Machine Learning for Predictive On-Demand Deployment of UAVs for Wireless Communications

Qianqian Zhang[1], Mohammad Mozaffari[1], Walid Saad[1], Mehdi Bennis[2], and Mérouane Debbah[3,4]
[1]Bradley Department of Electrical and Computer Engineering, Virginia Tech, VA, USA, Emails: {qqz93,mmozaff,walids}@vt.edu.
[2]Center for Wireless Communications, University of Oulu, Finland, Email: bennis@ee.oulu.fi.
[3]Mathematical and Algorithmic Sciences Lab, Huawei France R&D, Paris, France, Email: merouane.debbah@huawei.com.
[4]Large Systems and Networks Group, CentraleSupélec, Université Paris-Saclay, 3 rue Joliot-Curie, 91192 Gif-sur-Yvette, France.

***Abstract*—In this paper, a novel machine learning (ML) framework is proposed for enabling a predictive, efficient deployment of unmanned aerial vehicles (UAVs), acting as aerial base stations (BSs), to provide on-demand wireless service to cellular users. In order to have a comprehensive analysis of cellular traffic, an ML framework based on a Gaussian mixture model (GMM) and a weighted expectation maximization (WEM) algorithm is introduced to predict the potential network congestion. Then, the optimal deployment of UAVs is studied with the objective of minimizing the power needed for UAV transmission and mobility, given the predicted traffic. To this end, first, the optimal partition of service areas of each UAV is derived, based on a fairness principle. Next, the optimal location of each UAV that minimizes the total power consumption is derived. Simulation results show that the proposed ML approach can reduce the required downlink transmit power and mobility power by over 20% and 80%, respectively, compared with an optimal deployment of UAVs with no ML prediction.**

I. Introduction

The demand for cellular data is experiencing an unprecedented increase. The next generation, 5G wireless cellular network is estimated to support a 200 fold increase in wireless data traffic by 2030 [1]. To cope with this exponential increase in demand, there has been growing interest in network densification for cellular systems as a means to improve spectrum efficiency and cellular network capacity.

The need for additional base stations (BSs) is more pronounced in cellular hotspot areas that exhibit a steep surge in data demands during temporary events, such as concerts and football games. To satisfy such temporary surges in traffic, the use of an unmanned aerial vehicle (UAV) as an aerial BS can be a more flexible and cost-effective approach, compared with a traditional, ground BS [2]. A mobile UAV can intelligently change its position, which is suitable to provide on-demand wireless service to ground users, thus overcoming coverage holes and alleviating congestions [3].

In order to deploy UAVs in a timely and flexible manner, network operators must be able to predict potential hotspots and congestion events a priori. To this end, there is a need to apply machine learning (ML) techniques to analyze demand patterns [4]. The ability of ML to exploit big data analytics enables a comprehensive prediction of a network's traffic amount and data distribution. By using such predictions, aerial UAV BSs can be optimally deployed to the target area beforehand thus providing an on-demand, delay-free and power-efficient wireless service to ground users.

The use of UAVs as cellular BSs has been addressed in [3], [5]–[7]. Deploying UAVs to optimize downlink coverage is studied in [5]. Meanwhile, in [3], the authors studied the use of UAVs as flying BSs to provide energy-efficient service to wireless users. Moreover, the work [6] focuses on using UAVs as relays, and the work in [7] studies an energy-efficient trajectory design. However, most of the existing works assume a time-invariant wireless network, or a given distribution of cellular users. To properly analyze an on-demand deployment of UAVs, the temporal and spatial patterns of the cellular traffic data must be predicted so as to optimally deploy UAVs to satisfy a time-varying data demand.

There are existing woks, such as [6], [8], and [9], that apply ML techniques to optimize UAV deployment. In [6], a neural model is formulated to study the map of UAVs to each hotspot areas. The authors in [8] studied the trajectory optimization using neural networks, while a segmented regression approach is proposed in [9] for UAV channel modeling, based on the terrain topology. However, none of these works demonstrates the benefit of applying ML to deploy UAVs on-demand and improve power efficiency and network performance. In order to analyze the data traffic of cellular networks, the authors in [10] studied a BS sleeping strategy for minimizing power consumption. However, the authors focused only on a low-traffic cellular network, which is not scalable for the more practical, congested scenarios.

The main contribution of this paper is a novel machine learning framework that enables operators to predict congestions and hotspot events, and subsequently, deploy temporary UAV BSs to provide aerial wireless service to mobile users, while minimizing the UAV power needed for downlink communications and mobility. We consider a heterogeneous cellular network, in which ground BSs can offload the wireless service to aerial UAVs when the predicted data demand of mobile users exceeds the network capacity. To guarantee a no-delay wireless service, a Gaussian mixture model (GMM) is introduced based on a weighted expectation maximization (WEM) algorithm [11] to predict the cellular data traffic. Then, the optimal deployment of UAVs is studied to minimize the power needed for UAV transmission and mobility, given the predicted traffic. To this end, we first study the division of service areas, based on a fairness principle. Then, we derive the optimal UAV locations that can minimize the total power consumption of the network. To the best of our knowledge, this is the first work that leverages ML to

predictively deploy UAVs as aerial BSs. Simulation results show that the proposed ML approach can reduce the required downlink transmit power and mobility power by over 20% and 80%, respectively, compared with an optimal deployment of UAVs with no ML prediction.

The rest of this paper is organized as follows. Section II presents the system model and problem formulation. Section III outlines the proposed ML and UAV deployment framework. Simulation results are presented in Section IV, while conclusions are drawn in Section V.

II. System Model and Problem Formulation

Consider a time-variant heterogeneous cellular network that serves a group of cellular users distributed in a geographical area $\mathcal{A}$. The cellular network consists of a set $\mathcal{I}$ of I UAVs and a set $\mathcal{J}$ of J BSs. Each user can receive data from both ground BSs and UAVs. Initially, a traditional BS will be chosen to serve the wireless users. However, if the downlink of the ground cellular system is overloaded due to heavy traffic, the ground BS will request the deployment of UAV BSs to offload some of its users.

Ground BSs and UAVs employ different frequency bands for downlink communications. Each UAV is equipped with directional antennas that enable beamforming. Therefore, interference among UAVs is negligible. Furthermore, each UAV adopts a frequency division multiple access (FDMA) technique and assigns a dedicated channel to one of its downlink users. Hereinafter, we use the notion of an *aerial cell* to indicate the service area of each UAV, and *aerial cellular users* to indicate users that are served by UAV cellular BSs.

Each UAV has a limited energy resource, that must be efficiently used for joint communications and mobility. To this end, the UAVs should intelligently change their positions to meet the required users' data rates, as well as to minimize their transmission power. However, given the cellular network is time-variant, the cellular traffic demand will change over time, which complicates the efficient deployment. To guarantee timely aerial service without having UAVs continuously moving, the network operator can use ML techniques to predict its network's data demand, and then, request the deployment of UAV BSs to the predicted hotspot areas, before the congestion occurs.

A. Air-to-ground channel model

Given a typical ground receiver located at $(x, y) \in \mathcal{A}$ and a UAV $i \in \mathcal{I}$ located at (x_i, y_i, h_i), the path loss of the downlink communication from UAV i to the receiver will be [12]:

$$L_i(x,y)[dB] = 20\log\left(\frac{4\pi f_c d_i(x,y)}{c}\right) + \xi_i, \qquad (1)$$

where $d_i(x,y) = \sqrt{(x-x_i)^2 + (y-y_i)^2 + h_i^2}$ is the distance between the ground receiver and UAV i, f_c is the carrier frequency, c is the speed of light, and ξ_i is the average additional loss to the free space propagation loss which depends on the environment. If the wireless link between UAV i and a ground user is line-of-sight (LOS), $\xi_i^{\text{LOS}} \sim N(\mu_{\text{LOS}}, \sigma^2_{\text{LOS}})$; otherwise, the non-line-of-sight (NLOS) link has an additional loss of $\xi_i^{\text{NLOS}} \sim N(\mu_{\text{NLOS}}, \sigma^2_{\text{NLOS}})$. The NLOS link will experience a high path loss due to shadowing and reflection. The probability of existence of LOS links between UAV i and the ground user will then be [12]:

$$p_i^{\text{LOS}}(x,y) = \frac{1}{1 + a\exp\left(-b[\frac{180}{\pi}\theta_i(x,y) - a]\right)}, \qquad (2)$$

where a and b are constant values which depend on the environment, and $\theta_i(x,y) = \sin^{-1}(\frac{h_i}{d_i(x,y)})$ is the elevation angle of UAV i with respect to the receiver. Then, the probability of having a NLOS link is $p_i^{\text{NLOS}}(x,y) = 1 - p_i^{\text{LOS}}(x,y)$ [3].

Consequently, the average path loss from UAV i to the ground reciever at (x, y) in the linear scale can be given as

$$\bar{L}_i(x,y) = p_i^{\text{LOS}}(x,y)L_i^{\text{LOS}}(x,y) + p_i^{\text{NLOS}}(x,y)L_i^{\text{NLOS}}(x,y). \qquad (3)$$

Therefore, the downlink capacity that UAV i can provide to a mobile user located at (x, y) will be:

$$R_i(x,y) = W_i \log_2\left(1 + \frac{P_i(x,y)G_i(x,y)/\bar{L}_i(x,y)}{W_i n_0}\right), \qquad (4)$$

where W_i is the transmission bandwidth of UAV i, $P_i(x,y)$ is the transmission power, $G_i(x,y)$ is the antenna gain of UAV i, and n_o is the average noise power spectral density. For tractability, we assume a perfect beam alignment between the UAV and the mobile receiver, and each UAV has the same antenna gain. Therefore, $G_i(x,y) = G$, which is a constant for all $i \in \mathcal{I}$ and $(x,y) \in \mathcal{A}$. Assume that the total available bandwidth of UAV i is B_i and the number of mobile users associated with UAV i is N_i, then the downlink bandwidth of each channel will be $W_i = B_i/N_i$.

The number N_i of aerial users that are served by UAV i within its aerial cell is given by:

$$N_i = N \int\int_{\mathcal{A}_i} f(x,y)\,\mathrm{d}x\,\mathrm{d}y, \qquad (5)$$

where $N = \sum_{i\in\mathcal{I}} N_i$ is the total number of aerial users, $\mathcal{A}_i$ is the service area of UAV i, and $f(x,y)$ is the distribution of aerial users. In order to provide a universal wireless service, the aerial cells of all UAVs should fully cover the geographical area $\mathcal{A}$ without overlap. That is, $\cup_{i\in\mathcal{I}}\mathcal{A}_i = \mathcal{A}$, and for $i \neq j$ and $i, j \in \mathcal{I}$, $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$. However, note that, the value of N and the user distribution $f(x,y)$ will change, according to the offloading requests from ground BSs.

B. Cellular traffic analysis

To provide an on-demand service, network operators need to change the UAVs' locations frequently, according to the offload requests from ground BSs, to satisfy the instant traffic demand. However, such continuous movement will consume excessive power. To efficiently deploy UAVs while guaranteeing a no-delay wireless service, a dataset of the cellular traffic history can be exploited by the network operator for

traffic prediction. This dataset, represented by a matrix $\boldsymbol{Q}$, records discrete data during each time period T for M days:

$$\boldsymbol{Q} = [N(x, y, t),\ D(x, y, t) | \forall t \in \mathcal{T}, (x, y) \in \mathcal{A}], \quad (6)$$

where $\mathcal{T} = \{T, 2T, \cdots, 24M\}$ is a discrete set of time, and the unit of T is hour. The first item $N(x, y, t)$ represents the number of aerial users that are offloaded from a BS at (x, y) to a UAV during a time interval from t to $t + T$, and the second item $D(x, y, t)$ denotes the amount of cellular traffic that a UAV needs to provide for the aerial users from a BS at (x, y) during the period from t to $t + T$.

Let N be the total number of aerial users, D be the total amount of aerial cellular traffic, $f(x, y)$ be the spatial distributions of aerial users, and $g(x, y)$ be the spatial distribution of aerial data traffic in $\mathcal{A}$. Without a comprehensive analysis of $\boldsymbol{Q}$, the values of N, D, $f(x, y)$ and $g(x, y)$ will change over time, based on the offloading requests of ground BSs, which causes a frequent movement of UAVs to meet the instant traffic demand, and excessive power consumed on mobility.

Therefore, our goal is to develop a centralized ML approach to predict N and $f(x, y)$ based on $N(x, y, t)$, and D and $g(x, y)$ based on $D(x, y, t)$, such that at the beginning of each period T, network operators can optimally deploy UAVs to minimize the power consumptions, while during each interval the locations of UAVs remain fixed.

C. Data rate requirement

Given the predicted information on the total amount of aerial cellular traffic D, and the distribution of aerial cellular traffic $g(x, y)$, the average data rate requirement within a service area $\mathcal{A}_i$ of UAV i can be given by

$$\alpha_i = \frac{1}{T} \int \int_{\mathcal{A}_i} D \cdot g(x, y) \, \mathrm{d}x \, \mathrm{d}y. \quad (7)$$

Since the communication capacity of UAV i should be greater than or equal to the rate demand of all users in its aerial cell $\mathcal{A}_i$, we formulate the data rate requirement as follows,

$$\int \int_{\mathcal{A}_i} R_i(x, y) \, \mathrm{d}x \, \mathrm{d}y \geq \alpha_i, \quad (8)$$

i.e.,

$$R_i(x, y) \geq \frac{Dg(x, y)}{T}. \quad (9)$$

We define $\beta(x, y) = \frac{Dg(x,y)}{T}$ as the average minimum data rate requirement for the aerial user at (x, y). Based on (4) and (9), the minimum transmit power that UAV i should provide to communicate with the user at (x, y) will be:

$$P_i^{\min}(x, y) = \frac{B_i n_0 \bar{L}_i(x, y)}{G N_i} \left(2^{\beta(x,y) N_i / B_i} - 1 \right). \quad (10)$$

Note that, the values of N_i and $\beta(x, y)$ in (10) will depend on the output of the cellular traffic analysis.

Consequently, the total transmit power of all UAVs needed to satisfy the data demand of all aerial users in $\mathcal{A}$ will be:

$$P_c = \sum_{i \in \mathcal{I}} \int \int_{\mathcal{A}_i} P_i^{\min}(x, y) \, \mathrm{d}x \, \mathrm{d}y. \quad (11)$$

Without loss of generality, we assume that the maximum transmission power of UAVs is sufficient to meet the data demand of aerial users. Meanwhile, the total power for each UAV $i \in \mathcal{I}$ to move from its current location (x_i^o, y_i^o, h_i^o) to the new location (x_i, y_i, h_i) will be:

$$P_t = \gamma \sum_{i \in \mathcal{I}} \left[(x_i^o - x_i)^2 + (y_i^o - y_i)^2 + (h_i^o - h_i)^2 \right]^{\frac{1}{2}}, \quad (12)$$

where γ is the rate of energy consumption a UAV needs to move by one meter.

Then, the second objective is to jointly find the optimal location and the partition of the service area $\mathcal{A}_i$ for each UAV $i \in \mathcal{I}$, such that the total power used for downlink transmissions and mobility can be minimized, i.e.,

$$\min_{\mathcal{A}_i, x_i, y_i, h_i} \quad P_c + P_t, \quad (13a)$$

$$\text{s. t.} \quad \frac{\int \int_{\mathcal{A}_i} P_i^{\min}(x, y) \, \mathrm{d}x \, \mathrm{d}y}{P_i^{\mathrm{a}}} = \kappa, \forall i \in \mathcal{I}, \quad (13b)$$

$$\cup_{i \in \mathcal{I}} \mathcal{A}_i = \mathcal{A}, \quad (13c)$$

$$\mathcal{A}_i \cap \mathcal{A}_j = \emptyset, \forall i \neq j \in \mathcal{I}, \quad (13d)$$

where P_i^{a} is the available power of UAV i, and κ is a constant for all $i \in \mathcal{I}$. The first constraint represents a fairness principle, whereby the ratio of the data traffic offloaded to each UAV equals to the ratio of the available power of each UAV. The second and third constraints guarantee that the service areas of all UAVs fully cover $\mathcal{A}$ without overlap.

Note that, without an ML analysis, the function $P_i^{\min}$, as well as P_c, will change, based on the offloading requests of ground BSs. Thus, the network operator needs to reorganize the aerial cellular system to meet the instant traffic demand frequently. However, with the predicted information of cellular traffic, the optimal problem (13) is fixed within each period T. Therefore, at the beginning of each interval, UAVs are deployed according to the solution of (13), and within the period, the location and aerial cell of each UAV remain unchanged.

III. Proposed Prediction and UAV Deployment Framework

Next, we propose a novel approach to address the aforementioned problems. First, a centralized ML approach will be proposed to predict the values of N, D, $f(x, y)$ and $g(x, y)$ for each time interval T. With the prediction information, the power minimization problem in (13) will be solved to optimally deploy each UAV.

A. Cellular traffic prediction

In order to have a robust and practical analysis, we use the real dataset [1] of City Cellular Traffic Map [13], which records the time, the location of each BS, the number of mobile users, and the total amount of data that each BS serves during each hour, from Aug. 19 to Aug. 26, 2012, in a median-size city in China. We assume that the maximum number of mobile users that each BS can serve within one hour is a fixed

[1] Our approach can accommodate other datasets without loss of generality.

number of N_m, and the maximum amount of cellular data is a constant D_m for all BSs. Thus, a new dataset is generated to capture the traffic of the aerial cellular network as $\boldsymbol{Q}^{'} = [N(x,y,t)-N_m,\ D(x,y,t)-D_m | \forall t \in \mathcal{T}, (x,y) \in \mathcal{A}]$, in which $N(x,y,t)-N_m$ is the number of aerial users from hour t to $t+1$, and $D(x,y,t)-D_m$ is the amount of aerial cellular traffic. For notation simplicity, hereinafter, we use $\boldsymbol{Q}$ to denote the aerial traffic dataset, instead of $\boldsymbol{Q}^{'}$. Since $N(x,y,t)$ and $D(x,y,t)$ have an analogous data structure, a similar approach will be applied to analyze $N(x,y,t)$ and $D(x,y,t)$. For simplicity, we keep the following discussion only on $D(x,y,t)$. Therefore, the objective is to use ML to formulate the temporal and spatial pattern of $D(x,y,t)$.

There are three key assumptions in the following ML analysis. First, due to the periodicity of human activity, the cellular traffic presents a repetitive daily pattern [14]. Based on this observation, we assume that the total cellular traffic during a specific hour of different days follows the same distribution. Therefore, we divide the dataset into 24 subsets, by merging the data of the same hour from different days. Second, we assume that the traffic amount between each hour of one day is independent. Therefore, given the 24 sub-datasets, 24 independent models will be built to study the pattern of each objective value of each hour. Furthermore, we assume that the temporal feature of $D(x,y,t)$ is independent from the spatial distribution. As a result, two separate models will be studied to identify the temporal feature $D(t)$ and the spatial feature $g(x,y)$ of $D(x,y,t)$ for each hour.

The model to capture the temporal and spatial characteristics of $D(x,y,t)$ relies on a GMM, which assumes that the data distribution can be modeled by the sum of multiple Gaussians with different weights as [15] $p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$, where x is a general data point, $p(x)$ is the probability distributed at x, K is the number of individual Gaussian models in GMM, and $\sum_{k=1}^{K} \pi_k = 1$, $\pi_k \in [0,1]$ is the mixing coefficient for each Gaussian.

1) Spatial distribution model: First, we study the modeling approach of the spatial feature $g(x,y)$ of $D(x,y,t)$. Given a time $t \in \{1,2,\cdots,24\}$, the data distribution of the cellular traffic from t to $t+1$ can be calculated by

$$g^t(x,y) = \frac{D(x,y,t)}{\int\int_{\mathcal{A}} D(x,y,t)\,\mathrm{d}x\,\mathrm{d}y}. \tag{14}$$

Then, a dataset $\boldsymbol{G}^t$ is formed by all the distribution $g^t(x,y)$ of M days for the specific hour t, and we seek to build a GMM to capture a pattern of data distribution for time t as

$$g^t(\boldsymbol{x}) = \sum_{k=1}^{K^t} \pi_k^t \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k^t, \boldsymbol{\Sigma^t}_k), \tag{15}$$

where $\boldsymbol{x} = (x,y)$ is the location vector. To find the parameters of K^t, π_k^t, $\boldsymbol{\mu}_k^t$, and $\boldsymbol{\Sigma}_k^t$, for a given $t \in \{1,2,\cdots,24\}$, and $k \in \{1,\cdots,K^t\}$, first, a classification approach based on a weighted K-means method is used to group the data $\boldsymbol{x}$ into K clusters, and the weight $D(x,y,t)$ is the data amount at $\boldsymbol{x} = (x,y)$. Then, the WEM algorithm with a polynomial time complexity will be used to find the optimal parameters of GMM. The convergence of the WEM iterative approach can be evaluated by the log likelihood function as [14]

$$\ln \mathcal{L}(\boldsymbol{\Sigma}, \boldsymbol{\mu}, \pi) = \sum_n \ln \sum_k \pi_k D(\boldsymbol{x}_n) g^t(\boldsymbol{x}_n|\boldsymbol{\Sigma}_k, \boldsymbol{\mu}_k), \tag{16}$$

whose value will increase as the iteration time increases. Our detailed approach is summarized in Algorithm 1.

Algorithm 1 Proposed algorithm to find parameters of the spatial distribution model $g^t(x,y)$

Input: $\boldsymbol{G}^t$ for a given t
Output: $\{\pi_k\}, \{\boldsymbol{\mu}_k\}$ and $\{\boldsymbol{\Sigma}_k\}$, for each $k \in \{1,\cdots,K\}$

Part I: Weighted K-means for parameter initialization
Input: $\boldsymbol{G}^t$
Output: K, $\{\boldsymbol{\mu}_k\}_{k\in\{1,2,\cdots,K\}}$
A. **For** $K = K_{\min} : K_{\max}$
1. Randomly choose K initial values of $\{\boldsymbol{\mu}_k\}_{k\in\{1,\cdots,K\}}$,
2. **Loop**:
a. Calculate the weighted distance of each data point to each $\boldsymbol{\mu}_k$ by $D(x,y,t)|\boldsymbol{x} - \boldsymbol{\mu}_k|$,
b. Assign each data point $\boldsymbol{x}$ to cluster k^*, such that $k* = \arg\min_k D(x,y,t)|\boldsymbol{x} - \boldsymbol{\mu}_k|$,
c. Recalculate $\boldsymbol{\mu}_k$ by averaging the values of data points belonging to cluster k as $\boldsymbol{\mu}_k = \sum_{\mathcal{C}_k} D(x,y,t)\boldsymbol{x} / \sum_{\mathcal{C}_k} D(x,y,t)$,
d. Check convergence: if $\{\boldsymbol{\mu}_k\}_{k\in\{1,2,\cdots,K\}}$ changes.
B. Choose the value of K that minimizes the ratio of intra-cluster to inter-cluster distance [15].

Part II: Weighted EM iteration
Input: $\boldsymbol{G}^t$, K, $\{\boldsymbol{\mu}_k\}_{k\in\{1,2,\cdots,K\}}$
Output: $\{\pi_k\}, \{\boldsymbol{\mu}_k\}$ and $\{\boldsymbol{\Sigma}_k\}$, for each $k \in \{1,\cdots,K\}$
1. **Initialize** $\boldsymbol{\Sigma}_k$ to be an identical matrix, and $\pi_k = 1/K$.
2. **E step**: Calculate the posterior probability for each data point $\boldsymbol{x}_n$ belonging to each cluster k by $r_{nk} = \pi_k \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) / \sum_{i=1}^{K} \pi_i \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$,
3. **M step**: Recalculate the parameters using the posterior probability r_{nk}
a. $\boldsymbol{\mu}_k = \sum_{n=1}^{N} D(x,y,t) r_{nk} \boldsymbol{x}_n / N_k$
b. $\boldsymbol{\Sigma}_k = \sum_{n=1}^{N} D(x,y,t) r_{nk} (\boldsymbol{x}_n - \boldsymbol{\mu}_k)(\boldsymbol{x}_n - \boldsymbol{\mu}_k)^T / N_k$
c. $\pi_k = N_k / K$
where $N_k = \sum_{n=1}^{N} D(x,y,t) r_{nk}$.
4. Check the **convergence** by (16). If not converged, return to E step.

2) Temporal distribution model: Given a time t, the total aerial traffic amount in the system from t to $t+1$ can be calculated by $D^t = \sum_{(x,y)\in\mathcal{A}} D(x,y,t)$. By gathering the data D^t of M days, we have a dataset $\boldsymbol{D}^t = \{D_1^t, \cdots, D_M^t\}$. The GMM that captures the temporal pattern of $\boldsymbol{D}^t$ is $p(D^t) = \sum_{v=1}^{V^t} \pi_v^t \mathcal{N}(D^t|\mu_v^t, \Sigma_v^t)$. The approach to model the temporal distribution D^t for $D(x,y,t)$, is similar to the algorithm in Algorithm 1. However, both the K-means and EM algorithm do not add weight to each data point. As a result, by ignoring all $D(x,y,t)$ used in Table 1 and substituting its value by one, Algorithm 1 can be applied to find the temporal pattern D^t. The mixture Gaussian model $p(D^t)$ is a probability density function (pdf) over the cellular data amount, from which we can get the cumulative distribution function (CDF) as $C_t(d) = \int_{-\infty}^{d} p(D^t)\,\mathrm{d}D^t$. The predicted data amount can be estimated by the CDF with a threshold. For example, with a threshold of 60%, the predicted traffic amount over the aerial networks can be given by $D^t = C_t^{-1}(0.6)$. The ML analysis of the temporal feature $N(t)$ and the spatial feature $f(x,y)$ of $N(x,y,t)$ can follow the approach of Algorithm 1.

B. On-demand, optimal UAV deployment

In order to optimally deploy UAVs to minimize the total power, problem (13) is formulated, which jointly considers the aerial cell partition and the UAVs' locations. With the prediction information, network operators only need to move UAVs at the beginning of each time interval, according to the solution of (13). However, solving (13) is challenging due to the mutual dependence between (x_i, y_i, h_i) and $\mathcal{A}_i$ with N_i and $\beta(x, y)$. For tractability, we solve (13) in two sequential steps. First, given the current location of each UAV $i \in \mathcal{I}$, we seek to find the optimal partition of the service area $\mathcal{A}_i$ for each UAV, that minimizes the power for transmissions. Then, for each UAV i, given its fixed service area $\mathcal{A}_i$, the optimal location is derived to minimize the required power for downlink communications and mobility.

1) Optimal partition of service areas: Given the current location of each UAV $i \in \mathcal{I}$, we aim to find the best partition of service areas $\{\mathcal{A}_i\}_{i\in\mathcal{I}}$, such that the total power for downlink communications of all UAVs is minimized. The optimal partition problem can be formulated as follows,

$$\min_{\mathcal{A}_i} \quad P_c, \tag{17a}$$

$$\text{s. t.} \quad \int\int_{\mathcal{A}_i} P_i^{\min}(x, y)\, \mathrm{d}x\, \mathrm{d}y = \kappa P_i^{\mathrm{a}}, \forall i \in \mathcal{I}, \tag{17b}$$

$$\cup_{i\in\mathcal{I}} \mathcal{A}_i = \mathcal{A}, \tag{17c}$$

$$\mathcal{A}_i \cap \mathcal{A}_j = \emptyset, \forall i \neq j \in \mathcal{I}. \tag{17d}$$

To solve this problem, we use our previously developed gradient-based method in [16, Theorem 1, Algorithm 1].

2) Optimal locations: Given the optimal partition of the service area $\{\mathcal{A}_i\}_{i\in\mathcal{I}}$, the power minimization problem can be reduced into I subproblems for each UAV $i \in \mathcal{I}$ as

$$\min_{x_i, y_i, h_i} \quad P_i^c + P_i^t. \tag{18}$$

Based on [3, Theorem 1], we focus on two scenarios in the following discussions. One is a high-altitude UAV, where $h_i^2 \gg (x - x_i)^2 + (y - y_i)^2$, and the other is the low-altitude UAV, where $h_i^2 \ll (x - x_i)^2 + (y - y_i)^2$. In scenario one, the value of θ_i in (2) is approximately $90°$, thus, $p_i^{\mathrm{LOS}}(x, y) \approx 1$ and $\bar{L}_i(x, y) \approx L_i^{\mathrm{LOS}}(x, y)$. Then, P_i can be rewritten as

$$P_i^c \approx O_i \int\int_{\mathcal{A}_i} Z_i(x, y) \left(d_i^2(x, y) + 10^{0.1\xi_i^{\mathrm{LOS}}}\right) \mathrm{d}x\, \mathrm{d}y, \tag{19}$$

where $O_i = \left(\frac{4\pi f_c}{c}\right)^2 \frac{B_i n_0}{G N_i}$ is a coefficient that does not depend on (x, y), and $Z_i(x, y) = 2^{\beta(x,y)N_i/B_i} - 1$. It is obvious that P_i is a convex function with respect to x_i and y_i. By setting the first partial derivatives to be zero, we have the optimal locations for UAV i that minimize the transmission power P_i^c as

$$x_i^* = \frac{\int\int_{\mathcal{A}_i} x\ Z_i(x, y)\ \mathrm{d}x\, \mathrm{d}y}{\int\int_{\mathcal{A}_i} Z_i(x, y)\ \mathrm{d}x\, \mathrm{d}y}, \tag{20a}$$

$$y_i^* = \frac{\int\int_{\mathcal{A}_i} y\ Z_i(x, y)\ \mathrm{d}x\, \mathrm{d}y}{\int\int_{\mathcal{A}_i} Z_i(x, y)\ \mathrm{d}x\, \mathrm{d}y}. \tag{20b}$$

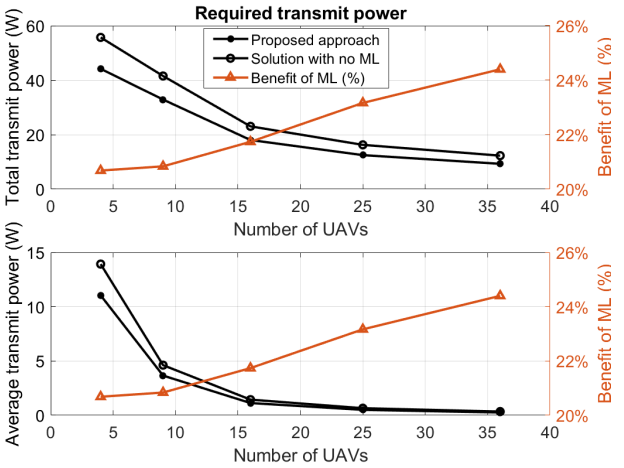

Fig. 1: Total and average required power for data transmission.

Although the objective function $P_i^c + P_i^t$ is convex with respect to x_i and y_i, deriving a closed-form solution of (18), which minimizes both the transmit and mobility power for each UAV, is challenging. However, it is easy to find the optimal solution of (18) based on a gradient-based algorithm. Using a similar approach, we can find the optimal location for scenario two.

IV. Simulation Results and Analysis

For simulations, we consider a UAV cellular network operating in a 5 GHz frequency band for downlink communications. The total available bandwidth for each UAV is 10 MHz. The noise power spectral is set to -174 dBm/Hz. For each UAV, the antenna gain is 10 dB, and the rate of energy consumption for moving per meter is $\gamma = 0.1$ Joules per meter. For ML, we use $\frac{7}{8}$ of the dataset to train the model, and the remaining $\frac{1}{8}$ data is used to evaluate the performance.

Fig. 1 shows the total and average communication power per UAV required to satisfy the users' data demands for two scenarios: the proposed approach and a solution with no ML predictions. In each case, the proposed optimal partition of service areas and the optimal location deployment are employed. Fig. 1 shows that, as the number of UAVs increases, both the total required power and the average communication power will decrease. When more UAVs are available, each aerial BS can serve a smaller coverage area, yielding a lower average path loss. Therefore, the needed total transmit power decreases, given a fixed amount of the total cellular traffic. As the total required transmit power decreases, the average power reduces accordingly. Fig. 1 further shows that compared with the solution without ML predictions, the proposed approach yields a significant improvement of power consumptions. The power reduction varies from 20.68% to 24.40%, as the number of UAVs increases from 9 to 36.

Fig. 2 shows the total and average mobility power of UAV BSs. As the number of UAVs increases, the total mobility power without ML predictions will increase significantly. By deploying more UAVs, network operator is more likely to send a UAV to meet an instant communication in a relatively far hotspot area, which causes more power consumed for

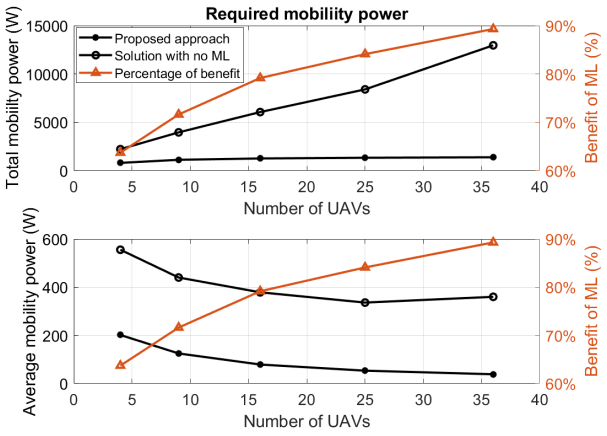

Fig. 2: Total and average required power for UAV mobility.

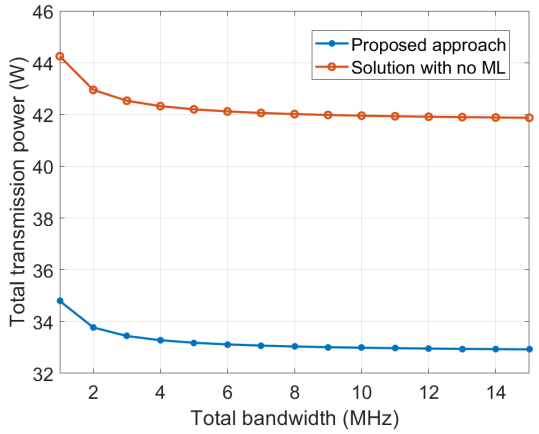

Fig. 3: Required transmit power as a function of the total bandwidth.

mobility. However, with ML prediction, UAV BSs do not need to travel constantly to satisfy the communication request. Therefore, the total mobility power of the proposed ML approach remains on a similar level. Also, as shown in Fig. 2, the average mobility power of each UAV will decrease, when the number of UAVs increases. This power decrease is caused by a smaller service area of each UAV. Fig. 2 further shows that compared with the solution without ML prediction, the proposed approach significantly reduces the required mobility power from 63.72% to 89.30%, as the number of UAVs increase from 9 to 36.

Fig. 3 shows the required transmit power as a function of the total bandwidth, assuming nine UAVs. As the available bandwidth increases, the transmit power will decrease. However, a wider bandwidth results in a higher noise power, which prevents the reduction of transmit power, especially when the bandwidth is greater than 5 MHz. For such noise-sensitive system, a lower spectrum efficiency cannot save additional power.

V. Conclusion

In this paper, we have proposed a novel approach for predictive deployment of UAV aerial BSs to provide an on-demand wireless service to the cellular users. We have formulated a power minimization problem to optimize the partition of the service area of each UAV, while minimizing the UAV power needed for downlink communications and mobility. In order to predict hotspots, a novel ML framework based on GMM and WEM has been developed. The results have shown that the proposed ML approach can reduce the required downlink transmit power and mobility power by over 20% and 80%, respectively, compared with an optimal deployment of UAVs with no ML prediction.

Acknowledgment

This research was supported by the U.S. National Science Foundation under Grant IIS-1633363.

References

[1] S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter-wave cellular wireless networks: Potentials and challenges," *Proceedings of the IEEE*, vol. 102, no. 3, pp. 366–385, Feb 2014.

[2] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *arXiv preprint arXiv:1803.00680*, 2018.

[3] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Optimal transport theory for power-efficient deployment of unmanned aerial vehicles," in *Proc. of IEEE International Conference on Communications (ICC)*, Kuala Lumpur, Malaysia, May 2016, pp. 1–6.

[4] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Machine learning for wireless networks with artificial intelligence: A tutorial on neural networks," *arXiv preprint arXiv:1710.02913*, 2017.

[5] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, "Efficient deployment of multiple unmanned aerial vehicles for optimal wireless coverage," *IEEE Communications Letters*, vol. 20, no. 8, pp. 1647–1650, Jun 2016.

[6] V. Sharma, M. Bennis, and R. Kumar, "UAV-assisted heterogeneous networks for capacity enhancement," *IEEE Communications Letters*, vol. 20, no. 6, pp. 1207–1210, 2016.

[7] Y. Zeng and R. Zhang, "Energy-efficient UAV communication with trajectory optimization," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3747–3760, Mar 2017.

[8] J. F. Horn, E. M. Schmidt, B. R. Geiger, and M. P. DeAngelo, "Neural network-based trajectory optimization for unmanned aerial vehicles," *Journal of Guidance, Control, and Dynamics*, vol. 35, no. 2, pp. 548–562, Mar-Apr 2012.

[9] J. Chen, U. Yatnalli, and D. Gesbert, "Learning radio maps for UAV-aided wireless networks: A segmented regression approach," in *Proc. of IEEE International Conference on Communications (ICC)*, Paris, France, July 2017, pp. 1–6.

[10] S. Zhang, S. Zhao, M. Yuan, J. Zeng, J. Yao, M. R. Lyu, and I. King, "Traffic prediction based power saving in cellular networks: A machine learning method," in *Proc. of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, Redondo Beach, CA, USA, Nov 2017, p. 29.

[11] I. D. Gebru, X. Alameda-Pineda, F. Forbes, and R. Horaud, "EM algorithms for weighted-data clustering with application to audio-visual scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 12, pp. 2402–2415, Jan 2016.

[12] 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, July 2014.

[13] "City cellular traffic map," https://github.com/caesar0301/city-cellular-traffic-map, accessed: 2016-10-05.

[14] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, Shanghai, China, Apr 2011, pp. 882–890.

[15] M. B. Christopher, *Pattern recognition and machine learning*. Springer-Verlag New York, 2016.

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