

Multi-objective UAV Positioning Mechanism for Sustainable Wireless Connectivity in Environments with Forbidden Flying Zones

İbrahim Atli*, Metin Ozturk*, Gianluca C. Valastro[†], Muhammad Zeeshan Asghar[‡]

*Faculty of Engineering and Natural Sciences, Ankara Yıldırım Beyazıt University, Ankara, Turkey

[†]National Inter-University Consortium for Telecommunications (CNIT), Italy

[‡]Department of Communications and Networking, Aalto University, Finland

Abstract—Unmanned aerial vehicles (UAVs)-based communication system is a promising solution to meet coverage and capacity requirements of future wireless networks. However, UAV-enabled communications is constrained with its coverage, energy consumption, and flying regulations, and the number of works focusing on the sustainability aspect of UAV-assisted networking has been limited in the literature so far. In this paper, we propose a solution to this limitation; particularly, we design a *Q*-learning-based UAV positioning scheme for sustainable wireless connectivity considering key constraints, that are, altitude regulations, non-flight zones, and transmit power. The objective is to find the optimal position of the UAV base station (BS) and minimize the energy consumption while maximizing the number of users covered. Moreover, a weighting mechanism is developed, where the energy consumption and number of users covered can be prioritized according to network/battery conditions. The proposed *Q*-learning-based solution is compared to the baseline *k*-means clustering method, where the UAV BS is positioned at the centroid location that minimizes the cumulative distance between the UAV BS and the users. The results demonstrate that the proposed solution outperforms the baseline *k*-means clustering-based method in terms of the number of users covered while achieving the desired minimization of the energy consumption.

Index Terms—Sustainable wireless connectivity, energy saving, UAV, communication system, 5G, positioning, reinforcement learning

I. INTRODUCTION

It has been a truism that the number of subscriptions to mobile communication networks has been increasing over the years, and that the newer generations have been becoming more dominant than the legacy after a few years of their first deployment. As seen in the report by Ericsson that the number of subscriptions for the fifth generation of mobile communications (5G) is projected to be around 3.5 billion around the world [1]. On the other hand, the mobile data traffic has also globally been on the rise; such that the same report reveals that the global mobile data traffic increased by 46% during 2020 and the monthly total global data traffic was more than 66 exabyte (EB) in the first quarter of 2021 [1].

All these statistics almost reach the same conclusion: the number of users and the amount of data consumed per user have been dramatically rising over the years. In 5G, on the other hand, such increase is more highlighted due to the fact that there are more demanding emerging applications, including tactile Internet, 4K video streaming, online gaming,

etc., and the concept of Internet of things (IoT) has been seriously proliferating, and pervading our daily lives with a large inclusion in various domains, such as healthcare, manufacturing, etc. [2]. This means that the two components of the increase in the data traffic (i.e., number of connected devices and the amount of data consumed per device) are more challenging issues in 5G, as the aforementioned data-hungry applications increases the amount of data consumed while IoT greatly contributes to the number of connected devices. Therefore, in 5G, the scale of the challenge is multiplied compared to that of the legacy networks, and thus there is more sophisticated solutions needed to tackle such level of unprecedented challenge.

There have been various concepts and technologies proposed in the literature in order to address the aforementioned capacity issues. The use of millimeter-wave (mmWave) frequencies, massive multi-input multi-output (mMIMO), and network densification are some of the most popular and practical ones among others [3]–[6]. Each of these technologies has different set of advantages and disadvantages, however they mainly target capacity enhancement in mobile communication networks. With mmWave communications, for example, an additional spectrum added to 5G networks—it has already been included in 5G New Radio (NR) as frequency range-2 [7]—, and thus the capacity is increased with this additional bandwidth. The use of higher carrier frequency also enable smaller antenna sizes, which subsequently enables mMIMO antenna array, enhancing the capacity further [3], [5]. Network densification, on the other hand, offers deployments of smaller base stations (BSs) with comparatively less antenna transmit power in order to reuse the frequency band, leading to a great deal of capacity enhancement [3].

Even though all these solutions are quite beneficial in enhancing the capacity of mobile communication networks, there is still room for improvements, since the spatio-temporal changes in wireless networks pose another type of challenges. More specifically, unusual circumstances, including expositions, sport competitions, and musical concerts, where much more people than normal gather together and significantly in-

crease the demand for wireless communications¹ are required to be tackled in a more sophisticated and intelligent manner. This is mainly because such kind of events do not happen pretty often (only few times a year); hence, it is not a good idea to design the network by taking them into consideration. In this regard, BS mounted on UAVs (which will be referred as to UAV BS hereafter) has been a promising solution to meet the strict user requirements of coverage, capacity, and quality of service (QoS). With the dawn of 5G network and related technologies, the user requirements are getting diverse as the users are from a diverse groups including conventional user equipments (UEs), IoT devices, machines, vehicles, etc. The UAV-assisted communication system is a solid use-case for the next generation of mobile communications given that UAVs are flexible, easy-to-deploy, and cost-efficient [8]. This provides a boost to the terrestrial cellular infrastructure as UAVs can be deployed to provide the extra coverage, or increase the capacity in the given area.

Indeed, terrestrial mobile BS mounted on terrestrial vehicles, including trucks, vans, and cars, would also offer flexibility to some extent, as they are also capable of moving according to the circumstances faced. However, UAV BSs are deemed as more advantageous than their terrestrial counterparts due to multiple additional benefits [3], [9]. First, while the terrestrial mobile BSs can also move in two dimensions, due to the fact that UAV BSs can also alter their altitude, giving them an extra degree-of-freedom [10]. Therefore, they can provide better connectivity since line-of-sight (LoS) becomes more likely with the required adjustments in terms of the altitude. Second, terrestrial mobile BSs are restricted with traffic regulations and with the planning of the environment (e.g., city); however, UAV BSs do not have such strict regulations apart from some altitude restrictions. To this end, UAV BSs are more suitable and feasible solutions in many scenarios if they are managed properly.

The primary research challenges regarding the UAV-assisted communication systems are as follows: i) determining optimal positions of the UAVs; ii) finding optimal UAV trajectory; and iii) meeting the regional regulations for UAVs. Energy efficient networking with UAV BSs is at the core of the discussion, since the UAVs are battery operated and have limited energy capacity. This is one of the most important limitations with UAV-assisted communications, requiring a proper management; otherwise, the concept can get infeasible if the flight time of UAVs cannot be sufficiently prolonged². Furthermore, UAV-based communication systems represent even a more complex case, because the total energy consumption depends on both the communication to the

ground users and flying on a predefined trajectory or simply hovering over a fixed point. Besides, to cover as much as users as possible while maintaining the energy efficiency is an important aspect of UAV-assisted networking, since the main objective is to enhance the capacity of the network in order to serve more users in temporally-dense networks. Put it another way, energy efficiency is required to realize the main objective, which is capacity enhancement in this case, and thus the idea is to keep the UAV BSs more in the air in order to serve more users in total. Therefore, in order to capture this phenomena in our work, we consider both the energy consumption and the total number of users covered as objective functions of our novel problem formulation.

No-fly-zones (NFZs), which are restricted or prohibited areas where UAV is not allowed to fly—such as military bases—, are considered as a practical constraint in the deployment phase of UAV BSs. This is because the optimal UAV trajectory and positioning are affected from NFZs given that UAVs will need to avoid those places even if they are optimal positions. In other words, UAV BSs are supposed to be positioned considering the NFZ constraints, bringing additional challenge to the optimization process. In addition, as aforementioned, there are also regulations on the minimum and maximum altitude of the UAVs; such that the UAVs are supposed to be within the allowed range in terms of their altitude³. In this regard, in order to seize this idea of NFZs and altitude regulations in this work, we consider them as constraints in the problem formulation.

Machine learning is a promising solution for various domains, such as agriculture [11], finance [12], healthcare [13], due to its strong capabilities in terms of convergence, dynamism, and agility. It also has a considerable place in optimizing wireless communications networks [14]–[18]. Moreover, it is envisioned that it will play rather a more critical role in the upcoming generations of mobile communication networks, such as sixth generation of mobile communication (6G) [19]–[21]. Reinforcement learning (RL) has a special place in machine learning, as it is structurally quite different than supervised and unsupervised learning methods. RL consists of a set of policy-based goal-seeking algorithms, where an agent takes actions in a given environment in order to maximize its reward or minimize the penalty, and therefore RL is predominantly used in optimization problems rather than time-series analysis, classification, or clustering as supervised and unsupervised learning algorithms do.

RL has unique advantageous characteristics making it more preferable than other types of optimization methodologies, including heuristics. First, RL algorithms, such as Q -learning and SARSA, are predominantly model-free, meaning that they do not require the model of the environment-of-interest in advance, they instead interact with the environment in order to capture its dynamics [22]. Moreover, since RL algorithms include learning in their body, they do not have to start from scratch every time there is a change in the environment, they rather adapt themselves to the changes, giving them a strength

¹It has been quite a normal habit that people go online during these events and broadcast live videos over their social media applications. Therefore, this critically affects the data traffic in those regions. In other words, the problem is not only that people accumulate around a specific region increasing the population intensity, they also tend to consume more bandwidth with these online video streaming applications.

²The sufficiency mentioned here should be discussed according to the scenario; i.e., the conditions of the networks and the requirements of mobile network operators as well as users, and hence it is quite hard to put a formal and strict definition and/or a numerical value for it. However, the main idea is to maximize the flight time of UAVs as much as possible.

³The altitude regulations vary for different countries and regions.

of optimization with reasonable time complexity. This is an essential feature for an optimization algorithm especially for dynamic scenarios, where network conditions change rapidly and frequently. To this end, we employ Q -learning, one of the most common RL algorithms, in this work in order to take the benefit of above-mentioned features.

A. Related Work

The literature on UAV-assisted communication systems will be thoroughly evaluated in this subsection. In recent years, numerous number of studies have been done in the field of UAV-assisted wireless networking [23]–[32], but only a few studies have looked at the flying regulations [33], [34]. In [8], the authors presented a survey on the most recent research possibilities and problems in the field of UAV aided wireless networks. The key difficulties in UAV-assisted networking are investigated, including 3D deployment, performance analysis, channel modeling, and energy efficiency.

For the coexistence of UAVs and under-laid device-to-device (D2D) communication networks, a tractable analytical framework is proposed in [29]. The authors showed that flying a UAV at the ideal altitude can result in the highest system sum-rate and coverage probability. Furthermore, an optimal trajectory design can reduce transmit power; however, networking under the UAV altitude regulations has not received enough attention in the literature. The lowest and maximum authorized altitudes of flying UAVs vary by country; for example, European laws for flying UAVs establish the limits of minimum and maximum allowed altitudes, which may fluctuate in different regions of the world, and in [33], the authors looked at the status of UAV-related regulations.

In the previous few years, numerous surveys and tutorials have been released. The findings reveal that creating air route networks is a scientifically sound and efficient way to standardize and improve the efficiency of low-altitude UAV operations. The most significant approach for UAV regulation in urban regions, in terms of safety and efficiency, is to enhance research that heavily relies on urban remote sensing and Geographic Information System (GIS) technology, as well as application demonstrations of low-altitude public air route networks [33]. In [34], the authors discussed the standardization initiatives for UAV-assisted UEs, UAV-assisted BSs, UAV communication prototypes, and UAV-assisted cellular communications cyber-physical security. The usage of UAV-assisted communication has been suggested as a possible approach for Internet of things (IoT) networks in the literature [8], [35]–[37]. In [35], it was demonstrated how to collect data in an energy-efficient manner for IoT networks, and the best way to deploy and move several UAVs was examined. The authors developed a framework for concurrently optimizing UAV three-dimensional (3D) positioning and mobility, device-UAV association, and uplink power regulation in their paper. First, the ideal UAV location is identified based on the locations of active IoT devices at each time instant. Next, the optimal UAV mobility patterns were studied to dynamically serve the IoT devices in a time-varying network. The goal is to utilize as little energy as

possible for the UAVs' mobility while serving IoT devices. For the coverage and rate analyses, a tractable analytical framework is developed [28], wherein the UAV's coexistence with a D2D communication network is taken into account.

The interfering UAVs are considered in [38], while in [39], the authors investigated the optimal 3D placement of multiple UAVs, that use directional antennas, to maximize total coverage area. The authors in [40] analyzed the impact of a UAV's altitude on the sum-rate maximization of a UAV-assisted terrestrial wireless network, and the 3D placement of drones with the goal of maximizing the number of ground users which are covered by the drone was investigated in [41]. The minimum number of drones needed for serving all the ground users within a given area was determined in [42]. In [43], evolutionary algorithms were employed to find the optimal placement of low-altitude platforms (LAPs) and portable BSs for disaster relief scenarios, by deploying the UAVs at the optimal locations, the number of BSs required to completely cover the desired area was minimized. The authors in [44] determined the optimal location of the UAV by maximizing the average rate while ensuring that the bit error rate will not exceed a specified.

Different considerations, such as flight time, energy limits, ground user demands, flying regulations, and avoiding NFZs, have a substantial impact on a UAV's trajectory. For maximizing the minimal average rate among ground users, the authors in [45] proposed a simultaneous optimization of user scheduling and UAV trajectory. While a number of jammers with unknown locations sent jamming signals, the authors in [45] presented a combination UAV and ground users' scheduling and transmit power allocation optimization technique. The optimal trajectory of UAVs with multiple antenna for maximum sum-rate in uplink communication was researched [46]. The throughput maximization problem in mobile relaying systems was investigated in [47] by optimizing the source/relay transmit power along with the relay trajectory, subject to practical mobility constraints such as UAV's speed and relay locations. The authors in [48] proposed an E-Spiral algorithm for accurate photogrammetry that considers the camera sensor and the flight altitude to apply the overlapping necessary to guarantee the mission success. This technique used an energy model to determine different optimal speeds for straight parts of the road, thereby lowering energy consumption and improving the energy model's ability to estimate overall path energy. To characterize the practical path planning requirements of UAVs in difficult situations, the authors in [49] developed an energy-aware multi-UAV multi-area coverage path planning model. A bipartite cooperative coevolution (BiCC) algorithm was suggested in this regard, which coevolves inter-area and intra-area path planning components to generate good solutions. In [50], the authors proposed a geometric planning-based iterative trajectory optimization technique. To begin, graph theory was used to generate all potential UAV-ground BS association sequences, and candidate association sequences were chosen based on the topological link between UAV and ground BSs. Following that, an iterative handover location

design based on the triangle inequality property is given to calculate the shortest flying route with quick convergence and minimal computation complexity. After that, by comparing all of the possible trajectories, the optimal flight trajectory can be determined. The authors presented a tradeoff between mission completion time and flight energy usage [50].

In addition, recent research looks on multi-objective optimization of UAV assisted communication [51], [52]. Over the course of a flight, a multi-objective optimization problem is constructed to jointly optimize three objectives [52]: 1) maximization of cumulative data rate, 2) maximization of total gathered energy, and 3) reduction of UAV energy consumption. Because these goals are incompatible, the authors suggested an enhanced deep deterministic policy gradient (DDPG) technique for learning UAV control policies with multiple goals. In [53], the authors developed a mathematical propulsion energy model for rotary-wing UAVs with the goal of minimizing the total energy consumption of the UAV while keeping all ground node data rates in consideration. The authors suggested a new path discretization method for converting the original optimization problem into a discretized equivalent with a finite number of optimization variables, for which the successive convex approximation technique yielded a high-quality suboptimal solution.

B. Contributions

In this paper, a smart UAV positioning mechanism is proposed by taking such regulation constraints into account to provide sustainable wireless coverage and services to the ground users under more realistic conditions. In particular, we propose a Q -learning-based approach for UAV-assisted communication systems. The optimal position of UAVs are determined under the constraints of altitude regulations, NFZs, and transmit power. The main contributions of the paper are as follows:

- A smart UAV positioning mechanism for a sustainable UAV communication system is proposed, under certain constraints.
- A multi-objective optimization model is formulated, that is, minimizing the energy consumption of UAV, while maximizing the number of users covered.
- A weighting mechanism is developed in order to prioritize the two objectives given in the previous item over each other for different scenarios.
- Q -learning based algorithm is used to find the optimal position of UAV. The convergence of the developed algorithm is first tested, followed by comparing its performance with the baseline k -means method in terms of number of users covered and energy consumption.

C. Organization of the Paper

The remainder of this paper is organized as follows. Section II describes the system model including propagation and energy consumption models, while Section III presents the problem formulation. Section IV presents the proposed Q -learning based UAV positioning mechanism, followed by discussing the simulation scenario and the results in Section V. Section VI concludes the the paper.

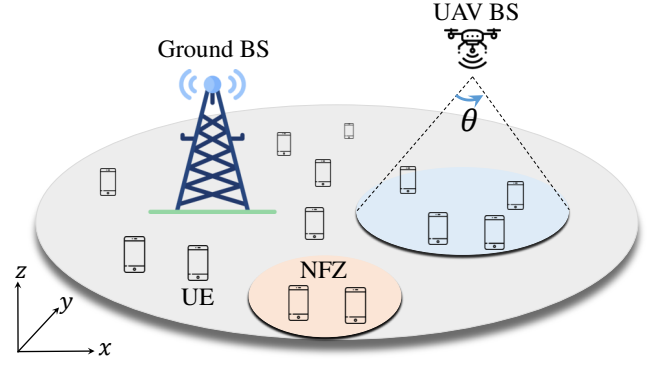


Fig. 1. The considered scenario depicting a ground macro BS that provides a wide-range coverage and UAV BS that provides additional capacity to the cellular network. A no-fly zone (NFZ), over which the UAV BSs are prohibited to fly, is also illustrated.

II. SYSTEM MODEL

In this section, we will elaborate on the system modeling of the work, including the scenario used, propagation and energy consumption modeling.

A. Scenario

We consider a UAV mounted BS to provide coverage to n_u ground users distributed over a rectangular geographical area of size $a \times b$ square-meter. Let $U = \{1, 2, 3, \dots, n_u\}$ be a set of n_u users, and the UAV can move in any direction (x , y , or z) to provide coverage to ground users based on the user density. The total time of service T_t (in mins) is divided into consecutive time-slots with equal duration of T_d (in mins), such that $n_{ts} = T_t/T_d$ is the number of time slots, and \vec{T} becomes a vector containing the consecutive time slots as $\vec{T} = [t_0, t_1, \dots, t_{n_{ts}}]$. The location of a user is represented by (x_u, y_u, z_u) , where $x_u \in \mathbb{R}^+$ is in the range of $x_u = [0, a]$, and similarly $y_u \in \mathbb{R}^+$ is in the range of $y_u = [0, b]$. $z_u \in \mathbb{R}^+$ is assumed to be a constant number as $z = h_u$, since we consider the conventional mobile handsets, which are carried around a similar height. In this work, we assume the height of UEs to be $h_u = 1.5$ meters.

The altitude of the UAV $h_d \in \mathbb{R}^+$ is in the range of $[h_{\min}, h_{\max}]$ where h_{\min} and h_{\max} are the minimum and maximum allowed altitude⁴ of the UAV, respectively. For instance, according to the European regulations for flying UAV, h_{\min} and h_{\max} are 30 and 120 meters, respectively.

B. Propagation Model

The propagation model is inspired from [26], [31], wherein the average path loss model for air-to-ground communication can be characterized in terms of LoS links and non-LoS (NLoS) links, given as

$$\begin{aligned} L_{\text{LoS}}^k &= 20 \log \left(\frac{4\pi f_c d_k}{2} \right) + \eta_{\text{LoS}}, \\ L_{\text{NLoS}}^k &= 20 \log \left(\frac{4\pi f_c d_k}{2} \right) + \eta_{\text{NLoS}}, \end{aligned} \quad (1)$$

⁴According to the regulations of concerning country/region.

where f_c is the carrier frequency and d_k is the Euclidian distance between the UAV and user k , c is the speed of light, η_{LoS} and η_{NLoS} are the mean value of the excessive path loss (in addition to the free-space path loss) for LoS and NLoS, respectively. The LoS link probability is given as

$$P_{\text{LoS}}^k(\vartheta_k) = \frac{1}{1 + \psi \exp(-\varsigma(\vartheta_k - \psi))}, \quad (2)$$

where ψ and ς are constant values depend on the environment, $\vartheta_k = \frac{180}{\pi} \arcsin(\frac{h_d}{d_k})$ is the elevation angle. Besides, the NLoS link probability can be calculated as

$$P_{\text{NLoS}}^k(\vartheta_k) = 1 - P_{\text{LoS}}^k(\vartheta_k). \quad (3)$$

Therefore, the average path loss can be expressed as

$$L_k(h_d, \vartheta_k) = 20 \log \left(\frac{4\pi f_c d_k}{2} \right) + P_{\text{LoS}}^k(\vartheta_k) \eta_{\text{LoS}} + P_{\text{NLoS}}^k(\vartheta_k) \eta_{\text{NLoS}}. \quad (4)$$

C. Energy Consumption Model

The energy consumption model is inspired from [31] where it is modeled as a combination of the energy consumption resulting from communication, UAV hovering, and UAV mobility.

1) *Communication Energy Consumption*: The communication energy is needed to communicate with the ground users; i.e., transmit/receive the signals to/from the users. As such, the communication energy consumption of UAV E_C can be calculated as follows:

$$E_C(t_j) = (n_{u,t_j} P_t + P_{\text{cu}}) t_{\text{cm}}, \quad (5)$$

where P_t is transmission power, P_{cu} is the on-board circuit power, t_{cm} is the duration to communication of UAV to user j , and n_{u,t_j} is the number of users served by UAV during time slot t_j .

D. Hovering Energy Consumption

The hovering energy is required to keep the UAV up in the air and stay at the right altitude, and the hovering energy consumption of the UAV during time slot t_j can be given as

$$E_H(t_j) = P_H t_H, \quad (6)$$

where t_H is the duration of hovering of UAV. P_H (in Watts) is the instantaneous hovering power consumption that can be determined by

$$P_H = \frac{MG^{\frac{3}{2}}}{\sqrt{2\rho\pi\beta^2}}, \quad (7)$$

where M is the number of rotors of the helicopter, G is the thrust (in Newton), ρ is the fluid density of the air, and β is the rotor disk radius.

E. Mobility Energy Consumption

The mobility energy is needed to move the UAV to the optimal position in order to serve the ground users. From [31], the mobility energy consumption of the UAV can be given as

$$E_M(t_j) = P_h \frac{d(t_j)}{v_h} + I(\Delta h(t_j)) P_a \frac{\Delta h(t_j)}{v_a} - (1 - I(\Delta h(t_j))) P_d \frac{\Delta h(t_j)}{v_d}, \quad (8)$$

where P_h is instantaneous power consumption for mobility in the horizontal direction, P_a is the ascending power, P_d is descending power. $d(t_j)$ the horizontal moving distance at t_j , while $\Delta h(t_j)$ is the changes in the altitude of the UAV at t_j . v_h , v_a , and v_d are the horizontal, vertical (ascending), and vertical (descending) velocities of the UAV, respectively⁵. $I(\Delta h(t_j))$ is the indicator function, such that [31]

$$I(\Delta h(t_j)) = \begin{cases} 1 & \Delta h(t_j) \geq 0, \\ 0 & \Delta h(t_j) < 0. \end{cases} \quad (9)$$

Lastly, the power consumption of the horizontal direction is as follows:

$$P_h = P_P + P_1, \quad (10)$$

where P_P is the parasitic power for overcoming the parasitic drag due to the aircraft's skin friction [31].

III. PROBLEM FORMULATION

The primary objective of this work is to maximize the number of connected users while minimizing the total energy consumption of the UAV BS in order to prolong its flight time. In this regard, we aim at finding the optimal position of the UAV BS and associate the ground users, which are normally out-of-service due to the congestion in the terrestrial network, to it so that the number of unconnected users are reduced; however, it is important to consider the total energy consumption of the UAV BS in order to maximize the service duration given they are battery operated and have limited flight time. We also consider certain constraints, including the NFZs (e.g., the UAV BS cannot fly over those forbidden regions), the altitude regulations for UAVS, etc., thereby determining the optimal positioning of the UAV BS by taken into account both the requirements and constraints becomes a non-trivial objective.

Theorem 1. *The number of connected users, n_c , can be controlled by the altitude of the UAV BS, h_d .*

Proof. Let K be a rectangular prism with the base area of $A_K = x_K y_K$, x_K and y_K are the x and y dimensions of the base of K that is placed on $z = 0$ plane. If we place the UAV BS—with a directivity angle of θ —at any point inside K , the radius of the footprint of the UAV BS can be calculated as follows [23]:

$$R_d = h_d \tan \left(\frac{\theta}{2} \right), \quad (11)$$

⁵For the details on the calculations of these velocities, please refer to [31].

where h_d is determined by

$$h_d = |\vec{N}|, \quad (12)$$

where $|\vec{N}|$ is the length of the normal vector, \vec{N} , from UAV BS to the $z = 0$ plane. Then, the footprint of the UAV BS can be found as

$$A_d = \pi R_d^2. \quad (13)$$

If a random point is selected on $z = 0$ plane, then the probability of falling inside the footprint of the UAV BS can be given as

$$p_r = \frac{A_d}{A_K} = \frac{\pi R_d^2}{x_K y_K}, \quad (14)$$

where A_K is the base area of the rectangular prism K . Let p_q is the probability of receiving sufficient signal-to-noise ratio (SNR⁶) for a UE, such that

$$p_q = P(S_r \geq T_{\min}), \quad (15)$$

where S_r is the received SNR, while S_{\min} is the minimum required SNR value to establish a connection between the UAV BS and the UE. We assume the ground BS uses a different frequency band than the UAV BS, thereby it does not create any interference to the UAV BS. Therefore, SNR is a better choice here. Moreover, note that T_s captures the receiver sensitivity of the user equipment (UE), and p_q encompasses small-scale and large-scale fading effects. Therefore, for a UE, the probability of being served by the UAV BS can be determined as

$$p_c = p_r p_q p_r, \quad (16)$$

where p_r is the probability of having enough resource for the UE at the UAV BS, such that $p_r = P(B_L \geq B_R)$, where B_L is the remaining radio resources at the UAV BS and B_R is the required radio resources for the UE. By substituting (11), (13), and (14) into (16), we get:

$$p_c = p_q p_r \frac{\pi z_d^2 \tan^2(\frac{\theta}{2})}{x_K y_K}. \quad (17)$$

Hence, it is obvious from (17) that the probability of being served by the UAV is a function of the height of the UAV BS and they have a direct proportionality. \square

Theorem 2. *The total energy consumption of the UAV BS, E_T , can be controlled by the altitude of the UAV BS, h_d .*

Proof. Let E_T be the total energy consumption of the UAV BS, E_C be the communication energy, E_M be the energy consumption during the UAV mobility. Suppose UAV moves to the optimal position and attain optimal altitude to serve the ground users.

The total energy consumption of the UAV BS can be calculated as follows:

$$E_T = E_C + E_H + E_M. \quad (18)$$

⁶It is worth noting that we assumed that the UAV BS is operating at an out-of-band spectrum, and thus it does not cause any interference to the terrestrial networks, or vice versa.

By substituting the E_M from (8) into (18), we get

$$E_T = E_C + E_H + P_h \frac{d(t_j)}{v_h} + I(\Delta h(t_j)) P_a \frac{\Delta h(t_j)}{v_a} - (1 - I(\Delta h(t_j))) P_d \frac{\Delta h(t_j)}{d}. \quad (19)$$

It is obvious from (19) that the total energy consumption of the UAV directly depends on the changes in the height of the UAV as well as the movement in the horizontal direction. \square

A. Optimization Problem Formulation

There are two primary objective functions considered in this work; namely, i) maximization of the number of served users by the UAV BS (n_c) and ii) minimization of the total energy consumption of the UAV BS (E_T). Therefore, these two objective functions can be formulated as follows:

1) *Maximization of Number of Served Users:* The number of served users by the UAV, n_c is supposed to be maximized at each time slot. Let F be the NFZ and a non-self-intersecting convex quadrilateral that is defined by its vertices as $\vec{V}_i = (x_i, y_i, z_i)$, where $i = 1, 2, 3, 4$. Moreover, let $\vec{C}_{3d} \in \mathbb{R}^3$ be a vector containing the 3-dimensional (3-D) coordinates of the UAV, and C_{2d} be a point in xy -plane, representing the projection of the UAV on the xy -plane, and imagine we draw straight lines from each vertex of F to the point C_{2d} . Then, the optimization problem can be modeled as follows:

$$\begin{aligned} & \max_{\vec{C}_{3d} \in \mathbb{R}^3} f(\vec{C}) \\ \text{s. t. } & \begin{aligned} & \mathbf{C}_1 : h_{\min} \leq h_d \leq h_{\max}, \\ & \mathbf{C}_2 : \angle A_1 + \angle A_2 + \angle A_3 + \angle A_4 < 2\pi, \\ & \mathbf{C}_3 : \theta < \pi, \\ & \mathbf{C}_4 : P_1 \leq P_{t_{\max}}, \end{aligned} \end{aligned} \quad (20)$$

where $\angle A_1 = \angle V_1 C_{2d} V_2$, $\angle A_2 = \angle V_2 C_{2d} V_3$, $\angle A_3 = \angle V_3 C_{2d} V_4$, and $\angle A_4 = \angle V_4 C_{2d} V_1$. $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ is the objective function, and $f(\vec{C}) = n_c$ in this case.

a) *Explanations of Constraints in (20):*

- **C₁:** The altitude of the UAV (h_d) is regulated in many countries and regions, such that the maximum (h_{\max}) and minimum (h_{\min}) altitudes that UAVs can flight are determined. Therefore, in this work, the UAV is supposed to obey these limitations in terms of the altitude.
- **C₂:** Since F is defined as the NFZ, it means that the UAV BS cannot fly over it. As such, this constraint confirms that the UAV BS is flying out of F , such that the projection of the UAV BS on the xy -plane, C_{2d} , is not within F .
- **C₃:** The directivity angle of the antenna of the UAV BS can be π at maximum⁷, but practically it should be less than that in order to have a better antenna gain. Though this could be normally not a hard constraint, in this work we deal with the case where the antenna angle

⁷The use of an isotropic antenna is not a good idea for UAV BSs as they are serving to the users under them in terms of height, and there is no sense to provide a radiation above the UAV BS. Therefore, we assume that the maximum antenna angle for the UAV BSs should be π .

is less than π , thereby this becomes a constraint for the optimization problem.

- **C₄**: Given that the maximum transmit power of the BSs are regulated, this constraint captures such regulations, meaning that the transmit power of the UAV BS has an upper bound.

2) *Minimization of Energy Consumption*: It is crucial to minimize the energy consumption of the UAV BS in order for it to stay in the air for a longer time so that the service that the ground users get is prolonged. Put it another way, the optimization objective elaborated in Section III-A1 focuses on maximizing the number f connected users, n_c , however, such objective is instantaneous (i.e., for a duration of a single time slot, T_d) and does not aim to maximize n_c for a period of time. The total number of connected users over a period of time considered can be calculated by

$$n_{c,t} = \sum_{i=1}^{n_{ts}} n_{c,i}, \quad (21)$$

where $n_{c,i}$ indicated the number of served users by the UAV BS during time slot i from \vec{T} . In (21), n_{ts} is a function of T_s , such that $n_{ts} = f(T_s)$, thereby although T_s is assumed to be fixed here, normally it is dependent on the energy stored in the UAV battery (i.e., battery capacity) as well as the energy consumption of the UAV BS. Since the battery capacity is fixed⁸, the only way left to prolong the UAV flight time is reducing the energy consumption. Therefore, the second objective of our problem formulation becomes the minimization of the total energy consumption of the UAV BS (E_T), and that can be modeled as follows:

$$\begin{aligned} \min_{\vec{C}_{3d} \in \mathbb{R}^3} \quad & g(\vec{C}) \\ \text{s. t.} \quad & \mathbf{C}_1 : h_{\min} \leq h_d \leq h_{\max}, \\ & \mathbf{C}_2 : \angle A_1 + \angle A_2 + \angle A_3 + \angle A_4 < 2\pi, \\ & \mathbf{C}_3 : \theta < \pi, \\ & \mathbf{C}_4 : P_t \leq P_{t_{\max}}, \end{aligned} \quad (22)$$

where $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ is the objective function, and $g(\vec{C}) = E_T$ in this case.

3) *Multi-objective Problem Formulation*: As detailed in Sections III-A1 and III-A2, there are two distinctive objectives included in our problem; i.e., maximization of connected users—as given in (20)—and minimization of the energy consumption of the UAV BSs—as given in (22). In this work, we aim at optimizing the both objectives—(20) and (22), simultaneously. In this regard, we developed the following

optimization model:

$$\begin{aligned} \max_{\vec{C}_{3d} \in \mathbb{R}^3} \quad & h(\vec{C}) \\ \text{s. t.} \quad & \mathbf{C}_1 : h_{\min} \leq h_d \leq h_{\max}, \\ & \mathbf{C}_2 : \angle A_1 + \angle A_2 + \angle A_3 + \angle A_4 < 2\pi, \\ & \mathbf{C}_3 : \theta < \pi, \\ & \mathbf{C}_4 : P_t \leq P_{t_{\max}}, \end{aligned} \quad (23)$$

where $h : \mathbb{R}^3 \rightarrow \mathbb{R}$ is the objective function, and $h(\vec{C}) = w_1 f(\vec{C}) - w_2 g(\vec{C}) = w_1 n_c - w_2 E_T$ in this case. Here, $w_1, w_2 \in \mathbb{R}$ are coefficients used for two purposes:

- To prioritize one objective over the other. For example, a mobile network operator may not be interested in the energy consumption much, and focuses only on covering as much as users as possible for a short duration, and it would choose $w_1 \gg w_2$. On the other hand, if the operator ranks both objectives equally, then it would choose $w_1 = w_2$. Therefore, w_1 and w_2 allow the operators to rank the objectives according to their requirements.
- To make the units of both $f(\vec{C})$ (unitless) and $g(\vec{C})$ (in Joules) the same, since $h(\vec{C})$ includes the summation of $f(\vec{C})$ and $g(\vec{C})$. To this end, while w_1 is chosen to be unitless, w_2 is in (1/Joules).

IV. PROPOSED Q-LEARNING BASED UAV POSITIONING MECHANISM

In RL, there is an agent taking actions to find the optimum policy for a given problem. Based on the action of the agent, first, corresponding state is observed, followed by evaluating the subsequent penalty/reward function. Then, the action-value function, storing calculated penalty/reward values for all the states and actions, is updated [22]. The agent takes action in two different ways: explore and exploit. In the initial phases of the implementation, the agent is expected to explore more in order to discover the environment better. However, after a sufficient exploration, the agent should start exploiting the available information to be able to focus on finding the best policy.

We adopted OpenAI Gym [54] tool for building environment for this study. It is based on episodic RL, where experience of each agent is divided into episodes. In initial state of each episode, we randomly localize the UAV BS and the users in a grid, and learning proceeds until the environment reaches one of the stopping criteria (this will be detailed in the following paragraphs). The main goal here is to maximize the total reward per episode and to decrease the number of episodes for achieving desired performance. RL steps in each episode are given in Algorithm 1, where s_t and s_{t+1} are the current and next states, respectively, and a_t is the current state while R_{t+1} is the expected value of the reward function.

In this study, states refer to the position of the UAV in the grid. The agent in the developed Q-learning algorithm has seven action values for each state, which denote the agent action a_t in UAV state of s_t at time t . The possible actions for

⁸We acknowledge that different UAVs have different battery capacities, but here we mean that once a UAV is selected out of various options, the battery capacity becomes something that cannot be changed/controlled.

Algorithm 1: Q -learning algorithm [22]

```

1 Initialise  $Q(s, a)$ ;
2 for each episode do
3   Initialise  $s$ 
4   done  $\leftarrow$  False ▷ A variable is introduced for a stopping criteria check
5   for each iteration do
6     while not done do
7       Select  $a_t$  ▷ Action is chosen according to the adopted policy
8       Take action,  $a_t$  ▷ The selected action is taken within the environment
9       Determine  $R_{t+1}$  ▷ Expected numeric value of the reward function is calculated
10      Obtain  $s_{t+1}$  ▷ The next state is determined
11      Update  $Q(s_t, a_t)$  ▷ The action-value function ( $Q$ -table) is updated
12       $s_t \leftarrow s_{t+1}$  ▷ The next state is assigned as the current state Check the stopping criteria
13      if Stopping criteria then
14        | done  $\leftarrow$  True ▷ If the stopping criteria is met, then execute the iterations and go the next episode
15      end
16    end
17  end
18 end

```

each state s_t are *hold, move up, move down, move left, move right, move forward, and move backward*. The agent follows ϵ -greedy [22] policy to take random actions initially—which is referred to as exploring—and decays ϵ through iterations—which is referred to as exploiting—for decreasing random actions. Given that the main goal of this study is to optimize energy consumption of the UAV along with maximizing user coverage, the reward function in the proposed method is inline with the objective function in (23), and depends on the energy consumption and coverage.

The components of the developed Q -learning algorithm for the problem of UAV BS positioning are detailed in the following paragraphs.

A. Environment

We create discrete environment with finite size (grid) representing the state of UAV in OpenAI-Gym. The size of the grids in the environment-of-interest in this study is (25, 25, 12), which is simulated with 10 meter-resolution in each axis. Therefore, the real environment size becomes (250, 250, 120) in meters. These certain dimensions of the environment are chosen by considering both the computational burden and the reality of the work; such that, the environment should be in a size of some realistic area (and the UAV BS should be able to have sufficient degree of freedom in movement) while not bringing much computational burden (the simulation time should be reasonable for us to make some tuning during the design of the algorithm). However, we intuitively confirm that the developed algorithm would work in any environment size, as the UA BS can only move slightly at one iteration thereby extending the size of the environment would not affect the performance of the algorithm other than prolonging the simulation time.

B. Agent

The UAV BS in the state s_t corresponds to the agent in this study. It will take an action, a_t , in state s_t , and it receives an observation and reward from the environment. Accordingly, it updates Q -table in order to learn the dynamics of the environment, and adapt itself to the changes. It is quite convenient to choose the UAV BS as the agent in the developed Q -learning algorithm, as it is the only one taking different actions; e.g., moving in different directions.

C. Actions

We consider seven different actions that agents can take. Let $\vec{C}_{3d} = (x_u, y_u, z_u)$ be the current position of the UAV BSs and $\vec{\tilde{C}}_{3d}$ be the position after an action taken, while r (in meters) denotes the step size in any direction. Then the list of the actions that the agent takes is as follows:

- $\vec{\tilde{C}}_{3d} = (x_u, y_u, z_u + r)$: Move up (in z direction)
- $\vec{\tilde{C}}_{3d} = (x_u, y_u, z_u - r)$: Move down (in z direction)
- $\vec{\tilde{C}}_{3d} = (x_u + r, y_u, z_u)$: Move left (in x direction)
- $\vec{\tilde{C}}_{3d} = (x_u - r, y_u, z_u)$: Move right (in x direction)
- $\vec{\tilde{C}}_{3d} = (x_u, y_u + r, z_u)$: Move forward (in y direction)
- $\vec{\tilde{C}}_{3d} = (x_u, y_u - r, z_u)$: Move backward (in y direction)
- $\vec{\tilde{C}}_{3d} = (x_u, y_u, z_u)$: Hold

D. States

We denote state s as the position of the UAV in a 3D space. We divide 3D space into grids (i.e., we discretized the state space) for having finite set of state that can be used in Q -learning. This state selection is inline with the criterion given in [22], such that the state should be affected by the actions that the agent takes. As such, the actions of the agent is fundamentally altering the 3D position of the UAV BS, which changes the state of the agent, which is also defined to be the 3D position of the agent.

E. Reward

In order to avoid the limitations of the work (or respect the constraints, in other words), a penalty mechanism is developed, such that the agents obtains a reward of -1 when the UAV BS

- goes beyond the dimensions of the environment,
- flies on the NFZ,
- does not respect any other constraint in (23).

On the other hand, a reward function is designed for the cases where the UAV BS is not in one of the states listed above. Since the main goal is to optimize energy consumption along with maximizing number of user covered, the reward is defined inline with the optimization objective in (23), such that

$$R_{t+1} = h(\vec{C}) = w_1 n_c - w_2 E_T. \quad (24)$$

The selection of the reward function as in (24) (i.e., making it equal to $h(\vec{C})$) is a legitimate decision, because the the objective of the developed Q -learning algorithm is to maximize the reward, R , and the objective function in (23) is the maximization of $h(\vec{C})$. Thus, making the reward equal to $h(\vec{C})$ is completely inline with the model in (23).

F. Policy

We follow an ϵ -greedy policy [22] in order to explore the environment by taking random actions in earlier iterations (exploration phase). As the iterations proceed (e.g., the number of the iterations get larger), we turn the exploration phase into the exploitation phase by decreasing ϵ with a decay-rate of 0.01. This is done in order to allow the agent explore and acquire new experiences during the exploration phase, while in the exploitation phase it uses the obtained experience to converge to an optimal value.

G. Q -table Update

We update the Q -table according the action a_t in the state s_t using:

$$Q(s, a) =: Q(s, a) + \alpha [R_{t+1} + \gamma \max(Q(s_{t+1})) - Q(s, a)] \quad (25)$$

where α is the learning-rate, and γ is the discount rate. Q -table update is crucial in storing the obtained experience as well as modifying it with the new data.

H. Initialization

In each episode, the UAV BS and the users are located randomly in the grid, so that the agent does not “memorize” (or it is called as overfitting in more technical terminology) a certain environment, instead produce a more generic model.

I. Episodes

The episode is considered as a snapshot of the environment in the problem formulation. The agent takes random actions in each episode and learn the environment using Q -table with (25) by evaluating the reward, R , through (24). When the the agent reaches stopping criteria, a new episode begins.

TABLE I
SIMULATION PARAMETERS

Parameter	Value
General	
Carrier frequency, f_c	1 GHz
Antenna directivity angle, θ	60°
Minimum UAV height, h_{\min}	30 m
Maximum UAV height, h_{\max}	120 m
Urban Area	250 times 250 m ²
Total number of users, n_t	100
Height from ground for all UEs	1.5 m
Speed of light, c	3e8 m/s
η_{LoS}	1.6 dB
η_{NLoS}	23 dB
Parameter of A2G path loss model, Ψ	12.08
Parameter of A2G path loss model, ζ	0.11
Number of rotors, M	4
Fluid density of the air, ρ	1.2 Kg/m ³
Rotor disk radius, β	0.25 m
Weight of the frame	1.5 Kg
Weight of the battery and payload	2 Kg
Bandwidth	180 kHz
Transmit power, P_t	30 dBm (1 W)
On-board circuit power, P_{cu}	0.01 W
Duration of hovering of UAV, t_h	1 s
Duration to communication of UAV, t_{cm}	1 s
Velocity of the UAV, v	30 m/s
Angular velocity, ω	40 rad/s
Drag coefficient	0.025
Rotor chord, c_b	0.022 m
Reference frontal area of the UAV	0.192 m ²
Q-learning	
Discount rate, γ	0.9
Epsilon, ϵ	1
Epsilon decay, ϵ -decay	0.95
Learning rate, α	0.9
Learning rate decay, α -decay	1e-4

J. Stopping Criteria

If the predefined maximum number of iterations is reached or all the users are covered by the UAV BS, the current episode is terminated, and the algorithm goes into a new episode. The maximum number of iteration is set to 2000 in this work.

V. PERFORMANCE EVALUATION

In this section, we present the performance evaluation of the proposed methodology. After describing the simulation scenario, we introduce the benchmark method as well as the performance metrics, followed by presenting the obtained results and corresponding discussions. The parameters used in the simulation campaigns are given in Table I.

A. Simulation Scenario

We implement a simulation scenario in order to evaluate the proposed Q -learning algorithm. An urban area of 250×250 m²—which is discretized by means of square-shaped grid—and $n_t = 100$ total number of users (that are normally out-of-service from the terrestrial network) are considered. Consequently, the UAV can move in the discretized 3D space in terms of x , y , and z coordinates. Furthermore, due to the regulations, we impose a minimum and a maximum altitude

of $h_{\min} = 30$ and $h_{\max} = 120$ meters, respectively, and a certain number of NFZs, corresponding to specific not allowed grids. As regards to the user mobility, we consider a random walk model, and the height from ground for all UEs is fixed to 1.5 meters. Furthermore, we assume the directivity angle $\theta = 60^\circ$ and the carrier frequency $f_c = 1$ GHz. An initialization procedure is performed in order to explicate the initial conditions. In particular, the values for all the involved parameters are determined, considering arbitrary positions for the UAV BS and the users. An outage threshold is calculated considering a required minimum received power to establish and maintain with a certain QoS a connection between the UE and the UAV BS, called $P_{r_{\min}}$. For a given transmitted power for the UAV BS, P_t —lower than the maximum allowed value, $P_{t_{\max}}$ —, and the above-mentioned $P_{r_{\min}}$, the path loss experienced by the UE, L , can be expressed by the following relation, $L = P_t - P_{r_{\min}}$. Considering the 2D position of the UAV BS (C_{2d}), the QoS constraint can be expressed in terms of L lower than L_{\max} [26], where L_{\max} is the path loss experienced by edge users. The footprint of the UAV BS, on the other hand, can be considered as a circle centered in the 2D position of the UAV BS (C_{2d}).

B. Benchmark and metrics

In this work, k -means algorithm is employed as a benchmark method, since it has been widely used in similar problems [23]. k -means is an unsupervised clustering algorithm, where the data points are clustered according to certain features. In k -means, a centroid is assigned to each cluster and the objective is to place the centroids to the position which yields minimum cumulative distance to the data points. In particular, in the initialization of the algorithm, the centroids are placed randomly and the data points are assigned to each centroid to form a cluster, and the assignment is done in a way that each data point is assigned to the cluster that is closest to it in terms of Euclidian distance. Then, the centroids are moved towards to the center of their clusters, and this process iteratively continues until the convergence, where the centroids cannot be moved anymore.

Therefore, as this algorithm finds the position of the centroid, where the cumulative distance between the centroid and the data points is the minimum, it serves as a strong benchmark for this problem. In particular, if the UAV BS is considered as the centroid, while the ground users are the data points, the k -means algorithm positions the UAV BS at a point where it is closest to the ground users in terms of distance. Given that the distance is the primary parameter affecting the link quality between the transmitter and receiver, k -means algorithm becomes a appropriate benchmark. With this algorithm, we compute the centroid position related to the actual ground users' positions. Consequently, the centroid corresponds to the best 2D position for UAV, in terms of respective distances between UAV and ground users.

Two different phases, one for training and one for testing, respectively, are performed in order to demonstrate the efficiency of the proposed Q -learning algorithm in terms of coverage and energy consumption. Regarding the coverage,

we count the number of ground users, which are normally out of service from the terrestrial network, connected to the UAV BS, and the more users covered means a better performance in terms of the coverage. In the energy consumption, we measure the total energy consumption, E_T , of the UAV BS while it is providing service to the users, and the less energy consumption refers to a better performance in terms of the energy consumption as the flight time of the UAV BS is prolonged.

During the training phase of the developed Q -learning algorithm, the simulation is conducted through a certain number of episodes, in order to populate the related Q -table and consequently achieve the needed learning. A trade-off between coverage and energy consumption prioritization is considered both in training and testing phases. Specifically, five different experiments are performed with different values of the weights (e.g., w_1 and w_2) that are responsible of prioritizing the coverage or the energy consumption.

C. Simulation Results

Fig. 2 demonstrates the averaged and normalized results in terms of energy consumption (orange bars) and covered users (green bars) for different altitudes through k -means algorithm. It is worth noting that, since the k -means algorithm determines the 2D position of the UAV BS, the altitude of the UAV BS should also be determined. Although there are different methods in order to determine the altitude, such as a trigonometric approach is used in [23], those usually do not consider the altitude regulations for the UAVs. In this work, on the other hand, considering such regulations, we used different fixed levels for the altitude of the UAV BS. In particular, three different altitude levels are considered: i) minimum allowed altitude ($h_{\min}=30$), maximum allowed altitude ($h_{\max}=120$ m), and the middle point between the two ($\frac{h_{\max} + h_{\min}}{2} = 75$ m).

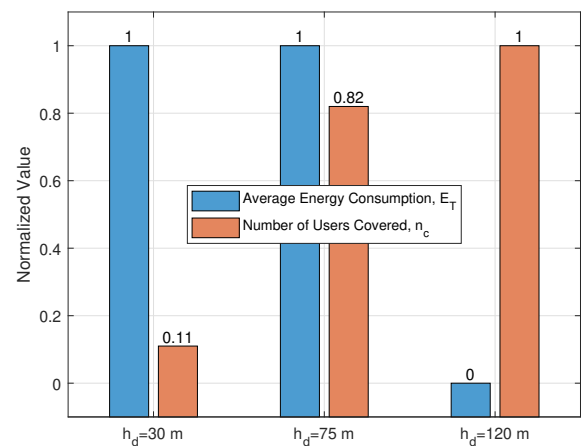


Fig. 2. Single UAV positioning for different altitudes through k -means.

From (11) it is understood that the value in the second case, i.e. with maximum allowed height of 120 meters, can

be assumed as the upper-bound in terms of coverage, since in this case the UAV is placed in the best 2D position with the maximum allowed height, that means the maximum achievable coverage area with respect to the size of the considered urban area, consequently obtaining the maximum number of covered users. A similar consideration can be done for this case (i.e., $h_{\max} = 120$ m) in terms of energy consumption. Since all ground users are covered at the first iteration, one of the stopping criteria is readily matched, thereby no movement is performed by the UAV, resulting in energy consumption due to mobility equals to zero. For the two remaining cases, with altitudes of 30 and 75 meters, respectively, the results in terms of coverage can be considered as lower-bound and median values, since as previously stated, the coverage area, and subsequently, the number of covered user are highly dependent on the considered altitude of the UAV BS. Lastly, considering the energy consumption results, the UAV BS exploits the maximum number of allowed iterations attempting to match the stopping criteria for coverage, consequently resulting in the maximum value for energy consumption due to mobility.

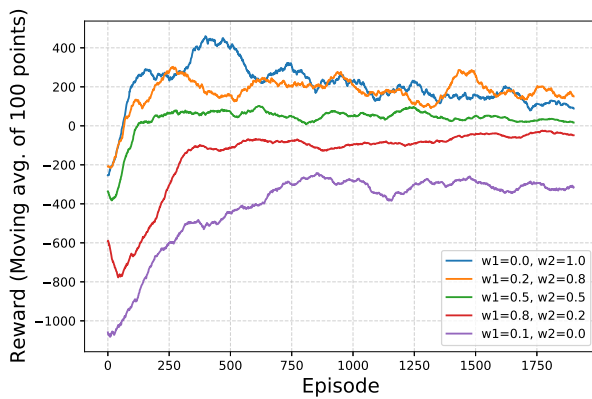


Fig. 3. Q -learning algorithm convergence in terms of rewards for different set of weights for 2000 episodes.

Fig. 3 presents the results in terms of achieved rewards, after an initial phase, the convergence of the Q -learning algorithm occurs, demonstrating the effectiveness of the Q -learning algorithm. One of the important takeaway from the findings in Fig. 3 is that, regardless of the weighting approach (e.g., for different w_1 and w_2 values), the designed Q -learning algorithm converges to the final reward. This confirms the proper design of the algorithm, and is a clear sign that it can work in various scenarios.

Following the above assumptions, the efficiency of the proposed Q -learning algorithm is verified through the testing phase, with the UAV BS positioning optimization through k -means as a benchmark. In this phase, the UAV BS is located in the above-mentioned simulated scenario and an arbitrary uniform distribution for the ground users is considered. The testing phase is conducted performing a certain number of runs, in order to average the results with regard to the specified parameters. Fig. 4 shows the single UAV BS position optimization for different set of weights and the results

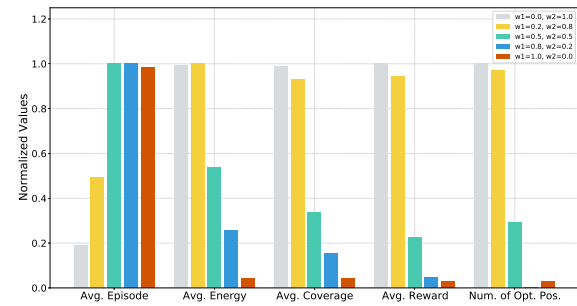


Fig. 4. Monte Carlo test results. Single UAV position optimisation comparison for different set of weights.

are the normalized version of the average values between 0 and 1. The goodness of considering a trade-off between coverage and energy consumption, achieved by means of the two different weights, is mostly visible in two of the five experiments. In particular, it can be seen that for the $w_1 = 0.2$ and $w_2 = 0.8$, the best average energy consumption is achieved, whereas for the weights $w_1 = 0.0$ and $w_2 = 1.0$, best overall coverage, and rewards are obtained. In other words, the proper performances of the above-mentioned trade-off can be observed from the results in Fig. 4. Effectively, when energy consumption is not prioritized, the UAV BS finds the optimum position in fewer episodes but at the expense of a higher energy consumption, conversely in the remaining cases. Therefore, it is worth mentioning that the designed weighting mechanism works well, as the performance of the Q -learning algorithm is deeply affected by the numerical values of the weights. However, these results do not only affirm that the weighting mechanism works, but also gives superiority to the proposed approach, as it can converge to a solution according to the requirements of the network operators.

VI. CONCLUSION

In this paper, a smart UAV BS positioning mechanism was proposed by taking altitude regulations as well as NFZs into account along with some hard constraints, including maximum transmit power and directivity of the UAV BS antenna, to provide sustainable wireless coverage and services to the ground users under more realistic conditions. As such, first, two different optimization models were developed for the minimization of the energy consumption and maximization of the number of users covered. Then, these two distinctive models are combined with a weighting mechanism, and a multi-objective optimization problem formulation was developed. With the developed weighting mechanism, wireless networks operators become capable of positioning the UAV BSs according to their requirements by relatively ranking the energy consumption and the number of users covered. We proposed a Q -learning-based approach for UAV-assisted communication systems, and the OpenAI Gym tool was used to build the RL environment. The objective is to find the optimal position of the UAV and minimize the

energy consumption while maximizing the number of users covered. The results demonstrate that the proposed solution outperforms the baseline k -means method in terms of covered users, while achieving the desired minimization of the energy consumption.

REFERENCES

- [1] Ericsson, "Ericsson Mobility Report," Ericsson, Tech. Rep., June 2021. [Online]. Available: <https://www.ericsson.com/4a03c2/assets/local/mobility-report/documents/2021/june-2021-ericsson-mobility-report.pdf>
- [2] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.
- [3] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A Tutorial Overview of Standards, Trials, Challenges, Deployment, and Practice," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 6, pp. 1201–1221, June 2017.
- [4] F. Jameel, Faisal, M. A. A. Haider, and A. A. Butt, "Massive MIMO: A survey of recent advances, research issues and future directions," in *2017 International Symposium on Recent Advances in Electrical Engineering (RAEE)*, Oct 2017, pp. 1–6.
- [5] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive mimo communication for future wireless systems: A survey," *IEEE Communications Surveys Tutorials*, vol. 20, no. 2, pp. 836–869, 2018.
- [6] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of Millimeter Wave Communications for Fifth-Generation (5G) Wireless Networks—With a Focus on Propagation Models," *IEEE Transactions on Antennas and Propagation*, vol. 65, no. 12, pp. 6213–6230, Dec 2017.
- [7] 3GPP, "5G; NR; Base Station (BS) radio transmission and reception," 3rd Generation Partnership Project (3GPP), Technical Specification (TS) 38.104, 10 2019, version 15.5.0.
- [8] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on uavs for wireless networks: Applications, challenges, and open problems," *IEEE Communications Surveys Tutorials*, vol. 21, no. 3, pp. 2334–2360, 2019.
- [9] I. Bor-Yaliniz and H. Yanikomeroglu, "The new frontier in ran heterogeneity: Multi-tier drone-cells," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 48–55, 2016.
- [10] M. Öztürk, "Cognitive networking for next generation of cellular communication systems," Ph.D. dissertation, University of Glasgow, 2020.
- [11] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70 – 90, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0168169917308803>
- [12] W. Lin, Y. Hu, and C. Tsai, "Machine learning in financial crisis prediction: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 421–436, July 2012.
- [13] J. Archenaa and E. M. Anita, "A survey of big data analytics in healthcare and government," *Procedia Computer Science*, vol. 50, pp. 408 – 413, 2015, Big Data, Cloud and Computing Challenges. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1877050915005220>
- [14] P. V. Klaine, M. A. Imran, O. Onireti, and R. D. Souza, "A Survey of Machine Learning Techniques Applied to Self-Organizing Cellular Networks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2392–2431, Fourthquarter 2017.
- [15] C. Zhang, P. Patras, and H. Haddadi, "Deep Learning in Mobile and Wireless Networking: A Survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2224–2287, thirdquarter 2019.
- [16] M. E. Morocho Cayamcela and W. Lim, "Artificial Intelligence in 5G Technology: A Survey," in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, Oct 2018, pp. 860–865.
- [17] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of Machine Learning in Wireless Networks: Key Techniques and Open Issues," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3072–3108, Fourthquarter 2019.
- [18] H. Zhang and L. Dai, "Mobility Prediction: A Survey on State-of-the-Art Schemes and Future Applications," *IEEE Access*, vol. 7, pp. 802–822, 2019.
- [19] A. Dogra, R. K. Jha, and S. Jain, "A Survey on beyond 5G network with the advent of 6G: Architecture and Emerging Technologies," *IEEE Access*, pp. 1–1, 2020.
- [20] —, "A Survey on beyond 5G network with the advent of 6G: Architecture and Emerging Technologies," *IEEE Access*, pp. 1–1, 2020.
- [21] Y. Lu and X. Zheng, "6G: A survey on technologies, scenarios, challenges, and the related issues," *Journal of Industrial Information Integration*, vol. 19, p. 100158, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2452414X20300339>
- [22] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*. MIT press, 2018.
- [23] M. Ozturk, J. P. B. Nadas, P. H. V. Klaine, S. Hussain, and M. A. Imran, "Clustering based uav base station positioning for enhanced network capacity," in *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)*, 2020, pp. 1–6.
- [24] X. Zhang and L. Duan, "Energy-saving deployment algorithms of UAV swarm for sustainable wireless coverage," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 10 320–10 335, 2020.
- [25] R. de Paula Parisotto, P. V. Klaine, J. P. B. Nadas, R. D. Souza, G. Brante, and M. A. Imran, "Drone base station positioning and power allocation using reinforcement learning," in *2019 16th International Symposium on Wireless Communication Systems (ISWCS)*, 2019, pp. 213–217.
- [26] L. Wang, B. Hu, and S. Chen, "Energy efficient placement of a drone base station for minimum required transmit power," *IEEE Wireless Communications Letters*, vol. 9, no. 12, pp. 2010–2014, 2020.
- [27] I. Donevski and J. J. Nielsen, "Dynamic standalone drone-mounted small cells," in *2020 European Conference on Networks and Communications (EuCNC)*, 2020, pp. 342–347.
- [28] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 569–572, 2014.
- [29] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Unmanned aerial vehicle with underlaid device-to-device communications: Performance and tradeoffs," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 3949–3963, 2016.
- [30] Z. Becvar, P. Mach, and M. Nikoouroo, "Reducing energy consumed by repositioning of flying base stations serving mobile users," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–7.
- [31] H. Qi, Z. Hu, H. Huang, X. Wen, and Z. Lu, "Energy efficient 3-D UAV control for persistent communication service and fairness: A deep reinforcement learning approach," *IEEE Access*, vol. 8, pp. 53 172–53 184, 2020.
- [32] K. Dorling, J. Heinrichs, G. G. Messier, and S. Magierowski, "Vehicle routing problems for drone delivery," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 1, pp. 70–85, 2017.
- [33] C. Xu, X. Liao, J. Tan, H. Ye, and H. Lu, "Recent research progress of unmanned aerial vehicle regulation policies and technologies in urban low altitude," *IEEE Access*, vol. 8, pp. 74 175–74 194, 2020.
- [34] A. Fotouhi, H. Qiang, M. Ding, M. Hassan, L. G. Giordano, A. Garcia-Rodriguez, and J. Yuan, "Survey on uav cellular communications: Practical aspects, standardization advancements, regulation, and security challenges," *IEEE Communications Surveys Tutorials*, vol. 21, no. 4, pp. 3417–3442, 2019.
- [35] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Mobile unmanned aerial vehicles (uavs) for energy-efficient internet of things communications," *IEEE Transactions on Wireless Communications*, vol. 16, no. 11, pp. 7574–7589, 2017.
- [36] Y. Pang, Y. Zhang, Y. Gu, M. Pan, Z. Han, and P. Li, "Efficient data collection for wireless rechargeable sensor clusters in harsh terrains using uavs," in *2014 IEEE Global Communications Conference*, 2014, pp. 234–239.
- [37] X. Liu and N. Ansari, "Resource allocation in uav-assisted m2m communications for disaster rescue," *IEEE Wireless Communications Letters*, vol. 8, no. 2, pp. 580–583, 2019.
- [38] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Drone small cells in the clouds: Design, deployment and performance analysis," in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–6.

- [39] —, “Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage,” *IEEE Communications Letters*, vol. 20, no. 8, pp. 1647–1650, 2016.
- [40] A. M. Hayajneh, S. A. R. Zaidi, D. C. McLernon, and M. Ghogho, “Drone empowered small cellular disaster recovery networks for resilient smart cities,” in *2016 IEEE International Conference on Sensing, Communication and Networking (SECON Workshops)*, 2016, pp. 1–6.
- [41] 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 *2016 IEEE International Conference on Communications (ICC)*, 2016, pp. 1–5.
- [42] E. Kalantari, H. Yanikomeroglu, and A. Yongacoglu, “On the number and 3d placement of drone base stations in wireless cellular networks,” in *2016 IEEE 84th Vehicular Technology Conference (VTC-Fall)*, 2016, pp. 1–6.
- [43] J. Košmerl and A. Vilhar, “Base stations placement optimization in wireless networks for emergency communications,” in *2014 IEEE International Conference on Communications Workshops (ICC)*, 2014, pp. 200–205.
- [44] P. Zhan, K. Yu, and A. Lee Swindlehurst, “Wireless relay communications using ann unmanned aerial vehicle,” in *2006 IEEE 7th Workshop on Signal Processing Advances in Wireless Communications*, 2006, pp. 1–5.
- [45] Q. Wu, Y. Zeng, and R. Zhang, “Joint trajectory and communication design for multi-uav enabled wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 2109–2121, 2018.
- [46] F. Jiang and A. L. Swindlehurst, “Optimization of uav heading for the ground-to-air uplink,” *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 5, pp. 993–1005, 2012.
- [47] Y. Zeng, R. Zhang, and T. J. Lim, “Throughput maximization for uav-enabled mobile relaying systems,” *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 4983–4996, 2016.
- [48] T. M. Cabreira, C. D. Franco, P. R. Ferreira, and G. C. Buttazzo, “Energy-aware spiral coverage path planning for uav photogrammetric applications,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3662–3668, 2018.
- [49] X. Shao, Y.-J. Gong, Z.-H. Zhan, and J. Zhang, “Bipartite cooperative coevolution for energy-aware coverage path planning of uavs,” *IEEE Transactions on Artificial Intelligence*, pp. 1–1, 2021.
- [50] D. Yang, Q. Dan, L. Xiao, C. Liu, and L. Cuthbert, “An efficient trajectory planning for cellular-connected uav under the connectivity constraint,” *China Communications*, vol. 18, no. 2, pp. 136–151, 2021.
- [51] G. Sun, J. Li, Y. Liu, S. Liang, and H. Kang, “Time and energy minimization communications based on collaborative beamforming for uav networks: A multi-objective optimization method,” *IEEE Journal on Selected Areas in Communications*, pp. 1–1, 2021.
- [52] Y. Yu, J. Tang, J. Huang, X. Zhang, D. K. C. So, and K.-K. Wong, “Multi-objective optimization for uav-assisted wireless powered iot networks based on extended ddpq algorithm,” *IEEE Transactions on Communications*, pp. 1–1, 2021.
- [53] Y. Zeng, J. Xu, and R. Zhang, “Energy minimization for wireless communication with rotary-wing uav,” *IEEE Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2329–2345, 2019.
- [54] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, “OpenAI gym,” *CoRR*, vol. abs/1606.01540, 2016. [Online]. Available: <http://arxiv.org/abs/1606.01540>