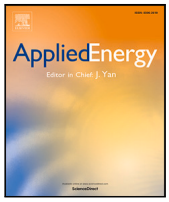
Applied Energy 313 (2022) 118812

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



Sizing ramping reserve using probabilistic solar forecasts: A data-driven method

Binghui Li [a], Cong Feng [b], Carlo Siebenschuh [c], Rui Zhang [c], Evangelia Spyrou [b], Venkat Krishnan [b], Benjamin F. Hobbs [d], Jie Zhang [a,*]

[a] The University of Texas at Dallas, Richardson, TX 75080, United States of America
[b] National Renewable Energy Laboratory, Golden, CO 80401, United States of America
[c] IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, United States of America
[d] The Johns Hopkins University, Baltimore, MD 21218, United States of America

ARTICLE INFO

Keywords:
Probabilistic forecast
k-nearest neighbors
Flexible ramping product
Solar power forecast
Flexibility
Reliability

ABSTRACT

Ramping products have been introduced or proposed in several U.S. power markets to mitigate the impact of load and renewable uncertainties on market efficiency and reliability. Current methods often rely on historical data to estimate the requirements of ramping products and fail to take into account the effects of the latest weather conditions and their uncertainties, which could lead to overly conservative or insufficient requirements. This study proposes a k-nearest-neighbor-based method to give weather-informed estimates of ramping needs based on short-term probabilistic solar irradiance forecasts. Forecasts from multiple sites are employed in conjunction with principal component analysis to derive numerical classifiers to characterize system-level weather conditions. In addition, we develop a data-driven method to optimize the model parameters in a rolling-forward manner. By using real-world data from the California Independent System Operator, we design two metrics to evaluate method performance: 1) frequency of shortage and 2) oversupply of ramping product. Our proposed method presents advantages in comparison with the baseline and a set of benchmark methods: without compromising system reliability, it reduces system ramping requirements by up to 25%, therefore improving both system reliability and economics.

1. Introduction

Ramping products have been introduced or proposed in several U.S. power markets to mitigate the impact of rapid net load changes caused by the increasing penetration of variable energy sources [1–5]. Like many ancillary services, a key step in the scheduling process is to determine the requirements of ramping product, which usually involves trade-offs between market efficiency and system reliability. Too low or too high requirements can pose a risk to system reliability or system costs, respectively.

Historically, studies that co-optimize energy and ancillary services can be divided into two categories based on how the requirements are determined: endogenous and exogenous [6,7]. Endogenous methods usually model the renewable and load uncertainties explicitly using a variety of techniques [8], such as stochastic optimization [9–11], robust optimization [12,13], and chance constraints [14,15]. These techniques typically represent load and renewable energy uncertainties in the form of probability distribution functions [16,17], scenarios [11,18], uncertainty sets [13], or intervals [19], depending on the algorithms used to solve the models. These models commonly include reliability targets as additional constraints, and the amount of required reserves are then given when the model is solved [20]. Typical reliability metrics include the loss-of-load-probability (LOLP) and the expected energy not served (EENS) [21–24]. These metrics can be used directly to construct model constraints, or used in conjunction with value of lost load (VOLL) to create additional terms in the model objective function by monetizing the loss [15]. The trade-offs between reliability and system economics are usually examined to select the optimal reserve requirements [25]. For example, in [14], a chance-constrained model was used to optimize the needs of flexibility reserve based on a combined metric that accounts for both procurement costs and loss costs due to wind spillage and load shedding. A two-stage stochastic optimization model was developed in [26] to explicitly consider the provision of operating reserves, which outperformed traditional methods where reserve requirements were given by certain percentages of peak load. In [15], flexible ramping requirements were endogenously modeled by considering the spatio-temporal correlation of wind power outputs

* Corresponding author.
E-mail address: jiezhang@utdallas.edu (J. Zhang).

https://doi.org/10.1016/j.apenergy.2022.118812
Received 5 November 2021; Received in revised form 28 January 2022; Accepted 20 February 2022
Available online 10 March 2022

Nomenclature

ANN	Artificial neural network
BAU	Business-as-usual
CAISO	California Independent System Operator
CART	Classification and regression tree
CNN	Convolutional neural network
DL	Deep learning
EENS	Expected energy not served
FRP	Flexible ramping product
GBM	Gradient boosting machine
GHI	Global horizontal irradiance
GRU	Gated recurrent unit
kNN	k-nearest-neighbor
LOLP	Loss-of-load probability
LSTM	Long short-term memory
MAE	Mean absolute error
ML	Machine learning
nMAE	Normalized mean absolute error
PCA	Principal component analysis
PV	Photovoltaics
RF	Random forest
RT	Real time
RTED	Real-time economic dispatch
RTUC	Real-time unit commitment
SVR	Support vector regression
VOLL	Value of lost load

and demand, which are represented by covariance matrices and solved as quadratically constrained programs. Ghaemi et al. [19] represented solar and wind power output forecasts using intervals and solved them by interval optimization. Nadermahmoudi et al. [18] represented load and wind uncertainties by scenarios and obtained the FRP provided by each generating unit in their studied system. In [21], a risk-limiting economic dispatch scheme optimized the dispatch and provision of ramping products under constraints upon LOLP. Although these models can securely co-optimize the provision of energy and reserves of all generating units, detailed uncertainty representation is usually required at the bus level, which is computationally prohibitive for real-world systems because of the curse of dimensionality. Note that although endogenous methods have been proposed by many studies, they are rarely adopted by real-world market operators.

Although these models can in theory identify a level of reserves that is optimal for an assumed mathematical representation of system reliability, in practice they can prove computationally prohibitive for real-world systems because of the high dimension of scenarios needed for uncertainty representation. The computational burden of endogenous methods is at least one of the reasons that explains why none of the US power system operators has followed such a method. Instead, power system operators rely on exogenous methods. The exogenous methods estimate system-level ancillary service requirements *ex ante* with lead times specific to particular markets. Unlike the endogenous methods, this approach has been widely adopted in real world and is constantly evolving. Traditionally, system operators use deterministic methods, where certain percentages of peak load are used as reserve margins [27–29]. This method, despite its simplicity, usually fails to satisfy system reliability standards owing to growing uncertainties caused by increasing renewable penetration. Consequently, probabilistic methods are sometimes adopted to estimate ancillary service needs based on the uncertainty levels of net load [30]. In these methods, reserve requirements are determined to cover predefined confidence intervals, which are usually selected to satisfy certain reliability standards [30]. Typically, the distributions are constructed from historical data. For example, historical area control errors are used to determine regulation requirements in the California Independent System Operator (CAISO), and historical net load forecast errors are used in the Texas power market to estimate non-spinning reserve requirements [31]. [17] used Gumbel copula to estimate the probabilistic distributions of net load and set FRP requirements based on the derived uncertainty demand curve.

Historical distributions are nevertheless highly static and usually fail to reflect the latest system states and weather information, resulting in overly conservative or insufficient requirements. CAISO conducted a simulation study for 2019 and estimated that the coverage of requirements determined through historical analysis was higher than 95% (target reliability level) [32]. The lack of weather information has motivated system operators in the U.S. [32] and abroad [33] to adopt methods that use wind, load, and solar forecasts as inputs to statistical approaches to estimate reserve requirements. In [32], CAISO used quantile regression with load, wind, and solar forecasts as inputs to estimate FRP requirements and results showed that it would result in higher or lower requirements compared to the historical approach depending on system conditions. In [34], a multiple linear regression model was developed to estimate regulating reserve requirements based on predicted system conditions, including load and wind power. In the recent decade, the value of probabilistic forecasts is being increasingly recognized in power system applications [35]. In [36], probabilistic wind and load forecasts were used to determine operating reserve requirements based on trade-offs between reliability indices and procurement costs. In [37,38], Etingov et al. showed how to reduce the regulation requirements by employing probabilistic net load forecasts, which were derived from probability distributions of multiple net load components by convolution. Instead of performing convolution, [39] also used probabilistic forecasts as input for generation of scenarios that were later used to determine operating reserve requirements.

Probabilistic forecasts explicitly account for the uncertainty information, which usually take a parametric form of probability distributions, or non-parametric forms such as quantiles, uncertainty intervals, and kernel density estimations [40,41]. Although probabilistic forecasting is adopted in some markets, it is used primarily to improve situational awareness and rarely plays any important role in the decision-making process. In this study, we develop a data-driven flexible ramping product (FRP) requirement estimation methodology based on probabilistic solar irradiance and power forecasts to dynamically determine the system-level FRP requirements of CAISO. Specifically, 2-hour-ahead (2HA) probabilistic solar forecasts at selected sites in the CAISO region are synthesized into a set of numerical classifiers to characterize future weather conditions. We employ principal component analysis (PCA) to reduce the dimension of high-dimensional classifiers. A k-nearest-neighbor-based (kNN-based) method is then used to construct distributions of net load forecast errors conditioned on the latest weather conditions, which are used to give weather-informed FRP requirements. Note that because of the high solar penetration in the CAISO market, this study focuses primarily on solar uncertainty, however, the proposed method can be extended by accounting for other components of net load uncertainty, e.g., wind and load [42].

Our work represents a practical method that is easily implementable and extensible by any power market operators. Particularly, the proposed kNN-based method can also be used to define requirements for other market products that depend on historical data, such as regulation and non-spinning reserves. The main contributions of this paper include: (1) developing a systematic way to prepare, process, and extract key features from probabilistic solar forecasts, which can be used by market operators in conjunction with data-driven methods to make weather-informed decisions; (2) demonstrating the trade-offs between system reliability and market efficiency as different kNN parameters are selected; and (3) modeling the selection of the optimal kNN parameters

as a multi-objective optimization problem and developing a method to dynamically select the optimal solutions.

The remainder of this paper is organized as follows. Section 2 describes the current implementation of FRP in CAISO (baseline) and the proposed kNN-based method. Section 3 further details the experimental setup, including the definition of numerical classifiers, reduction of multi-dimensional classifiers, dynamic selection of model parameters, and benchmark models. Section 4 presents the results using metrics indicative of system reliability and oversupply. Section 5 concludes.

2. Methodology

The FRP is designed to help power systems mitigate the impact of net load changes on system operations. Specifically, market operators reserve sufficient rampable capacities from controllable resources for future dispatch intervals to counteract expected and unexpected net load changes and maintain power balance in compliance with reliability standards. By definition, the FRP requirements consist of two separate parts: (1) the variation part, which is the system-wide net load movement between two adjacent time intervals, and (2) the uncertainty part, which represents the net load forecast error [1,3]. Currently, both CAISO and the Midcontinent Independent System Operator determine upward and downward FRP requirements independently, with separate prices and quantities, and they rely on historical data to estimate the uncertainty requirements. This study focuses on the uncertainty component of the FRP in the 15-min real-time (RT) market in CAISO. Note that the method can be extended to the other markets with similar definitions.

Because net load is expressed as a linear combination of gross system-wide load, wind power, and solar power, net load forecast errors are driven by forecast errors of all terms. Forecast errors of renewable resources, however, typically account for disproportionally greater shares. It is estimated that CAISO's solar power uncertainty contributes at least half of overall net load uncertainty [43], therefore, we focus on the uncertainty component caused by solar power only, and we assume perfect foresight for wind and load. Although neglecting wind and load uncertainties could lead to underestimation of total net load uncertainty, it allows us to better demonstrate our method, and we do not think it introduces systematic errors.

2.1. Baseline: From the current CAISO implementation

In CAISO, the FRP in the 15-min RT market aims to cover the difference between net load forecasts made in the 15-min real-time unit commitment (RTUC) and in the 5-min real-time economic dispatch (RTED). Because each 15-min interval consists of three 5-min intervals, CAISO calculates three net load forecast errors for each 15-min interval—i.e., the differences between the binding net load forecasts of the constituent 5-min intervals and the advisory net load forecast of the 15-min interval, as shown in Fig. 1. Then, CAISO uses the maximum positive error as the upward error ($\epsilon^U_{d,t}$) and the minimum negative error as the downward error ($\epsilon^D_{d,t}$), which are also the FRP needs. They can be expressed as:

$$\epsilon^U_{d,t} = \max_{i=0,\dots,2} (\widehat{NL}^5_{d,3t-i} - \widehat{NL}^{15}_{d,t}) \tag{1}$$

$$\epsilon^D_{d,t} = \min_{i=0,\dots,2} (\widehat{NL}^5_{d,3t-i} - \widehat{NL}^{15}_{d,t}) \tag{2}$$

where t is the index of the 15-min RTUC interval in day d; and $3t-i, i = 0,\dots,2$ are the indices of the three constituent RTED intervals. The net load forecast for the RTUC and RTED intervals are denoted by $\widehat{NL}^{15}_{d,t}$ and $\widehat{NL}^5_{d,3t-i}$, respectively.

Using the net load forecast errors defined, we can construct separate histograms from history data to determine the upward and downward FRP requirements. In CAISO, histograms are constructed hourly by using historical net load forecast errors in the same hour from 40 previous weekdays if d is a weekday, or 20 previous weekends if d is a weekend. CAISO distinguishes between weekdays and weekends because human behaviors result in different load profiles for these two sets of days. Because our study assumes perfect foresight for load, and because solar uncertainty is not affected by human behavior, we make no such distinction in our baseline, i.e., the number of historical days K is the same for both weekdays and weekends. Note that although K is fixed in CAISO's implementation, we wish to examine the sensitivity of the baseline to K over a range of different values, therefore, given hour h in day d, we construct a histogram of upward errors from K previous days and derive the following empirical distribution function:

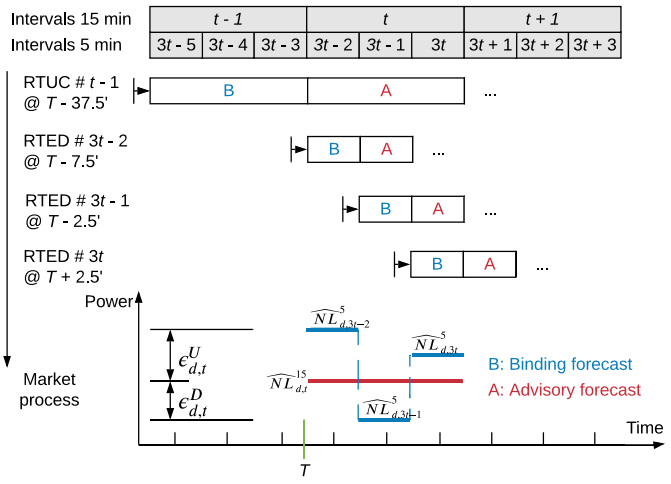


Fig. 1. Net load forecast error in CAISO's 15-min RT market. Note that T is the starting time of the tth 15-min interval. FRP requirements are needed at least 37.5 min prior to T to run RTUC # $t-1$.

$$\hat{F}^U_{d,h}(x) = \mathbb{E}_{\epsilon \in \mathcal{E}^U_{d,h}} [\mathbf{1}(x \geq \epsilon)] \tag{3}$$

where $\mathbf{1}(x \geq \epsilon)$ is the indicator function, $\mathbb{E}$ is the expectation function, and $\mathcal{E}^U_{d,h}$ represents the set of $4K$ historical samples of maximum net load forecast errors in hour h from K previous days (note that one hour includes four upward errors):

$$\mathcal{E}^U_{d,h} = \{\epsilon^U_{d-i,4h-j} | i = 1,\dots,K, j = 0,\dots,3\} \tag{4}$$

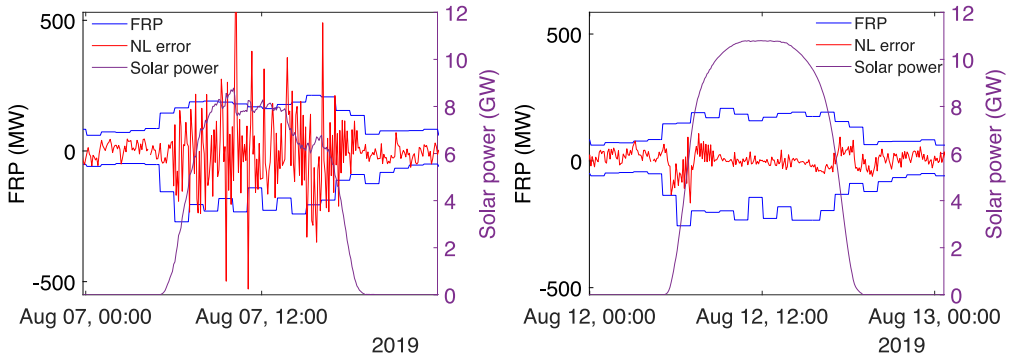
We use the 97.5th percentile as the upward FRP requirement by taking the inverse of $\hat{F}^U_{d,h}$:

$$\widehat{FRP}^U_{d,h} = \hat{F}^{U-1}_{d,h}(0.975) \tag{5}$$

Similarly, we construct the empirical distribution function of downward errors $\hat{F}^D_{d,h}(x)$ and use the 2.5th percentile as the downward FRP requirement:

$$\widehat{FRP}^D_{d,h} = \hat{F}^{D-1}_{d,h}(0.025) \tag{6}$$

CAISO uses historical data to construct histograms of net load forecast errors because of its simplicity to implement; however, the historical data often fails to reflect the latest weather conditions, potentially resulting in overly conservative or insufficient ramping needs. Fig. 2 compares the realized net load forecast errors with the published FRP requirements from 2 days in August 2019 in CAISO. Because the 2 days are close in time (5 days apart), the FRP requirements differ by less than 1% because of similar histograms in use, whereas the solar power profiles imply drastically different weather conditions, which potentially explain the greater uncertainty needs in the cloudy day than the sunny day; therefore, if similar FRP amounts are procured, the system could experience a shortage of FRP in the cloudy day, hence a compromised reliability level, and the over-procurement in the sunny



(a) A cloudy day. (b) A sunny day.

Fig. 2. Comparison of realized net load forecast errors with the published FRP requirements in CAISO.

day could incur unnecessary costs. This observation motivates the need for the latest weather information in estimating FRP requirements, which we pursue in this paper using probabilistic solar irradiance forecasts.

2.2. The kNN-based method

The kNN-based method can be viewed as a direct extension of CAISO's original implementation since both methods rely on historical data, however, in contrast to the baseline, the kNN-based method constructs weather-conditioned histograms by using probabilistic solar forecasts. As a non-parametric classification technique, kNN has been widely adopted to improve the performance of forecasts in many fields, such as wind speed, wind power, solar irradiance, and solar power [44–48]. The main idea of kNN is to label all training data with select features and pinpoint a new data point by using the K closest neighbors in the feature space [49]. Using the closest neighbors instead of the entire training data set can potentially improve the forecast performance.

In our case, the data points to be predicted are net load forecast errors, which are affected by weather conditions; therefore, given prior net load forecast errors and their associated weather conditions, we use the weather forecasts of a future hour to identify the K most similar historical hours and use their net load forecast errors to construct predictive distributions of the future FRP requirements. The process of the kNN-based method is given below.

2.2.1. Characterize the weather conditions

Power output from photovoltaic (PV) plants is affected by many factors, including technical specifications of solar panel arrays and meteorological parameters. Because the conversion from solar irradiance to electricity is completely deterministic and predictable, solar irradiance uncertainty is the primary source of uncertainty. Here, we use probabilistic solar irradiance forecasts at selected sites in CAISO to characterize the weather conditions. Given hour h in day d, we define a set of numerical classifiers $\mathbf{x}^c_{d,h} \in \mathbb{R}^n$ based on the forecasts within that hour. Detailed derivation of all classifiers are given in 3.1.

2.2.2. Distance measures

To use the kNN method, we need to choose a distance measure to characterize the similarity of any two classifiers. Because classifiers are in the form of n-dim vectors, we use Euclidean distance in this study. Using classifier c, the similarity of the weather conditions between the target hour (d_0, h) and a historical hour (d_i, h) is:

$$\ell^c_{(d_0,d_i),h} = \|\mathbf{x}^c_{d_0,h} - \mathbf{x}^c_{d_i,h}\|_2 \tag{7}$$

2.2.3. Determine the FRP requirements

Upon calculation of the distances, we sort all historical days in an ascending order based on their similarities to the target hour in terms of $\ell^c_{(d_0,d_i),h}$, and we select the top K days:

$$\mathcal{D}^{c,K}_{d_0,h} = \{d_1, \dots, d_K | \ell^c_{(d_0,d_1),h} \le \ell^c_{(d_0,d_2),h} \le \cdots \le \ell^c_{(d_0,d_K),h}\} \tag{8}$$

Then, the set of upward and downward net load forecast errors conditioned on the solar forecasts are given by:

$$\begin{aligned} \mathcal{E}^{U,c,K}_{d_0,h} &= \{\epsilon^U_{d,4h-j} | d \in \mathcal{D}^{c,K}_{d_0,h}, j = 0, \dots, 3\} \\ \mathcal{E}^{D,c,K}_{d_0,h} &= \{\epsilon^D_{d,4h-j} | d \in \mathcal{D}^{c,K}_{d_0,h}, j = 0, \dots, 3\} \end{aligned} \tag{9}$$

Last, $\hat{F}^U_{d,h}(x)$ and $\hat{F}^D_{d,h}(x)$ can be given by (3), from which $\widehat{FRP}^U_{d,h}$ and $\widehat{FRP}^D_{d,h}$ can be calculated.

3. Experiments

3.1. Definition of classifiers

Ground solar irradiance is jointly determined by sun position and local cloudiness levels. Because the sun positions can be accurately calculated given the time of year, cloud-induced uncertainty is the major contributing factor to solar irradiance uncertainty. In the literature, the clear-sky index is often used to reflect cloudiness levels [50]. It is obtained by normalizing measured global horizontal irradiance (GHI) by clear-sky GHI. Here, the predicted clear-sky index in a 15-min interval t in day d can be given by:

$$k^\alpha_{d,t} = \frac{I^\alpha_{d,t}}{I^0_{d,t}} \tag{10}$$

where $I^0_{d,t}$ denotes clear-sky GHI and $I^\alpha_{d,t}$ is the αth percentile of the GHI prediction. Because $k^\alpha_{d,t}$ could approach infinity as a result of low $I^0_{d,t}$ magnitude during sunrise and sunset, we consider only the time intervals when the solar elevation is greater than 3°. Here, we use the Ineichen model to calculate clear-sky GHI [50].

Although the clear-sky index can reflect the cloudiness level, it is usually insufficient to accurately characterize the weather conditions. For example, a cloudy sky and an overcast sky might have similar clear-sky indices, whereas solar irradiance is more uncertain in a cloudy day. The other indicator of uncertainty levels in probabilistic forecasts is the size of predictive intervals, because larger predictive intervals usually indicate greater uncertainty levels. Since solar irradiance is a function of time-of-day, its predictive interval is also affected by time. To remove the time dependency of solar irradiance, we use the predictive intervals of clear-sky indices to measure the forecast uncertainties. Specifically, we use the difference between the 25th and 75th percentiles as the width of the predictive intervals:

$$w_{d,t} = k^{75}_{d,t} - k^{25}_{d,t} \tag{11}$$

Table 1
Summary of numerical classifiers. Note that the variable in a parenthesis indicates that the classifier is a function of that variable, e.g., the k in classifier 1 to 3 indicates that they are functions of the clear-sky index.

c	$\mathbf{x}^c_{d,h}$	c	$\mathbf{x}^c_{d,h}$	c	$\mathbf{x}^c_{d,h}$	c	$\mathbf{x}^c_{d,h}$
1	$\mu_{d,h}(k)$	4	$\mu_{d,h}(w)$	7	$\mu_{d,h}(k^{PV})$	10	$\mu_{d,h}(w^{PV})$
2	$\sigma_{d,h}(k)$	5	$\sigma_{d,h}(w)$	8	$\sigma_{d,h}(k^{PV})$	11	$\sigma_{d,h}(w^{PV})$
3	$v_{d,h}(k)$	6	$v_{d,h}(w)$	9	$v_{d,h}(k^{PV})$	12	$v_{d,h}(w^{PV})$

The clear-sky indices and their width are all instantaneous values – i.e., they only reflect the cloudiness levels at a certain time point – while FRP is given hourly. To measure the overall uncertainty level within an hour, we introduce three operations: mean ($\mu_{d,h}(\cdot)$), standard variation ($\sigma_{d,h}(\cdot)$), and variability ($v_{d,h}(\cdot)$). Taking clear-sky index for example, the hourly mean measure the overall cloudiness level and is given by:

$$\mu_{d,h}(k) = \frac{1}{4}\sum_{i=0}^{3} k^{50}_{d,4h-i} \tag{12}$$

The standard deviation and variability both measure the cloud-induced fluctuations during a time period, where greater fluctuations usually lead to higher forecast errors [50]:

$$\sigma_{d,h}(k) = \sqrt{\frac{1}{3}\sum_{i=0}^{3}\left[k^{50}_{d,4h-i} - \mu_{d,h}(k)\right]^2} \tag{13}$$

$$v_{d,h}(k) = \sqrt{\frac{1}{4}\sum_{i=0}^{3}\left[k^{50}_{d,4h-i} - k^{50}_{d,4h-i-1}\right]^2} \tag{14}$$

By obtaining the hourly mean, standard deviation, and variability of clear-sky index and the size of predictive interval, we have defined 6 numerical classifiers based on the probabilistic solar irradiance forecasts ($c = 1$ to 6), as listed in Table 1.

Another indicator of local cloudiness level is the clear-sky power index, which is the ratio of solar power forecasts over solar power productions under clear-sky conditions. The term reflects cloudiness levels from a solar power production perspective and has been used in existing studies [51], and the width of clear-sky power index can also be defined accordingly:

$$k^{PV,\alpha}_{d,t} = \frac{p^{\alpha}_{d,t}}{p^{0}_{d,t}}, w^{PV} = k^{PV,75}_{d,t} - k^{PV,25}_{d,t} \tag{15}$$

In this study, we convert solar irradiance to solar power by using PVLib, an open-source simulation tool for solar PV systems [52]; therefore, we can define another six classifiers based on $k^{PV,\alpha}_{d,t}$ ($c = 7$ to 12), as listed in Table 1.

3.2. Reduction of multi-dimensional classifiers

Because of the complexity of cloud movements, 1-dim classifiers derived from a single site are usually insufficient to correctly characterize the solar power uncertainty over the entire CAISO region, and multi-dimensional classifiers based on multiple sites are more appropriate. Real-world system operators can use high-resolution raster forecasts to prepare multi-dimensional classifiers, however, solar irradiance from locations within close proximity can be subject to strong spatial–temporal correlations, and the increased dimension of kNN classifiers can increase the computational intensity. Here, we use principal components analysis (PCA) to reduce the dimension of multi-dimensional classifiers.

PCA is a widely used unsupervised learning method for deriving a low-dimensional set of features from a large set of variables. PCA defines a linear transformation to map the raw data into a new coordinate system such that the new coordinates are those along which the data presents the greatest variance. In the kNN-based method, given an n-dim numerical classifier, the returned principal components represent a set of linear combinations of the original n dimensions, which are ranked based on the amount of explained variance. Our analysis based on single sites suggests that the first three principal components can explain 70%–80% of the total variance of all twelve 1-dim classifiers defined in Table 1.

We start PCA by normalizing the original classifier. Suppose $\mathbf{y}^c_{d,h}$ represents the normalized $\mathbf{x}^c_{d,h}$, and $\mathbf{Y}^s_{d,h} = (\mathbf{y}^1_{d,h}, \dots, \mathbf{y}^C_{d,h}) \in \mathbb{R}^C$ represents a C-dim vector consisting of C normalized 1-dim classifiers from site s, $\mathbf{Y}_{d,h} = (\mathbf{Y}^1_{d,h}, \dots, \mathbf{Y}^S_{d,h}) \in \mathbb{R}^{CS}$ includes all normalized 1-dim classifiers from S sites. We use history data to derive the coefficient matrix, or the loadings, $\mathbf{T} \in \mathbb{R}^{CS\times CS}$, and obtain the principal components $\mathbf{Z}_{d,h} = \mathbf{Y}_{d,h}\cdot\mathbf{T}$, where the ith component is $\mathbf{z}^{(i)}_{d,h}$. In our analysis, We use the first three components as the reduced multi-site classifier, which are then used to identify similar days and construct predictive distributions of FRP.

3.3. FRP requirement evaluation metrics

Given that a shortage of ramping capacities can affect system reliability, and because an oversupply could incur unnecessary production costs, we evaluate the FRP requirements from two perspectives: system reliability and market efficiency. Although typical metrics for system reliability include the loss-of-load probability and expected energy not supplied [36], these metrics might be inappropriate for evaluating the FRP requirements because a lack of ramping capacities does not necessarily cause the loss-of-load event [53]; hence, we follow [6] to use the probability of FRP shortage to assess the adequacy of the FRP requirements. In addition, market efficiency is usually measured in terms of production costs given by market simulations; however, without knowing detailed techno-economic characteristics of individual generators and bus-level demand in the real world, it is out of our scope. For simplicity, we use the amount of system-level FRP oversupply as our metric for market efficiency.

In summary, we adopt two metrics [29]: the frequency of FRP shortage and the amount of FRP oversupply. Based on the definition of FRP, the actual upward and downward requirements equal the upward and downward net load forecast errors in each 15-min interval, i.e., $\epsilon^U_{d,4h-i}$ and $\epsilon^D_{d,4h-i}$, $i = 0, \dots, 3$; therefore, the differences between the estimated FRP requirements and the actual needs are:

$$\begin{aligned}\delta^U_{d,4h-i} &= \widehat{FRP}^U_{d,h} - \epsilon^U_{d,4h-i}\\ \delta^D_{d,4h-i} &= \widehat{FRP}^D_{d,h} - \epsilon^D_{d,4h-i}\end{aligned} \tag{16}$$

The frequencies of reserve shortages during hour h in day d are defined as:

$$\begin{aligned}p^U_{d,h} &= \tfrac{1}{4}\textstyle\sum_{i=0}^{3}\mathbf{1}(\delta^U_{d,4h-i} < 0)\\ p^D_{d,h} &= \tfrac{1}{4}\textstyle\sum_{i=0}^{3}\mathbf{1}(\delta^D_{d,4h-i} > 0)\end{aligned} \tag{17}$$

Note that although the frequency of the FRP shortage is calculated at a 15-min resolution, the FRP requirements are defined on an hourly basis [3]; therefore, the actual hourly FRP needs are given by the following equations:

$$\begin{aligned}FRP^U_{d,h} &= \max_{i=0,\dots,3}(\epsilon^U_{d,4h-i})\\ FRP^D_{d,h} &= \min_{i=0,\dots,3}(\epsilon^D_{d,4h-i})\end{aligned} \tag{18}$$

The oversupplies of up and down FRP during hour d, h are:

$$\begin{aligned}q^U_{d,h} &= \left|\max(0, \widehat{FRP}^U_{d,h} - FRP^U_{d,h})\right|\\ q^D_{d,h} &= \left|\min(0, \widehat{FRP}^D_{d,h} - FRP^D_{d,h})\right|\end{aligned} \tag{19}$$

To measure the performance during an extended period of time, we define the following overall evaluation metrics during a given set of

hours $\mathcal{M} = \{(d, h)\}$:

$$\begin{aligned} P^U_{\mathcal{M}} &= \mathbb{E}_{d,h\in\mathcal{M}}\left[p^U_{d,h}\right] \\ P^D_{\mathcal{M}} &= \mathbb{E}_{d,h\in\mathcal{M}}\left[p^D_{d,h}\right] \\ Q^U_{\mathcal{M}} &= \sum_{d,h\in\mathcal{M}} q^U_{d,h} \\ Q^D_{\mathcal{M}} &= \sum_{d,h\in\mathcal{M}} q^D_{d,h} \end{aligned} \tag{20}$$

where P and Q denote the frequencies of shortages and total oversupplies over period $\mathcal{M}$, respectively.

3.4. Dynamic selection of kNN parameters

A key step in the kNN method is to select an appropriate classifier (c) and an appropriate number of closest neighbors (K). Ideally, in an hour (d_0, h_0), the optimal c and K should be selected such that both the frequencies of reserve shortage ($p^U_{d_0,h_0}$ or $p^D_{d_0,h_0}$) and reserve oversupplies ($q^U_{d_0,h_0}$ or $q^D_{d_0,h_0}$) are minimized. This forms the following multi-objective optimization (taking upward FRP as an example):

$$\min_{c,K} (p^{U,c,K}_{d_0,h_0}, q^{U,c,K}_{d_0,h_0}) \text{ s.t. } c \in C, K \in \mathbb{Z}^+ \tag{21}$$

Because the FRP requirements must be made available to market participants before (d_0, h_0), and because the forecast errors can only be known afterwards, we use the metrics calculated using N previous days as an empirical approximation. The optimization becomes:

$$\min_{c,K} (P^{U,c,K}_{\mathcal{M}}, Q^{U,c,K}_{\mathcal{M}}) \text{ s.t. } c \in C, K \in \mathbb{Z}^+ \tag{22}$$

where $\mathcal{M}$ represents a validation set that includes hour h from N previous days: $\mathcal{M} = \{(d_0 - i, h_0)|i = 1, \ldots, N\}$. Because of the limited search space in our study, it is possible to solve the multi-objective optimization by an exhaustive search over all pairs of c and K and by selecting the optimal solution which minimizes the objective of both $P^{U,c,K}_{\mathcal{M}}$ and $Q^{U,c,K}_{\mathcal{M}}$. Note that each hour could have a different set of optimal c and K, which are dynamically selected by the optimization model; therefore, we propose a two-step kNN-based method to estimate the FRP requirements of hour (d_0, h_0), as summarized in Fig. 3. In Step 1, the kNN parameters (c^*, K^*) are optimized based on historical performance from a validation set consisting of N previous days. In Step 2, the optimal classifier $\mathbf{x}^{c*}_{d_0,h_0}$ is calculated based on solar forecasts, and the FRP requirements are estimated based on the K most similar historical days. Note that the method is applied in a rolling-forward manner, where the validation set is updated everyday by adding the latest day and dropping the earliest day. The dynamic construction of a validation set is aimed to improve the performance of the kNN-based method, as shown in Fig. 8, the dynamic case outperforms most other cases with constant kNN parameters. Furthermore, a dynamic validation simulates how operators such as the CAISO commonly use similar methods in practice (as detailed in Section 2.1), re-estimating parameters and requirements daily, based on updated data sets (dropping the oldest day and adding the newest day to the dataset).

3.5. Machine/deep-learning-based benchmarks

A collection of machine/deep-learning-based (ML/DL-based) benchmarks are set to compare with the proposed kNN-based approach. Different from the proposed approach, these ML/DL-based benchmarks are modeled as supervised learning regression problems, which do not follow the CAISO FRP implementation procedure. The general idea of the FRP prediction is to fit a mapping between the deployed FRP and the input at the training stage:

$$\begin{aligned} \widehat{FRP}^U &= \hat{\Phi}^U(\mathbf{x}, \mathbf{W}^U) \\ \widehat{FRP}^D &= \hat{\Phi}^D(\mathbf{x}, \mathbf{W}^D) \end{aligned} \tag{23}$$

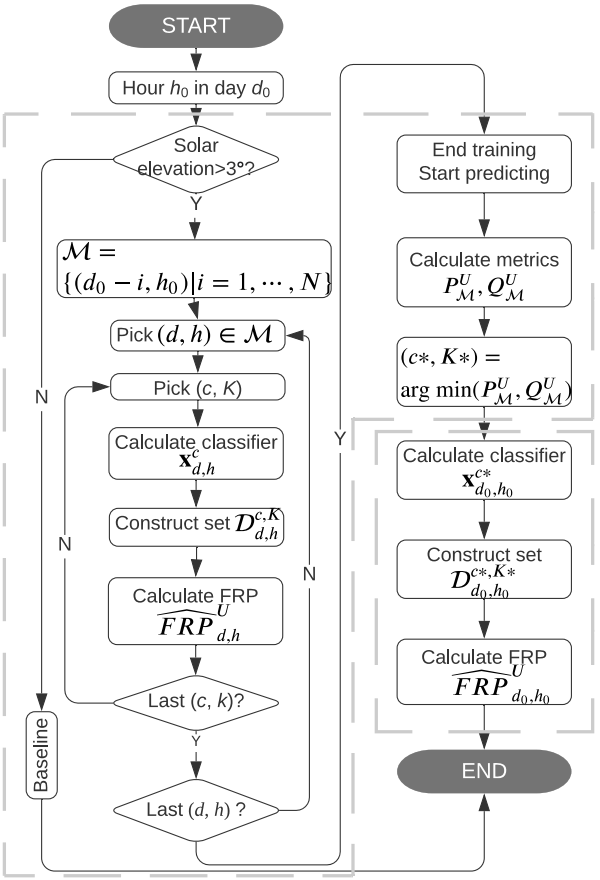


Fig. 3. Estimate FRP by dynamically selecting kNN parameters. We use upward FRP to demonstrate the workflow and downward FRP follows similar process. Note that the $\mathbf{x}^c_{d,h}$ can be any 1-dim classifier from Table 1 or PCA classifiers.

where $\mathbf{x}$ is an input variable, $\mathbf{W}^U$ and $\mathbf{W}^U$ are trainable parameters of ML/DL-based models, and $\hat{\Phi}^U$ and $\hat{\Phi}^D$ are the mapping relationships between the input and the upward and downward FRP, respectively. Note that the input $\mathbf{x}$ can be any classifier used in the kNN-based method. Table 2 lists the selected ML/DL-based benchmarks for comparison, including one convolutional neural network (CNN), three shallow neural networks (ANNs), three support vector regression (SVR) models, three gradient boosting machine (GBM) models, and a random forest (RF) model. Note that no recurrent neural networks, e.g., long short-term memory (LSTM) and gated recurrent unit (GRU), are included in the benchmark because of the discontinuous time series used in this research, and because our preliminary analysis based on LSTM and GRU fails to show improvements over other ML/DL-based models. The hyperparameters/parameters are tuned by cross-validation by splitting validation subsets from the training dataset.

CNNs were first developed for image classifications, but later were introduced to deep regression problems. CNNs have been widely used in the power and energy domain, such as forecasting [54], infrastructure detection [55], and individual behavior understanding [56]. The 1-dim CNN model consists of three layers: one convolutional layer followed by two densely-connected layers. This architecture is determined by the cross-validation optimization. An example of the CNN architecture optimization is shown in Fig. 4d. There are 32 filters with a kernel size of 5 in the convolutional layer. The two densely-connected layers contain 10 and 1 neurons. Adaptive moment estimation (Adam) is used for training, and the learning rate is 0.01. The sigmoid function is used in all layers.

Four shallow machine learning algorithms with different training strategies, kernels, or distributions are also included as benchmarks.

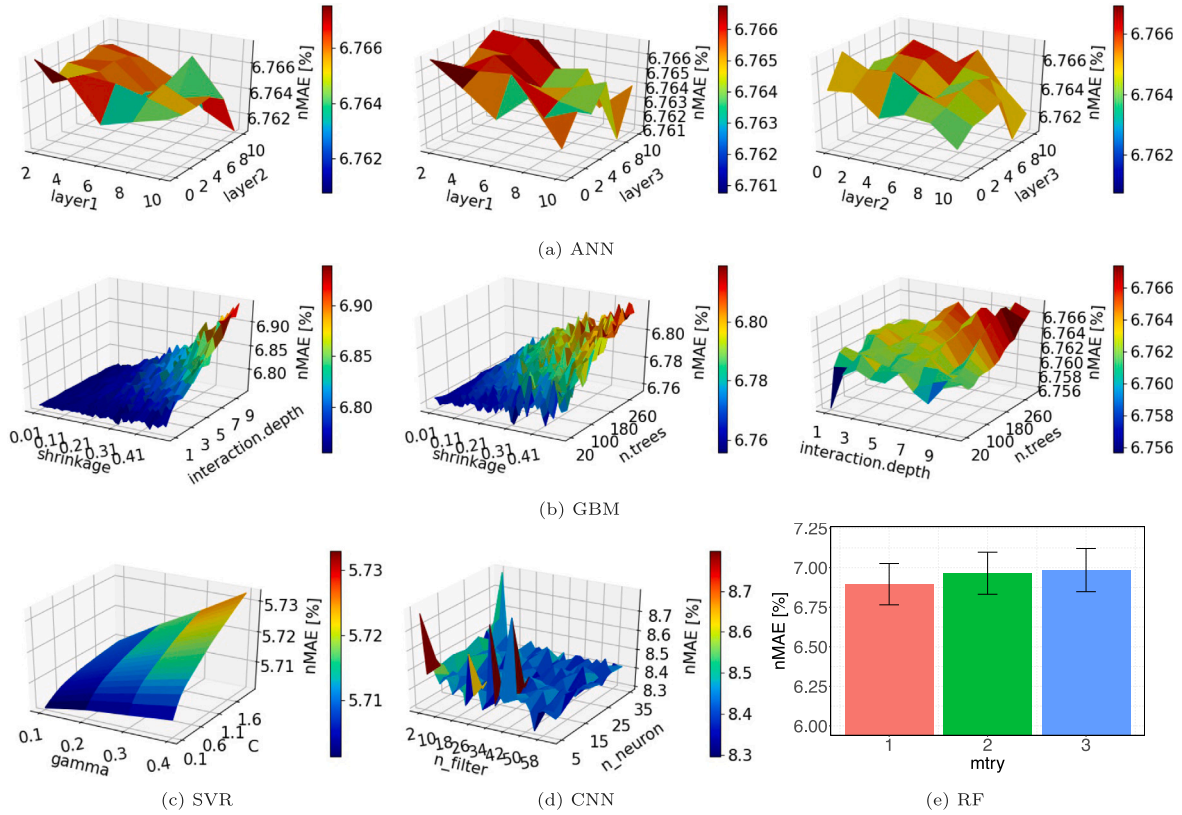


Fig. 4. Parameter/hyperparameter optimization examples. (a) An ANN optimization example using B_2, where the number neurons in each layer is optimized. (b) A GBM optimization example using B_8. (c) A SVR optimization example using B_7. (d) A CNN optimization example using B_1. (e) An RF optimization example using B_{11}. Note that nMAE denotes normalized mean absolute error.

Table 2
Deep/machine learning benchmarks.

Algorithm	Model	Training algorithm and kernel function
CNN	B_1	Adaptive moment estimation
ANN	B_2	Standard back-propagation
	B_3	Momentum-enhanced back-propagation
	B_4	Resilient back-propagation
SVR	B_5	Linear kernel
	B_6	Polynomial kernel
	B_7	Radial basis function kernel
GBM	B_8	Squared loss
	B_9	Laplace loss
	B_{10}	T-distribution loss
RF	B_{11}	Classification and regression tree (CART) aggregation

These models are also widely used in various regression problems in power systems [57–59]. ANN models are trained with a learning rate of 0.01. The momentum in B_2 is set to be 0.9. The minimum and maximum update rate in B_3 are 1×10^{-6} and 50, respectively. Similar to CNN, the hyperparameters of ANNs are also determined by the grid-search optimization, which is illustrated by an example in Fig. 4a. The number of layers and the number of neurons in each layer are determined in the optimization process based on the validation dataset. The kernel coefficient (gamma) and the regularization parameter (C) in SVR models are the tuning parameters and their tuning process is shown in Fig. 4c. In GBM models, the number of trees (n.trees), the number of splits (interaction.depth), and the learning rate (shrinkage) are the most important parameters. Their selection is shown in Fig. 4b. The number of variables randomly sampled as candidates at each split (mtry) in the RF model is tuned, as shown in Fig. 4e.

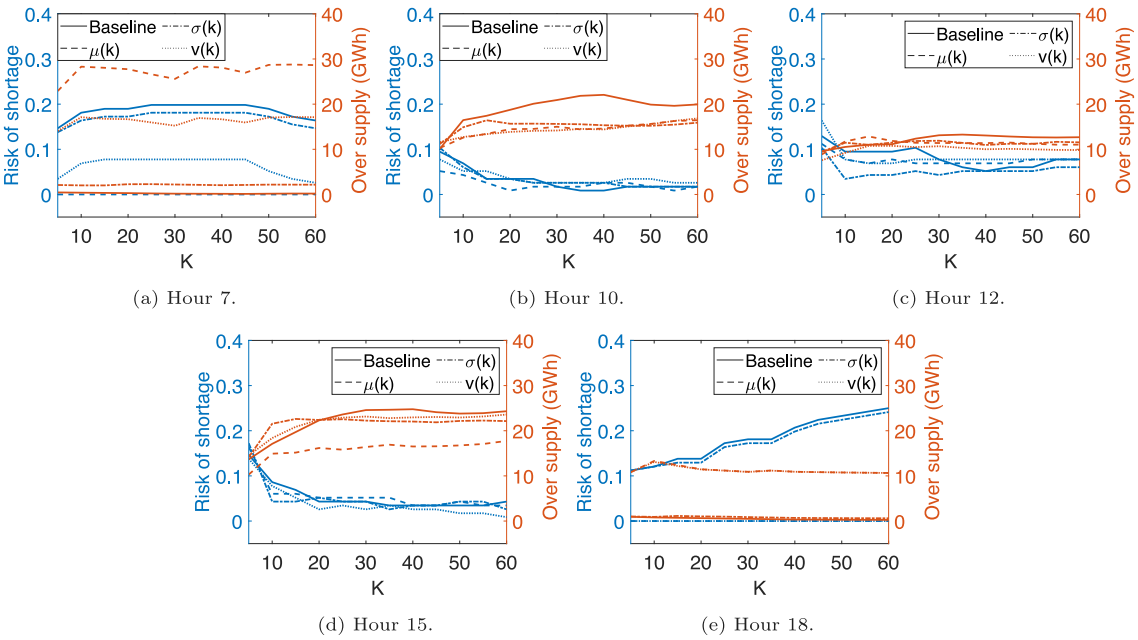
4. Case studies and results

4.1. Test cases

The data used in our case studies are drawn from two sources: We download the historical system-level net load forecast errors from CAISO's Open Access Same-time Information System, which publishes real-time and historical data related to CAISO's daily market operations; IBM's Watt-Sun forecasting system provides 2HA probabilistic solar irradiance forecasts for a selected number of sites [60]. Note that IBM's sites are located at places where ground observations are available. Given that CAISO's solar PV power plants are mainly concentrated in the Southern California region, we use the forecasts from five sites ($s = 1$ to 5) in Southern California to characterize the weather conditions. The forecasts are issued every 15 min and include the 5th, 25th, 50th, 75th, and 95th percentiles of mean GHI during each 15-minute interval. The forecasts have NaN values during nighttime, therefore are removed. The forecasts and historical CAISO data span from April 1, 2019 to December 31, 2020. We select three months from 2020 as our test sets: February, August, and October, which represent winter, summer, and transitional season, respectively. Data from the remaining 18 months is used to train and validate our models.

4.2. Trade-offs between reliability and oversupply

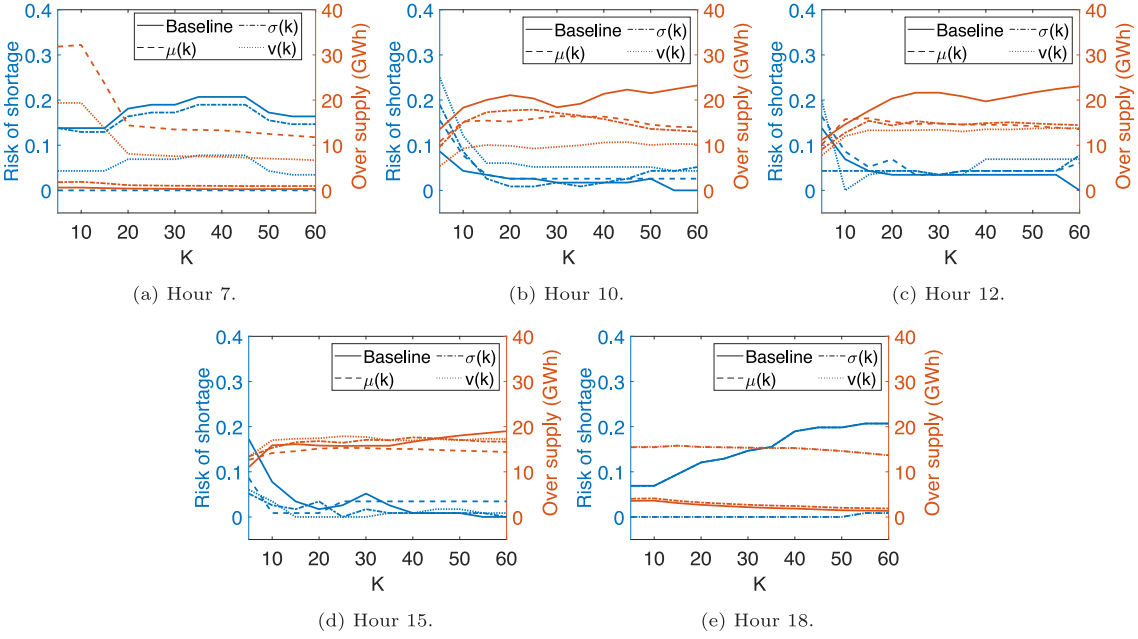
We first study the performance in terms of the frequency of FRP shortage and oversupply as a function of classifiers (c) and the number of nearest neighbors (K). Figs. 5 and 6 show the frequencies of FRP shortage ($P_{\mathcal{M}}^{U,c,K}$, $P_{\mathcal{M}}^{D,c,K}$) and total oversupplies ($Q_{\mathcal{M}}^{U,c,K}$, $Q_{\mathcal{M}}^{D,c,K}$) in both directions during 5 representative hours in February—i.e., h is fixed, and d includes all days in February in $\mathcal{M}$. The 5 selected hours represent five different times of day: early morning ($h = 7$), midmorning



(a) Hour 7. (b) Hour 10. (c) Hour 12.

(d) Hour 15. (e) Hour 18.

Fig. 5. Upward FRP oversupply and risk of shortage as a function of k in February 2020 by time of day. Site 2 is used. Note that line types in the legend only represent classifiers, and line colors indicate whether a line represents "risk of shortage" (left axis) or "oversupply" (right axis).



(a) Hour 7. (b) Hour 10. (c) Hour 12.

(d) Hour 15. (e) Hour 18.

Fig. 6. Downward FRP oversupply and risk of shortage as a function of k in February 2020 by time of day. Site 2 is used. Note that line types in the legend only represent classifiers, and line colors indicate whether a line represents "risk of shortage" (left axis) or "oversupply" (right axis).

($h = 10$), noon ($h = 12$), midafternoon ($h = 15$), and late afternoon ($h = 18$). For brevity, we show only the results for the baseline and three 1-dim classifiers ($c = 1, 2, 3$) because the other 1-dim classifiers ($c = 4$ to 12) present similar trends. In many cases, the frequency of the FRP shortage in both directions is reduced as K increases for hours between mid-morning and mid-afternoon; at those times, the total oversupply slightly increases as K increases. For example, the total oversupply in the baseline during hour 10 increases from 10 GWh to 20 GWh in the upward direction, and from 13 GWh to 23 GWh in the downward direction, as K increases from 5 to 60. Meanwhile, the frequency of the FRP shortage decreases from 10% to less than 5% in both directions because of greater coverage. Similar trends can also be found in the three kNN-based cases by using 1-dim classifiers.

A comparison between the baseline and the kNN cases indicates that the kNN-based method can reduce the total oversupply without significantly compromising the reliability level. For example, by using $\mu(k)$ as a classifier ($c = 1$) and $K = 30$, the total oversupply of the upward FRP at hour 10 is brought down to 15 GWh compared to 20 GWh in the baseline, whereas the frequency of the FRP shortage remains the same. In addition, the baseline presents greater risks of FRP shortage in both directions during early morning and late afternoon: the frequencies are greater than 10% at hour 7 and 18. In contrast, the frequencies of the upward FRP shortage are 10% and 7% when using $v(k)$ and $\mu(k)$ as a classifier, respectively. This indicates that the kNN-based method can significantly improve the reliability level.

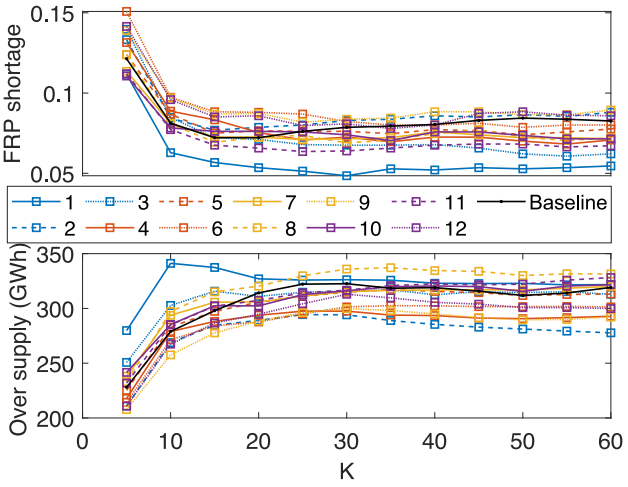


Fig. 7. The frequency of the FRP shortage and the FRP oversupply as a function of K in February 2020. One-dim classifiers ($c = 1$ to 12) from site 2 are used.

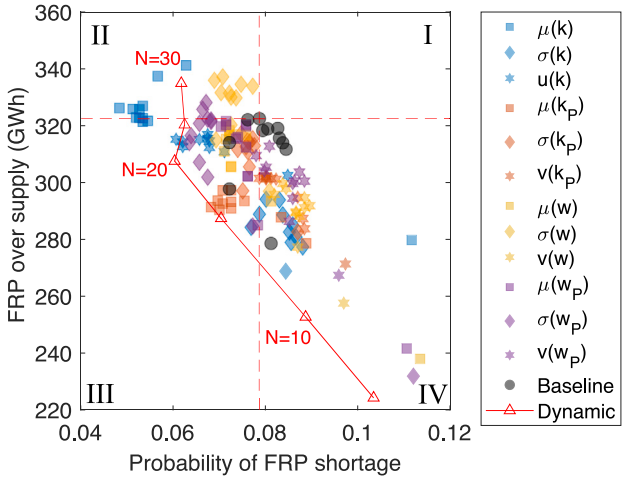


Fig. 8. Trade-offs between reliability and oversupply in February 2020. One-dim classifiers from Site 2 are used.

Next, we examine the overall performance in terms of reliability and oversupply. As Fig. 7 shows, all 1-dim classifiers present similar trends as K increases, i.e., the FRP oversupply increases sharply when $K \leq 20$, which in turn results in a considerable drop of the frequency of the FRP shortage. When $K \geq 20$, however, both metrics remain relatively constant. For example, when $\mu(k)$ is used as a classifier, the total FRP oversupply increases to 320 GWh and the frequency of FRP shortage decreases to 6% as K increases to 20, and both stay constant thereafter. This phenomenon suggests that a sufficient number of days ($K \geq 20$) are required to give a reliable estimation of FRP requirements across all classifiers. Note that in the baseline, the frequency of the FRP shortage increases slightly as K increases from 20 to 60, implying that 20 days could yield better performance than CAISO's business-as-usual (BAU) implementation, where $K = 30$. In addition, similar to the previous observation in Figs. 5 and 6, Fig. 7 also illustrates the trade-off between reliability and oversupply: using $\mu(k)$ as a classifier leads to high FRP oversupplies across the whole K range; however, the conservative FRP requirements result in the lowest frequencies of FRP shortage.

Fig. 8 shows the trade-off between reliability and oversupply in the form of Pareto frontiers. The point at the intersection of two dashed lines represents CAISO's BAU implementation. The two dashed lines divide the plane into four quadrants (I, II, III, and IV), where points in Quadrant III indicate an improvement in both dimensions, points in Quadrant I indicate a degradation in both metrics, and Quadrants II and IV represent noninferior solutions, i.e., an improvement in one objective requires a degradation of the other. In Fig. 8, a significant percentage of kNN points fall into Quadrant III and no point falls within Quadrant I, suggesting that the kNN-based method can result in more economic and reliable solutions than the baseline.

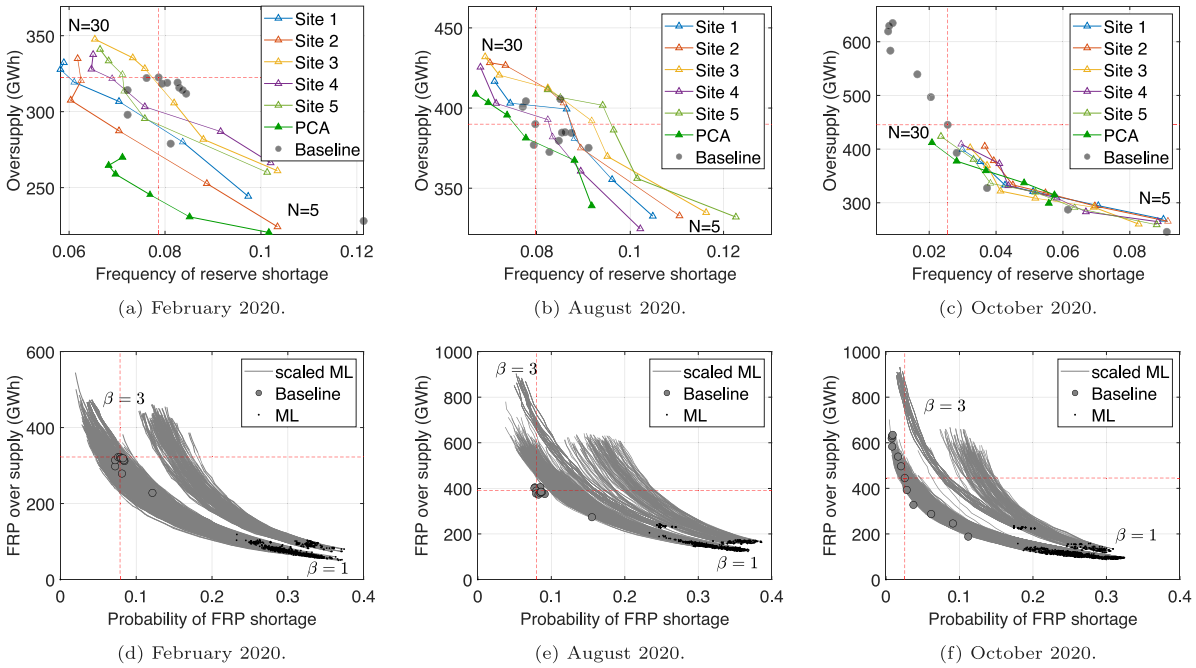
4.3. Dynamic selection of kNN parameters

Until now, we have demonstrated that the kNN-based method can potentially improve the estimates of FRP requirements; however, it could also lead to the deterioration of reliability or oversupply when an inappropriate classifier or number of days is selected. In addition, the performance of classifiers can vary significantly with the time of day. As shown in Figs. 5a, 5e, 6a, and 6e, $\mu(k)$ yields the highest reliability levels during early morning and late afternoon in both directions; whereas Figs. 5c and 6c suggest that $\sigma(k)$ outperforms $\mu(k)$ during noon in terms of reliability. The inconsistent performance presented by Figs. 5 and 6 implies the challenges associated with the selection of an appropriate classifier, which motivates the dynamic selection of the optimal kNN parameters.

Fig. 8 shows the results when the kNN parameters are dynamically selected from all 1-dim classifiers. We observe a similar trade-off between reliability and oversupply when the size of the validation set (N) changes. In addition, Fig. 8 shows that the results are significantly improved when the kNN parameters are dynamically selected. For example, when $N = 20$, the frequency of the FRP shortage is reduced to 6%, less than any other runs with similar total oversupplies. Another way to view the improvements resulting from the proposed method is to consider the reduction in the value on one axis in Fig. 8, holding the other axis at its baseline value. For example, holding the oversupply at 322 GWh (CAISO's BAU value), the dynamic selection method brings the probability of FRP shortage down to 6%. Similarly, Table 3 shows that when the risk is held constant at 8%, the oversupply is reduced to 272 GWh from 322 GWh in CAISO's BAU case, a 16% reduction.

Fig. 9 shows the optimized Pareto frontiers of all 1-site cases and the multi-site case using PCA-kNN from February, August, and October 2020. All optimal Pareto frontiers present similar trends as in Fig. 8 when N varies. Although most 1-site cases present better performance than the baseline, their performance varies because of geographic differences. The variation of performance across sites implies the challenge of using one single site to characterize the weather condition of the whole CAISO region. By contrast, the optimized PCA-kNN frontier accounts for the weather conditions from all 5 sites, and results in less variation of performance. The optimized PCA-kNN frontier presents better performance in terms of both dimensions compared to the baseline and all 1-site cases in February and August. As Table 3 suggests, when the risk of FRP shortage is held at 8%, the PCA-kNN case brings down the oversupply in February to 242 GWh, a 25% reduction compared to CAISO's BAU case.

To compare the results from the kNN-based methods with other data-driven methods, the evaluation metrics from the benchmark ML/DL-based models are calculated and shown in Fig. 9d-f. Note that these models do not require the use of historical data to construct predictive distributions of FRP, therefore are not affected by the number of neighbors K. As displayed in Fig. 9d-f, the black dots that represent the original benchmark methods are all concentrated in Quadrant IV, indicating reduced oversupplies yet lower reliability levels. For example, the minimum FRP oversupply from February given by the PCA-kNN case is 220 GWh (the points at the bottom as shown in Fig. 9a), which is still significantly greater than the maximum oversupply from the ML/DL benchmarks, 106 GWh; however, the smallest frequency of FRP shortage in the benchmarks is 22.8%, which is greater than not only all kNN-based methods but also the baseline.



(a) February 2020. (b) August 2020. (c) October 2020.

(d) February 2020. (e) August 2020. (f) October 2020.

Fig. 9. (a)–(c) Pareto frontiers of all 1-site cases and the multi-site PCA-kNN case. (d)–(f) Frontiers from the ML/DL-based methods. The black dots represent the original results from ML/DL-based benchmarks – i.e., $\beta = 1$ – and the gray curves represent the Pareto frontiers when the factor β ranges from 1 to 3.

The significantly greater chances of FRP shortage of the benchmark methods result in an unfair comparison, therefore, to make a comparison based on a more level playing field, we adjust the FRP requirements from the benchmark methods by multiplying them with a factor β, where $\beta = 1$ indicates the original results. Owing to the trade-off between the oversupply and the chance of FRP shortage, a greater β results in more conservative FRP requirements, i.e., more FRP oversupplies and lower probabilities of FRP shortage. As Fig. 9d shows, the profiles of the adjusted benchmark results resemble the previous kNN profiles. Although several curves fall into Quadrant I, a considerable fraction of the adjusted benchmark results fall into Quadrant III, indicating potential improvement in both dimensions. Particularly, Table 3 indicates that when the reliability level is held constant at CAISO's BAU level, the oversupplies from the best adjusted benchmark results are even lower than the optimized PCA-kNN method in February and October.

We also evaluate our results by calculating mean absolute error (MAE), a commonly used evaluation metric as defined below:

$$\text{MAE} = \frac{1}{|\mathcal{M}|} \cdot \sum_{(d,h)\in\mathcal{M}} \left| \widehat{FRP}^{D}_{d,h} - FRP^{D}_{d,h} \right| \tag{24}$$

where set $\mathcal{M}$ represents an evaluation period and is one month in our study. Note that (24) only gives the definition of MAE of downward FRP and the MAE of upward FRP is in similar form. Table 4 shows MAE of all cases. It shows that although single-site kNN at times results in greater MAE values than the baseline, the adoption of PCA appreciably reduces the MAE values and result in better performance. In addition, although the ML/DL method results in the lowest MAE values, as shown in Fig. 9, it also leads to greater chances of FRP shortage. The fact that MAE fails to take into account system reliability is a limitation, and highlights the importance of considering both oversupply and reliability in the decision-making process.

Table 3
Comparison of oversupplies when the probabilities of FRP shortage are fixed at CAISO's BAU levels.

	February	August	October
Probability of FRP shortage, CAISO BAU	0.078	0.080	0.026
Oversupply, CAISO BAU	322	390	445
Oversupply, PCA kNN	242	379	392
Oversupply, best adjusted ML/DL	226	445	387
Oversupply, best single-site	272	395	415
Number of the best single-site	2	4	5

Note: Oversupplies are in GWh.

Table 4
MAE of the studied cases. Note that the baseline and single-site MAE values are mean over all K values, PCA-kNN values are mean over all N values, and ML/DL is the mean over all included methods.

	Baseline (MW)	Single-site kNN (MW)	PCA-kNN (MW)	ML/DL (MW)
FRP, downward, February 2020	358	364	296	223
FRP, upward, February 2020	393	433	315	225
FRP, downward, August 2020	247	342	278	165
FRP, upward, August 2020	453	443	405	346
FRP, downward, October 2020	389	420	291	199
FRP, upward, October 2020	598	763	460	219

5. Conclusion

This paper proposed a data-driven method to improve the estimation of the FRP requirements by using probabilistic solar forecasts. We developed a variety of numerical classifiers as labels of weather conditions based on the predicted uncertainty, variability, and cloudiness index of solar irradiance and power. In addition, we adopted PCA to reduce the dimension of multi-dimensional classifiers, which better reflects system-level uncertainties by including classifiers from multiple sites across CAISO. We applied a kNN-based method to identify historical days of similar weather condition and used the realized FRP requirements from these days to construct predictive distributions of the FRP needs, which can be used to give weather-informed estimations of FRP requirements.

By applying the proposed method to the CAISO 15-min RT market, our results from three representative months in 2020 suggest that the proposed method can improve the performance compared to the baseline in terms of both system reliability and oversupplies. Because of the performance sensitivity to the kNN parameters, the paper also

proposed a selection method to dynamically identify the best kNN parameters per hour of the day, which allows power system operators to maximize the benefits of the kNN-based method. Key insights from our results include:

1. The performance of the kNN-based method varies across classifiers and the kNN parameters. Generally, a greater K results in more conservative estimations, i.e., greater FRP requirements and increased reliability levels.
2. Despite variations in performance, the weather-informed decisions perform better in terms of reliability and market efficiency most of the time, especially during early morning and late afternoon, when sunrise and sunset cause greater net load uncertainties. As Fig. 9a shows, by using multi-site classifiers, the PCA-kNN method can reduce FRP oversupply by up to 25% without increasing the risk of reserve shortage.
3. Although different classifiers from a single site explain weather uncertainty from different perspectives, they also present considerable correlations. In addition, a comparison of results using forecasts from different sites suggests that geographical variations can affect performance. In response, the PCA-based method effectively characterizes the system-level weather uncertainty by using only a reduced set of principal components, and results in improved performance when compared with all single-site cases. In our case, using 3 principal components can explain 70%–80% of total variance, and as shown in Fig. 9a, using PCA classifiers results in an additional 11% of reduction in total FRP oversupply compared with the best single-site classifier. The proposed approach is extensible to include more sites (e.g., the entire CAISO region) to further improve the performance, therefore promises greater benefits in real world.
4. Different from the kNN-based method that is bounded by similar days, the ML/DL-based methods learn to estimate reserves with greater variations. Moreover, as shown in Fig. 9, the ML/DL-based methods tend to underestimate FRP requirements, which further result in increased frequencies of FRP shortages. This is because FRP requirements are subject to a lower bound of 0 MW but no upper bound, and because the ML/DL-based methods put the same weights on FRP shortages and oversupplies, underestimating FRP requirements result in smaller prediction errors. This implies that a better objective function is required in the ML/DL-based methods to take into account trade-offs between reliability and economics in real world. In addition, it is interesting to train different ML/DL models for different weather classifications with more years of data.
5. Although the ML/DL-based benchmarks can potentially reduce the oversupply of FRP, our methods present better performance in terms of reliability levels. In addition, Table 3 indicates that our optimized PCA-kNN results outperform the best adjusted benchmark results in August, indicating superior and robust performance. In the other two months, the best adjusted benchmark results show lower FRP oversupplies, however, Fig. 9 shows that a significant percentage of adjusted benchmarks also result in greater FRP oversupplies that CAISO's BAU results, indicating great variation of performance. By contrast, our optimized PCA-kNN results present better performance than CAISO's BAU across all months.

The kNN-based method represents a generic approach to providing weather-informed ramping requirements in power market scheduling and operations, which can also be extended to determine the requirements of other forms of ancillary services in electricity markets. As mentioned, the method is also extensible to include other components of net load probabilistic forecasts. One limitation of our analysis is that we assume perfect foresight in load and wind power to focus on solar power, therefore, future studies should include additional uncertainties associated with load and wind power to better reflect real-world rules.

In future work, we will also explore quantifying the economic and reliability benefits using look-ahead market clearing runs, i.e., using UC and ED models. Currently, the selection of kNN parameters is formulated as a multi-objective optimization model that considers reliability and economic separately, as shown in (21). In a multi-objective problem like this, a single solution can be chosen if the user can select weights reflecting the relative priority of the two objectives. Such weights would be user and situation dependent, and would best be selected after studying the trade-offs embodied in the set of efficient (Pareto) alternatives. In addition, a critical caveat in our analysis is that the solar power output is converted from solar irradiance by using simulation tools, which may not reflect the actual irradiance-to-power conversion because of imperfect knowledge of technical parameters of real-world solar power plants, such as inverter capacities, PV panel capacities, and degradation rates. Future work can calibrate the simulated result by comparing it with real-world data or conducting sensitivity analyses to address parametric uncertainties. Comparative studies can also be conducted to demonstrate the impact on the model performance by using simulated PV power as classifiers.

CRediT authorship contribution statement

Binghui Li: Conceptualization, Software, Formal analysis, Investigation, Methodology, Validation, Writing – original draft. **Cong Feng:** Software, Formal analysis, Investigation, Methodology, Validation, Writing – original draft. **Carlo Siebenschuh:** Software. **Rui Zhang:** Software. **Evangelia Spyrou:** Formal analysis, Writing – review & editing. **Venkat Krishnan:** Writing – review & editing. **Benjamin F. Hobbs:** Methodology, Project administration, Supervision, Writing – review & editing, Funding acquisition. **Jie Zhang:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number DE-EE0008215.

Disclaimer

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

References

[1] Navid N, Rosenwald G. Ramp capability product design for MISO markets. 2013, Accessed: 2020-01-27.
[2] Parker N. Ramp product design. 2015, Accessed: 2020-01-27.
[3] California Independent System Operator. Business practice manual for market operations. 2020, Accessed: 2020-01-27.
[4] Villar J, Bessa R, Matos M. Flexibility products and markets: Literature review. Electr Power Syst Res 2018;154:329–40.
[5] Wang Q, Hodge B-M. Enhancing power system operational flexibility with flexible ramping products: A review. IEEE Trans Ind Inf 2017;13(4):1652–64.
[6] Thatte AA, Xie L. A metric and market construct of inter-temporal flexibility in time-coupled economic dispatch. IEEE Trans Power Syst 2016;31(5):3437–46.
[7] Dranka GG, Ferreira P, Vaz AIF. A review of co-optimization approaches for operational and planning problems in the energy sector. Appl Energy 2021;304:117703.
[8] Cui M, Zhang J. Estimating ramping requirements with solar-friendly flexible ramping product in multi-timescale power system operations. Appl Energy 2018;225:27–41.
[9] Wang B, Hobbs BF. Real-time markets for flexiramp: A stochastic unit commitment-based analysis. IEEE Trans Power Syst 2016;31(2):846–60.
[10] Yamujala S, Kushwaha P, Jain A, Bhakar R, Wu J, Mathur J. A stochastic multi-interval scheduling framework to quantify operational flexibility in low carbon power systems. Appl Energy 2021;304:117763.
[11] Zhang X, Che L, Shahidehpour M, Alabdulwahab A, Abusorrah A. Electricity-natural gas operation planning with hourly demand response for deployment of flexible ramp. IEEE Trans Sustain Energy 2016;7(3):996–1004.
[12] Ye H, Li Z. Deliverable robust ramping products in real-time markets. IEEE Trans Power Syst 2018;33(1):5–18.
[13] Huang C, Ma H, Yan Z, Chen S, Li M. Portfolio management for a wind-storage system based on distributionally robust optimisation considering a flexible ramping product. IET Renew Power Gener 2020;14(16):3192–9.
[14] Wang Z, Shen C, Liu F, Wang J, Wu X. An adjustable chance-constrained approach for flexible ramping capacity allocation. IEEE Trans Sustain Energy 2018;9(4):1798–811.
[15] Fang X, Sedzro KS, Yuan H, Ye H, Hodge B-M. Deliverable flexible ramping products considering spatiotemporal correlation of wind generation and demand uncertainties. IEEE Trans Power Syst 2019;35(4):2561–74.
[16] Nosair H, Bouffard F. Energy-centric flexibility management in power systems. IEEE Trans Power Syst 2016;31(6):5071–81.
[17] Sreekumar S, Sharma KC, Bhakar R. Gumbel copula based multi interval ramp product for power system flexibility enhancement. Int J Electr Power Energy Syst 2019;112:417–27.
[18] Nadermahmoudi E, Amraee T, Oskouee SS. Stochastic very short-term economic dispatch for wind power operation using flexible ramp reserve. Int Trans Electr Energy Syst 2020;30(8):e12454.
[19] Ghaemi S, Salehi J, Moeini-Aghtaie M. Developing a market-oriented approach for supplying flexibility ramping products in a multimicrogrid distribution system. IEEE Trans Ind Inf 2021;17(10):6765–75.
[20] Mohandes B, Moursi MSE, Hatziargyriou N, Khatib SE. A review of power system flexibility with high penetration of renewables. IEEE Trans Power Syst 2019;34(4):3140–55.
[21] Wu C, Hug G, Kar S. Risk-limiting economic dispatch for electricity markets with flexible ramping products. IEEE Trans Power Syst 2016;31(3):1990–2003.
[22] Wu H, Shahidehpour M, Li Z, Tian W. Chance-constrained day-ahead scheduling in stochastic power system operation. IEEE Trans Power Syst 2014;29(4):1583–91.
[23] Zhao P, Wu H, Gu C, Hernando-Gil I. Optimal home energy management under hybrid photovoltaic-storage uncertainty: A distributionally robust chance-constrained approach. IET Renew Power Gener 2019;13(11):1911–9.
[24] Zhao S, Fang Y, Wei Z. Stochastic optimal dispatch of integrating concentrating solar power plants with wind farms. Int J Electr Power Energy Syst 2019;109:575–83.
[25] Bavafa F, Niknam T, Azizipanah-Abarghooee R, Terzija V. A new biobjective probabilistic risk-based wind-thermal unit commitment using heuristic techniques. IEEE Trans Ind Inf 2017;13(1):115–24.
[26] Papavasiliou A, Oren SS, O'Neill RP. Reserve requirements for wind power integration: A scenario-based stochastic programming framework. IEEE Trans Power Syst 2011;26(4):2197–206.
[27] Rebours Y, Kirschen D. A survey of definitions and specifications of reserve services. Report, University of Manchester; 2005, p. 1–38.
[28] Ela E, Hytowitz R, Helman U. Ancillary services in the United States: Technical requirements, market designs, and price trends. Tech. rep., Palo Alto, CA: Electric Power Research Institute; 2019.
[29] De Vos K, Stevens N, Devolder O, Papavasiliou A, Hebb B, Matthys-Donnadieu J. Dynamic dimensioning approach for operating reserves: Proof of concept in Belgium. Energy Policy 2019;124:272–85.
[30] Holttinen H, Milligan M, Ela E, Menemenlis N, Dobschinski J, Rawn B, Bessa RJ, Flynn D, Gomez-Lazaro E, Detlefsen NK. Methodologies to determine operating reserves due to increased wind power. IEEE Trans Sustain Energy 2012;3(4):713–23.
[31] Zhou Z, Levin T, Conzelmann G. Survey of US ancillary services markets. Tech. rep., Argonne, IL (United States): Argonne National Lab.(ANL); 2016.
[32] Operator CIS. Flexible ramping product refinements initiative appendix C–Quantile regression approach. 2020, Accessed: 2022-01-25.
[33] De Vos K. Methodology for the dimensioning of the aFRR needs. Tech. rep., Elia Transmission Belgium SA; 2020.
[34] Zhang G, McCalley JD. Estimation of regulation reserve requirement based on control performance standard. IEEE Trans Power Syst 2018;33(2):1173–83.
[35] Li B, Zhang J. A review on the integration of probabilistic solar forecasting in power systems. Sol Energy 2020;210:68–86, Special Issue on Grid Integration.
[36] Matos MA, Bessa RJ. Setting the operating reserve using probabilistic wind power forecasts. IEEE Trans Power Syst 2010;26(2):594–603.
[37] Etingov P, Miller L, Hou Z, Makarov Y, Pennock K, Beaucage P, Loutan C, Motley A. Balancing needs assessment using advanced probabilistic forecasts. In: 2018 IEEE international conference on probabilistic methods applied to power systems (PMAPS). IEEE; 2018, p. 1–6.
[38] Etingov P, Miller L, Hou Z, Makarov Y, Loutan C, Katzenstein W. Improving BA control performance through advanced regulation requirements prediction. In: 2018 IEEE power & energy society general meeting (PESGM). 2018.
[39] Costilla-Enriquez N, Ortega-Vazquez M, Tuohy A, Motley A, Webb R. Operating dynamic reserve dimensioning using probabilistic forecasts. 2021, arXiv preprint arXiv:2108.09362.
[40] Van der Meer DW, Widén J, Munkhammar J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renew Sustain Energy Rev 2018;81:1484–512.
[41] Ahmed A, Khalid M. A review on the selected applications of forecasting models in renewable power systems. Renew Sustain Energy Rev 2019;100:9–21.
[42] Li B, Zhang J, Hobbs BF. A copula enhanced convolution for uncertainty aggregation. In: 2020 IEEE power energy society innovative smart grid technologies conference (ISGT). 2020, p. 1–5.
[43] CAISO. Flexible ramping product refinements draft final proposal. 2020, Accessed: 2020-08-10.
[44] Yesilbudak M, Sagiroglu S, Colak I. A new approach to very short term wind speed prediction using k-nearest neighbor classification. Energy Convers Manage 2013;69:77–86.
[45] Zhang Y, Wang J. K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting. Int J Forecast 2016;32(3):1074–80.
[46] Pedro HT, Coimbra CF. Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. Renew Energy 2015;80:770–82.
[47] Zhang Y, Wang J. GEFCom2014 probabilistic solar power forecasting based on k-nearest neighbor and kernel density estimator. In: 2015 IEEE power energy society general meeting. 2015, p. 1–5.
[48] Liu Z, Zhang Z. Solar forecasting by K-nearest neighbors method with weather classification and physical model. In: 2016 North American power symposium (NAPS). 2016.
[49] Friedman J, Hastie T, Tibshirani R. The elements of statistical learning: Data mining, inference, and prediction. Springer Science & Business Media; 2009.
[50] Inman RH, Pedro HT, Coimbra CF. Solar forecasting methods for renewable energy integration. Prog Energy Combust 2013;39(6):535–76.
[51] Engerer N, Mills F. KPV: A clear-sky index for photovoltaics. Sol Energy 2014;105:679–93.
[52] Holmgren WF, Hansen CW, Mikofski MA. Pvlib Python: A python package for modeling solar energy systems. J Open Source Softw 2018;3(29):884.
[53] Spyrou E, Krishnan V, Xu Q, Hobbs BF. What is the value of alternative methods for estimating ramping needs? In: 2020 IEEE green technologies conference(GreenTech). 2020, p. 159–64.
[54] Feng C, Zhang J. SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. Sol Energy 2020;204:71–8.
[55] Malof JM, Bradbury K, Collins LM, Newell RG. Automatic detection of solar photovoltaic arrays in high resolution aerial imagery. Appl Energy 2016;183:229–40.
[56] Feng C, Mehmani A, Zhang J. Deep learning-based real-time building occupancy detection using ami data. IEEE Trans Smart Grid 2020;11(5):4490–501.
[57] Feng C, Cui M, Hodge B-M, Zhang J. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. Appl Energy 2017;190:1245–57.
[58] Feng C, Cui M, Hodge B-M, Lu S, Hamann HF, Zhang J. Unsupervised clustering-based short-term solar forecasting. IEEE Trans Sustain Energy 2018;10(4):2174–85.
[59] Feng C, Sun M, Zhang J. Reinforced deterministic and probabilistic load forecasting via Q-learning dynamic model selection. IEEE Trans Smart Grid 2019;11(2):1377–86.
[60] Lu S, Hwang Y, Khabibrakhmanov I, Marianno FJ, Shao X, Zhang J, Hodge B-M, Hamann HF. Machine learning based multi-physical-model blending for enhancing renewable energy forecast - improvement via situation dependent error correction. In: 2015 European control conference (ECC). 2015, p. 283–90.