IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON 5G AND BEYOND MOBILE WIRELESS COMMUNICATIONS
ENABLING INTELLIGENT MOBILITY

Received August 24, 2019, accepted October 9, 2019, date of publication October 15, 2019, date of current version October 25, 2019.

*Digital Object Identifier 10.1109/ACCESS.2019.2947546*

# A Two-Step Environment-Learning-Based Method for Optimal UAV Deployment

**XINRAN LUO, YAN ZHANG, (Member, IEEE), ZUNWEN HE, (Member, IEEE), GUANSHU YANG, AND ZIJIE JI, (Student Member, IEEE)**
School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

Corresponding author: Yan Zhang (zhangy@bit.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61871035, and in part by the National High Technology Research and Development Program of China under Grant 2015AA01A708.

**ABSTRACT** Unmanned aerial vehicles (UAVs) can be used as low-altitude flight base stations to satisfy the coverage requirements of wireless users in various scenarios. In practical applications, since the transmitted power and energy resources of the UAVs are limited and the propagation environments are complicated and time-variant, it is challenging to control a group of UAVs to ensure coverage performance while preserving the connectivity and safety of the UAV networks. To this end, a two-step environment-learning-based method is proposed for the intelligent deployment of the UAVs. First, a machine learning algorithm is used to establish an accurate prediction model of the link qualities from the UAVs to the users under a specific scenario for the next step. Then, a modified deep deterministic policy gradient (DDPG) algorithm is employed to control the movements of the UAVs according to the predicted link qualities and to maximize the proportion of covered users. The prioritized experience replay mechanism is introduced to the standard DDPG algorithm to accelerate the deployment procedure. The coverage performance is analyzed in both the interference-free situation and the situation with co-channel interference. Simulation results have shown that the proposed method has a higher convergence speed than the standard DDPG method. Additionally, the proposed deployment method can achieve higher coverage performance and better adaptability to the dynamic environment than three commonly used methods, the random method, the K-means-based method, and the statistical-channel-model-based method.

**INDEX TERMS** Coverage performance, environment-learning-based method, link quality, optimal deployment, unmanned aerial vehicle networks.

## I. INTRODUCTION

In recent years, unmanned aerial vehicles (UAVs) have attracted great attention due to small size, low price, and high flexibility. The UAVs with variable positions can establish line-of-sight (LoS) communication links to the users. Therefore, the UAVs are suitable to be used as low-altitude flight base stations (BSs) to reduce the signal attenuation and improve the coverage performance [1], [2]. For example, in the case of a terrestrial BS failure, the UAV-BSs can be rapidly deployed to satisfy temporary coverage demands for wireless services [3]–[5]. The cellular networks can also be assisted by the UAV-BSs in the temporary hotspot area [6]–[9]. The deployment problem of the UAVs is more complicated than that of the terrestrial BSs. First, in practical application scenarios, due to the limited transmitted power and energy resources, multiple UAVs are often required to be deployed together to ensure the large coverage. Second, since the propagation environment is complex and changeable, the UAVs are expected to have certain adaptability to the environment to rapidly satisfy the coverage requirements of the users. In addition, multiple UAVs should be set with a certain distance limitation to maintain both the connectivity to ensure the robustness of the network and the security to prevent collisions caused by unexpected situations. Therefore, how to effectively deploy the positions of the UAVs to improve the coverage performance is a challenging problem.

Many research works have extensively finished for the deployment of the UAVs in wireless networks. The optimal deployment was mainly designed to realize the maximum

The associate editor coordinating the review of this manuscript and approving it for publication was Cesar Briso.

throughput of the users [10]–[12], the minimum transmitted power [13], [14], the trajectory optimization [15]–[17], and the optimal coverage performance of the UAV networks [18]–[26]. For the coverage problem, the authors in [18] considered the influence of distance and LoS probability and obtained the optimal height with maximum coverage radius for a single UAV via theoretical derivations. In [19], an equivalent quadratically-constrained mixed-integer nonlinear optimization method was proposed for maximizing the revenue of the network. The authors in [20] decoupled the UAV deployment problem in vertical and horizontal dimensions and modeled the UAV deployment problem as a placement problem with the smallest enclosing circles. The aforementioned works only considered a single UAV. However, the deployment of multiple UAVs is often required to complete the coverage task in a large area.

In [21], the authors investigated the influence of flying altitude on the combined transmitted power and coverage range of two UAVs in both the presence of interference and interference-free scenarios. The work of [22] provided the optimal separation distance between UAVs for mitigating co-channel interference and maximizing overall coverage performance in both suburban and urban environments. In [23], the authors formulated the UAV deployment problem as a continuous control task and proposed a deep reinforcement learning method for maximizing the energy efficiency of the UAV network with the joint consideration of coverage, fairness, energy consumption, and connectivity. Nevertheless, the distribution and movement of ground users were neglected in their works. In [24], a centralized deployment algorithm and a distributed motion control algorithm were proposed to realize the on-demand coverage of the UAVs. The authors in [25] offered an improved multi-population genetic algorithm to maximize the number of users with different quality of service requirements. A method of deploying multiple UAV-BSs was proposed in [26] to achieve a maximum number of covered users and avoid inter-cell interference.

However, these works mentioned above [24]–[26] assumed that the qualities of the communication links between the UAVs and the ground users could be estimated by means of statistical results, which were used to determine whether the users can be covered. In the practical propagation environments, due to the random distribution of the users and scatterers, the statistical results often have poor prediction accuracy on the link quality. Therefore, statistical results can hardly reflect the specific environmental details. Moreover, the random movements of the users will cause continuous changes of the propagation conditions, which requires the UAV network to have sufficient adaptability to the uncertain environment.

To this end, we propose a two-step environment-learning-based method to realize the optimal deployment of the UAV network for maximizing the coverage performance in both the interference-free situation and the situation with co-channel interference. The concept of environment learning in this paper is proposed for obtaining the mapping relationship between dynamic environments and the UAV deployment decisions. The UAV network can be deployed online in the actual application environment after sufficient learning.

A two-step learning method is proposed to perform the learning procedure. A typical machine learning algorithm, random forest, is first employed to learn the underlying relationship between the propagation environment and the link qualities from the UAVs to the users. Then, an accurate prediction model of the link qualities is established to provide accurate environment information for the next learning step. A modified deep deterministic policy gradient (DDPG) algorithm is second employed to learn the impact of the user distribution on the coverage performance according to the prediction model. In this process, due to the dynamic changes of the user locations, the UAV network is required to efficiently adapt to the time-variant environment. Therefore, the prioritized experience replay mechanism is introduced to the DDPG algorithm for the purpose of accelerating the deployment procedure. After learning, the deployment decisions of the UAV network are obtained for maximizing the proportion of covered users under the premise of ensuring connectivity and security.

In summary, the main contributions of this paper are listed as follows.

- A two-step environment-learning-based method is proposed to achieve the optimal deployment of the UAV network for maximizing the coverage performance. The method integrates the mechanisms of machine learning and reinforcement learning to obtain a mapping of the dynamic application environment to deployment decisions.
- The machine learning algorithm is used to predict the link qualities from the UAVs to the users, which provides accurate environmental status information for deployment decisions.
- A modified DDPG algorithm is employed to autonomously find the optimal deployment of the UAVs through continuous learning from the environment. The prioritized experience replay mechanism is introduced to improve the adaptability of the UAV network to the time-variant environment and accelerate the learning process.
- The coverage performance of the proposed two-step environment-learning-based method in both the interference-free and the situation with co-channel interference is evaluated in the simulations. The standard DDPG method, the random method, the K-means-based method, and the statistical-channel-model-based deployment method are employed for comparison. The results show that the proposed method can achieve high coverage performance and fast deployment speed.

The remainder of this paper is organized as follows. Section II presents the system model. The proposed two-step environment-learning-based method for the UAV deployment problem is described in Section III. Section IV shows the

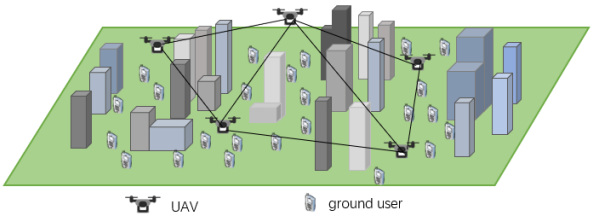



**FIGURE 1.** The scenario of the UAVs serving as the base stations.

simulation results and analysis. Finally, Section V concludes the paper.

## II. SYSTEM MODEL

As illustrated in Fig. 1, we consider a low-altitude network with $N$ UAVs in an urban area $\boldsymbol{R}$ to provide temporary communication services for the ground users. The motions of the UAVs are activated by one controller, which can get the global environment information and transmit control commands to all the UAVs. The proposed method is executed by the central controller instead of being configured on the UAV platforms, which ensures the light load of the UAVs to save energy resources and extend the flight time. For the sake of simplicity, we assume that the UAVs have the same fixed flight altitude $h$ and are all equipped with omnidirectional antennas with the gain $G = 1$. There are dense buildings on both sides of the streets, and $K$ users are distributed randomly along each street with a height of $z$. The user equipments in the systems are equipped with location-aware devices like global positioning system (GPS) chips, and the locations of the users are reported to the central controller via the UAVs. The downlink quality is analyzed as an example.

The coverage performance of the UAV network is investigated in both the interference-free situation and the situation with co-channel interference. The proportion of covered users is used to evaluate the coverage performance.

In the interference-free situation, the multiple UAVs work at different frequency points, so there is no interference between signals from different UAVs. For the ground user $k$, if the received signal-to-noise ratio (SNR) from the UAV $i$ exceeds a threshold $N_{\text{th}}$, the user $k$ is covered and its communication requirements can be supported by the UAV $i$. The noise in the channel is additive white Gaussion noise. The SNR of the user $k$ is given as

$$\text{SNR}_{i,k}(\text{dB}) = 10\log_{10}(\frac{p_{r\,i,k}}{p_n}), \tag{1}$$

where $p_n$ is the noise power in mW. $p_{r\,i,k}$ is the received power of user $k$ from UAV $i$ in mW and can be expressed by

$$p_{r\,i,k} = 10^{\frac{P_t - \text{PL}_{i,k}}{10}}, \tag{2}$$

where $P_t$ is the transmitted power in dBm, and $\text{PL}_{i,k}$ in dB indicates the path loss between the UAV $i$ and the user $k$.

In order to ensure the connectivity of the UAV network, the distance between two connected UAVs must not exceed the maximum sensed radius $R_s$, which is determined by the sensors equipped on each UAV [24]. We assume that each UAV needs to connected with at least two others to ensure the robustness of the network. At the same time, the distance of any two UAVs must be no less than the minimum distance $R_{\min}$ to prevent possible collisions. In this system, the user receives the signals of multiple UAVs and chooses to access the UAV with the best link quality. The ultimate goal is to find the optimal deployment of the multiple UAVs for maximizing the proportion of covered users on the premise of ensuring the connectivity and safety of the UAV network. The problem can be expressed by

$$\begin{aligned} &\max_{\substack{(x_i,y_i)\in\boldsymbol{R}\\ i\in[1,N]}} \frac{1}{K}\text{card}(k|k\in[1,K]), \quad \max_i(\text{SNR}_{i,k}) \geq N_{\text{th}}) \\ &\text{s.t. } R_{\min} \leq d_{i,j} \leq R_s, \end{aligned} \tag{3}$$

where $\text{card}(\cdot)$ represents the number of elements in the aggregation, $(x_i, y_i)$ represents the location of the UAV $i$ ,and $d_{i,j}$ is the distance between the UAV $i$ and UAV $j$.

Besides the interference-free situation, the situation where the UAVs interfere with each other during transmission should also be taken into considerations. Due to the scarcity of spectrum resources, sometimes multiple UAVs might need to reuse frequency band and transmit over the same channel, which will cause the co-channel interference in the UAV network [21]. It is assumed that the user can receive signals from multiple UAVs, which work at the same frequency. Therefore, when the user accesses the UAV with the best link quality, the signals from other UAVs will bring the co-channel interference. In this case, the user $k$ can be covered by UAV $i$ if the received signal-to-interference-plus-noise ratio (SINR) exceeds a threshold $I_{\text{th}}$. According to [27], the SINR of user $k$ can be expressed as

$$\text{SINR}_{i,k}(\text{dB}) = 10\log_{10}(\frac{p_{r\,i,k}}{\sum_{j\neq k} p_{r\,j,k} + p_n}). \tag{4}$$

The optimization problem in the situation with co-channel interference can be expressed by

$$\begin{aligned} &\max_{\substack{(x_i,y_i)\in\boldsymbol{R}\\ i\in[1,N]}} \frac{1}{K}\text{card}(k|k\in[1,K]), \quad \max_i(\text{SINR}_{i,k}) \geq I_{\text{th}}) \\ &\text{s.t. } R_{\min} \leq d_{i,j} \leq R_s. \end{aligned} \tag{5}$$

In (1) and (4), if the power values of the transmitted signals and noise are constant, the SNR or SINR is only determined by the path loss. In general, the path loss is defined as a function of the probabilities of LoS and non-line-of-sight (NLoS) links and can be expressed as [18]

$$\text{PL}(f, d) = P(\text{LoS}) \times \text{PL}_{\text{LoS}} + P(\text{NLoS}) \times \text{PL}_{\text{NLoS}}, \tag{6}$$

where the $P(\text{LoS})$ and $P(\text{NLoS}) = 1 - P(\text{LoS})$ represent the probabilities of LoS and NLoS connections, which depend on the environment and the elevation angle between the link and the horizontal plane. $\text{PL}_{\text{LoS}}$ and $\text{PL}_{\text{NLoS}}$ represent the path loss values of LoS and NLoS links based on the statistical experience, respectively.

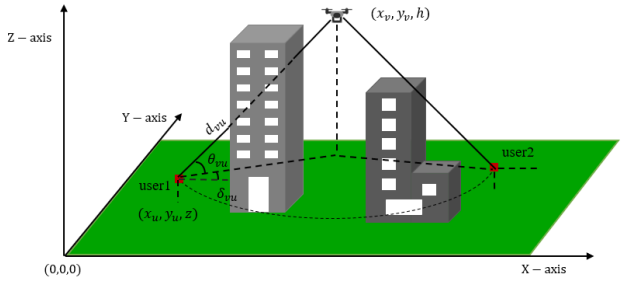



**FIGURE 2.** The scenario where user1 and user2 have the same distance and elevation angle but have different path loss. The notations in the figure are used as the input features for the random forest.

However, the aforementioned statistical model cannot accurately reflect the practical propagation characteristics. For example, in the scenario shown in Fig. 2, although the user1 and user2 have the same distances and elevation angles to the UAV, due to the obstruction of the building, the path loss of the user1 is larger than that of the user2 who has an LoS connection with the UAV. Therefore, the statistical model cannot accurately describe the link qualities of users and thus may affect the decisions of the UAV locations. Moreover, because of the randomness of the user locations and movements, the deployed locations of the UAVs need to be constantly adjusted, i.e., the UAV network needs to be adaptive to the time-variant environment. These are the motivations of proposing our environment-learning-based method.

## III. TWO-STEP ENVIRONMENT-LEARNING-BASED UAV DEPLOYMENT METHOD

In this section, we propose a two-step environment-learning-based method to solve the optimization problems. The problems can be separated into two parts which are solved sequentially. The procedure of the proposed method is shown in Fig. 3. A machine learning algorithm is first employed to learn the propagation characteristics of the air-to-ground (A2G) channel and to provide an accurate prediction model of the link qualities under the specific scenario. Second, a modified DDPG algorithm is proposed to learn the mapping relationship between the dynamic propagation environment and the UAV deployment. The model trained in the first step is used here to predict the accurate link qualities based on practical user positions and further to generate the coverage ratio. The central controller continuously updates the positions of the UAVs according to the coverage performance until it finally obtains the optimal deployment decision.

### A. PREDICTION MODEL OF A2G LINK QUALITY BASED ON THE RANDOM FOREST ALGORITHM

The first learning step is to learn the propagation characteristics of the A2G channel in a specific environment and to build an accurate prediction model for link qualities. The link quality can reflect the practical coverage performance and further guide the UAV deployment decisions. In this study, the path loss is chosen to evaluate the link quality. As mentioned above, the statistical model is difficult to accurately

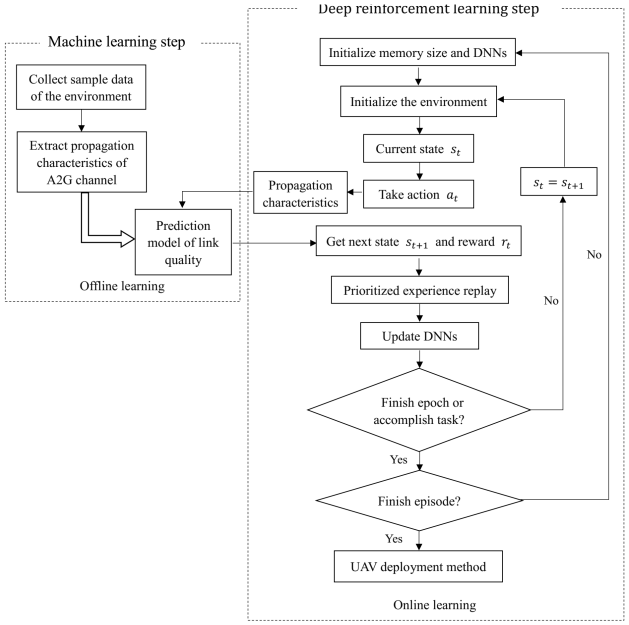



**FIGURE 3.** The procedure of the two-step environment-learning-based method.

describe the path loss and therefore the link quality of users in a complicated environment.

In order to improve the prediction accuracy of the link qualities, machine learning algorithms can be employed in the first step. In this paper, the random forest algorithm is selected because of its good performance in generalization and training speed. Random forest is a highly flexible machine learning algorithm and is optimized from the decision tree model [28]. It takes advantage of integrated learning and integrates multiple decision trees to form a forest. The forest randomly selects the training samples of each decision tree via the bootstrapping method. The features of each decision tree are also selected randomly from the initial feature set. These random processes ensure the irrelevance of the decision trees and further improve the stability of the random forest model. For regression or prediction problems, the output of the random forest is the average of the outputs of all the decision trees. We have demonstrated in previous works [29]–[31] that the random forest has high accuracy in the estimation of channel qualities.

To deploy the UAVs efficiently and precisely, the prediction model should have the capability to accurately predict the quality of any link. Therefore, it is necessary to acquire enough position samples of both the UAV and the user as the training data. Offline measurement campaigns can be carried out to collect these training samples at different positions of the UAV and user. The label of the sample is the actual path loss value, and the inputs are environmental features, which are listed as follows.

- The horizontal position of the UAV $(x_v, y_v)$, where $v \in \{1, 2, ..., V\}$ represents the $v$th position of the UAV.
- The horizontal position of the user $(x_u, y_u)$, where $u \in \{1, 2, ..., U\}$ represents the $u$th position of the user.

- The plane angle between the line of the projection of the transmission path and the x-axis direction $\sigma_{vu}$.
- The propagation distance from the UAV to the user $d_{vu}$.
- The elevation angle between the line of the transmission path and the horizontal plane $\omega_{vu}$.

Then, the prediction model of A2G link quality can be expressed as

$$\mathrm{PL}_{vu} = f(x_v, y_v, x_u, y_u, \sigma_{vu}, d_{vu}, \omega_{vu}) \tag{7}$$

where $f(\cdot)$ represents the mapping from the channel characteristics to the link qualities. Totally $UV$ samples can be collected for the training purpose. Then, the random forest algorithm is used to train the prediction model offline.

### B. DEPLOYMENT OF THE UAV NETWORK BASED ON A MODIFIED DDPG ALGORITHM

The second learning step is to realize the optimal deployment of the UAV network through a modified deep reinforcement learning (DRL) algorithm for maximizing the coverage ratio of the users. To handle the problems in high dimensional space, the DRL uses the deep neural networks (DNNs) to replace the state-action value function $Q(\cdot)$. In this study, the central controller of the UAV network is modeled as a DRL learning agent to explore the environment. The agent can acquire the practical positions of the UAVs and the users via UAVs. Then, the link qualities can be computed by the prediction model which has been learned in the first step. Based on the predicted link qualities, the agent can derive the coverage performance as the current state and accordingly takes actions to control the movements of the UAVs. Then, the agent acquires the corresponding reward and computes the next state. It constantly interacts with the environment in this kind of trial-and-error manner. The final purpose is to learn a policy $\pi(s)$ that maps any of a state to an action to maximize the discounted cumulative reward [32]

$$R_t = \sum_{t=1}^{T} \gamma r(s_t, a_t), \tag{8}$$

where $s_t$ and $a_t$ represent the state and action of the agent at epoch $t$, respectively. $T$ is the total number of epochs. $r(\cdot)$ is the reward function and $\gamma \in [0, 1]$ is the discount factor. The state-action value function

$$Q(s_t, a_t) = E_\pi [R_t | s_t, a_t] \tag{9}$$

indicates the expected value from the state $s_t$ when taking the action $a_t$ following the strategy $\pi(s_t)$.

In our proposed method, the state, action, and reward of the agent are defined as follows.

1) State $s_t = (\boldsymbol{x_t}, \boldsymbol{y_t}, c_t, \boldsymbol{e_t}, o_t, z_t, q_t)$:

- $\boldsymbol{x_t} = \{x_1, x_2, \ldots, x_N\}$ and $\boldsymbol{y_t} = \{y_1, y_2, \ldots, y_N\}$ indicate the positions of the $N$ UAVs in the Cartesian coordinate, and the values are normalized in order to be used as the input features of the DNNs.
- $c_t \in [0, 1]$ indicates the proportion of covered users in the target area.
- $\boldsymbol{e_t} = \{e_1, e_2, \ldots, e_L\}$ is the proportion of covered users in $L$ divided cells. The target area is evenly divided into $L$ cells in order to avoid the lack of local state changes. $e_l \in [0, 1]$ indicates the proportion of covered users in cell $l$, where $l = \{1, 2, ..., L\}$.
- $o_t = \{0, 1\}$ is 0 if the UAV flies out of range and otherwise is 1.
- $z_t = \{0, 1\}$ is 0 if the UAV network is disconnected or a collision occurs between the UAVs and otherwise is 1.
- $q_t = \{0, 1\}$ is 1 if all the users are covered by the UAV network and otherwise is 0.

It is assumed that the positions of the ground users are distributed randomly. For the initial state, the positions of $N$ UAVs can be specified randomly or by the K-means method. The random initialization method is to randomly specify the positions of the UAVs. The K-means initialization method is to divide users into different clusters and to initialize each UAV at the center of each cluster. The path loss of each link can be calculated by the prediction model in (7) and then the received SNR or SINR can be computed. The user is considered to successfully access the UAV if the SNR or SINR is larger than the minimum reception threshold. The state is continuously updated according to the movements of the UAVs. It should be noted that $o_t$ and $z_t$ are set as 1 in the initial state.

2) Action $a_t = \{m_1, m_2, \ldots, m_N\}$, where $m_i \in [0, 2\pi]$ indicates the flight angle of UAV $i$. The values are also normalized for the usage in the DNNs. In the algorithm, the flight decisions of the UAVs are updated with a fixed distance of $d_m$. According to the angles selected by the algorithm, the new positions of the UAVs at time $t+1$ can be calculated as

$$(x_{t+1}, y_{t+1}) = (x_t + d_m \cos(a_t), y_t + d_m \sin(a_t)). \tag{10}$$

3) Reward $r_t$ is defined as

$$r_t = \begin{cases} p_1 & \text{if out of range,} \\ p_2 & \text{if distance is not satisfied,} \\ p_3 & \text{if all users are covered,} \\ c_t + (c_t - c_{t-1}) \times 5 & \text{otherwise,} \end{cases} \tag{11}$$

where $p_1$ represents the penalty value of the UAVs moving out of range, $p_2$ is the penalty value of unsatisfied UAV distance, and $p_3$ is the reward value of the UAV network achieving full coverage. First, the reward function considers whether the action leads to violations of boundaries or distance constraints. If the flight decision results in an out-of-bounds, the network is given a penalty of $r_t = p_1$ and the movement of the corresponding UAV is canceled. Similarly, if an action causes disconnection or possible collisions of the UAVs, the penalty of the network is $r_t = p_2$. Second, if the UAV network has already been connected, the reward is given as $r_t = c_t + (c_t - c_{t-1}) \times 5$, which represents the sum of the current coverage and the change of coverage caused by the current action. Especially, if the users are all covered, the network is given $r_t = p_3$ as a reward for completing the

coverage task, and then the UAVs keep hovering at the fixed positions to provide full coverage to the users.

The action space of the UAVs is designed as a continuous space to ensure the accuracy of the deployment. For control in the continuous space, DDPG is a commonly used DRL method. This method uses the DNNs with parameters $\theta^{\mu}$ and $\theta^{Q}$ to represent the deterministic strategy $a = \pi(s|\theta^{\mu})$ and the Q function $Q(s, a|\theta^{Q})$ based on the actor-critic framework. In addition, target networks which have the same structure as the main networks are employed to solve the instability problem. The actor is a policy network, whose objective function is defined as the total rewards with a discount [32]

$$J(\theta^{\mu}) = \mathrm{E}[r_1 + \gamma r_2 + \gamma^2 r_3 + \cdots]. \tag{12}$$

The actor network is updated along the increasing direction of the Q value

$$\frac{\partial J(\theta^{\mu})}{\partial \theta^{\mu}} = \mathrm{E}[\frac{\partial Q(s, a|\theta^{Q})}{\partial a} \frac{\partial \pi(s|\theta^{\mu})}{\partial \theta^{\mu}}]. \tag{13}$$

The critic is a value network that approximates the value function of the state-action pair and provides gradient information. It is updated by minimizing the following loss function

$$L\left(\theta^{Q}\right) = \mathrm{E}\left[(Y_t - Q(s_t, a_t|\theta^{Q})\right], \tag{14}$$

where $Y_t$ is the target value and can be estimated by

$$Y_t = r_t + \gamma Q'(s_{t+1}, \pi(s_{t+1}|\theta^{\mu'})|\theta^{Q'}), \tag{15}$$

where $r_t$ is the reward at epoch $t$. $\theta^{\mu'}$ and $\theta^{Q'}$ represent the parameters of the target policy network and the target value network, respectively.

The DDPG algorithm uses a replay buffer to store experiences of $\{s_t, a_t, r_t, s_{t+1}\}$ and randomly selects a minibatch of them to update the neural networks at each learning epoch. The replay buffer may result in a few selecting opportunities for many experiences with large rewards and successful attempts, which will affect the convergence speed of the algorithm. However, in the practical environment, the distribution of user positions is dynamically changing, which requires the UAV network to realize rapid deployment for adapting to the time-variant environment. Therefore, instead of replaying all experiences uniformly, we introduce the prioritized experience replay mechanism [33] to improve the priorities of more valuable samples. The priority of the sample $n$ is decided by the temporal difference error (TD-error) $\delta_n$, which is defined as

$$\delta_n = Y_t - Q(s_t, a_t|\theta^{Q}). \tag{16}$$

The samples with large TD-errors have higher priorities in the replay buffer. The priority of sample $n$ is defined as $\rho_n = 1/\mathrm{rank}(n)$, where $\mathrm{rank}(\cdot)$ is the rank of the sample $n$ decided by $\delta_n$. Then, the sampled probability of $n$ is given as

$$P(n) = \frac{\rho_n^{\varphi}}{\sum_C \rho_C^{\varphi}}, \tag{17}$$

where $\varphi$ determines the weight of converting TD-error to priority and $C$ is the data size of minibatch. The prioritized replay mechanism may cause bias because it changes the state visitation frequency, and further changes the decisions. To handle the problem, importance-sampling weights are calculated as

$$W_n = \frac{1}{M^{\beta} P(n)^{\beta}}, \tag{18}$$

where $M$ is the size of the replay buffer and the parameter $\beta$ determines the extent of correction. With the prioritized replay mechanism, the modified DDPG is more efficient than the standard DDPG algorithm.

### C. SUMMARY OF THE PROPOSED METHOD

The specific steps of the proposed environment-learning-based method are presented in Algorithm 1. It is a detailed description of the procedure in Fig. 3. First, the random forest algorithm is used to establish the prediction model of link qualities (Line 1–3). Second, the DNNs are initialized. The movements of UAVs are controlled according to the outputs of the neural network and then the corresponding reward and the next state are obtained. Particularly, random noise is added to the actions for exploration, which follows a normal distribution with a mean of zero and a variance of $\varepsilon$. Here, $\varepsilon$ decays with a rate of 0.9995 over each learning epoch until the minimum variance $\varepsilon_m$ is reached, and then the algorithm can exploit the learning results and choose the optimal strategy (Line 4–22). The algorithm uses the prioritized experience replay mechanism for updating the networks. The critic network is trained by minimizing the loss function and the actor network is trained by computing the gradient function. (Line 23–29).

In practical environments, there may be several deployment decisions that can achieve full coverage of users. In this case, the average SNR or SINR can be added to the outputs of the well-trained model and further considered in the method. When the full coverage can be obtained by more than one deployment decisions, the one with the highest value of the average SNR or SINR can be selected as the optimal solution.

## IV. RESULTS AND ANALYSIS

In this section, simulations are conducted to evaluate the performance of the proposed UAV deployment method. We first describe the environment settings and then analyze the simulation results.

### A. ENVIRONMENT SETTINGS

Wireless Insite software [34] is used in this paper to generate accurate downlink channel data via the ray-tracing method. The ray-tracing method calculates the amplitude, phase, delay, and polarization of each possible ray path in a multipath channel based on the theory of wave propagation and obtains a coherent synthesis of all rays at the receiving point [35]. It has already been proved that the software can accurately and reliably calculate the wave propagation characteristics [34]. A dense urban scenario in Helsinki, Finland, is selected as the target region with a size of 900 m $\times$ 450 m

**Algorithm 1** Two-Step Environment-Learning-Based Method for Optimal UAV Deployment

Collect historical channel data or perform link qualitiy measurements in advance;
Extract channel characteristics data as the traning data set;
Establish the prediction model of link qualities based on the random forest algorithm;
Initialize the main and target DNNs with replay memory $M$ and weights $\theta$, $\theta'$;
**for** episode $= 1$ **do**
  Initialize the positions of the UAVs and the users;
  Calculate the initial state $s_t$ through the prediction model;
  **for** epoch $t = 1$ **do**
    Select an action $a_t$ with random noise;
    Extract current channel characteristics;
    Calculate the link qualities by (7) and obtain the current coverage performance;
    **if** UAV $i$ flies beyond the border **then**
      $r_t = p_1$ and cancel the corresponding action
    **end if**
    **if** the distance of UAV network is not satisfied **then**
      $r_t = p_2$
    **end if**
    **if** the users are are all covered **then**
      $r_t = p_3$ and the UAVs hover at the fixed positions;
      break
    **end if**
    Calculate the next state $s_{t+1}$ and the corresponding reward $r_t$ based on the current performance;
    Store transition sample $(s_t, a_t, r_t, s_{t+1})$ into $M$ with maximal priority $\rho_t = \max_{(f<t)} \rho_f$;
    Select random minibatch of $C$ samples $(s_n, a_n, r_n, s_{n+1})$ from $M$ with probability $P(n)$;
    Calculate importance-sampling weight $W_n$ and TD-error $\delta_n$;
    Update transition priority according to $\delta_n$;
    Set $Y_n$ by (15);
    Update the main networks $\theta^Q$ and $\theta^\pi$ by (13) and (14)
    Every $S$ steps update the target networks $\theta^{Q'}$, $\theta^{\pi'}$ to the main networks.
  **end for**
**end for**

and is shown in Fig. 4. There are 43 buildings with a height of 50 m unevenly distributed on both sides of the streets. The UAVs and the users are all equipped with omnidirectional antennas. The carrier frequency is 2.4 GHz and the bandwidth is 10 MHz. We consider 4 UAVs and 350 users in the scenario, where the users are randomly distributed on 12 streets. The streets and the possible locations for the UAVs are also shown

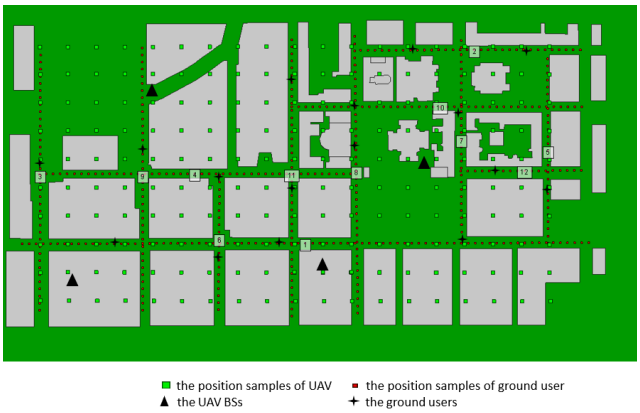



**FIGURE 4.** Simulation environment.

**TABLE 1.** Parameters of the simulation environment.

| Parameter | Explanation | Value |
|---|---|---|
| $N$ | number of UAVs | 4 |
| $K$ | number of users | 350 |
| $h$ | height of UAVs | 90 m |
| $z$ | height of users | 1.5 m |
| $R_s$ | sense range of UAVs | 500 m |
| $R_{\min}$ | minimum distance requirement of UAVs | 50 m |
| $N_{\text{th}}$ | minimum SNR threshold of users | -7 dBm |
| $I_{\text{th}}$ | minimum SINR threshold of users | -10 dBm |
| $d_m$ | UAV flight distance | 20 m |
| $B$ | bandwidth of A2G channel | 10 MHz |
| $W$ | power spectral density of Gaussian white noise | -174 dBm/Hz |
| $P_t$ | transmitted power of UAVs | 20 dBm |
| $V$ | number of UAV position samples | 190 |
| $U$ | number of user position samples | 451 |

in Fig. 4. For the clarity of the figure, only a small number of possible user locations are marked. The parameters of the simulation environment are illustrated in Table 1.

### B. PERFORMANCE OF THE MACHINE-LEARNING-BASED LINK QUALITY PREDICTION MODEL

In the first learning step, enough sample points are required to generate a reliable link quality prediction model. In this simulation, in order to mimic the actual urban environment, the samples of user positions are evenly collected on the 12 streets with 1.5 m height, and there are 451 samples of user positions in total. Taking the time consumption and data size into account, a UAV is set to move along the 10 horizontal routes with 90 m height, and the positions of the UAV are sampled every 40 m. Each route has 19 samples of UAV positions, and there is a total of 190 position samples which are shown in Fig. 4. The propagation characteristics in (7) are obtained from the ray-tracing method and the data sample can be expressed as $\{x_v, y_v, x_u, y_u, \sigma_{vu}, d_{vu}, \omega_{vu}, \text{PL}_{vu}\}$. The number of propagation paths includes a maximum of 10 reflection paths, 1 direct path, and 1 scatter path. The number of samples is 85,690 and these samples are divided into the training data set (80 percent) and the test data set (20 percent). The random forest algorithm is employed to predict the A2G link qualities of the users. 200 decision trees with a maximum

depth of 22 are selected and the rest parameters are set as default values. Three statistical metrics, including mean absolute error (MAE), standard deviation (STD), and root mean square error (RMSE) [36] are used as indicators to measure the accuracy of the prediction results. These indicators can be calculated by comparing the actual values of the test data set with the predicted values and can be defined as

$$\begin{aligned} \mathrm{MAE} &= \frac{1}{B}\sum_{b=1}^{B}\left|\mathrm{PL}_b - \mathrm{PL}'_b\right| \\ \mathrm{STD} &= \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\left|\mathrm{PL}_b - \mathrm{PL}'_b\right| - \mathrm{MAE}\right)^2} \\ \mathrm{RMSE} &= \sqrt{\frac{1}{B}\sum_{b=1}^{B}\left(\mathrm{PL}_b - \mathrm{PL}'_b\right)^2} \end{aligned} \tag{19}$$

where $B$ is the total number of samples in the test data set, $\mathrm{PL}_b$ is the actual value of the $b$th sample, and $\mathrm{PL}'_b$ is the predicted value.

The typical A2G urban statistical model in (6) is used here for comparison. The probability of LoS link $P$(LoS) is defined as

$$P\left(\mathrm{LoS}\right) = \frac{1}{1 + A\exp(-D[\omega - A])}, \tag{20}$$

where $A$ and $D$ are constants depended on the environment, $\omega$ is the elevation angle between the link and the horizontal plane. The path loss models of LoS and NLoS links are respectively given as

$$\begin{aligned} \mathrm{PL}_{\mathrm{LoS}} &= 20\log_{10}d + 20\log_{10}f + 20\log_{10}(4\pi/c) + \eta_{\mathrm{LoS}} \\ \mathrm{PL}_{\mathrm{NLoS}} &= 20\log_{10}d + 20\log_{10}f + 20\log_{10}(4\pi/c) + \eta_{\mathrm{NLoS}} \end{aligned} \tag{21}$$

where $f$ is the carrier frequency in MHz, $c$ is the speed of light, and $d$ is the propagation distance in km. $\eta_{\mathrm{LoS}}$ and $\eta_{\mathrm{NLoS}}$ are the additional losses for LoS and NLoS links, respectively. The environment parameters of the model are set as $A = 9.6$, $D = 0.28$, $\eta_{\mathrm{LoS}} = 1$, $\eta_{\mathrm{NLoS}} = 20$ according to [18].

Table 2 lists the prediction accuracy of the two models on the test data set. The RMSE and STD of the statistical model are about 15 dB higher than those of the random forest, and its MAE value is about 13 dB higher. The results show that the random forest model has smaller errors and fits better with the actual urban environment than the statistical model. It is because the traditional statistical A2G model loses the details of the environment and thus may bring large errors to the prediction results of the link qualities. As we mentioned above, such predictions cannot reflect the practical coverage of users and severely affect the optimal deployment of the UAVs.

### C. CONVERGENCE PERFORMANCE OF THE MODIFIED DDPG DEPLOYMENT METHOD IN THE TRAINING PROCESS

In the second step, according to the link quality prediction model, the modified DDPG method is trained in the dynamic environment for rapidly and adaptively obtaining the optimal deployment. The actor network of the modified DDPG is a two-layer fully-connected feedforward neural network, which includes 500 and 400 neurons in the two layers and utilizes the tanh(·) function for activation. The critic network is also fully-connected with two layers, the neuron numbers of the two layers are 700 and 20.

**TABLE 2.** Comparison of the random forest model and the statistical model in prediction accuracy.

| Method | STD (dB) | MAE (dB) | RMSE (dB) |
|---|---|---|---|
| Random Forest Model | 3.97 | 2.80 | 3.97 |
| Statistic A2G Model | 19.10 | 16.29 | 19.25 |

**TABLE 3.** Configurations of the parameters in the proposed deployment method.

| Parameter | Explanation | Value |
|---|---|---|
| $M$ | memory size | 50000 |
| $C$ | batch size | 128 |
| $\alpha$ | learning rate | 0.0005 |
| $\varphi$ | weight of converting TD-error to priority | 0.6 |
| $\beta$ | weight of importance-sampling | 0.4 |
| $\gamma$ | reward discount | 0.99 |
| $p_1$ | penalty of unsatisfied distance | -1 |
| $p_2$ | penalty of moving out of range | -2 |
| $p_3$ | reward of full coverage | 2 |
| $\varepsilon$ | variance of action noise | 1.5 |
| $\varepsilon_m$ | minimum of variance | 0.01 |
| $S$ | steps for updating target networks | 1500 |
| $L$ | number of divided cells | 9 |

The proposed method is trained for 1800 episodes, each of which has 150 epochs. The positions of the UAVs and the users are initialized at the beginning of each episode. In the proposed method, the initialization process of the UAV positions can be carried out in two ways, i.e., random initialization and K-means initialization. In each epoch, the UAVs are controlled to move according to the current state and get a corresponding reward and the next state. The parameters of the proposed deployment method are shown in Table 3.

The UAV network obtains the optimal deployment decision by maximizing the environmental cumulative reward. The accumulated reward of each episode in the training process is used to evaluate the convergence of the proposed method. Fig. 5 shows the accumulated reward values, which are averaged in every three episodes for the clarity of the figure. Especially, if the UAV network achieves full coverage during one episode, the reward of each remaining epoch in this episode, $r_t$, defaults to 2. The highest value of 300 indicates a special case when the UAV network can achieve full coverage at the initial state. In the beginning, the UAV controller is in a stage of randomly exploring the unknown environment. It begins to learn after about 400 episodes, and then the accumulated reward starts to increase.

The accumulated reward of the proposed method with random initialization gradually converges to a stable result after about 800 episodes, which means that the UAV network has already learned a good strategy. Due to the randomness of the

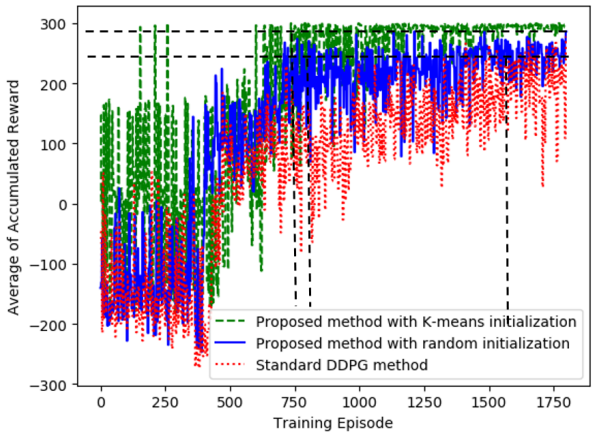



**FIGURE 5.** The average of the accumulated reward of the proposed method with random initialization, the proposed method with K-means initialization, and the standard DDPG method in the training process.

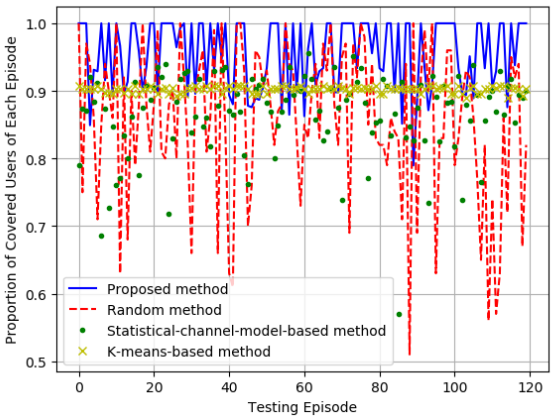



**FIGURE 6.** The coverage performance of the proposed method in the interference-free situation when compared with three methods, including the random method, the K-means-based method, and the statistical-channel-model-based method in the test process.

initial state, the accumulated reward of each episode is not a fixed value. With the K-means initialization, the accumulated reward of the proposed modified DDPG method converges to a stable value after about 750 episodes and gradually converges to the value close to 300. The results show that the converge rate of the proposed method with K-means initialization can be improved compared to that with random initialization. This is because the clustered positions obtained by the K-means method are close to the optimal deployment positions. In this case, the controller may quickly find the optimal locations and learn a corresponding deployment decision. At the same time, the K-means initialization will introduce additional clustering calculations.

The standard DDPG is used here for comparison. The accumulated reward of standard DDPG has a small drop at about 800th episode, which means the valuable strategies have not been well learned and utilized. The standard DDPG gets a similar stable result as the proposed method with random initialization after about 1600 episodes. The results indicate that the proposed methods with random initialization and K-means initialization both show faster convergence speed than the standard DDPG method.

In order to adapt to the real-time nature of the communication scenario, the complexity of the deployment method needs to be discussed. Within the proposed method, the channel characteristics are learned offline in the first step. Then, the prediction model of link qualities is built through the machine learning method before deployment. In the process of real-time deployment, the model can be used directly to predict the link qualities according to the user positions. Therefore, the complexity of the proposed method mainly depends on that of reinforcement learning in the second step, which mainly lies in the need to fully explore the dynamic environment during the training process. The training process can be done first based on offline learning, and then the training results can be updated according to a small amount of online learning to improve accuracy. Therefore, based on the well-trained model, the best deployment decisions can be obtained under different distributions of the user positions in real-time.

### D. COVERAGE PERFORMANCE OF THE MODIFIED DDPG DEPLOYMENT METHOD IN THE INTERFERENCE-FREE SITUATION

In this subsection, we verify the coverage performance of the proposed method in the interference-free situation. After training, the users and the UAVs are randomly initialized in each episode to test the coverage performance of the proposed method. The proposed method with random initialization is used here as an example. The UAV network first needs to ensure its connectivity and security when providing communication services to the users. However, due to the time-varying characteristics of the environment, the UAVs need to constantly move to ensure coverage, which may cause a certain probability of connection failure. The failure probability is tested in 1000 episodes by using the well-trained method and the result is 2.5 percent.

Under the premise of ensuring connectivity and safety, the coverage performance of the proposed method in 120 test episodes is shown in Fig. 6. The random deployment method, the K-means-based deployment method, and the statistical-channel-model-based deployment method are employed for comparison. The random deployment method [23] selects the actions randomly for the UAVs at each epoch. Its decisions are irrelevant to the current state of the environment. If the UAVs fly beyond the area or do not meet the distance requirements, the actions are abandoned and the UAVs make no movements. The K-means-based deployment method is to deploy the UAVs directly above the clusters of the users [37]. The statistical-channel-model-based deployment method decides the UAV locations via the modified DDPG algorithm and updates the state information based on the statistical A2G channel model in [18].

We average the coverage ratio results of the 120 episodes. The average coverage ratio of our proposed method can

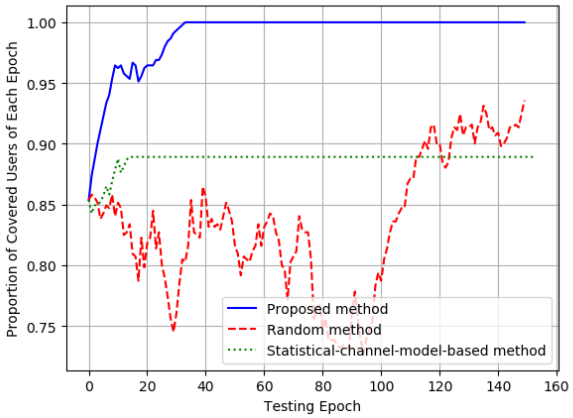



**FIGURE 7.** The change of the coverage performance under each epoch in one episode of the proposed method in the interference-free situation when compared with the random method and the statistical-channel-model-based method.

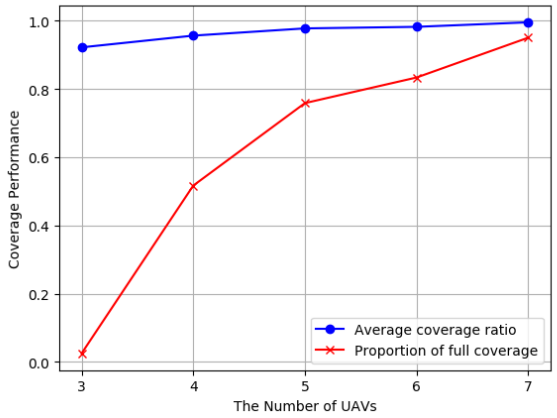



**FIGURE 8.** The impact of the number of UAVs on the coverage performance in the interference-free situation.

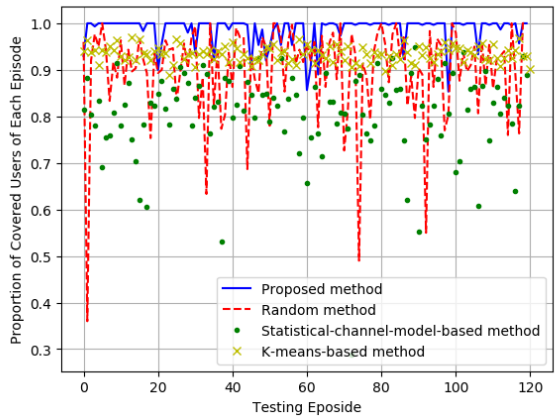



**FIGURE 9.** The coverage performance of the proposed method in the situation with co-channel interference when compared with three methods, including the random method, the K-means-based method, and the statistical-channel-model-based method in the test process.

reach 95.59 percent. Moreover, 53 percent of the episodes can achieve full coverage of users. There are three possible reasons for not reaching full coverage. First, the initial state of the environment is too poor, and the UAVs cannot move to the optimal positions within the limit of 150 epochs. Second, the state and action space of the environment are both continuous. The UAV network has the possibility of moving to an unexplored state during the movements and needs to learn the state. Third, the predicted link quality still has a certain error, which leads to the deviation of the deployment result, so that the users with poor link qualities may not be covered.

The average coverage ratio of the random deployment method is 85.55 percent, and the full coverage can be obtained in only 9 percent of the episodes. The reason is that the optimal deployment is realized only by simple random traversal, which is difficult to find the optimal deployment for multiple UAVs. The average coverage ratio of the K-means-based method is 90.22 percent, which is higher than that of the random deployment method. This is because the K-means algorithm ensures the users of each cluster can be located within the coverage radius of the corresponding UAV. However, the UAVs at these deployment positions cannot cover all the users due to the obstruction of the buildings. The average coverage ratio of the statistical-channel-model-based method is 86.79 percent, and the full coverage also cannot be realized. That is due to the fact that the statistical channel model has poor accuracy in estimating the link qualities. This model misleads the central controller to believe that the selected locations of the UAVs can achieve the optimal deployment. However, the coverage requirements are not satisfied because of the inaccurate estimations of link qualities. The results show that our proposed method outperforms the three other methods in terms of coverage performance.

Moreover, Fig. 7 shows the change of coverage ratio under each epoch in a randomly selected episode. Since the user positions are fixed during one episode, the K-means-based method is not considered here. The coverage ratio of our proposed method begins to increase after 5 epochs. After about 60 epochs, the proposed method achieves the full coverage. The coverage ratio of the random deployment method indicates that the method has no convergence and requires more explorations to find the optimal solution. The coverage of the statistical-channel-model-based deployment method is finally stable at about 0.89 due to the prediction error of link qualities.

Fig. 8 shows the impact of the number of UAVs on the coverage performance. The results of the average coverage ratio and the proportion of full coverage are tested in 120 episodes. The average coverage ratios for different numbers of the UAVs are all higher than 0.9. The proportion of full coverage is only 0.025 for 3 UAVs and increases to 0.98 when the number of UAVs is 7.

### *E. COVERAGE PERFORMANCE OF THE MODIFIED DDPG DEPLOYMENT METHOD IN THE SITUATION WITH CO-CHANNEL INTERFERENCE*

In the situation with co-channel interference, the proposed method is employed to learn the optimal deployment decision

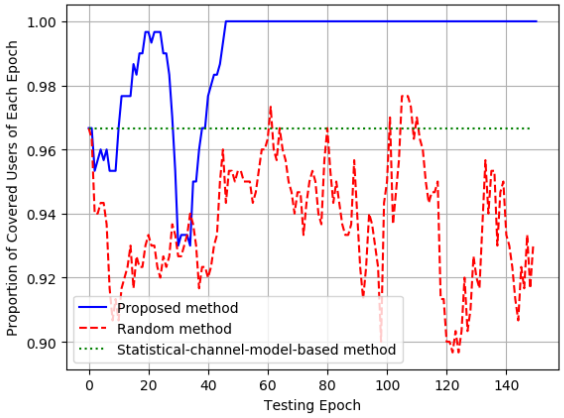



**FIGURE 10.** The change of the coverage performance under each epoch in one episode of the proposed method in the situation with co-channel interference when compared with the random method and the statistical-channel-model-based method.

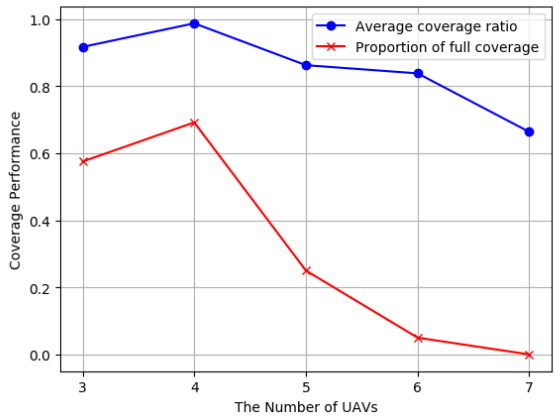



**FIGURE 11.** The impact of the number of UAVs on the coverage performance in the situation with co-channel interference.

to solve the problem in (5). After training, similar to the interference-free situation, the test results of the coverage performance are shown in Fig. 9, Fig. 10, and Fig. 11. Fig. 9 shows the proportion of covered users in 120 test episodes. The average coverage ratio of the proposed method is 98.69 percent, and 63 percent of the episodes can achieve full coverage of users. Fig. 10 illustrates the change of coverage ratio under each epoch in one randomly selected episode. The decrease of the coverage ratio at about 25th epoch may due to the action noise or the lack of learning at the current positions. The controller agent quickly corrects its wrong actions and achieves the full coverage after about 45 epochs. The results indicate that our proposed method still shows better performance in the situation with co-channel interference than the other methods.

Fig. 11 shows the impact of the number of UAVs on the coverage performance. The results are different from those in the interference-free situation. The two indicators of the coverage performance are both improved when the number of UAVs increases from 3 to 4. However, when the number of UAVs is larger than 4, the coverage performance becomes worse as the number of UAVs increases in the considered environment. This is because when the number of UAVs increases, the interference signal received by the user also increases. The SINR of the user is deteriorated as the interference power becomes strong, so the coverage performance is degraded.

## V. CONCLUSION

The UAV networks can be used as low-altitude BSs to flexibly and efficiently satisfy the communication demands of users. In this paper, we have proposed an efficient two-step environment-learning-based method for optimal UAV deployment. The method has maximized the coverage performance under the premise of ensuring the connectivity and safety of the network. In the first learning step, the A2G channel characteristics have been learned to generate an accurate prediction model of the link qualities from the UAVs to the users. The well-trained prediction model has provided a reliable coverage ratio. In the second learning step, according to the predicted link qualities, the deployment decisions of the UAV network have been learned for maximizing the proportion of covered users. The machine learning algorithm and the DRL algorithm have been applied respectively in the two learning steps. To improve the efficiency of deployment, a prioritized experience replay mechanism has been introduced to the second learning step. We have conducted simulations for evaluating the performance of the proposed method in both the presence of interference and interference-free scenarios. It has shown that the proposed method has a higher convergence speed than the standard DDPG deployment method. Three commonly used methods, the random deployment method, the K-means-based deployment method, and the statistical-channel-model-based method, have also been considered for comparison. The simulation results have shown that the proposed deployment method has better adaptability to the environmental changes and has higher coverage performance.

## REFERENCES

[1] I. Bucaille, S. Héthuin, A. Munari, R. Hermenier, T. Rasheed, and S. Allsopp, "Rapidly deployable network for tactical applications: Aerial base station with opportunistic links for unattended and temporary events absolute example," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, San Diego, CA, USA, Nov. 2013, pp. 1116–1120.

[2] B. Li, Z. Fei, and Y. Zhang, "UAV communications for 5G and beyond: Recent advances and future trends," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2241–2263, Apr. 2019.

[3] Y. Zeng, R. Zhang, and T. J. Lim, "Wireless communications with unmanned aerial vehicles: Opportunities and challenges," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.

[4] I. Bor-Yaliniz and H. Yanikomeroglu, "The new frontier in RAN heterogeneity: Multi-tier drone-cells," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 48–55, Nov. 2016.

[5] E. Kalantari, M. Z. Shakir, H. Yanikomeroglu, and A. Yongacoglu, "Backhaul-aware robust 3D drone placement in 5G+ wireless networks," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Paris, France, May 2017, pp. 1–6.

[6] Z. Khosravi, M. Gerasimenko, S. Andreev, and Y. Koucheryavy, "Performance evaluation of UAV-assisted mmWave operation in mobility-enabled urban deployments," in *Proc. Int. Conf. Telecommun. Signal Process. (TSP)*, Athens, Greece, Jul. 2018, pp. 1–5.

[7] F. Lagum, I. Bor-Yaliniz, and H. Yanikomeroglu, "Strategic densification with UAV-BSs in cellular networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 384–387, Jun. 2018.

[8] J. Liu, M. Sheng, R. Lyu, and J. Li, "Performance analysis and optimization of UAV integrated terrestrial cellular network," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1841–1855, Apr. 2019.

[9] A. V. Savkin and H. Huang, "Deployment of unmanned aerial vehicle base stations for optimal quality of coverage," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 321–324, Feb. 2019.

[10] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Wireless communication using unmanned aerial vehicles (UAVs): Optimal transport theory for hover time optimization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8052–8066, Dec. 2017.

[11] E. Kalantari, H. Yanikomeroglu, and A. Yongacoglu, "On the number and 3D placement of drone base stations in wireless cellular networks," in *Proc. IEEE Veh. Technol. Conf. (VTC-Fall)*, Montreal, QC, Canada, Sep. 2016, pp. 1–6.

[12] Z. Wang, L. Duan, and R. Zhang, "Traffic-aware adaptive deployment for UAV-aided communication networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

[13] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Optimal transport theory for power-efficient deployment of unmanned aerial vehicles," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 22–27.

[14] L. Ruan, J. Wang, J. Chen, Y. Xu, Y. Yang, H. Jiang, Y. Zhang, and Y. Xu, "Energy-efficient multi-UAV coverage deployment in UAV networks: A game-theoretic framework," *China Commun.*, vol. 15, no. 10, pp. 194–209, Oct. 2018.

[15] Q. Wu and R. Zhang, "Common throughput maximization in UAV-enabled OFDMA systems with delay consideration," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6614–6627, Dec. 2018.

[16] Q. Wu, Y. Zeng, and R. Zhang, "Joint trajectory and communication design for multi-UAV enabled wireless networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2109–2121, Mar. 2018.

[17] Q. Wu, L. Liu, and R. Zhang, "Fundamental trade-offs in communication and trajectory design for UAV-enabled wireless network," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 36–44, Feb. 2019.

[18] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.

[19] R. I. Bor-Yaliniz, A. El-Keyi, and H. Yanikomeroglu, "Efficient 3-D placement of an aerial base station in next generation cellular networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–5.

[20] M. Alzenad, A. El-Keyi, F. Lagum, and H. Yanikomeroglu, "3-D placement of an unmanned aerial vehicle base station (UAV-BS) for energy-efficient maximal coverage," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 434–437, Aug. 2017.

[21] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Drone small cells in the clouds: Design, deployment and performance analysis," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[22] A. A. Khuwaja, G. Zheng, Y. Chen, and W. Feng, "Optimum deployment of multiple UAVs for coverage area maximization in the presence of co-channel interference," *IEEE Access*, vol. 7, pp. 85203–85212, Jun. 2019.

[23] C. H. Liu, Z. Chen, J. Tang, J. Xu, and C. Piao, "Energy-efficient UAV control for effective and fair communication coverage: A deep reinforcement learning approach," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 2059–2070, Sep. 2018.

[24] H. Zhao, H. Wang, W. Wu, and J. Wei, "Deployment algorithms for UAV airborne networks toward on-demand coverage," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 9, pp. 2015–2031, Sep. 2018.

[25] Y. Chen, N. Li, C. Wang, W. Xie, and J. Xv, "A 3D placement of unmanned aerial vehicle base station based on multi-population genetic algorithm for maximizing users with different QoS requirements," in *Proc. IEEE Int. Conf. Commun. Technol. (ICCT)*, Chongqing, China, Oct. 2018, pp. 967–972.

[26] J. Sun and C. Masouros, "Deployment strategies of multiple aerial BSs for user coverage and power efficiency maximization," *IEEE Trans. Commun.*, vol. 67, no. 4, pp. 2981–2994, Apr. 2019.

[27] L. Zhou, L. Wei, Z. G. Sheng, X. P. Hu, H. T. Zhao, J. B. Wei, and V. C. M. Leung, "Green cell planning and deployment for small cell networks in smart cities," *Ad Hoc Netw.*, vol. 43, pp. 30–42, Jun. 2016.

[28] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[29] Y. Zhang, J. Wen, G. Yang, Z. He, and J. Wang, "Path loss prediction based on machine learning: Principle, method, and data expansion," *Appl. Sci.*, vol. 9, no. 9, May 2019, Art. no. 1908.

[30] G. Yang, Y. Zhang, Z. He, J. Wen, Z. Ji, and Y. Li, "Machine-learning-based prediction methods for path loss and delay spread in air-to-ground millimetre-wave channels," *IET Microw., Antennas Propag.*, vol. 13, no. 8, pp. 1113–1121, Jul. 2019.

[31] Y. Zhang, J. Wen, G. Yang, Z. He, and X. Luo, "Air-to-air path loss prediction based on machine learning methods in urban environments," *Wireless Commun. Mobile Comput.*, vol. 2018, Jun. 2018, Art. no. 8489326.

[32] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Beijing, China, 2014, pp. 387–395.

[33] Y. Hou, L. Liu, Q. Wei, X. Xu, and C. Chen, "A novel DDPG method with prioritized experience replay," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Banff, AB, Canada, Oct. 2017, pp. 316–321.

[34] P. Medeđović, M. Veletić, and Z. Blagojević, "Wireless insite software verification via analysis and comparison of simulation and measurement results," in *Proc. Int. Conv. MIPRO*, Opatija, Croatia, May 2012, pp. 776–781.

[35] Y. Wu, Z. Gao, C. Chen, L. Huang, H.-P. Chiang, Y. Huang, and H. Sun, "Ray tracing based wireless channel modeling over the sea surface near diaoyu islands," in *Proc. Int. Conf. Comput. Intell. Theory, Syst. Appl. (CCITSA)*, Yilan, Taiwan, Dec. 2015, pp. 124–128.

[36] J. Isabona and V. M. Srivastava, "Hybrid neural network approach for predicting signal propagation loss in urban microcells," in *Proc. IEEE Region 10 Humanitarian Technol. Conf. (R10-HTC)*, Agra, India, Dec. 2016, pp. 1–5.

[37] X. Liu, Y. Liu, and Y. Chen, "Deployment and movement for multiple aerial base stations by reinforcement learning," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Abu Dhabi, United Arab Emirates, Dec. 2018, pp. 1–6.

**XINRAN LUO** received the B.S. degree in electronic information countermeasure from the Beijing Institute of Technology, Beijing, China, in 2018, where she is currently pursuing the M.S. degree in information and communication engineering with the School of Information and Electronics. Her research interests include wireless channel modeling, machine learning, and reinforcement learning.

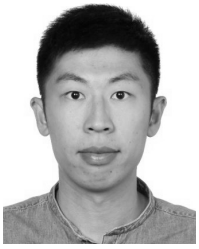
**YAN ZHANG** (S'06–M'10) received the B.S. degree in information engineering from the Beijing Institute of Technology, Beijing, China, in 2005, and the Ph.D. degree in information and communication engineering from Tsinghua University, Beijing, in 2010.

From 2010 to 2013, he was a Postdoctoral Researcher with the Department of Electronic Engineering, Tsinghua University, Beijing. From 2014 to 2015, he was a Research Assistant with the School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh, U.K. He is currently an Associate Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing. His main research interests include wireless channel modeling, physical security, and 5G mobile communication systems.

**ZUNWEN HE** received the B.S. and M.S. degrees in electromechanical engineering, and the Ph.D. degree in information and communication engineering from the Beijing Institute of Technology, Beijing, China, in 1986, 1989, and 2004, respectively.

He is currently an Associate Professor and the Vice Dean of the School of Information and Electronics, Beijing Institute of Technology, Beijing. His research interests include physical layer security, wireless sensor networks, and information systems.

**GUANSHU YANG** received the B.E. degree in communication engineering from Jilin University, Changchun, China, in 2017. She is currently pursuing the master's degree in information and communication engineering with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. Her research interest includes wireless channel modeling based on machine learning.



**ZIJIE JI** (S'18) received the B.S. degree in telecommunication engineering from the Beijing Institute of Technology, Beijing, China, in 2016, where he is currently pursuing the Ph.D. degree in information and communication engineering with the School of Information and Electronics, Beijing Institute of Technology, Beijing. His research interests include wireless communication security, secret key generation, and machine learning.