energies

MDPI

Article

Multi-Agent Cooperation Based Reduced-Dimension Q(λ) Learning for Optimal Carbon-Energy Combined-Flow

Huazhen Cao [1], Chong Gao [1], Xuan He [1], Yang Li [1] and Tao Yu [2,*]

[1] Power Grid Planning Center of Guangdong Power Grid Co., Ltd., Guangzhou 510640, China; 13922120994@139.com (H.C.); gao__chong@163.com (C.G.); hexuan1216@163.com (X.H.); liyang2010vip@163.com (Y.L.)

[2] School of Electric Power, South China University of Technology, Guangzhou 510640, China

* Correspondence: taoyu1@scut.edu.cn; Tel.: +86-20-2223-6205

Received: 17 June 2020; Accepted: 31 August 2020; Published: 14 September 2020

Abstract: This paper builds an optimal carbon-energy combined-flow (OCECF) model to optimize the carbon emission and energy losses of power grids simultaneously. A novel multi-agent cooperative reduced-dimension Q(λ) (MCR-Q(λ)) is proposed for solving the model. Firstly, on the basis of the traditional single-objective Q(λ) algorithm, the solution space is reduced effectively to shrink the size of *Q*-value matrices. Then, based on the concept of ant cooperative cooperation, multi-agents are used to update the *Q*-value matrices iteratively, which can significantly improve the updating rate. The simulation in the IEEE 118-bus system indicates that the proposed technique can decrease the convergence speed by hundreds of times as compared with conventional Q(λ), keeping high global stability, which is very suitable for dynamic OCECF in a large and complex power grid compared with other algorithms.

Keywords: multi-agent cooperation; reduced-dimension Q(λ); optimal carbon-energy combined-flow

1. Introduction

With the increasing impact of the greenhouse effect on the environment, low-carbon economy has gradually become the key development direction of various energy consumption industries. As the largest CO_2 emitter, the electric power industry will play an important role in low-carbon economic development [1]. All kinds of energy-consuming enterprises have also commenced on focusing on the control of carbon emissions, especially in the power industry, which makes up approximately 40% of CO_2 emissions in the whole world [2]. Generally speaking, low-carbon power involves four sectors: generation, transmission, distribution and consumption. Therefore, how to reduce the carbon emissions of transmission and distribution sectors in the power grid industry has turned into an instant issue to be solved [3,4].

Up to now, numerous scholars have carried out research on all aspects of low-carbon power, including optimal power flow (OPF) [5–7], economic emission dispatching [8,9], low-carbon power system dispatch [10], unit commitment [11,12], carbon storage and capture [13,14] and other issues. However, the previous studies mainly focused on the carbon emissions of the generation side, with a lack of research on how to reduce the carbon emissions of the power network (i.e., the transmission and distribution sides). Therefore, the optimal carbon-energy combined-flow (OCECF) model, which can reflect the energy flow and carbon flow distribution of the power grid, is further established in this paper. Basically, the OCECF is on the basis of the conventional reactive power optimization model, which should not only attempt to minimize the power loss and voltage deviation, but also aim to minimize the carbon emission of the power network while satisfying the various operating constraints of power systems.

Obviously, the OCECF is a complicated nonlinear planning problem considering the carbon flow losses of power grids, which can be solved by traditional optimization strategies including nonlinear planning [15], the Newton method [16] and the interior point method [17]. However, due to the strong nonlinearity of power systems, the discontinuity of the objective function and constraint conditions, as well as the existence of multiple local optimal solutions, usually hinder the effectiveness or applications of the classical optimization methods. On the other hand, meta-heuristic algorithms including the genetic algorithm (GA) [18], particle swarm optimization (PSO) [19,20], grouped grey wolf optimizer (GWO) [21] and the memetic salp swarm algorithm (MSSA) [22] have relatively low dependence on specific models, and can obtain relatively satisfactory results when solving such problems. However, due to the low convergence stability of the algorithm, these algorithms may only converge to a local optimal solution. Thus, the conventional Q(λ) reinforcement learning algorithm with better convergence robustness and stability is proposed in [23]. Nevertheless, because of the search ergodicity of the single agent Q(λ) algorithm, its convergence is relatively long for large-scale system optimization due to the low learning efficiency, while the "dimension disaster" problem with the increasing number of variables can also occur. Moreover, the on-line optimization requirement of the OCECF is also difficult to be met.

Therefore, the author of ant colony optimization (ACO) introduces the concept of ant colony in the classical Q-learning algorithm and puts forward the multiagent Ant-Q algorithm with a faster optimization speed [24]. Based on this, a new multi-agent cooperation-based reduced-dimension Q(λ) (MCR-Q(λ)) learning is proposed for OCECE in this paper, which mainly contains the following contributions:

(i) Most of existing low-carbon power studies did not consider the carbon emissions of the power network due to the energy flow and carbon flow from the generation side to the load side, which cannot satisfy the low-carbon requirement from the viewpoint of the power network. In contrast, the presented OCECF can further reduce the carbon emissions of the power network, which can improve the benefit of the power grid company in a carbon trading market.

(ii) The proposed MCR-Q(λ) can effectively shorten the dimension of the solution space of the Q algorithm to solve the OCECF problem by introducing the eligibility trace (λ) returns mechanism [23]. Besides, it also can accelerate the convergence rate and avoid trapping into a low-quality optimum for OCECE via multi-agent cooperation.

The framework of this paper mainly includes: firstly, Section 2 which concludes the related work; Section 3 presents the establishment of the OCECF mathematical model; then, the principle of MCR-Q(λ) learning is described in Section 4; Section 5 gives the concrete steps of solving the OCECF problem; Section 6 undertakes simulation studies on the IEEE 118 node system to verify the convergence and stability of MCR-Q(λ) learning. Finally, the conclusion of the whole paper is presented in Section 7.

2. Related Work

2.1. Low-Carbon Power

To achieve a low-carbon operation of a power system, extensive studies were devoted to addressing the environmental economic dispatch (EED). In EED, the minimization of emissions [25] is generally designed as one part of the objective function. To further improve the operation economy, the uncertainty of wind power was considered in [26,27], in which the power output of a wind turbine was evaluated based on a probability distribution function of the wind speed. Besides, a modified EED, by combining heat and power economic dispatch, was presented in [28], which can achieve an optimal operation for the heat and power system simultaneously. Furthermore, a coordinated operation of an integrated regional energy system with various energies (e.g., a CO_2-capture-based power) was proposed in [29], while the demand response was also introduced in EED. To further reduce carbon emissions, the CO_2 emission trading system was combined into the daily operation of an energy system. In [30], a decentralized economic dispatch was proposed by considering the carbon capture power plants with carbon emission trading. Moreover, the power uncertainty of wind and photovoltaic energy was fully taken into account in [31,32] based on carbon emission trading. For the purpose of clarifying the internal relation between energy consumption and carbon emissions from power grids, the concept of carbon emission flow is put forward for the first time in reference [33]. On this basis, the authors of [34–36] carried out a theoretical analysis and case verification on the carbon emission flow calculation and the carbon flow tracking of a power system, respectively.

2.2. *Application of Meta-Heuristic Algorithms*

In fact, the optimal low-carbon operation of a power system faces with various complex and difficult optimization problems, e.g., EED. Hence, various meta-heuristic algorithms have been employed for these optimization problems due to their strong searching ability and high application flexibility. In [25], an improved PSO combining the differential evolution algorithms was designed for EED. In [26], a so-called exchange market algorithm was used for EED due to its fast convergence and strong global searching ability. In [27], a population-based honey bee mating optimization with an online learning mechanism was presented. Inspired by the well-known tag-team game in India, the novel Kho-Kho optimization algorithm [28] with an excellent optimization performance was proposed for EED. To achieve a distributed optimization for real-time power dispatch, a novel adaptive distributed auction-based algorithm with a varying swap size was proposed in [37]. On the other hand, the reinforcement learning-based optimization attracted many investigations for optimal operations of power systems. In [23], a distributed multi-step Q(λ) learning was proposed for the complex OPF of a large-scale power system. To satisfy the requirement of multi-objective optimization, an approximate ideal multi-objective solution Q(λ) learning was presented in [36] via a design of multiple Q matrices for different objective functions.

3. OCECF Mathematical Model

3.1. *Carbon-Energy Combined-Flow*

The carbon-energy combined-flow (CECF) of the power grid is a comprehensive network flow [36], which combines the power flow of the power grid with the carbon emission flow attached to the power flow of the power grid. Among them, the energy flow is the actual network flow, and the carbon emission flow is the virtual network flow, which can be referred to as the carbon flow in the power system. Carbon flow is generated in the power generation, which represents the concept that the carbon emission is transferred from the generation side to the demand side. The energy flow transfers from the power supply end to the receiving end, but unlike the energy flow, only the power supply that produces carbon emissions at the power supply end can be called a carbon source, as shown in Figure 1. For a given carbon source, the carbon emission is equivalent to the product of the energy flow and the carbon emission rate of the corresponding power generation side [35].

Energy flow is the transmission of electric energy in the power grid. In the process of transmission, there will be power losses, commonly known as network losses, which are generally described as follows:

$$P_{\text{loss}} = \sum_{i,j \in N_{\text{L}}} g_{ij}\left[V_i^2 + V_j^2 - 2V_iV_j\cos\theta_{ij}\right] \tag{1}$$

where V_i and V_j are the voltage amplitudes of the interconnection node i and j, respectively; θ_{ij} means the voltage phase angle difference between node i and j; g_{ij} denotes the conductance between node i and j; N_{L} denotes the branch set of the power network.

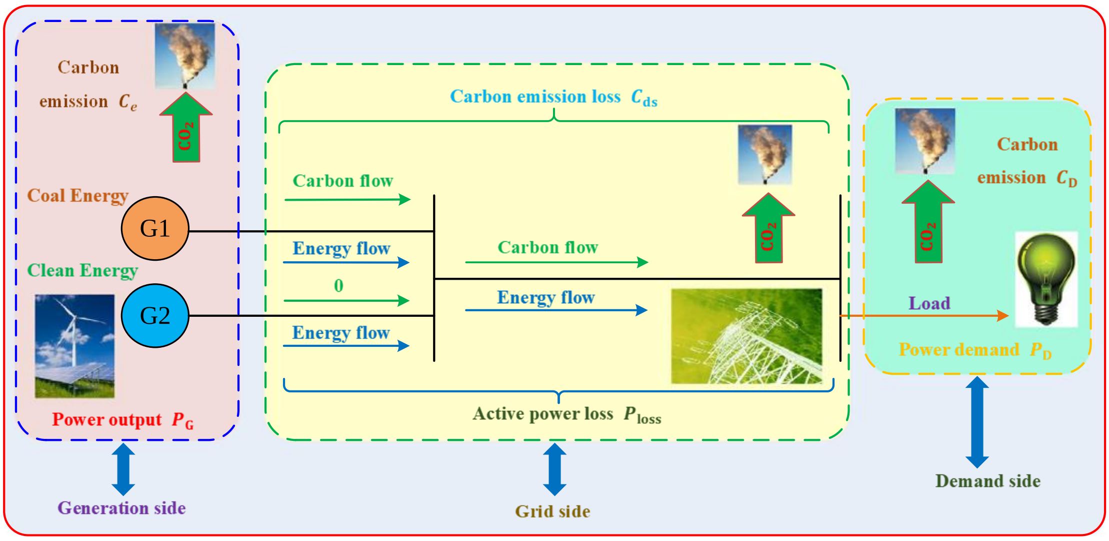


Figure 1. The carbon-energy combined-flow (CECF) structure in power systems.

In the process of power transmission, the energy flow should bear the corresponding amount of carbon flow losses. The tracking of the grid carbon emission flow is based on load flow tracking, and the source of network loss is traced in light of the proportional sharing rule [35]. The ratio of the *w*th generator to the whole active power injected at node *j* is

$$\beta_{wj} = \frac{a_{jw}^{(-1)} P_{sw}}{P'_{nj}} \tag{2}$$

where P_{sw} is the active output of the *w*th generator; P'_{nj} represents the whole active power injection of the *j* node in the equivalent lossless network; $a_{jw}^{(-1)}$ means the active power injection weight of the *w*th generator at node *j*, its specific derivation process can be found in [23].

The proportion of the *w*th generator outgoing line at node *j* is the same, and the line loss is decomposed according to the utilization share of the carbon source to the line. Hence, β_{wj} is the component ratio of the active power losses of the *w*th generator in line *i*–*j*. Here, the active power losses of line *i*–*j* can be expressed as follows:

$$\Delta P_{ij} = \sum_{w \in W} \left(\frac{a_{jw}^{(-1)} \Delta P_{ij}}{P'_{nj}} \right) P_{sw} \tag{3}$$

where W denotes the generator set.

Therefore, the total carbon flow losses of the power grid can be described by

$$\mathrm{C}_{\mathrm{ds}} = \sum_{i,j \in N_L} \sum_{w \in W} \left(\frac{a_{jw}^{(-1)} \Delta P_{ij}}{P'_{nj}} \right) P_{sw} \delta_{sw} \tag{4}$$

where δ_{sw} denotes the carbon emission rate of the *w*th generator.

3.2. OCECF Model

The OCECF model aims to reduce the network losses and carbon flow losses as much as possible according to satisfying the constraints of the power grid and maintaining the stability of the power system voltage. Therefore, the OCECF model is able to describe as follows [23,36]:

$$\begin{cases} min\ \mu_1 f_1(x) + \mu_2 f_2(x) + (1 - \mu_1 - \mu_2)V_{\mathrm{d}} \\ s.t.P_{\mathrm{G}i} - P_{\mathrm{D}i} - V_i \sum\limits_{j\epsilon N_i} V_j\Big(g_{ij}\cos\theta_{ij} + b_{ij}\sin\theta_{ij}\Big) = 0 \\ Q_{\mathrm{G}i} - Q_{\mathrm{D}i} - V_i \sum\limits_{j\epsilon N_i} V_j\Big(g_{ij}\sin\theta_{ij} + b_{ij}\cos\theta_{ij}\Big) = 0 \\ P_{\mathrm{G}i}^{\min} \le P_{\mathrm{G}i} \le P_{\mathrm{G}i}^{\max}\ i \in N_{\mathrm{G}} \\ Q_{\mathrm{G}i}^{\min} \le Q_{\mathrm{G}i} \le Q_{\mathrm{G}i}^{\max}\ i \in N_{\mathrm{G}} \\ V_i^{\min} \le V_i \le V_i^{\max}\ i \in N_{\mathrm{B}} \\ Q_{\mathrm{C}i}^{\min} \le Q_{\mathrm{C}i} \le Q_{\mathrm{C}i}^{\max}\ i \in N_{\mathrm{C}} \\ k_{\mathrm{t}i}^{\min} \le k_{\mathrm{t}i} \le k_{\mathrm{t}i}^{\max}\ i \in N_{\mathrm{k}} \\ |S_i| \le S_i^{\max}\ i \in N_{\mathrm{L}} \end{cases} \tag{5}$$

where nonlinear functions $f_1(x)$ and $f_2(x)$ are the components of carbon flow loss and active power loss; V_{d} is the voltage stability component; μ_1 and μ_2 are the weight coefficients, $\mu_1 \in [0,\ 1]$, $\mu_2 \in [0,\ 1]$, $\mu_1 + \mu_2 \le 1$; $x = [V,\ \theta,\ k_{\mathrm{t}},\ Q_{\mathrm{C}}]^{\mathrm{T}}$ corresponds to the voltage value of each node of the power grid V, the phase angle of each node θ and the on-load tap changer (OTLC) ratio k_{t}, reactive power compensation Q_{C}. The remaining variables can be referenced in the nomenclature and V_{d} can be described as [23]

$$V_{\mathrm{d}} = \sum_{j=1}^{n}\left|\frac{2V_j - V_{j\max} - V_{j\min}}{V_{j\max} - V_{j\min}}\right| \tag{6}$$

where n represents the number of load nodes; V_j is the node voltage of load node j; and $V_{j\max}$ and $V_{j\min}$ denote the maximal and minimal voltage ranges of load node j, respectively.

4. MCR-Q(λ) Learning

4.1. *Q(λ) Learning*

Multi-step backtrack Q(λ) learning is a conventional algorithm of RL, in which Q-learning combines the idea multi-step TD(λ) returns [38] and introduces the eligibility trace, such that the convergence speed of the algorithm can be improved to a certain extent. The eligibility trace can be described as [38]

$$e_k(s,a) = \begin{cases} \gamma\lambda e_{k-1}(s,a) + 1, \text{ if } (s,a) = (s_k, a_k) \\ \gamma\lambda e_{k-1}(s,a), \text{otherwise} \end{cases} \tag{7}$$

where $e_k(s,a)$ stands for the eligibility trace under a state-action pair (s, a) corresponding to the kth iteration; (s_k, a_k) denotes the actual state-action pair of the kth iteration; γ means the discount factor; and λ represents the trace-decay factor.

The eligibility trace (λ) uses the "backward estimation" mechanism to approximate the optimal value function matrix Q^*, and sets Q_k as the kth iterative value of the estimated value Q^*, thus the value function of the algorithm can be updated iteratively as follows [39]:

$$\rho_k = R(s_k, s_{k+1}, a_k) + \gamma Q_k(s_{k+1}, a_{\mathrm{g}}) - Q_k(s_k, a_k) \tag{8}$$

$$\delta_k = R(s_k, s_{k+1}, a_k) + \gamma Q_k(s_{k+1}, a_{\mathrm{g}}) - Q_k(s_k, a_{\mathrm{g}}) \tag{9}$$

$$Q_{k+1}(s,a) = Q_k(s,a) + \alpha\delta_k e_k(s,a) \tag{10}$$

$$Q_{k+1}(s_k, a_k) = Q_{k+1}(s_k, a_k) + \alpha\rho_k \tag{11}$$

where α is the learning factor; $R(s_k, s_{k+1}, a_k)$ is the reward function value of the kth iterative time environment from state s_k to s_{k+1} through the selected action a_k; and a_{g} is the greedy action strategy, which also represents the action corresponding to the highest Q-value in the current state, which can be written by [39]

$$a_g = \underset{a \in A}{\operatorname{argmax}} Q_k(s_{k+1}, a) \tag{12}$$

where A represents the action set, which is also the alternative action set for each variable.

4.2. MCR-Q(λ) Learning

4.2.1. Reduced-Dimension of Solution Space

As shown in Figure 2, the traditional single-objective Q(λ) algorithm does not decompose the action space of all the variables. Assume that the ith variable x_i has m_i alternative solutions, the number of action set elements $|A| = m_1 m_2 \cdots m_n$, when the number of variables n is large, the alternative action combination will increase accordingly, which leads to a slow convergence and difficulties in the iterative calculation. Up to now, the most usual way to work out this "dimension disaster" issue is hierarchical reinforcement learning (HRL) [40]. However, it is difficult to determine the hierarchical design and connection, which usually leads to the convergence of the algorithm to the local optimal solution.

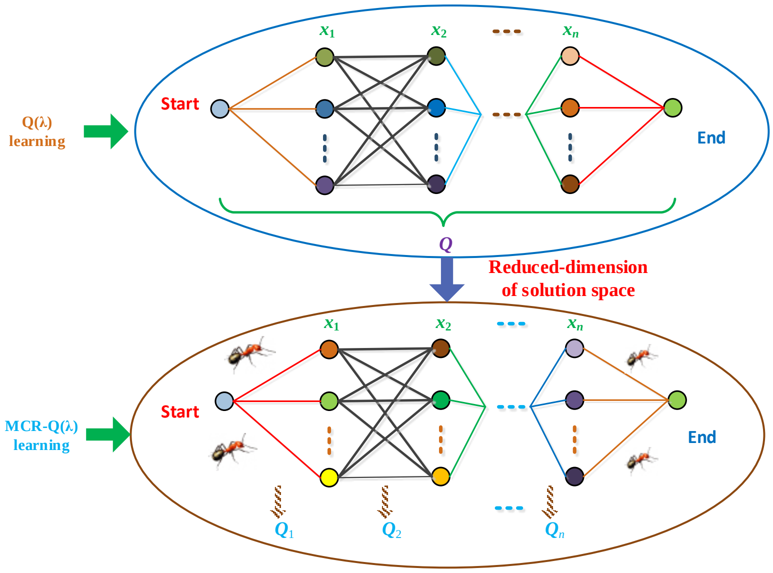


Figure 2. Difference between Q(λ) and MCR-Q(λ).

Under the framework of the proposed MCR-Q(λ) learning algorithm, each variable has a corresponding value function Q_i matrix, and the action set is respectively divided into $(A_1,\ A_2,\ \cdots,\ A_n)$ with $|A_i| = m_i$. In the iterative optimization of each Q matrix, the difficulty of optimization is greatly reduced due to the action space being obviously smaller. Meanwhile, the action space of each variable is the state space of the next variable, which enhances the internal relationship between variables, as can be illustrated in Figure 2. The state space of the first variable is divided according to the load scenario.

4.2.2. Multi-Agent Cooperative Search

In the iterative optimization of Q(λ) learning, which only employs a single agent for exploration and exploitation, the Q matrix is less efficient at updating just one element per iteration. On the contrary, in MCR-Q(λ) learning, there are multiple agents for exploration and exploitation at the same time, in which multiple elements of the Q matrix can be updated at each iteration, and the update speed of the Q matrix is greatly improved. Here, the value function of MCR-Q(λ) learning can be updated iteratively as follows [23]:

$$\rho_k^{ij} = R^{ij}\left(s_k^{ij}, s_{k+1}^{ij}, a_k^{ij}\right) + \gamma Q_k^i\left(s_{k+1}^{ij}, a_g^i\right) - Q_k^i\left(s_k^{ij}, a_g^i\right) \tag{13}$$

$$\delta_k^{ij} = R^{ij}\left(s_k^{ij}, s_{k+1}^{ij}, a_k^{ij}\right) + \gamma Q_k^i\left(s_{k+1}^{ij}, a_g^i\right) - Q_k^i\left(s_k^{ij}, a_g^i\right) \tag{14}$$

$$Q_{k+1}^i\left(s^i, a^i\right) = Q_k^i\left(s^i, a^i\right) + \alpha\delta_k^{ij} e_k^i\left(s^i, a^i\right) \tag{15}$$

$$Q_{k+1}^i\left(s_k^{ij}, a_k^{ij}\right) = Q_{k+1}^i(s_k, a_k) + \alpha\rho_k^{ij} \tag{16}$$

where the superscript i represents the ith variable or the ith Q-value matrix; the superscript j represents the jth objective; $e_k^i\left(s^i, a^i\right)$ and a_g^i are similar to Equations (7) and (12), respectively.

As with the Ant-Q algorithm, MCR-Q(λ) does not calculate the global reward function after each individual selects all the variables, i.e., from the start to the end, as shown in Figure 2. The reward function value can be calculated as follows [24]:

$$R^{ij}\left(s_k^{ij}, s_{k+1}^{ij}, a_k^{ij}\right) = \begin{cases} \frac{W}{L_{Best}}, \text{ if } \left(s_k^{ij}, a_k^{ij}\right) \in SA_{Best} \\ 0, \text{ otherwise} \end{cases} \tag{17}$$

where L_{Best} represents the function value of an individual (i.e., the best individual) that has the lowest value of the objective function value at the kth iteration; W is a positive constant; SA_{Best} denotes the state-action pair set of the optimal individual executed at the kth iteration.

4.2.3. Action Selections

As all individuals are exploring and learning, they are faced with action selections. When the individual j prepares to determine the variable x_i, its action selection is based on the following equation [41]:

$$a_{k+1}^{ij} = \begin{cases} \underset{a^i \in A_i}{\operatorname{argmax}} Q_{k+1}^i\left(s_{k+1}^{ij}, a^i\right), \text{ if } q \le q_0 \\ a_S, \text{ otherwise} \end{cases} \tag{18}$$

where q is a random number; q_0 is a positive constant for determining the probability of a pseudo-random selection; a_s denotes the action determined by the pseudo-random selection. In this paper, the rotary selection method is adopted to determine the action to be selected according to the P_k^i distribution of the action probability matrix, and the probability matrix is calculated as follows:

$$P_{k+1}^i\left(s_{k+1}^{ij}, a_{k+1}^i\right) = \frac{Q_{k+1}^i\left(s_{k+1}^{ij}, a_{k+1}^i\right)}{\sum_{a^i \in A_i} Q_{k+1}^i\left(s_{k+1}^{ij}, a^i\right)} \tag{19}$$

When an individual finds the best value of the objective function, the probability of its state-action for the corresponding action will be increased, which will attract other individuals to perform the same action. When the algorithm converges, all individuals will perform the same state-action pair when selecting all variables from the start to the end.

5. OCECF Based on MCR-Q(λ) Learning

5.1. Design of State and Action

As mentioned above, the action space of each variable is designed to be the state space of the next variable, in which the state space of the first variable is designed to be the state set of the environment (i.e., the power grid). For OCECF, the power grid load scenario can be designed as the state of the first variable, where a load scenario is divided at every 15 min and the scenarios with similar loads are set to the same state, e.g., the power grid load scenarios with different loads at 11:00 a.m. and 11:15 a.m. can be regarded as two different states.

In addition, OCECF mainly optimizes the carbon emissions on the power grid side, and the variables in the model are mainly divided into two categories: (a) reactive power compensation device and (b) the OTLC ratio. Thus, the action set corresponding to each variable is a discrete optional action of the reactive power compensation quantity or transformer changer ratio.

5.2. Design of Reward Function

As shown in Equation (17), L_{Best} represents the optimal objective function value of all individuals. According to the OCECF model described by Equation (5), the inequality constraint is brought in by the objective function, and then the objective function value obtained by the individual j becomes [41]

$$L^j = \mu_1 f_1(x^j) + \mu_2 f_2(x^j) + (1 - \mu_1 - \mu_2) V_d^j + N^j \tag{20}$$

$$L_{Best} = \min_{j \in J} L^j \tag{21}$$

where N^j denotes the number of unsatisfied inequality constraints calculated by the power flow after the individual j determines the variable, and J is the number of groups.

5.3. Parameter Setting

In MCR-Q(λ) learning, six parameters γ, λ, α, q_0, J and W, have great influence on the effect of the algorithm [36]. After a large number of simulation tests using trial-and-error, all the parameters can be set as indicated in Table 1.

5.4. Algorithm Flow of the OCECF

Generally speaking, the algorithm flow of OCECF based on MCR-Q(λ) learning is shown in Algorithm 1.

Algorithm 1 Flow of MCR-Q(λ) Learning for OCECF

```
Initialization: functions $Q^I$, action probability $P^i$, eligibility trace matrices $e^i$, and $i = 1, 2, \cdots, n$;
Input power flow calculation result;
Calculate fitness values of all individuals;
Set $k: = 0$;
WHILE $k < k_{\max}$;
  FOR $i = 1$ to $n$
    According to Equations (18) and (19), individual $j$ selects the corresponding action $a_k^i$ of each variable in turn and records the next state;
    Calculate power flow for all variables $x$ determined by individuals;
  END FOR
  According to Equations (1) and (4)–(6) respectively calculate the linear loss $P_{\mathrm{loss}}$, the carbon loss $C_{\mathrm{ds}}$, the number of constraints $N$ of dissatisfaction inequality, and the voltage stable component $V_{\mathrm{d}}$;
  Calculate the reward function $R^{ij}$ from Equations (17)–(21);
  Update the $Q$-value functions by Equations (13)–(16);
END WHILE
Output: optimal variable $x$ and corresponding optimal function value.
```

Table 1. Parameter setting of MCR-Q(λ) learning.

Parameters	Range	Value
γ	$0 < \gamma < 1$	0.1
λ	$0 < \lambda < 1$	0.5
α	$0 < \alpha < 1$	0.1
q_0	$0 < q_0 < 1$	0.8
J	$J > 1$	20
W	$W > 0$	1

6. Case Studies

For purpose of testing the optimization performance of MCR-Q(λ) learning, the simulation results of Q(λ) learning, Q learning [41], quantum genetic algorithm (QGA) [42], GA [43], PSO [44], ant colony system (ACS) [45], group search optimizer (GSO) [46] and artificial bee colony (ABC) [47] were also introduced for comparison. Note that the weight coefficient in Equation (5) can be adjusted according to the preference on different components of the objective function. In the simulation analysis, since three components of the objective function in Equation (5) have the same preferences, and the weight coefficient in Equation (5) is set to be 1/3, both the testing IEEE 118-bus system and IEEE 300-bus system are referenced from the tool called MATPOWER [48], in which the detailed parameters can be found in [49]. Besides, it assumes that both the wind and solar energy outputs can be accurately acquired by using effective forecasting techniques, e.g., the deep long-short-term memory recurrent neural network [50]. Among them, the algorithms are simulated and tested in Matlab 2016b by a personal computer with an Intel(R) Core TM i5-4210 CPU at 2.6 GHz with 8 GB of RAM.

6.1. Case Study of IEEE 118-Bus System

6.1.1. Simulation Model

According to different generator types, the carbon emission rate δ_{sw} of each unit in the IEEE 118-bus system is summarized in Table 2. Besides, this paper adopts the same benchmark model of IEEE 118-bus system in all case studies, related detail parameters can be referenced in [36].

Moreover, the system load of the IEEE 118-bus system is mainly divided into five scenarios, as shown in Table 3. Particularly, the scenarios from 1 to 5 represent the system with different load demands, where the load demand gradually increases from scenarios 1 to 5 for all the presented nodes in Table 3. As mentioned above, Tables 2 and 3 are obtained under the same benchmark model of IEEE 118-bus system [36].

In fact, reactive power compensation can be designed for the nodes with generators or load demand to provide adequate reactive power, while the OLTC ratio can be selected for the line with two different voltage nodes. According to this rule, the reactive power compensation of nodes 45, 79, and 105, and the OLTC ratio of lines 8–5, 26–25, 30–17, 63–59, and 64–61 are respectively selected as controllable variables, which are defined in sequence as (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8), with

(1) The reactive power compensation is divided into five configurations as {−40%, −20%, 0%, 20%, 40%} with its reference value;

(2) The OLTC ratio is divided into three grades, which are {0.98, 1.00, 1.02}.

Hence, the optimization variables of the IEEE 118-bus system can be found in Table 4, where the variables can be divided into two types, i.e., the reactive power compensation and OLTC ratio; the “no. of bus” represents the location of each variable in the power network; the “action space” denotes the set of the alternative control actions for each variable; and the “variable number” is the number of all the optimization variables.

Table 2. Carbon emission rate of the IEEE 118-bus system.

Generator Node	Generator Type	δ_{sw} (kg/kW·h)	Generator Node	Generator Type	δ_{sw} (kg/kW·h)
1	Gas	0.5	65	Hydro	0
4	Hydro	0	66	Wind	0
6	Coal	1.06	69	Gas	0.5
8	Coal	1.01	70	Hydro	0
10	Coal	0.95	72	Coal	1.06
12	Coal	1.5	73	Coal	1.01
15	Coal	0.7	74	Coal	0.95
18	Gas	0.5	76	Coal	1.5
19	Hydro	0	77	Coal	0.7
24	Hydro	0	80	Hydro	0
25	Coal	1.01	85	Hydro	0
26	Coal	0.95	87	Gas	0
27	Coal	1.5	89	Wind	0
31	Wind	0	90	Gas	1.01
32	Coal	1.06	91	Coal	0.95
34	Coal	1.01	92	Coal	1.5
36	Coal	0.95	99	Coal	0
40	Coal	1.5	100	Hydro	0
42	Coal	0.7	103	Hydro	0
46	Hydro	0	104	Gas	1.06
49	Hydro	0	105	Coal	1.01
54	Gas	0.5	107	Coal	0.95
55	Photovoltaic	0	110	Coal	1.5
56	Coal	1.01	111	Coal	0.7
59	Coal	0.95	112	Coal	0
61	Coal	1.5	113	Hydro	0
62	Hydro	0	116	Hydro	0

Table 3. Load statistical conditions employed in five scenarios.

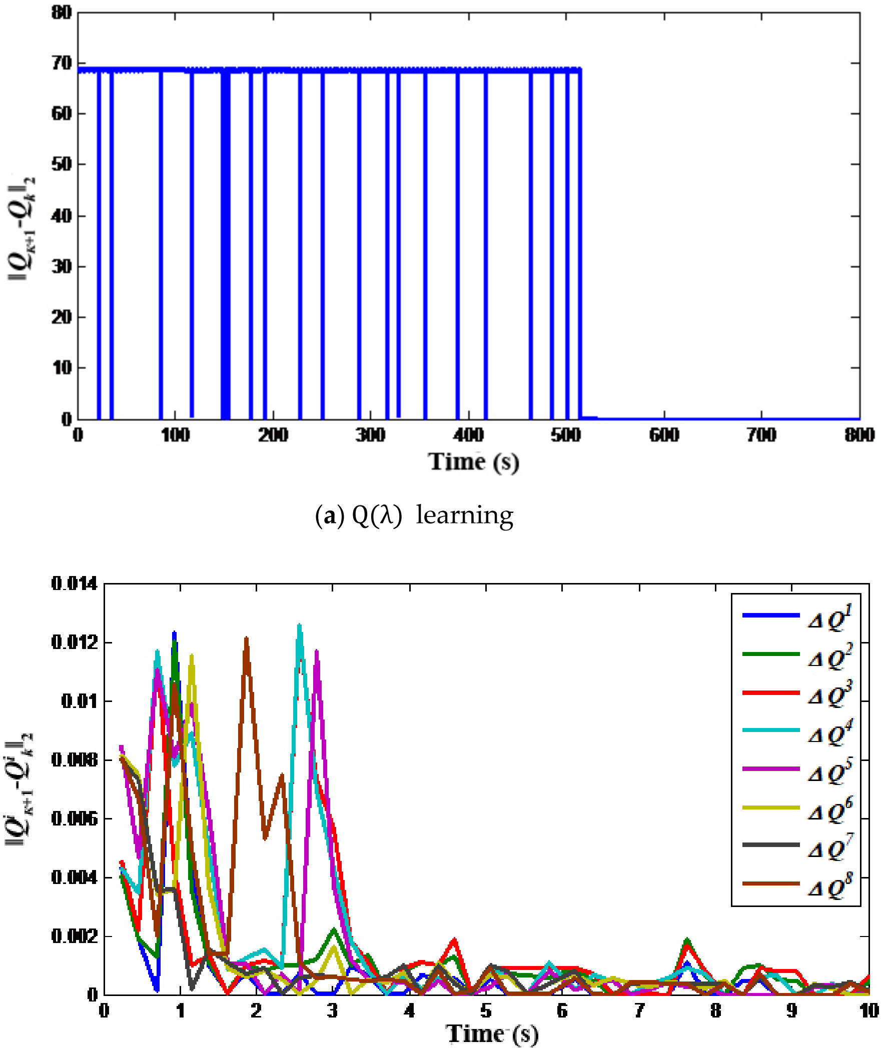
Scenarios	Active Power (MW)				
	Node 54	Node 59	Node 80	Node 90	Node 116
1	91	221	105	131	148
2	102	249	118	147	166
3	113	277	131	163	184
4	124	305	144	179	202
5	135	333	157	192	220

Table 4. Optimization variables of the IEEE 118-bus system.

Variable Type	Number of Bus	Action Space	Variable Number
Reactive power compensation	45, 79, 105	{−40%, −20%, 0%, 20%, 40%}	3
OLTC ratio	8–5, 26–25, 30–17, 63–59, 64–61	{0.98, 1.00, 1.02}	5

6.1.2. Convergence Analysis

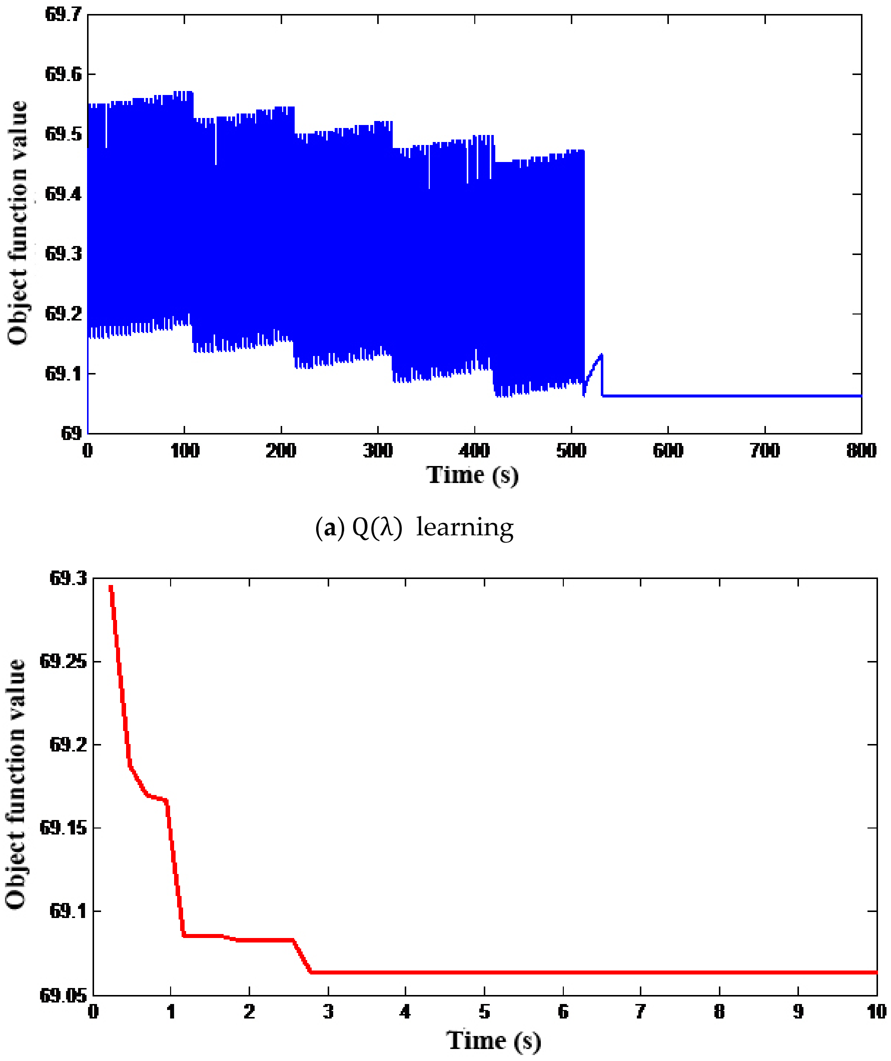
Figure 3 illustrates the convergence process of the Q-value deviation between Q(λ) learning and MCR-Q(λ) learning under scenario 1, where the Q-value deviation is defined as the 2-norm of matrix $(Q_{k+1} - Q_k)$, that is, $\|Q_{k+1} - Q_k\|_2$. As obtained from Figure 3a, since the Q matrix of single-objective Q(λ) learning is large and the updating speed is slow, the algorithm can converge to the optimal Q^* matrix through a variety of trial-and-error explorations, while the convergence time is about 530s. In contrast, after reducing the dimension of the solution space of MCR-Q(λ) learning, the Q^i matrix corresponding to each variable is very small, and 20 objectives are updated at the same time. The optimization speed is more than 100 times of that of Q(λ) learning, which can converge after about 3.5 s, as shown in Figure 3b. Moreover, it can be obtained from the convergence of the objective function values in Figure 4 that the optimization speed of MCR-Q(λ) learning is much faster, and both algorithms can converge to the global optimal solution.



(**a**) Q(λ) learning

(**b**) MCR-Q(λ) learning

Figure 3. Q-value difference convergence.



(**a**) Q(λ) learning

(**b**) MCR-Q(λ) learning

Figure 4. Convergence process of the objective function value.

When MCR-Q(λ) learning converges, the value function matrix Q^i and probability matrix P^i corresponding to all variables will prefer a state-action pair, and all individuals will tend to be consistent in selecting the action, as demonstrated in Figure 5.

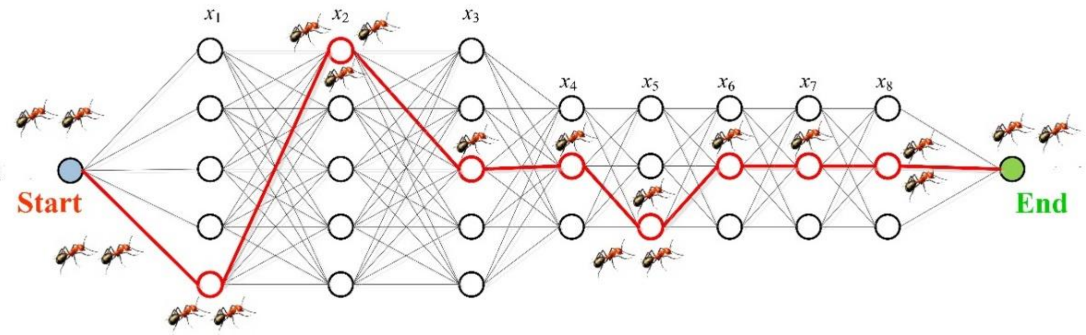


Figure 5. Convergent results of state-action pairs by MCR-Q(λ) learning.

6.1.3. Comparative Analysis of Simulation Results

For the purpose of evaluating the optimization capability of MCR-Q(λ) learning, this section applies all the algorithms to solve the OCECF model for 10 repetitions. For each method, the objective function value is directly taken to evaluate the quality of a solution during the searching process, which is the most crucial index to evaluate the optimization performance.

Table 5 indicates the average convergence results of 10 repetitions for the different algorithms, and it can be found that:

(a) The optimal solution obtained by Q learning and Q(λ) learning is the best, but the optimization time is also the longest, which also shows the strong ergodicity of RL;

(b) The convergence objective value of MCR-Q learning and MCR-Q(λ) learning is the closest to Q learning and Q(λ) learning, and the convergence time is the shortest, while the convergence speed is about 100 times that of single-objective Q learning and Q(λ) learning;

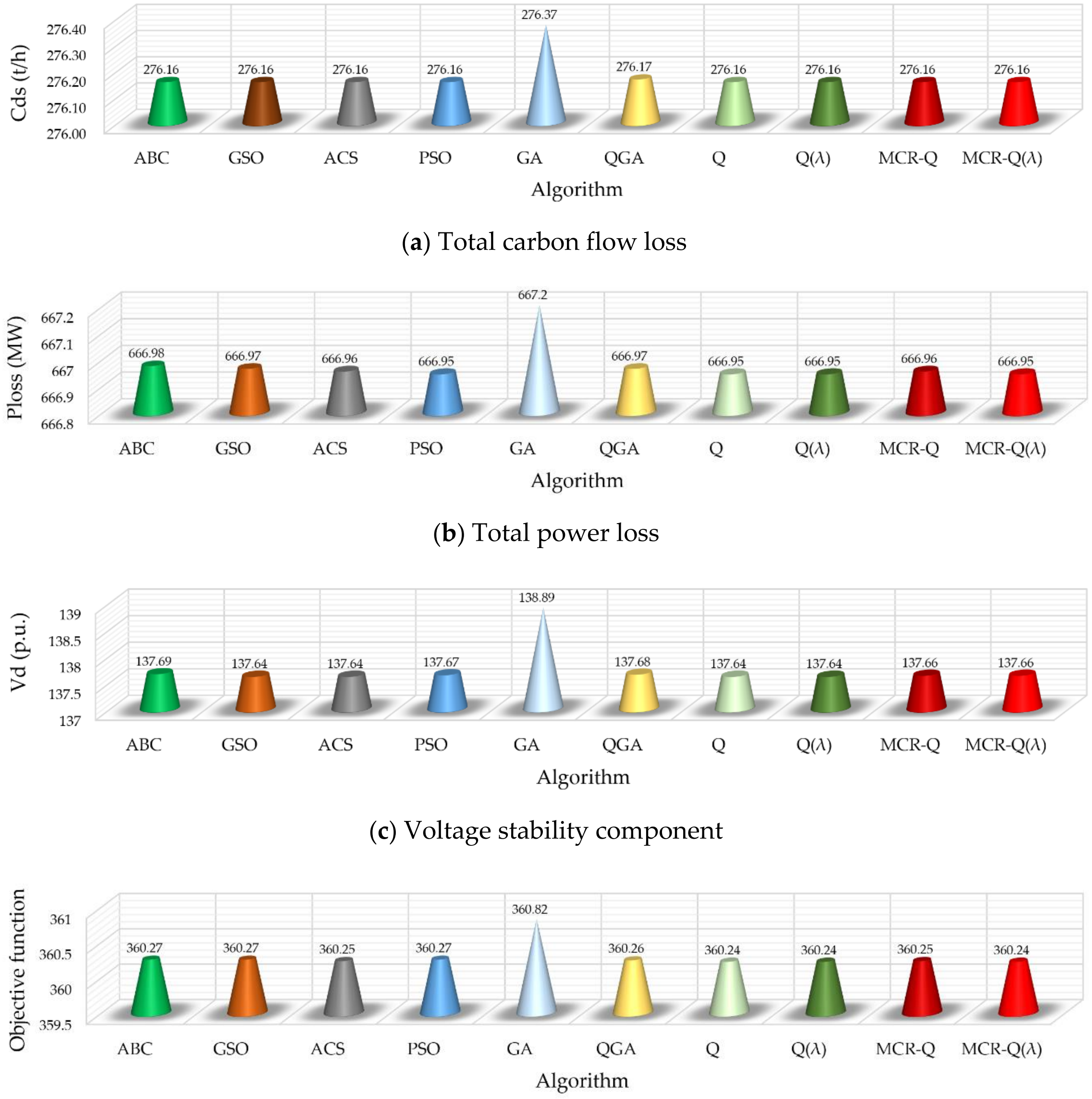
(c) RL improves the algorithmic speed by up to 37.13% with the introduction of the eligibility trace (λ) returns mechanism;

(d) With the increase in the load scenario, the line losses and carbon losses of the power grid will also increase correspondingly. However, since the power system has a sufficient reactive power supply, its voltage stability component just changes slightly.

Table 5. Average results of different algorithms on the IEEE 118-bus system in 10 runs.

Scenarios	Indexes	ABC	GSO	ACS	PSO	GA	QGA	Q	Q(λ)	MCR-Q	MCR-Q(λ)
1	Time (s)	55.08	13.30	13.68	31.44	17.14	20.53	660.00	608.00	5.75	5.27
	C_{ds} (t/h)	50.71	50.71	50.71	50.71	50.77	50.71	50.71	50.71	50.71	50.71
	P_{loss} (MW)	128.85	128.85	128.85	128.85	128.91	128.85	128.85	128.85	128.85	128.85
	V_{d}	27.65	27.63	27.63	27.64	27.86	27.65	27.63	27.63	27.63	27.64
	Objective	69.07	69.07	69.06	69.07	69.18	69.07	69.06	69.06	69.06	69.06
2	Time (s)	65.73	15.83	8.93	29.72	16.44	16.52	646.00	450.00	4.14	3.43
	C_{ds} (t/h)	52.69	52.69	52.69	52.69	52.73	52.70	52.69	52.69	52.69	52.69
	P_{loss} (MW)	130.24	130.23	130.23	130.23	130.28	130.24	130.23	130.23	130.23	130.23
	V_{d}	27.58	27.56	27.56	27.57	27.70	27.58	27.56	27.56	27.57	27.57
	Objective	70.17	70.16	70.16	70.17	70.23	70.17	70.16	70.16	70.17	70.16
3	Time (s)	36.75	12.66	23.69	49.40	15.57	12.35	671.00	445.00	4.92	3.09
	C_{ds} (t/h)	54.92	54.92	54.92	54.92	54.95	54.92	54.92	54.92	54.92	54.92
	P_{loss} (MW)	132.50	132.50	132.49	132.49	132.53	132.49	132.49	132.49	132.49	132.49
	V_{d}	27.52	27.52	27.52	27.53	27.74	27.52	27.52	27.52	27.53	27.52
	Objective	71.65	71.65	71.64	71.65	71.74	71.64	71.64	71.64	71.64	71.64
4	Time (s)	44.11	16.65	10.16	52.77	15.93	14.33	663.00	447.00	4.70	4.30
	C_{ds} (t/h)	57.48	57.48	57.48	57.48	57.52	57.48	57.48	57.48	57.48	57.48
	P_{loss} (MW)	135.66	135.66	135.66	135.66	135.72	135.66	135.66	135.66	135.66	135.66
	V_{d}	27.49	27.48	27.48	27.48	27.85	27.48	27.48	27.48	27.48	27.48
	Objective	73.54	73.54	73.54	73.54	73.70	73.54	73.54	73.54	73.54	73.54
5	Time (s)	26.43	18.41	7.67	42.65	14.27	12.92	658.00	441.00	6.37	5.01
	C_{ds} (t/h)	60.36	60.36	60.36	60.36	60.40	60.36	60.36	60.36	60.36	60.36
	P_{loss} (MW)	139.73	139.73	139.73	139.72	139.76	139.73	139.72	139.72	139.73	139.72
	V_{d}	27.45	27.45	27.45	27.45	27.74	27.45	27.45	27.45	27.45	27.45
	Objective	75.84	75.85	75.85	75.84	75.97	75.84	75.84	75.84	75.84	75.84

Figure 6 gives the results comparison between different methods, where each value is the average of the sum value of five scenarios in 10 runs. It is obvious that the result obtained by GA is the worst among all the methods due to its premature convergence. On the other hand, the proposed MCR-Q(λ) learning only has a slight improvement on each index compared with the other methods, but it also can obtain the lowest total carbon flow loss and objective function. It verifies that the proposed method can effectively satisfy the low-carbon requirement from the viewpoint of power networks.



(**a**) Total carbon flow loss

(**b**) Total power loss

(**c**) Voltage stability component

(**d**) Objective function

Figure 6. Comparison of results obtained by different methods in the IEEE 118-bus system.

Lastly, Table 6 gives the statistic convergence results of 10 repetitions for the different algorithms, and it can be found that:

(a) The Q learning and Q(λ) learning have the highest convergence stability and can converge to the global optimal solution every time;

(b) The statistical variance and standard deviation of MCR-Q(λ) learning are the closest to Q learning and Q(λ) learning, which have a relatively high convergence stability;

(c) Except RL, other algorithms are more likely to trap at a local optimum because of the parameter setting and the lack of learning ability.

Table 6. Distribution statistics of the objective function under different algorithms in the IEEE 118-bus system in 10 runs.

Scenarios	Criteria	ABC	GSO	ACS	PSO	GA	QGA	Q	Q(λ)	MCR-Q	MCR-Q(λ)
1	Best	69.06	69.06	69.06	69.06	69.06	69.06	69.06	69.06	69.06	69.06
	Worst	69.09	69.08	69.07	69.09	69.36	69.11	69.06	69.06	69.06	69.07
	Variance	1.2×10^{-4}	2.7×10^{-5}	5.9×10^{-6}	5.5×10^{-5}	8.4×10^{-3}	2.2×10^{-4}	0	0	0	1.6×10^{-6}
	Standard deviation	1.1×10^{-2}	5.2×10^{-3}	2.4×10^{-3}	7.4×10^{-3}	9.1×10^{-2}	1.5×10^{-2}	0	0	0	1.3×10^{-3}
2	Best	70.16	70.16	70.16	70.16	70.16	70.16	70.16	70.16	70.16	70.16
	Worst	70.22	70.17	70.16	70.19	70.33	70.20	70.16	70.16	70.20	70.17
	Variance	3.0×10^{-4}	5.7×10^{-6}	0	5.5×10^{-5}	4.1×10^{-3}	2.3×10^{-4}	0	0	1.5×10^{-4}	2.8×10^{-6}
	Standard deviation	1.7×10^{-2}	2.4×10^{-3}	0	7.4×10^{-3}	6.4×10^{-2}	1.5×10^{-2}	0	0	1.2×10^{-2}	1.7×10^{-3}
3	Best	71.64	71.64	71.64	71.64	71.64	71.64	71.64	71.64	71.64	71.64
	Worst	71.66	71.65	71.66	71.69	71.96	71.65	71.64	71.64	71.65	71.64
	Variance	2.4×10^{-5}	1.3×10^{-5}	2.3×10^{-5}	2.7×10^{-4}	1.0×10^{-2}	5.6×10^{-6}	0	0	8.2×10^{-6}	0
	Standard deviation	5.3×10^{-3}	3.6×10^{-3}	4.8×10^{-3}	1.6×10^{-2}	1.0×10^{-1}	2.4×10^{-3}	0	0	2.9×10^{-3}	0
4	Best	73.54	73.54	73.54	73.54	73.54	73.54	73.54	73.54	73.54	73.54
	Worst	73.57	73.55	73.55	73.54	73.87	73.54	73.54	73.54	73.54	73.54
	Variance	7.6×10^{-5}	5.7×10^{-6}	2.3×10^{-5}	0	1.0×10^{-2}	0	0	0	0	0
	Standard deviation	8.7×10^{-3}	2.4×10^{-3}	4.8×10^{-3}	0	1.0×10^{-1}	0	0	0	0	0
5	Best	75.84	75.84	75.84	75.84	75.84	75.84	75.84	75.84	75.84	75.84
	Worst	75.85	75.85	75.86	75.84	76.12	75.85	75.84	75.84	75.85	75.84
	Variance	5.7×10^{-6}	1.3×10^{-5}	2.6×10^{-5}	0	8.7×10^{-3}	1.6×10^{-6}	0	0	5.7×10^{-6}	0
	Standard deviation	2.4×10^{-3}	3.6×10^{-3}	5.1×10^{-3}	0	9.3×10^{-2}	1.3×10^{-3}	0	0	2.4×10^{-3}	0

6.2. Case Study of the IEEE 300-Bus System

6.2.1. Simulation Model

According to different generator types, the carbon emission rate δ_{sw} of each unit in the IEEE 300-bus system is summarized in Table 7. Besides, 96 different load scenarios are designed to simulate different optimization tasks in a day for the IEEE 300-bus system, as shown in Figure 7. Moreover, the optimization variables are given in Table 8.

Table 7. Carbon emission rate of the IEEE 300-bus system.

Generator Node	Generator Type	δ_{sw} (kg/kWh)	Generator Node	Generator Type	δ_{sw} (kg/kWh)	Generator Node	Generator Type	δ_{sw} (kg/kWh)
8	Hydro	0	171	Hydro	0	7002	Hydro	0
10	Photovoltaics	0	176	Hydro	0	7003	Coal	1.06
20	Coal	1.01	177	Hydro	0	7011	Coal	1.5
63	Coal	0.95	185	Coal	1.01	7012	Coal	0.7
76	Coal	1.5	186	Coal	0.95	7017	Photovoltaics	0
84	Coal	0.7	187	Coal	1.5	7023	Gas	0.5
91	Coal	0.95	190	Hydro	0	7024	Hydro	0
92	Coal	1.5	191	Hydro	0	7039	Wind	0
98	Coal	0.7	198	Hydro	0	7044	Coal	1.5
108	Hydro	0	213	Hydro	0	7049	Coal	0.7
119	Gas	0.5	220	Wind	0	7055	Hydro	0
124	Coal	1.06	221	Gas	0.5	7057	Wind	0
125	Coal	1.01	222	Coal	1.06	7061	Coal	1.06
138	Hydro	0	227	Coal	1.01	7062	Coal	1.01
141	Hydro	0	230	Coal	0.95	7071	Coal	1.01
143	Coal	1.06	233	Coal	1.5	7130	Hydro	0
146	Coal	1.01	236	Coal	0.7	7139	Hydro	0
147	Coal	0.95	238	Coal	0.95	7166	Coal	0.7
149	Coal	1.5	239	Hydro	0	9002	Gas	0.5
152	Hydro	0	241	Hydro	0	9051	Coal	1.06
153	Photovoltaics	0	242	Coal	0.95	9053	Coal	1.01
156	Coal	1.06	243	Coal	1.5	9054	Hydro	0
170	Coal	0.95	7001	Coal	0.95	9055	Photovoltaics	0

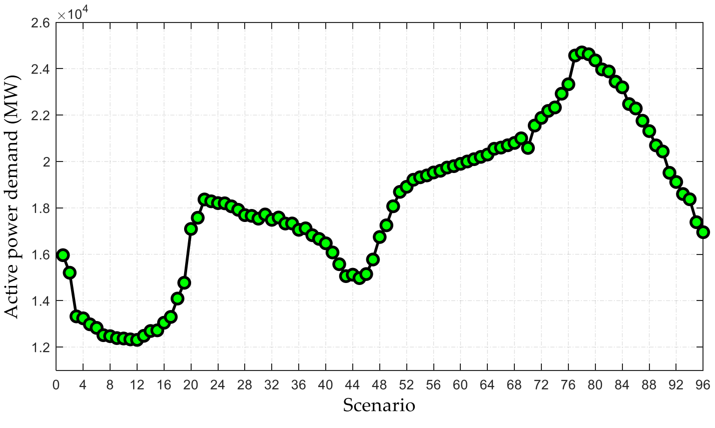


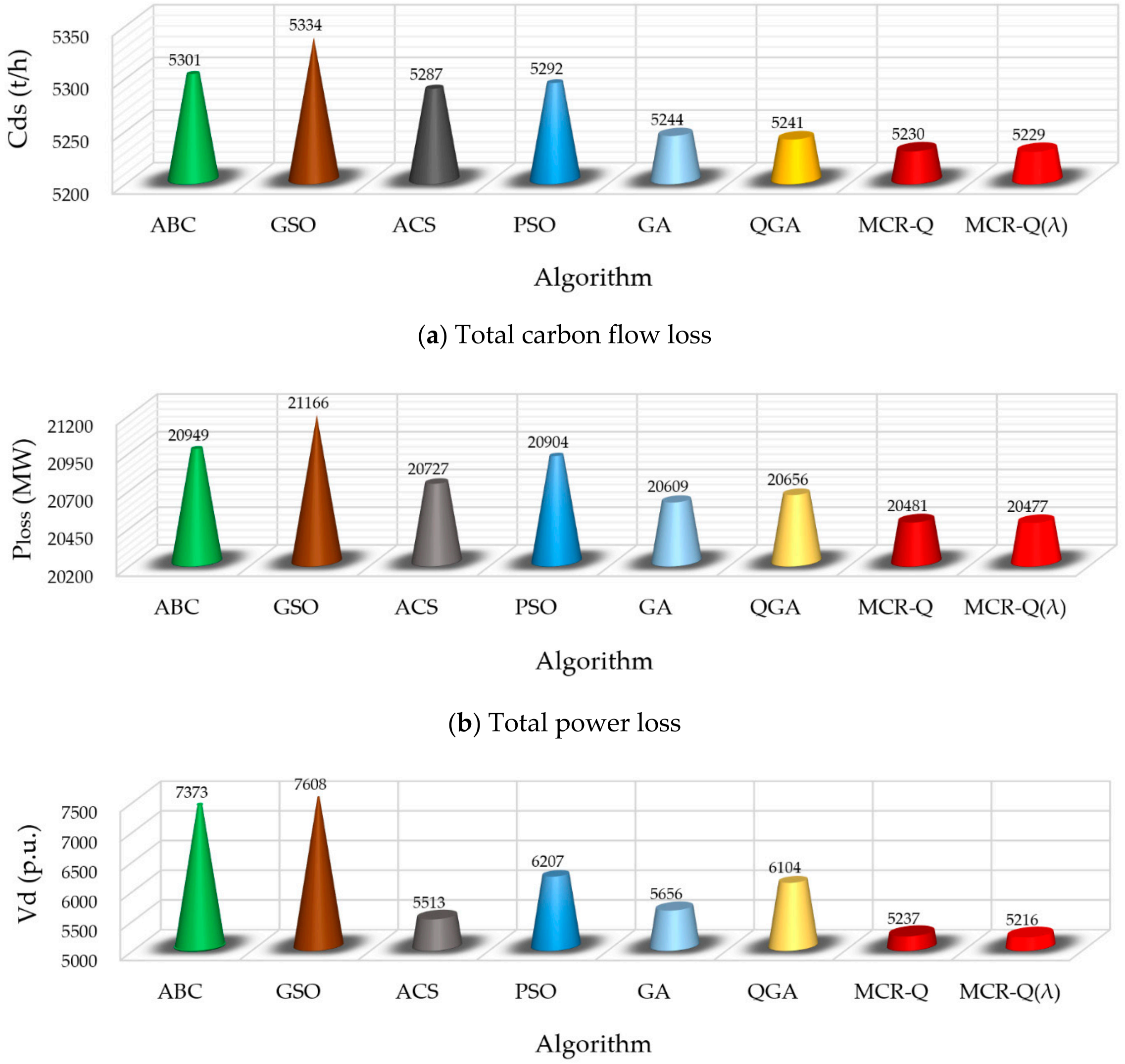
Figure 7. The load scenarios of the IEEE 300-bus system.

Table 8. Optimization variables of the IEEE 300-bus system.

Variable Type	Number of Bus	Action Space	Variable Number
Reactive power compensation	117, 120, 154, 164, 166, 173, 190, 231, 238, 240, 248	{−40%, −20%, 0%, 20%, 40%}	11
OLTC ratio	9021–9022, 9002–9024, 9023–9025, 9023–9026, 9007–9071, 9007–9072, 9003–9031, 9003–9032, 9003–9033, 9004–9041, 9004–9042, 9004–9043, 9003–9034, 9003–9035, 9003–9036, 9003–9037, 9003–9038, 213–214, 222–237, 227–231, 241–237, 45–46, 73–74, 81–88, 85–99, 86–102, 122–157, 142–175, 145–180, 200–248, 211–212, 223–224, 196–2040, 7003–3, 7003–61, 7166–166, 7024–24, 7001–1, 7130–130, 7011–11, 7023–23, 7049–49, 7139–139, 7012–12	{0.98, 1.00, 1.02}	44

6.2.2. Comparative Analysis of Simulation Results

For the purpose of evaluating the optimization capability of MCR-Q(λ) learning, this section applies all the algorithms to solve the OCECF model for 10 runs. Since the number of optimization variables of the IEEE 300-bus system dramatically increases, the conventional Q and Q(λ) algorithms cannot implement an optimization due to the dimension disaster. Figure 8 provides the results comparison between different methods, where each value is the average of the sum value of a day in 10 runs. It can be found that the proposed MCR-Q(λ) learning significantly outperforms other methods on the total carbon flow loss, total power loss, voltage stability component and the objective function. Hence, the MCR-Q(λ) learning-based OCECF can achieve a low-carbon operation for the power network. Particularly, these values obtained by MCR-Q(λ) learning are 2.0%, 3.4%, 45.9% and 10.3% lower than that obtained by GSO. It verifies that the optimization performance of MCR-Q(λ) is much better than other conventional meta-heuristic algorithms as the system scale increases.

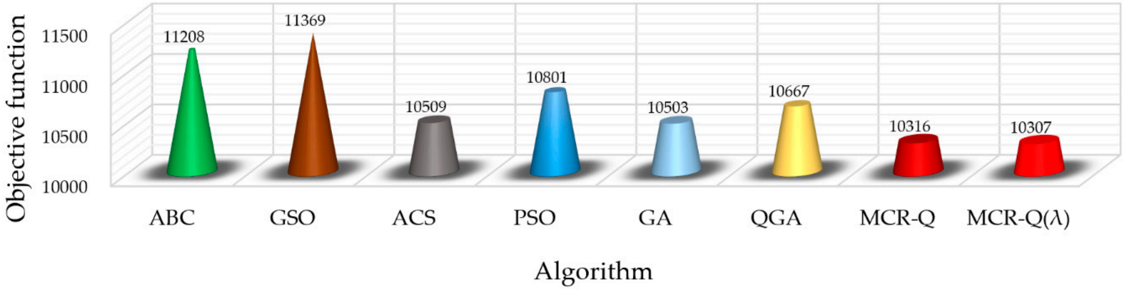


(**a**) Total carbon flow loss

(**b**) Total power loss

(**c**) Voltage stability component

Figure 8. *Cont.*



(**d**) Objective function

Figure 8. Comparison of results obtained by different methods in the IEEE 300-bus system.

Besides, Table 9 gives the distribution statistics of the objective function under different algorithms in the IEEE 300-bus system, where each value is the sum value of the objective function of a day in 10 runs; the best, worst, variance and standard deviation (Std. Dev.) are calculated to evaluate the convergence stability [51]. It can be seen from Table 9 that the convergence stability of MCR-Q(λ) learning is the highest among all the methods with the smallest variance and standard deviation of the objective function.

Table 9. Distribution statistics of the objective function under different algorithms in the IEEE 300-bus system in 10 runs.

Criteria	ABC	GSO	ACS	PSO	GA	QGA	MCR-Q	MCR-Q(λ)
Best	11,182.97	11,328.38	10,505.58	10,795.73	10,495.03	10,658.28	10,312.84	10,305.54
Worst	11,229.61	11,404.35	10,513.40	10,812.54	10,509.03	10,675.34	10,320.86	10,308.30
Variance	246.73	541.79	10.25	45.62	23.96	32.57	10.12	1.20
Standard deviation	15.71	23.28	3.20	6.75	4.89	5.71	3.18	1.09

7. Conclusions

This paper builds an OCECF model to optimize the carbon emission and energy losses of power grids simultaneously and proposes a new MCR-Q(λ) learning to solve this problem, which has the following four contributions/novelties:

(1) The OCECF model carefully considers the distribution of carbon flow in the power grid, which effectively resolves the carbon emission optimization at the power grid side;

(2) MCR-Q(λ) learning is proposed for the first time, which largely reduces the dimension of the solution space, and significantly accelerates the updating rate of the *Q*-value matrix via multi-agent cooperative exploration learning, such that the optimization speed can be considerably accelerated;

(3) Compared with Q(λ) learning, the convergence rate of MCR-Q(λ) learning can be increased by about 100 times, while a higher global convergence stability is guaranteed. Hence, it is very suitable for resolving dynamic OCECF in a large and complex power grid compared with other algorithms;

(4) Like ACO, MCR-Q(λ) learning is also suitable for solving various complex optimization problems.

To further improve the operation benefit of power grids, future works can focus on the carbon trading system-based optimal power flow and the Pareto-based multi-objective learning methods, while a decentralized optimization will be studied for high operation privacy and reliability.

Author Contributions: H.C. established the model, implemented the simulation and wrote this article; C.G. guided and revised the paper and refined the language; X.H. collected references; Y.L. guided the research; T.Y. assisted in writing algorithms. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Technical Projects of China Southern Power Grid grant number [GDK JXM20173256].

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

Nomenclature

P_{Gi}, Q_{Gi}	active and reactive power generation of the *i*th node
$\mathrm{P_{Di}}$, $\mathrm{Q_{Di}}$	active and reactive power demand of the *i*th node
V_i, V_j	voltage magnitude of the *i*th and *j*th node
b_{ij}	susceptance of line *i*–*j*
S_i	apparent power flow of the *i*th transmission line
N_i	node set
N_L	set of branches of the power network
N_G	set of units
$\mathrm{N_H}$	set of hydro units
N_B	set of PQ nodes
N_C	set of compensation equipment
N_K	set of on-load transformers
k_t	on-load tap changer ratio
Q_c	reactive power compensation
θ	phase angle of each node
V_d	component of voltage stability
$V_{j\min}$, $V_{j\max}$	minimum and maximum voltage limit of load node *j*
μ_1, μ_2	weight coefficients
W	generator set
(s_k, a_k)	actual state-action pair of the *k*th iteration
$\mathrm{\delta_k}$, $\mathrm{\rho_k}$	estimates of Q-function errors
$R(s_k, s_{k+1}, a_k)$	reward function value of the *k*th iterative time environment from state s_k to s_{k+1} through a selected action a_k
$\mathrm{a_g}$	greedy action strategy
A	action set
L_{Best}	function value of an individual (i.e., the best individual) that has the least value of the target function value at the *k*th iteration
SA_{Best}	state-action pair set of the best individual executed at the *k*th iteration
γ	discount factor
λ	trace-decay factor
α	learning factor
J	number of groups

Abbreviations

OCECF	optimal carbon-energy combined-flow
OTLC	on-load tap changer
MCR-Q(λ)	multi-agent cooperative reduced-dimension Q(λ)
HRL	hierarchical reinforcement learning
EED	environmental economic dispatch

References

1. Yang, B.; Jiang, L.; Wang, L.; Yao, W.; Wu, Q.H. Nonlinear maximum power point tracking control and modal analysis of DFIG based wind turbine. *Int. J. Electr. Power Energy Syst.* **2016**, *74*, 429–436. [CrossRef]
2. Yang, Y.D.; Song, A.J.; Liu, H.; Qin, Z.J.; Deng, J.; Qi, J.J. Parallel computing of multi-contingency optimal power flow with transient stability constraints. *Prot. Control Mod. Power Syst.* **2018**, *3*, 204–213. [CrossRef]
3. Yang, B.; Yu, T.; Zhang, X.S.; Li, H.F.; Shu, H.C.; Sang, Y.Y.; Jiang, L. Dynamic leader based collective intelligence for maximum power point tracking of PV systems affected by partial shading condition. *Energy Convers. Manag.* **2019**, *179*, 286–303. [CrossRef]

4. Yang, B.; Wang, J.B.; Zhang, X.S.; Yu, T.; Yao, W.; Shu, H.C.; Zeng, F.; Sun, L.M. Comprehensive overview of meta-heuristic algorithm applications on pv cell parameter identification. *Energy Convers. Manag.* **2020**, *208*, 112595. [CrossRef]
5. Yang, B.; Yu, T.; Shu, H.C.; Zhang, Y.M.; Chen, J.; Sang, Y.Y.; Jiang, L. Passivity-based sliding-mode control design for optimal power extraction of a PMSG based variable speed wind turbine. *Renew. Energy* **2018**, *119*, 577–589. [CrossRef]
6. Badal, F.R.; Das, P.; Sarker, S.K.; Das, S.K. A survey on control issues in renewable energy integration and microgrid. *Prot. Control Mod. Power Syst.* **2019**, *4*, 87–113. [CrossRef]
7. Li, Y.; Li, Y. Two-step many-objective optimal power flow based on knee point-driven evolutionary algorithm. *Processes* **2018**, *6*, 250. [CrossRef]
8. Li, G.Y.; Li, G.D.; Zhou, M. Comprehensive evaluation model of wind power accommodation ability based on macroscopic and microscopic indicators. *Prot. Control Mod. Power Syst.* **2019**, *4*, 215–226. [CrossRef]
9. Yang, B.; Yu, T.; Shu, H.C.; Dong, J.; Jiang, L. Robust sliding-mode control of wind energy conversion systems for optimal power extraction via nonlinear perturbation observers. *Appl. Energy* **2018**, *210*, 711–723. [CrossRef]
10. Ji, Z.; Kang, C.; Chen, Q.; Xia, Q.; Jiang, C.; Chen, Z. Low-carbon power system dispatch incorporating carbon capture power plants. *IEEE Trans. Power Syst.* **2013**, *28*, 4615–4623. [CrossRef]
11. Kuo, C.C.; Lee, C.Y.; Sheim, Y.C. Unit commitment with energy dispatch using a computationally efficient encoding structure. *Energy Convers Manag.* **2011**, *52*, 1575–1582. [CrossRef]
12. Ji, B.; Yuan, X.; Li, X.; Huang, Y.; Li, W. Application of quantum-inspired binary gravitational search algorithm for thermal unit commitment with wind power integration. *Energy Convers Manag.* **2014**, *87*, 589–598. [CrossRef]
13. Wall, T.; Stanger, R.; Santos, S. Demonstrations of coal-fired oxy-fuel technology for carbon capture and storage and issues with commercial deployment. *Int. J. Greenh. Gas Control* **2011**, *5*, S5–S15. [CrossRef]
14. Coninck, H.D.; Benson, S.M. Carbon Dioxide capture and storage: Issues and prospects. *Annu. Rev. Environ. Resour.* **2014**, *39*, 243–270. [CrossRef]
15. Chen, J.; Yao, W.; Zhang, C.K.; Ren, Y.; Jiang, L. Design of robust MPPT controller for grid-connected PMSG-Based wind turbine via perturbation observation based nonlinear adaptive control. *Renew. Energy* **2019**, *134*, 478–495. [CrossRef]
16. Giras, T.C.; Talukdar, S.N. Quasi-newton method for optimal power flows. *Int. J. Electr. Power Energy Syst.* **1981**, *3*, 59–64. [CrossRef]
17. Zhang, X.S.; Xu, Z.; Yu, T.; Yang, B.; Wang, H. Optimal mileage based AGC dispatch of a GenCo. *IEEE Trans. Power Syst.* **2020**, *35*, 2516–2526. [CrossRef]
18. Azzam, M.; Mousa, A.A. Using genetic algorithm and TOPSIS technique for multi-objective reactive power compensation. *Electr. Power Syst. Res.* **2010**, *80*, 675–681. [CrossRef]
19. Juan, L.I.; Yang, L.; Liu, J.L.; Yang, D.L.; Zhang, C. Multi-objective reactive power optimization based on adaptive chaos particle swarm optimization algorithm. *Power Syst. Prot. Control* **2011**, *39*, 26–31.
20. Han, P.P.; Fan, G.J.; Sun, W.Z.; Shi, B.L.; Zhang, X.A. Research on identification of LVRT characteristics of photovoltaic inverters based on data testing and PSO algorithm. *Processes* **2019**, *7*, 250. [CrossRef]
21. Yang, B.; Zhang, X.S.; Yu, T.; Shu, H.C.; Fang, Z.H. Grouped grey wolf optimizer for maximum power point tracking of doubly-fed induction generator based wind turbine. *Energy Convers. Manag.* **2017**, *133*, 427–443. [CrossRef]
22. Yang, B.; Zhong, L.E.; Yu, T.; Li, H.F.; Zhang, X.S.; Shu, H.C.; Sang, Y.Y.; Jiang, L. Novel bio-inspired memetic salp swarm algorithm and application to MPPT for PV systems considering partial shading condition. *J. Clean. Prod.* **2019**, *215*, 1203–1222. [CrossRef]
23. Yu, T.; Liu, J.; Chan, K.W.; Wang, J.J. Distributed multi-step Q(λ) learning for optimal power flow of large-scale power grids. *Int. J. Electr. Power Energy Syst.* **2012**, *42*, 614–620. [CrossRef]
24. Liana, M.; Roberto, S. The Ant-Q algorithm applied to the nuclear reload problem. *Ann. Nucl. Energy* **2002**, *29*, 1455–1470.
25. Zhao, X.-G.; Liang, L.; Meng, J.; Zhou, Y. An improved quantum particle swarm optimization algorithm for environmental economic dispatch. *Expert Syst. Appl.* **2020**, *152*, 113370.
26. Hagh, M.T.; Kalajahi, S.M.S.; Ghorbani, N. Solution to economic emission dispatch problem including wind farms using exchange market algorithm method. *Appl. Soft Comput. J.* **2020**, *88*, 106044. [CrossRef]

27. Ghasemi, A.; Gheydi, M.; Golkar, M.J.; Eslami, M. Modeling of wind/environment/economic dispatch in power system and solving via an online learning meta-heuristic method. *Appl. Soft Comput.* **2016**, *43*, 454–468. [CrossRef]
28. Srivastava, A.; Das, D.K. A new Kho-Kho optimization algorithm: An application to solve combined emission economic dispatch and combined heat and power economic dispatch problem. *Eng. Appl. Artif. Intell.* **2020**, *94*, 103763. [CrossRef]
29. He, L.; Lu, Z.; Geng, L.; Zhang, J.; Li, X.; Guo, X. Environmental economic dispatch of integrated regional energy system considering integrated demand response. *Int. J. Electr. Power Energy Syst.* **2020**, *116*, 105525. [CrossRef]
30. Zhang, R.; Yan, K.; Li, G.; Jiang, T.; Li, X.; Chen, H. Privacy-preserving decentralized power system economic dispatch considering carbon capture power plants and carbon emission trading scheme via over-relaxed ADMM. *Int. J. Electr. Power Energy Syst.* **2020**, *121*, 106094. [CrossRef]
31. Jin, J.; Zhou, P.; Li, C.; Guo, X.; Zhang, M. Low-carbon power dispatch with wind power based on carbon trading mechanism. *Energy* **2019**, *170*, 250–260. [CrossRef]
32. Tan, Q.; Ding, Y.; Ye, Q.; Mei, S.; Zhang, Y.; Wei, Y. Optimization and evaluation of a dispatch model for an integrated wind-photovoltaic-thermal power system based on dynamic carbon emissions trading. *Appl. Energy* **2019**, *253*, 113598. [CrossRef]
33. Kang, C.; Zhou, T.; Chen, Q. Carbon emission flow in network. *Sci. Rep.* **2012**, *2*, 479. [CrossRef] [PubMed]
34. Zhou, T.; Kang, C.; Qianyao, X.U.; Chen, Q. Preliminary Investigation on a Method for carbon emission flow calculation of power system. *Autom. Electr. Power Syst.* **2012**, *36*, 44–49.
35. Li, B.; Song, Y.; Hu, Z. Carbon flow tracing method for assessment of demand side carbon emissions obligation. *IEEE Trans. Sustain. Energy* **2013**, *4*, 1100–1107. [CrossRef]
36. Zhang, X.S.; Yu, T.; Yang, B.; Zheng, L.M.; Huang, L.N. Approximate ideal multi-objective solution Q(λ) learning for optimal carbon-energy combined-flow in multi-energy power systems. *Energy Convers. Manag.* **2015**, *106*, 543–556. [CrossRef]
37. Zhang, X.; Tan, T.; Zhou, B.; Yu, T.; Yang, B.; Huang, X. Adaptive distributed auction-based algorithm for optimal mileage based AGC dispatch with high participation of renewable energy. *Int. J. Electr. Power Energy Syst.* **2021**, *124*, 106371. [CrossRef]
38. Sutton, R.S. Learning to predict by the methods of temporal differences. *Mach. Learn.* **1988**, *3*, 9–44. [CrossRef]
39. Tao, Y.U.; Wang, Y.M.; Zhen, W.G.; Wenjia, Y.E.; Liu, Q.J. Multi-step backtrack Q-learning based dynamic optimal algorithm for auto generation control order dispatch. *Control Theory Appl.* **2011**, *28*, 58–64.
40. Ghavamzadeh, M.; Mahadevan, S. Hierarchical average reward reinforcement learning. *J. Mach. Learn. Res.* **2017**, *8*, 2629–2669.
41. Cao, H.Z.; Yu, T.; Zhang, X.S.; Yang, B.; Wu, Y.X. Reactive power optimization of large-scale power system: A transfer bees optimizer application. *Processes* **2019**, *7*, 321. [CrossRef]
42. Xiong, Y.; Chen, H.H.; Miao, F.Y.; Wang, X.F. A quantum genetic algorithm to solve combinatorial optimization problem. *Acta Electron. Sin.* **2004**, *32*, 1855–1858.
43. Kumari, M.S.; Maheswarapu, S. Enhanced genetic algorithm based computation technique for multi-objective optimal power flow solution. *Int. J. Electr. Power Energy Syst.* **2010**, *32*, 736–742. [CrossRef]
44. Hazra, J.; Sinha, A.K. A multi-objective optimal power flow using particle swarm optimization. *Eur. Trans. Electr. Power* **2011**, *21*, 1028–1045. [CrossRef]
45. Han, Y.; Shi, P. An improved ant colony algorithm for fuzzy clustering in image segmentation. *Neurocomputing* **2007**, *70*, 665–671. [CrossRef]
46. Basu, M. Group search optimization for solution of different optimal power flow problems. *Electr. Mach. Power Syst.* **2016**, *44*, 10. [CrossRef]
47. Karaboga, D.; Basturk, B. On the performance of artificial bee colony (ABC) algorithm. *Appl. Soft Comput.* **2008**, *8*, 687–697. [CrossRef]
48. Zimmerman, R.D.; Murillo-Sánchez, C.E.; Thomas, R.J. MATPOWER: Steady-state operations, planning and analysis tool for power systems research and education. *IEEE Trans. Power Syst.* **2011**, *26*, 12–19. [CrossRef]
49. MATPOWER—Free, Open-Source Tools for Electric Power System Simulation and Optimization. Available online: https://matpower.org/ (accessed on 21 May 2020).

50. Mahmoud, K.; Abdel-Nasser, M.; Mustafa, E.; Ali, Z.M. Improved salp-swarm optimizer and accurate forecasting model for dynamic economic dispatch in sustainable power systems. *Sustainability* **2020**, *12*, 576. [CrossRef]
51. Zhang, X.S.; Yu, T.; Yang, B.; Cheng, L.F. Accelerating bio-inspired optimizer with transfer reinforcement learning for reactive power optimization. *Knowl. Based Syst.* **2017**, *116*, 26–38. [CrossRef]