold an amount of reserve that exactly covers a prediction interval of the expected net demand forecast error, which suggests an equivalence between forecast error and reserve requirement. Details on the model structure can be found in Appendix A 1.

As outlined schematically in Fig. 1, the model is trained on forecasts from near-past and near-future 15-minute and 5-minute intervals. The model predicts the forecast error between the 15-minute forecast at  $T_0 + 15$  and the three 5-minute forecasts that span the same time window. When deployed, the trained model would ingest near-past 5-minute forecasts as well as near-past and near-future 15-minute forecasts that would be available at the time at which the prediction is executed ( $T_0$ ). The trained model uses 34 input data points to produce each set of quantile forecasts.

We use a full year of historical 2019 demand, solar production, and wind production forecast data from the CAISO OASIS database as the basis of our forecast error predictions.<sup>28</sup> Only binding interval forecasts are used; advisory interval forecasts are not publicly available. The data are cleaned in a manual process that fills in short intervals of missing data via linear interpolation and corrects other data errors.

We are able to use the same year of data for both model training and model validation because we split the data via a tenfold cross validation procedure. For each model configuration, we train ten separate model instances on a unique subset of nine folds and test or "validate" each model instance using the remaining, tenth data fold. In this way, we measure variance in model performance due to the choice of training data set and robustly compare the performance of different model configurations, while also ensuring that each individual model's test set data has not been used in the training process. Details on model training can be found in Appendixes A 2 and A 4.

RESERVE is built using TensorFlow,<sup>29</sup> which is an open-source Python library. We have made RESERVE publicly available on GitHub.<sup>30</sup> RESERVE is structured as an artificial neural network with two hidden layers and ten neurons per layer. Validation loss is calculated after each epoch of training. The Adam optimizer<sup>31</sup> is used to identify weights and biases in the model that minimize the pinball loss.

We train models to simultaneously predict the conditional quantiles of forecast error for four outputs: net demand, demand, solar production, and wind production. These multi-objective models minimize a weighted average of the pinball loss corresponding to these four outputs, the weights being  $\sqrt{3}$ , 1, 1, and 1, respectively. A separate model is trained for seven different target quantiles: P2.5, P5, P25, P50, P75, P95, and P97.5. Furthermore, for each target quantile, a separate model is trained for each of the ten training dataset folds, resulting in 70 model instances in total. To evaluate performance at a particular target quantile (e.g., P97.5), we combine the validation set predictions of the ten model instances corresponding to that quantile, and compute performance metrics using the set of combined validation set predictions. In Sec. IV, Sec. V, and Appendix A, we refer to the collection of predictions from many model instances simply as "the model."



J. Renewable Sustainable Energy **14**, 036303 (2022); doi: 10.1063/5.0087144 Published under an exclusive license by AIP Publishing

**TABLE I.** Performance of neural network machine learning at achieving the desired target (input) coverage of net demand forecast errors. Coverage is defined as the percentage of forecast periods in which the realized forecast error falls below the model prediction and is evaluated here for an entire year of observations from 2019.

Target coverage (%)	Achieved coverage (%)	
2.5	2.3	
5	4.7	
25	24.6	
50	50.5	
75	74.8	
95	95.0	
97.5	97.5	

#### IV. MACHINE LEARNING RESULTS

#### A. Results compared to target metrics

As shown in Table I, the timeseries estimates of forecast error closely replicate the target coverage over a wide range of possible levels of forecast error. The model is, therefore, well suited to evaluate both extreme under- and over-forecast events, as well as more moderate levels of forecast error. Predictions with 50% coverage could help oper- ators understand potential biases in their underlying point forecasts.

Figure 2(a) visualizes the results of machine learning predictions for an example day in 2019, showing net demand forecast error predictions at different target quantiles, as well as the simultaneous predictions of the individual components of net demand forecasts: demand, wind, and solar. We observe that the P25–P75 prediction interval generally tracks the historical forecast error over the course of the day, but as expected does not cover all of the observed forecast



FIG. 2. (a) Forecast error quantile predictions for a sample day (7 July 2019) for net demand, demand, solar, and wind. (b) Forecast error quantile predictions for a sample day (22 December 2019) for net demand, during which net demand forecast error exceeded the machine learning 97.5% forecast error quantile prediction. In both (a) and (b), we observe sawtooth-shaped historical forecast error when the underlying forecast is changing rapidly; this is because we are comparing a 15-minute forecast to three 5-minute forecasts, which follows a convention adopted by the California Independent System Operator. The solar figure in (a) shows moderate levels of forecast error at night; this could be reduced or eliminated via either further refinement of the machine learning model and/or a day/night filter imposed outside of the model.

errors. As we move to prediction intervals that are expected to cover larger fractions of the forecast errors (P5–P95, and then P2.5–P97.5), we see more of the observed forecast errors being covered by the machine learning prediction. The P2.5–P97.5 prediction interval covers almost all observed forecast errors. As can be seen in Fig. 2(a), the forecast errors and prediction intervals of the individual demand, wind, and solar components can help to explain the underlying drivers of net demand forecast error. Including these individual components can help grid operators to build intuition for and confidence in the net demand forecast errors.

Figure 2(b) shows a day on which the machine learning model did not fully cover the net demand forecast error during sunrise and sunset. The model's prediction of forecast error increases shortly after the high net demand forecast error events, which suggests that the model can respond to recently observed periods of high error by increasing the predicted error going forward.

## B. Machine learning model performance comparison to incumbent CAISO method

As a point of reference for our neural network machine learning model, we compare our results to the incumbent "histogram" method used by CAISO to calculate ramping requirements in 2019. The CAISO histogram method looks backward in time approximately one month to observe extreme (P2.5 and P97.5) forecast error events. It then uses these events, grouped by hour of the day and day type (weekend or weekday), as the ramping requirement for the current day. The CAISO histogram method determines ramping requirements by comparing the 15-minute forecast to the maximum (or minimum) 5-minute forecast,<sup>6</sup> our machine learning algorithm predicts the forecast error between the 15-minute forecast and each 5-minute forecast. To render the comparison between machine learning and histogram methods easier, we take the maximum (or minimum depending on error direction) of the P2.5 and P97.5 machine learning 5-minute

predictions. We further adjust the machine learning predictions by setting the headroom or foot room ramping requirements to zero in any interval that would have had a negative requirement.

We evaluate the performance of our machine learning model using several performance metrics that are well suited to probabilistic forecasting models. Many of these metrics have been used by CAISO to quantify the performance of ramping requirement calculation methodologies.<sup>32</sup> Each of these performance metrics, shown in Table II and described further in Appendix A 3, quantifies an aspect of desirable model performance but must be considered in the context of the full suite of performance metrics to understand tradeoffs between different aspects of model performance. For example, it is desirable for a model to achieve a low average requirement in order to reduce the total cost of ramping capacity procurement. However, the ramping requirement should not be lowered at the expense of coverage, which should remain close to the targeted conditional quantile level to ensure that sufficient ramping capacity is held to maintain system reliability. While the observed forecast error should occasionally exceed the predicted forecast error because the ramping requirement target is 97.5% coverage (not 100% coverage), it is desirable to keep the average and maximum values for any exceedance as low as possible; higher values of exceedance represent larger reliability risks.

Machine learning methods can improve historical lookback uncertainty calculations by adapting to near-real time conditions. Comparing the Machine Learning and Histogram columns of Table II shows that machine learning methods can lower both the average requirement and the size of exceedance events. Exceedance events are of particular concern for system operators, as these events can drive a balancing area to be in violation of North American Electric Reliability Corporation reliability standards<sup>33</sup> and can cause the operators to lean on neighboring balancing areas for flexibility. Machine learning shows improvements in both the headroom and foot room directions, further demonstrating that it can successfully represent both under- and overforecast events. The pinball losses of machine learning predictions are

TABLE II. Performance metrics for machine learning net demand forecast error prediction based on 2019 data. Headroom/upward ramping requirements ensure that enough resources are available to provide extra power in periods with under-forecasted net demand; foot room/downward requirements prepare for over-forecasted net demand. The coverage metrics for machine learning presented in this table differ slightly from those presented in Table I because the 5-minute machine learning predictions are transformed into 15-minute requirements by taking the maximum or minimum error within each 15-minute interval. To ensure comparability between machine learning and histogram columns, the histogram performance metrics were re-calculated from those depicted in CAISO's documentation.

			Headr (upward r	oom amping)	Foot room (downward ramping)	
Performance metric	Definition	Units	Histogram	Machine learning	Histogram	Machine learning
Coverage	Percent of forecast errors covered by reserve requirement (target is 97.5% coverage)	%	94.4	97.3	92.8	96.6
Average requirement	Average of predicted forecast error at targeted quantile	MW	776	614	786	726
Average exceeding	The average size of excesses when observed forecast error exceeds the model prediction	MW	234	152	220	175
Maximum exceeding	Maximum size of excess when observed forecast error exceeds the model prediction	MW	3,353	1,705	2,652	1,983
Pinball loss	The expectation of the pinball loss function is mini- mized by the true conditional quantiles	MW	29.6	16.6	31.5	19.9

generally lower than those of the histogram method, indicating that machine learning net demand uncertainty estimates more closely resemble the "true" conditional quantiles for which the prediction is made.

The machine learning estimate of solar forecast error has the most uncertainty in the middle of the solar generation output range, which is frequently partially cloudy periods, as shown in the top left panel of Fig. 3. The forecast error for demand, in contrast, is largest when demand is high (top right panel). Both phenomena are reflected in the net demand forecast, shown in the bottom panels, with the highest net demand forecast errors occurring during medium solar generation and high load hours. The histogram method, shown in the solid red lines in the bottom panels, entirely misses these dynamics, resulting in uncertainty requirements that are not strongly dependent on either the solar forecast or the load forecast. This increases the risk of over-procurement of ramping capacity during periods in which net demand uncertainty is low and under-procurement during periods when net demand uncertainty is high.

Extreme under-forecast events are of special concern for system operators because their ability to start and ramp up generation can be limited within the operating hour. Figure 4 examines the performance of the machine learning and histogram methods during the 1% of intervals with the highest net demand under-forecasts. Neither method results in reserve requirements that cover all of the errors, but the machine learning reserve requirements are 1,000–1,500 MW

higher than the histogram method for every instance of forecast error greater than 2,500 MW. This indicates that a benefit of the machine learning method is that it better prepares the system to cover extreme forecast errors.

In Sec. V, we describe the formulation of our production cost model, which we used to demonstrate cost savings related to the utilization of machine learning requirements in the place of historical lookback (histogram) requirements. The results from the production cost modeling are desribed in Sec. VI. We note that CAISO is developing a quantile regression methodology as the replacement for histogrambased ramping requirements.<sup>34</sup> While CAISO's quantile regression method does not rely on machine learning, it is similar to ramping requirements based on real-time conditions. The quantile regression method would, therefore, be expected to provide some of the benefits relative to the histogram method that we show via PLEXOS modeling of machine learning requirements.

## V. EXPERIMENTAL DESIGN FOR PRODUCTION COST SIMULATIONS

#### A. Production cost model setup

CAISO, like other grid operators, runs multiple markets on different timeframes to help commit and dispatch resources under uncertainty, including a day-ahead market for the CAISO footprint and



FIG. 3. Machine learning quantiles as a function of solar generation forecast and demand forecast. Histogram ramping requirements for net demand are shown for comparison.

J. Renewable Sustainable Energy 14, 036303 (2022); doi: 10.1063/5.0087144



FIG. 4. Machine learning and histogram upward ramping reserve requirements for periods of extreme (top 1%) net demand under-forecast events, plotted against the magnitude of the historical net demand forecast error. Each data point represents an individual observation from 2019.

real-time markets for the wider footprint of Western Energy Imbalance Market (EIM). The CAISO market features co-optimized procurement of energy as well as regulation, spinning, and nonspinning reserves.<sup>35</sup> The EIM includes 15-minute market (FMM) and 5-minute real-time dispatch (RTD) timeframes and features flexible ramping products for the FMM and RTD in the upward and downward directions. Flexible ramping capacity ensures that unit commitment and dispatch prepare for uncertainty in the net demand forecast.

As shown in Fig. 5, we use a commercial production cost model, PLEXOS,<sup>5</sup> to simulate different timeframes to explore how different flexible ramping requirements on the 15-minute timeframe impact the cost, greenhouse gas emissions, and reliability of the CAISO system. Each day is simulated three times with progressively lower levels of forecast uncertainty and thermal generator commitment flexibility.

The first timeframe that we model in PLEXOS represents the combined impact of CAISO day-ahead and hour-ahead scheduling and is simulated with an hourly time step resolution. Unit commitments from slower-moving steam units (both standalone and steam units that are part of combined cycle units) are fixed after the first stage and cannot be changed in subsequent stages. While the on/off status for each of these units cannot be changed, subsequent stages can re-dispatch the setpoint of each online unit within its operational parameters. The second timeframe represents the FMM and is simulated with a 15-minute time step resolution. For the FMM timeframe, profiles for demand, wind, solar, and imports/exports are updated to reflect new forecast information, and 15-minute flexible ramping



J. Renewable Sustainable Energy 14, 036303 (2022); doi: 10.1063/5.0087144

Published under an exclusive license by AIP Publishing

requirements are enforced. All remaining thermal units are committed in the second stage, but their dispatch setpoints can still be redispatched in the next stage. The third timeframe represents the RTD market and is simulated with a 5-minute time step resolution. Profiles for demand, wind, solar, and imports/exports are updated again, as well as profiles for flexible ramping requirements.

In all stages of modeling, regulation up, regulation down, and spinning reserves are modeled as part of the unit commitment and dispatch simulation. Reserve and flexible ramping requirements are enforced at the CAISO-wide level. We model unit commitment and dispatch on a unit-by-unit basis for the CAISO system itself and include interactions with external entities via fixed import and export schedules at fixed prices based on historical data. Energy dispatch is performed at the zonal level with transmission constraints enforced between zones that represent the three investor-owned utilities served by CAISO: Pacific Gas and Electric, Southern California Edison, and San Diego Gas & Electric.

We simulate multiple resource portfolios in PLEXOS, three of which are presented in detail. The first is a near-past historical retrospective of 2019 in which the resource portfolio, demand, imports, and exports are set at 2019 historical levels. The second, labeled "2030 High Battery," reflects a 2030 resource portfolio that the California Public Utilities Commission adopted for CAISO's 2021-2022 transmission planning cycle.<sup>36</sup> As shown in Fig. 6(a), the largest differences between 2019 and 2030 High Battery portfolios result from the growth of solar and battery resources. Since the batteries are added largely to meet California's resource adequacy needs and GHG reduction targets (rather than the operational needs that are the subject of this paper), we are also interested in understanding how the model performs on a future system with high penetration of renewable resources but lower penetration of battery storage. Accordingly, we model a "2030 Low Battery" resource portfolio, which removes 90% (12.5 GW) of battery capacity from the 2030 High Battery portfolio and replaces it with 12.5 GW of combustion turbine capacity.

To explore the value of machine learning-derived ramping requirements for net demand forecast uncertainty on the future CAISO electricity system, we perform two separate simulations for each of the 2030 Low and High Battery portfolios: one with machine-learning derived ramping requirements, and one with histogram ramping requirements that are scaled up from 2019 values to be consistent with the 2030 resource portfolio (see Appendix B for more detail). Machine learning based 15-minute ramping requirements are calculated by taking the P2.5 and P97.5 values of forecast error of net demand from the RESERVE model, which is produced at a 5-minute time resolution, and taking the maximum (or minimum depending on error direction) of the 5-minute values within each 15-minute (see Sec. IV B). Figure 6(b) compares 15-minute flexible ramping requirements for 2019 and 2030.

#### VI. PRODUCTION COST SAVINGS RESULTS

#### A. Low battery capacity case results

One of our principal findings is that machine learning-derived ramping requirements can provide a meaningful reduction in production costs, GHG emissions, natural gas generation, and renewable curtailment in systems with lower levels of battery capacity. As shown in Table III, the 2019 and 2030 Low Battery portfolios show 0.3% and 0.5% production cost savings, respectively. Adding production cost savings and savings from renewable energy certificate procurement costs<sup>37</sup> results in total savings of \$18.5M/yr (2019) and \$42.2M/yr (2030 Low Battery) for the CAISO footprint. Machine learning ramping requirements reduce natural gas generation by replacing that generation with curtailed renewable generation, which decreases GHG emissions.

Our results in Table III may overestimate the possible savings from machine learning requirements in the actual CAISO system because we simulate CAISO's flexible ramping needs as a standalone entity. The geographic diversity of solar, wind, and demand across the entire EIM market footprint reduces CAISO's actual flexible ramping needs; if the diversity benefit of EIM-wide balancing were to be factored in, a reduction in the ramping requirement of ~47% would be expected in 2019.<sup>38</sup> We also do not simulate procurement of ramping capacity from non-CAISO resources to meet CAISO-area ramping requirements, which would be expected to reduce savings from machine learning requirements relative to what we have modeled. At the same time, our PLEXOS production simulation model understates the value of operational flexibility due to a lack of unplanned unit



FIG. 6. (a) CAISO resource portfolios simulated in PLEXOS for 2019 and 2030. (b) 15-minute flexible ramping requirements for 2019 and 2030, shown as the average requirement in each hour of the day based on a year of time series data.

TABLE III. Savings from machine learning 15-minute flexible ramping requirements relative to histogram requirements. Savings from PLEXOS runs are calculated as the change in the 5-minute RTD stage that results from a change in the flexible ramping requirement in the upstream FMM stage.

		Difference: Histogram minus machine learning				
Metric	Units	2019	2030 low battery	2030 high battery		
Production cost savings	% of annual production cost	0.3%	0.4%	0.0%		
	\$M/yr	14.5	12.9	0.1		
Total cost savings (renewable curtailment reduction valued at \$18/MWh)	\$M/yr	18.5	42.2	0.2		
GHG savings	% of annual emissions	0.2%	0.6%	0.0%		
	MMTCO <sub>2</sub> /yr	0.1	0.3	0.0		
Natural gas generation reduction	% of annual natural gas generation	0.4%	1.6%	0.2%		
	GWh/yr	225	832	81		
Curtailment reduction	% of wind and solar generation potential	0.6%	0.9%	0.0%		
	GWh/yr	224	813	7		
Decrease in frequency of RT5 energy prices above \$150/MWh (negative indicates increase)	% of 5-minute intervals	0.0%	0.0%	-0.2%		

outages, perfect-foresight dispatch of energy storage within a single simulation stage, zonal (as opposed to nodal) modeling of transmission constraints, and a variety of other factors. These real-world considerations would tend to increase the observed savings and reliability benefits from machine learning reserves.

Because we cannot observe reliability events or near-events directly in PLEXOS, we instead observe if machine learning reserves affect the frequency of high energy price events. We have already seen that machine learning flexible ramping requirements are significantly higher during extreme net load under-forecast events relative to the histogram method, likely decreasing the frequency of high energy prices that are driven by under-forecasts. However, the lower average machine learning requirements have the potential to increase the frequency of high price events because fewer MW of flexibility is required on average. To explore the net impact of these two factors, we calculate the difference in frequency of 5-minute intervals in the RTD model stage that have an energy price above \$150/MWh between machine learning and histogram PLEXOS simulations. \$150/MWh is chosen as the threshold for a "high" price because it is roughly five times greater than the average price. We observe that the histogram and machine learning PLEXOS results have minimal differences in the frequency of high energy prices. Thus, we infer that the machine learning requirements adjust to the underlying uncertainty in the net demand forecast in a way that saves cost by reducing the ramping requirement in times where there is less forecast uncertainty but does not significantly increase the frequency of challenging balancing intervals.

#### B. High battery capacity case results

Comparing the 2030 Low Battery and 2030 High Battery columns of Table III, we see that increasing the capacity of batteries on the 2030 CAISO system results in a steep decrease in the benefits of machine learning ramping requirements relative to histogram requirements. With higher battery capacity, the cost to provide ramping requirements is frequently near zero, which reduces the potential benefits of improvements to the ramping requirement itself. This dynamic is explained in depth below.

Figure 7 demonstrates that on a system with abundant solar generation but little supporting battery capacity (2030 Low Battery), it could be challenging and costly to meet ramping requirements. Higher levels of solar generation relative to 2019 increase the net demand ramping requirements because the magnitude of solar generation uncertainty grows with solar capacity. Meeting the ramping requirements is challenging largely due to the relationship between reserve provision and curtailment. The 2030 Low Battery panel of Fig. 8 shows that renewable curtailment can persist during daylight hours; increasing thermal generation commitment to provide ramping flexibility during these hours would increase curtailment and fuel consumption because many thermal generators must be online and generating to provide operational flexibility. The relatively small battery capacity in the 2030 Low Battery portfolio forces the batteries to choose whether to provide reserves and ramping or to charge and discharge to perform energy arbitrage; this choice creates an opportunity cost for batteries to provide flexible ramping. On the example day in Fig. 8, the charging and discharging schedules in the 2030 Low Battery column show that batteries do not fully cycle their state of charge despite persistent renewable curtailment, thereby largely forgoing energy arbitrage opportunities. Instead, the batteries use most of their capacity to provide both upward and downward reserves. Because of high opportunity costs from thermal resources and batteries, we see positive flexible ramping prices; downward flexible ramping is especially challenging to provide on this day, with high positive prices throughout daylight hours.

Comparing the 2030 High Battery and 2030 Low Battery lines in Fig. 7 demonstrates that adding batteries drastically reduces the cost of providing 15-minute ramping flexibility. We calculate the market size for flexible ramping (in both the upward and downward directions) as the product of 15-minute FRP price and 15-minute FRP quantity over the course of the year. The market size for 15-minute FRP for the 2030 Low Battery portfolio is \$20M/yr but drops to only \$4M/yr for the



**FIG. 7.** 2030 prices for 15-minute flexible ramping requirements as a function of battery capacity (with low and high battery cases as bookends), presented as an hourly average over the year. Intermediate battery capacity portfolios (2.8, 4.2, and 7.0 GW) depict the pace at which additional battery capacity reduces flexible ramping prices. These portfolios, like the 2030 Low Battery portfolio, remove a portion of the battery capacity from the 2030 High Battery portfolio and replace it with equivalent combustion turbine capacity.



FIG. 8. Example dispatch day from June 2030. Simulations with machine learning ramping requirements are shown. In the Battery Operations panels, percentages can go above 100% because batteries provide energy and reserves with the full operational range from charging to discharging.

2030 High Battery portfolio largely because the extra battery capacity in the 2030 High Battery portfolio reduces the marginal cost to provide within-hour ramping capacity to near zero in most hours.

On the example day in Fig. 8, the impact of higher battery capacity on both reserve and energy prices is shown. After sunset in the 2030 Low Battery column, the price of upward flexible ramping tracks the shape of the energy price, showing that the opportunity cost to provide ramping flexibility can be tied to the cost to provide energy on this day. In contrast, the price of upward flexible ramping in the 2030 High Battery column is near zero in most hours and the energy price after sunset is flat.

Batteries are expected to flatten energy prices by discharging during the highest price intervals first, followed by the next highest price intervals, and so on. Comparing the energy prices in Fig. 8, we see that the addition of battery capacity between the 2030 Low Battery and 2030 High Battery portfolios has resulted in prices from sunset to midnight converging on a single, flat price. Higher priced combustion turbine generation shown in the 2030 Low Battery column is replaced with battery dispatch and lower priced combined cycle gas generation in the 2030 High Battery column. The occurrence of many consecutive hours of similar prices enables batteries to provide reserve and ramping flexibility at low or zero cost because a battery is indifferent to exactly when it produces energy at night, as long as it discharges before low or zero price solar production hours occur.

Four-hour duration batteries, including those modeled in PLEXOS here, can be particularly effective at providing ramping capacity in a system with abundant solar energy because the duration of high- and low-price hours is largely set by the diurnal schedule of solar generation. Solar resources generate for roughly half the day, creating low-priced periods longer than four hours in duration, followed by nighttime periods of higher prices that are also longer than four hours. Due to their limited energy capacity, the batteries cannot charge at their full power rating in all of the low-price hours, nor can they discharge at their full power rating in all of the high price hours. This can be seen in the charging and discharging schedules of batteries in Fig. 9, where on average each four-hour battery is not being scheduled at 100% charge/discharge in the 2030 High Battery portfolio. Because the batteries can generally be scheduled such that they have spare capacity available to provide grid flexibility within the operating hour, they can provide reserve and ramping capacity at a low or near-zero marginal cost. We expect, but do not demonstrate, that similar results would be observed with different storage durations, as long as the average duration of the storage resources is shorter than the duration of high- and low-price periods.

In the capacity expansion modeling that designed the 2030 High Battery portfolio, resource adequacy and greenhouse reduction constraints frequently drive portfolio selection.<sup>39</sup> Four-hour batteries installed primarily to provide resource adequacy are available to provide arbitrage and ramping in off-peak hours. Similarly, batteries installed primarily to drive GHG reductions by moving solar energy to the nighttime can also provide ramping and resource adequacy. Thus, on average batteries are not utilized at 100% capacity in each hour (Fig. 9), but they are utilized heavily during peak periods and hours of curtailment. Because batteries have slack capacity available to provide ramping and reserve capacity in most hours, the marginal cost of providing these products decreases as more batteries are added to the



FIG. 9. Battery utilization for the 2030 Low Battery and 2030 High Battery portfolios is presented as an hourly average over the year. Both cases use machine learning flexible ramping requirements.

scitation.org/journal/rse

system (Fig. 7). A recent report, which also simulates a California power system corroborates our finding that batteries can have a large impact on operating reserve prices, and in the extreme, can minimize the cost to provide operating reserves.<sup>27</sup>

While the results of our 2030 High Battery scenario suggest that operating flexibility may be a less important constraint on systems with high battery penetration, this portfolio mix may be somewhat unique to the southwestern portion of the United States, which has a practically unlimited supply of high-quality solar resources, somewhat limited availability of portfolio-diversifying wind resources, and a summer-peaking load pattern for which battery storage paired with solar has a high resource adequacy value. On these systems, solardriven price swings allow short-duration batteries to maintain the state of charge windows necessary to provide within-hour balancing services at low cost as a by-product of charging and discharging to perform diurnal energy arbitrage. Northern systems with high wintertime loads and a higher reliance on wind power, which does not have such a predictable diurnal output pattern, are likely to see less development of battery storage in the next decade. These systems would continue to benefit from machine-learning-derived reserve requirements, as would systems in the southern United States that have not yet achieved battery saturation. Moreover, while we have not explored this concept in this paper, we believe that the machine-learning techniques that we describe here could help system operators and battery storage project owners maintain a state of charge and maximize the value of batteries to the system.

Finally, the value of machine learning in reserve calculations is not limited to production cost savings. As shown above in Fig. 4, machine learning-derived reserves are much better at identifying potentially large under-forecast errors. Even on systems with high battery storage penetration, early and accurate identification of these potential events could help the system operator to ensure that the system's batteries are charged and ready to help ensure system reliability.

#### VII. CONCLUSIONS

Our results suggest that in systems with relatively constrained sub-hourly flexibility, both machine learning-derived ramping requirements and/or provision of ramping capacity from variable renewables (see Appendix C) can provide value to the system in the form of reduced fuel combustion during most hours and better reliability during periods of high forecast error. As renewable penetration increases, the potential value of the machine learning model grows as well.

Our study focuses narrowly on one flexible product of the CAISO system: the 15-minute flexible ramping product. However, other balancing products such as regulation up and down could also benefit from machine learning techniques to finely tune procurement requirements. Regulation is currently a much larger cost driver for the current CAISO system than FRP; hence, significant additional savings may be available to the CAISO and other system operators from the machine learning model. Future studies should evaluate the benefits of applying machine learning techniques for these products.

We also observe that the presence of large quantities of battery storage can reduce the importance of reserving within-hour flexibility in grid operations ahead of real-time operations. However, until such time as these resources are present everywhere—likely years or decades in the future—operators can derive value from revising ramping requirements and expanding the number of resources that can contribute to ramping needs. Even in electricity systems with much higher levels of storage, high-quality information on net demand uncertainty can be valuable to ensure that storage resources maintain a state of charge that enables them to perform during critical periods, particularly periods of high net demand forecast error. Future studies should evaluate the benefits of the machine learning model for other systems, particularly northern systems with higher reliance on wind rather than solar and fewer batteries.

#### ACKNOWLEDGMENTS

The work presented herein was funded in part by the Advanced Research Projects Agency-Energy (ARPA-E), U.S. Department of Energy, under Award No. DE-AR0001275. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof, nor do they state or reflect the views of the CAISO.

The study team appreciates feedback throughout the study process from Joseph King, Ashley Arigoni, Richard Wilson, and Richard O'Neill from ARPA-E and Guillermo Bautista Alderete, Hong Zhou, Amber Motley, Peter Klauer, and Clyde Loutan from the CAISO.

#### AUTHOR DECLARATIONS Conflict of Interest

#### Conflict of Interes

Funding for the work presented herein originates exclusively from ARPA-E and Energy and Environmental Economics, Inc. (E3). E3 is a consulting company that works with a diverse set of clients on energy and electricity issues, including but not limited to governmental agencies, developers, utilities, and non-governmental agencies. E3 clients did not influence or fund the work presented herein.

#### **Author Contributions**

Yuchi Sun: Conceptualization, methodology, software, validation, investigation, writing-review and editing, visualization, and funding Acquisition. James H. Nelson: Conceptualization, methodology, writing-original draft, writing-review and editing, visualization, and funding acquisition. John C. Stevens: Conceptualization, methodology, software, data curation, writing-original draft, writing-review and editing, visualization, project administration, and funding acquisition. Adrian H. Au: Methodology, software, validation, investigation, and visualization. Vignesh Venugopal: Contributed in methodology, software, validation, investigation, writing-original draft, and visualization. Charles Gulian: Methodology, software, validation, formal analysis, investigation, writing-original draft, and visualization. Saamrat Kasina: Conceptualization, methodology, and investigation. Patrick O'Neill: Methodology and investigation. Mengyao Yuan: Data curation, software, and investigation. Arne Olson: Conceptualization, methodology, writing-original draft, writing-review and editing, supervision, project administration, and funding acquisition. Y.S., J.H.N., and J.C.S. contributed equally to this work.

#### DATA AVAILABILITY

The data that support the findings of this study are openly available in CAISO OASIS at http://oasis.caiso.com/mrioasis/logon.do, Ref. 28, and in Energy and Environmental Economics at https:// github.com/e3-/RESERVE, Ref. 30. The authors used a PLEXOS model and Western Electricity Coordinating Council (WECC) dataset in this study as a base from which to build our system. This base PLEXOS model and dataset are not publicly available but can be licensed from Energy Exemplar (https://www.energyexemplar.com/plexos). The data that support the findings of this study are openly available in CAISO OASIS at http://www.caiso.com/TodaysOutlook/Pages/default.aspx, and the CAISO managing oversupply dataset (http://www.caiso.com/ Documents/ProductionAndCurtailmentsData\_2019.xlsx) historical interchange data from the EIA's Hourly Electric Grid Monitor (https://www.eia.gov/electricity/gridmonitor/dashboard/electric\_ overview/US48/US48). PLEXOS Model Benchmarking was performed using the EPA's CEMS database (https://www.epa.gov/emc/ emc-continuous-emission-monitoring-systems). The public data that support the findings of this study are available from the corresponding author upon reasonable request.

#### APPENDIX A: SUPPLEMENTAL RESERVE MODELING DETAILS

#### 1. RESERVE model structure

The RESERVE model is a multi-layer perceptron neural network with two layers and ten neurons per layer. We experimented with the number of hidden layers and neurons. This did not result in major changes to model performance. Figure 10 presents an illustrative diagram of the RESERVE neural network structure.

The RESERVE model presented in this manuscript utilizes 34 inputs:

- 15-minute market forecasts for demand, solar, and wind at  $T_0 30$ ,  $T_0 15$ ,  $T_0$ , and  $T_0 + 15$  (3 × 4 = 12 inputs);
- 5-minute market forecasts for demand, solar, and wind at  $T_0 30$ ,  $T_0 25$ ,  $T_0 20$ ,  $T_0 15$ ,  $T_0 10$ , and  $T_0 5$  (3 × 6=18 inputs); and
- Calendar-based inputs: solar hour angle, solar day angle, days since start of training period, and an index for each 5-minute interval within the 15-minute interval at  $T_0 + 15$  (four inputs).



FIG. 10. Illustrative diagram of the RESERVE neural network.

J. Renewable Sustainable Energy **14**, 036303 (2022); doi: 10.1063/5.0087144 Published under an exclusive license by AIP Publishing

The 5-minute index is necessary because RESERVE predicts forecast error for each of the three 5-minute intervals within each 15-minute interval. All inputs other than the 5-minute interval index remain static as we progress through the three 5-minute intervals within each 15-minute interval.

The inputs described above are fed into the input layer, which is then subjected to operations that transform them into the values of each neuron, and ultimately the desired outputs. The input streams are first normalized with the built-in normalization layer function in TensorFlow. This first subtracts the mean of each feature from the time-varying feature values. Then, this difference is divided by its standard deviation. The mean and standard deviation is calculated from all training-validation data and remains constant for each cross-validation fold and the testing process. Each layer of operation is a linear recombination of the previous layer with an activation layered on top, in the following mathematical form:

$$H_{k+1,i} = ReLU\left(\sum_{j} H_{k,j} w_{k,i,j} + b_{k,i,j}\right),\tag{A1}$$

in which  $H_{k,j}$  denotes the value of neuron j in layer k.  $w_{k,i,j}$  is referred to as the weight that connects the neuron i in layer k to neuron j in layer k + 1, while  $b_{k,i,j}$  is the bias term that connects the two same neurons.

On top of the linear combination operation defined by the weights and biases, an activation function applies non-linearity to the system. REctified Linear Unit<sup>40</sup> (ReLU) is chosen as the activation function, which takes the following form:

$$ReLU(x) = \begin{cases} 0, & x < 0, \\ x, & x \ge 0, \end{cases}$$
(A2)

in which x is the value passed from previous layer or input. ReLU has become an industry standard with its demonstrated performance in multiple disciplines and resistance against gradient explosion.

Taking the values of the output neurons, the loss function forms the minimization target. In RESERVE's case, the Pinball loss function of each output neuron can be written as

PinballLoss
$$(y, \tau) = w_o \begin{cases} \tau(y' - y), & y < y', \\ (1 - \tau)(y - y'), & y \ge y', \end{cases}$$
 (A3)

in which  $\tau$  is an input parameter representing the quantile for which we would like the model to generate a prediction. y and y' are the predicted and actual values of this output, respectively.  $w_o$  is the weight for this output in the total objective calculation. Note that when  $\tau = 0.5$ , the pinball loss becomes the mean absolute error. RESERVE minimizes the weighted-average pinball loss associated with net demand, demand, solar, and wind forecast errors; the weighting factors that we use are  $\sqrt{3}$ , 1, 1, and 1, respectively. We use  $\sqrt{3}$  (equal to about 1.73) for net demand to prioritize this output over the others because net demand forecast error will ultimately be used to set reserve levels. Individual demand, solar, and wind forecast error estimates can be used to understand drivers of net demand forecast error but are not directly used in our study.

As discussed in Sec. IV B, we perform two transformations on RESERVE's predicted forecast error values to create reserve requirements that are used in production simulation. The first transformation identifies the maximum (or minimum depending on forecast error quantile) of the three 5-minute forecast error values in each 15-minute interval to produce 15-minute ramping requirement values. This was done at the CAISO's request to match the calculation process used by CAISO's histogram method. The second transformation adjusts the machine learning predictions by setting the headroom or foot room ramping requirements to zero in any interval that would have had a negative requirement-an infrequent occurrence that implies a persistent bias toward either underor over-forecasting in the underlying demand, wind, or solar forecasts. Note that the performance metrics that we present in Appendix A4 have not had these two transformations applied because in that section, we quantify the ability of the RESERVE model to achieve the desired forecast error prediction targets.

#### 2. Training procedure

RESERVE uses a tenfold cross validation procedure with the ADAM optimizer.

During training, a full year of data from 2019 is used, which theoretically should yield 1 year \* 365 days/year \* 24 h/day \* 12 5-minute intervals/hour = 105,120 training samples. After removing a small number of samples with incomplete data, this translates to roughly 90,000 training samples and 10,000 validation samples in each of the training and validation iterations. On the computer we employed when training RESERVE, which includes an Intel Core i7-8665U processor and 16 GB of RAM, it takes 10-15 epochs or 2 to 5 minutes to train one model for one quantile. With ten cross validation iterations and seven quantiles, it takes 3 to 4 hours to finish training.

To train each of the ten models for a given input quantile, the model is fed batches of data to calculate gradients and make updates to weights and biases. In our application, each batch consists of 64 data samples. In practice, the size of the batch is chosen such that it is small enough to not throttle the data transfer between RAM and the processing unit. In practice, 64 and multiples of it are often chosen as a convenient default.<sup>41</sup>

For each cross-validation fold, RESERVE trains until no significant improvement are observed in three epochs (TensorFlow parameter: patience = 3). Here, we use epoch to describe when enough batches of data have passed through the model that the model has seen all training data exactly once. Significant improvement is defined as an improvement at least 0.5 MW in pinball loss (TensorFlow parameter: min\_delta = 0.5).

A summary of our training, evaluation, and deployment process can be found in Fig. 11.

#### 3. Evaluation metrics

As described in Table II, we use coverage, average requirement, average exceeding, maximum exceeding, and pinball loss to evaluate our model's performance. While the main text is focused qualitative description, here we give mathematical definitions for completeness. For coverage,

$$C(y, y') = \begin{cases} 1, & y < y', \\ 0, & y \ge y', \end{cases} \quad \text{coverage} = \frac{1}{N} \sum_{i} C(y_i, y'_i), \quad (A4)$$

in which  $C(y_i, y'_i)$  is a function that keeps a tally of each instance where forecasted quantile value is smaller than the real value. N is the total number of samples. For a perfect model, coverage would be exactly the target quantile.

For average requirement,

Requirement 
$$=\frac{1}{N}\sum_{i}|y_{i}|.$$
 (A5)

It is a simple average of the absolute value of the predicted forecast error for the input quantile. We use the term "requirement" here because in our PLEXOS study, we use forecast error quantile data for flexible ramping requirements.

For average exceedance,

$$E^{avg} = \frac{1}{M} \sum_{i} \left( y_i - y'_i \right) \quad \forall y_i > y, \tag{A6}$$

in which M is the total number of instances where predicted quantile y' is smaller than the true value y.



J. Renewable Sustainable Energy 14, 036303 (2022); doi: 10.1063/5.0087144

Published under an exclusive license by AIP Publishing

For maximum exceedance,

$$E^{max} = \max(y_i - y'_i) \quad \forall y_i > y. \tag{A7}$$

Maximum exceedance is the largest of all exceedance instances.

For the mathematical definition of the pinball loss, readers are referred to Appendix A 2.

To provide a broader range of metrics, we also calculate and present below two additional metrics using 2019 data: reliability and sharpness. Reliability and sharpness are similar metrics to the coverage and requirement metrics (respectively) presented in the main text of the manuscript, except that coverage and requirement focus on one side of a quantile, while reliability and sharpness focus on an opposing quantile pair.

Reliability, sometimes also called validity or calibration, is defined as

$$C(y, y_1, y_2) = \begin{cases} 1, & y_1 < y < y_2, \\ 0, & y \ge y_2 \text{ or } y < y_1, \end{cases}$$
Reliability =  $\frac{1}{N} \sum_i C(y_i, y_{i,\tau}, y_{i,1-\tau}),$ 
(A8)

in which  $y_{i,\tau}$ ,  $y_{i,1-\tau}$  are predictions from opposing quantiles  $\tau$  and  $1 - \tau$  on the same sample. The term "reliability" here should be interpreted as the statistical metric, not the power systems concept of reliable system operations. Reliability indicates how frequently the observed forecast errors fall within the range between two quantiles. Here, we calculate the reliability of RESERVE's predictions between the quantile pair 2.5% and 97.5%. A perfect reliability result here would be 95%, and the probability that forecast errors fall between 2.5% and 97.5%. The reliability of RESERVE's predictions for 2019 data is found to be 93.9%, improving on the CAISO histogram method for the same quantiles, which has a reliability of 87.2%.

Sharpness, sometimes also called efficiency or width, is defined as

Sharpness 
$$=$$
  $\frac{1}{N} \sum_{i} |y_{\tau} - y_{1-\tau}|.$  (A9)

Sharpness measures the average distance between a pair of quantile predictions. All else equal, a smaller sharpness value is desired because higher reserve requirements increase power system dispatch costs. Here, we measure sharpness in units of MW between the 2.5% and 97.5% quantiles. We find that RESERVE's sharpness for 2019 data is 1,340 MW, improving on the calculated CAISO histogram method value of 1,562 MW.

#### 4. Supplemental RESERVE results

In this section, we provide coverage and pinball loss results for each prediction target and target quantile for 2019 data, coverage and pinball loss results for each of the 10 folds for the net demand prediction target for 2019 data, and quantile crossing frequency and magnitude for the net demand prediction target for 2019 data.

## APPENDIX B: SUPPLEMENTAL PLEXOS MODELING DETAILS AND RESULTS

In this section, we provide additional detail on aspects of PLEXOS production simulation modeling.

#### 1. 2019 Benchmarking

To ensure that our simulations of the 2019 CAISO system adequately reflect conditions experienced in 2019, we benchmark results of the 2019 histogram ramping requirement PLEXOS simulation to historical CAISO data. It is our goal to model a realistic set of historical conditions, but some deviations are expected between a production cost simulation and actual CAISO operations.

In Fig. 12, we observe good agreement between the 2019 CAISO historical and the PLEXOS simulation in terms of annual average energy production, broken out by energy source. In Fig. 13, we show that PLEXOS largely reproduces historical trends on a month-hour average basis for 5-minute energy prices, thermal generation, and renewable curtailment. Of note, energy price differences in February are higher in 2019 CAISO historical data due to a gas pipeline outage that caused higher than usual natural gas prices. We do not recreate this outage in PLEXOS because we do not want our analysis to reflect anomalous gas grid conditions, and therefore, we see lower energy prices in February.

#### 2. Battery state of charge constraints

In our PLEXOS modeling, batteries can provide flexible ramping capacity, regulation reserves, and spinning reserves in addition

TABLE IV. Coverage and pinball loss results for each prediction target and target quantile for 2019 data.

Metric			Target coverage (%)							
	Prediction target	2.5%	5%	25%	50%	75%	95%	97.5%		
Achieved coverage (%)	Net demand	2.5%	4.8%	24.4%	49.9%	75.3%	95.0%	97.5%		
	Demand	2.4%	4.8%	24.8%	50.8%	74.9%	94.9%	97.5%		
	Solar	3.1%	5.1%	23.0%	51.2%	73.9%	94.9%	97.3%		
	Wind	2.6%	4.9%	25.7%	48.6%	75.2%	95.3%	97.3%		
Pinball loss (MW)	Net demand	12.7	19.5	46.5	56.3	44.8	18.6	13.2		
	Demand	17.6	25.8	67.1	81.1	66.0	26.2	20.4		
	Solar	10.9	15.0	36.7	43.8	36.1	17.2	13.7		
	Wind	9.7	12.4	28.0	33.0	26.5	10.2	8.1		

J. Renewable Sustainable Energy 14, 036303 (2022); doi: 10.1063/5.0087144

Published under an exclusive license by AIP Publishing

scitation.org/journal/rse

		Target coverage of net demand (%)							
Metric	Fold #	2.5%	5%	25%	50%	75%	95%	97.5%	
Achieved coverage (%)	1	2.2%	3.7%	23.4%	53.7%	75.4%	95.7%	97.8%	
	2	1.5%	4.5%	24.7%	49.6%	78.0%	95.3%	97.8%	
	3	2.0%	5.7%	22.6%	47.0%	73.0%	94.8%	98.1%	
	4	3.3%	6.3%	26.9%	52.5%	74.0%	94.7%	97.2%	
	5	2.7%	4.3%	24.5%	49.2%	73.6%	94.0%	96.9%	
	6	2.0%	4.5%	23.6%	50.9%	74.4%	94.3%	97.0%	
	7	1.7%	3.8%	22.6%	47.6%	75.6%	94.5%	96.9%	
	8	3.3%	4.6%	26.9%	54.6%	73.4%	95.3%	97.8%	
	9	2.4%	5.2%	26.9%	52.2%	77.6%	95.5%	98.0%	
	10	2.6%	5.2%	25.5%	50.7%	74.1%	94.6%	97.6%	
Pinball loss (MW)	1	13.9	23.2	61.1	75.3	59.9	22.2	17.1	
	2	18.5	24.5	63.3	76.0	63.0	25.3	18.4	
	3	15.3	24.3	67.9	83.2	67.5	27.5	20.0	
	4	23.9	30.8	69.9	84.4	69.1	25.7	19.8	
	5	16.8	28.4	68.8	83.2	66.6	28.6	29.6	
	6	18.0	24.2	68.3	81.4	66.7	28.4	20.8	
	7	15.8	25.1	67.5	80.8	67.0	25.6	19.3	
	8	18.4	26.5	67.0	80.7	66.5	26.5	19.6	
	9	18.4	25.0	67.1	81.2	65.4	25.9	19.2	
	10	16.6	25.8	69.7	84.7	68.1	26.2	19.9	

TABLE V. Coverage and pinball loss results for each of the 10 folds for the net demand prediction target for 2019 data.

to charging and discharging. We impose a state of charge constraint on battery reserve and ramping capacity provision that ensures that the capacity could be called upon continuously for one hour at the reserved capacity. This assumption ensures that batteries are prepared for forecast errors or contingency events that span many consecutive 15- or 5-minute intervals but is not so restrictive as to exclude batteries from providing reserve and ramping capacity. For upward products (upward flexible ramping, spinning reserve, and regulation up), a state of charge constraint ensures that the storage resource has at least 1 MWh of energy that could be discharged for every MW of capacity committed. For downward products (downward flexible ramping and regulation down), a state of charge constraint ensures that the storage resource could accept least 1 MWh of charge for every MW of capacity committed. The impact of the battery state of charge constraint can be seen in the flexible ramping price charts in Fig. 7. In the 2030 High Battery Case, we observe flexible ramping prices that are materially above zero only during sunrise and sunset hours. We believe that these prices result from the economic incentive to fully discharge and charge batteries before sunrise and sunset, respectively. The battery state of charge requirements creates an opportunity cost for the battery to provide ramping and reserve flexibility because the battery must choose between charging/discharging and holding capacity for sub-hourly ramping/reserve needs. The flexible ramping price increases at sunrise and sunset in the upward and downward direction, respectively, are consistent with batteries trading off between energy and reserve scheduling.

#### 3. Implementation of flexible ramping requirements

We implement CAISO flexible ramping requirements as two reserve products in PLEXOS: upward flexible ramping and downward flexible ramping. The upward and downward requirements represent the uncertainty of the net demand forecast, as derived using either histogram or machine learning methods. Flexible ramping requirements are modeled for both the 15-minute and 5-minute dispatch stages; the flexible ramping requirement time series data are updated between the 15-minute and 5-minute stage. Because net demand uncertainty is typically smaller on the 5-minute timeframe, the 15-minute requirements are generally much larger than the 5-minute requirements. We focus on 15-minute flexible ramping requirements in this study and, therefore, use 2019 5-minute histogram requirements from CAISO OASIS for all 5-minute model runs.

Hydroelectric (including pumped storage), battery, and combined cycle gas turbines (CCGTs) can provide flexible ramping capacity in our PLEXOS model. CCGT and hydroelectric resources are limited in the amount of flexible ramping capacity that they can provide by generator ramp rate limits and the timeframe of the flexible ramping product (either 15- or 5-minute). Flexible ramping capacity reserved on each resource is separate and mutually exclusive from the three other reserve products modeled in PLEXOS: regulation up, regulation down, and spinning reserve.

Similar to CAISO's implementation of FRP constraints in their market optimization software, we include a price cap on the provision of FRP capacity. This cap ensures that energy dispatch is

ARTICLE

**TABLE VI.** Quantile crossing frequency and magnitude for the net demand prediction target for 2019 data. Blank cells indicate that no quantile crossings were observed. Quantile crossing is a comparison between two different prediction targets; diagonal terms are shaded to indicate that a comparison of a quantile to itself is not meaningful. While we observe some quantile crossing events between nearby quantiles, the frequency of these events decreases quickly as the distance between target quantiles increases. Our production cost study utilizes only a single extreme quantile in each direction (P2.5 and P97.5); quantile crossing is not observed between these quantiles.

		Target coverage of net demand (%)							
Metric	Target coverage of net demand (%)	2.5%	5%	25%	50%	75%	95%	97.5%	
Frequency of quantile	2.5%		20.8%	0.2%					
crossing (% of	5%	20.8%		1.0%					
15-minute intervals)	25%	0.2%	1.0%		2.1%	0.1%			
	50%			2.1%		3.1%	0.1%	0.6%	
	75%			0.1%	3.1%		0.6%	1.5%	
	95%				0.1%	0.6%		11.2%	
	97.5%				0.6%	1.5%	11.2%		
Average size of	2.5%		58	27					
quantile crossing	5%	58		34					
(MW)	25%	27	34		31	22			
	50%			31		25	41	109	
	75%			22	25		56	121	
	95%				41	56		78	
	97.5%				109	121	78		
Maximum size of	2.5%		491	217					
quantile crossing	5%	491		467					
(MW)	25%	217	467		239	89			
	50%			239		263	149	638	
	75%			89	263		322	985	
	95%				149	322		923	
	97.5%				638	985	923		



FIG. 12. Annual generation comparison between historical 2019 CAISO operations and PLEXOS benchmark simulation of 2019. PLEXOS data originates from the 5-minute dispatch stage.

prioritized over ramping capacity. We model a FRP price cap of \$1,000/MWh in the upward direction and \$155/MWh in the downward direction in the 15-minute stage and 5-minute stage.

## 4. 2030 histogram 15-minute ramping requirement approximation

We create a time series of 15-minute flexible ramping requirements that approximate what CAISO histogram flexible ramping requirements would be with a 2030 resource portfolio and demand forecast. As described below, 2030 histogram reserves are derived by scaling up the 2019 histogram reserves using the expected growth in net load forecast error. This growth in the forecast error is driven by the expected growth in load, wind and solar from 2019 to 2030. This growth in solar and wind capacity is shown in Fig. 6(a).

CAISO currently uses the difference between 15- and 5-minute market net load forecasts to derive histogram reserves. We used CAISO's OASIS portal to download binding interval 2019 15- and 5-minute flexible ramping product requirements, as well as load, wind, and solar profiles for the 15- and 5-minute markets. The wind and solar time series data from 2019 are, respectively, adjusted to 2030 levels by linearly scaling up 2019 output profiles in all intervals using the ratio of 2030 wind and solar capacities to the



FIG. 13. Month-hour average comparison between historical 2019 CAISO operations and PLEXOS benchmark simulation of 2019. PLEXOS data originates from the 5-minute dispatch stage.

2019 wind and solar capacities from the CAISO Master Control Area Generating Capability List.<sup>28,36</sup> We use linear scaling because there were already large amounts of solar and wind capacity online in 2019 in CAISO in geographically diverse regions, and the diversity benefit of new solar and wind additions is likely to be small as a result. The change in demand from 2019 to 2030 is derived by linearly scaling up 2019 demand from CAISO OASIS using the ratio of the month-hourly average demand in the 2030 year to the 2020 year, though the overall changes are small relative to that resulting from the growth in renewable generating capacity.<sup>23</sup> Using these scaled load, wind, and solar profiles, the 2019 CAISO histogram 15-minute flexible ramping requirement time series is scaled up to 2030 levels using the change in month-hourly average net demand forecast error between 2019 and 2030. 2030 15-minute ramping requirements are summarized on a month-hour basis in Fig. 6(b). We recognize that CAISO is planning to improve the histogram calculation method well before 2030. We did not have access to sample timeseries for the improved quantile regression method that CAISO is planning to release in the near term nor can we predict what longer term improvements CAISO will make by 2030. Thus, we employed this approximate method to scale up incumbent reserves.

#### 5. Hydroelectric and pumped hydroelectric resources

Historical 2019 data from the CAISO Daily Renewables Watch<sup>42</sup> is used to derive limitations on the aggregate fleet of CAISO hydroelectric resources. Hydroelectric and pumped hydroelectric resources, collectively referred to as "hydro" here, are modeled in aggregate in PLEXOS. The total energy production from hydro and pumped storage resources on each day from 2019 is used to set daily energy production constraints in PLEXOS. In addition, daily maximum and minimum fleet-wide output levels observed on each day in 2019 are used to set the operational range of the aggregated hydro units in PLEXOS. A maximum ramping rate for hydro is set using historical 2019 output data. The fleetwide energy budget, operational range, and ramp rate limits are assigned to Southern California Edison (SCE) and Pacific Gas and Electric (PG&E) resources on a hydroelectric plant capacityweighted share. San Diego Gas and Electric (SDG&E) hydro is not modeled because the nameplate capacity of SDG&E hydro resources is small relative to the CAISO fleet-wide capacity. Hydroelectric operational parameters derived from 2019 data are also used for 2030 simulations.

scitation.org/journal/rse

#### 6. Import and export schedules

We focus on our analysis on the CAISO system and perform unit commitment and dispatch for resources within the CAISO footprint. To represent ties with neighboring regions, we include fixed import and export schedules for each modeling stage based on historical 2019 interchanges. Two import/export schedules are derived: imports and exports from the Southwest (SW) are connected to the SCE zone and imports and exports from the Northwest (NW) are connected to the PG&E zone. Total net imports to CAISO from external balancing areas are derived from the EIA Form 930.43 Fifteen- and 5-minute market net imports are derived from the CAISO OASIS data<sup>28</sup> with Day + Hour ahead stage net imports calculated as the difference between total and sub-hourly net imports. Balancing area interchange schedules from CAISO OASIS are aggregated into NW and SW zones, creating two import/export schedules. Historical market prices from the Mid-Columbia and Palo Verde nodes are used as a component of the total production costs shown in Table III and Table VII; these historical market prices do not influence import and export dynamics because import/export schedules are not allowed to change in PLEXOS. We keep import/export schedules constant at 2019 levels for 2030 simulations.

#### 7. Cost treatment

All costs are reported in \$2019.

#### APPENDIX C: SUPPLEMENTAL PLEXOS ANALYSIS: SOLAR PROVIDING FLEXIBLE RAMPING CAPACITY

In this section, we provide an exploration of the value of utility-scale solar resources providing flexible ramping capacity.

As the amount of energy produced by variable renewable resources increases over time, so does the potential to use these resources for balancing and ramping. In the PLEXOS modeling presented in the main paper, we have not included variable renewable resources (wind and solar) as resources that can contribute to ramping requirements. Previous work has highlighted that variable renewable resources have the technical capabilities to provide shortterm balancing services<sup>44–46</sup> and has also demonstrated production cost and greenhouse gas emissions savings associated with their participation.<sup>25,47</sup> Despite these capabilities and potential savings, variable renewable resources do not currently provide meaningful contributions to ramping or operational reserve capacity in the CAISO or other organized electricity markets in the United States.

We use PLEXOS to simulate utility-scale solar, the variable renewable resource in the CAISO system with the largest installed capacity, providing 15-minute and 5-minute flexible ramping capacity (Solar FRP) in both 2019 and 2030. All Solar FRP simulations use machine learning flexible ramping requirements. We limit downward FRP from utility-scale solar to the solar production potential minus any curtailment, which is equivalent to the energy setpoint of the solar resource. We limit upward FRP from utility-scale solar to the level of solar curtailment because solar must be able to increase output to provide upward FRP and cannot do so if it is not curtailed. We do not limit solar FRP provision based on solar forecast error, which implies that our results are an upper bound on the benefits of solar FRP; in practice, system operators would need to consider solar forecast error when committing ramping capacity. In the future, it may be possible to use short-term probabilistic forecasts of variable renewable uncertainty, potentially provided by a machine learning model like the one presented in this paper, to bound the ability of variable renewable resources to provide flexibility to ramping and reserve products based on solar forecast uncertainty.

Production costs are reduced by 0.5% when solar provides flexible ramping for the 2030 Low Battery portfolio (Table VII). Solar can be particularly effective at providing downward ramping, as solar resources do not need to be pre-curtailed to do so. The marginal cost of meeting the downward flexible ramping requirement

TABLE VII. Savings from including utility-scale solar as a resource that can provide 15-minute flexible ramping capacity. Savings are calculated as the change in the 5-minute RTD stage that results from including solar as a resource that can provide flexible ramping in the upstream FMM stage as well as the 5-minute RTD stage.

		Difference: solar cannot provide flexible ramping minus solar can provide flexible ramping				
Metric	Units	2019	2030 low battery	2030 high battery		
Production cost savings	% of annual production cost	0.3%	0.5%	0.0%		
	\$M/yr	\$16	\$29	\$1		
Total cost savings (renewable curtailment reduction valued at \$18/MWh)	\$M/yr	\$21	\$43	\$2		
GHG savings	% of annual emissions	0.3%	0.7%	0.0%		
	MMTCO <sub>2</sub> /yr	0.1	0.3	0.0		
Natural gas generation	% of annual natural gas generation	0.6%	1.2%	0.2%		
reduction	GWh/yr	332	746	76		
Curtailment reduction	% of wind and solar generation potential	0.9%	1.0%	0.1%		
	GWh/yr	247	750	61		
Decrease in frequency of RT5 energy prices above \$150/MWh	% of 5-minute intervals	0.1%	0.1%	0.1%		

ARTICLE

(the downward FRP shadow price) drops steeply between a model run where solar cannot provide FRP and one in which it can (Fig. 14, top left), indicating that it is challenging to provide downward reserves in a highly renewable grid without large contributions from either solar or battery resources, especially during periods of curtailment. For a solar resource to provide upward ramping, it must be pre-curtailed; this requirement is frequently met in the 2030 Low Battery portfolio because the relative lack of battery capacity creates long periods of daytime renewable curtailment. With the 2030 Low Battery portfolio, solar FRP is effective at reducing the cost to provide upward ramping capacity (Fig. 14, top right) because if already curtailed, solar generation has no marginal cost to provide upward ramping.

When we increase the capacity of battery storage by moving to the 2030 High Battery portfolio, the incremental value of solarprovided FRP drops (Table VII), which is a result that is broadly consistent with recent work from the National Renewable Energy Laboratory.<sup>27</sup> As we have previously discussed, the marginal cost to provide flexible ramping capacity can approach zero in most hours with high enough levels of battery capacity; adding solar flexible ramping to a system with low FRP procurement costs results in low benefits of doing so. The bottom left panel of Fig. 14 shows that solar FRP reduces downward FRP prices during sunset; by sunset, batteries need to be fully charged to prepare for a full discharge cycle in the nighttime. Batteries cannot simultaneously be fully charged and provide downward ramping because providing downward ramping requires the batteries to be prepared for additional charging; solar can lower costs by providing downward ramping near sunset. Solar is not frequently curtailed near sunrise and is, therefore, not particularly effective at reducing the early morning upward FRP cost (bottom right panel of Fig. 14).

Reflecting a moderate near-term level of value, savings from solar FRP with the 2019 resource portfolio are found to be intermediate relative to the 2030 Low and High Battery portfolios (Tables VII).

#### APPENDIX D: SUPPLEMENTAL PLEXOS RESULTS: PRODUCTION COST SAVINGS BY RESOURCE CATEGORY

Production cost savings by resource category for results shown in Tables III and VII.



FIG. 14. 15-minute flexible ramping requirement prices for the 2030 Low Battery (top) and High Battery portfolios (bottom), with and without solar as a resource that can provide 15-minute flexible ramping capacity (FRP). Prices are presented as an hourly average over the year.

ARTICLE

TABLE VIII. Production cost savings by resource category for results shown in Tables III and VII. Imports and export cost differences are not shown because the level of imports and exports is held constant and, therefore, have zero cost difference between cases. Totals may deviate slightly from individual cost components due to independent rounding.

	Cost difference (\$M/yr): histogram minus machine learning (Table III)				Cost difference (\$M/yr): solar cannot provide flexible ramping minus solar can provide flexible ramping (Table VII)			
Cost category	2019	2030 low battery	2030 high battery	2019	2030 low battery	2030 high battery		
Gas combined cycle	33	1	5	19	38	6		
Gas combustion turbine	-12	12	-5	-1	-2	-5		
Gas combined heat and power	-7	0	0	-2	-6	0		
Production cost subtotal	15	13	0	16	29	1		
Renewable curtailment	4	29	0	5	14	1		
Total	19	42	0	21	43	2		

#### REFERENCES

- <sup>1</sup>J. Williams, B. Haley, F. Kahrl, J. Moore, A. D. Jones, M. S. Torn, and H. McJeon, "Pathways to deep decarbonization in the United States." The U.S. report of the Deep Decarbonization Pathways Project of the Sustainable Development Solutions Network and the Institute for Sustainable Development and International Relations (2015), see https://irp-cdn.multiscreensite.com/ be6d1d56/files/uploaded/DDPP 2015 REPORT.pdf.
- <sup>2</sup>J. Williams, A. DeBenedictis, R. Ghanadan, A. Mahone, J. Moore, W. Morrow III, S. Price, and M. Torn, "The technology path to deep greenhouse gas emissions cuts by 2050: The pivotal role of electricity," Science 335(6064), 53-59 (2012).
- <sup>3</sup>Y. Sun, G. Szűcs, and A. R. Brandt, "Solar PV output prediction from video streams using convolutional neural networks," Energy Environ. Sci. 11, 1811 (2018).
- <sup>4</sup>R. Marquez and C. Coimbra, "Intra-hour DNI forecasting based on cloud tracking images 515 analysis," Sol. Energy 91, 327 (2013).
- <sup>5</sup>See Energy Exemplar, https://energyexemplar.com/products/plexos for "PLEXOS Software version 8.3, 2021;" accessed 21 December 2021.
- <sup>6</sup>See K. Westendorf, https://www.caiso.com/Documents/FlexibleRamping ProductUncertaintyCalculationImplementationIssues.pdf for "Flexible ramping product uncertainty calculation and implementation issues, 2018;" accessed 21 December 2021.
- <sup>7</sup>I. Krad, E. Ibanez, and W. Gao, "A comprehensive comparison of current operating reserve methodologies," in 2016 IEEE/PES Transmission and Distribution Conference and Exposition (T&D), 2016.
- <sup>8</sup>E. Spyrou, V. Krishnan, Q. Xu, and B. F. Hobbs, "What is the value of alternative methods for estimating ramping needs?," in 2020 IEEE Green Technologies Conference (IEEE, 2020), pp. 159-164.
- <sup>9</sup>See Electricity Reliability Council of Texas ERCOT, http://www.ercot.com/content/wcm/key\_documents\_lists/137978/9\_2019\_Methodology\_for\_Determining\_ Minimum\_Ancillary\_Service\_Requirements.pdf for "Item 9: 2019 methodology for determining minimum ancillary service requirements;" accessed 21 December 2021.
- <sup>10</sup>S. A. Fatemi, A. Kuh, and M. Fripp, "Parametric methods for probabilistic forecasting of solar irradiance," Renewable Energy 129, 666 (2018).
- $^{11}$ I. Takeuchi, Q. Le, T. Sears, and A. Smola, "Nonparametric quantile estimation," J. Mach. Learn. Res. 7, 1231 (2006), see https://www.jmlr.org/ papers/volume7/takeuchi06a/takeuchi06a.pdf. <sup>12</sup>P. Lauret, M. David, and H. T. Pedro, "Probabilistic solar forecasting using
- quantile regression models," Energies 10(10), 1591 (2017).
- <sup>13</sup>A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," IEEE Trans. Neural Networks 22(3), 337 (2010).
- <sup>14</sup>F. Liu, C. Li, Y. Xu, G. Tang, and Y. Xie, "A new lower and upper bound estimation model using gradient descend training method for wind speed interval prediction," Wind Energy 24(3), 290 (2021).

- 15 F. Golestaneh, P. Pinson, and H. B. Gooi, "Very short-term nonparametric probabilistic forecasting of renewable energy generation-With application to solar energy," IEEE Trans. Power Syst. 31(5), 3850 (2016).
- <sup>16</sup>N. Meinshausen and G. Ridgeway, "Quantile regression forests," J. Mach. Learn. Res. 7, 983 (2006), see https://www.jmlr.org/papers/volume7/meinshausen06a/ meinshausen06a.pdf.
- <sup>17</sup>A. Mayr, "Prediction inference with ensemble methods," Doctoral dissertation (Institut Für Statistik, 2010).
- 18 See Electric Power Research Institute, https://www.epri.com/research/programs/067417/ for "Dynamic assessment and determination of operating reserve;" accessed 21 December 2021.
- <sup>19</sup>C. Zhao and C. Wan, "Operating reserve quantification using prediction intervals of wind power: An integrated probabilistic forecasting and decision methodology," IEEE Trans. Power Syst. 36(4), 3701 (2021).
- <sup>20</sup>M. Mattos, R. Bessa, A. Botterud, and Z. Zhou, *Renewable Energy Forecasting*: From Models to Applications (Woodhead Publishing, 2017), pp. 279-308.
- <sup>21</sup>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 124, 272 (2019).
- 22See J. Novacheck, G. Brinkman, and G. Porro, https://www.nrel.gov/docs/ fy18osti/71465.pdf for "Operational analysis of the eastern interconnection at very high renewable penetrations, 2018;" accessed 5 May 2022.
- 23 A. Mileva, J. Johnston, J. H. Nelson, and D. Kammen, "Power system balancing for deep decarbonization of the electricity sector," Appl. Energy 162, 1001-1009 (2016).
- 24See Energy and Environmental Economics, Inc., https://www.ethree.com/wpcontent/uploads/2017/02/PacifiCorp-ISOEnergyImbalanceMarketBenefits-1.pdf for "Pacificorp-ISO energy imbalance market benefits, 2013," accessed 5 May 2022.
- 25 J. Nelson, S. Kasina, J. Stevens, J. Moore, and A. Olson, "Investigating the economic value of flexible solar power plant operation" (Energy and Environmental Economics, Inc., 2018), see https://www.ethree.com/wpcontent/uploads/2018/10/Investigating-the-Economic-Value-of-Flexible-Solar-Power-Plant-Operation.pdf.
- <sup>26</sup>See J. Nelson and L. Wisland, https://www.ucsusa.org/sites/default/files/attach/ 2015/08/Achieving-50-Percent-Renewable-Electricity-In-California.pdf for "Achieving 50 percent renewable electricity in California," (2015); accessed 5 May 2022.
- <sup>27</sup>B. Frew, B. Sergi, P. Denholm, W. Cole, N. Gates, D. Levie, and R. Margolis, "The curtailment paradox in the transition to high solar power systems," Joule 5, 1143 (2021).
- 28 See California Independent System Operator, http://oasis.caiso.com/mrioasis/ logon.do for "California ISO open access same-time information system, 2021;" accessed 21 December 2021.
- 29 See Tensorflow, https://www.tensorflow.org/ for "An end-to-end open source platform for machine learning;" accessed 27 December 2021.
- 30 See Energy and Environmental Economics, Inc., https://github.com/e3-/ RESERVE for "E3 RESERVE model;" accessed 21 December 2021.
- J. Renewable Sustainable Energy 14, 036303 (2022); doi: 10.1063/5.0087144

- <sup>31</sup>D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in International Conference on Learning Representations, 2015.
- <sup>32</sup>See California Independent System Operator, http://www.caiso.com/Initiative Documents/AppendixC-QuantileRegressionApproach-FlexibleRampingProduct Requirements.pdf for "Flexible ramping product refinements initiative, appendix C—Quantile regression approach, 2020;" accessed 21 December 2021.
- <sup>33</sup>See North American Electric Reliability Corporation, https://www.nerc.com/ pa/Stand/Reliability%20Standards/BAL-001-2.pdf for "Standard BAL-001-2— Real power balancing control performance, 2020;" accessed 21 December 2021.
- <sup>34</sup>See California Independent System Operator, http://www.caiso.com/Initiative Documents/Analysis-FlexibleRampingUncertaintyCalculationintheWesternEnergy ImbalanceMarket.pdf for "Flexible ramping uncertainty calculation in the western energy imbalance market (EIM), 2022;" accessed 9 May 2022.
- <sup>35</sup>See California Independent System Operator, https://bpmcm.caiso.com/Lists/ PRR%20Details/Attachments/479/

Market%20Operations%20BPM%2072hr% 20RUC%20PRR.pdf for "Market processes and products;" accessed 21 December 2021.

- <sup>36</sup>See California Public Utilities Commission, https://www.cpuc.ca.gov/industriesand-topics/electrical-energy/electric-power-procurement/long-term-procurementplanning/2019-20-irp-events-and-materials/portfolios-and-modeling-assumptionsfor-the-2021-2022-transmission-planning-process for "RESOLVE transmission preferred portfolio public release 2020–2021: 46 MMT scenario, 2021;" accessed 21 December 2021.
- <sup>37</sup>See Southern California Edison, https://www.sce.com/regulatory/tariff-books/ rates-pricing-choices/renewable-energy-credit for "Renewable Energy Credit (REC);" accessed 16 November 2021.
- <sup>38</sup>See Western Energy Imbalance Market, https://www.westerneim.com/Pages/ About/QuarterlyBenefits.aspx for "ISO EIM benefits report Q1–Q4 2019;" accessed 21 December 2021.
- <sup>39</sup>See California Public Utilities Commission Energy Division, https:// www.cpuc.ca.gov/-/media/cpuc-website/divisions/energy-division/documents/

integrated-resource-plan-and-long-term-procurement-plan-irp-ltpp/2019-2020-irpevents-and-materials/2019-irp-proposed-reference-system-plan\_20191106.pdf for "2019–20 IRP: Proposed reference system plan, 2019;" accessed 16 November 2021.

- <sup>40</sup>V. Nair and G. Hinton, "Rectified linear units improve restricted Boltzmann machines," in Proceedings of the 27th International Conference on Machine Learning, 2010.
- <sup>41</sup>Y. Bengio, "Practical recommendations for gradient-based training of deep architectures," *Neural Networks: Tricks of the Trade* (Springer, Berlin/ Heidelberg, 2012), pp. 437–478.
- <sup>42</sup>See CAISO, http://www.caiso.com/market/Pages/ReportsBulletins/Renewables Reporting.aspx for "Renewables and emissions reports, 2021;" accessed 21 December 2021.
- <sup>43</sup>See U.S. Department of Energy's Energy Information Agency, https://www.eia. gov/electricity/gridmonitor/about for "Hourly grid electricity monitor" (2020).
- <sup>44</sup>C. Loutan, P. Klauer, S. Chowdhury, S. Hall, M. Morjaria, V. Chadliev, N. Milam, C. Milan, and V. Gevorgian, "Demonstration of essential reliability services by a 300-MW solar photovoltaic power plant," Report No. NREL/TP-5D00-67799 (NREL, Boulder, CO, 2017).
- <sup>45</sup>C. Loutan, V. Gevorgian, S. Chowdhury, M. Bosanac, E. Kester, D. Hummel, R. Leonard, M. Rutemiller, J. Fregoe, D. Pittman, and C. Kosuth, "Avangrid renewables Tule Wind Farm: Demonstration of capability to provide essential grid services" (CAISO, Folsom, CA, 2020), see http://www.caiso.com/ documents/ windpowerplanttestresults.pdf.
- <sup>46</sup>P. L. Denholm, Y. Sun, and T. T. Mai, "An introduction to grid services: Concepts, technical requirements, and provision from wind," Report No. NREL/TP-6A20-72578 (National Renewable Energy Lab, 2019).
- <sup>47</sup>F. Kahrl, J. H. Kim, A. Mills, R. Wiser, C. C. Montañés, and W. Gorman, "Variable renewable energy participation in U.S. ancillary services markets" (Lawrence Berkeley National Laboratory, 2021), see <u>https://escholarship.org/uc/item/29r4z5mx</u>.