A Hybrid Electricity Price Scenario Generation Method for Stochastic Virtual Bidding in the Electricity Market

Dongliang Xiao, *Member, IEEE*, and Wei Qiao, *Fellow, IEEE*

***Abstract*---Stochastic optimization can be used to generate optimal bidding strategies for virtual bidders in which the uncertain electricity prices are represented by using scenarios. This paper proposes a hybrid scenario generation method for electricity price using a seasonal autoregressive integrated moving average (SARIMA) model and historical data. The electricity price spikes are first identified by using an outlier detection method. Then, the historical data are decomposed into base and spike components. Next, the base and spike component scenarios are generated by using the SARIMA- and historical data-based methods, respectively. Finally, the electricity price scenarios are obtained by combining the base and spike component scenarios. Case studies are carried out for a virtual bidder in the PJM electricity market to validate the proposed method. The optimal bidding strategies of the virtual bidder are generated by solving a stochastic optimization problem using the electricity price scenarios generated by the proposed, the SARIMA, and a historical data-based method, respectively. Case study results show that the proposed method is better than the SARIMA method in preserving statistical properties of the electricity price in the generated scenarios and is better than the historical data-based method in predicting the future trend of the electricity price and, therefore, can help the virtual bidder earn more profit in the electricity market.**

***Index Terms*—Electricity market, electricity price, scenario generation, stochastic optimization, virtual bidding.**

I. Nomenclature

A. Indices and Sets:

t	Index of time periods, running from 1 to T.
w	Index of scenarios, running from 1 to Ω.
g	Index of the autoregressive terms in an SARIMA model, running from 1 to G.
h	Index the moving average terms in an SARIMA model, running from 1 to H.
i	Index of the seasonal autoregressive terms in an SARIMA model, running from 1 to P.
j	Index of the seasonal moving average terms in an SARIMA model, running from 1 to Q.
m	Index of the elements in a dataset, running from 1 to M.
Λ_t^D/Λ_t^R	Set of the scenarios for day-ahead (DA)/real-time (RT) electricity price in a time period t.
$\Lambda_t^{Db}/\Lambda_t^{Rb}$	Set of the base component scenarios for DA/RT electricity price in a time period t.
$\Lambda_t^{Ds}/\Lambda_t^{Rs}$	Set of the spike component scenarios for DA/RT electricity price in a time period t.
X^D/X^R	Set of historical DA/RT electricity price data.
X^{Db}/X^{Rb}	Set of the base component of historical DA/RT electricity price data.
X^{Ds}/X^{Rs}	Set of the spike component of historical DA/RT electricity price data.

B. Decision Variables:

P_{tw}^{VI}/P_{tw}^{VD}	Power sold/bought by a virtual bidder in the DA market for a scenario w in a time period t when an incremental/decremental bidding curve is used.

C. Parameters and Constants:

P^{Vmax}	Maximum bidding capacity of a virtual bidder.
pr_{tw}	Probability of the occurrence of a scenario w in a time period t.
$\lambda_{tw}^D/\lambda_{tw}^R$	DA/RT electricity price for a scenario w in a time period t in the scenario set Λ_t^D/Λ_t^R
$\lambda_{tw}^{Db}/\lambda_{tw}^{Rb}$	The wth base component scenario of the DA/RT electricity price scenario set $\Lambda_t^{Db}/\Lambda_t^{Rb}$
$\lambda_{tw}^{Ds}/\lambda_{tw}^{Rs}$	The wth spike component scenario of the DA/RT electricity price scenario set $\Lambda_t^{Ds}/\Lambda_t^{Rs}$
x_m^D/x_m^R	The mth DA/RT electricity price data in the historical dataset X^D/X^R
x_m^{Db}/x_m^{Rb}	The mth base component of the DA/RT electricity price data in the historical dataset X^{Db}/X^{Rb}
x_m^{Ds}/x_m^{Rs}	The mth spike component of DA/RT electricity price data in the historical dataset X^{Ds}/X^{Rs}
z_m^D/z_m^R	The mth binary parameter in the vector Z^D/Z^R, which is equal to 1 if x_m^D/x_m^R is identified as a price spike, and is equal to 0 otherwise.
Z^D/Z^R	The vector used to mark the price spikes in the dataset X^D/X^R.
S	Seasonality order in an SARIMA model.
d	Differentiation order in an SARIMA model.
ϕ_g	The gth autoregressive parameter in an SARIMA model.
θ_h	The hth moving average parameter in an SARIMA model.
D	Seasonal differentiation order in an SARIMA model.
Φ_i	The ith seasonal autoregressive parameter in an SARIMA model.

This work was supported in part by the Nebraska Public Power District through the Nebraska Center for Energy Sciences Research.

Dongliang Xiao is with the School of Electric Power, South China University of Technology, Guangzhou 510641, China (email: dxiao@huskers.unl.edu).

Wei Qiao (corresponding author) is with the Power and Energy Systems Laboratory, Department of Electrical and Computer Engineering, University of Nebraska-Lincoln, Lincoln, NE 68588-0511 USA (e-mail: wqiao3@unl.edu).

DOI: 10.17775/CSEEJPES.2021.00890

Θ_j The jth seasonal moving-average parameter in an SARIMA model.

$\varepsilon_{tw}^{Db}/\varepsilon_{tw}^{Rb}$ Independent error term of a base component scenario w for the DA/RT electricity price in the independent error vector E_t^{Db}/E_t^{Rb}.

E_t^{Db}/E_t^{Rb} Independent error vector of the base component scenario sets for the DA/RT electricity price.

$\bar{\varepsilon}_{tw}^{Db}/\bar{\varepsilon}_{tw}^{Rb}$ Dependent error term of a base component scenario w for the DA/RT electricity price in the dependent error vector $\bar{E}_t^{Db}/\bar{E}_t^{Rb}$.

$\bar{E}_t^{Db}/\bar{E}_t^{Rb}$ Dependent error vector of the base component scenario sets for the DA/RT electricity price.

α Skewness of a dataset.

β Kurtosis of a dataset.

$\mathrm{K}_{X^{Db},X^{Rb}}$ Variance-covariance matrix of the base component datasets X^{Db} and X^{Rb}.

$\Sigma_{X^{Db},X^{Db}}$ Variance of the dataset X^{Db}.

$\Sigma_{X^{Db},X^{Rb}}$ Covariance of the base component datasets X^{Db} and X^{Rb}.

$L_{\Lambda_t^{Db},\Lambda_t^{Db}}$ Transformation matrix used for correlating the error terms of the base component scenario sets Λ_t^{Db} and Λ_t^{Rb}.

$\overline{X^D}/\overline{X^R}$ Sample mean value of the dataset X^D/X^R.

II. Introduction

MOST wholesale electricity markets in the United States have a two-settlement structure, which includes a DA and an RT markets. In the DA market, the participants, such as power producers and load serving entities, submit bids one day before the operating day based on their DA schedules; the market is then cleared and the cleared powers are settled at DA prices. In the RT market, the power deviations from the DA schedules are settled at RT prices on the operating day [1]. In addition to the market participants that have physical assets on the demand or generation sides, pure financial participants can also buy or sell power at DA prices in the DA market and their DA commitments are settled at RT prices in the RT market on the next day. This kind of transaction is called virtual bidding or convergence bidding; and these pure financial participants are referred to as virtual bidders whose profitability is related to the difference between DA and RT electricity prices [2]. Virtual bidding was first used in the PJM market in 2000 and currently is available in most U.S. electricity markets [3].

The main purposes of allowing virtual bidders to participate in electricity markets are to increase the liquidity and reduce price difference between DA and RT markets. The benefits and drawbacks that virtual bidding may bring to the market were discussed in [4]. By analyzing the historical data in California electricity market, the work [5] and [6] concluded that virtual bidding could reduce the difference between DA and RT electricity prices. However, virtual bidding might not improve market efficiency if used by a financial transmission right holder [7] or a cyber attacker [8]. Additionally, the impact of virtual bidding on market efficiency depends on the forecast accuracy of the virtual bidder. The study in [2] showed that the virtual bidders with perfect forecast results could improve the efficiency of electricity markets. However, the authors of [9] addressed that the virtual bidders with bad forecast results would decrease the total social welfare and should be screened out of the electricity markets. Thus, the virtual bidders' forecast accuracy affects both the profitability of virtual bidding and the efficiency of electricity markets.

The stochastic optimization technique can be used to generate optimal bidding strategies for virtual bidders while addressing the electricity price uncertainties via scenarios. In this circumstance, the forecasted hourly electricity price is represented by a set of scenarios with certain probabilities instead of a deterministic value, and the accuracy of the generated scenarios affects the profit of the virtual bidder significantly. In the literature, scenario generation methods based on statistical models [10]-[16] or historical data [17] have been reported for power system applications. In [10]-[12], the SARIMA model was used to generate scenarios for electricity price and renewable energy productions, respectively. In [13], the wind power scenarios were generated based on a multivariate normal distribution and the variability of the wind power was characterized by a range parameter in the covariance function. In [14], a generalized dynamic factor model was used to generate dependent load and wind power scenarios. In [15], a quantile regression forest model was employed to generate scenarios for wind, photovoltaic, and small hydro power productions. In [16], Weibull distribution was considered to generate wind speed scenarios, and transfer component analysis was utilized to improve the effectiveness of the scenario generation method.

Statistical models, such as SARIMA, can usually provide satisfactory scenario generation results if the historical data are stable and normally distributed [12]. However, the electricity price data are usually volatile and contain spikes caused by unexpected factors, such as power outages [18] and strategic behaviors of market participants [19]. In the United States, the average annual volatility of electricity price is 359.8%, which is much higher than the price volatility of natural gas (48.5%), financial assets (37.8%), metals (21.8%), and agriculture (49.1%) [20]. Moreover, there are much more positive price spikes than negative ones. Thus, the distribution of the electricity price data is asymmetric and very different from a normal distribution. The scenarios generated by the commonly used statistical models may not be able to capture and asymmetry and spikes of the electricity price.

To avoid the drawback of statistical models, the work [17] generated electricity price and wind power scenarios by using their historical data in different time periods directly for the stochastic wind power bidding. The generated scenarios were assigned with an equal probability. However, this method does not sufficiently utilize temporal correlations between historical data, which are considered in the SARIMA model and shown to be helpful for predicting the future trends of uncertain parameters [12]. Thus, the historical data-based method may not forecast the future trend of uncertain electricity price as accurately as the SARIMA-based method.

This paper proposes a hybrid scenario generation method that utilizes the temporal correlations of historical electricity price data without the need for any assumption of the historical data's distribution, such as the normal distribution assumption in the SARIMA-based method. In the proposed method, the spikes contained in the historical electricity price data are first identified. Then, the historical electricity price data is decomposed into base and spike components. Next, the

SARIMA method is used to generate the base component scenarios; and the spike component scenarios are generated from the historical data of the spike components directly. Finally, the base and spike scenarios are combined and used for the stochastic optimization problem for the virtual bidder. The main contributions of this paper are the following:

1) A hybrid method of generating electricity price scenarios based on the SARIMA model and historical data is proposed. By decomposing the historical electricity price data into base and spike components and generating their scenarios using different methods separately, both the trend and variations of the future electricity price can be captured in the generated scenarios without the need for any assumption of the historical data's distribution. The improved scenario generation by using the proposed method can help the virtual bidder earn more profits in the electricity market.

2) The statistical properties of the scenarios generated by different methods are studied in detail. It is found that the scenarios generated by the proposed method can characterize the volatility, asymmetry, and heavy tails of electricity prices more accurately than those generated by the SARIMA-based method without outlier detection.

The remainder of this paper is organized as follows. Section III presents the stochastic optimization problem of generating optimal virtual bidding strategies in electricity markets. Section IV presents the proposed hybrid electricity price scenario generation method. Section V compares different scenario generation methods. Section VI presents results of case studies. Section VII concludes the paper.

III. Stochastic Virtual Bidding Strategy

A. Market Framework

Fig. 1 shows the typical time frame of a two-settlement electricity market widely used in the U.S [21]. On the day before the operating day, virtual bidders submit incremental and decremental virtual bids before the submission closure time for the DA market; and both the cleared virtual bids and prices are determined by the DA market clearing process. On the operating day, the RT power balance is ensured through a RT market clearing process; and the deviations caused by the DA virtual bids need to be settled at RT electricity prices on this trading floor. To make the virtual bidding profitable in the electricity market, the virtual bidders need to forecast DA and RT electricity prices accurately using the latest historical data.

In contrast to the conventional power producers or electricity retailers that have physical resources in the power grid, the virtual bidders are pure financial participants without any physical resources. The maximum bidding capacity of a virtual bidder is determined by the available credit in its trading account, and a virtual bidder with higher credit would have a larger virtual bidding capacity.

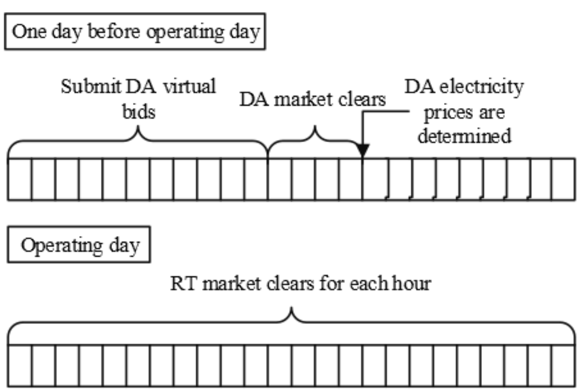

Fig. 1. Time frame of a typical two-settlement electricity market.

B. Stochastic Optimization Model for Virtual Bidding

In this paper, the virtual bidder is assumed to be a price-taker in the electricity market and its bidding capacity is not large enough to influence DA or RT electricity price. The mathematical model for optimizing the expected profits of a virtual bidder in a time period *t* is as follows:

$$\underset{P_{tw}^{VI},P_{tw}^{VD}}{\text{Max}}\ \pi_t = \sum_{w=1}^{\Omega} pr_{tw}[(\lambda_{tw}^D - \lambda_{tw}^R)P_{tw}^{VI} + (\lambda_{tw}^R - \lambda_{tw}^D)P_{tw}^{VD}] \tag{1}$$

Subject to:

$$0 \le P_{tw}^{VI},\ \forall t, \omega \tag{2}$$
$$0 \le P_{tw}^{VD},\ \forall t, \omega \tag{3}$$
$$P_{tw}^{VD} + P_{tw}^{VI} \le P^{Vmax},\ \forall t \tag{4}$$
$$P_{tw}^{VI} = P_{tw'}^{VI},\ \forall t, \omega, \omega': \lambda_{tw}^D = \lambda_{tw'}^D \tag{5}$$
$$P_{tw}^{VD} = P_{tw'}^{VD},\ \forall t, \omega, \omega': \lambda_{tw}^D = \lambda_{tw'}^D \tag{6}$$
$$\left(\lambda_{tw}^D - \lambda_{tw'}^D\right)\left(P_{tw}^{VI} - P_{tw'}^{VI}\right) \ge 0,\ \forall t, \omega \tag{7}$$
$$\left(\lambda_{tw}^D - \lambda_{tw'}^D\right)\left(P_{tw}^{VD} - P_{tw'}^{VD}\right) \le 0, \forall t, \omega \tag{8}$$

where the objective function is the expected profit of the virtual bidding in the time period *t*; Constraints (2)-(4) limit the virtual bidding capacities in the DA market; Constraints (5) and (6) ensure that the scenarios with the same DA electricity price have the same DA virtual bid capacity on the bidding curves; and Constraints (7) and (8) constitute the non-decreasing and non-increasing properties for the incremental and decremental bidding curves, respectively.

In the stochastic optimization problem of the virtual bidder, since the expected profit is calculated based on the scenarios of uncertain DA and RT electricity prices, the scenario values significantly affect the profitability of the DA virtual bidding strategy obtained by solving the optimization problem. If the future trend or some probabilistic properties of the uncertain prices cannot be captured by the generated scenario sets $\Lambda_t^D = \{\lambda_{tw}^D\}_{w=1}^{\Omega}$ and $\Lambda_t^R = \{\lambda_{tw}^R\}_{w=1}^{\Omega}$ sufficiently, the expected profit calculated by using (1) will deviate significantly from the actual profit. This circumstance indicates that the objective function and constraints containing the uncertain parameters of the stochastic optimization problem are incorrect so that the generated DA bidding strategy may not help the virtual bidder earn profit effectively in the electricity market.

Additionally, as shown in (1), a virtual bidder's profit depends on the absolute value of the difference between DA and RT prices. If DA electricity price is close to RT electricity price, the virtual bidder could not earn much profit no matter what scenario generation method is adopted, because there are no arbitrage opportunities in DA and RT electricity markets.

IV. Proposed Hybrid Scenario Generation Method

This section presents the overall framework and detailed procedure of the proposed hybrid scenario generation method for electricity price. The generated scenarios are used in the stochastic optimization problem (1)-(8) given in Section III.

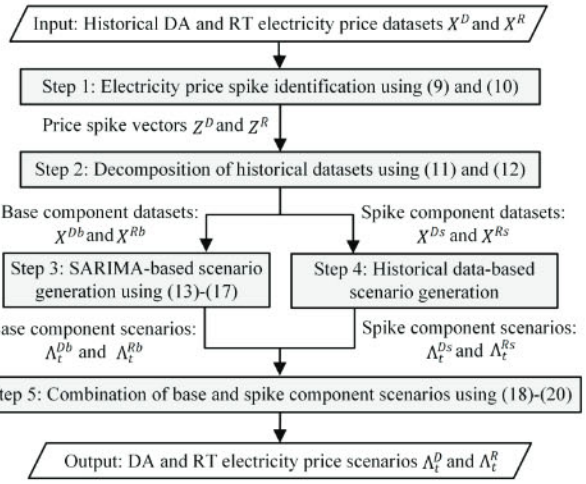

Fig. 2. Framework of the proposed hybrid scenario generation method.

A. Overall Framework

Fig. 2 shows the flowchart of the proposed method. The scenario sets of the DA electricity price $\Lambda_t^D = \{\lambda_{tw}^D\}_{w=1}^{\Omega}$ and RT electricity price $\Lambda_t^R = \{\lambda_{tw}^R\}_{w=1}^{\Omega}$ are generated jointly based on the historical DA electricity price dataset $X^D = \{x_m^D\}_{m=1}^M$ and RT electricity price dataset $X^R = \{x_m^R\}_{m=1}^M$ while considering the dependency between DA and RT prices.

First, the spikes contained in the historical electricity price datasets are identified using an outlier detection algorithm. Then, the original historical DA price dataset X^D is decomposed into a base component dataset X^{Db} and a spike component dataset X^{Ds}. Similarly, the original historical RT price dataset X^R is decomposed into a base component dataset X^{Rb} and a spike component dataset X^{Rs}. Next, an SARIMA-based method is designed to generate the base component scenario sets Λ_t^{Db} and Λ_t^{Rb}; and the spike component scenario sets Λ_t^{Ds} and Λ_t^{Rs} are generated from the spike component datasets directly. Finally, by adding the base and spike component scenarios, the final DA and RT price scenario sets Λ_t^D and Λ_t^R are obtained. Each step of the proposed hybrid method for generating the DA and RT electricity price scenario sets Λ_t^D and Λ_t^R from the historical datasets X^D and X^R, respectively, is presented in detail as follows.

B. Step 1: Identification of Price Spikes

Figs. 3 and 4 provide the histograms of the historical DA and RT electricity prices for a certain month in the PJM electricity market, respectively. Some DA/RT electricity price data patterns deviate significantly from their mean or median value and, thus, can be regarded as price spikes or outliers from a statistical perspective. Therefore, an outlier detection method based on the median absolute deviation (MAD) is designed to identify the price spikes.

Specifically, a data pattern is identified as a spike if it deviates more than three times the MAD from the median value of the historical data [22]. The MAD of the dataset X^D can be calculated as follows.

$$MAD(X^D) = 1.4826\ median(|X^D - median(X^D)|) \tag{9}$$

where $median()$ is the function of calculating the median of a dataset. Then, each DA electricity price spike is identified and marked using a binary parameter z_m^D as follows.

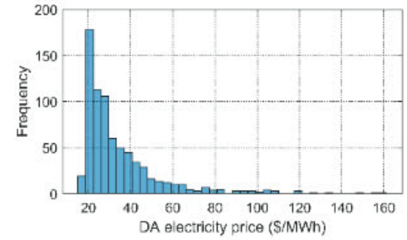

Fig. 3. Histogram of the DA electricity price for a certain month.

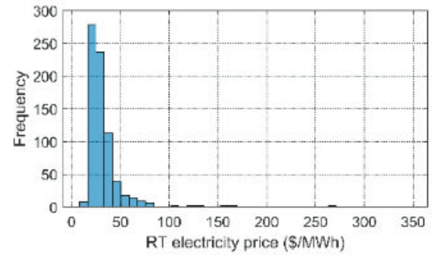

Fig. 4. Histogram of the RT electricity price for a certain month.

$$z_m^D = \begin{cases} 0, & |x_m^D - median(X^D)| \le 3\text{MAD}(X^D) \\ 1, & |x_m^D - median(X^D)| > 3\text{MAD}(X^D) \end{cases} \tag{10}$$

where z_m^D is equal to 1 if the mth data pattern is identified as a price spike. Finally, a vector $Z^D = \{z_m^D\}_{m=1}^M$ can be obtained to mark all the DA electricity price spikes in the dataset X^D.

The RT electricity price spikes in the dataset X^R can be identified using the method similar to (9) and (10) and marked by another vector $Z^R = \{z_m^R\}_{m=1}^M$. The proposed outlier detection method is based on MAD instead of standard deviation, because MAD is robust to outliers and can be used to measure the dispersion of the data more accurately than standard deviation.

C. Step 2: Decomposition of Historical Dataset

After the price spikes have been identified, the historical DA price dataset $X^D = \{x_m^D\}_{m=1}^M$ is decomposed into a base component dataset $X^{D\,b} = \{x_m^{Db}\}_{m=1}^M$ and a spike component dataset $X^{Ds} = \{x_m^{Ds}\}_{m=1}^M$. For each historical price data pattern identified as a spike, the base component is equal to the median value of the historical data, and the spike component is equal to the original price data pattern minus the median value. For each historical data pattern that is not identified as a spike, the base component is equal to the original data pattern, and the spike component is zero. The formulas for calculating the base and spike components of the historical DA electricity price data are expressed as (11) and (12), respectively.

$$x_m^{Db} = \begin{cases} x_m^D, & \text{if } z_m^D = 0 \\ median(X^D), & \text{if } z_m^D = 1 \end{cases} \tag{11}$$

$$x_m^{Ds} = \begin{cases} 0, & \text{if } z_m^D = 0 \\ x_m^D - median(X^D), & \text{if } z_m^D = 1 \end{cases} \tag{12}$$

The historical RT electricity price dataset X^R is also decomposed into a base component dataset $X^{Rb} = \{x_m^{Rb}\}_{m=1}^M$ and a spike component dataset $X^{Rs} = \{x_m^{Rs}\}_{m=1}^M$ using formulas similar to (11) and (12), respectively.

D. Step 3: Base Component Scenario Generation

To generate scenarios for multiple uncertain parameters by using statistical models, the joint probability distribution needs to be estimated first, which is generally a complex work. However, if the uncertain parameters are assumed to follow a multivariate Gaussian distribution, the scenario generation process can be simplified by using the univariate SARIMA model and variance-covariance matrices [12].

The base component scenarios of the DA electricity price λ_{tw}^{Db} in a certain time period t can be generated using the SARIMA model, which are expressed as follows.

$$\left(1-\sum_{g=1}^{G}\phi_g B^g\right)\left(1-\sum_{i=1}^{P}\Phi_i B^{iS}\right)(1-B)^d(1-B^S)^D\lambda_{tw}^{Db} = \left(1-\sum_{h=1}^{H}\theta_h B^h\right)\left(1-\sum_{j=1}^{Q}\Theta_j B^{jS}\right)\varepsilon_{tw}^{Db} \tag{13}$$

where S is the seasonality order, $\phi_1,\phi_2,\cdots,\phi_G$ are G autoregressive parameters; $\theta_1,\theta_2,\cdots,\theta_H$ are H moving average parameters; $\Phi_1,\Phi_2,\cdots,\Phi_P$ are P seasonal autoregressive parameters; and $\Theta_1,\Theta_2,\cdots,\Theta_Q$ are Q seasonal moving average parameters; ε_{tw}^{Db} represents the forecast error for the scenario w, which follows an independent normal probability distribution for the SARIMA model; and B is the backward shift operator, whose function is given as follows.

$$B^d\lambda_{tw}^{Db} = \lambda_{t-d,w}^{Db} \tag{14}$$

Based on the historical base component dataset X^{Db}, the parameters of (13), which include ϕ_g, Φ_i, θ_h, and Θ_j, can be estimated by using the maximum likelihood method. To generate a scenario λ_{tw}^{Db} in the time period t, an error term ε_{tw}^{Db} needs to be first sampled from a normal probability distribution. The scenario λ_{tw}^{Rb} and error term ε_{tw}^{Rb} for the RT price can be generated in a way similar to (13) and (14).

To consider the dependency between DA and RT prices in the scenario generation process, their correlation is modeled by using the variance-covariance matrix $\mathrm{K}_{X^{Db},X^{Rb}}$ of the DA price base component dataset X^{Db} and the RT price base component dataset X^{Rb} expressed as follows.

$$\mathrm{K}_{X^{Db},X^{Rb}} = \begin{bmatrix}\Sigma_{X^{Db},X^{Db}} & \Sigma_{X^{Db},X^{Rb}}\\ \Sigma_{X^{Db},X^{Rb}} & \Sigma_{X^{Rb},X^{Rb}}\end{bmatrix} \tag{15}$$

where $\Sigma_{X^{Db},X^{Db}}$ is the variance of X^{Db}, $\Sigma_{X^{Rb},X^{Rb}}$ is the variance of X^{Rb}, and $\Sigma_{X^{Db},X^{Rb}}$ is the covariance of X^{Db} and X^{Rb}. Then, the Cholesky decomposition is performed for the variance-covariance matrix as follows.

$$\mathrm{K}_{X^{Db},X^{Rb}} = L_{\Lambda_t^{Db},\Lambda_t^{Db}} L_{\Lambda_t^{Db},\Lambda_t^{Db}}^T \tag{16}$$

where $L_{\Lambda_t^{Db},\Lambda_t^{Db}}$ is the transformation matrix used for correlating the error terms of the scenario sets Λ_t^{Db} and Λ_t^{Rb}. Let $E_t^{Db} = [\varepsilon_{tw}^{Db}]_{\Omega\times 1}$ and $E_t^{Rb} = [\varepsilon_{tw}^{Rb}]_{\Omega\times 1}$ be the column vectors containing Ω independent error terms of the scenario sets Λ_t^{Db} and Λ_t^{Rb}, respectively. Then, the independent error vectors E_t^{Db} and E_t^{Rb} are transformed to be the dependent error vectors $\bar{E}_t^{Db} = [\bar{\varepsilon}_{tw}^{Db}]_{\Omega\times 1}$ and $\bar{E}_t^{Rb} = [\bar{\varepsilon}_{tw}^{Rb}]_{\Omega\times 1}$ as follows.

$$\begin{bmatrix}\bar{E}_t^{Db}\\ \bar{E}_t^{Rb}\end{bmatrix} = L_{\Lambda_t^{Db},\Lambda_t^{Db}}\begin{bmatrix}E_t^{Db}\\ E_t^{Rb}\end{bmatrix} \tag{17}$$

Then, the SARIMA model (13) is modified by replacing the independent error term ε_{tw}^{Db} with the dependent error term $\bar{\varepsilon}_{tw}^{Db}$. Finally, the base component scenario sets $\Lambda_t^{Db} = \{\lambda_{tw}^{Db}\}_{w=1}^{\Omega}$ and $\Lambda_t^{Rb} = \{\lambda_{tw}^{Rb}\}_{w=1}^{\Omega}$ are generated by using the modified SARIMA model.

E. Step 4: Spike Component Scenario Generation

The price spikes in electricity markets can affect the market participants' economic benefits significantly. However, since price spikes are usually caused by some unexpected events in the power system, the spike component data are usually highly volatile and difficult to be forecasted using the SARIMA model. This work proposes to generate the spike component scenario sets Λ_t^{Ds} and Λ_t^{Rs} using the samples of the historical spike component datasets X^{Ds} and X^{Rs} in the tth hour, respectively, i.e., $\Lambda_t^{Ds} = \{\lambda_{tw}^{Ds}\}_{w=1}^{\Omega}$ and $\Lambda_t^{Rs} = \{\lambda_{tw}^{Rs}\}_{w=1}^{\Omega}$, where $\lambda_{tw}^{Ds}\in X^{Ds}$, $\lambda_{tw}^{Rs}\in X^{Rs}$, and $\Omega \le M$.

F. Step 5: Combination of Base and Spike Component Scenarios

Since the base and spike component scenario sets generated in Step 3 and 4, respectively, have the same number of scenarios, they can be added directly to generate the final DA price scenario set Λ_t^D and RT price scenario set Λ_t^R as follows.

$$\Lambda_t^D = \Lambda_t^{Db} + \Lambda_t^{Ds} \tag{18}$$

$$\Lambda_t^R = \Lambda_t^{Rb} + \Lambda_t^{Rs} \tag{19}$$

Since the sequence of the scenarios in each scenario set is random, the base and spike component scenarios are combined randomly. Additionally, since all of the scenarios in Λ_t^D and Λ_t^R are assigned with an equal probability, respectively, the probability pr_{tw} of each scenario w of the final DA and RT electricity price scenario sets is

$$pr_{tw} = \frac{1}{\Omega} \tag{20}$$

where Ω is the total number of scenarios generated for the stochastic optimization problem for the virtual bidder.

V. Comparison of Existing and Proposed Scenario Generation Methods

The SARIMA- and historical data-based scenario generation methods have been widely used in practice. However, both methods have disadvantages when used for electricity price scenario generation.

For the SARIMA-based scenario generation method, both the error term ε_{tw}^{Db} and the generated scenario λ_{tw}^{Db} in (13) are assumed to follow symmetric normal distributions, which may not be correct for actual probability distribution of electricity price. As shown in Figs. 3 and 4, the DA and RT price data are asymmetric and have heavy tails and, thus, are different from normal distributions and cannot be simply characterized by using mean and variance.

To measure the symmetricity and heavy-tailedness of a probability distribution, skewness and kurtosis are commonly used [23]. The skewness $\alpha(X^D)$ and kurtosis $\beta(X^D)$ of the dataset X^D are calculated as follows.

$$\alpha(X^D) = \frac{\frac{1}{M}\sum_{m=1}^{M}\left(x_m^D-\overline{X^D}\right)^3}{\left(\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(x_m^D-\overline{X^D}\right)^2}\right)^3} \tag{21}$$

$$\beta(X^D) = \frac{\frac{1}{M}\sum_{m=1}^{M}\left(x_m^D-\overline{X^D}\right)^4}{\left(\frac{1}{M}\sum_{m=1}^{M}\left(x_m^D-\overline{X^D}\right)^2\right)^2} \tag{22}$$

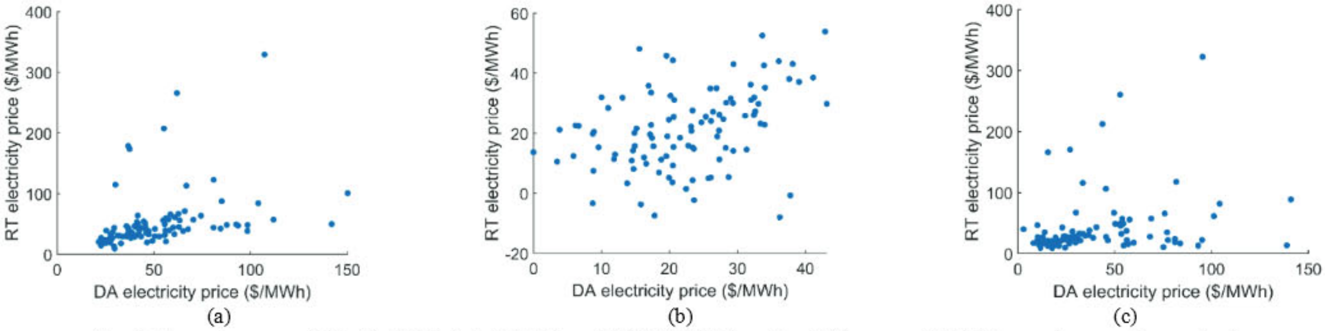

Fig. 5. Scenarios generated by the (a) historical data-based, (b) SARIMA-based, and (c) proposed hybrid scenario generation methods.

In the SARIMA model, since the normal distributions used for characterizing the error terms are symmetric, their skewness values are 0. Additionally, since the tails of the normal distributions have the same shape, their kurtosis values are the same, which is 3 [23]. However, for the DA and RT price data shown in Figs. 3 and 4, their skewness values are 2.62 and 5.88, respectively; and their kurtosis values are 11.78 and 44.5, respectively, which are much larger than 3. These values indicate that the DA and RT price data are asymmetric and more heavy-tailed than normal distributions. Thus, it is not accurate to generate the electricity price scenarios by using the SARIMA method, which cannot characterize some key statistical properties of the price data correctly.

Compared to the SARIMA method, the scenarios generated by using the historical data directly can preserve more statistical properties, such as asymmetry and heavy-tailedness, of the electricity price, but do not fully utilize the temporal correlations of the historical data. In contrast, the SARIMA model (13) considers the temporal correlations of the historical data of the uncertain electricity price and, therefore, is superior to the historical data-based method for predicting the future trend of the uncertain electricity price [12].

The proposed hybrid scenario generation method utilizes the advantages of both the SARIMA and the historical data-based methods. On one hand, the base component data is more stable than the original data. Thus, it is more suitable to use the SARIMA model to generate the base component scenarios. On the other hand, the spike scenarios generated by the historical data-based method preserve more statistical properties of the historical data without any assumption for its distribution. Thus, the scenarios generated by the proposed hybrid method can capture the future trend and preserve important statistical properties, such as skewness and kurtosis, of the electricity price, which can help increase the profit of the virtual bidding strategy obtained by solving the stochastic optimization problem (1)-(8).

VI. Case Studies And Results

A. Simulation Setup

The proposed hybrid electricity price scenario generation method is validated via case studies for a virtual bidder in the PJM electricity market. The virtual bidder has the maximum virtual capacity of 30 MW and is assumed to submit DA virtual bids at the Eastern Hub in the PJM electricity market. The hourly DA and RT electricity price data at the trading hub are publicly available on the PJM website. The historical data from June 2018 to May 2019 are used for the case studies. For each operating day, the historical data of the last three months are used to generate the electricity price scenarios and stochastic virtual bidding strategies for different hours on the next day. For instance, the scenarios and virtual bidding strategies for September 1, 2018 are generated based on the historical data from June 1, 2018 to August 31, 2018. The parameters of the SARIMA model are estimated using the MATLAB econometric toolbox. Since the stochastic optimization problem (1)-(8) for the virtual bidder is a linear programing (LP) problem, it can be solved efficiently by using the Yalmip toolbox [24] and Gurobi in MATLAB [25].

B. Results of the Scenarios Generated by Different Methods

The results of 100 scenarios generated by using the historical data-based, SARIMA-based, and proposed hybrid methods in a certain hour are shown in Fig. 5. Fig. 5(a) shows that there are many positive price spikes in the DA and RT markets, whose values deviate significantly from the mean values of the data. For instance, the mean value of the RT electricity price data in Fig. 5(a) is 51.5 $/MWh; while the RT price spikes are as high as 329.2 $/MWh, which makes the distribution of the historical price data pretty asymmetric. In this circumstance, a statistical model, such as the SARIMA model, is not capable of fully characterizing the statistical properties of the data because the error terms in the SARIMA model (13) are assumed to follow a normal distribution, As shown in Fig. 5(b), the scenarios generated by the SARIMA model are symmetric and their maximum deviation from the mean is much smaller than that in Fig. 5(a). Thus, the information of the price spikes is lost in the scenarios of Fig. 5(b). When using the proposed hybrid scenario generation method, some important statistical properties of the historical data, such as symmetricity and heavy-tailedness, are preserved. As shown in Fig. 5(c), quite a few DA and RT electricity price scenarios generated by the proposed method deviate significantly from the mean values, which is similar to the result in Fig. 5(a). However, the minimum DA electricity price scenario value in Fig. 5(c) is close to that in Fig. 5(b) but is lower than that in Fig. 5(a), because the base components of the scenarios in Fig. 5(c) are generated using the SARIMA model instead of the historical data.

To further compare the proposed method with the historical data-based and the SARIMA methods, four key statistical parameters, including mean, variance, skewness, and kurtosis,

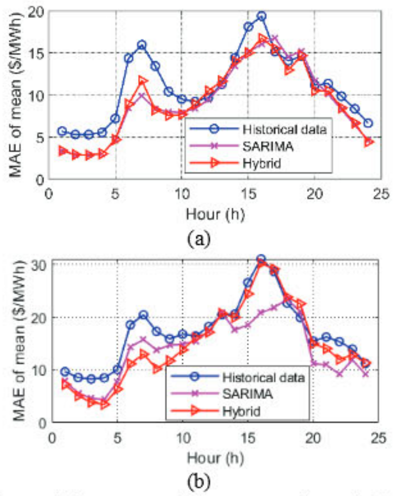

Fig. 6. Comparison of the MAEs of the mean values of (a) DA and (b) RT electricity price scenarios generated by different methods.

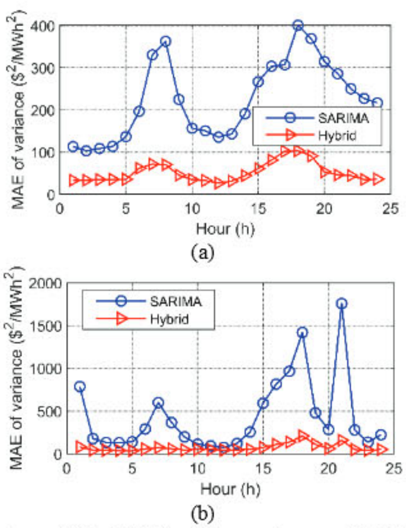

Fig. 7. Comparison of the MAEs of the variance of (a) DA and (b) RT electricity price scenarios generated by different methods.

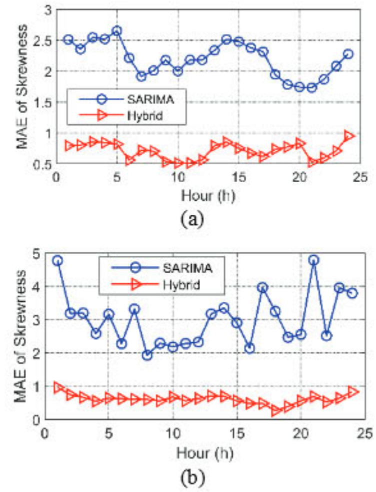

Fig. 8. Comparison of the MAEs of the skewness of (a) DA and (b) RT electricity price scenarios generated by different methods.

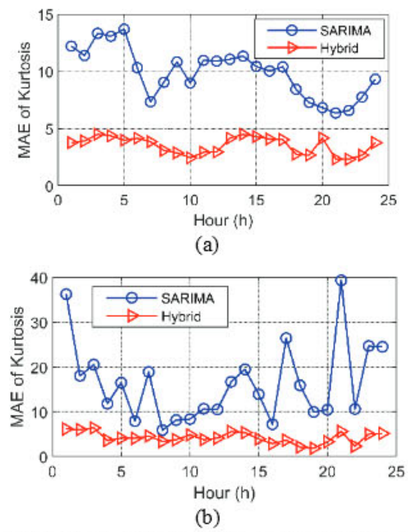

Fig. 9. Comparison of the MAEs of the kurtosis of (a) DA and (b) RT electricity price scenarios generated by different methods.

of the scenarios generated for each hour of a day over 8 months from June 2018 to May 2019 (called hourly scenarios) are calculated. First, the mean absolute error (MAE) between the mean value of the hourly DA/RT electricity price scenarios generated by each of the three methods and the actual DA/RT electricity price values over the 8 months is calculated for each hour of a day. The resulting MAEs of the mean values of the DA and RT electricity price scenarios generated by the three different methods for the 24 hours of a day are compared in Fig. 6. The results show that in most hours, the MAEs of the historical data-based method are larger than those of the other two methods that use the SARIMA model. Thus, the SARIMA and proposed hybrid methods can forecast the future trend of the DA/RT electricity price, which can be represented by the mean value of its scenarios generated for each hour, more accurately than the historical data-based method.

Next, the MAEs of the variance values of the DA/RT electricity price scenarios generated by the SARIMA method and the proposed method, respectively, with respect to that generated by the historical data-based method over the 8 months are calculated for each hour of a day. The resulting MAEs of the variance values of the DA and RT electricity price scenarios generated by the SARIMA and proposed methods for the 24 hours of a day are compared in Fig. 7. Similarly, the MAEs of the skewness and kurtosis values of the DA and RT electricity price scenarios generated by the SARIMA and proposed methods for the 24 hours of a day are calculated and compared in Figs. 8 and 9, respectively. The results show that the MAEs of variance, skewness, and kurtosis of the proposed method are lower than those of the SARIMA-based method, respectively, in all of the 24 hours. The results indicate that the scenarios generated by the proposed method can characterize the volatility, asymmetry, and heavy tails of the electricity prices more accurately than those generated by the SARIMA-based method.

C. Results of Different Stochastic Virtual Bidding Strategies

To study the impacts of different scenario generation methods on the stochastic virtual bidding strategies, 100 scenarios are generated for dependent DA and RT electricity prices, respectively, by using the three different methods and are used in the stochastic optimization problem (1)-(8), respectively. Then, by solving the problem (1)-(8), the optimal virtual bidding strategies are generated. Fig. 10 shows that the

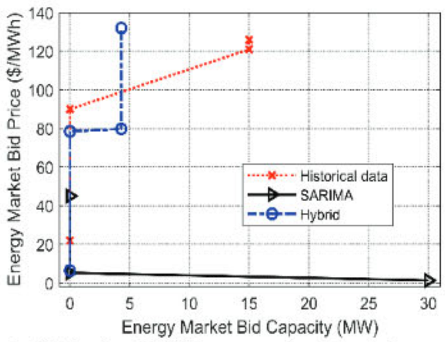

Fig. 10. Optimal DA virtual bidding curves in a certain hour generated by using different scenario generation methods.

TABLE I
ACTUAL PROFITS OF THE VIRTUAL BIDDER OBTAINED BY USING DIFFERENT SCENARIO GENERATION METHODS

Method	Historical data	SARIMA	Hybrid
Profit in October 2018 ($)	-823.55	-3676.55	-171.16
Profit in November 2018 ($)	3098.20	6234.21	843.49
Profit in December 2018 ($)	1259.48	27513.91	11702.89
Profit in January 2019 ($)	11422.61	8402.69	19254.58
Profit in February 2019 ($)	28757.2	15742.25	21922.16
Profit in March 2019 ($)	-42.45	-1394.07	81.27
Profit in April 2019 ($)	-283.34	-753.63	-151.95
Profit in May 2019 ($)	-445.46	-1027.54	98.74
Total profit of eight months ($)	42892.71	51041.27	53417.48

DA virtual bidding curves are different when using different scenario generation methods. When using the SARIMA method, the virtual bidder submits a decremental bidding curve to buy power in the DA market. However, when using the historical data-based and proposed hybrid scenario generation methods, the virtual bidder submits incremental bidding curves to sell power in the DA market.

Based on the generated virtual bidding curves and actual DA and RT electricity prices, the actual profits of the virtual bidder are calculated. Table I compares the actual monthly profits of the virtual bidder from October, 2018 to May, 2019 obtained using the three different scenario generation methods. The total profits of the 8 months obtained by using the three different scenario generation methods are all positive. This indicates that the virtual bidder can make profit in the electricity market regardless what scenario generation method is used. When using the proposed method, the virtual bidding is profitable in 6 months. However, when using the SARIMA and historical data-based scenario generation methods, the virtual bidding is only profitable in 4 months. Moreover, the proposed method outperforms the other two methods in 6 of the 8 months, and the total profit obtained using the proposed method is 24.54% and 4.66% higher than those obtained by using the historical data-based and SARIMA-based scenario generation methods, respectively. On the other hand, Table I shows that the historical data-based and SARIMA-based scenario generation methods outperform the proposed hybrid method in 2 of the 8 months. This is because certain extreme scenarios of the uncertain electricity prices caused by some unexpected events, such as sudden power outages, abnormal weather, etc. cannot be predicted accurately by using the historical data or the SARIMA model. In those circumstances, the actual profits obtained by using the three methods tend to be random. It should be pointed out that the performance of a scenario generation method should be evaluated by using a sufficiently large number of data samples, which are hourly virtual bidding results over a sufficiently long time, such as 8 months, in this paper. As the result, the proposed method is shown to be statistically better than the other two methods, which, however, does not guarantee that the proposed method is always better than the other two methods in all hours, days, or months for this virtual bidder's stochastic decision-making problem.

VII. CONCLUSIONS

This paper proposed a hybrid electricity price scenario generation method for generating bidding strategies for virtual bidders via stochastic optimization. In the proposed method, price spikes were identified using an outlier detection method; based on the spikes identified, the historical price data was decomposed into base and spike components. Then, the base component scenarios were generated by using the SARIMA model; and the spike component scenarios were generated by using the historical data-based method. The final electricity price scenarios were obtained by adding the base and spike component scenarios together.

The proposed method was validated and compared with the historical data and SARIMA-based methods for generating bidding strategies for a virtual bidder in the PJM market. The case study results showed that the scenarios generated by the proposed method could characterize the volatility, asymmetry, and heavy tails of the historical data more accurately than the SARIMA-based method. The total profit obtained using the proposed method was 24.54% and 4.66% higher than those obtained by using the historical data-based and SARIMA scenario generation methods, respectively.

In the future work, the proposed outlier detection-based hybrid scenario generation method can be extended to other decision-making problems with uncertainties, such as power system planning and renewable energy trading in electricity markets. Additionally, a price-maker virtual bidder can be studied by using a bilevel optimization model in which the scenarios of other market participants' bidding strategies can be generated by using the proposed method and the impact of the virtual bidder's bidding capacity on market outcomes can be analyzed.

REFERENCES

[1] A. J. Conejo, M. Carrion, and J. M. Morales, *Decision Making Under Uncertainty in Electricity Market*, New York, NY, USA: Springer, 2010.

[2] M. Kohansal, E. Samani and H. Mohsenian-Rad, "Understanding the Structural Characteristics of Convergence Bidding in Nodal Electricity Markets," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 1, pp. 124-134, Jan. 2021.

[3] D. Xiao, J. C. do Prado, and W. Qiao, "Optimal joint demand and virtual bidding for a strategic retailer in the short-term electricity market," *Electric Power Systems Research*, vol. 190, Jan. 2021.

[4] M. Celebi, A. Hajos, and P. Q. Hanser, "Virtual bidding: The good, the bad and the ugly," *The Electricity Journal*, vol. 23, no. 5, pp. 16–25, Jun. 2010.

[5] R. Li, A. J. Svoboda, and S. S. Oren, "Efficiency impact of convergence bidding in the California electricity market," *Journal of Regulatory Economics*, vol. 48, no.3, pp. 245-284, Dec. 2015.

[6] C. K. Woo, J. Zarnikau, E. Cutter, S. T. Ho, and H. Y. Leung, "Virtual bidding, wind generation and California's day-ahead electricity forward premium," *The Electricity Journal*, vol. 28, no. 1, pp. 29–48, Feb. 2015

[7] C. Lo Prete, N. Guo, and U. V. Shanbhag, "Virtual bidding and financial transmission rights: an equilibrium model for cross-product manipulation in electricity markets," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 953-967, Mar. 2019.
[8] S. Ahmadian, X. Tang, H. A. Malki, and Z. Han, "Modelling cyber attacks on electricity market using mathematical programming with equilibrium constraints," *IEEE Access*, vol. 7, pp. 27376-27388, Feb. 2019.
[9] W. Tang, R. Rajagopal, K. Poolla, and P. Varaiya, "Model and data analysis of two-settlement electricity market with virtual bidding," in *Proc. IEEE Conference on Decision and Control*, Dec. 2016, pp. 6645–6650.
[10] D. Xiao, W. Qiao, and L. Qu, "Risk-constrained stochastic virtual bidding in two-settlement electricity markets," in *Proc. IEEE Power and Energy Society General*, Aug. 2018, pp. 1-5.
[11] D. Xiao, W. Qiao, and L. Qu, "Risk-averse offer strategy of a photovoltaic solar power plant with virtual bidding in electricity markets," in *Proc. IEEE Power and Energy Society Innovative Smart Grid Technologies Conference*, Aug. 2019, pp. 1-5.
[12] J. M. Morales, R. Mnguez, and A. J. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Applied Energy*, vol. 87, no. 3, pp. 843-855, Mar. 2010.
[13] X.-Y. Ma, Y.-Z. Sun, and H.-L. Fang, "Scenario generation of wind power based on statistical uncertainty and variability," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 4, pp. 894–904, Oct. 2013.
[14] D. Lee and R. Baldick, "Load and wind power scenario generation through the generalized dynamic factor model," *IEEE Transactions on Power Systems*, vol. 32, no. 1, pp. 400–410, Jan. 2017.
[15] S. Camal, F. Teng, A. Michiorri, G. Kariniotakis, L. Badesa, "Scenario generation of aggregated wind, photovoltaics and small hydro production for power systems applications," *Applied Energy*, vol. 242, pp. 1396-1406, May 2019.
[16] J. Hu, H. Li, and Z. Liu, "A novel scenario generation framework based on the knowledge of existing wind power plants," *IEEE Transactions on Sustainable Energy*, to be published.
[17] L. Baringo and A. J. Conejo, "Strategic offering for a wind power producer," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4645–4654, Nov. 2013.
[18] C. L. Anderson and M. Davison, "A hybrid system-econometric model for electricity spot prices: Considering spike sensitivity to forced outage distributions," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 927–937, Aug. 2008.
[19] X. Guan, Y.-C. Ho, and D. L. Pepyne, "Gaming and price spikes in electric power markets," *IEEE Transactions on Power Systems*, vol. 16, no. 3, pp. 402–408, Aug. 2001.
[20] M. K. AlAshery, D. Xiao, and W. Qiao, "Second-order stochastic dominance constraints for risk management of a wind power producer's optimal bidding Strategy," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 3, pp. 1404-1413, Jul. 2020.
[21] PJM market [Online]. Available: https://www.pjm.com/
[22] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata, "Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median," *Journal of Experimental Social Psychology*, vol. 49, no. 4, pp. 764–766, 2013
[23] V. K. Rohatgi, *An Introduction to Probability Theory and Mathematical Statistics*, New York, NY, USA: Wiley, 1976.
[24] J. Lofberg, "YALMIP: A toolbox for modeling and optimization in MATLAB," in *Proc. IEEE International Conference on Robotics and Automation*, Sept. 2004, pp. 284-289.
[25] W. Yin, *Gurobi mex: a MATLAB interface for gurobi*, Nov. 2011 [Online].Available:http://www.convexoptimization.com/wikimization/index.php/Gurobi_Mex:_A_MATLAB_interface_for_Gurobi