sensors

MDPI

Article

Sensor Clustering Using a *K*-Means Algorithm in Combination with Optimized Unmanned Aerial Vehicle Trajectory in Wireless Sensor Networks

Thanh-Nam Tran [1,†], Thanh-Long Nguyen [2,*,†], Vinh Truong Hoang [3] and Miroslav Voznak [4]

[1] Data Science Laboratory, Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City 700000, Vietnam
[2] Faculty of Information Technology, Ho Chi Minh City University of Food Industry, Ho Chi Minh City 700000, Vietnam
[3] Faculty of Computer Science, Ho Chi Minh City Open University, Ho Chi Minh City 700000, Vietnam
[4] Faculty of Electrical Engineering and Computer Science, VSB-Technical University of Ostrava, 17. listopadu 2172/15, 708 00 Ostrava, Czech Republic
* Correspondence: longnt@hufi.edu.vn
† These authors contributed equally to this work.

Abstract: We examine a general wireless sensor network (WSN) model which incorporates a large number of sensors distributed over a large and complex geographical area. The study proposes solutions for a flexible deployment, low cost and high reliability in a wireless sensor network. To achieve these aims, we propose the application of an unmanned aerial vehicle (UAV) as a flying relay to receive and forward signals that employ nonorthogonal multiple access (NOMA) for a high spectral sharing efficiency. To obtain an optimal number of subclusters and optimal UAV positioning, we apply a sensor clustering method based on *K*-means unsupervised machine learning in combination with the gap statistic method. The study proposes an algorithm to optimize the trajectory of the UAV, i.e., the centroid-to-next-nearest-centroid (CNNC) path. Because a subcluster containing multiple sensors produces cochannel interference which affects the signal decoding performance at the UAV, we propose a diagonal matrix as a phase-shift framework at the UAV to separate and decode the messages received from the sensors. The study examines the outage probability performance of an individual WSN and provides results based on Monte Carlo simulations and analyses. The investigated results verified the benefits of the *K*-means algorithm in deploying the WSN.

Keywords: wireless sensor network (WSN); unnamed aerial vehicle (UAV); optimal UAV positioning; *K*-means clustering; gap statistic method; centroid-to-next-nearest-centroid (CNNC) trajectory

Citation: Tran, T.-N.; Nguyen, T.-L.; Hoang, V.T.; Voznak, M. Sensor Clustering Using a *K*-Means Algorithm in Combination with Optimized Unmanned Aerial Vehicle Trajectory in Wireless Sensor Networks. *Sensors* **2023**, *23*, 2345. https://doi.org/10.3390/s23042345

Academic Editor: Alberto Gotta

Received: 3 January 2023
Revised: 23 January 2023
Accepted: 16 February 2023
Published: 20 February 2023

1. Introduction

Deployments of wireless sensor networks (WSNs) are increasing because of their beneficial applications. For example, WSNs can be deployed to monitor or collect environmental data (meteorological information such as precipitation, wind speed and direction, air pressure, humidity, temperature, etc.) in remote or difficult terrain [1–6]. A major challenge in deploying a WSN is distributing a large number of wireless sensors over a large and complex geographical area. Wireless sensors are generally low cost, have low power consumption and are highly flexible in their application. However, transmitting a signal from a wireless sensor directly to a control centre presents an important challenge [7], especially if a large number of wireless sensors are deployed to directly collect data. Using terrestrial infrastructure for the purpose of collecting data from wireless sensors is impractical because of high deployment costs and a low flexibility. A potential solution to this problem is using an environmental monitoring system which dispatches an unmanned aerial vehicle (UAV) to a geographical area to retrieve the collected sensor data.

To deploy a WSN, Heinzelman et al. [8] proposed a low-energy adaptive clustering hierarchy (LEACH), a clustering method now considered the most well-known clustering protocol for WSNs. In a hierarchical topology, clusters contain two types of node: cluster members and cluster heads. Member nodes are grouped into different clusters, and in each cluster, a single node is designated a cluster head. The cluster head has the most important role in the cluster, tasked with receiving signals from cluster members and forwarding those signals to other cluster heads [9] or the base station [10].

UAVs have gained increasing consideration as aerial relays which deliver mobility and on-demand wireless connections in areas with complex topography and no network coverage. A UAV's online time, however, is limited by its own on-board energy limitations. The evolution of UAV-assisted WSNs is compelling the scientific community to search for new ways of performing energy harvesting (EH) from external power sources to prolong the online time of UAVs. A variety of effective solutions have been proposed, grouped according to two main types of technique, namely, simultaneous wireless information and power transfer (SWIPT) and techniques for determining the optimal positions for the UAV.

Radio frequency EH shows promise as a potential solution for UAV-assisted WSNs. Initial studies on radio frequency EH were used in a technology termed wireless power transfer (WPT) to recharge the wireless sensors in the WSN. A new radiofrequency EH technique, termed SWIPT, introduced significant benefits to WPT [11]. Many authors have studied SWIPT over the last two decades [12–15], investigating the performance difference between time switching and power splitting in SWIPT protocols [16]. The current study applied a time-switching protocol because it supports a phase for EH.

Applied to all beyond 5G/6G wireless communications, nonorthogonal multiple access (NOMA) provides massive connections, low latency and high reliability [14,17–21]. In the current study, we took advantage of NOMA's benefits, applying NOMA at the UAV to superimpose coding of the data received from wireless sensors and to forward this superimposed signal to a mobile data centre.

1.1. Motivation

The use of machine learning in practical applications is escalating. The authors in [6] applied an artificial neural network (ANN) for sensor clustering. By contrast, some wireless sensors in the study in [6] were clustered as separate single-member clusters. In [22], the authors proposed a distance- and energy-constrained *K*-means clustering scheme (DEKCS) for cluster head selection to prolong the lifetime of underwater WSNs. With this new clustering algorithm, a prospective cluster head was selected according to its position in the cluster and its residual battery level. The authors dynamically updated residual energy thresholds set for prospective cluster heads to ensure that the network fully depleted its energy before disconnection. In this manner, cluster heads could be drained of energy and become inactive/dead sensors. The current study applied the *K*-means algorithm and the gap statistic method first introduced in [23] to obtain an optimal number of subclusters. To our best knowledge, the gap statistic method has not been applied for WSN clustering in any previous study.

In [24], the authors examined a UAV-assisted data collection WSN. The UAV's trajectory was optimized by applying the travelling salesman problem. Note that in [1], the UAV in the proposed network visited every wireless sensor, while in [24], the optimal serving order for sensors was determined according to a standard travelling salesman problem algorithm, which can be optimally solved with the efficient cutting-plane method (i.e., the shortest path from the start point to the end point). The authors also proposed an algorithm which used the pattern search method to solve the problem of optimizing the UAV position and sensor uploading power. In [1], the UAV could be exhausted as a consequence of long flight distances. The authors in [24] used a UAV that navigated the shortest path from the start point to end point, but it consequently ignored/missed some wireless sensors. In another study, the authors addressed the UAV's trajectory problem by jointly optimizing the UAV's velocity, hovering positions and visiting sequence [25]. The

scientific community is very interested in studying UAVs' trajectories for the significant potential gains in aerial network performance. Researchers have applied several types of trajectory, for example, straight trajectory [26,27], circular trajectory [10,28,29] and spiral trajectory [30–32]. The authors in [25] introduced an interesting UAV trajectory scheme (Figure 4), where the UAV visited all N monitoring areas and then found suitable positions to transmit the collected data. In [25], the UAV collected data in four stages: (i) UAV data collection flight, (ii) UAV data collection processing, (iii) UAV data transmission flight and (iv) UAV data transmission processing. That UAV's operating schedule is illustrated in Figure 5 [25].

The current study proposes the use of a UAV for its high mobility, quick implementation and low cost. The main drawback to a small-sized, lightweight aircraft such as a UAV, however, is its limited on-board energy. UAVs are therefore not suitable for flying close to each sensor to collect data, as proposed in [25]. The current study therefore investigated the application of a K-means algorithm to cluster wireless sensors into multiple, optimized subclusters.

1.2. Contribution

Inspired by the studies mentioned in the previous section, we employed a UAV as an aerial relay to provide a sustainable, functional solution for a WSN. The main contributions of the current study are:

- The use of three-dimensional Cartesian coordinates for a WSN which contains a random number of randomly distributed wireless sensors.
- The decomposition of the UAV trajectory optimization into two subproblems: (i) the global WSN cluster is divided into multiple subclusters whose number is optimized with unsupervised machine learning which applies K-means clustering in combination with the gap statistic method; (ii) a centroid-to-next-nearest-centroid algorithm is then applied to find the shortest path for travel through every subcluster.
- An analysis of the system performance of the WSN over Rayleigh distributions and a presentation of the derived closed-form expressions for the outage probability at the UAV and mobile base station.
- Outage probability results for the UAV and mobile base station derived from Monte Carlo simulations and verified with an analysis.

The remainder of the paper is organized as follows: Section 2 introduces the WSN model, wireless sensor clustering algorithm, joint UAV trajectory, free-space channel modelling, and joint UAV operating schedule; Section 3 provides an analysis of the WSN's performance based on outage probability and presents the closed form expressions for outage probability at the UAV and mobile base station; Section 4 examines and plots the investigated results; Section 5 discusses conclusions.

For clarity, Table 1 presents the notation used in the paper.

Table 1. List of important notations.

Notations	Describe	Conditions
N	Random number of sensors	$20 \leq N \leq 50$
K	Optimal number of clusters given by the K-means algorithm, where the number of subclusters K is optimized	$k_{min} \leq K \leq \frac{N}{k_{min}}$
S_n	nth sensor node, where a lower value for n has higher priority	$n = \{1, \ldots, N\}$
$\mathbf{C}$	Global wireless sensor cluster	$\mathbf{C} \supset S_n$ s.t. $\forall n \in N, \lvert\mathbf{C}\rvert = N$

Table 1. *Cont.*

Notations	Describe	Conditions
$\mathbf{C}_k$	kth subcluster	$k = \{1, \ldots, K\}, \mathcal{N}_k = \lvert\mathbf{C}_k\rvert$
$S_n^{(i,k)}$	nth sensor is ith member of the kth subcluster, where a lower value for i has higher priority	$i = \{1, \ldots, \mathcal{N}_K\}, \mathcal{N}_K = \lvert\mathbf{C}_k\rvert$
T	Global transmission time period	$T \in \mathbb{Z}^+$
t	UAV time period	$t = mod(T, K) \vee K$
A_{S_n}, A_U, A_B	Number of antennae at the sensors, UAV and mobile base station	$A_{S_n} \geq 1$, $A_U \geq 1$, and $A_B \geq 1$
ε	Path-loss exponent factor	$\varepsilon \geq 2$
$\tilde{\mathbf{C}}$	Visited cluster set	Updated after the UAV visits the centroid of a subcluster $\mathbf{C}_k$ as given by $\tilde{\mathbf{C}} \leftarrow \tilde{\mathbf{C}} \cup \mathbf{C}_k$
$\mathbf{H}_{S_n,U}, \mathbf{H}_{U,B}$	Precoding fading channel matrices from sensors to the UAV and from the UAV to the mobile base station	$\mathbf{H}_{S_n,U} \in \mathbb{C}^{A_{S_n} \times A_U}$ and $\mathbf{H}_{U,B} \in \mathbb{C}^{A_U \times A_B}$ have sizes of $A_{S_n} \times A_U$ and $A_U \times A_B$, respectively
$\sigma_{S_n,U}, \sigma_{U,B}$	Channel gains	$\sigma_{S_n,U} = E\left\{\left\lvert h_{S_n,U}^{(\cdot,\cdot)}\right\rvert^2\right\}, \sigma_{U,B} = E\left\{\left\lvert h_{U,B}^{(\cdot,\cdot)}\right\rvert^2\right\}$
$\alpha_{S_n^{(i,k)}}$	Power allocation factor for sensor S_n, indexed ith in subcluster $\mathbf{C}_k$	$\alpha_{S_n^{(1,k)}} + \ldots + \alpha_{S_n^{(\mathcal{N}_k,k)}} = 1$ and $\alpha_{S_n^{(1,k)}} > \ldots > \alpha_{S_n^{(\mathcal{N}_k,k)}}$
P_{S_n}, P_U, P_B	Respective power domains at the sensors, UAV and mobile base station B	Let $P_{S_1} = \ldots = P_{S_N}$ dB
$\mathcal{R}$	Predefined bit-rate threshold for sensors	bps/Hz
$\gamma_{U-x_{S_n^{(i,k)}}}, \gamma_{B-x_{S_n^{(i,k)}}}$	SINR reached at UAV U and B when message $x_{S_n^{(i,k)}}$ of sensor S_n is decoded	SIC decodes the message with the biggest power allocation factor by treating other messages and AWGN as interference
$R_{U-x_{S_n^{(i,k)}}}, R_{B-x_{S_n^{(i,k)}}}$	Instantaneous bit rate reached at UAV U and mobile base station B when message $x_{S_n^{(i,k)}}$ of sensor S_n is decoded	bps/Hz
OP_U, OP_B	Outage probabilities at UAV U and mobile base station B	$0 \leq OP_U \leq 1$, $0 \leq OP_B \leq 1$, a lower outage probability result is better performance

2. WSN Model

The current study examines a general WSN with a randomly distributed number of wireless sensors. Figure 1 depicts a WSN with a random number of sensors $N = 42$ positioned at the Cartesian coordinate (x, y, z) in three dimensions. Let us assume that a mobile base station B is positioned at $B(0, 0, 0)$ and each wireless sensor S_n for $n = \{1, \ldots, N\}$ is positioned randomly at coordinate $S_n(x, y, 0)$, where $x = \{0.1, \ldots, 1\}$ and $y = \{0.1, \ldots, 1\}$ as shown in Table 2. For simplicity, we assume that the wireless sensors and mobile base station are positioned relative to a flat earth.

Definition 1. *We denote the global set* $\mathbf{C}$ *as containing all wireless sensors.* $\lvert\mathbf{C}\rvert$ *returns* N*, the total number of wireless sensor nodes (i.e.,* $\lvert\mathbf{C}\rvert = N$*). Let us assume that the data observations (i.e., wireless sensor positioning) are clustered into* K *subclusters, i.e.,* $\mathbf{C} \supseteq \mathbf{C}_1 \cup \ldots \cup \mathbf{C}_K$ *and* $N = \lvert\mathbf{C}\rvert = \lvert\mathbf{C}_1\rvert + \ldots + \lvert\mathbf{C}_K\rvert = \sum_{k=1}^{K} \mathcal{N}_k$*, where* $\mathcal{N}_k = \lvert\mathbf{C}_k\rvert$*.*

Figure 1 illustrates a random distribution of wireless sensors. Each wireless sensor is allocated a given index by the subscript n, where $n = \{1, \ldots, N\}$ and a lower index n has a higher priority. For clarity, sensor S_n has a higher priority than sensor S_{n+1} (e.g., sensor S_1 has a higher priority than sensor S_2).

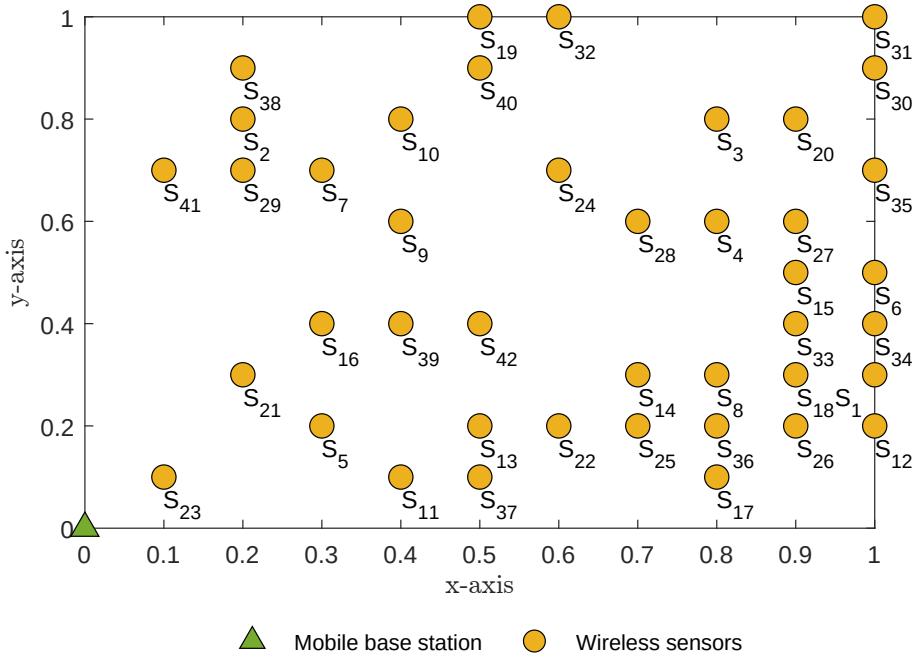


Figure 1. Random positioning of sensor nodes, where $20 \leq N \leq 50$.

Table 2. Wireless sensor positions are distributed randomly.

Sensors	x-Coordinate	y-Coordinate	Sensors	x-Coordinate	y-Coordinate
S_1	1	0.3	S_2	0.2	0.8
S_3	0.8	0.8	S_4	0.8	0.6
S_5	0.3	0.2	S_6	1	0.5
S_7	0.3	0.7	S_8	0.8	0.3
S_9	0.4	0.6	S_{10}	0.4	0.8
S_{11}	0.4	0.1	S_{12}	1	0.2
S_{13}	0.5	0.2	S_{14}	0.7	0.3
S_{15}	0.9	0.5	S_{16}	0.3	0.4
S_{17}	0.8	0.1	S_{18}	0.9	0.3
S_{19}	0.5	1	S_{20}	0.9	0.8
S_{21}	0.2	0.3	S_{22}	0.6	0.2
S_{23}	0.1	0.1	S_{24}	0.6	0.7
S_{25}	0.7	0.2	S_{26}	0.9	0.2
S_{27}	0.9	0.6	S_{28}	0.7	0.6
S_{29}	0.2	0.7	S_{30}	1	0.9
S_{31}	1	1	S_{32}	0.6	1
S_{33}	0.9	0.4	S_{34}	1	0.4
S_{35}	1	0.7	S_{36}	0.8	0.2
S_{37}	0.5	0.1	S_{38}	0.2	0.9
S_{39}	0.4	0.4	S_{40}	0.5	0.9
S_{41}	0.1	0.7	S_{42}	0.5	0.4

Note: Two or more wireless sensors will never occupy the same position; each position is allocated only one wireless sensor, as illustrated in Figure 1.

2.1. WSN Clustering

Remark 1. *Because the optimization problem is complex, we propose breaking it down into several subproblems and observing the random distribution of wireless sensors over a large geographical area, as illustrated in Figure 1. The global wireless sensor cluster can be divided into multiple subclusters. The number of subclusters can be optimized by applying the gap statistic method, and the wireless sensors can be assigned to a subcluster using the K-means algorithm. To solve these problems, we propose a solution in Proposition 1.*

Proposition 1. *The optimal number of subclusters is yielded as follows:*

- *Observing the latitudes (x-axis) and longitudes (y-axis) of the wireless sensors, we determine the optimal number of subclusters* $K \leftarrow k_{optimal}$. *The gap statistic method is applied to the number of subclusters k to compute the corresponding total within the intracluster variation* W_k, *i.e., the sum of squares function, given by*

$$W_k = \sum_{\kappa=1}^{k} \frac{1}{2|\mathbf{C}_\kappa|} \sum_{S_i, S_j \in \mathbf{C}_\kappa} d_{S_i,S_i}, \tag{1}$$

  where $\kappa = \{1, \ldots, k\}$, $|\mathbf{C}_\kappa|$ *returns the number of wireless sensor nodes in cluster* $\mathbf{C}_\kappa$ *and* d_{S_i,S_j} *is the squared Euclidean distance of all pairwise sensor nodes in the cluster* $\mathbf{C}_\kappa$ *for* $S_i, S_j \in \mathbf{C}_\kappa$ *and* $i \neq j$. *It is important to note that we assume* $k_{min} = 4$ *and* $k_{max} = \frac{|\mathbf{C}|}{k_{min}} = \frac{N}{k_{min}}$, *where* $|\mathbf{C}|$ *is the number of observations (or the number of sensors within the global cluster* $\mathbf{C}$*). Let us briefly consider factors* k_{min} *and* k_{max}. *We define* $k_{min} = 4$ *to prevent a uniform distribution of wireless sensors positions throughout the area; the gap statistic method thus returns the optimal number of clusters* $k_{optimal} = 1$. *For example, Figure 2a,b in [23] plots the distribution of sensors spread throughout a region and the corresponding optimal number of clusters at* $K = 1$, *respectively. However, we also define* $k_{max} = \frac{N}{k_{min}}$ *to prevent each wireless sensor owning a private cluster, a problem that would lead to a UAV visiting every wireless sensor to collect data.*
- *Reference data sets* Ω *with a random uniform distribution are generated. Each reference data set* ω *of these reference data sets* Ω *is clustered with a variable number of clusters* $k = \{k_{min}, \ldots, k_{max}\}$. *The corresponding total is computed within the intracluster variation* $W_{\kappa\omega}$ *given in the dispersion metrics for* $\kappa = \{1, \ldots, k\}$ *and* $\omega = \{1, \ldots, \Omega\}$.
- *The estimated gap statistic is computed as the deviation of the observed* W_k *value from its expected value* $W_{\kappa\omega}$ *under the null hypothesis* $Gap(k) = \frac{1}{\Omega} \sum_{\omega=1}^{\Omega} \log(W_{\kappa\omega}) - \log(W_k)$. *Let* $l = \frac{1}{\Omega} \sum_{\omega=1}^{\Omega} \log(W_{\kappa\omega})$. *The standard deviation (sd) of the statistics is then computed, given by* $sd_k = \sqrt{\frac{1}{\Omega} \sum_{\omega=1}^{\Omega} (\log(W_{\kappa\omega}) - l)^2}$.
- *Using the gap statistic method, the smallest value of* κ *is selected as the optimal number of clusters, the gap statistic being within one standard deviation of the gap statistic at* $\kappa + 1$, *given* $k_{optimal} = \min\{k\}$ *and* $Gap(k) \geq Gap(k+1) - \theta_{k+1}$, *where* $\theta_{k+1} = sd_{k+1}\sqrt{1 + \frac{1}{\Omega}}$.

For example, Figure 2 indicates the optimal number of clusters at $k_{optimal} = 4$, determined by the gap statistic algorithm according to the randomly positioned sensor nodes shown in Figure 1.

The *K*-means algorithm was used to calculate the position for each centroid, with an optimal number of clusters $K \leftarrow k_{optimal}$. The computed centroids of four subclusters ($K = 4$) are listed in Table 3. Figure 3 illustrates all wireless sensor nodes after clustering to *K* subclusters. After clustering, each sensor node is grouped into a subcluster; for example, S_1^3 indicates that sensor S_1 is a member of subcluster $\mathbf{C}_3$ (Figure 3).

Table 3. Centroids after clustering.

Centroids	x-Axis	y-Axis	Centroids	x-Axis	y-Axis
$\mathbf{C}_1$	0.38	0.24	$\mathbf{C}_2$	0.3636	0.8
$\mathbf{C}_3$	0.8769	0.3	$\mathbf{C}_4$	0.8875	0.75

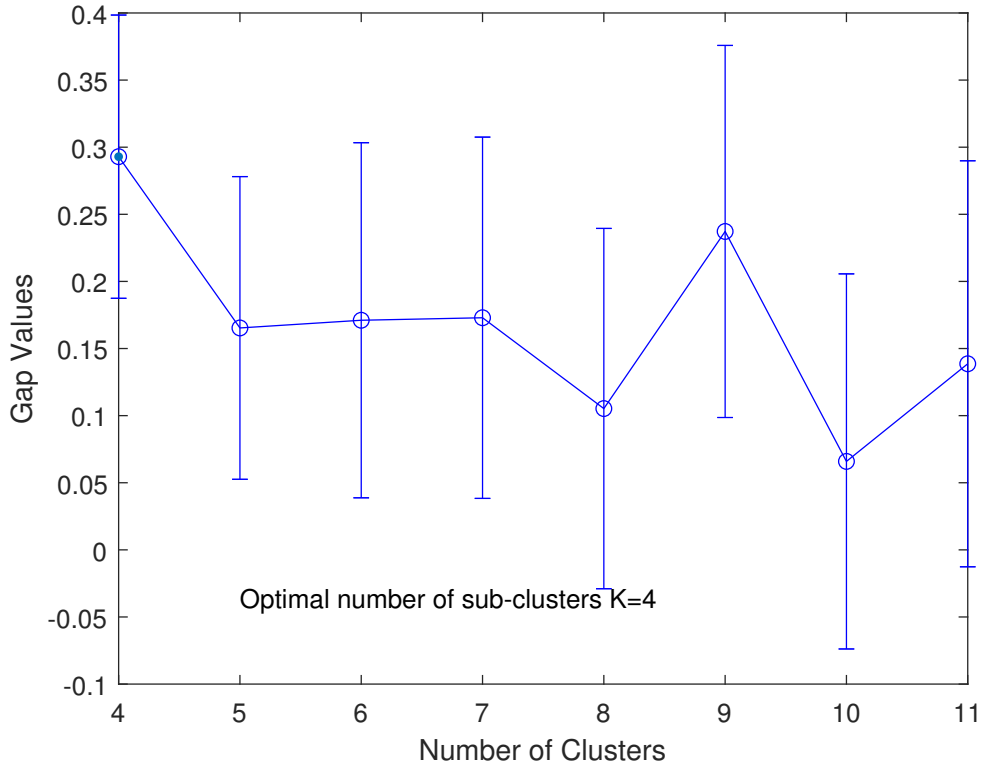


Figure 2. Optimized number of subclusters using the gap statistic method, the optimal number of clusters at $K = 4$ satisfying the first maximum standard error.

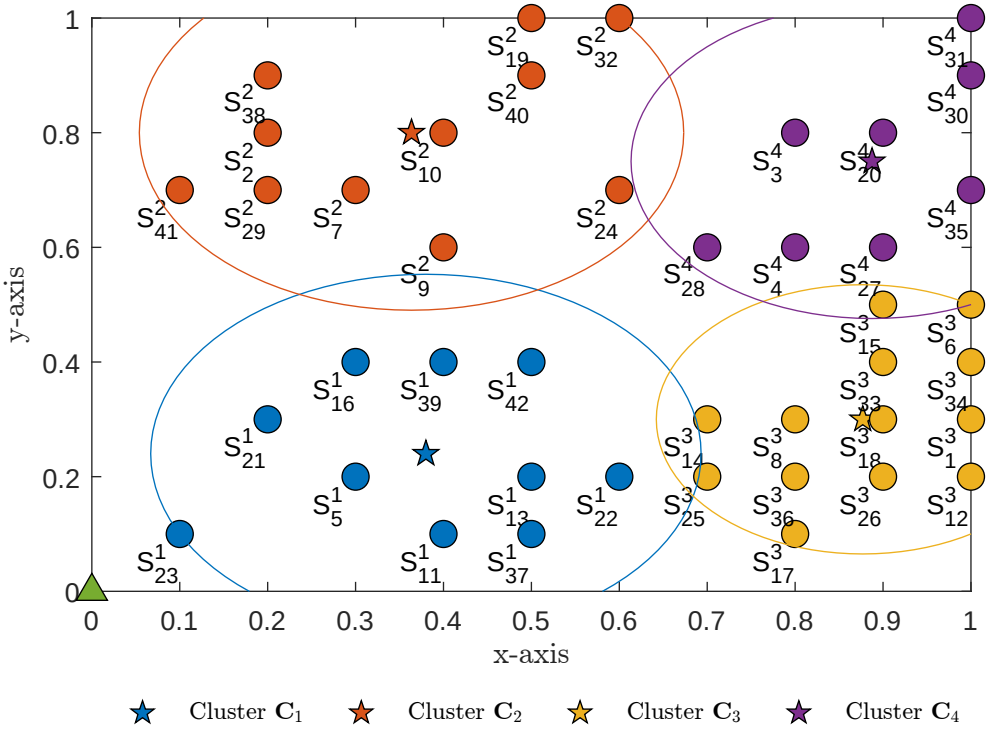


Figure 3. Sensor clustering using the *K*-means algorithm, with optimal number of clusters $K = 4$.

2.2. Joint UAV Trajectory

Definition 2. *We introduce a novel joint UAV trajectory algorithm to compute the centroid to next nearest centroid.*

Problem 1. *Operating as a flying relay, the UAV has the advantage of a high mobility and is able to fly close to wireless sensors to receive and forward a superimposed signal. This, however, leads to a long flight path, and the other wireless sensors must wait to be served. We minimized the flight time/path of the UAV according to the cluster centroid positions shown in Table 3. How to obtain the shortest flight path is outlined in Proposition 2.*

Proposition 2. *The centroid-to-next-nearest-centroid trajectory was computed as follows:*

- *Step 1: To determine the nearest centroid from the mobile base station B, we calculate the smallest pairwise Cartesian distance from the mobile base station to each subcluster centroid.*
- *Step 2: The UAV selects the next nearest cluster centroid. In this case, the UAV considers candidate centroids without regard to any of the previously selected cluster centroids in* $\tilde{\mathbf{C}}$*. It is important that the centroids contained in the visited set* $\tilde{\mathbf{C}}$ *be removed from the candidate list to prevent the UAV returning to the previous subcluster* $\tilde{\mathbf{C}}$*. The UAV repeats Step 2 (i.e.,* $\mathbf{C} \backslash \tilde{\mathbf{C}} \neq \emptyset$*) until the list of candidate subclusters is empty (i.e.,* $\mathbf{C} \backslash \tilde{\mathbf{C}} = \emptyset$*).*

Without loss of generality, we examined a single round trip of the UAV. Table 4 lists the next nearest subcluster centroids determined from the above selection strategy. The results in Figure 3 indicate that subcluster $\mathbf{C}_1$ was the nearest to the mobile base station B compared to the other subclusters. Subcluster $\mathbf{C}_1$ was therefore selected at block period time $T = 1$. The UAV visited subcluster $\mathbf{C}_1$ first to collect data from all sensor members in subcluster $\mathbf{C}_1$. The visited set $\tilde{\mathbf{C}} \leftarrow \tilde{\mathbf{C}} \cup \mathbf{C}_1$ was then updated. After all data from the sensor members in subcluster $\mathbf{C}_1$ were collected, the UAV selected subcluster $\mathbf{C}_3$ because it contained the next nearest subcluster centroid. The UAV then visited subcluster $\mathbf{C}_3$ at global time period $T = 2$ to collect data from all sensor members in the subcluster. The visited set $\tilde{\mathbf{C}} \leftarrow \tilde{\mathbf{C}} \cup \mathbf{C}_3$ was again updated. The UAV continued to follow this procedure, selecting the next nearest centroid and updating the visited set, until all data have been collected from each subcluster. In this manner, the UAV followed the shortest possible flight path, as shown in Figure 4. After travelling through all K subclusters ($\tilde{\mathbf{C}} \equiv \mathbf{C}$) and collecting all data from wireless sensor members in each subcluster, the UAV's task was complete and it returned to the mobile base station. For the real-time application of a UAV-assisted WSN, the UAV would repeat the round trips summarized in Table 5.

Table 4. Pairwise centroid-to-centroid distance based on Cartesian distances.

	$\mathbf{C}_1$	$\mathbf{C}_2$	$\mathbf{C}_3$	$\mathbf{C}_4$
$\mathbf{C}_1$	~~0~~	0.5602	**0.5005**	0.7195
$\mathbf{C}_3$	~~0.5005~~	0.7166	~~0~~	**0.4501**
$\mathbf{C}_4$	~~0.7195~~	**0.5262**	~~0.4501~~	~~0~~
$\mathbf{C}_2$	~~0.5602~~	~~0~~	~~0.7166~~	~~0.5262~~

Observing Table 4, notice that numbers with inclined lines (e.g., ~~0~~) and numbers with bold (e.g., **0.5005**) mean visited clusters and next nearest clusters. For clarity, when UAV visited cluster $\mathbf{C}_1$ (row with $\mathbf{C}_1$), the UAV selects the next-nearest cluster (i.e., $\mathbf{C}_3$) and ignores cluster $\mathbf{C}_1$. Next, the UAV visited cluster $\mathbf{C}_3$ (row with $\mathbf{C}_3$), the UAV selects the next-nearest cluster (i.e., $\mathbf{C}_4$) and ignores clusters $\mathbf{C}_1$ and $\mathbf{C}_3$. The remaining rows in Table 4 have the same meaning.

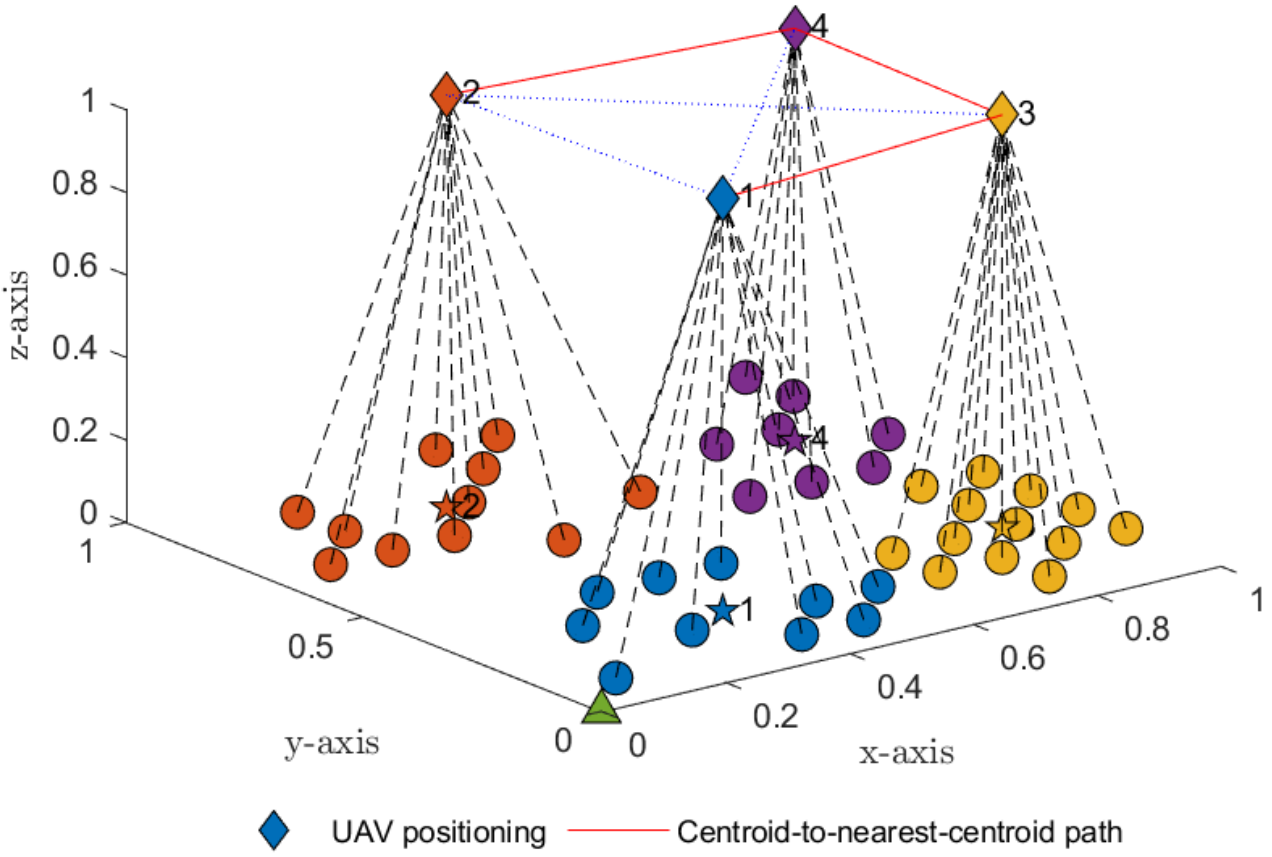


Figure 4. Joint UAV trajectory and the shortest path based on the centroid-to-next-nearest-centroid distance given by Algorithm 1 (i.e., $\mathbf{C}_1 \rightarrow \mathbf{C}_3 \rightarrow \mathbf{C}_4 \rightarrow \mathbf{C}_2$).

Algorithm 1 *K*-means clustering for the optimal number of subclusters and shortest path determined from a centroid to the next nearest centroid

Input: Generate a wireless sensor network with a number N of randomly positioned wireless sensors;

Output: An optimal number of subclusters K and subcluster centroids;

1: Initialize variables $k_{min} = 4$, $k_{max} = \frac{N}{k_{min}}$;
2: Attempt $\forall k = \{k_{\min}, \ldots, k_{\max}\}$ to find the optimal number K of subclusters, computed according to Proposition 1;
3: Find the centroid positions for K subclusters by applying K-means clustering;
4: Compute the pairwise distances between the mobile base station B and subcluster centroids;
5: Select the nearest centroid and update $\tilde{\mathbf{C}}$;
6: **while** $\left|\mathbf{C} \cap \tilde{\mathbf{C}}\right| \neq |\mathbf{C}|$ **do**
7: Compute the pairwise distances between the current centroid and other centroids;
8: Select the nearest centroid and then update $\tilde{\mathbf{C}}$.
9: **end while**
10: **return** the number of subclusters K, centroid positions $\mathbf{C}_k(x, y)$ for $k \in K$ and the shortest path.

Table 5. Joint trajectory schedule for global transmission time period T and optimal number of clusters K, where the UAV period $t = \{(mod\{T,K\}|mod\{T,K\} \neq 0) \vee (K|mod\{T,K\} = 0)\}$.

Global period T	1	2	3	4	5 ...
UAV period t	1	2	3	4	1 ...
Clusters $\mathbf{C}_k|k \in K$	$\mathbf{C}_1$	$\mathbf{C}_3$	$\mathbf{C}_4$	$\mathbf{C}_2$	$\mathbf{C}_1$ ...
No. members $\mathcal{N}_k = \|\mathbf{C}_k\|$	10	13	8	11	10 ...
Members $S_n^{(i,k)}$	$S_5^{(1,1)}, S_{11}^{(2,1)}, S_{13}^{(3,1)}, S_{16}^{(4,1)}, S_{21}^{(5,1)}, S_{22}^{(6,1)}, S_{23}^{(7,1)}, S_{37}^{(8,1)}, S_{39}^{(9,1)}, S_{42}^{(10,1)}$	$S_1^{(1,3)}, S_6^{(2,3)}, S_8^{(3,3)}, S_{12}^{(4,3)}, S_{14}^{(5,3)}, S_{15}^{(6,3)}, S_{17}^{(7,3)}, S_{18}^{(8,3)}, S_{25}^{(9,3)}, S_{26}^{(10,3)}, S_{33}^{(11,3)}, S_{34}^{(12,3)}, S_{36}^{(13,3)}$	$S_3^{(1,4)}, S_4^{(2,4)}, S_{20}^{(3,4)}, S_{27}^{(4,4)}, S_{28}^{(5,4)}, S_{30}^{(6,4)}, S_{31}^{(7,4)}, S_{35}^{(8,4)}$	$S_2^{(1,2)}, S_7^{(2,2)}, S_9^{(3,2)}, S_{10}^{(4,2)}, S_{19}^{(5,2)}, S_{24}^{(6,2)}, S_{29}^{(7,2)}, S_{32}^{(8,2)}, S_{38}^{(9,2)}, S_{40}^{(10,2)}, S_{41}^{(11,2)}$	$S_5^{(1,1)}, S_{11}^{(2,1)}, S_{13}^{(3,1)}, S_{16}^{(4,1)}, S_{21}^{(5,1)}, S_{22}^{(6,1)}, S_{23}^{(7,1)}, S_{37}^{(8,1)}, S_{39}^{(9,1)}, S_{42}^{(10,1)}$

2.3. Channel Modelling for a UAV-Assisted WSN

In our previous work [33], we considered free space (i.e., air-to-ground (A2G), ground-to-air (G2A) and air-to-air (A2A)) and first introduced the flat-earth distance based on real latitudes and longitudes. The proposed solutions were effective in determining and tracking the optimal positions for the UAV. A separate study [33] examined the problems related to channel modelling in a WSN which contained multiple subclusters. In the current study, we address the uplinks, i.e., the channels from the wireless sensors to the UAV ($\mathbf{H}_{S_n,U}$) and the channels from the UAV to the mobile base station ($\mathbf{H}_{U,B}$). The precoding channel matrices $\mathbf{H}_{S_n,U}$ and $\mathbf{H}_{U,B}$ are expressed by

$$\mathbf{H}_{S_n,U} = \begin{bmatrix} h_{S_n,U}^{(1,1)} & \cdots & h_{S_n,U}^{(1,A_U)} \\ \vdots & \ddots & \vdots \\ h_{S_n,U}^{(A_{S_n},1)} & \cdots & h_{S_n,U}^{(A_{S_n},A_U)} \end{bmatrix} \in \mathbb{C}^{A_{S_n} \times A_U}, \tag{2}$$

where A_{S_n} and A_U are the number of antennae on the wireless sensor S_n and UAV U, respectively; the channel coefficient $h_{S_n,U}^{(.,.)} \in \mathbf{H}_{S_n,U}$ is formulated according to $h_{S_n,U}^{(.,.)} = g\left(d_{S_n,U}^{G2A}\right)^{-\varepsilon}$, where g is the Rayleigh fading channel, ε is the path-loss exponent, and $d_{S_n,U}^{G2A}$ is the G2A distance from the sensor node S_n to UAV U. Note that the free-space distance based on latitude and longitude is given by expression ([33], Equation (2)). For simplicity and without loss of generality, all wireless sensors are allocated with Cartesian coordinates in a three-dimensional space. The G2A distance from wireless sensor S_n to UAV U is therefore given by $d_{S_n,U}^{(\mathrm{G2A})} = \sqrt{|x_{S_n} - x_U|^2 + |y_{S_n} - y_U|^2 + |z_{S_n} - z_U|^2}$, where the x, y and z axes represent latitude, longitude and altitude, respectively, on a flat earth.

Similarly, the precoding channel matrix $\mathbf{H}_{U,B}$ is expressed as

$$\mathbf{H}_{U,B} = \begin{bmatrix} h_{U,B}^{(1,1)} & \cdots & h_{U,B}^{(1,A_U)} \\ \vdots & \ddots & \vdots \\ h_{U,B}^{(A_B,1)} & \cdots & h_{U,B}^{(A_B,A_U)} \end{bmatrix} \in \mathbb{C}^{A_B \times A_U}, \tag{3}$$

where A_B is the number of antennae at the mobile base station, and the channel coefficient $h_{U,B}^{(.,.)} \in \mathbf{H}_{U,B}$ is formulated according to $h_{U,B}^{(.,.)} = g\left(d_{S_n,U}^{A2G}\right)^{-\varepsilon}$, where $d_{U,B}^{A2G}$ is the A2G distance from the UAV U to the mobile base station B and given by $d_{U,B}^{(A2G)} = \sqrt{|x_U - x_B|^2 + |y_U - y_B|^2 + |z_U - z_B|^2}$.

2.4. UAV Joint Schedule

This study introduces a novel scheduling protocol for a UAV-assisted WSN. The coefficient t is the time required to complete a transmission cycle of three phases, i.e., λ_1, λ_2 and λ_3, where λ_1 is the first phase during which the UAV receives signals from the sensor nodes in a cluster, λ_2 is the second phase during which the UAV receives radiofrequency energy from the mobile base station, and λ_3 is the third phase during which the UAV transmits the superimposed signals to the mobile base station for data analysis. Figure 5 depicts an electronic control unit (ECU) which performs a task corresponding to a predefined operation in a common UAV schedule.

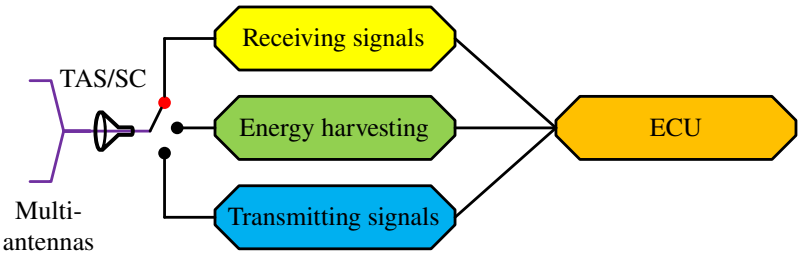


Figure 5. Joint schedule.

The ECU implements an electronic switch which applies three successive modes during the same transmission block t:

- In phase λ_1, the interface of the receiving signal circuit is active while the other interfaces are inactive. The UAV receives signals from the sensor nodes in the currently visited subcluster, given by (5).
- In phase λ_2, the interface of the EH circuit is active while the other interfaces are inactive. The UAV receives radiofrequency energy from the mobile base station B, given by (10), while the ECU decodes the messages from the signals received from the wireless sensors.
- In phase λ_3, the interface of the transmitting signal circuit is active while the other interfaces are inactive. The UAV encodes the messages received in the first phase and forwards the superimposed signals to the mobile base station B, given by (11).

2.4.1. Phase 1: Uplinks between Wireless Sensors and the UAV

In the first phase, having selected the next nearest subcluster centroid, the UAV visits and hovers at the selected centroid and receives signals wirelessly from the subcluster's sensors. Figure 6 illustrates the procedure of receiving signals and processing data at the UAV during the first phase.

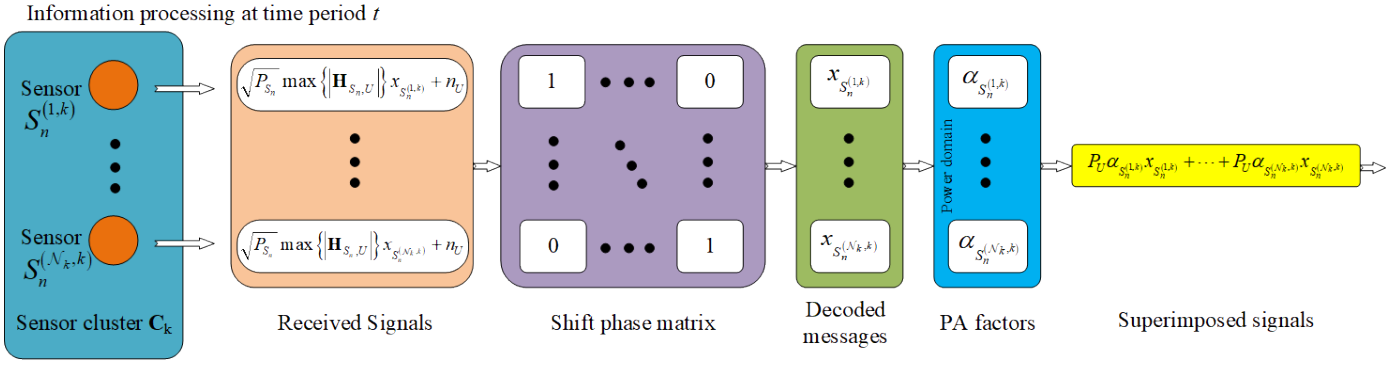


Figure 6. Procedure of processing data at the UAV.

Based on the number of subclusters K and the global period T, we calculate the UAV period t as follows:

$$t = \begin{cases} mod\{T,K\}, & \text{s.t. } mod\{T,K\} > 0, \\ K, & \text{s.t. } mod\{T,K\} = 0, \end{cases} \tag{4}$$

where the $mod\{T,K\}$ function refers to the modulo value between T and K (e.g., for $T = 10$ and $K = 4$, $t = mod\{T,K\} = 2$). At each UAV period t, the UAV selects the next nearest centroid using the centroid-to-next-nearest-centroid algorithm. According to the trajectory mapped in Table 5, for $T = 10$, $K = 4$ and $t = 2$, the UAV selects the subcluster $\mathbf{C}_3$ and serves wireless sensors $S_1^{(1,3)}$, $S_6^{(2,3)}$, $S_8^{(3,3)}$, $S_{12}^{(4,3)}$, $S_{14}^{(5,3)}$, $S_{15}^{(6,3)}$, $S_{17}^{(7,3)}$, $S_{18}^{(8,3)}$, $S_{25}^{(9,3)}$, $S_{26}^{(10,3)}$, $S_{33}^{(11,3)}$, $S_{34}^{(12,3)}$ and $S_{36}^{(13,3)}$. The signals received from the wireless sensors in subcluster $\mathbf{C}_k$ over UAV period t are given by

$$\max_{[A_{S_n} \times A_U]} \left\{ y_{S_n^{(i,k)}}(T,N,K) \right\} = \sqrt{P_{S_n}} \max_{[A_{S_n} \times A_U]} \left\{ \left| \mathbf{H}_{S_n,U} \right| \right\} x_{S_n^{(i,k)}} + n_U, \tag{5}$$

$$s.t. \qquad 1 \le n \le N, 1 \le k \le K, 1 \le i \le \mathcal{N}_k, \mathcal{N}_k = |\mathbf{C}_k|, \tag{6}$$

where t is obtained from (4), k is mapped as in Table 5, and P_{S_n} is the transmit power of sensor S_n. For simplicity, we assume that $P_{S_1} = \ldots = P_{S_N}$.

Let us denote $\mathbf{D}(T,N,K)$, which is the mathematical description of a diagonal matrix, as follows:

$$\mathbf{D}(T,N,K) = diag(1,\ldots,1)_{\mathcal{N}_k \times \mathcal{N}_k} = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}_{|\mathbf{C}_k| \times |\mathbf{C}_k|}, \tag{7}$$

where the diagonal matrix $\mathbf{D}$ has the size $|\mathbf{C}_k| \times |\mathbf{C}_k|$ for transmission period T and all nondiagonal elements are zero, as indicated in Figure 6. The predecoded matrix obtained at the UAV is derived by multiplying the received signal matrix (5) with the diagonal matrix (7); thus, $preDecode(T,N,K) = y_{S_n^{(i,k)}}(T,N,K) \times \mathbf{D}(T,N,K)$. The UAV selects each element in the predecoded matrix to obtain the SINR at the point when the UAV decodes message x_{S_n} from sensor $S_n \in \mathbf{C}_k$, as follows:

$$\max_{[A_{S_n} \times A_U]} \left\{ \gamma_{U - x_{S_n^{(i,k)}}}(T,N,K) \right\} = \rho_{S_n} \max_{[A_{S_n} \times A_U]} \left\{ |\mathbf{H}_{S_n,U}|^2 \right\}, \tag{8}$$

where the signal-to-noise ratio (SNR) $\rho_{S_n} = P_{S_n}/N_0$. For simplicity, we assume that $\rho_{S_1} = \ldots = \rho_{S_N}$.

We then obtain the instantaneous bit rate at the point when the UAV decodes message $x_{S_n^{(i,k)}}$ from sensor $S_n^{(i,k)}$, as follows:

$$\max_{[A_{S_n} \times A_U]} \left\{ R_{U - x_{S_n^{(i,k)}}}(T,N,K) \right\} = \frac{1}{2} \log_2 \left(1 + \max_{[A_{S_n} \times A_U]} \left\{ \gamma_{U - x_n^{(i,k)}}(T,N,K) \right\} \right). \tag{9}$$

2.4.2. Phase 2: Prolong the UAV's Online Time with EH

The most challenging aspect of deploying a UAV is managing its power limitations as a small, lightweight aircraft. We propose applying SWIPT techniques to prolong the UAV's online time. In a previous study [33] (Figure 4), we adopted a power splitting protocol. In the current study, we applied a time-switching technique for its advantages in a WSN (Figure 5); the technique is different from the proposed time-switching models in [33]

(Figure 4). In phase λ_2, the UAV harvests radiofrequency energy from the mobile base station according to

$$EH(T, N, K) = \eta P_B \sigma_{B,U}, \tag{10}$$

where P_B is the power domain at the mobile base station, and $\sigma_{B,U}$ is the expected channel gain between the mobile base station and the UAV at its current UAV position. It is important to note that η is the collected energy factor and that we assume $\eta = 1$ for simplicity.

2.4.3. Phase 3: Transmitting Signals

In [10], the authors applied the amplify-and-forward protocol at the UAV to receive and forward signals to a single device. In the current study, we implemented a decode-and-forward protocol at the UAV to ensure that the UAV received, decoded and encoded messages successfully before forwarding the superimposed signals to the mobile base station. To improve latency, we applied the emerging NOMA technique for its high spectral efficiency. The UAV U encoded the messages $\forall x_{S_n{}^{(i,k)}} \in \mathbf{X}_k$ from the sensors in the current subcluster $\mathbf{C}_k$ and superimposed them into the signal by sharing the power domain P_U and using different power allocation factor $\alpha_{S_n}^{(i,k)}$. From the precoding matrix $\mathbf{H}_{U,B}$, as given by (3), only the best channel was selected for signal transmission.

In the third phase λ_3 of transmission block t, the mobile base station B received radiofrequency signals as follows:

$$\max_{[A_U \times A_B]} \{y_B(T, N, K)\} = \max_{[A_U \times A_B]} \{|\mathbf{H}_{U,B}|\} \sum_{\forall S_n^{(i,k)} \in \mathbf{C}_k} \sqrt{P_U \alpha_{S_n^{(i,k)}}} x_{S_n^{(i,k)}} + n_B, \tag{11}$$

where n_B is the AWGN (i.e., $n_B \sim CN(0, N_0)$ with zero mean and variance N_0) at the mobile base station B, P_U is the power domain at UAV U, and $\alpha_{S_n^{(i,k)}}$ is the power allocation factor for message $x_{S_n^{(i,k)}}$ of wireless sensor $S_n^{(i,k)}$. The NOMA technique applies superimposed coding by sharing the power domain and therefore, the power allocation strategy strongly affects the success or failure of decoding a message. Previous studies [17,21,34] have also applied power allocation strategies; the current study, however, addresses a WSN divided into multiple subclusters and therefore proposes the novel power allocation strategy described below.

Proposition 3. *The power allocation strategy for transmitting messages from wireless sensors over UAV transmission period t while the UAV visits subcluster* $\mathbf{C}_k$ *is expressed as follows:*

$$\alpha_{S_n^{(i,k)}} = \frac{\mathcal{N}_k - i + 1}{\sum_{j=1}^{\mathcal{N}_k} j}, \tag{12}$$

where a sensor with a higher priority is allocated a larger power allocation factor; for example, sensor S_1*, which has the highest priority and is the first member in subcluster* $\mathbf{C_3}$*, is allocated the largest power allocation factor* $\alpha_{S_1^{(1,3)}} = 0.1428$*, whereas sensor* S_{36}*, which has the lowest priority and is the last member in the subcluster* $\mathbf{C}_3$*, is allocated the smallest power allocation factor* $\alpha_{S_{36}^{(13,3)}} = 0.011$*. For clarity, we applied the power allocation factors presented in Table 6. From Equation (12), the power allocation strategy in the subcluster is constrained such that* $\alpha_{S_n^{(i,k)}} > \ldots > \alpha_{S_n^{(1,k)}}$ *and* $\alpha_{S_n^{(i,k)}} + \ldots + \alpha_{S_n^{(1,k)}} = 1$.

Table 6. Power allocation factors at wireless sensors for transmitting messages, arranged according to subclusters.

$\mathbf{C}_1$	$\alpha_{S_5^{(1,1)}} = 0.18182$, $\alpha_{S_{11}^{(2,1)}} = 0.16364$, $\alpha_{S_{13}^{(3,1)}} = 0.14545$, $\alpha_{S_{16}^{(4,1)}} = 0.12727$, $\alpha_{S_{21}^{(5,1)}} = 0.10909$, $\alpha_{S_{22}^{(6,1)}} = 0.090909$, $\alpha_{S_{23}^{(7,1)}} = 0.072727$, $\alpha_{S_{37}^{(8,1)}} = 0.054545$, $\alpha_{S_{39}^{(9,1)}} = 0.036364$, $\alpha_{S_{42}^{(10,1)}} = 0.018182$
$\mathbf{C}_2$	$\alpha_{S_2^{(1,2)}} = 0.16667$, $\alpha_{S_7^{(2,2)}} = 0.15152$, $\alpha_{S_9^{(3,2)}} = 0.13636$, $\alpha_{S_{10}^{(4,2)}} = 0.12121$, $\alpha_{S_{19}^{(5,2)}} = 0.10606$, $\alpha_{S_{24}^{(6,2)}} = 0.090909$, $\alpha_{S_{29}^{(7,2)}} = 0.075758$, $\alpha_{S_{32}^{(8,2)}} = 0.060606$, $\alpha_{S_{38}^{(9,2)}} = 0.045455$, $\alpha_{S_{40}^{(10,2)}} = 0.030303$, $\alpha_{S_{41}^{(11,2)}} = 0.015152$
$\mathbf{C}_3$	$\alpha_{S_1^{(1,3)}} = 0.14286$, $\alpha_{S_6^{(2,3)}} = 0.13187$, $\alpha_{S_8^{(3,3)}} = 0.12088$, $\alpha_{S_{12}^{(4,3)}} = 0.10989$, $\alpha_{S_{14}^{(5,3)}} = 0.098901$, $\alpha_{S_{15}^{(6,3)}} = 0.087912$, $\alpha_{S_{17}^{(7,3)}} = 0.076923$, $\alpha_{S_{18}^{(8,3)}} = 0.065934$, $\alpha_{S_{25}^{(9,3)}} = 0.054945$, $\alpha_{S_{26}^{(10,3)}} = 0.043956$, $\alpha_{S_{33}^{(11,3)}} = 0.032967$, $\alpha_{S_{34}^{(12,3)}} = 0.021978$, $\alpha_{S_{36}^{(13,3)}} = 0.010989$
$\mathbf{C}_4$	$\alpha_{S_3^{(1,4)}} = 0.22222$, $\alpha_{S_4^{(2,4)}} = 0.19444$, $\alpha_{S_{20}^{(3,4)}} = 0.16667$, $\alpha_{S_{27}^{(4,4)}} = 0.13889$, $\alpha_{S_{28}^{(5,4)}} = 0.11111$, $\alpha_{S_{30}^{(6,4)}} = 0.083333$, $\alpha_{S_{31}^{(7,4)}} = 0.055556$, $\alpha_{S_{35}^{(8,4)}} = 0.027778$

The SINR at the mobile base station B when B decodes message $x_{S_n^{(i,k)}} \in \mathbf{X}_k$ treats other messages $x_{S_n^{(j,k)}} \in \mathbf{X}_k$, where $\alpha_{S_n^{(j,k)}} < \alpha_{S_n^{(i,k)}}$, and AWGN n_B as interference by applying SIC:

$$\max_{[A_U \times A_B]} \left\{ \gamma_{B-x_{S_n^{(i,k)}}}(T,N,K) \right\} = \frac{\max\limits_{[A_U \times A_B]} \left\{ |\mathbf{H}_{U,B}|^2 \right\} \alpha_{S_n^{(i,k)}} \rho_U \sigma_{U,B}}{\max\limits_{[A_U \times A_B]} \left\{ |\mathbf{H}_{U,B}|^2 \right\} \rho_U \sigma_{U,B} \sum\limits_{j=i+1}^{\mathcal{N}_k} \alpha_{S_n^{(j,k)}} + 1}, \tag{13}$$

$$= \max_{[A_U \times A_B]} \left\{ |\mathbf{H}_{U,B}|^2 \right\} \alpha_{S_n^{(\mathcal{N}_k,k)}} \rho_U \sigma_{U,B}, \tag{14}$$

where $i < \mathcal{N}_k$ in (13) and $i = \mathcal{N}_k$ in (14).

The maximum instantaneous bit-rate threshold attained if the mobile base station decodes message $x_{S_n^{(i,k)}}$ in the best-received signal, given by (11), is expressed as:

$$\max_{[A_U \times A_B]} \left\{ R_{B-x_{S_n^{(i,k)}}}(T,N,K) \right\} = \frac{1}{2} \log_2 \left(1 + \max_{[A_U \times A_B]} \left\{ \gamma_{B-x_{S_n^{(i,k)}}}(T,N,K) \right\} \right), \tag{15}$$

where $\forall x_{S_n^{(i,k)}} \in \mathbf{X}_k$, and $\forall i = \{1, \ldots, \mathcal{N}_k\}$.

3. System Performance Analysis

In this section, we derive the novel closed-form expressions for the independent outage probability at the UAV and the dependent outage probability at the mobile base station.

3.1. Outage Probability Performance at the UAV

Theorem 1. *The independent outage probability at the UAV U relates to the UAV's unsuccessful decoding of the message in the received signal, given by (5). In other words, the maximum instantaneous bit-rate threshold, given by (9), cannot reach the predefined bit-rate threshold $\mathcal{R}$. The independent outage probability at the UAV U in transmission block t is therefore expressed as*

$$OP_{U-x_{S_n^{(i,k)}}}(T,N,K) = 1 - \Pr\left\{ \max_{[A_{S_n} \times A_U]} \left\{ R_{U-x_{S_n^{(i,k)}}}(T,N,K) \right\} \geq \mathcal{R} \right\}. \tag{16}$$

Based on Equation (16), we propose Algorithm 2 to calculate the Monte Carlo simulations for the outage probability at the UAV U.

Algorithm 2 Calculate the outage probability at the UAV U from (16) for transmission block t.

Input: Initialize the parameters as in Table 1 and randomly generate 10^6 samples of each fading channel over a Rayleigh distribution

Output: Simulate (Sim) the results for outage probability at the UAV U in transmission block t

for $k = 1$ to the optimal number of subclusters K **do**
  for $i = 1$ to the number $\mathcal{N}_k$ of sensor members within the subcluster $\mathbf{C}_k$ **do**
    Calculate the SINR at the UAV from (8);
    Calculate the achievable maximum instantaneous bit-rate from (9);
    Initialize variable $count \leftarrow 0$;
    for $l = 1$ to 10^6 samples **do**
      if $\left(\min\limits_{S_n^{(i,k)} \in \mathbf{C}_k} \max\limits_{[A_{S_n} \times A_U]} \left\{ R_{S_n^{(i,k)}}(T,N,K) \right\} \geq \mathcal{R} \right)$ **then**
        $count \leftarrow count + 1$;
      end if
    end for
    $OP_{U-x_{S_n}^{(i,k)}}(T,N,K) = 1 - \frac{count}{10^6}$;
  end for
  $OP_U(T,N,K) = \frac{1}{\mathcal{N}_k} \sum\limits_{k=1}^{\mathcal{N}_k} OP_{U-x_{S_n}^{(i,k)}}(T,N,K)$;
end for
return Outage probabilities at UAV $OP_U(T,N,K)$;

Remark 2. *From expression* (16)*, we obtain the outage probability at the UAV over Rayleigh distributions:*

$$OP_{U-x_{S_n^{(i,k)}}}(T,N,K) = \sum_{\psi=0}^{A_{S_n}A_U} \frac{(-1)^{\psi}(A_{S_n}A_U)!}{\psi!(A_{S_n}A_U-\psi)!} \exp\left(-\frac{\psi\gamma}{\rho_{S_n}\sigma_{S_n,U}}\right), \tag{17}$$

where the SINR threshold is given by $\gamma = 2^{2\mathcal{R}} - 1$*. It is important to note that Equation (17) obtains the independent outage probabilities at the UAV. Generally, the outage probability at the UAV is calculated from* $OP_U(T,N,K) = \frac{1}{\mathcal{N}_k} \sum\limits_{i=1}^{\mathcal{N}_k} OP_{U-x_{S_n^{(i,k)}}}(T,N,K)$.

See Appendix C for the proof.

3.2. Outage Probability at the Mobile Base Station

Theorem 2. *The dependent outage event at the mobile base station occurs when the flying relay (FR)-UAV either cannot decode at least the message* $x_{S_n^{(i,k)}} \in \mathbf{X}_k$ *or the mobile base station B cannot decode at least the message* $x_{S_n^{(i,k)}} \in \mathbf{X}_k$ *from the best-received signal* $y_B(T,N,K)$*, given by (11). The outage probability at the mobile base station with an underlying U-assisted multi-input multioutput (MIMO)-NOMA network is therefore expressed as*

$$OP_B(T,K,N) = 1 - \Pr\left\{ \min_{x_{S_n^{(i,k)}} \in \mathbf{X}_k} \max_{[A_{S_n} \times A_U]} \left\{ R_{U-x_{S_n^{(i,k)}}}(T,N,K) \right\} \geq \mathcal{R}, \right.$$
$$\left. \min_{x_{S_n^{(i,k)}} \in \mathbf{X}_k} \max_{[A_U \times A_B]} \left\{ R_{B-x_{S_n^{(i,k)}}}(T,N,K) \right\} \geq \mathcal{R} \right\}. \tag{18}$$

Based on Equation (18), we propose Algorithm 3 to calculate the Monte Carlo simulations for outage probability at the mobile base station for transmission block t over Rayleigh distributions.

Algorithm 3 Calculate the outage probability at the mobile base station from (18) for transmission block t over Rayleigh distributions.

Input: Initialize the parameters as in Table 1 and randomly generate 10^6 samples of each fading channel over a Rayleigh distribution;
Output: Simulate (Sim) the results for outage probability at the mobile base station B;

for $k = 1$ to the optimal number K of the subcluster **do**
  for $i = 1$ to the number of sensors $\mathcal{N}_k$ **do**
    Calculate the SINR at the UAV U from (8);
    Calculate the achievable maximum instantaneous bit-rate at the UAV U from (9);
    Calculate the minimum-maximum instantaneous bit-rate threshold at the UAV U from (9);
    Calculate the SINR at the mobile base station from (13) or (14);
    Calculate the achievable maximum bit-rate at the mobile base station from (15);
    Calculate the achievable minimum-maximum bit-rate at the mobile base station from (15);
    Initialize variable $count \leftarrow 0$;
    for $l = 1$ to 10^6 samples **do**
      if $\left(\min\left\{ \min\limits_{x_{S_n^{(i,k)}} \in \mathbf{X}_k} \max\limits_{[A_S \times A_U]} \left\{ R_{U - x_{S_n^{(i,k)}}}(T,N,K) \right\}, \min\limits_{x_{S_n^{(i,k)}} \in \mathbf{X}_k} \max\limits_{[A_U \times A_B]} \left\{ R_{B - x_{S_n^{(i,k)}}}(T,N,K) \right\} \right\} \geq \mathcal{R} \right)$ **then**
        $count \leftarrow count + 1$;
      end if
    end for
    $OP_{B - x_{S_n}^{(i,k)}}(T,N,K) = 1 - \frac{count}{10^6}$;
  end for
  $OP_B(T,N,K) = \frac{1}{\mathcal{N}_k} \sum\limits_{k=1}^{\mathcal{N}_k} OP_{B - x_{S_n}^{(i,k)}}(T,N,K)$;
end for
return Dependent outage probability at the mobile base station $OP_B(T,N,K)$;

Remark 3. *The outage probability at the mobile base station in transmission block t is given by (18) from Theorem 1 and expressed in novel closed-form as follows:*

$$OP_B^{(i)}(T,N,K) = \max\left\{ \sum_{\psi=0}^{A_S A_U} \frac{(-1)^\psi (A_S A_U)!}{\psi!(A_S A_U - \psi)!} \exp\left(- \frac{\psi\gamma}{\min\left\{\rho_{S_n} \sigma_{S_n,U}\right\}} \right), \sum_{\psi=0}^{A_U A_B} \frac{(-1)^\psi (A_U A_B)!}{\psi!(A_U A_B - \psi)!} \exp\left(- \frac{\psi\gamma}{\beta \rho_U \sigma_{U,B}} \right) \right\}, \tag{19}$$

$$s.t. \quad \beta_i = \alpha_{S_n^{(i,k)}} - \gamma \sum_{j=i+1}^{\mathcal{N}_k} \alpha_{S_n^{(j,k)}}, \tag{20}$$

$$\beta = \min_{i=\{1,\ldots,\mathcal{N}_k\}} \{\beta_i\}, \tag{21}$$

where SINR threshold $\gamma = 2^{2\mathcal{R}} - 1$.

See Appendix D for the proof.

4. Numerical Results and Discussion

In this section, we examine the individual WSN and discuss the results of the study. For the purposes of the analysis and the Monte Carlo simulations, a random number of wireless sensors N was generated and randomly distributed according to the positions il-

lustrated in Figure 1. Unless specified otherwise, we assumed that the mobile base station's position was at coordinate $B(0,0,0)$ and that the UAV's position $U(x,y,1)$ determined by K-means clustering had a fixed altitude at $z = 1$. The number of antennae equipped at the wireless sensors, UAV and mobile base station was $A_{S_n} = A_U = A_B = 2$. The K-means algorithm determined the optimal number of subclusters as $K = 4$. The path-loss exponent factor was $\varepsilon = 4$. The list of pairwise distances from each subcluster centroid to the mobile base station was $d^{A2G}_{\mathbf{C}_1,B} = 0.494$, $d^{A2G}_{\mathbf{C}_2,B} = 0.8788$, $d^{A2G}_{\mathbf{C}_3,B} = 0.9268$ and $d^{A2G}_{\mathbf{C}_4,B} = 1.1620$. The nearest subcluster to the mobile base station was therefore $\mathbf{C}_1$. The UAV selected subcluster $\mathbf{C}_1$ for the global period $T = \{1,5,9,13,\ldots\}$. At global period $T = \{2,6,10,14,\ldots\}$, the UAV then selected the next nearest subcluster to its current subcluster $\mathbf{C_1}$, i.e., subcluster $\mathbf{C_3}$, because distances $d^{A2A}_{\mathbf{C}_1,\mathbf{C}_3} = 0.5005 < d^{A2A}_{\mathbf{C}_1,\mathbf{C}_2} = 0.5602 < d^{A2A}_{\mathbf{C}_1,\mathbf{C}_4} = 0.7195$. At global period $T = \{3,7,11,15,\ldots\}$, the UAV again selected the next nearest subcluster to subcluster $\mathbf{C_3}$, i.e., subcluster $\mathbf{C_4}$, since distances $d^{A2A}_{\mathbf{C}_3,\mathbf{C}_4} = 0.4501 < d^{A2A}_{\mathbf{C}_3,\mathbf{C}_4} = 0.7166$. At global period $T = \{4,8,12,17,\ldots\}$, the UAV selected the next nearest subcluster to subcluster $\mathbf{C_4}$, i.e., subcluster $\mathbf{C_2}$, because the final distances $d^{A2A}_{\mathbf{C}_4,\mathbf{C}_2} = 0.5262$. The UAV thus selected the shortest trajectory $\mathbf{C_1} \rightarrow \mathbf{C}_3 \rightarrow \mathbf{C}_4 \rightarrow \mathbf{C}_2$. Without loss of generality and for simplicity, we assumed that $\lambda_1 = \lambda_2 = \lambda_3 = \frac{t}{3}$ for a single round trip of the UAV and global period $T = \{1,2,3,4\}$.

4.1. Numerical Results

Figure 7a–d plot the outage probabilities at the UAV at the point when it decoded the received signals from wireless sensors in subclusters $\mathbf{C}_1$, $\mathbf{C_3}$, $\mathbf{C_4}$ and $\mathbf{C_2}$, respectively. The bit-rate threshold for all wireless sensors was $\mathcal{R} = 1.5$ bps/Hz. The outage probabilities at the majority of wireless sensors were very similar as SNR $\rho_{S_n} \rightarrow \infty$; however, the graphs in Figure 7a indicate that the outage probability of sensor S_{23} was worse than the outage probability at the other sensors of the same subcluster $\mathbf{C}_1$. Figure 3 indicates that wireless sensor S_{23} was the farthest from the subcluster centroid $\mathbf{C}_1$ ($d^{G2A}_{S^{(7,1)}_{23}} = 1.0479$). The results verified the efficiency of the K-means algorithm. It is important that the bit-rate threshold $\mathcal{R}$ for the wireless sensors was set to $\mathcal{R} = 1.5$ bps/Hz. However, the UAV successfully decoded most of the messages from the wireless sensors, achieving a high outage probability performance (Figure 7). We conclude that the outage probability performance of the majority of wireless sensors in each subcluster was equal since they were evenly distributed around the subcluster's centroid. The Monte Carlo simulations given by (16) were also verified by the analysis results given by (17).

Next, we examined the results for the mobile base station and obtained its outage probability performance at the points when the UAV visited subclusters $\mathbf{C}_1$, $\mathbf{C}_3$, $\mathbf{C}_4$ and $\mathbf{C}_2$ and forwarded the superimposed signals, given by (11), to the base station (Figure 8a–d). The outage probability performance of the mobile base station was poorer than the outage probability performance at the UAV (Figure 7a–d), even though the bit-rate threshold was set to $\mathcal{R} = 0.1$ bps/Hz. This may have been because the UAV was deployed with NOMA and therefore, the sensors were forced to share the power domain to transmit the messages in the superimposed signal. This means that the last member in the subcluster was allocated a very small power allocation factor, given by (12). These power allocation factors are presented in Table 6. Subcluster $\mathbf{C}_3$ contained $\mathcal{N}_3 = 13$ wireless sensors, and the last wireless sensor in $\mathbf{C}_3$ ($S^{(13,3)}_{36}$) was allocated the lowest power allocation factor ($\alpha_{S^{(13,3)}_{36}}$). Therefore, despite a global optimization of the subclusters and the positions of the cluster centroids by the K-means algorithm, the large number of wireless sensors in the subcluster unfortunately led to unsatisfactory results.

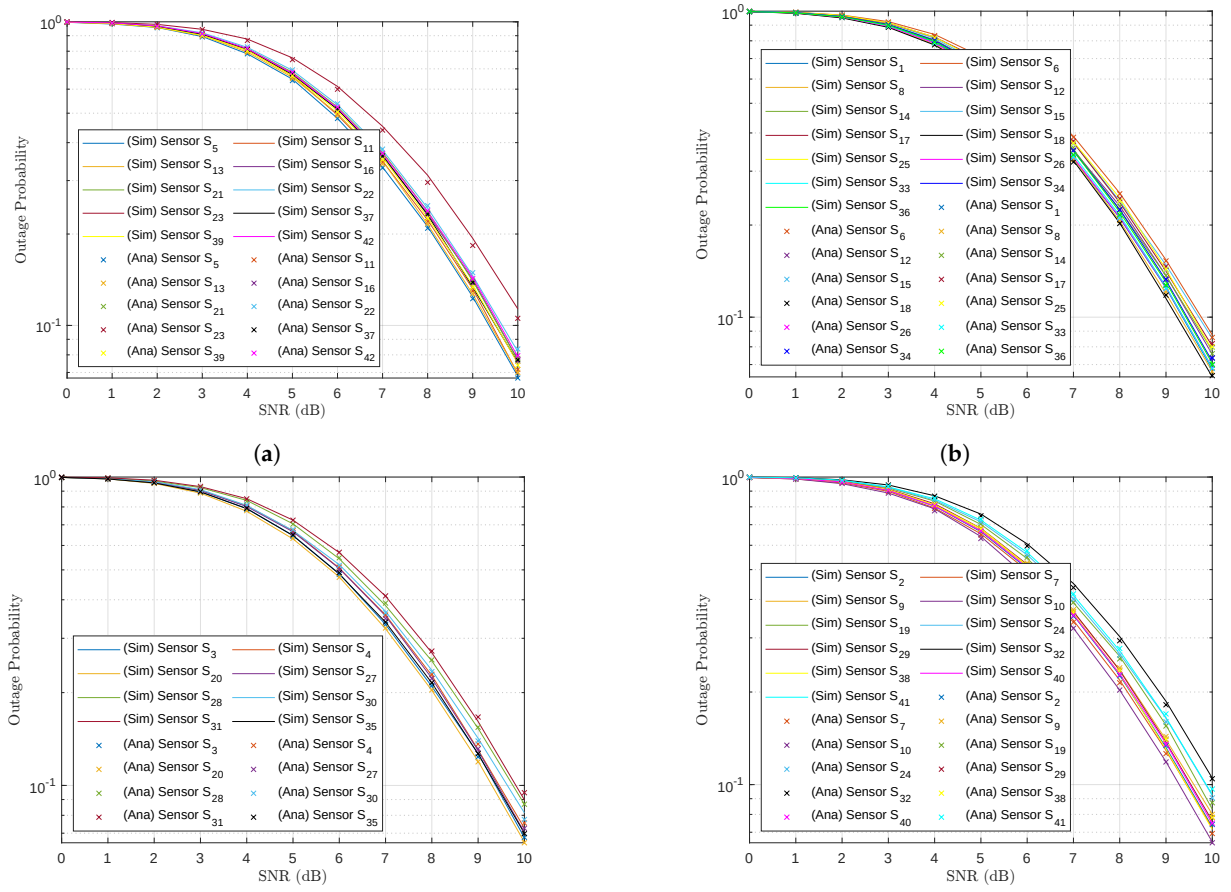


Figure 7. Outage probability at the UAV for the UAV's subcluster trajectory sequence (**a**) $\mathbf{C}_1$, (**b**) $\mathbf{C}_3$, (**c**) $\mathbf{C}_4$ and (**d**) $\mathbf{C}_2$.

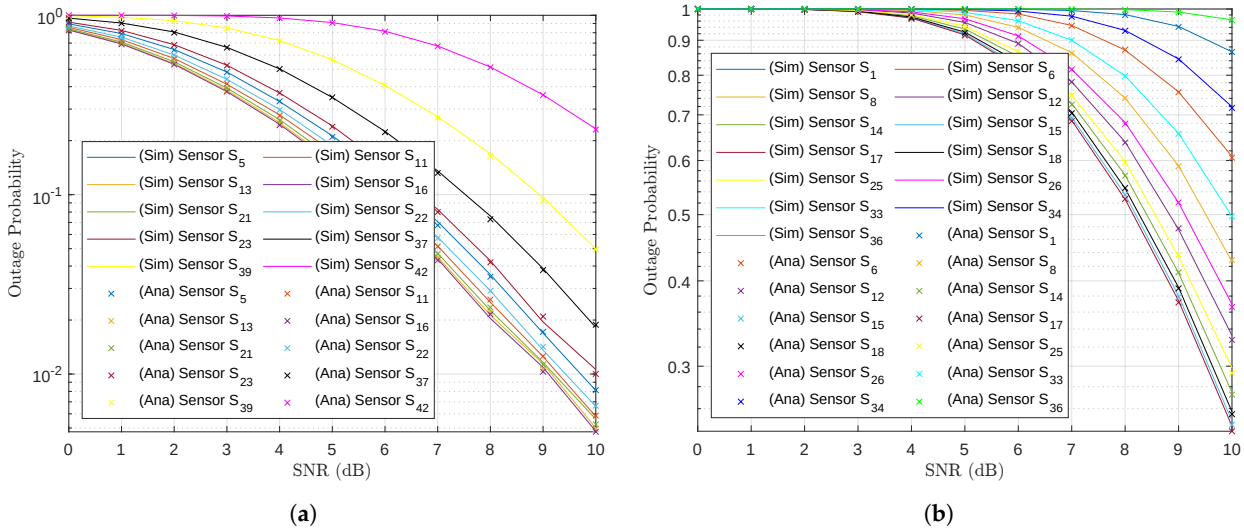
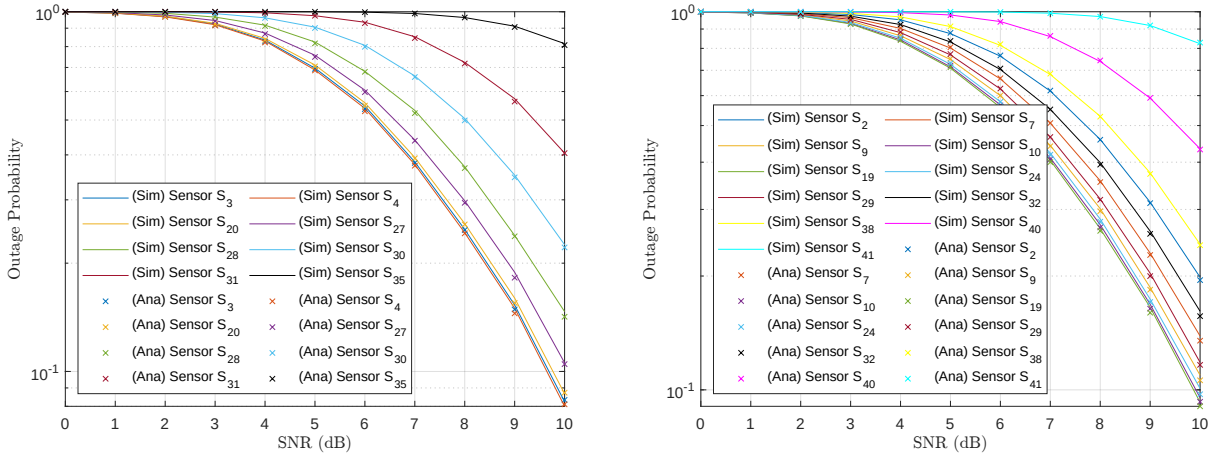


Figure 8. *Cont.*

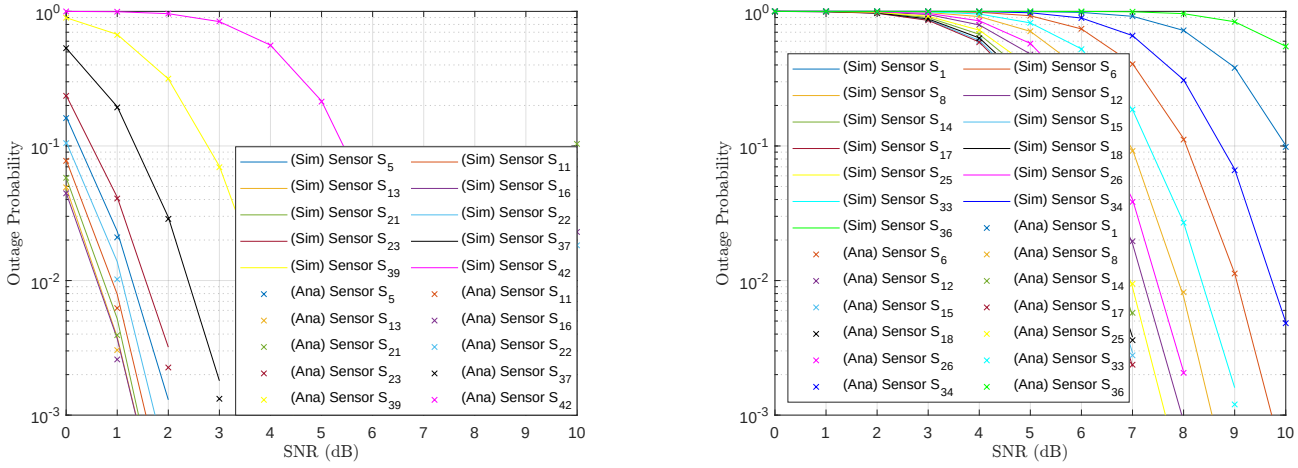


(**c**) (**d**)

Figure 8. Outage probability at the mobile base station for the UAV's subcluster trajectory sequence (**a**) $\mathbf{C}_1$, (**b**) $\mathbf{C}_3$, (**c**) $\mathbf{C}_4$ and (**d**) $\mathbf{C}_2$.

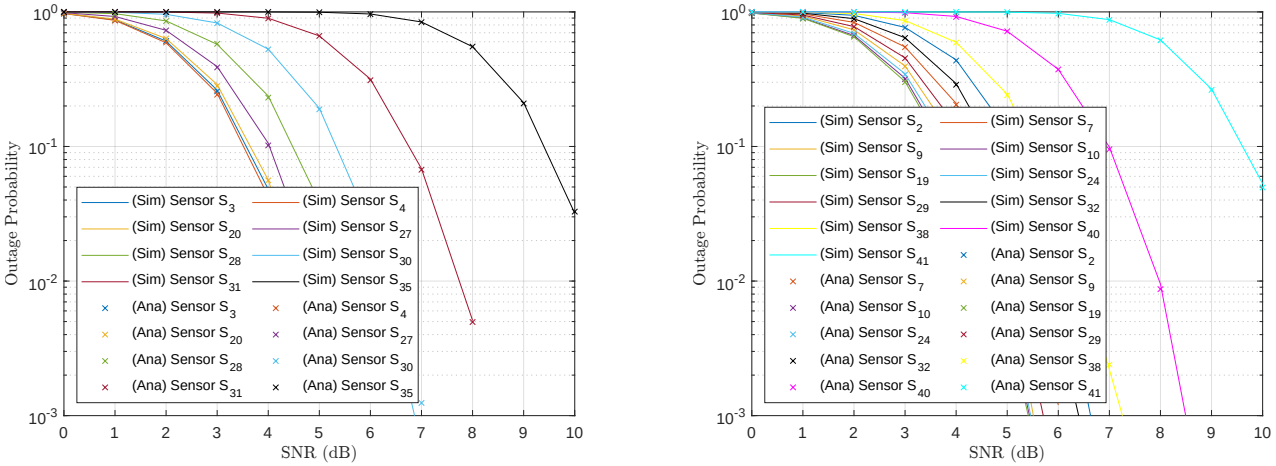
4.2. Discussion

The outage probability performance at the mobile base station was strongly affected by several factors, such as the UAV's transmit power P_U, the distance of the UAV from the mobile base station $d_{U,B}$ and the number $\mathcal{N}_k$ of messages transmitted in the superimposed signals. The UAV was not able to increase the transmit power P_U, however, because of its power limitations. A large number of antennae at both the wireless sensors and the UAV could not be equipped since the constraints for a small size, light weight and low cost did not permit it. It was also not possible to reduce the distance from the UAV to the mobile base station because of obstructions in the terrain. To address these conditions, we equipped a larger number of antennae at the mobile base station, as the mobile base station incorporated a generator and energy was not a significant problem. Therefore, equipping $A_B = 32$ antennae at the mobile base station instead of the same number at the UAV $A_{S_n} = A_U = A_B = 2$ (Figure 8a–d) yielded the results in Figure 9a–d. It is clear that the outage probability improved significantly at the mobile base station for $A_B = 32$ while $A_{S_n} = A_U = 2$. It is also clear that the outage probability performance at the mobile base station when the UAV visited subcluster $\mathbf{C}_1$ improved greatly since this subcluster $\mathbf{C}_1$ was closer than the other subclusters. The outage probabilities at the other subclusters also improved as the SNR $\rho_U \to \infty$.



(**a**) (**b**)

Figure 9. *Cont.*



(**c**) (**d**)

Figure 9. Improved outage probability at the mobile base station equipped with $A_B = 32$ antennae for the UAV trajectory subcluster sequence (**a**) $\mathbf{C}_1$, (**b**) $\mathbf{C}_3$, (**c**) $\mathbf{C}_4$ and (**d**) $\mathbf{C}_2$.

5. Conclusions

This study presented a general WSN containing a randomly distributed number of wireless sensors with three-dimensional Cartesian coordinates. To improve the WSN's performance, we applied a K-means algorithm and gap statistic method to optimize sensor clustering into a number of subclusters K. The UAV's trajectory was calculated with an algorithm which determined the shortest path between the subcluster centroids. The aims of the study were achieved (i.e., flexible deployment, low cost and high reliability) through the effective proposed solutions, and the results were verified with both Monte Carlo simulations and theoretical analysis. Although the study provided some benefits from the application of the K-means algorithm for wireless sensor clustering, some problems still persisted that can be studied in future work. Future studies can investigate the problems with (1) fragmented power resources created by an imbalance in the number of subcluster sensors and (2) some clusters covering a larger geographic area than others as a result of sparsely distributed sensors. As a potential solution, we propose dividing the network into larger clusters when the number of sensors reaches a certain threshold.

Author Contributions: T.-N.T. and T.-L.N. proposed ideas and formulation of overarching research aims, designed of WSN model, applied mathematical to analyze the proposed model, wrote the initial draft. V.T.H. contributed to preparation and presentation of the published work by those from the original research group, specifically critical review, commentary and revision—including pre- and post-publication stages. M.V. contributed to evolution of overarching research goals, management activities to annotate, oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team, acquisition of the financial support for the project leading to this publication. All authors have read and agreed to the published version of the manuscript.

Funding: This research received funding from the Ministry of Education, Youth and Sports under grant Reg. No. SP2021/25 and partially under the Large Infrastructures for Research, Experimental Development and Innovations project Reg. No. LM2018140.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study were randomly generated.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study, the collection, analysis and interpretation of data, writing of the manuscript, or the decision to publish the results.

Abbreviations

A2A	air-to-air
A2G	air-to-ground
AWGN	additive white Gaussian noise
CDF	cumulative distribution function
CSI	channel state information
EH	energy harvesting
FR	flying relay
G2A	ground-to-air
MIMO	multi-input multioutput
NOMA	nonorthogonal multiple access
PDF	probability density function
SIC	successive interference cancellation
SINR	signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
SWIPT	simultaneous wireless information and power transfer
UAV	unmanned aerial vehicle
WPT	wireless power transfer
WSN	wireless sensor network

Appendix A

The probability density function (PDF) and cumulative distribution function (CDF) of the Rayleigh distribution are expressed, respectively, as:

$$f_{|h_{src,des}|^2}(x) = \frac{1}{\sigma_{src,des}} \exp\left(-\frac{x}{\sigma_{src,des}}\right), \tag{A1}$$

and

$$F_{|h_{src,des}|^2}(x) = 1 - \exp\left(-\frac{x}{\sigma_{src,des}}\right), \tag{A2}$$

where $|h_{src,des}|^2$ are random independent variables, i.e., x in (A1) and (A2). In addition, $\sigma_{src,des}$ is the expected channel gain, where $\sigma_{src,des} = E\left[|h_{src,des}|^2\right]$ between the source (*src*) and destination (*des*).

Appendix B

We studied an individual WSN (Figure 1) which contained an optimal number of subclusters $K = 4$ (Figure 2) and wireless sensors allocated according to those subclusters (Figure 3). A UAV travelled the shortest path from each subcluster centroid through all subclusters (Figure 4). The results for the outage probability at the UAV indicated a high performance (Figure 7a–d); however, the outage probability at the mobile base station was comparatively poorer (Figure 8a–d). We hypothesized that the MIMO technique could improve the OP performance and therefore equipped a larger number of antennas at the mobile base station ($A_B = 32$); the results showed a significant improvement in outage probability at the mobile base station (Figure 8a,c,d), except when the mobile base station decoded the message of the last sensor $S_{36}^{(13,3)}$ of subcluster $\mathbf{C}_3$, which demonstrated a worse performance (Figure 8b), probably because the subcluster contained $|\mathbf{C}_3| = 13$ sensors. Another individually generated WSN (Figure A1a) contained an optimal number of $K = 10$ subclusters (Figure A1b) and wireless sensors allocated according to those subclusters (Figure A1c). This configuration also contained an optimal UAV trajectory (Figure A1d). It is important to note that the subclusters depicted in Figure A1c contained fewer sensors than the subclusters in Figure 3. The wireless sensors depicted in Figure A1c were consequently better served than the wireless sensors in Figure 3. However, fewer sensors in each subcluster, and thus a greater number of subclusters, led to a longer UAV flight path.

For example, the results in Figure 4 indicate that the UAV's period to visit all subclusters was $t = 4$; the results in Figure A1d, however, indicate a UAV period of $t = 10$. We conclude that fewer sensors in each subcluster deliver better results, but at the expense of the UAV consuming more energy to travel longer distances.

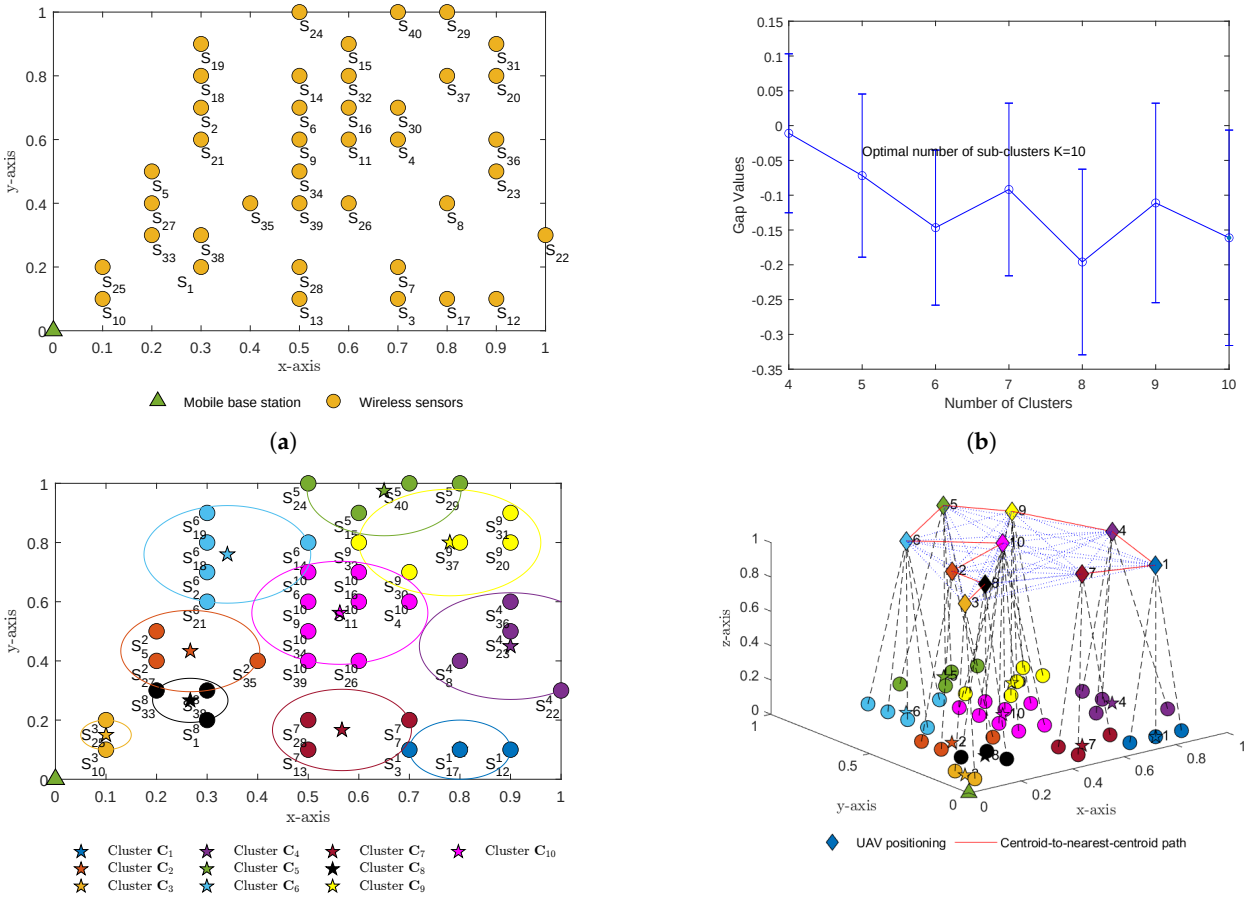


Figure A1. The randomly distributed WSN (**a**), determined optimal number of subclusters K (**b**), division into subclusters (**c**) and centroid-to-next-nearest-centroid trajectory (**d**).

Appendix C

By substituting the SINR given by (8) into (9) and then substituting (9) into Theorem 1, as given (16), we thus obtain the independent outage probability at the UAV:

$$OP_{U-x_{S_n^{(i,k)}}} = 1 - \Pr\left\{\max\left\{|\mathbf{H}_{S_n,U}|^2\right\} \geq \frac{\gamma}{\rho_{S_n}}\right\}. \tag{A3}$$

From the precoding matrix $|\mathbf{H}_{S_n,U}|^2$ given (2) and the PDF given by (A1), we obtain

$$\begin{aligned} OP_{U-x_{S_n^{(i,k)}}} &= 1 - \left(1 - \sum_{\psi=0}^{A_{S_n}A_U} (-1)^{\psi} \binom{A_{S_n}A_U}{\psi} \int_{\gamma/\rho_{S_n}}^{+\infty} \frac{1}{\sigma_{S_n,U}} \exp\left(-\frac{\psi x}{\sigma_{S_n,U}}\right) dx\right) \\ &= \sum_{\psi=0}^{A_{S_n}A_U} \frac{(-1)^{\psi}(A_{S_n}A_U)!}{\psi!(A_{S_n}A_U - \psi)!} \exp\left(-\frac{\psi\gamma}{\rho_{S_n}\sigma_{S_n,U}}\right). \end{aligned} \tag{A4}$$

Note that Equation (A4) evaluates the independent outage probability at the UAV at the point when the UAV decodes the received signal from wireless sensor $S_n^{(i,k)}$ in

subcluster $\mathbf{C}_k$ unsuccessfully. The outage probability at the UAV is generally expressed as $OP_U(T,N,K) = \frac{1}{\mathcal{N}_k}\sum_{k=1}^{\mathcal{N}_k} OP_{U-x_{S_n}^{(i,k)}}(T,N,K)$.

Appendix D

Expression (18) is rewritten as follows:

$$\begin{aligned} OP_B(T,K,N) = 1 - \Pr\Bigg\{ & \min_{x_{S_n^{(i,k)}} \in \mathbf{X}_k} \max_{[A_{S_n} \times A_U]} \left\{ R_{U-x_{S_n^{(i,k)}}}(T,N,K) \right\} \geq \mathcal{R}, \\ & \min_{x_{S_n^{(i,k)}} \in \mathbf{X}_k} \max_{[A_U \times A_B]} \left\{ R_{B-x_{S_n^{(i,k)}}}(T,N,K) \right\} \geq \mathcal{R} \Bigg\} \\ = \max\Bigg\{ & 1 - \Pr\left\{ \min_{x_{S_n^{(i,k)}} \in \mathbf{X}_k} \max_{[A_{S_n} \times A_U]} \left\{ R_{U-x_{S_n^{(i,k)}}}(T,N,K) \right\} \geq \mathcal{R} \right\}, \\ & 1 - \Pr\left\{ \min_{x_{S_n^{(i,k)}} \in \mathbf{X}_k} \max_{[A_U \times A_B]} \left\{ R_{B-x_{S_n^{(i,k)}}}(T,N,K) \right\} \geq \mathcal{R} \right\} \Bigg\}. \end{aligned} \tag{A5}$$

Let $\vartheta = 1 - \Pr\left\{ \min_{x_{S_n^{(i,k)}} \in \mathbf{X}_k} \max_{[A_{S_n} \times A_U]} \left\{ R_{U-x_{S_n^{(i,k)}}}(T,N,K) \right\} \geq \mathcal{R} \right\}$. By substituting the SINR given by (8) into (9) and then substituting (9) into ϑ, we obtain

$$\begin{aligned} \vartheta &= 1 - \left(1 - \min_{i=\{1,\ldots,\mathcal{N}_k\}} \left\{ \sum_{\psi=0}^{A_{S_n}A_U} (-1)^{\psi} \binom{A_{S_n}A_U}{\psi} \int_{\gamma/\rho_{s_n}}^{+\infty} \frac{1}{\sigma_{S_n^{(i,k)},U}} \exp\left(-\frac{\psi x}{\sigma_{S_n^{(i,k)},U}} \right) dx \right\} \right) \\ &= \min_{i=\{1,\ldots,\mathcal{N}_k\}} \left\{ \sum_{\psi=0}^{A_{S_n}A_U} \frac{(-1)^{\psi}(A_{S_n}A_U)!}{\psi!(A_{S_n}A_U-\psi)!} \int_{\gamma/\rho_{s_n}}^{+\infty} \frac{1}{\sigma_{S_n^{(i,k)},U}} \exp\left(-\frac{\psi x}{\sigma_{S_n^{(i,k)},U}} \right) dx \right\} \\ &= \sum_{\psi=0}^{A_{S_n}A_U} \frac{(-1)^{\psi}(A_{S_n}A_U)!}{\psi!(A_{S_n}A_U-\psi)!} \exp\left(-\frac{\psi\gamma}{\rho_{s_n} \min\limits_{i=\{1,\ldots,\mathcal{N}_k\}} \left\{ \sigma_{S_n^{(i,k)},U} \right\}} \right). \end{aligned} \tag{A6}$$

Similarly, let $\xi = 1 - \Pr\left\{ \min_{x_{S_n^{(i,k)}} \in \mathbf{X}_k} \max_{[A_U \times A_B]} \left\{ R_{B-x_{S_n^{(i,k)}}}(T,N,K) \right\} \geq \mathcal{R} \right\}$. By substituting the SINR given by (13) or (14) into (15) and then substituting (15) into ξ, we obtain

$$\begin{aligned}\xi &= 1 - \Pr\left\{ \min_{x_{S_n^{(i,k)}} \in \mathbf{X}_k} \max_{[A_U \times A_B]} \left\{ R_{B - x_{S_n^{(i,k)}}}(T,N,K) \right\} \geq \mathcal{R} \right\} \\ &= 1 - \left(1 - \min_{i=\{1,\ldots,\mathcal{N}_k\}} \left\{ \sum_{\psi=0}^{A_U A_B} (-1)^{\psi} \binom{A_U A_B}{\psi} \int_{\gamma / \rho_U \left(\alpha_{S_n^{(i,k)}} - \gamma \sum_{j=i+1}^{\mathcal{N}_k} \alpha_{S_n^{(j,k)}} \right)}^{+\infty} \frac{1}{\sigma_{U,B}} \exp\left(-\frac{\psi x}{\sigma_{U,B}} \right) dx \right\} \right) \\ &= \sum_{\psi=0}^{A_U A_B} \frac{(-1)^{\psi} (A_U A_B)!}{\psi! (A_U A_B - \psi)!} \exp\left(- \frac{\psi \gamma}{\rho_U \min\limits_{i=\{1,\ldots,\mathcal{N}_k\}} \left\{ \alpha_{S_n^{(i,k)}} - \gamma \sum\limits_{j=i+1}^{\mathcal{N}_k} \alpha_{S_n^{(j,k)}} \right\} \sigma_{U,B}} \right). \end{aligned} \tag{A7}$$

From expression (A5), we conclude that the outage probability at the mobile base station is independent since it belongs to the outage probability at UAV. The study verified that $\vartheta < \xi$ and therefore, the outage probability at the mobile base station given by (A5) refers to $OP_B(T,N,K) = \max\{\vartheta, \xi\} = \xi$. Observing the individual WSN model depicted in Figure 4, a possible explanation is that the UAV was close to the wireless sensors, which thus strongly owned the channel state information (CSI) $h_{S_n,U}$, even though they simply transmitted their own messages x_{S_n} under their own power domain P_{S_n}. However, the UAV travelled a long distance since the subcluster was far from the mobile base station; therefore, it forwarded messages in the superimposed signal by sharing the power domain P_U, and thus $OP_B(T,N,K) = \max\{\vartheta, \xi\} = \xi$. To improve $OP_B(T,N,K)$, both ϑ and ξ must be improved. To improve ϑ, the number of antennae at the wireless sensors or their transmit power must be increased. However, wireless sensors have certain constraints, such as having a low cost and low power. These solutions are therefore not practical or even obtainable. To address these conditions, we attempted to improve ξ by increasing the number of antennae at the mobile base station ($A_B = 32$). The outage probabilities at the mobile base station improved (Figure 9a–d) over the previous results (Figure 8a–d).

References

1. Gong, J.; Chang, T.H.; Shen, C.; Chen, X. Flight Time Minimization of UAV for Data Collection over Wireless Sensor Networks. *IEEE J. Sel. Areas Commun.* **2018**, *36*, 1942–1954. [CrossRef]
2. Li, J.; Zhao, H.; Wang, H.; Gu, F.; Wei, J.; Yin, H.; Ren, B. Joint Optimization on Trajectory, Altitude, Velocity, and Link Scheduling for Minimum Mission Time in UAV-Aided Data Collection. *IEEE Int. Things J.* **2020**, *7*, 1464–1475. [CrossRef]
3. Zhan, C.; Zeng, Y.; Zhang, R. Energy-Efficient Data Collection in UAV Enabled Wireless Sensor Network. *IEEE Wirel. Commun. Lett.* **2018**, *7*, 328–331. [CrossRef]
4. Zhan, C.; Zeng, Y. Completion Time Minimization for Multi-UAV-Enabled Data Collection. *IEEE Trans. Wirel. Commun.* **2019**, *18*, 4859–4872. [CrossRef]
5. Wang, Z.; Liu, R.; Liu, Q.; Thompson, J.S.; Kadoch, M. Energy-Efficient Data Collection and Device Positioning in UAV-Assisted IoT. *IEEE Int. Things J.* **2020**, *7*, 1122–1139. [CrossRef]
6. Kong, P.Y. Distributed Sensor Clustering Using Artificial Neural Network with Local Information. *IEEE Int. Things J.* **2022**, *9*, 21851–21861. [CrossRef]
7. Ur Rahman, S.; Kim, G.H.; Cho, Y.Z.; Khan, A. Positioning of UAVs for throughput maximization in software-defined disaster area UAV communication networks. *J. Commun. Netw.* **2018**, *20*, 452–463. [CrossRef]
8. Heinzelman, W.; Chandrakasan, A.; Balakrishnan, H. An application-specific protocol architecture for wireless microsensor networks. *IEEE Trans. Wirel. Commun.* **2002**, *1*, 660–670. [CrossRef]
9. Dargie, W.; Wen, J. A Simple Clustering Strategy for Wireless Sensor Networks. *IEEE Sens. Lett.* **2020**, *4*, 1–4. [CrossRef]
10. Zhang, S.; Zhang, H.; He, Q.; Bian, K.; Song, L. Joint Trajectory and Power Optimization for UAV Relay Networks. *IEEE Commun. Lett.* **2018**, *22*, 161–164. [CrossRef]

11. Jayakody, D.N.K.; Thompson, J.; Chatzinotas, S.; Durrani, S. *Wireless Information and Power Transfer: A New Paradigm for Green Communications*; Springer: Berlin/Heidelberg, Germany, 2017.
12. Zhang, R.; Ho, C.K. MIMO Broadcasting for Simultaneous Wireless Information and Power Transfer. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 1989–2001. [CrossRef]
13. Zhou, X.; Zhang, R.; Ho, C.K. Wireless Information and Power Transfer: Architecture Design and Rate-Energy Tradeoff. *IEEE Trans. Commun.* **2013**, *61*, 4754–4767. [CrossRef]
14. Tran, T.N.; Voznak, M.; Fazio, P.; Ho, V.C. Emerging cooperative MIMO-NOMA networks combining TAS and SWIPT protocols assisted by an AF-VG relaying protocol with instantaneous amplifying factor maximization. *AEU-Int. J. Electron. Commun.* **2021**, *135*, 153695. [CrossRef]
15. Tran, T.N.; Vo, T.P.; Fazio, P.; Voznak, M. SWIPT model adopting a PS framework to aid IoT networks inspired by the emerging cooperative NOMA technique. *IEEE Access* **2021**, *9*, 61489–61512. [CrossRef]
16. Perera, T.D.P.; Jayakody, D.N.K. Analysis of time-switching and power-splitting protocols in wireless-powered cooperative communication system. *Phys. Commun.* **2018**, *31*, 141–151. [CrossRef]
17. Ding, Z.; Yang, Z.; Fan, P.; Poor, H.V. On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. *IEEE Signal Process. Lett.* **2014**, *21*, 1501–1505. [CrossRef]
18. Timotheou, S.; Krikidis, I. Fairness for non-orthogonal multiple access in 5G systems. *IEEE Signal Process. Lett.* **2015**, *22*, 1647–1651. [CrossRef]
19. Xiao, Y.; Hao, L.; Ma, Z.; Ding, Z.; Zhang, Z.; Fan, P. Forwarding strategy selection in dual-hop NOMA relaying systems. *IEEE Commun. Lett.* **2018**, *22*, 1644–1647. [CrossRef]
20. Tang, X.; An, K.; Guo, K.; Wang, S.; Wang, X.; Li, J.; Zhou, F. On the performance of two-way multiple relay non-orthogonal multiple access-based networks with hardware impairments. *IEEE Access* **2019**, *7*, 128896–128909. [CrossRef]
21. Tran, T.N.; Voznak, M. Adaptive multiple access assists multiple users over multiple-input-multiple-output non-orthogonal multiple access wireless networks. *Int. J. Commun. Syst.* **2021**, *34*, e4803. [CrossRef]
22. Omeke, K.G.; Mollel, M.S.; Ozturk, M.; Ansari, S.; Zhang, L.; Abbasi, Q.H.; Imran, M.A. DEKCS: A Dynamic Clustering Protocol to Prolong Underwater Sensor Networks. *IEEE Sens. J.* **2021**, *21*, 9457–9464. [CrossRef]
23. Tibshirani, R.; Walther, G.; Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B* **2001**, *63*, 411–423. [CrossRef]
24. Wang, Y.; Chen, M.; Pan, C.; Wang, K.; Pan, Y. Joint Optimization of UAV Trajectory and Sensor Uploading Powers for UAV-Assisted Data Collection in Wireless Sensor Networks. *IEEE Int. Things J.* **2022**, *9*, 11214–11226. [CrossRef]
25. Liu, K.; Zheng, J. UAV Trajectory Optimization for Time-Constrained Data Collection in UAV-Enabled Environmental Monitoring Systems. *IEEE Int. Things J.* **2022**, *9*, 24300–24314. [CrossRef]
26. Ma, Y.; Tang, Y.; Tao, J.; Zhang, D.; Tao, S.; Li, W. Energy-Efficient Transmit Power And Straight Trajectory Optimization In Uav-Aided Wireless Sensor Networks. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring), Antwerp, Belgium, 25–28 May 2020; pp. 1–7. [CrossRef]
27. Yuan, X.; Yang, T.; Hu, Y.; Xu, J.; Schmeink, A. Trajectory Design for UAV-Enabled Multiuser Wireless Power Transfer with Nonlinear Energy Harvesting. *IEEE Trans. Wirel. Commun.* **2021**, *20*, 1105–1121. [CrossRef]
28. Li, B.; Qi, X.; Yu, B.; Liu, L. Trajectory Planning for UAV Based on Improved ACO Algorithm. *IEEE Access* **2020**, *8*, 2995–3006. [CrossRef]
29. Wu, Q.; Zeng, Y.; Zhang, R. Joint Trajectory and Communication Design for Multi-UAV Enabled Wireless Networks. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 2109–2121. [CrossRef]
30. Ji, J.; Zhu, K.; Niyato, D.; Wang, R. Probabilistic Cache Placement in UAV-Assisted Networks with D2D Connections: Performance Analysis and Trajectory Optimization. *IEEE Trans. Commun.* **2020**, *68*, 6331–6345. [CrossRef]
31. Jafari, B.; Saeedi, H.; Enayati, S.; Pishro-Nik, H. Energy-Optimized Path Planning for Moving Aerial Base Stations: A Non User-Oriented Framework. *IEEE Commun. Lett.* **2022**, *26*, 672–676. [CrossRef]
32. Vanegas, G.; Armesto, L.; Girbés-Juan, V.; Pérez, J. Smooth Three-Dimensional Route Planning for Fixed-Wing Unmanned Aerial Vehicles with Double Continuous Curvature. *IEEE Access* **2022**, *10*, 94262–94272. [CrossRef]
33. Tran, T.N.; Nguyen, T.L.; Voznak, M. Approaching K-Means for Multiantenna UAV Positioning in Combination with a Max-SIC-Min-Rate Framework to Enable Aerial IoT Networks. *IEEE Access* **2022**, *10*, 115157–115178. [CrossRef]
34. Tran, T.N.; Voznak, M. On secure system performance over SISO, MISO and MIMO-NOMA wireless networks equipped a multiple antenna based on TAS protocol. *EURASIP J. Wirel. Commun. Netw.* **2020**, *2020*, 11. [CrossRef]