

Article

Day-Ahead Market Modeling for Strategic Wind Power Producers under Robust Market Clearing

Huiru Zhao, Yuwei Wang, Mingrui Zhao, Qingkun Tan and Sen Guo *

School of Economics and Management, North China Electric Power University, Beijing 102206, China; zhaohuiru@ncepu.edu.cn (H.Z.); wangyuwei2017666@gmail.com (Y.W.); 1142206037@ncepu.edu.cn (M.Z.); tanqingkun@163.com (Q.T.)

* Correspondence: guosen324@163.com; Tel.: +86-010-61773084

Received: 21 March 2017; Accepted: 28 June 2017; Published: 4 July 2017

Abstract: In this paper, considering real time wind power uncertainties, the strategic behaviors of wind power producers adopting two different bidding modes in day-ahead electricity market is modeled and experimentally compared. These two different bidding modes only provide a wind power output plan and a bidding curve consisting of bidding price and power output, respectively. On the one hand, to significantly improve wind power accommodation, a robust market clearing model is employed for day-ahead market clearing implemented by an independent system operator. On the other hand, since the Least Squares Continuous Actor-Critic algorithm is demonstrated as an effective method in dealing with Markov decision-making problems with continuous state and action sets, we propose the Least Squares Continuous Actor-Critic-based approaches to model and simulate the dynamic bidding interaction processes of many wind power producers adopting two different bidding modes in the day-head electricity market under robust market clearing conditions, respectively. Simulations are implemented on the IEEE 30-bus test system with five strategic wind power producers, which verify the rationality of our proposed approaches. Moreover, the quantitative analysis and comparisons conducted in our simulations put forward some suggestions about leading wind power producers to reasonably bid in market and bidding mode selections.

Keywords: day-ahead electricity market; wind power producer; wind power uncertainty; robust market clearing; Least Squares Continuous Actor-Critic algorithm

1. Introduction

The day-ahead electricity market (EM) is a crucial component in the EM system [1]. In recent years, wind power resources have experienced an unprecedented growth in the day-ahead EMs worldwide. Studies on wind power bidding in the day-ahead EM with wind power penetration are too numerous to enumerate one by one. References [2–4] etc., for reasons such as low marginal cost of wind power producer (WPP) etc., hold that the bidding mode (BM) of a WPP is to only send the independent system operator (ISO) its power output plan for each period of the next day (namely, BM 1). The ISO ensures the wind power accommodation according to every WPP's day-ahead power output plan, but a WPP should be financially punished when its real time power output deviates from the day-ahead bidding one [3]. References [5–9] etc., based on actual EMs such as PJM (Pennsylvania-New Jersey-Maryland) etc., point out that the BM of a WPP is to provide ISO a bidding curve for each period of the next day (namely, BM 2). A bidding curve consists of bidding price and power output. According to these day-ahead bidding curves provided by WPPs, ISO, within a certain range of forecasted power outputs corresponding to each WPP, can dispatch the power outputs of WPPs in the day-ahead EM. However, a WPP should also be financially punished when its real time power output deviates from the day-ahead scheduled one [9]. In this work, we believed that different BMs adopted by WPPs may lead to different market results such as profits, clearing prices and operation cost of the power system.

Hence, one motivation of this work is to experimentally compare those two BMs adopted by WPPs in day-ahead EM.

In addition, owing to the inherent intermittence, fluctuation, and low predictability of wind power output, uncertainties grow significantly with the increasing penetration of wind power resources, which pose major challenges to wind power accommodation in the EM [10]. Therefore, many references propose that the wind power accommodation can be improved by modifying the market clearing model (MCM) corresponding to ISO in day-ahead EM. References [9,11–13] applied the stochastic optimization (SO) method in the market clearing procedure. The SO-MCM significantly increases the number of constraints in MCM by generating real time wind power output stochastic scenarios (WPOSSs) based on real time wind power output probability distributions. The market clearing results (scheduled power results and clearing price of every node) can be obtained by optimizing the expected value of the objective function in SO-MCM [9]. This approach takes into account different security constraints of power system under different WPOSSs, and improves, to a certain extent, the wind power accommodation capacity of power system. However, SO-MCM still has the following shortcomings, thereby greatly reducing the feasibility and rationality of this method [14,15]: (1) in practice, the probability distribution of real time wind power output is difficult to obtain; (2) a small number of real time WPOSSs may lead to a reduction of the ability of the power system to resist random real time wind power output deviations from its day-ahead (bidding or scheduled) one; (3) a large number of WPOSSs may significantly increase the computational complexity of the model, thereby resulting in solving difficulties. In order to overcome these abovementioned shortcomings, recently, robust optimization (RO) methods are applied to the construction of power system dispatch models by many studies. Reference [16] proposed a two-stage robust security constraint unit-commitment (SCUC) model. The key idea of this two-stage robust SCUC model is to determine the optimal unit-commitment (UC) solution in the first stage which leads to the least operation cost for the worst wind power output scenario (WPOS) in the second stage. However, this approach is very conservative due to the optimization for the operation cost of the worst WPOS in the second stage. In Reference [17] the authors combined the stochastic and robust approaches using a weight factor in the objective function to address the conservativeness issue. Reference [18] employed the Affine Policy (AP) to formulate and solve the robust security constraint economic dispatch (SCED) model. Reference [19] proposed a robust optimization framework for robust SCUC and robust SCED which repeatedly calculates the UC and ED solutions in the first stage to optimize the operation cost of the basic WPOS but to pass the security test in the second stage. RO-based power system dispatch models do not need the probability distribution of real time wind power output. The number of constraints need not be significantly increased with the increase of the size of uncertainty set [19]. The optimal robust UC and ED solutions can satisfy every unit-wise and system-wise constraint under the worst WPOS [19], which means RO-based power system dispatch models can not only improve wind power accommodation but also maintain low computational complexity so as to promote the application of these models in practice. However, the approaches in [16–19] cannot be introduced directly for modification of MCM in day-ahead EM because it is not mentioned in those studies how to price the power outputs, loads, reserves and deviations (uncertainties). Recently, in study of Ye et al. [20], this shortcoming is made up by combining cost causation principle and locational marginal price (LMP) in robust SCUC and robust SCED modeling approaches so as to successfully modify the MCM in day-ahead EM by using a RO method. Therefore, inspired by [20], in this work, no matter which BM WPPs adopt in day-ahead EM bidding, the MCM corresponding to ISO will be modified by using a RO-method in order to make the power system accommodate any deviation caused by real time wind power uncertainties within a certain range.

Finally, a WPP participating in EM bidding aims at profit maximization. In a day-ahead EM, there are many participants competing with each other. In addition to real time wind power uncertainties, WPPs, like other conventional generation companies (GenCOs), are faced with complex market environment conditions, such as imperfect and incomplete information. Hence, there are many

similarities between EM modeling approaches with and without WPPs participating in market bidding. EM modeling approaches proposed in [21–25] are based on game theory. EM modeling approaches proposed in [26–31] are based on machine learning algorithms. Recently, many relevant studies take renewable energy (i.e., wind power) bidding into account. Reference [32] supposed that a WPP strategically bids with BM 2 in day-ahead EM, and put forward a closed-form analysis on WPP's strategic behavior based on the Stackelberg game model. Reference [33] proposed an autoregressive integrated moving average (ARIMA) model to obtain the optimal bidding strategy for a WPP who bids with BM 2 in day-ahead EM. The authors of [34] studied the behaviors of strategic WPPs bidding in EM with BM 1 based on Cournot game model. In Reference [35] an imbalance cost minimization bidding strategy for a BM 1 adopted WPP through forecasting the real time wind power probability distribution functions was proposed. Reference [36] analyzed the strategic behavior of a BM 1 adopted WPP in day-ahead EM based on Roth-Erev reinforcement learning algorithm. In [37], a stochastic programming problem was proposed for obtaining the optimal offering and operating strategy for a large wind-storage system adopting BM 2. Reference [38] considered the uncertainty on electricity price through a set of exogenous scenarios and solved the bidding problem of a BM 1 adopted thermal-wind power producer by using a stochastic mixed-integer linear programming approach. The authors in reference [39] proposed a two-stage stochastic bidding model based on kernel density estimation (KDE) for a BM 2 adopted WPP to obtain the optimal day-ahead bidding strategy. The approaches in [24,25,33,35,37–39] resulted in repeatedly solving multi-level mathematical programming models for every participant, the computational complexities of which limit their applications in more realistic situations. The methods proposed in [21–23,32,34] produced sets of nonlinear equations which are difficult to solve or have no solutions. The approaches in [26–31,36] belong to the agent-based EM modeling approaches, in which every bidding participant is considered as an agent who has the ability of adaptive learning so as to improve its profit during process of repeated bidding in market. Table-based reinforcement learning (TBRL) algorithms are usually proposed in depicting agents' adaptive learning approaches in EM bidding, such as the Q-learning-based approach proposed in [26,27], simulated annealing Q-learning-based approach proposed in [28], Roth-Erev reinforcement learning-based EM test bed (called Multi-Agent Simulator of Competitive Electricity Markets or MASCEM) proposed in [29,36], SARSA (state-action-reward-state-action)-based approach proposed in [30], fuzzy Q-learning-based approach proposed in [31], etc. By using agent-based EM modeling approaches, it is neither necessary to repeatedly solve multi-level mathematical programming models for every agent, nor to establish sets of nonlinear equations which are difficult to solve or have no solutions. Low computational complexity and low reliance on common knowledge make these approaches more applicable in EM modeling [26]. However, in TBRL algorithms, both an agent's action (i.e., bidding strategy) and state (i.e., market environment) sets must be assumed as discrete, otherwise it will cause the problem of “curse of dimensionality”, which does not conform to the actual situation of the day-ahead EM and hinders a strategic WPP to obtain its globally optimal bidding strategies no matter which BM it adopts. So far as we know, there is no reasonable way to solve this issue in the published literature studying wind power and other renewable energy bidding in EMs. Recently, Chen [40] proposed a modified reinforcement learning (RL) algorithm called Least Squares Continuous Actor-Critic (LSCAC) algorithm which can make both the action and state sets continuous without causing the problem of “curse of dimensionality”. Therefore, another motivation of this work was to apply for the first time the LSCAC algorithm for modeling the strategic bidding behaviors of WPPs in day-ahead EMs. On the one hand, this approach properly solved the contradiction of making every agent's action and state sets continuous and causing the “curse of dimensionality” problem. On the other hand, it can provide a reasonable EM test bed to simulate and experimentally compare those two BMs adopted by WPPs in a day-ahead EM.

Therefore, the main novelty of this paper can be summarized as to firstly propose the LSCAC-based day-ahead EM modeling approach for strategic WPPs under robust market clearing conditions. The purpose for employing robust MCM is to improve wind power accommodation by

reconstructing the market clearing mechanism. The motivation of proposing the LSCAC-based EM modeling approach is to assist strategic WPPs to make more appropriate bidding decisions so as to improve both the WPPs' profits and the economic efficiency of the whole market compared with TBRL-based approaches. Moreover, comparison between different BMs can offer some suggestions about improving wind power resources development and market economic efficiency.

The rest of this paper is organized as follows: in Section 2, the concrete mathematic formulations of WPPs' different BMs and the robust day-head MCM are proposed. Section 3 puts forward the proposed LSCAC-based day-ahead EM modeling approach for WPPs. Section 4 conducts the simulations and comparisons. Section 5 concludes the paper.

2. Problem Description

2.1. Model Assumptions

According to [41], the day-ahead EM is actually a dynamic complex system in which dynamic (direct and indirect) interactions exist among all participants. When considering WPPs' strategic behaviors in the day-ahead EM, on one hand, the MCM of the ISO should be modified in order to accommodate the deviations caused by real time wind power output uncertainties [10,19,20], while on the other hand, EM modeling approaches should be proposed to help with obtaining WPPs' reasonable strategies under different BMs.

In this section, strategic WPPs' different BMs and the robust day-head MCM are mathematically formulated. For the sake of simplicity and without loss of generality, we make some assumptions listed as follows before conducting any further research:

- Like [9], in our study, the problem of SCUC is assumed to have been solved exogenously in advance, and consequently, the UC constraints (i.e., ramping rates, startup costs/times, minimum down-times) are not considered. However, the proposed single period EM modeling approach containing single period robust MCM can be extended to a multi-period one. Moreover, network loss is ignored, and the shift factor matrix is constant.
- Because we mainly consider the strategic behaviors of WPPs in a day-ahead EM, the bidding strategy of any other conventional generators is neglected [9,32,33]. We also assume that load in any bus is inelastic without load shedding [9].
- Uncertainties are only caused by WPPs, The uncertainty-set can be truly formulated by the ISO [10,19,20]. The marginal cost of every WPP is neglected [3,31]. Hence, when a WPP is in BM 2, there is only one bidding price in this WPP's bidding curve.

2.2. Different BMs of WPP

Real-time wind power output cannot be accurately predicted from the day-ahead horizon, which forces WPPs and the ISO to carefully consider these strong uncertainties when bidding and implementing the day-ahead market clearing, respectively. However, the ISO (WPP) can predict the real time wind power output interval of a certain WPP more accurately than the real time wind power output prediction. If the number of WPPs in a power system is N_W , consistent with [10,19,20], the uncertainty set corresponding to those N_W WPPs' real time power outputs can be modeled as:

$$\mathbf{U} = \left\{ \mathbf{Pw} = (Pw_1, Pw_2, \dots, Pw_i, \dots, Pw_{N_W}) : \right. \\ \left. lw_i \leq Pw_i \leq uw_i, \forall i, \sum_{i=1}^{N_W} \frac{|2Pw_i - (uw_i + lw_i)|}{(uw_i - lw_i)} \leq \Lambda \right\} \quad (1)$$

where, i is the index for WPP, Pw_i represents the actual power output of the i -th WPP, lw_i , uw_i ($0 \leq lw_i \leq uw_i$) represent the lower and upper bounds of Pw_i , Λ is the budget parameter and assumed as an integer [20]. Moreover, accurate prediction of lw_i and uw_i can provide a valuable reference for WPPs' bidding decision-making. Therefore:

- When BM 1 is adopted by WPP_i, the only bidding parameter is its power output plan (bidding power output) Pw_i^b which must satisfy Equation (2):

$$lw_i \leq Pw_i^b \leq uw_i, \forall i \quad (2)$$

Under this BM, the bidding strategy of WPP_i is to adjust the value of Pw_i^b .

- When BM 2 is adopted by WPP_i, it provides to ISO the bidding curve which is as follows:

$$\rho(Pw_i^b) = \rho_i^b, lw_i \leq Pw_i^b \leq uw_i, \rho_{low,i} \leq \rho_i^b \leq \rho_{upp,i} \quad (3)$$

where, ρ_i^b is WPP_i's bidding price. $\rho_{low,i}, \rho_{upp,i}$ ($0 \leq \rho_{low,i} \leq \rho_{upp,i}$) represent the lower and upper limits of ρ_i^b . Under this BM, the bidding strategy of WPP_i is to adjust the value of ρ_i^b .

2.3. Robust Market Clearing Models under Different BMs of WPPs

According to Section 2.1, because the problem of SCUC is assumed to have been solved exogenously in advance, we propose a single period day-head robust MCM which is mainly focused on an ED procedure. The purpose of doing so is to make the power system accommodate any wind power deviation caused by real time wind power output uncertainty within a certain uncertainty set. With this robust MCM, the ISO, based on the day-ahead biddings (curves) of all participants, desires to get the optimal robust ED solution in the base-case scenario [19]. Under the optimal robust ED solution, the ISO can re-dispatch the flexible resources, such as adjustable conventional generators with fast ramping capabilities, etc., to follow the load when a deviation occurs. The method of obtaining an optimal robust ED solution in the base-case scenario is significantly less conservative than that in the worst-case scenario [16]. Moreover, this robust MCM can reasonably generate prices for power outputs, loads, reserves and deviations which are the byproducts of the optimal robust ED solution [10,20].

The mathematical formulation of this robust MCM can be described as follows:

$$(\text{Problem}) \quad \min_{P_j, Pw_i^{disp}, \forall j,i} \sum_{j=1}^{N_G} c_j P_j + \sum_{i=1}^{N_W} \rho_i^b Pw_i^{disp} \quad (4)$$

$$s.t. \sum_{j=1}^{N_G} P_j + \sum_{i=1}^{N_W} Pw_i^{disp} = \sum_{m=1}^{N_{bus}} d_m \quad (5)$$

$$P_j \leq P_j^{\max}, \forall j \quad (6)$$

$$-P_j \leq -P_j^{\min}, \forall j \quad (7)$$

$$\sum_{m=1}^{N_{bus}} \Gamma_{l,m} \sum_{i,j \in \Theta(m)} (P_j + Pw_i^{disp} - d_m) \leq F_l, \forall l \quad (8)$$

$$- \sum_{m=1}^{N_{bus}} \Gamma_{l,m} \sum_{i,j \in \Theta(m)} (P_j + Pw_i^{disp} - d_m) \leq F_l, \forall l \quad (9)$$

$$\text{and } \Omega = \left\{ P_{GW} = (P_1, P_2, \dots, P_j, \dots, P_{N_G}, Pw_1^{disp}, Pw_2^{disp}, \dots, Pw_i^{disp}, \dots, Pw_{N_W}^{disp}) : \right. \\ \left. \forall Pw \in \Omega, \exists \Delta P = (\Delta P_1, \Delta P_2, \dots, \Delta P_j, \dots, \Delta P_{N_G}) \text{ such that} \right. \quad (10) \\ \left. \sum_{j=1}^{N_G} \Delta P_j + \sum_{i=1}^{N_W} (Pw_i - Pw_i^{disp}) = 0 \right.$$

$$P_j + \Delta P_j \leq P_j^{\max}, \forall j \quad (11)$$

$$-P_j - \Delta P_j \leq -P_j^{\min}, \forall j \quad (12)$$

$$\Delta P_j \leq r_j^u, \forall j \quad (13)$$

$$-\Delta P_j \leq r_j^d, \forall j \quad (14)$$

$$\sum_{m=1}^{N_{bus}} \Gamma_{l,m} \sum_{i,j \in \Theta(m)} (P_j + \Delta P_j + Pw_i - d_m) \leq F_l, \forall l \quad (15)$$

$$-\sum_{m=1}^{N_{bus}} \Gamma_{l,m} \sum_{i,j \in \Theta(m)} (P_j + \Delta P_j + Pw_i - d_m) \leq F_l, \forall l \quad (16)$$

where, j and N_G represent the index and number of conventional generators that are determined as being the state of start-up in advance, respectively. Because it is assumed that the UC solution is fixed exogenously, N_G can also be considered as the number of conventional generators in power system for simplicity. P_j is the dispatched power output for the j -th generator, c_j is the cost coefficient of the j -th generator, m and N_{bus} represent the index and number of buses in power system. In the basic-case scenario, Equation (5) indicates power balance of the system, Equations (6) and (7) show the power limits of generators, Equations (8) and (9) stand for the transmission constraints of all lines in system. In case of wind power deviations, Equation (10) indicates power balance of the system in re-dispatch, Equations (11) and (12) show the power limits of generators in re-dispatch, Equations (13) and (14) are constraints for power re-dispatch variables ΔP_j s, Equations (15) and (16) stand for the transmission constraints of all lines in system in re-dispatch.

- If BM 1 is adopted by WPPs, then:

$$\rho_i^b = 0, \forall i \quad (17)$$

$$Pw_i^{disp} = Pw_i^b, \forall i \quad (18)$$

{Equations (4)–(9), (17), (18)} $\cup \Omega$ constitute the day-ahead robust MCM for ISO under BM 1 (namely, RMCM 1).

- If BM 2 is adopted by WPPs, ISO dispatches WPPs' day-ahead power output schedules which should at least satisfy [9]:

$$Pw_i^{disp} \leq uw_i, \forall i \quad (19)$$

$$-Pw_i^{disp} \leq lw_i, \forall i \quad (20)$$

{Equations (4)–(9), (19), (20)} $\cup \Omega$ constitute the day-ahead robust MCM for ISO under BM 2 (namely, RMCM 2).

2.4. Robust MCM Reformulation

By solving RMCM 1 or RMCM 2, the obtained optimal robust ED solution $P_{GW} \in \Omega$ is immunized against any uncertainty $\forall Pw \in U$ [19]. When uncertainty Pw occurs, deviations caused by Pw can be accommodated by the power re-dispatch ΔP . However, it should be noted that both RMCM 1 and RMCM 2 cannot be directly solved. Similar to [19,20], reformulation is adopted to solve the two RMCMs. In order to facilitate the description, reformulation of {Equations (4)–(9)} $\cup \Omega$ which contains a master problem (MP) and a sub-problem (SP) will first be established as follows:

$$(MP) \min_{P_j, Pw_i^{disp}, \forall j, i} \sum_{j=1}^{N_G} c_j P_j + \sum_{i=1}^{N_W} \rho_i^b Pw_i^{disp} \quad (21)$$

$$s.t. (5)–(9)$$

$$\sum_{j=1}^{N_G} \Delta P_{jk} + \sum_{i=1}^{N_W} (Pw_{ik} - Pw_i^{disp}) = 0, \forall k \in \kappa \quad (22)$$

$$P_j + \Delta P_{jk} \leq p_j^{\max}, \forall j, \forall k \in \kappa \quad (23)$$

$$-P_j - \Delta P_{jk} \leq -p_j^{\min}, \forall j, \forall k \in \kappa \quad (24)$$

$$\Delta P_{jk} \leq r_j^u, \forall j, \forall k \in \kappa \quad (25)$$

$$-\Delta P_{jk} \leq r_j^d, \forall j, \forall k \in \kappa \quad (26)$$

$$\sum_{m=1}^{N_{bus}} \Gamma_{l,m} \sum_{i,j \in \Theta(m)} (P_j + \Delta P_{jk} + Pw_{ik} - d_m) \leq F_l, \forall l, \forall k \in \kappa \quad (27)$$

$$-\sum_{m=1}^{N_{bus}} \Gamma_{l,m} \sum_{i,j \in \Theta(m)} (P_j + \Delta P_{jk} + Pw_{ik} - d_m) \leq F_l, \forall l, \forall k \in \kappa \quad (28)$$

and:

$$(\text{SP})\Psi = \max_{Pw \in \mathbf{U}} \min_{s^+, s^-, \Delta P} \sum_{i=1}^{N_W} (s_i^+ + s_i^-) \quad (29)$$

$$\text{s.t.} \sum_{j=1}^{N_G} \Delta P_j + \sum_{i=1}^{N_W} (Pw_i - Pw_i^{\text{disp}} + s_i^+ - s_i^-) = 0 \quad (30)$$

$$-s_i^+ \leq 0, \quad \forall i \quad (31)$$

$$-s_i^- \leq 0, \quad \forall i \quad (32)$$

$$\sum_{m=1}^{N_{bus}} \Gamma_{l,m} \sum_{i,j \in \Theta(m)} (P_j + \Delta P_j + Pw_i^{\text{disp}} + Pw_i - Pw_i^{\text{disp}} + s_i^+ - s_i^- - d_m) \leq F_l, \forall l \quad (33)$$

$$-\sum_{m=1}^{N_{bus}} \Gamma_{l,m} \sum_{i,j \in \Theta(m)} (P_j + \Delta P_j + Pw_i^{\text{disp}} + Pw_i - Pw_i^{\text{disp}} + s_i^+ - s_i^- - d_m) \leq F_l, \forall l \quad (11)-(14) \quad (34)$$

where, κ is the index set for worst uncertainty points Pw_k s which are dynamically generated in (SP) during the solution procedure.

If BM 1 is adopted by WPPs, Equations (17) and (18) should be added to (MP). If BM 2 is adopted by WPPs, Equations (19) and (20) should be added to (MP). According to References [19,20], the objective function in (SP) contains the summation of non-negative slack variables s_i^+ s and s_i^- s, which evaluates the violation associated with the solution from (MP). s_i^+ s and s_i^- s can be explained as un-followed uncertainties (i.e., generation shedding etc.) due to system limitations. Hence, to solve (SP) is to find the worst point Pw_k in \mathbf{U} given ED solutions. The solution procedure is [19]:

- (1) $\kappa \leftarrow \phi, k \leftarrow 1, \Psi \leftarrow +\infty$, define feasibility tolerance Δ ;
- (2) **while** $\Psi \geq \Delta$ **do**
- (3) Solve (MP), obtain optimal $P_j, Pw_i^{\text{disp}}, \forall j, i$;
- (4) Solve (SP) with $P_j, Pw_i^{\text{disp}}, \forall j, i$, get solution (Ψ, Pw_k) ;
- (5) $\kappa \leftarrow \kappa \cup k, k \leftarrow k + 1$;
- (6) **end while**.

2.5. Clearing Price Mechanism

After the convergence of the abovementioned solution procedure, the optimal robust ED solution can be obtained by solving (MP) for the last time. We set $\lambda, \beta_{j1}(\forall j), \beta_{j2}(\forall j), \eta_{l1}(\forall l), \eta_{l2}(\forall l)$ to represent the generalized Lagrange multipliers (GLMs) for Equations (5)–(9), respectively, $\mu_{i1}(\forall i), \mu_{i2}(\forall i)$ to represent the GLMs for Equations (19) and (20), respectively, and $\lambda_k, \beta_{j1k}(\forall j, \forall k \in \kappa), \beta_{j2k}(\forall j, \forall k \in \kappa), \zeta_{j1k}(\forall j, \forall k \in \kappa), \zeta_{j2k}(\forall j, \forall k \in \kappa), \eta_{l1k}(\forall l, \forall k \in \kappa), \eta_{l2k}(\forall l, \forall k \in \kappa)$ to represent the GLMs for

Equations (22)–(28), respectively. Consistent with the cost causation principle and LMP calculation method mentioned in [10,20], when solving (MP) for the last time, the clearing price mechanism of our proposed RMCMs can be described as follows:

- No matter which BM WPPs adopt, the day-ahead LMP for energy credit and load payment at bus m can be calculated as:

$$\begin{aligned}\pi_m &= \frac{\partial L(P_{GW}, \Delta P, \lambda, \beta, \zeta, \eta)}{\partial d_m} = \frac{\partial L(P_{GW}, \Delta P, \lambda, \beta, \mu, \zeta, \eta)}{\partial d_m} \\ &= \lambda - \sum_l \Gamma_{l,m}(\eta_{l1} - \eta_{l2}) - \sum_l \sum_{k \in \kappa} \Gamma_{l,m}(\eta_{l1k} - \eta_{l2k})\end{aligned}\quad (35)$$

where, $L(P_{GW}, \Delta P, \lambda, \beta, \zeta, \eta)$ is the generalized lagrange function (GMF) for (MP) under BM 1 adopted by WPPs, $L(P_{GW}, \Delta P, \lambda, \beta, \mu, \zeta, \eta)$ is the GMF for (MP) under BM 2 adopted by WPPs.

- Defining uncertainty marginal price (UMP) as [20]: the marginal cost of immunizing the next unit increment of uncertainty, then no matter which BM WPPs adopt, for the deviation $Pw_k - Pw^{disp}$ corresponding to a worst point Pw_k , the UMP for reserve credit and deviation payment at bus m is:

$$\begin{aligned}\pi_{mk} &= \frac{\partial L(P_{GW}, \Delta P, \lambda, \beta, \zeta, \eta)}{\partial \varepsilon_{mk}} = \frac{\partial L(P_{GW}, \Delta P, \lambda, \beta, \mu, \zeta, \eta)}{\partial \varepsilon_{mk}} \\ &= \lambda_k - \sum_l \Gamma_{l,m}(\eta_{l1k} - \eta_{l2k})\end{aligned}\quad (36)$$

where, $\varepsilon_{mk} = \sum_{i \in \Theta(m)} (Pw_i^{disp} - Pw_{ik})$. Moreover, when $\pi_{mk} > 0$, it can be illustrated that the direction of power re-dispatch corresponding to worst point Pw_k at bus m is upward ($\Delta P_{jk} \geq 0, \forall j \in \Theta(m)$), and when $\pi_{mk} < 0$, it can be illustrated that direction of power re-dispatch corresponding to worst point Pw_k at bus m is downward ($\Delta P_{jk} \leq 0, \forall j \in \Theta(m)$).

Moreover, the structural differences between RMCM 1 and RMCM 2 may make the primal solutions $(P_{GW}, Pw^{disp}, \Delta P)$ and the dual solutions $(\lambda, \beta, (\mu), \alpha, \eta)$ obtained by solving RMCM 1 and RMCM 2 differ from each other. Therefore, although the clearing price formulas under 2 BMs have no difference according to Equations (35) and (36), different primal-dual solutions obtained by solving RMCM 1 and RMCM 2 still make the obtained clearing prices (π_m, π_{mk}) under 2 BMs different.

In summary, no matter which BM is adopted by WPPs, the estimated profit of WPP _{i} ($\forall i$) in one day-ahead bidding can be calculated as:

$$R_i = \pi_m Pw_i^{disp} - \sum_{k \in \kappa} \pi_{mk} (Pw_i^{disp} - Pw_{ik}), i \in \Theta(m) \quad (37)$$

Hence, the objective of WPP _{i} ($\forall i$) bidding in day-ahead EM is to maximize R_i .

3. LSCAC-Based Day-Ahead EM Modeling Approach for WPPs

3.1. Definitions

Although the BM and MCM in a day-ahead EM can be specified by relevant regulators in advance, a strategic WPP in EM still has limited information about other rivals. Owing to this fact of incomplete and imperfect information in the day-ahead EM [41], strategic WPPs must dynamically improve their profits through repeatedly bidding in day-ahead EM, which is actually a dynamic multi-participant decision-making process. This work intends to propose a LSCAC-based day-head EM modeling approach to simulate this dynamic multi-WPP decision-making process. Hence, similar to Reference [41], some necessary definitions are organized as follows:

- Agent: we consider every WPP as an agent who, for the purpose of improving its profit, has the adaptive learning ability to dynamically adjust its bidding strategy according to its accumulated experiences through repeated bidding. Hence, the multi-WPP decision-making process can be

also considered as multi-agent decision-making process. In our work, LSCAC algorithm is applied to depict this adaptive learning ability and to assist every WPP in bidding decision making.

- Iteration: since the market is assumed to be cleared in day-ahead single period basis, we consider each transaction day as an iteration T .
- State variable: in iteration T , the LMP and UMPs in bus m cleared in iteration $T - 1$ are considered as the market environment states for WPP_{*i*} connected in bus m ($i \in \Theta(m)$), which is because WPP_{*i*} ($i \in \Theta(m)$) actually has no idea about other market information. Taking $x_{i,T}$ to represent the state variable vector for WPP_{*i*} in iteration T , the relationship between $x_{i,T}$ and clearing prices is as follows [20]:

$$x_{i,T} = (\pi_{m,T-1}, \pi_{m,T-1}^{up}, \pi_{m,T-1}^{down}), i \in \Theta(m) \quad (38)$$

$$\pi_{m,T-1}^{up} = \sum_{k \in \kappa_{m,T-1}^{up}} \pi_{mk,T-1}; \pi_{m,T-1}^{down} = \sum_{k \in \kappa_{m,T-1}^{down}} \pi_{mk,T-1} \quad (39)$$

$$\kappa_{m,T-1}^{up} = \{k : \pi_{mk,T-1} \geq 0\}; \kappa_{m,T-1}^{down} = \{k : \pi_{mk,T-1} < 0\} \quad (40)$$

where, $\pi_{m,T-1}$ and $\pi_{mk,T-1}$ represent the LMP and the k -th UMP in bus m cleared in iteration $T - 1$, respectively.

- Action variable: in iteration T , the bidding strategy of WPP_{*i*} is considered as its action. Taking $a_{i,T}$ to represent the action variable for WPP_{*i*}. If WPP_{*i*} bids in EM under BM 1, the relationship between $a_{i,T}$ and bidding strategy is as follow:

$$a_{i,T} = Pw_{i,T}^b \quad (41)$$

where, $Pw_{i,T}^b$ is the bidding output (strategy) of WPP_{*i*} in iteration T . If WPP_{*i*} bids in EM under BM 2, the relationship between $a_{i,T}$ and bidding strategy is as follows:

$$a_{i,T} = \rho_{i,T}^b \quad (42)$$

where, $\rho_{i,T}^b$ is the bidding price (strategy) of WPP_{*i*} in iteration T .

- Reward: in iteration T , WPP_{*i*}'s reward is:

$$r_{i,T} = R_{i,T} \quad (43)$$

where, $R_{i,T}$ is WPP_{*i*}'s estimated profit obtained from bidding in iteration T .

3.2. LSCAC Algorithm

In TBRL-based EM modeling approaches [26–31,36], both the state and action sets should be assumed as discrete, otherwise the problem of “curse of dimensionality” will be caused so as to significantly hinder agents from improving their profits. However, according to Section 3.1 and [41], in day-ahead EM with many strategic WPPs, both $x_{i,T}(\forall i)$ and $a_{i,T}(\forall i)$ are within the continuous, bounded and closed sets (spaces). Therefore, a modified RL algorithm must be applied in day-ahead EM modeling for the study of strategic behaviors of WPPs.

In our work, we apply the LSCAC algorithm to this issue for the first time. The LSCAC algorithm is a modified actor-critic based RL algorithm which can rapidly tackle the dynamic multi-agent decision-making problem with continuous action and state sets. In the LSCAC algorithm, state value function and policy function of every agent are approximated by using linear combinations of basis functions. Linear parameters in state value functions corresponding to agents' critic parts are updated online by using the temporal difference error (TD(0))-based method, the specific procedure of which can be found in [41]. The online updating procedure of linear parameters in policy functions corresponding to agents' actor parts is described as follows [40]: by using a linear function, we estimate

and repeatedly update in an agent's actor part an optimal policy function $\hat{I} : X \rightarrow A$ defined on the continuous state space X :

$$a_x^{(optimal)} = \hat{I}(x) = \phi(x)^T \omega = \sum_{h=1}^n \phi_h(x) \omega_h \quad x \in X \quad (44)$$

where, $\phi_h : X \rightarrow R(h = 1, 2, \dots, n)$ represents the h -th basis function of state $x \in X$. A represents the continuous action set of an agent, $a_x^{(optimal)} \in A$ represents the optimal action in face of state x . The linear parameter vector ω can be described as: $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T \in R^n$.

An agent must generate a corresponding action $a \in A$ in face of any state $x \in X$ based on the policy maintained and repeatedly updated by its actor part. The policy is actually an action generating model which has the ability of balancing the exploration and exploitation, and can be mathematically formulated as follows [40,41]:

$$pro(x, a) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(a - \phi(x)^T \omega)^2\right\} \quad (45)$$

where, $\sigma > 0$ is a standard deviation parameter which represents the exploring ability of the LSCAC algorithm.

Hence, the MSE function of ω is defined as [40,41]:

$$MSE(\omega) = \frac{1}{2} \int_{x \in X} P^{(pro)}(x) \int_{a \in A} sig[\delta(x, a)] [\phi(x)^T \omega - a]^2 da dx \quad (46)$$

where, $P^{(pro)}(x)$ is the probability distribution function of x under policy pro , $sig[\delta(x, a)]$ is the sigmoid function of $\delta(x, a)$ which means the TD(0) error of selecting action a in face of state x . Its formulation is as follows [40,41]:

$$sig[\delta(x, a)] = \frac{1}{1 + e^{-q\delta(x, a)}} \quad q > 0 \quad (47)$$

In iteration T , using δ_T to replace $\delta(x_T, a_T)$, formulation of δ_T is as follows [40,41]:

$$\delta_T = r_T + \gamma \phi(x_{T+1})^T \theta_T - \phi(x_T)^T \theta_T \quad (48)$$

where, linear vector θ_T is composed of linear parameters in value functions in iteration T [41], $0 \leq \gamma \leq 1$ is a discount factor.

Let the derivative of Equation (46) on ω equal to 0, then:

$$\int_{x \in X} P^{(pro)}(x) \int_{a \in A} \frac{1}{1 + e^{-q\delta(x, a)}} [\phi(x)^T \omega - a] \phi(x) da dx = 0 \quad (49)$$

It should be noted that the integral formula in the left side of Equation (49) is hard to calculate. If the sample points from iteration 0 to iteration N are $(x_0, a_0, r_0, x_1), (x_1, a_1, r_1, x_2), \dots, (x_N, a_N, r_N, x_{N+1})$, Equation (46) can be approximately replaced by:

$$\sum_{T=0}^N \frac{1}{1 + e^{-q\delta_T}} (\phi(x_T)^T \omega - a_T) \phi(x_T) = 0 \quad (50)$$

The reformulation of Equation (50) is:

$$\left[\sum_{T=0}^N \frac{1}{1 + e^{-q\delta_T}} \phi(x_T) \phi(x_T)^T \right] \omega = \sum_{T=1}^N \frac{1}{1 + e^{-q\delta_T}} a_T \phi(x_T) \quad (51)$$

Define a n -order matrix \mathbf{A}_N and a n -dimensional vector \mathbf{b}_N , respectively:

$$\mathbf{A}_N = \sum_{T=0}^N \frac{1}{1 + e^{-q\delta_T}} \boldsymbol{\phi}(x_T) \boldsymbol{\phi}(x_T)^T \quad (52)$$

$$\mathbf{b}_N = \sum_{T=1}^N \frac{1}{1 + e^{-q\delta_T}} a_T \boldsymbol{\phi}(x_T) \quad (53)$$

hence:

$$\boldsymbol{\omega} = (\mathbf{A}_N)^{-1} \mathbf{b}_N \quad (54)$$

Because $(\mathbf{A}_N)^{-1}$ may not exist, $(\mathbf{A}_N)^{-1}$ can be approximately replaced by $(\mathbf{A}_N + \Pi \mathbf{I})^{-1}$ based on the method of ridge regression [40], where Π ($\Pi > 0$) is a smaller constant, \mathbf{I} is an n -order identity matrix. When N is large, the calculation of parameter vector $\boldsymbol{\omega}$ in Equation (54) may be unstable. Reference [40] has proposed a new calculation formula for parameter vector $\boldsymbol{\omega}$, which is as follows:

$$\boldsymbol{\omega} = \left[\frac{1}{N} (\mathbf{A}_N + \Pi \mathbf{I}) \right]^{-1} \left(\frac{1}{N} \mathbf{b}_N \right) \quad (55)$$

3.3. The Step-by-Step Procedure of the Proposed Approach

In summary, the step-by-step procedure of LSCAC-based day-ahead EM modeling approach for WPPs (under two BMs) can be described as follows:

- (1) **Input:** basis function vector $\boldsymbol{\phi}: X \rightarrow R^n$, step length factor series $\{\alpha_{i,T}\}_{T=0}^{\infty}$ where $\sum_{T=0}^{\infty} \alpha_{i,T} = \infty$, $\sum_{T=0}^{\infty} (\alpha_{i,T})^2 < \infty$, and values of σ, q, γ (for $\text{WPP}_i(\forall i)$).
- (2) $T = 0, N = 0$, set the iterative termination condition such as the maximum iterations (T_{\max}).
- (3) set $\boldsymbol{\theta}_{i,T} = \boldsymbol{\theta}_{i,0}$ and $\boldsymbol{\omega}_{i,T} = \boldsymbol{\omega}_{i,0}$, $\mathbf{A}_{i,N} = \mathbf{A}_{i,0} = \Pi_i \mathbf{I}$, $\Pi_i > 0$ and $\mathbf{b}_{i,N} = \mathbf{b}_{i,0} = \mathbf{0}$ (for $\text{WPP}_i(\forall i)$).
- (4) set $x_{i,T} = x_{i,0}$ (for $\text{WPP}_i(\forall i)$).
- (5) In iteration T , generate $a_{i,T} \sim N(\boldsymbol{\phi}(x_{i,T})^T \boldsymbol{\omega}_{i,T}, \sigma^2)$ where $a_{i,T}$ represents $Pw_{i,T}^b$ or $\rho_{i,T}^b$ (for $\text{WPP}_i(\forall i)$) and then ISO implements the robust MCM represented by RMCM 1 or RMCM 2.
- (6) After market clearing, $\text{WPP}_i(\forall i)$ obtains the immediate reward $r_{i,T}$ using Equation (37) and a new market state $x_{i,T+1}$ which can be generated by the ISO using Equations (35), (36), (38)–(40).
- (7) $\delta_{i,T} = r_{i,T} + \gamma \boldsymbol{\phi}(x_{i,T+1})^T \boldsymbol{\theta}_{i,T} - \boldsymbol{\phi}(x_{i,T})^T \boldsymbol{\theta}_{i,T}, \forall i$.
- (8) $\boldsymbol{\theta}_{i,T+1} = \boldsymbol{\theta}_{i,T} + \alpha_{i,T} \delta_{i,T} \boldsymbol{\phi}(x_{i,T}), \forall i$.
- (9) $\mathbf{A}_{i,N+1} = \mathbf{A}_{i,N} + \frac{1}{1 + e^{-q\delta_{i,T}}} \boldsymbol{\phi}(x_{i,T}) \boldsymbol{\phi}(x_{i,T})^T, \forall i$.
- (10) $\mathbf{b}_{i,N+1} = \mathbf{b}_{i,N} + \frac{1}{1 + e^{-q\delta_{i,T}}} a_{i,T} \boldsymbol{\phi}(x_{i,T}), \forall i$.
- (11) $\boldsymbol{\omega}_{i,T+1} = \left[\frac{1}{N+1} (\mathbf{A}_{i,N+1}) \right]^{-1} \left(\frac{1}{N+1} \mathbf{b}_{i,N+1} \right), \forall i$.
- (12) $T = T + 1, N = N + 1$.
- (13) Checking the iterative termination condition, if our procedure achieved the iterative termination condition, go on to step (14), otherwise, return to step (5).
- (14) **Output:** $\boldsymbol{\theta}_i^* = \boldsymbol{\theta}_{i,T}, \boldsymbol{\omega}_i^* = \boldsymbol{\omega}_{i,T}$ (for $\text{WPP}_i(\forall i)$), based on which $\text{WPP}_i(\forall i)$ can select the optimal bidding strategy (under BM 1 or 2) in face of whichever market state is.

Moreover, According to [40], we choose a Gaussian radial basis function as $\boldsymbol{\phi}(x)$.

4. Simulations and Discussions

4.1. System Data

In this section, by implementing the robust MCM mentioned in Section 2, our proposed day-ahead EM modeling approaches under different BMs are simulated on the IEEE 30-bus test system with five strategic WPPs [9]. Matlab R2014a is utilized to conduct our simulations. Figure 1 shows the schematic structure of the test system. Table 1 depicts the predicted single period loads distributed in different buses [42]. Consistent with the assumptions in Section 2, any uncertainties in this test system are not caused by loads. Parameters of conventional generators can be seen in Table 2, and the predicted power output intervals of the five WPPs which are the crucial components of the uncertainty set, are listed in Table 3. For the sake of simplicity and without losing generality, we assume the power output interval of $WPP_i(\forall i)$ predicted by ISO is the same as that predicted by $WPP_i(\forall i)$ itself.

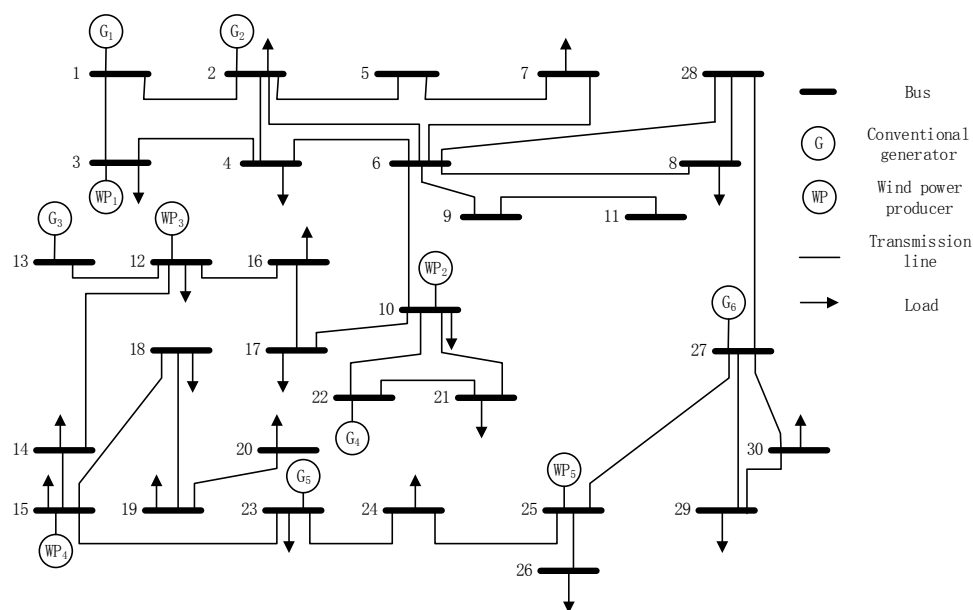


Figure 1. Diagram of the test system (Note: For the sake of simplicity, here it is assumed that the maximum congestion constraints in all transmission lines are 30 MW).

Table 1. Values of un-elastic single period loads.

Bus Number	2	3	4	7	8
Load (MW)	26.7	7.4	12.6	27.8	35
Bus	10	12	14	15	16
Load (MW)	10.8	16.2	11.2	13.2	8.5
Bus number	17	18	19	20	21
Load (MW)	14	8.2	14.5	7.2	22.5
Bus	23	24	26	29	30
Load (MW)	8.2	13.7	8.5	7.4	15.6

Table 2. Parameters of conventional generators.

Bus	Generators	c_j (10^3 \$/MWh)	P^{min} (MW)	P^{max} (MW)	$ramp^{up,max}$ (MW)	$ramp^{down,max}$ (MW)
1	G ₁	36	0	80	24	24
2	G ₂	31.5	0	80	24	24
13	G ₃	41.25	0	50	12	12
22	G ₄	37.087	0	55	12	12
23	G ₅	37.5	0	40	10	10
27	G ₆	40	0	40	10	10

Note: Because the bidding strategies of conventional generators are neglected, the value of c_j ($\forall j$) in Table 2 is obtained by using marginal cost of the j -th conventional generator when its power output reaches P_j^{max} . Moreover, parameters of every conventional generator's marginal cost can be seen in Reference [43].

Table 3. Predicted power output intervals of WPPs.

Bus	WPP	Lw (MW)	Uw (MW)
3	WPP ₁	5	20
10	WPP ₂	10	25
12	WPP ₃	10	25
15	WPP ₄	20	30
25	WPP ₅	5	20

4.2. Robust MCM Testing

The value of budget parameter Λ is related to the size of the uncertainty set. The smaller Λ is, the smaller is the size of uncertainty set estimated by ISO. That is to say the day-ahead market clearing procedure of ISO tends to be more deterministic with the decrease of the value of Λ ($\Lambda \geq 0$). When $\Lambda = 0$, it means the predicted power output of every WPP is deemed by ISO as a definite value which, according to Equation (1) and [10,19,20], is equal to the intermediate value of its power output interval, and the day-ahead MCM of ISO is completely turned into a conventional deterministic MCM similar to [30]. However, uncertainties exist objectively in a power system with WPPs. If the market clearing procedure were implemented by ISO without considering enough uncertainties, the reserve capacity of the system dispatched in day-ahead might find it hard to accommodate deviations caused by WPPs' power output uncertainties in real time, which can seriously affect the security of the system and cause huge extra costs such as wind-abandonment, etc. Therefore, market clearing results under different Λ values must be compared so as to verify the necessity of proposing robust MCM in day-ahead EM with wind power penetrations.

In this section, in order to facilitate the market clearing comparison, we assume every WPP is under BM 1 and sends the ISO the intermediate value of its power output interval. In fact, the same key conclusions obtained from market clearing comparisons can also be generated with other BM and strategies. Moreover, no matter what the ISO thinks the value of Λ is, the actual value of Λ which represents the objective existence of uncertainty is fixed, by us, to the number of WPPs in the system ($\Lambda = 5$). Table 4 shows the market clearing results under different Λ values.

Table 4. Market clearing results under different Λ values.

Λ Value	Operation Cost (10^3 \$)	Uncertainty That Cannot Be Accommodated	Number of Uncertainty Poles That Cannot Be Accommodated
0	5.8955	yes	22
1	6.0087	yes	14
2	6.1023	yes	9
3	6.1851	yes	3
4	6.2078	no	0
5	6.2078	no	0

In Table 4, because the marginal cost of every WPP is neglected [3,31], the “operation cost” in column 2 can be calculated by using $\sum_{j=1}^{N_G} c_j P_j$ when the optimal ED solution is obtained. Moreover, the “uncertainty that cannot be accommodated” in column 3 means whether there exist uncertainty points in $U(\Lambda = 5)$ that cannot be accommodated when the optimal ED solution is obtained. The “number of uncertainty poles that cannot be accommodated” in column 4 denotes the number of poles in $U(\Lambda = 5)$ that cannot be accommodated when the optimal ED solution is obtained. From Table 4, it can be concluded that:

- The total operation cost increases with the increase of Λ . However, uncertainties that cannot be accommodated tend to be eliminated by increasing Λ . On one hand, it means the conservatism of ISO is improved with the increase of Λ , which reduces the economic efficiency of scheduling to a certain extent. On the other hand, the operation cost is calculated based on the basic scenario (WPPs’ day-ahead bids), in which extra cost caused by uncertainties that cannot be accommodated is not taken into account. Although that extra cost is hard to be specifically calculated due to many reasons such as missing information about the real-time occurrence of uncertainty from day-ahead horizon etc., it can be considerable once any uncertainty that cannot be accommodated occurs in practice. Therefore, it is necessary to eliminate uncertainties that cannot be accommodated by reasonably increasing the value of Λ .
- When $\Lambda = 0$, it means the ISO clears the market using the conventional deterministic MCM. The number of uncertainty poles that cannot be accommodated in case $\Lambda = 0$ is 22 which is significantly more than any other cases listed in Table 4 (actually number of uncertainties that cannot be accommodated in case $\Lambda = 0$ is infinite). That is to say it is necessary to employ a modified MCM, such as our proposed robust MCM, in day-ahead EM with considerable uncertainties (i.e., WPPs).
- Comparing cases of $\Lambda = 4$ and $\Lambda = 5$, on one hand, there are no uncertainties that cannot be accommodated in both of the two cases; on the other hand, operation cost in case $\Lambda = 4$ is equal to that in case $\Lambda = 5$. Moreover, increasing Λ means to increase the computational complexity of solving the robust MCM [15,16,19,20]. Hence, the proposed robust MCM with $\Lambda = 4$ is applied for market clearing in our subsequent simulations.

4.3. LSCAC-Based EM Modeling Approach Testing

In this section and our subsequent simulations, no matter under which BM, every WPP (agent) will start with experiencing a training process of 3000 iterations. During this training process, all WPPs consider the balance of exploration and exploitation when selecting bidding strategies (actions) in each iteration [41]. After the training process, decision making process of 500 iterations will be implemented by every WPP, in which only greedy policy will be adopted when selecting actions in face of any state of the market. Moreover, we randomly set the action for every WPP at the beginning of the first training iteration because every WPP starts with limited experience in strategy selecting.

Testing and verifying whether our proposed LSCAC-based day-head EM approach under BM 1 reaches to dynamic stability or not after 3000 training iterations can be shown in Figures 2–4. Moreover, Testing and verifying whether our proposed LSCAC-based day-head EM approach under BM 2 reaches to dynamic stability or not after 3000 training iterations can be shown in Figures 5–7. In Figures 4 and 7, summed UMP at bus m in each iteration can be calculated by using Equation (39).

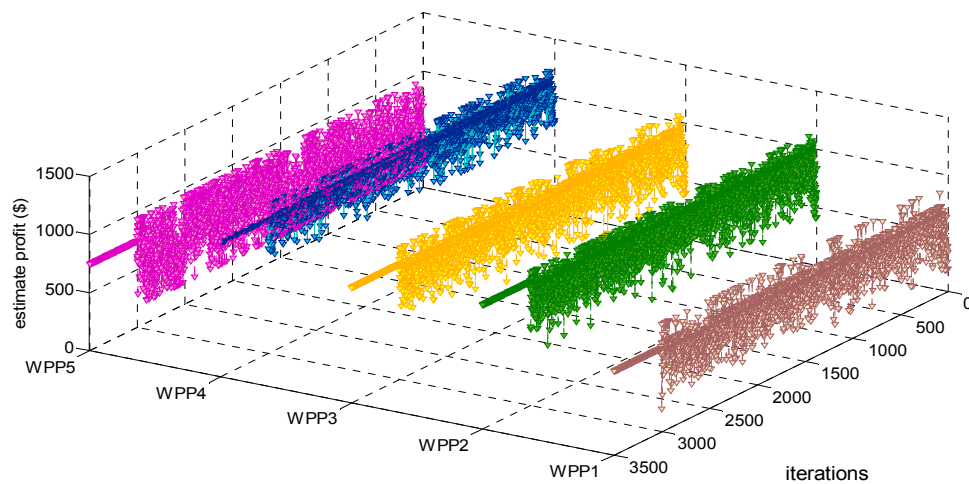


Figure 2. Dynamic adjustment process of estimated profit corresponding to every WPP under BM 1.

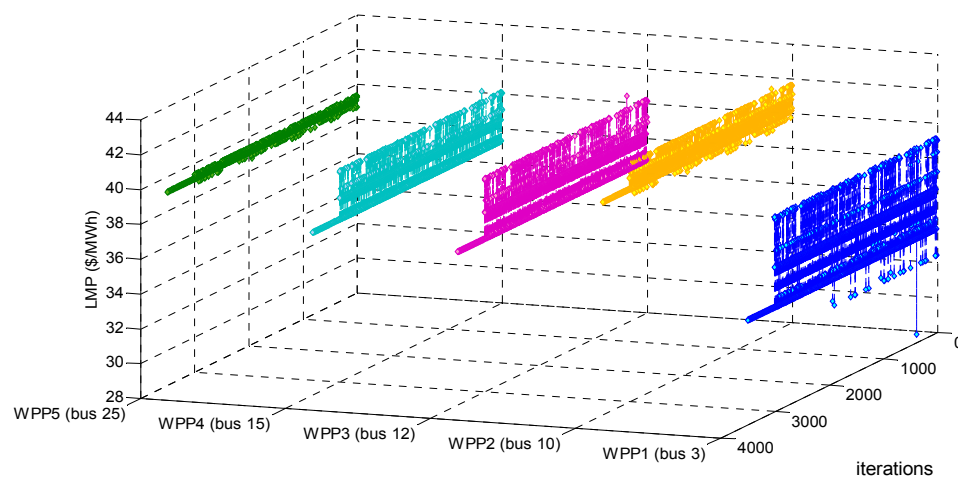


Figure 3. Dynamic adjustment process of LMP corresponding to every WPP under BM 1.

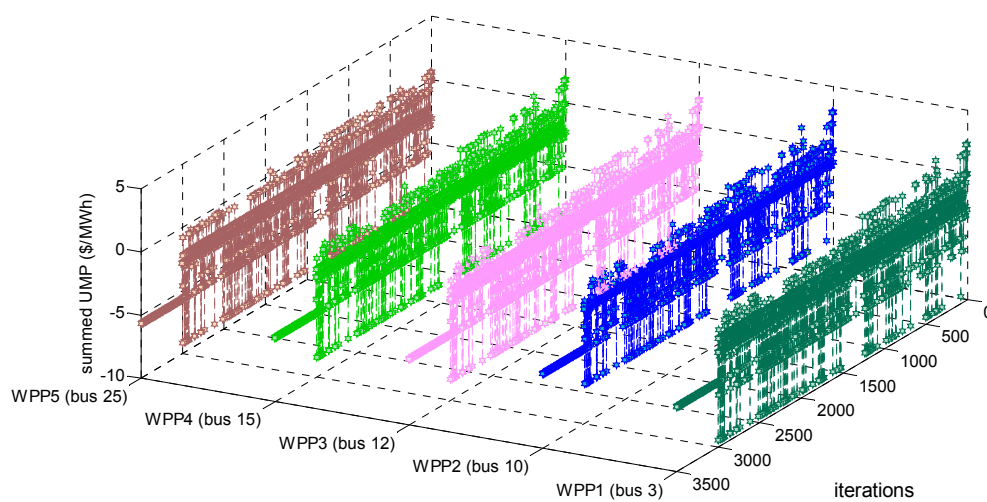


Figure 4. Dynamic adjustment process of summed UMP corresponding to every WPP under BM 1.

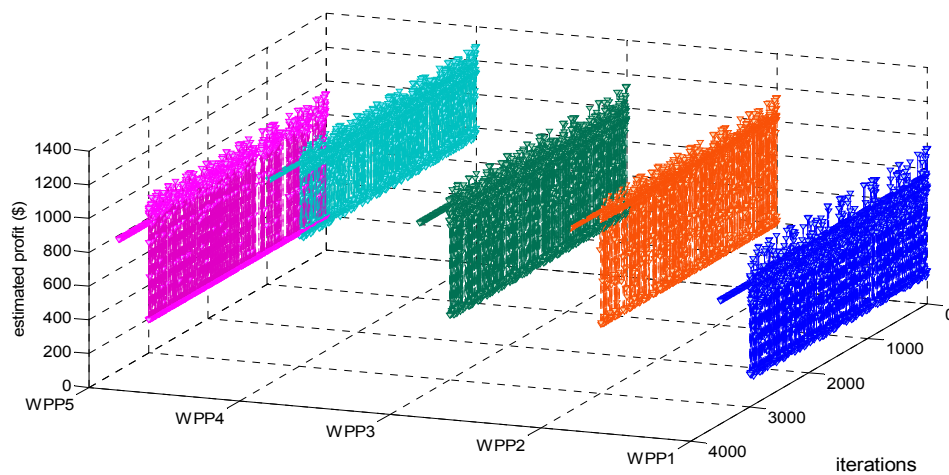


Figure 5. Dynamic adjustment process of estimated profit corresponding to every WPP under BM 2.

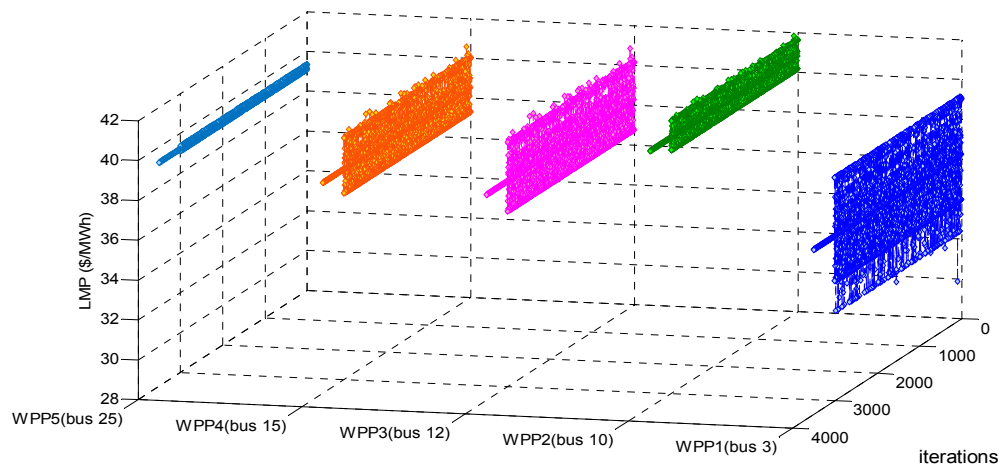


Figure 6. Dynamic adjustment process of LMP corresponding to every WPP under BM 2.

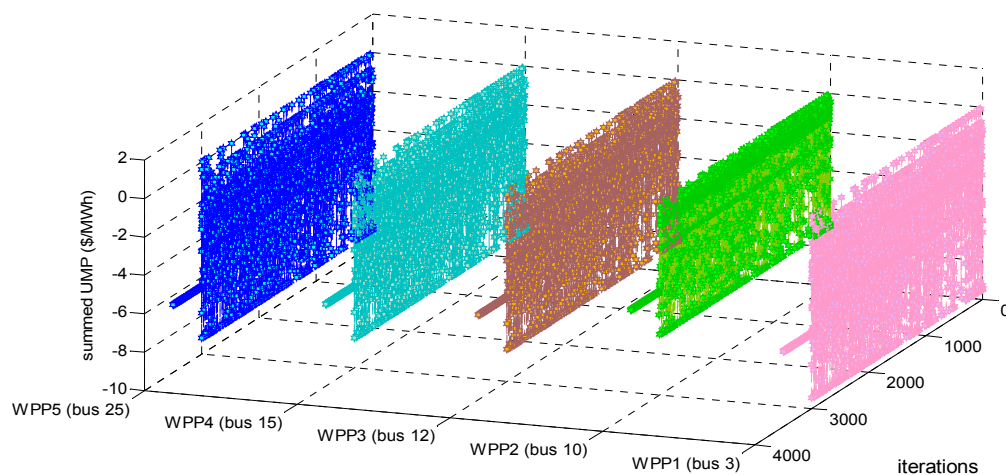


Figure 7. Dynamic adjustment process of summed UMP corresponding to every WPP under BM 2.

Before we analyze BMs and strategies for WPPs by using our proposed LSCAC-based day-head EM modeling approach, it should be tested first whether our proposed approaches under different BMs converge to dynamic stabilities after every WPP experiences enough iterations of on line training. If the

convergence was verified, the market state and obtained action of every WPP would no longer change after enough training iterations. It should be noted that in the existing TBRL-based approaches [26–31,36], the action set of every agent is discrete and finite, and the optimality of an agent's final obtained action can be easily verified by using method mentioned in [31], which is to compare profits brought from all actions in this agent's action set while fixing the actions of other agents. However, in our proposed LSCAC-based approach, the action set of every agent is continuous. It is impossible to directly test the optimality of an agent's final obtained action because there are infinite actions other than this final obtained one. Therefore, we propose the following three steps to further test the performance of the LSCAC-based day-head EM modeling approach:

- To test the optimality of a WPP's final obtained strategy in TBRL-based (i.e., Q-Learning algorithm [26,27]) day-head EM modeling approaches after converging to dynamic stabilities by comparing profits brought from all this WPP's strategies while fixing other WPPs' obtained strategies. The specific optimality test method can be seen in [31].
- To test whether a WPP can obtain more profit by using LSCAC algorithm than TBRL algorithm (Q-Learning algorithm [26,27]) or not, after converging to dynamic stabilities.
- To test whether the whole market can reach lower operation cost in our proposed LSCAC-based approach than TBRL-based (Q-Learning algorithm [26,27]) one or not, after converging to dynamic stabilities.

The related parameters of our LSCAC-based day-head EM modeling approach are listed in Table 5.

Table 5. Related parameters in LSCAC-based day-head EM modeling approach.

Under BM 1							
LSCAC-Based Agents	EM State Set (\$/MWh)		Action Set	γ	α	σ	q
	X_1	X_2					
WPP ₁	[0,100]	[−50, 50]	[5,20] MW	0.5	0.1	4.5	1
WPP ₂	[0,100]	[−50, 50]	[10,25] MW	0.5	0.1	4.5	1
WPP ₃	[0,100]	[−50, 50]	[10,25] MW	0.5	0.1	4.5	1
WPP ₅	[0,100]	[−50, 50]	[20,30] MW	0.5	0.1	3	1
WPP ₅	[0,100]	[−50, 50]	[5,20] MW	0.5	0.1	4.5	1
Under BM 2							
LSCAC-Based Agents	EM State Set (\$/MWh)		Action Set	γ	α	σ	q
	X_1	X_2					
WPP ₁	[0,100]	[−50, 50]	[30,50] \$/MWh	0.5	0.1	4.5	1
WPP ₂	[0,100]	[−50, 50]	[30,50] \$/MWh	0.5	0.1	4.5	1
WPP ₃	[0,100]	[−50, 50]	[30,50] \$/MWh	0.5	0.1	4.5	1
WPP ₅	[0,100]	[−50, 50]	[30,50] \$/MWh	0.5	0.1	3	1
WPP ₅	[0,100]	[−50, 50]	[30,50] \$/MWh	0.5	0.1	4.5	1
Central point sets in the Gauss radial basis function corresponding with X_1 and X_2		X_1	{0, 5, 10, 15, ... , 100}				
		X_2	{−50, −45, −40, ... , 50}				

From Figures 2–4, it can be seen that, after randomly fluctuating in 3000 training iterations, the adjustment processes of estimated profit, LMP and summed UMP of every WPP remain constant during 500 decision-making iterations. Actually, other adjustment processes such as that of operation cost, every WPP's bidding strategy, etc. also become constant after 3000 training iterations. Therefore, our proposed approach under BM 1 can converge to dynamic stability after every WPP experiences 3000 iterations of online training.

From Figures 5–7, it can be seen that, after randomly fluctuating in 3000 training iterations, the adjustment processes of estimated profit, LMP and summed UMP of every WPP remain constant

during 500 decision-making iterations. Actually, other adjustment processes such as that of operation cost, every WPP's bidding strategy, etc. also become constant after 3000 training iterations. Therefore, our proposed approach under BM 2 can converge to dynamic stability after every WPP experiences 3000 iterations of online training.

The main reason about the fluctuating trends in the 3000 training iterations in Figures 2–7 is that in order to balance the exploration and exploitation during these 3000 training iterations, every WPP must maintain the ability of exploration which is to randomly select bidding strategies according to the repeatedly updated Equation (45), all WPPs' insufficient experiences and unstable action selecting policies make the dynamic training process of EM fluctuate randomly. The main reason about the constant trends in 500 decision-making iterations in Figures 2–7 is that after accumulating enough experiences, every WPP adopts the greedy policy which is to only select its considered optimal bidding strategy in face of any observed EM state in each of the 500 decision-making iterations, all WPPs' sufficient experiences and stable action selecting policies make the dynamic decision-making process of EM converge to stability. Therefore, it may be concluded that enough training iterations considering the balance of exploration and exploitation, as well as the greedy action selecting policy adopted in decision-making iterations are two main factors resulting in EM dynamic stability. Taking EM approach under BM 1 for example, Figure 8 shows the dynamic adjusting process of WPP₁'s estimated profit when every WPP experiences 1000 training iterations and 500 decision-making iterations, and Figure 9 shows the dynamic adjustment process of WPP₁'s estimated profit when every WPP experiences 3500 training iterations without greedy action selecting policy.

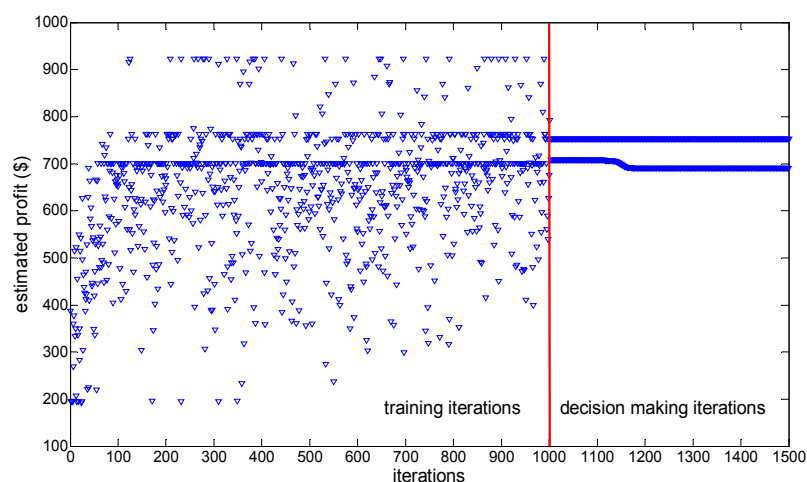


Figure 8. Dynamic adjusting process of WPP₁'s estimated profit under BM 1 when every WPP experiences 1000 training iterations and 500 decision-making iterations.

From Figure 8, it is shown that although the greedy action selecting policies are adopted by WPPs in decision-making iterations, insufficient training iterations, which mean insufficient experiences accumulated, still make WPP₁'s estimated profit fluctuate during decision-making process. Actually, the dynamic adjustment processes of other WPPs' estimated profits also fluctuate during decision-making iterations.

From Figure 9, it is shown that although more than 3000 training iterations considering the balance of exploration and exploitation are conducted, WPP₁'s estimated profit still fluctuates during the last 500 iterations due to its lack of a greedy action selection policy. Actually, the dynamic adjustment processes of other WPPs' estimated profits also fluctuate during the last 500 iterations. Moreover, WPP₁'s estimated profit in Figure 9 is much more volatile during the last 500 iterations than that in Figure 8, which is mainly because WPPs lacking greedy action selection policies tend to bid more randomly in the EM.

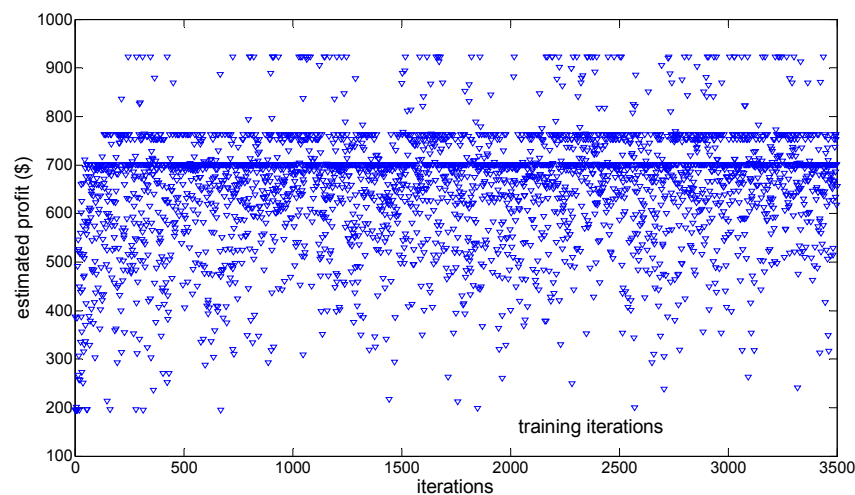


Figure 9. Dynamic adjustment process of WPP₁'s estimated profit under BM 1 when every WPP experiences 3500 training iterations without greedy action selecting policy.

Therefore, EM dynamic stability cannot be reached whether there are insufficient training iterations or the greedy action selecting policy is not considered, which, to a certain extent, on the one hand verifies our conclusions about the two main factors resulting in EM dynamic stability, and on the other hand, suggests that the proposed 3000 training iterations and 500 decision-making iterations are comparatively reasonable for our proposed LSCAC-based approach to reach EM dynamic stability.

To further test the performance of our proposed approaches under different BMs, two Q-learning-based day-ahead EM approaches (QDEMAs) are taken for comparison. In these QDEMAs, some WPPs are designated as Q-learning-based agents while other undesignated WPPs are still the LSCAC-based ones. A Q-learning-based agent dynamically adjusts its action based on Q-learning algorithm which use ϵ -greedy policy [26] to balance exploration and exploitation in 3000 training iterations, and greedy policy in 500 decision-making iterations. Difference among these QDEMAs is only reflected in the number of Q-learning-based agents. Parameters related to these two QDEMAs are listed in Table 6.

Table 6. Related parameters of Q-learning-based agents in two QDEMAs.

QDEMA 1 under BM 1						
Q-Learning-Based Agents	EM State Set (\$/MWh)		Action Set (MW)	ϵ	γ	α
	X_1	X_2				
WPP ₁	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{5, 6, ..., 20}	0.1	0.5	0.1
WPP ₂	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{10, 11, ..., 25}	0.1	0.5	0.1
WPP ₃	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{10, 11, ..., 25}	0.1	0.5	0.1
WPP ₄	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{20, 21, ..., 30}	0.1	0.5	0.1
WPP ₅	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{5, 6, ..., 20}	0.1	0.5	0.1
QDEMA 1 under BM 2						
Q-Learning-Based Agents	EM State Set (\$/MWh)		Action Set (\$/MWh)	ϵ	γ	α
	X_1	X_2				
WPP ₁	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{30, 31, 32, ..., 50}	0.1	0.5	0.1
WPP ₂	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{30, 31, 32, ..., 50}	0.1	0.5	0.1
WPP ₃	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{30, 31, 32, ..., 50}	0.1	0.5	0.1
WPP ₄	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{30, 31, 32, ..., 50}	0.1	0.5	0.1
WPP ₅	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{30, 31, 32, ..., 50}	0.1	0.5	0.1

Table 6. Cont.

QDEMA 2 under BM 1						
Q-Learning-Based Agents	EM State Set (\$/MWh)		Action Set (MW)	ϵ	γ	α
	X_1	X_2				
WPP ₃	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{10, 11, ..., 25}	0.1	0.5	0.1
WPP ₄	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{20, 21, ..., 30}	0.1	0.5	0.1
WPP ₅	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{5, 6, ..., 20}	0.1	0.5	0.1

QDEMA 1 under BM 2						
Q-Learning-Based Agents	EM State Set (\$/MWh)		Action Set (\$/MWh)	ϵ	γ	α
	X_1	X_2				
WPP ₃	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{30,31,32, ..., 50}	0.1	0.5	0.1
WPP ₄	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{30,31,32, ..., 50}	0.1	0.5	0.1
WPP ₅	{0, 5, 10, ..., 100}	{−50, −45, −40, ..., 50}	{30,31,32, ..., 50}	0.1	0.5	0.1

After 3000 training and 500 decision-making iterations, the obtained market results of those two QDEMA and our proposed approach under BM 1 are listed in Table 7, and the obtained results of those two QDEMA and our proposed approach under BM 2 are listed in Table 8. Moreover, like our proposed LSCAC-based approach, no matter under which BM, both these QDEMA can converge to dynamic stability after every WPP experiences 3000 iterations of online training. That means those results listed in Tables 7 and 8 are not obtained accidentally, a LSCAC-based or Q-learning-based WPP does not change its strategy when market state affected by all WPPs' strategies keeps unchanged.

Table 7. Obtained market results of those two QDEMA and our proposed approach under BM 1 (10^3 \$).

Approaches	Profit of					Operation Cost
	WPP ₁	WPP ₂	WPP ₃	WPP ₄	WPP ₅	
QDEMA 1	0.6497	0.9183	0.8195	1.0683	0.6631	6.1012
QDEMA 2	0.8034	1.0859	0.7706	0.9175	0.5989	5.9973
LSCAC-based approach	0.7185	1.0667	0.9917	1.1578	0.7432	5.8309

Table 8. Obtained market results of those two QDEMA and our proposed approach under BM 2 (10^3 \$).

Approaches	Profit of					Operation Cost
	WPP ₁	WPP ₂	WPP ₃	WPP ₄	WPP ₅	
QDEMA 1	0.6816	0.9643	0.8740	1.1375	0.7012	6.9908
QDEMA 2	0.8319	1.1083	0.8172	1.0864	0.6398	6.7637
LSCAC-based approach	0.7318	1.0797	1.0316	1.2073	0.7726	6.4134

From Tables 7 and 8, it can be conclude that:

- By using the optimality test method in [31], no matter under which BM and in which QDEMA, every Q-learning-based WPP's final obtained strategy can be verified as its optimal one in its discrete action set, which can bring it the most profit when other WPPs' strategies are fixed.
- No matter under which BM, on the one hand, estimated profits of WPP₁ and WPP₂ in QDEMA 2 are higher than those in QDEMA 1, respectively, and estimated profits of WPP₃, WPP₄ and WPP₅ in our proposed LSCAC-based approach are higher than those in QDEMA 2, respectively, which, to some extent, indicates one can get more profit by using the LSCAC algorithm to bid in EM than the Q-learning one within the same conditions; on the other hand, the operation cost in our proposed LSCAC-based approach is lower than that in QDEMA 2, and the operation cost in

QDEMA 2 is lower than that in QDEMA 1, which, to some extent, indicates that with the increase in the number of LSCAC-based agents in EM, the operation cost of whole system can be reduced.

In conclusion, no matter under which BM, Q-learning-based WPPs can finally find their optimal bidding strategies from their discrete and finite action sets. If these WPPs are transformed into LSCAC-based ones, they can finally find their more applicable strategies from their continuous action sets, which not only bring more profits for themselves but also bring lower operation cost for the whole system than that based on Q-learning method. Hence, although it is hard to directly test the optimality of every LSCAC-based WPP's final obtained strategy, our further test has, to some extent, verified the rationality and scientific basis of applying our proposed LSCAC-based approach in day-ahead EM modeling for strategic WPPs.

Moreover, no matter under which BM, simulation of our proposed approach on IEEE 30 bus test system with five strategic WPPs takes only about 43 seconds to reach the final results (after 3500 iterations). That is to say, the time complexity of our proposed approach is relatively low so that we can extend it to the modeling and simulation of more realistic and more complex EM system.

4.4. BMs Analysis for WPPs

In this section, our proposed LSCAC-based approach is applied to analyze the obtained market results under different BMs after 3000 training iterations and 500 decision-making iterations.

Moreover, it should be noted that under BM 2, in order to lead $WPP_i(\forall i)$ to reasonably bid in market, we set lower and upper limits $\rho_{low,i}$ and $\rho_{upp,i}$ (\$/MWh) for its bidding price ρ_i^b . Values of $\rho_{low,i}$ and $\rho_{upp,i}(\forall i)$ may affect the obtained market results such as final obtained LMPs, estimated profits of all WPPs and operation cost of the system etc. after 3500 iterations. Hence, different values of $\rho_{low,i}$ and $\rho_{upp,i}(\forall i)$ should be taken into account when considering BMs. For the sake of simplicity and without losing generality, we set $\rho_{low,i} = \rho_{low}, \rho_{upp,i} = \rho_{upp} \forall i$, and different values of ρ_{low} and ρ_{upp} are considered.

After 3500 iterations, considering different values of ρ_{low} (while fixing the upper limit ρ_{upp} to 50 \$/MWh, the same as what listed in Table 5, Table 9 is listed for the comparison of obtained market results under different BMs.

Table 9. Comparison of obtained market results under different BMs by considering different values of ρ_{low} .

BMs	Profit of (10^3 \$)					Operation Cost (10^{-3} \$)	Average LMP (\$/MWh)
	ρ_{low}	WPP ₁	WPP ₂	WPP ₃	WPP ₄	WPP ₅	
Under BM 2	0	0.7318	1.0797	1.0316	1.2073	0.7726	6.4134
	10	0.7318	1.0797	1.0316	1.2073	0.7726	6.4134
	20	0.7318	1.0797	1.0316	1.2073	0.7726	6.4134
	30	0.7318	1.0797	1.0316	1.2073	0.7726	6.4134
Under BM 1		0.7185	1.0667	0.9917	1.1578	0.7432	5.8309

After 3500 iterations, considering different values of ρ_{upp} (while fixing the upper limit ρ_{low} to 30 \$/MWh, the same as what was listed in Table 5, Table 10 is provided for the comparison of obtained market results under different BMs.

Table 10. Comparison of obtained market results under different BMs by considering different values of ρ_{upp} .

BMs	Profit of ($10^3\$$)					Operation Cost ($10^3\$$)	Average LMP (\$/MWh)
	ρ_{upp}	WPP ₁	WPP ₂	WPP ₃	WPP ₄	WPP ₅	
Under BM 2	50	0.7318	1.0797	1.0316	1.2073	0.7726	6.4134
	60	0.7646	1.1082	1.0656	1.2322	0.7953	6.5973
	70	0.8769	1.1231	1.0921	1.2547	0.9080	6.8231
	80	0.9228	1.1439	1.1292	1.2861	0.9242	7.0197
Under BM 1		0.7185	1.0667	0.9917	1.1578	0.7432	5.8909
							38.7490

From Tables 9 and 10, it can be seen that:

- In Table 9, when values of ρ_{low} are 0, 10, 20 and 30 (\$/MWh), respectively, the obtained profit of every WPP, operation cost and average LMP of 30 buses under BM 2 remain unchanged. Actually, if $\rho_{low} \leq 30$ \$/MWh, the obtained bidding price (strategy) of every WPP under BM 2 is higher than 30 (\$/MWh), which means values of ρ_{low} lower than 30 (\$/MWh) cannot affect every WPP's bidding decision-making. Therefore, in our opinion, it is hard to weaken the market power of every WPP by only reducing the value of lower limit of every WPP's bidding price while WPPs provide ISO their bidding curves consisting of bidding prices and power outputs for the next day.
- In Table 10, the obtained profit of every WPP, operation cost and average LMP of 30 buses under BM 2 increases with the increase of the value of ρ_{upp} . That may be mainly because the more the value of ρ_{upp} is, the greater market power WPPs have. Therefore, in our opinion, the upper limit of every WPP's bidding price should not be set too high while WPPs provide ISO their bidding curves consisting of bidding prices and power outputs for the next day.
- In most cases of ρ_{low} and ρ_{upp} , WPPs under BM 2 can get more profits than under BM 1, which may be because WPPs under BM 2 can obtain greater market power by directly adjusting their bidding prices so as to further improve their profits compared with WPPs under BM 1. However, from the perspective of the whole market, both the obtained operation cost and average LMP under BM 2 are higher than that under BM 1, which, to some extent, indicates WPPs adopting BM 2 cause lower economic efficiency in the whole market than adopting BM 1. Therefore, in our opinion, if the purpose of permitting WPPs to bid is to promote the development of wind power resources by improving WPPs' profits, providing ISO their bidding curves is more applicable, and if the purpose of permitting WPPs to bid is to improve the economic efficiency of the whole market, only sending their power output plans is more applicable.

5. Conclusions

In this paper, we present a LSCAC-based day-head EM modeling approach with a robust market clearing mechanism embedded in it, and strategic behaviors of WPPs under two different BMs are successively mimicked and compared by using our proposed approach. With employing the robust MCM, day-head ED solution of the market can be immunized against any uncertainty within the real time wind power uncertainty set estimated by ISO, which not only ensures the ability of wind power accommodation in EM, but also reasonably generates LMPs for energy credit and load payment as well as UMPs for reserve credit and deviation payment. By employing the LSACA algorithm, every WPP can significantly improve its profit and the operation cost of the system can also be remarkably decreased compared with employing the TBRL (i.e., Q-learning) algorithms. Low computational time (taking only about 43 seconds for our simulation on IEEE 30 bus test system to reach the final results) makes that our proposed approach easily extendible to provide a reasonable test bed for simulation of more realistic and more complex EM systems. Moreover, by conducting comparisons on market results under different BMs in simulation, some suggestions leading WPPs to reasonably bid in market

and BM selections are put forward for the purposes of promoting the development of wind power resources and improving the economic efficiency of the whole market.

Acknowledgments: This study is supported by the National Natural Science Foundation of China under Grant No. 71373076, the Fundamental Research Funds for the Central Universities under Grant No. JB2016183, and the National Key R&D Program of China under Grand No. 2016YFB0900501.

Author Contributions: Huiru Zhao guided the research; Yuwei Wang established the model, implemented the simulation and wrote this article; Mingrui Zhao, Qingkun Tan and Sen Guo collected references.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

Acronym

EM	Electricity market
WPP	Wind power producer
BM	Bidding mode
ISO	Independent system operator
MCM	Market clearing model
SO	Stochastic optimization
WPOSS/WPOS	Wind power output stochastic scenario/wind power output scenario
RO	Robust optimization
SCUC/UC	security constraint unit-commitment/unit-commitment
SCED/ED	security constraint economic-dispatch/economic-dispatch
AP	Affine policy
GenCO	generation company
ARIMA	Autoregressive integrated moving average
SARSA	State-action-reward-state-action
KDE	Kernel density estimation
TBRL	Table-based reinforcement learning
MASCEM	Multi-agent simulator of competitive electricity market
LSCAC	Least square continuous actor-critic
RMCM	Robust market clearing model
MP/SP	Master problem/sub-problem
LMP	Locational marginal price
UMP	Uncertainty marginal price
GLM/F	Generalized lagrange multiplier/fuction
TD	Time difference
QDEMA	Q-learning-based day-ahead EM approach

Index

i, j, m	Indices for WPP, conventional generator and bus
k	index of the worst point for uncertainty (element in κ)
l	Index for transmission line
T, N	Indices for iteration in LSCAC-based EM approach
h	Index for the dimension of basis function vector

Set and Function

\mathbf{U}	Real-time wind power output uncertainty set
Ω	Robust feasible region for every day-ahead ED solution which is immunized against any real-time uncertainties within \mathbf{U}
$\Theta(m)$	Set of WPPs, conventional generators and loads connected in bus m
κ	Index set for worst uncertainty points
$\kappa_{m,T-1}^{up}, \kappa_{m,T-1}^{down}$	Set of indices k for upward and downward UMPs at bus m iteration $T - 1$
\mathbf{X}/\mathbf{A}	State space/action space
$\rho(\cdot)$	WPP's bidding function (curve) under BM 2, which is assumed to be identically equal to a certain bidding price ρ_i^b no matter what its bidding power output Pw_i^b is

$L(\cdot)$	Generalized lagrange function
$\hat{I}(\cdot)$	Optimal action selecting policy function
$\phi_h(\cdot)/\boldsymbol{\phi}(\cdot)$	h -th basic function/basic function vector
$pro(\cdot, \cdot)$	Probability density function for selecting action a under state x (representing action selecting policy during training iterations)
$MSE(\cdot)$	Mean square function
$P^{(pro)}(\cdot)$	Probability distribution function of state x under policy pro
$sig(\cdot)$	Sigmoid function
$\delta(\cdot, \cdot)$	TD error function for selecting action a under state x
Constant	
N_w, N_G, N_{bus}	The numbers of WPPs, conventional generators and buses
lw_i, uw_i	lower and upper bounds for i -th WPP's real-time wind power output
$\rho_{low,i}, \rho_{upp,i}$	lower and upper limits for i -th WPP's bidding price
Λ	Budget parameter relating to the size of uncertainty set
c_j	cost coefficient of j -th conventional generator
d_m	Aggregated equivalent load at bus m
P_j^{\max}, P_j^{\min}	Maximum and minimum generation outputs
$\Gamma_{l,m}$	Shift factor for line l with respect to bus m
F_l	Transmission line flow limit for line l
r_j^u, r_j^d	j -th conventional generator's ramping-up/down limits for uncertainty accommodation
Δ	Feasibility tolerance
n	dimension number of basis function vector
q	Coefficient in sigmoid function
Π	Small positive constant for calculating the linear parameter vector ω in $\hat{I}(x)$ based on ridge regression
α	Step length factor
σ	Standard deviation
γ	Discount factor
Parametric Variable	
Pw_i^b	i -th WPP's bidding power output parameter
ρ_i^b	i -th WPP's bidding price parameter
θ	Linear parameter vector in TD error function (and value function)
ω	Linear parameter vector in optimal action selecting policy function
θ^*	Convergence value of θ
ω^*	Convergence value of ω
A_N/b_N	Intermediate parameter matrix/vector in LSCAC iterations for obtaining convergence value of ω
Variable	
Pw/Pw_i	Random variable vector representing the uncertainty of WPPs' joint real-time power outputs/random variable representing the uncertainty of i -th WPP's real-time power output, moreover, Pw/Pw_i can also be considered as part of the decision variables (vector) in SP
Pw_k	k -th worst uncertainty point of WPPs' joint real-time power outputs, moreover, Pw_k also represents part of SP's solutions when solving SP for the k -th time
$P_j, Pw_i^{disp}, \Delta P_{jk}$	Decision variables in MP, representing day-ahead dispatched power output of j -th conventional generator, i -th WPP, as well as real-time power re-dispatch incremental result of j -th conventional generator under k -th worst uncertainty point, respectively
$P_{GW}, Pw^{disp}, \Delta P$	Variable vector consisting of $P_j (\forall j)$ and $Pw_i^{disp} (\forall i)$, variable vector consisting of $Pw_i^{disp} (\forall i)$, variable vector consisting of $\Delta P_{jk} (\forall j, \forall k)$
$s_i^+, s_i^-, \Delta P_j$	Other decision variables in SP, s_i^+, s_i^- are non-negative slack variables, and the sum of which evaluate the violation associated with the solution from MP, ΔP_j represents real time power re-dispatch approach of j -th conventional generator for accommodating uncertainties within U

Ψ	Value of SP's objective function
$\lambda, \beta, \mu, \zeta, \eta$	Generalized lagrange multiplier vectors
ε_{mk}	Deviation of the real-time power output generated by WPPs connecting in bus m under k -th worst uncertainty point from the day-ahead bidding (dispatched) one
π_m, π_{mk}	LMP in bus m , UMP in bus m under k -th worst uncertainty point
R_i	Estimated profit of i -th WPP
x, a, r	State variable, action variable, reward

References

1. Prabavathi, M.; Gnanadass, R. Energy bidding strategies for restructured electricity market. *Int. J. Electr. Power Energy Syst.* **2015**, *64*, 956–966. [[CrossRef](#)]
2. Majumder, S.; Khaparde, S.A. Revenue and ancillary benefit maximization of multiple non-collocated wind power producers considering uncertainties. *IET Gener. Trans. Distrib.* **2016**, *10*, 789–797. [[CrossRef](#)]
3. Li, J.; Wan, C.; Xu, Z. Robust offering strategy for a wind power producer under uncertainties. In Proceedings of 2016 IEEE International Conference on Smart Grid Communications (SmartGridComm), Sydney, Australia, 6–9 November 2016; pp. 752–757.
4. Zhao, Q.; Shen, Y.; Li, M. Control and bidding strategy for virtual power plants with renewable generation and inelastic demand in electricity markets. *IEEE Trans. Sustain. Energy* **2016**, *7*, 562–575. [[CrossRef](#)]
5. Zugno, M.; Morales, J.M.; Pinson, P.; Madsen, H. Pool strategy of a price-maker wind power producer. *IEEE Trans. Power Syst.* **2013**, *28*, 3440–3450. [[CrossRef](#)]
6. Shafie-khah, M.; Heydarian-Forushani, E.; Golshan, M.E.H.; Moghaddam, M.P.; Sheikh-El-Eslami, M.K.; Catalão, J.P.S. Strategic offering for a price-maker wind power producer in oligopoly markets considering demand response exchange. *IEEE Trans. Ind. Inf.* **2015**, *11*, 1542–1553. [[CrossRef](#)]
7. Delikaraoglou, S.; Papakonstantinou, A.; Ordoudis, C.; Pinson, P. Price-maker wind power producer participating in a joint day-ahead and real-time market. In Proceedings of the 12th International Conference on the European Energy Market (EEM), Lisbon, Portugal, 19–22 May 2015; pp. 1–5.
8. De la Nieta, A.A.S.; Contreras, J.; Muñoz, J.I.; O'Malley, M. Modeling the impact of a wind power producer as a price-maker. *IEEE Trans. Power Syst.* **2014**, *29*, 2723–2732. [[CrossRef](#)]
9. Lei, M.; Zhang, J.; Dong, X.; Ye, J.J. Modeling the bids of wind power producers in the day-ahead market with stochastic market clearing. *Sustain. Energy Technol. Assessm.* **2016**, *16*, 151–161. [[CrossRef](#)]
10. Ye, H.; Ge, Y.; Shahidehpour, M.; Li, Z. Pricing energy and flexibility in robust Security-Constrained Unit Commitment model. In Proceedings of the 2016 IEEE Power and Energy Society General Meeting (PESGM), Boston, MA, USA, 17–21 July 2016; pp. 1–5.
11. Morales, J.M.; Conejo, A.J.; Perez-Ruiz, J. Economic valuation of reserves in power systems with high penetration of wind power. *IEEE Trans. Power Syst.* **2009**, *24*, 900–910. [[CrossRef](#)]
12. Morales, J.M.; Conejo, A.J.; Liu, K.; Zhong, J. Pricing electricity in pools with wind producers. *IEEE Trans. Power Syst.* **2012**, *27*, 1366–1376. [[CrossRef](#)]
13. Catalao, J.P.S.; Pousinho, H.M.I.; Mendes, V.M.F. Optimal offering strategies for wind power producers considering uncertainty and risk. *IEEE Syst. J.* **2012**, *6*, 270–277. [[CrossRef](#)]
14. Zugno, M.; Conejo, A.J. A robust optimization approach to energy and reserve dispatch in electricity markets. *Eur. J. Oper. Res.* **2015**, *247*, 659–671. [[CrossRef](#)]
15. Wei, W.; Liu, F.; Mei, S. Robust and economical scheduling methodology for power systems. Part one: Theoretical foundations. *Autom. Electr. Power Syst.* **2013**, *37*, 37–43.
16. Jiang, R.; Wang, J.; Guan, Y. Robust unit commitment with wind power and pumped storage hydro. *IEEE Trans. Power Syst.* **2012**, *27*, 800–810. [[CrossRef](#)]
17. Zhao, C.; Guan, Y. Unified stochastic and robust unit commitment. *IEEE Trans. Power Syst.* **2013**, *28*, 3353–3361. [[CrossRef](#)]
18. Warrington, J.; Goulart, P.; Mariéthoz, S.; Morari, M. Policy-based reserves for power systems. *IEEE Trans. Power Syst.* **2013**, *28*, 4427–4437. [[CrossRef](#)]
19. Ye, H.; Wang, J.; Li, Z. MIP reformulation for max-min problems in two-stage robust SCUC. *IEEE Trans. Power Syst.* **2017**, *32*, 1237–1247. [[CrossRef](#)]

20. Ye, H.; Ge, Y.; Shahidepour, M.; Li, Z. Uncertainty marginal price, transmission reserve, and day-ahead market clearing with robust unit commitment. *IEEE Trans. Power Syst.* **2016**, *32*, 1782–1795. [CrossRef]
21. Langary, D.; Sadati, N.; Ranjbar, A.M. Direct approach in computing robust Nash strategies for generating companies in electricity markets. *Int. J. Electr. Power Energy Syst.* **2014**, *54*, 442–453. [CrossRef]
22. Salem, Y.; Agtash, A. Supply curve bidding of electricity in constrained power networks. *Energy* **2010**, *35*, 2886–2892.
23. Min, C.G.; Kim, M.K.; Park, J.K.; Yoon, Y.T. Game-theory-based generation maintenance scheduling in electricity markets. *Energy* **2013**, *55*, 310–318. [CrossRef]
24. Wang, J.; Zhou, Z.; Botterud, A. An evolutionary game approach to analyzing bidding strategies in electricity markets with elastic demand. *Energy* **2011**, *36*, 3459–3467. [CrossRef]
25. Shivaie, M.; Ameli, M.T. An environmental/techno-economic approach for bidding strategy in security-constrained electricity markets by a bi-level harmony search algorithm. *Renew. Energy* **2015**, *83*, 881–896. [CrossRef]
26. Rahimiyan, M.; Mashhadi, H.R. Supplier's optimal bidding strategy in electricity pay-as-bid auction: Comparison of the Q-learning and a model-based approach. *Electr. Power Syst. Res.* **2008**, *78*, 165–175. [CrossRef]
27. Xiong, G.; Hashiyama, T.; Okuma, S. An electricity supplier bidding strategy through Q-Learning. In Proceedings of the IEEE Power Engineering Society Summer Meeting, Chicago, IL, USA, 21–25 July 2002; pp. 1516–1521.
28. Ziogos, N.P.; Tellidou, A.C. An agent-based FTR auction simulator. *Electr. Power Syst. Res.* **2011**, *81*, 1239–1246. [CrossRef]
29. Santos, G.; Fernandes, R.; Pinto, T.; Praça, I.; Vale, Z.; Morais, H. MASCEM: EPEX SPOT Day-Ahead market integration and simulation. In Proceedings of the 18th International Conference on Intelligent System Application to Power Systems (ISAP), Porto, Portugal, 11–16 September 2015; pp. 1–5.
30. Bach, T. Using Reinforcement Learning to Study the Features of the Participants' Behavior in Wholesale Power Market. Available online: https://www.science-definition.com/whatis/Using_Reinforcement_Learning_to_Study_the_Features_of_the_Participants%20A1%20AF_Behavior_in_Wholesale_Power_Market (accessed on 29 June 2017).
31. Salehizadeh, M.R.; Soltaniyan, S. Application of fuzzy Q-learning for electricity market modeling by considering renewable power penetration. *Renew. Sustain. Energy Rev.* **2016**, *56*, 1172–1181. [CrossRef]
32. Xiao, Y.; Wang, X.; Wang, X.; Dang, C.; Lu, M. Behavior analysis of wind power producer in electricity market. *Appl. Energy* **2016**, *171*, 325–335. [CrossRef]
33. Ravnaas, K.W.; Doorman, G.; Farahmand, H. Optimal wind farm bids under different balancing market arrangements. In Proceedings of the IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems, Singapore, 14–17 June 2010; pp. 30–35.
34. Sharma, K.C.; Bhakar, R.; Tiwari, H.P. Strategic bidding for wind power producers in electricity markets. *Energy Conv. Manag.* **2014**, *86*, 259–267. [CrossRef]
35. Matevosyan, J.; Soder, L. Minimization of imbalance cost trading wind power on the short term power market. *IEEE Trans. Power Syst.* **2005**, *21*, 1–7.
36. Soares, T.; Santos, G.; Pinto, T.; Morais, H.; Pinson, P.; Vale, Z. Analysis of strategic wind power participation in energy market using MASCEM simulator. In Proceedings of the 2015 18th International Conference on Intelligent System Application to Power Systems (ISAP), Porto, Portugal, 11–16 September 2015; pp. 1–6.
37. Ding, H.; Pinson, P.; Hu, Z.; Wang, J.; Song, Y. Optimal offering and operating strategy for a large wind-storage system as a price maker. *IEEE Trans. Power Syst.* **2017**, *PP*, 1. [CrossRef]
38. Laia, R.; Pousinho, H.M.I.; Melico, R.; Mendes, V.M.F. Bidding strategy of wind-thermal energy producers. *Renew. Energy* **2016**, *99*, 673–681. [CrossRef]
39. Vilim, M.; Botterud, A. Wind power bidding in electricity markets with high wind penetration. *Appl. Energy* **2014**, *118*, 141–155. [CrossRef]
40. Chen, G. Research on Value Function Approximation Methods in Reinforcement Learning. Master's Thesis, Soochow University, Jiangsu, China, 2014.
41. Zhao, H.; Wang, Y.; Guo, S.; Zhao, M.; Zhang, C. Application of gradient descent continuous actor-critic algorithm for double-side day-ahead electricity market modeling. *Energies* **2016**, *9*, 725. [CrossRef]

42. Index of /Data. Available online: <http://motor.ece.iit.edu/Data/> (accessed on 29 June 2017).
43. Buygi, M.O.; Zareipour, H.; Rosehart, W.D. Impacts of large-scale integration of intermittent resources on electricity markets: A supply function equilibrium approach. *IEEE Syst. J.* **2012**, *6*, 220–232. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).