Contents

List of Figures

List of Tables

1 Introduction

The landscape of telecommunications is in a state of constant evolution, shaped by the rapid advancements in technology and the ever-growing demand for connectivity. In recent years, this evolution has accelerated exponentially, driven by the widespread adoption of smartphones, tablets, and a plethora of cellular applications. This surge in consumer demand has sparked an unprecedented explosion in data traffic, fundamentally reshaping the way we communicate, work, and live.

At the heart of this revolution lies the proliferation of 4G wireless connections [1], which have revolutionized the accessibility and availability of data services. With the promise of faster speeds, lower latency, and enhanced reliability, 4G networks have become the backbone of modern telecommunications infrastructure. This resulted in a digital ecosystem where data flows freely, empowering individuals and businesses alike to connect, collaborate, and innovate like never before.

The impact of this transformation is evident in the staggering statistics: studies show that subscribers now consume 1.7 times more data on 4G networks than on their 3G counterparts, and a staggering 11 times more than on 2G networks [1]. Moreover, global data traffic witnessed a remarkable 34% increase in 2020 compared to the previous year, with projections indicating a further surge in the years to come. This exponential growth in data consumption underscores the critical role that telecommunications plays in our increasingly interconnected world [2].

As we look ahead to the dawn of the fifth generation (5G) of wireless technology, the opportunities and challenges before us are unparalleled. With the proliferation of Internet of Things (IoT) devices and the emergence of new use cases such as autonomous vehicles and smart cities, the demand for connectivity is poised to skyrocket [3]. However, with this surge in demand comes a host of challenges, from network congestion to spectrum scarcity, that threaten to undermine the promise of 5G.

To address these challenges, there is an urgent need for innovative solutions that can augment existing infrastructure and unlock the full potential of 5G technology. One such solution lies in the integration of unmanned aerial vehicles (UAVs) into cellular networks. The rapid advancements in UAV technology have unlocked a world of possibilities, from aerial photography and surveillance to disaster response and infrastructure inspection [1].

But perhaps the most promising application of UAV technology lies in its potential to revolutionize wireless communication. By deploying UAVs as flying base stations, we can extend network coverage to remote and underserved areas, enhance signal strength in urban environments, and improve network reliability in times of crisis. Moreover, UAVs can serve as agile, flexible platforms for deploying small cells and other network infrastructure, enabling operators to quickly adapt to changing demand and optimize network performance. However, the integration of UAVs into cellular networks is not without its challenges. From regulatory hurdles to technical limitations, there are numerous obstacles that must be overcome to realize the full potential of this transformative technology [1], [3], [4].

This project seeks to address these challenges head-on by exploring the feasibility and efficiency of UAV-assisted cellular networks comprehensively. Through a combination of theoretical analysis, computer simulations, and real-world experiments, we aim to address the abilities of UAVs to act as base stations.

2 Related Work

The first study [5] examines resource allocation strategies for 5G-UAV-based emergency wireless communications, emphasizing the importance of efficient resource allocation for reliable and timely communication in emergency scenarios. The second study [6] explores the energy harvesting and data transmission capabilities of SWIPT-enabled cellular-connected UAVs, highlighting their potential to harvest energy from the environment and enable sustainable UAV operations. Lastly, the third study [7] investigates UAV-aided communication schemes, specifically focusing on the benefits of full-duplex non-orthogonal multiple access (FD-NOMA) for energy-efficient, reliable, and low-latency communications. All these researches mainly focus on fronthaul connectivity of the UAV and conventional base stations but for reliable and efficient communication the backhaul and placement of the UAV has the same priority as fronthaul connection. As a result of the above studies and observations, The focus of our paper is expanded to include backhaul and fronthaul connectivity through optimal placement of UAVs. The same issue has been addressed in different studies for example the study [8] focuses on a process where a UAV-BS system is introduced to an environment where they use a high-capacity wireless backhaul that uses millimeter-wave(mm-Wave) frequency bands. This increases the maximum throughput achievable as well as increases the rate of data delivery.

In another study [9], the authors explore the possibility of using tethered unmanned aerial vehicles (UAVs) in networks that resemble Terragraphs. In order to do this, they have used a unique deep reinforcement learning (DRL) framework that optimizes the number of Conventional BSs (CBSs) and UAVs needed, the multi-hop backhauling topology, and the hovering positions of deployed UAVs to minimize the overall deployment cost. The proposed framework is formulated based on the max-Bellman optimality equation in order to maximize the maximum reward. But the down of side above studies is they make use of tethered UAVs which restricts the range and height of the UAV. To utilize the full potential of UAVs it should be untethered, but this opens up many new challenges like energy optimization, placement of the UAV, and association of the UAVs to Conventional Base Stations(CBS). The studies

[10], [11] address the issue of the association of the UAVs to CBS but assume a fixed location of the UAVs instead of evaluating for different locations.

For the following study, we have utilised different data and approaches to find the optimal locations of UAVs and then associate the UAVs with the CBS.

3 Data and Methods

3.1 Data

The project requires the locations of CBS to generate the optimal locations of UAVs which are further referred to as controller-UAVs, for the study we have made use of two different CBS data.

i) The first dataset includes the base stations geolocation data that are situated in New York City. The Federal Communications Commission (FCC) Antenna Structure official website is the source of this dataset. It offers a wide range of diverse antennas made especially for using with radiofrequency (RF) signals, either sending or receiving. We have acquired 85 different geolocation datasets for New York City, as shown in Fig-1. Additionally, this collection is known as NYC Data.

ii) The second CBS distribution is generated using a statistical approach called a Matern type-I hard-core process [12]. This process ensures that CBSs are positioned at appropriate distances from each other to avoid interference. The approach makes use of different parameters which include the Geographical Area in which the network will be deployed R, d is the desired density of CBSs per unit area, and D_{min} is the minimum distance between CBSs. The Matern process outputs the coordinates (v_i, w_i) representing the locations of CBSs, satisfying the specified density and separation distance criteria. The total number of CBS generated is based on the condition presented in Equation (1).

$$T^{avg} = d \cdot exp(\text{-}d\pi {D^2}_{min}) \cdot R. \tag{1}$$

For better comparison and analysis of different data, the number of CBS in the NYC Data is restricted according to the Area of the network and Minimum Distance

Figure 1. Base Stations geolocations.

conditions mentioned in the above process.

3.2 Methodology

The project assumes a 5G heterogeneous network which includes a Ground Network, CBS, User Equipment, and UAVs. The architecture involves 3 layers in the network, the first layer includes the distribution of CBS and Ground Network, the second layer includes a swarm of UAVs called controller-UAVs which are located at a height varying in the range 300 to 800 meters from ground level, these controller-UAVs associated with the CBS, and for the final layer a UAV called master-UAV has deployed much higher altitude than the controller-UAVs has capabilities of both fronthaul and backhaul connections, the controller-UAVs are connected to the master-UAV which in turn has a backhaul connection with the Ground Network.

To maximize the association process the controller-UAVs should be deployed in optimal locations. To identify the optimal locations we compare 3 different approaches:

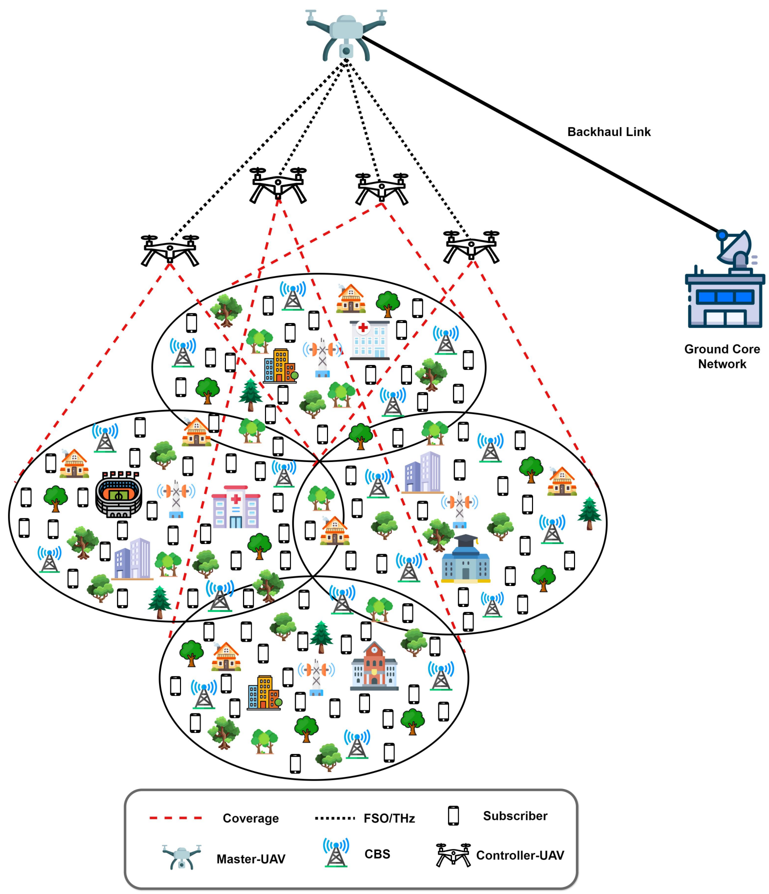

Figure 2. Illustration depicting a UAV-enabled fronthaul network with consideration for backhaul.

- K-Means,
- K-Mediods,
- Genetic Algorithm(GA).

The association between CBS and controller-UAVs includes different parameters and conditions. For path loss between CBS and controller-UAVs, we have considered the commonly known model Air-to-Ground which is presented in [13] and [14]. The equation (2) represents the path loss model.

$$PL = F_0 + P_{LoS} * PL_{LoS} + P_{NLoS} * PL_{NLoS}, \tag{2}$$

where $P_{LoS} = 1/(1 + a * exp(-b(\theta - a)))$, $P_{NLoS} = 1 - P_{LoS}$ in which a and b are environment constants which vary depending on the environment we want to simulate (Urban, Rural and Suburban).

$$F^0{}_{i,j}(dB) = \gamma \cdot \log_{10}\left(\frac{4\pi \cdot S_{i,j}}{\lambda_{\text{carrier}}}\right), \tag{3}$$

$$\lambda_{\text{carrier}} = \left(\frac{c}{f_{carrier}}\right), \tag{4}$$

where S is the distance between the CBS and UAVs.

$$p^r{}_{i,j}(dB) = 10 \cdot \log_{10}(\Omega_{\text{i,j}}) + N_{i,j} - I_{i,j}, \tag{5}$$

$$N_{i,j}(dB) = P_{i,j}^{LoS} \cdot Z^0 + P_{i,j}^{NLoS} \cdot Z^1. \tag{6}$$

We produce the path loss for each CBS and UAV pair. Using the Interference Threshold multiplied by path loss and divided by the square of the channel gain, we determine the ideal power for the UAV. The ideal power level ought to be lower than the maximum transmit power. Equation (6) is used to determine the fading using the ideal power and Nakagami distribution channel coefficients Z_0 and Z_1. Equation (5) is used to determine the power received once all of the values for each pair of CBS and controller-UAV have been determined. Equation (7) uses the power received to get the SINR value.

$$\eta_{\text{i,j}} = \frac{p_{i,j}^r}{\sigma_{\text{n}}{}^2 + I_s}. \tag{7}$$

where

- σ_n^2 - Noise Power for the particular Link,
- I - Interference from all the other controller-UAV for that particular CBS.

We have applied the previously computed channel gain to the interference.

3.3 Association of CBS and Controller-UAV

Theoretically, every UAV broadcasts a signal of its optimal power toward the CBS, at the SINR and is calculated and verified if it is greater than the Minimum SINR value (η_{min}). The association happens between controller-UAVs which provides the highest SINR value for a particular CBS. To present the same concept in the simulation we initialize an Association Matrix called A with the number of CBS M as columns and the Number of controller-UAVs N as the rows. The matrix is initialized with zeros, later using the above-generated data we identify a single pair of every CBS which provides the maximum SINR and also satisfies the Minimum SINR condition. We populated the pairs in the matrix with 1. To calculate the required bandwidth(b) for each CBS we make use of both data rates and SINR values generated from equation (8).

$$b_{i,j} = \frac{{r^d}_{i,j}}{\log_2(1 + \eta_{i,j})}, \tag{8}$$

where r_d - Data Rate required by the CBS.

The next step in the process is for the UAV to allocate the bandwidth accordingly but to have a fair generalization of the real world and also to have energy-efficient usage of UAV we introduce some conditions and constraints:

i) Each CBS can associate with only one controller-UAV.

ii) Each controller-UAV can only associate up to a maximum limited number of links(l_{max}).

iii) Each controller-UAV can allocate a certain Maximum Bandwidth (B), no controller-UAV cannot exceed the limit.

iv) The allocation of the bandwidth is done using the Spectral Efficiency value calculated through the equation, the higher the value of spectral efficiency the more priority to allocate the bandwidth.

$$\sum_{i=1}^{T} b_{i,j} \cdot a_{i,j} \leq B_j, \tag{9}$$

$$\sum_{i=1}^{T} a_{i,j} \leq l_j^{max}. \tag{10}$$

The above constraints are formulated using equations (9), and (10).

If any pair in the association matrix fails to follow the condition and constraints mentioned above the value is updated to 0.

The association process is summarized in the Algorithm 1.

Algorithm 1 Association between CBS and Controller-UAV

Input: $M, N, l_{\max}, b_{i,j}, r_{i,j}, \eta_{i,j}$
Output: $\mathbf{A}$
//Initialize matrix $\mathbf{A}$ with all zeroes
Initialize $\mathbf{A} = \phi$ ▷ **Step-1 at each CBS**
  for $i = 1$ **to** M **do**
    Select Controller-UAV from which an i^{th} CBS receives maximum SINR, η^{max}. Also, i^{th} CBS validates if it satisfies the constraint (21), Then CBS updates $a_{i,j} = 1$. Otherwise, i^{th} CBS updates $a_{i,j} = 0$. ▷ **Step-2 at each Controller**
    for $j = 1$ **to** N **do**
      Initialize counter: $X_l = 0$, $X_b = 0$
      while $X_l < l_{max} \wedge X_b < B$ **do**
        Find highest spectral efficient CBS
        if $X_b + b_{i,j} \leq B$ **then**
          Update $X_l = X_l + 1$ and $X_b = X_b + b_{i,j}$
        else
          Update $a_{i,j} = 0$

The count of '1's within the matrix may vary between Step -1 and Step 2 in Algorithm 1.

3.4 Backhaul Connectivity

The main objective of the paper was to introduce backhaul connectivity in UAV cellular communication. As presented before we fly a master-UAV above a swarm of UAVs which is connected to the Ground Network. The main objective of the master-UAV is to have a fair simulation that is near to a real-world scenario. We use equation (11) to get the matrix's sum rate (F_s) after updating Matrix A in step 2 of the algorithm.

$$F_s = \sum_{i=1}^{T} \sum_{j=1}^{U} {r^d}_{i,j} \cdot a_{i,j}. \tag{11}$$

We verify if the sum rate is less than the capacity of the Backhaul Connection(R_B). If the conditions fail we select the association pair with minimum demanding data rate and de-associate the pair, this is done by updating the index value in the matrix from 1 to 0 after the constraint is verified again and if it fails again we de-associate another pair for the controller-UAV which demands fewer data rate, this process iterates till the constraint is passed. The reason to de-associate the lower demanding data rate and lesser association is to introduce energy efficiency and maximize the overall data rate. The process is summarized in the Algorithm 2.

Algorithm 2 Task at Master-UAV

Initialize: F_s as the total sum rate of associated CBSs

while $F_s > R_B$ **do**

Select Controller-UAV with max associated CBSs

// This approach will introduce fairness for all regions Due to the reason of selection (may or may not) of Different Controller-UAV in each iteration.

Select CBS with minimum data rate, $min(r^d_{i,j})$, demand

De-associate the selected pair (i^{th} CBS of j^{th} Controller-UAV) and

Update $a_{i,j} = 0, X_l = X_l - 1, F_s = F_s - r^d_{i,j}$ and $X_b = X_b - b_{i,j}$

3.5 Optimal Locations of Controller-UAV

Finding optimal locations for the controller-UAV to maximise the sum rate was one of the study's other goals. As previously stated, we examine three methodologies for two distinct datasets in order to determine which is the most effective. We follow the described process for both CBS distributions even if the datasets are different.

3.5.1 K-Means

K-Means algorithm is used to maximize the objective function i.e., equation (11). The main aim of the K-Means is to split the M CBS locations into N clusters by minimizing the Euclidean distances within the cluster [10]. The output of the K-Means is the locations of the centroids for different clusters. These centroids are the optimal locations for controller-UAVs.

The number of clusters is defined by equation (12).

$$N = ceil(M/l_{max}), \tag{12}$$

where l_{max} is the maximum number of links per controller-UAV .

For both the dataset we initially ran the K-Means Clustering Algorithm to identify the optimal controller-UAV locations. After obtaining controller-UAV locations we run Algorithm-1 and Algorithm-2 separately for both distributions.

3.5.2 K-Medoids

The objective of K-Medoids is also to maximize the sum rate using the same clustering method but to create disjoint subsets of clusters and to minimize the within-cluster squared error. The splitting is done in the same way as K-Means. The centroids are generated after successfully running the algorithm. We run Algorithm-1 and Algorithm-2 to generate the output and different association matrices.

3.5.3 Genetic Algorithm(GA)

Genetic Algorithm which is inspired from natural genes, updates itself to satisfy the objective. So to maximize our objective function we use a Genetic Algorithm. There are 3 important involved in the genetic algorithm we are separately simulated for both datasets:

Step-1: Using both datasets we initially generate random locations for the controller-UAV .We generate 50 different locations with altitude included which is randomly updated in between a range of 300 to 800 m. Each location is called Chromosome Y. We run Algorithm 1 and Algorithm 2 for all the different Chromosomes. Later the chromosomes are arranged in descending order of Sum Rate and No. of 1s in the Final Matrix A.

Step-2: We select the first 40% of the chromosomes as Top Priority Parents and the remaining 60% are selected based on the Roulette Wheel Selection. The chromosome with a higher sum rate has a chance of being selected, then mutation operation is applied to the 60% of chromosomes. The mutation operation updates one or more locations of the UAV present in the current chromosome, this updation is done according to equation (13).

$$\boldsymbol{Y}_e{}^{mtd} = \boldsymbol{Y}_e \pm (M_{Rate} \times \boldsymbol{g}_{\text{value}}^{\text{mtd}}). \tag{13}$$

After updating, both the 40% and 60% are combined.

Step-3: The same steps 1 and step 2 are iterated multiple times and if for 5 iterations there is no change in Maximum Sum Rate the loop breaks out and provides the output chromosome number which is the optimal location of controller-UAVs and generates the sum rate. The no.of iterations is 50 but because hardware lacking we could run upto 10 iterations.

3.6 Simulation Parameters

Parameter	Description	Value
$\boldsymbol{f_{carrier}}$	Carrier Frequency	2GHz
$\boldsymbol{a,\ b}$	Environment Constants	9.61, 0.61
$\boldsymbol{PL_{LoS},\ PL_{NLoS}}$	Efficiency of LoS and NLoS Links	1, 20 (dB)
$\boldsymbol{d}$	Density of CBSs	$2 \times 10^{-6}/m^2$
$\boldsymbol{R}$	Total Area	16Km^2
$\boldsymbol{D_{Min}}$	Minimum Distance between CBSs	250m
$\boldsymbol{h_{min},\ h_{max}}$	Altitude Range of UAV	300m, 800m
$\boldsymbol{r^d}$	Data rate of CBS	20, 40, 40, 80, 100
$\boldsymbol{I_t}$	Interference Threshold	1.1943^{-14} Watt
${\sigma_n}^2$	Noise Power	-125 dB
$\boldsymbol{l_{max}}$	Maximum No. of Links	7
η_{min}	Minimum SINR	-10 dB
$\boldsymbol{R_B}$	Bachhaul Capacity	1.66Gbps
$\boldsymbol{M_{Rate}}$	Mutation Rate	9%
$\boldsymbol{g^{mtd}}$	Mutation Genes	4

Table 1: Simulation Parameters

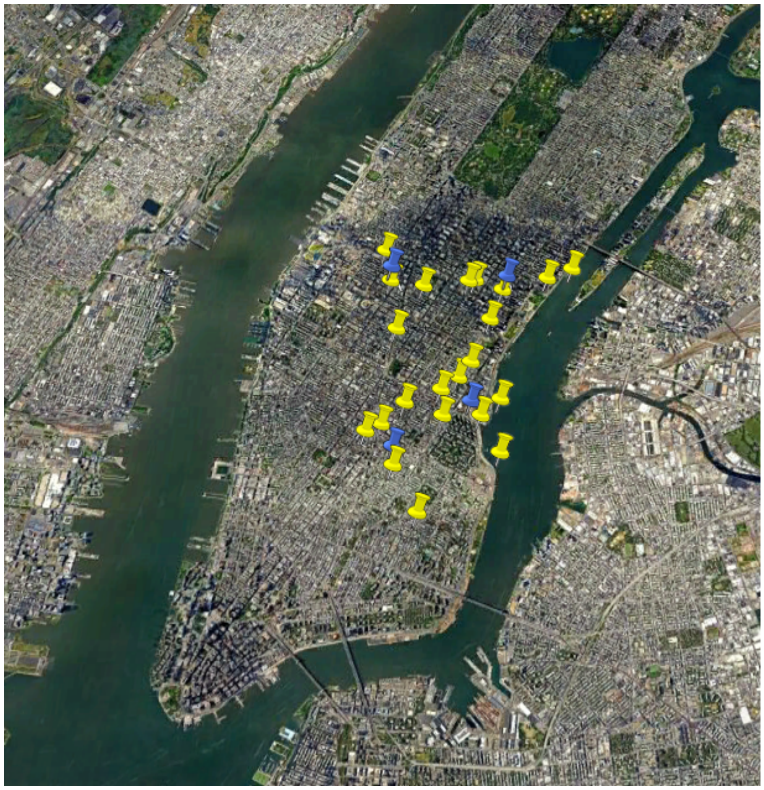

Figure 3. K-Means locations.

4 Results and Discussions

The depicted Figure-3, provides a visual representation of the results obtained from the application of the K-means algorithm. This algorithm was employed to determine the optimal locations for deploying controller-UAVs within a given geographic area. In this specific scenario, the dataset utilized originates from the New York base station dataset, which serves as a foundation for identifying potential sites for these controller-UAVs.

Within the figure, distinct visual markers are utilized to convey key information. The yellow pins symbolize the locations of existing conventional base stations (CBSs) extracted from the New York base station dataset. These CBSs represent established infrastructure within the area under consideration.

Within the figure, distinct visual markers are utilized to convey key information. The yellow pins symbolize the locations of existing conventional base stations (CBSs) extracted from the New York base station dataset. These CBSs represent established infrastructure within the area under consideration.

In contrast, the blue pins denote the optimal locations determined by the K-means algorithm for deploying controller-UAVs. These locations are strategically identified to maximize coverage, minimize interference, and optimize network performance. By leveraging the spatial distribution of existing CBSs and employing advanced algorithmic techniques, the K-means algorithm effectively identifies locations where deploying controller-UAVs can complement existing infrastructure and enhance overall network capabilities.

Through the integration of geographical data analysis and machine learning methodologies, the figure highlights the potential of leveraging UAV technology to augment and optimize existing telecommunications networks. By strategically positioning controller-UAVs based on algorithmic insights derived from data-driven analysis, operators can enhance network coverage, improve service quality, and meet the evolving demands of users in urban environments such as New York City.

Furthermore, the utilization of the K-means algorithm underscores the importance of data-driven decision-making in the design and deployment of UAV-assisted cellular networks. By leveraging large-scale datasets and advanced analytical techniques, operators can identify optimal deployment strategies that maximize network efficiency while minimizing deployment costs.

In summary, the figure serves as a visual representation of the synergistic relationship between existing terrestrial infrastructure and emerging UAV technology in the context of cellular network optimization. It illustrates how data-driven analysis and advanced algorithms can inform strategic decision-making to enhance network performance and meet the ever-growing demands of modern telecommunications users.

The depicted Fig-4 provides a visual representation of the outcomes derived from the application of the K-Medoids algorithm, specifically employed to ascertain the optimal placements for controller-UAVs within a designated geographical area. In

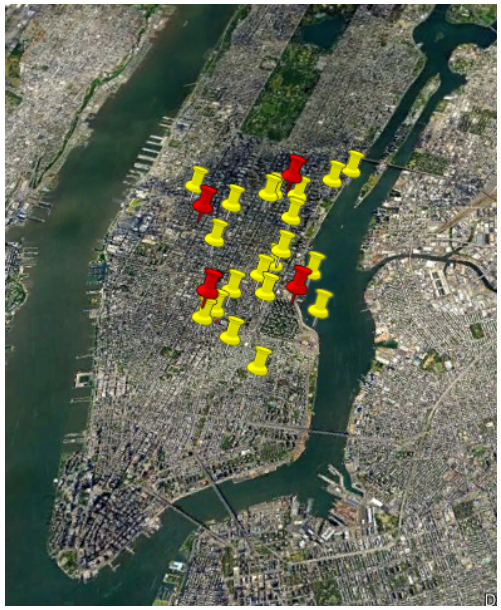

Figure 4. K-Medoids locations.

this particular context, the dataset utilized originates from the New York base station dataset, forming the basis for identifying potential sites for these controller-UAVs.

Distinct visual markers are employed within the figure to convey essential information. The yellow pins denote the locations of existing conventional base stations (CBSs) extracted from the New York base station dataset. These CBSs serve as established infrastructure within the geographical area under consideration.

Conversely, the red pins symbolize the optimal locations determined by the K-Medoids algorithm for deploying controller-UAVs. These locations are strategically identified to maximize coverage, minimize interference, and optimize network performance. By leveraging the spatial distribution of existing CBSs and employing advanced algorithmic techniques, the K-Medoids algorithm effectively identifies locations where deploying controller-UAVs can complement existing infrastructure and enhance overall network capabilities.

Through the integration of geographical data analysis and machine learning methodologies, the figure underscores the potential of utilizing UAV technology to augment and optimize existing telecommunications networks. By strategically

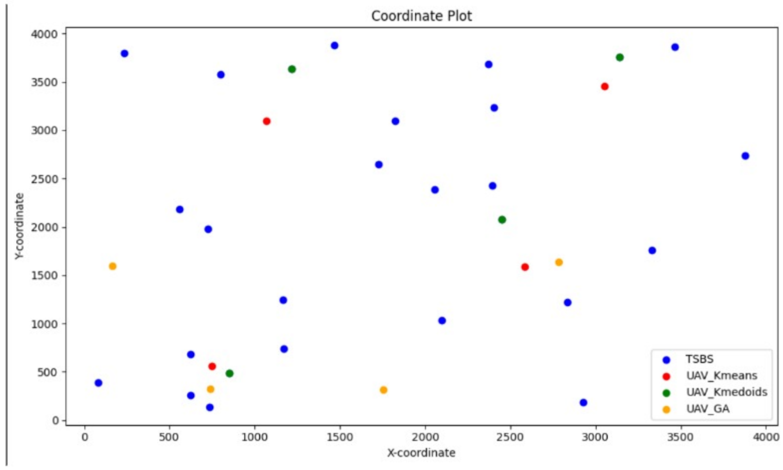

Figure 5. Matern Type-1 Hard Core Process.

positioning controller-UAVs based on algorithmic insights derived from data-driven analysis, operators can enhance network coverage, improve service quality, and address the evolving demands of users in urban environments like New York City.

Furthermore, the utilization of the K-Medoids algorithm highlights the significance of data-driven decision-making in the design and deployment of UAV-assisted cellular networks. By leveraging large-scale datasets and advanced analytical techniques, operators can identify optimal deployment strategies that maximize network efficiency while minimizing deployment costs.

In essence, the figure serves as a visual representation of the symbiotic relationship between existing terrestrial infrastructure and emerging UAV technology within the realm of cellular network optimization. It illustrates how data-driven analysis and advanced algorithms can inform strategic decision-making to enhance network performance and meet the evolving demands of modern telecommunications users.

In Figure-5, various algorithms have been applied to determine optimal placements for Controller UAVs within a designated geographical area. The figure incorporates locations generated by different algorithms, with each algorithm's outcomes represented by distinct visual markers:

1) Yellow Dots (Genetic Algorithm): These dots represent locations identified by the Genetic Algorithm (GA). GA employs principles of natural selection to iteratively optimize solutions. In this context, yellow dots signify potential deployment sites for Controller UAVs identified through the Genetic Algorithm.

2) Red Dots (K-Medoids): The red dots symbolize the locations determined by the K-Medoids algorithm. Unlike K-means, which identify cluster centroids, K-Medoids select actual data points within clusters as representatives. These red dots denote optimal locations for deploying Controller UAVs identified through the K-Medoids algorithm.

3) Green Dots (K-means): The green dots represent locations identified by the K-means algorithm. K-means partitions the geographical area into clusters based on proximity, aiming to optimize network performance by strategically positioning Controller UAVs. Green dots indicate potential deployment sites identified as cluster centroids by the K-means algorithm.

4) Blue Dots (Conventional Base Stations - CBS): These dots denote the locations of existing conventional base stations (CBS) within the geographical area. CBSs represent established infrastructure and serve as reference points for optimizing the deployment of Controller UAVs. Blue dots signify the spatial distribution of CBSs extracted from the New York base station dataset.

By incorporating locations generated by multiple algorithms, along with existing base stations, within a single image, Fig-5 offers a comprehensive overview of potential deployment sites for Controller UAVs and the distribution of existing infrastructure. Stakeholders can use this comparative representation to evaluate different algorithmic approaches' efficacy in optimizing cellular network deployment strategies and meeting users' evolving demands.

In Fig-6, a visualization of 3D coordinates of UAVs is presented, generated utilizing the locations derived from the Matérn process CBS distribution. The figure provides a spatial representation of the optimal placements for deploying UAVs within a designated area, leveraging the distribution of conventional base stations (CBS) as determined by the Matérn process.

Fig-7 illustrates the relationship between sum rate (F_s) and maximum bandwidth, specifically focusing on results from the K-means algorithm. The x-axis shows the maximum bandwidth, while the y-axis displays the sum rate. Key features:

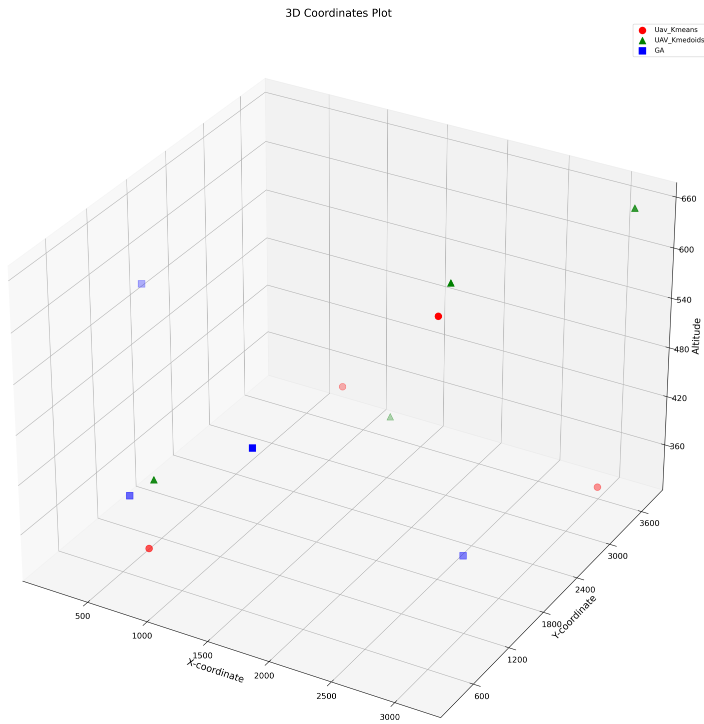

Figure 6. 3D coordinates of UAV generated using the locations of Matern process CBS Distribution

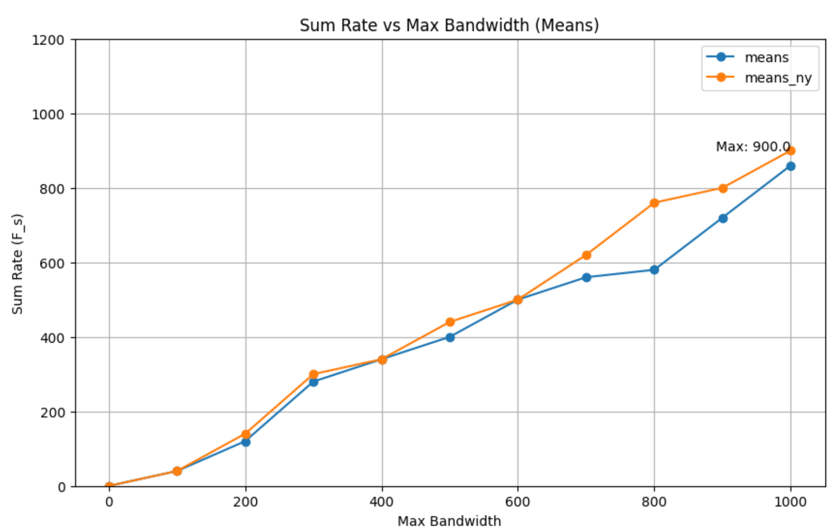

Figure 7. K-Means Sum rate vs Max Bandwidth

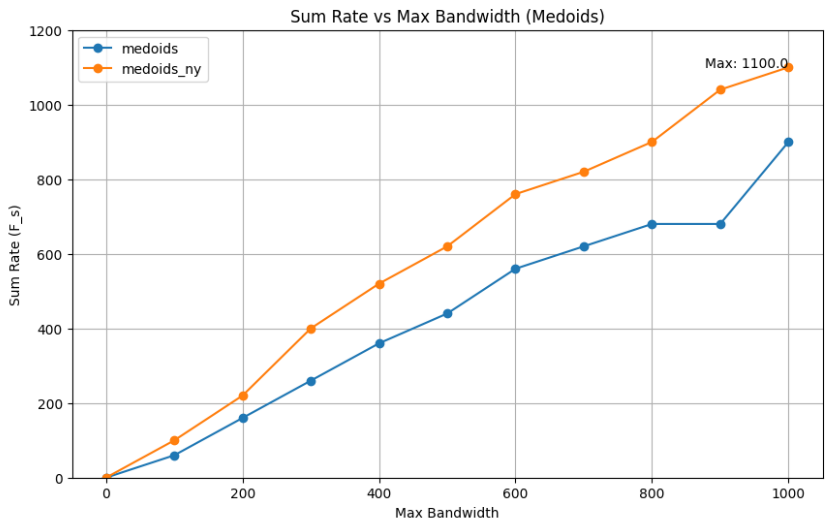

Figure 8. K-Medoids Sum rate vs Max Bandwidth

1) Means (Blue Line - K-means on Matern Hard Core Data): Average sum rate from K-means applied to Matern hardcore data, varying bandwidth.

2) Means-ny (Orange Line - K-means on New York Data): Average sum rate from K-means on New York Base Station Dataset, varying bandwidth.

This comparison aids in understanding how bandwidth variations impact the sum rate across different datasets.

Fig-8 illustrates the relationship between sum rate (F_s) and maximum bandwidth, focusing on outcomes generated by the K-Medoids algorithm. The x-axis represents maximum bandwidth, while the y-axis denotes the sum rate. Key features:

1) Medoids (Blue Line - K-Medoids on Matern Hard Core Data): Average sum rate from K-Medoids algorithm applied to Matern hardcore data, with varying bandwidth.

2) Medoids-ny (Orange Line - K-Medoids on New York Data): Average sum rate from K-Medoids algorithm on New York Base Station Dataset, with varying bandwidth.

This comparison allows for an analysis of how changes in maximum bandwidth affect the sum rate across different datasets when using the K-Medoids algorithm.

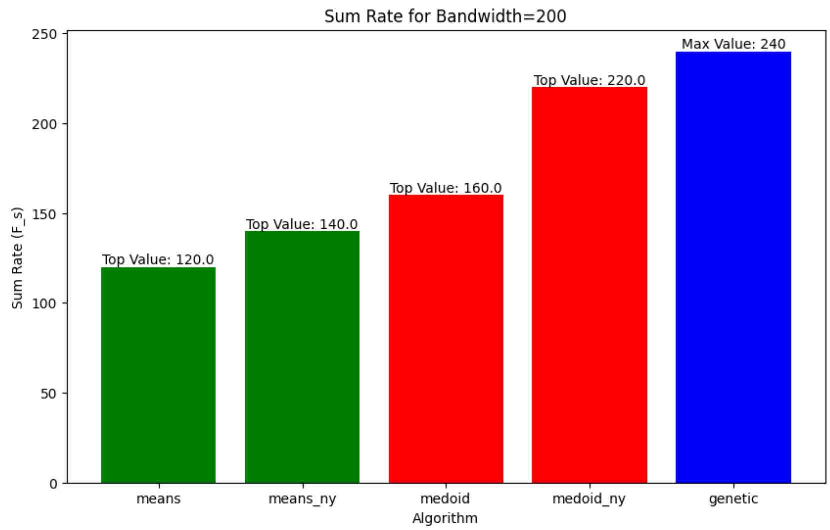

Figure 9. Maximum Sum Rate achieved

Fig-9 illustrates the comparison between the sum rate of different algorithms for a bandwidth of 200 and the genetic algorithm for 50 iterations.

Fig-10 illustrates the relationship between sum rate (F_s) and maximum bandwidth, focusing on outcomes generated by the Genetic Algorithm. The x-axis represents maximum bandwidth, while the y-axis denotes the sum rate. Key Features:

1) Genetic (Blue Line - Genetic Algorithm on Matern Hard Core Data): Average sum rate from the Genetic Algorithm applied to Matern hardcore data, with varying bandwidth.

2) Genetic-NY (Orange Line - Genetic Algorithm on New York Data): Average sum rate from the Genetic Algorithm on New York Base Station Dataset, with varying bandwidth.

This comparison allows for an analysis of how changes in maximum bandwidth affect the sum rate across different datasets when using the Genetic Algorithm.

Figure-11 shows the maximum number of connections that can be made for the controller UAV.

Figure-12 illustrates the relationship between sum rate (F_s) and maximum bandwidth, specifically focusing on results from applying the PPP(Poison Point

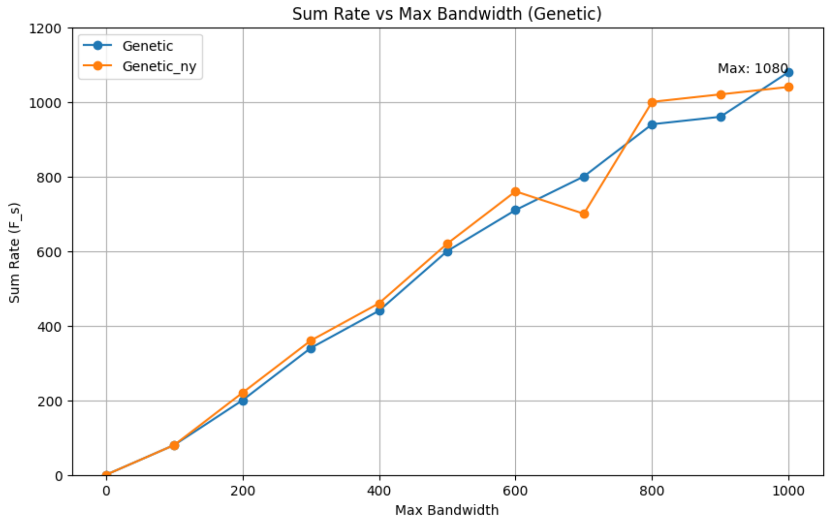

Figure 10. Genetic Algorithm Sum Rate vs Max Bandwidth

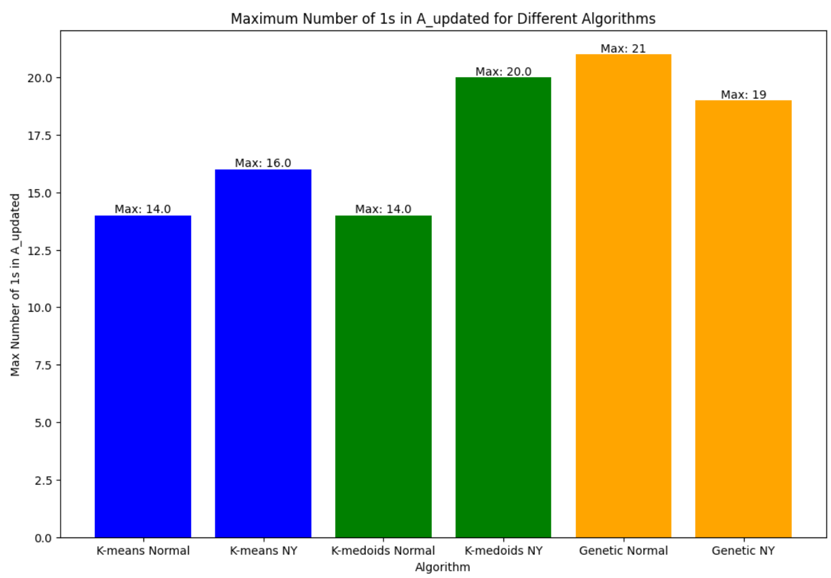

Figure 11. Maximum no. of 1s updated

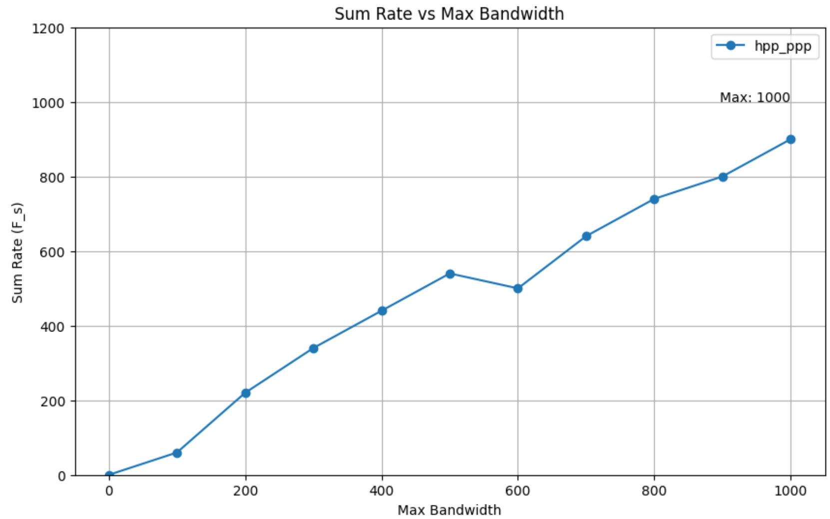

Figure 12. HPP Sum rate vs Max Bandwidth

Process) algorithm for the deployment of controller UAV and HPP(Matern Type-1 Hardcore Point Process) for master. The x-axis shows the maximum bandwidth, while the y-axis displays the sum rate.

The graph's dip can be attributed to the possibility that there exists an optimal position for the controller UAV at every bandwidth. We have adopted a standard procedure whereby key point processes are deployed to the same places regardless of bandwidth; therefore, there is a chance that these sites may not be ideal for a variety of bandwidth.

The relationship between cumulative rate (F_s) and maximum bandwidth is depicted in Figure 13, which focuses on the outcomes of employing the PPP for controller UAVs and the HPP method for the master UAV's deployment. The maximum bandwidth is displayed on the x-axis, while the total rate is displayed on the y-axis.

The maximum number of connections made by the controller UAV when the PPP method is applied is shown in Figure 14.

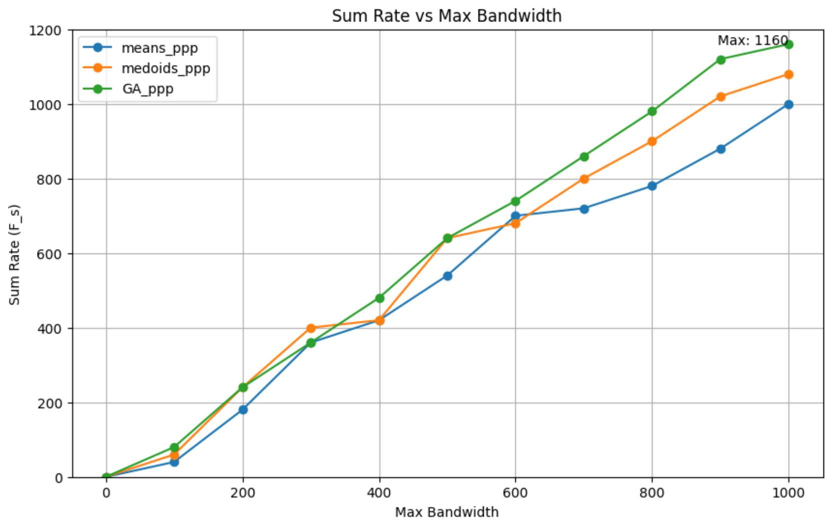

Figure 13. PPP models Sum rate vs Max Bandwidth

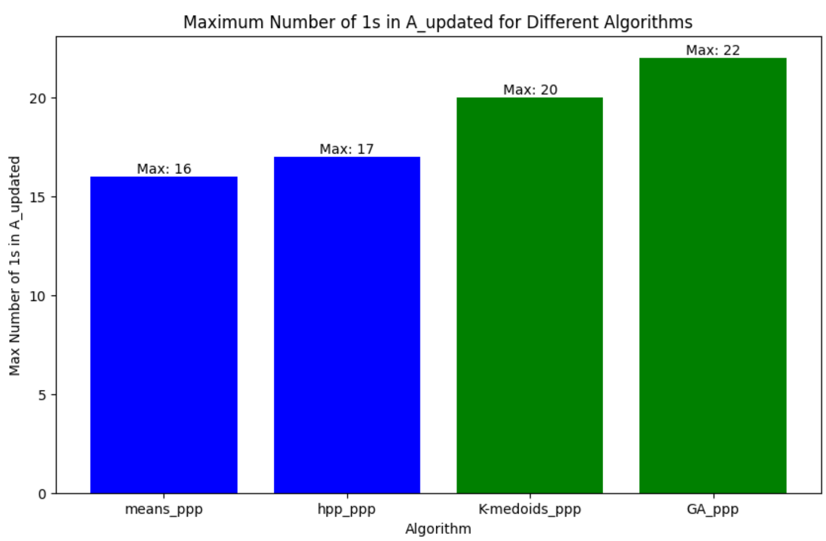

Figure 14. Maximum no. of 1s updated for PPP

Max Bandwidth	**Sum Rate (F_s) Means**	**Sum Rate (F_s) Medoids**	**Sum Rate (F_s) Genetic**
0	0	0	0
100	40	60	80
200	120	160	200
300	280	260	340
400	340	360	440
500	400	440	600
600	500	560	710

Table 2: Hardcore Point Process

Max Bandwidth	**Sum Rate (F_s) Means**	**Sum Rate (F_s) Medoids**	**Sum Rate (F_s) Genetic**
0	0	0	0
100	40	60	80
200	140	160	220
300	300	260	360
400	340	360	460
500	440	440	620
600	500	560	760

Table 3: New York city dataset

Max Bandwidth	**Sum Rate (F_s) Means**	**Sum Rate (F_s) Medoids**	**Sum Rate (F_s) Genetic**
0	0	0	0
100	40	60	80
200	180	240	240
300	360	400	360
400	420	420	480
500	540	640	640
600	700	680	740

Table 4: Poison Point Process

5 Conclusion

The main motive of the project is to establish the backhaul connectivity for UAV deployed,we have identified that the Genetic Algorithm in general for any kind of data produces better results compared to K-Medoids and K-Means algorithm which involves sum rate, No.of associations but if we look into time consumed, the computation required Genetic Algorithm take more time to generate the optimal locations and also requires more computation compared to K-Medoids which generates results that are close to Genetic Algorithm with less time and computation.For any real-life application K-Medoids can employed for distribution of controller-UAVs.The future aspect can be using reinforcement learning for the tethered drones.For backhaul connectivity, we can also use mm-Wave configuration.

References

[1] Amponis, G., T. Lagkas, M. Zevgara, G. Katsikas, T. Xirofotos, I. Moscholios, and P. Sarigiannidis. "Drones in B5G/6G Networks as Flying Base Stations. Drones 2022, 6, 39." (2022).

[2] Yaacoub, Elias, and Mohamed-Slim Alouini. "A key 6G challenge and opportunity—Connecting the base of the pyramid: A survey on rural connectivity." Proceedings of the IEEE 108, no. 4 (2020): 533-582.

[3] Geraci, Giovanni, Adrian Garcia-Rodriguez, M. Mahdi Azari, Angel Lozano, Marco Mezzavilla, Symeon Chatzinotas, Yun Chen, Sundeep Rangan, and Marco Di Renzo. "What will the future of UAV cellular communications be? A flight from 5G to 6G." IEEE communications surveys tutorials 24, no. 3 (2022): 1304-1335.

[4] Alfaia, Rodrigo Dias, Anderson Vinicius de Freitas Souto, Evelin Helena Silva Cardoso, Jasmine Priscyla Leite de Araújo, and Carlos Renato Lisboa Francês. "Resource management in 5G networks assisted by UAV base stations: Machine learning for overloaded Macrocell prediction based on users' temporal and spatial flow." Drones 6, no. 6 (2022): 145.

[5] Katwe, Mayur, Rasika Deshpande, Keshav Singh, and Chih-Peng Li. "UAV aided FD-NOMA for Energy-efficient Ultra-Reliable Low-latency Communications." In MILCOM 2022-2022 IEEE Military Communications Conference (MILCOM), pp. 31-36. IEEE, 2022.

[6] Li, Xinran, Xiaohui Zhou, and Salman Durrani. "SWIPT-Enabled Cellular-Connected UAV: Energy Harvesting and Data Transmission." In 2022 IEEE International Conference on Communications Workshops (ICC Workshops), pp. 1107-1112. IEEE, 2022.

[7] Yao, Zhuohui, Wenchi Cheng, Wei Zhang, and Hailin Zhang. "Resource allocation for 5G-UAV-based emergency wireless communications." IEEE Journal on Selected Areas in Communications 39, no. 11 (2021): 3395-3410.

[8] Moon, Jangwon, Junwoo Kim, Hoon Lee, Youngjin Moon, Yongsu Lee, Youngjo Bang, Kyungyeol Sohn et al. ”Implementation of mmWave long-range backhaul for UAV-BS.” ETRI Journal 45, no. 5 (2023): 781-794.

[9] Zhang, Yongqiang, Mustafa A. Kishk, and Mohamed-Slim Alouini. ”Deployment optimization of tethered drone-assisted integrated access and backhaul networks.” IEEE Transactions on Wireless Communications (2023).

[10] M. K. Shehzad, S. A. Hassan, A. Mahmood, and M. Gidlund, “On the association of small cell base stations with UAVs using unsupervised learning,” in Proc. IEEE 89th Veh. Technol. Conf., 2019, pp. 1–5.

[11] M. K. Shehzad, S. A. Hassan, M. Luque-Nieto, J. Poncela, and H. Jung, “Energy efficient placement of UAVs in wireless backhaul networks,” in Proc. 2nd ACM MobiCom Workshop Drone Assist. Wireless Commun. 5G Beyond, 2020, pp. 1–6.

[12] B. Matern, Spatial variation, vol. 36, Springer Lecture Notes in Statistics. Berlin, Germany: Springer, 1986.

[13] Zhang, Hongliang, Lingyang Song, and Zhu Han. Unmanned aerial vehicle applications over cellular networks for 5G and beyond. Cham, Switzerland: Springer International Publishing, 2020.

[14] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, “Modeling air-to ground path loss for low altitude platforms in urban environments,” in Proc. IEEE Glob. Commun. Conf., 2014, pp. 2898–2904.