# Energy-Optimal Data Collection for WSN-based Agricultural Monitoring with UAV: A Clustering Compressed Sampling Approach

*Abstract*—In this paper, we propose a hierarchical data collection scheme, towards the realization of unmanned aerial vehicles (UAV)-aided industrial wireless sensor networks (IWSN). The particular application is that of agricultural monitoring. For that, we propose the use of hybrid compressed sampling (CS) through exact and greedy approaches. With the exact approach – to model the energy-optimal formulation – it is utilized an improved linear programming formulation of the minimum cost flow problem. The greedy approach is based on a proposed *balance factor* parameter, consisting of data sparsity and the distance from the cluster head to the normal nodes. To improve the node clustering efficiency, a hierarchical data collection scheme is implemented, by which, the nodes in different layers are adaptively clustered, and the UAV can be scheduled to perform energy-efficient data collection. Simulation results show that our method can effectively collect the data and plan the path for the UAV at a low energy cost.

*Index Terms*—artificial intelligence, unmanned aerial vehicles (UAV), industrial wireless sensor network (IWSN), agricultural monitoring system, intelligent signal processing.

## I. INTRODUCTION

**T**HE agricultural industry plays a vital role in global economy growth and contributes to the employment of significant sections of the population. According to a survey from the United Nations-Food and Agriculture Organisations, worldwide food production will increase by 70% in 2050 – due to the human population expansion [1]. However, the decline in the quality of arable land, freshwater resources shortage, environmental degradation and natural disasters have exacerbated the already existing challenges in the development of agriculture. Simultaneously, our natural environment is undergoing unpredictable dramatic changes, making the profitability of future crop seasons very challenging to forecast; agricultural outputs are highly dependent on the soil condition, water quality, temperature, etc [2].

New technology advances in robotics and *artificial intelligence* open new ways to attempt to tackle the shortcomings in the prediction of complex systems – by collecting and analysing multi-sensor data with networks of sensors and *intelligent signal processing* [3], [4]. In open environments such as agricultural fields, unmanned aerial vehicles (UAV) present a feasible and cost effective opportunity to implement intelligent data collection [5].

In addition, by way of parallel rapid developments in information and communications technology (ICT), the Internet of Things (IoT) – especially industrial IoT (IIoT) technologies – has emerged as a catalyst towards the realization of the Industry 4.0 framework [6], [7]. The IIoT technology – e.g., industrial wireless sensor networks (IWSN) – has been widely promoted in various fields, ranging from smart mine, smart city, smart ocean, smart factory, and even in the area of intelligent agriculture [8], [9]. What characterizes an IWSN is its ability to coordinate all the units towards a common goal; that of maximizing the amount of data and its consistency. For agricultural industry purposes, an IWSN essentially consists on a multiplicity of sensors that collect information, such as soil components, temperature, PH, etc. The data are then transferred to a centralized computing unit (an edge computing platform), so that agriculture management policies can be effectively determined [10].

Thus UAVs open a new path for gathering the data from the sensors in IWSN, upgrading the IWSN to UAV-aided IWSN, i.e., UAV-IWSN. Instead of recording data based on active data routing, UAV units can fly through each sensing field and collect data from the sensors, thus minimizing the infrastructure needed for implementing intelligent farming methods. UAV-IWSNs are aimed to be deployed over the extension of latifundia, where the dedicated sensors are far from each other, but still act as coordinated data sensing components. The UAVs perform specific data collection missions by flying over planned paths. Utilizing UAV to collect data benefits from the following advantages: i) compared with the vehicles on the ground, UAVs can collect data at some special places where the normal collector cannot reach to [11]; ii) UAVs typically displace faster than autonomous ground vehicles, and can return to base in case of adverse weather promptly, so that the expensive equipment is not damaged. iii) UAVs are normally equipped with a high-performance computing unit and large battery capacity, and can sometimes provide the ability of lightweight data computation. iv) UAVs can be easily adapted for tasks they were not designed for, such as the detecting the incidence of wildfires and rural crime. v) UAVs do not interfere with the livestock.

A challenge of UAV-IWSN are the deployment costs. As a consequence not all sensor units may be equipped with all the available sensing and computing capacities. It may be more beneficial to distribute the capabilities optimally among the units, so that some units may have high-performance computing or battery-UAV communication, while others just perform the task of basic data collectors [12]. Hence, the concept of node clustering can be employed to improve the applicability and energy efficiency of UAV-IWSNs, where the cluster head (CH) is in charge of gathering the sensed data, and then transferring them to the UAV – while the other units are only performing data sensing tasks. Further, to improve the data gathering performance, the signal processing efficiency needs to be considered [8].

Succinctly, in order to perform energy-efficient data collection in UAV-IWSN, an intelligent data collection scheme integrating a node clustering algorithm, intelligent signal processing methods together with path planning policies is advantageous. With this goal in mind, in this article, the compressed sampling (CS) technique [13], [14] is employed, and we propose a hierarchical data collection scheme for an UAV-IWSN based agricultural monitoring system. Our contribution can be summarized as follows:

i) We introduce the concept of cluster-based CS and propose a hybrid CS-based node clustering model. Based on the model, we propose a hierarchical data collection scheme integrating an exact approach and a greedy approach;

ii) For the exact approach, an improved version of the minimum cost flow problem – to model the energy-optimality – is utilized. This can be formulated as an improved linear programming (LP) method;

iii) In the greedy approach, we introduce a *balance factor* parameter, accounting for both, data sparsity and the distance from the cluster head to the sensing nodes;

iv) The proposed scheme is tuned through an ant colony optimization (ACO) algorithm-based path planning policy.

The remainder of this paper is organized as follows. Related work is surveyed in Sec. II. Sec. III introduces the problem statement of this work. Sec. IV presents our proposed approach, while Sec. V shows some evaluation results for demonstrating and validating the scheme. In Sec. VI conclusions are drawn.

## II. RELATED WORK

In this section, a review of the current research progress in IoT and IIoT-based agricultural monitoring systems is provided. As stated, the concept of IIoT is becoming ubiquitous in a wide range of industrial environments. For agricultural applications, the motivation towards using IIoT technology is the need to embed industrial sensors into farming spaces – to monitor the irrigating system, crop, weather, soil information. For instance, in [15] the authors used industrial WSN technology to design crop field monitoring systems; by considering three categories of industrial sensors that survey temperature, humidity and crop images. By analyzing these parameters, the informed decisions for maintaining the crop's health can be made. Their proposal also contains a remote monitoring/controlling platform, more suitable for areas where the water is scarce. In [16], the authors proposed an intelligent agriculture monitoring system based on GSM, to enable automatic deployment of some important agriculture missions. By way of a GSM module, the centralized data analysis system can gather the sensed data from the sensors, and transmit the analysis result to the irrigating system to perform automatic irrigation. Their proposal is very significant to the design of cellular network-based or online agricultural monitoring systems. In [17], the authors focused on a real-time and clock-shared rainfall agricultural monitoring system for IWSN, which also served the purpose of protecting the crop from animal attacks.

In [18], novel sensors for monitoring the leaf area index (LAI) of the crops were introduced. Due to the advantages of energy-efficient, scaled-down and low cost sensor technology, the theoretical principles behind the IEEE 802.15.4 protocol towards IWSN can be deployed in a realistic scenario, such as a large-scale cornfield. In particular, the authors showcased a project for designing a feasible, low-cost modification of commercial off-the-shelf photosynthetically active radiation (PAR) sensors, which can be deployed in the LAI monitoring system. In [19], the authors propose a clustering routing algorithm based on Dijkstra algorithm (C.R.D.A). We note that, in the clustering phase, the C.R.D.A divides the clusters by following the ELBOW method, and utilizes K-means to allocate each node to the cluster, according to the length from the normal nodes to the cluster heads, through a greedy approach. Their work gives a compressing clustering algorithm from the aspect of relative distance between the normal nodes and the cluster heads, ignoring the transmitted data among the nodes. In [20], to measure the long-term evolutionary trend of daily average soil temperature within a specified period, the authors introduced an IWSN-based monitoring system to acquire the spatio-temporal variation of daily soil temperature. However the maximum monitoring duration or the techniques to improve the lifetime of the monitoring network were not highlighted. In [21], the authors respectively utilize RFID, QR code technology, and cameras to build an IWSN-based irrigation facilities management system. Their proposal can provide extremely rewarding information to assist the facilities manager with irrigation policy determination. In [22], the authors focused on the analysis of the agronomic variables of the cassava crops, and proposed an IWSN technology-based agricultural system. In particular, each sensor in IWSN was equipped with advanced soil moisture and temperature modules, and the network performance is capable of automatic checking by evaluating the received signal strength indicator (RSSI), link quality indicator (LQI), and network convergence time. In [23] low-altitude remote sensing, bio-sensors, chlorophyll meter, multispectral/near-infrared camera and fluorescence spectrometer were used. to build an IWSN-based monitoring system to study the relative canopy chlorophyll content in citrus orchards – by generating spatial distribution maps. It should be noted about that the authors utilized the univariate and multiple linear regressions together with the partial least squares method to propose a leaf *soil and plant analyzer development* (SPAD) value predictive model. In [24], the authors focus on the energy consumption efficiency issue and propose an adaptive energy consumption model for the IWSN-based agricultural monitoring system. Their proposal can be referred to compute the optimal number of the nodes in the network, while the connectivity and coverage of the entire network are guaranteed.

Recently, some scholars have proposed the use of UAV to assist with the agricultural information collection. For instance, UAVs can be deployed to obtain spectral or infrared images of plants to perform farmland of large scale monitoring or hazard/irrigation prediction. In [25], the authors utilize UAVs equipped with infrared and visible sensors to perform

remote water stress monitoring, and compare its monitoring information with the data collected from the sensors on the ground, such that joint analysis result can be achieved. In [26], the authors propose a agricultural monitoring system integrating IWSN and solar powered UAV to monitor $CO_2$ and $CO_4$ at the greenhouse. The UAV measures the data by hovering above the monitoring field, while the sensors of IWSN are collecting the data on the ground, leading to a 3D monitoring environment.

To further extend the applicability of UAV-IWSN based agricultural monitoring systems, we introduce the data-driven node clustering policy especially from the aspect of data CS, enabling an UAV-assisted data collection framework.

## III. PROBLEM STATEMENT
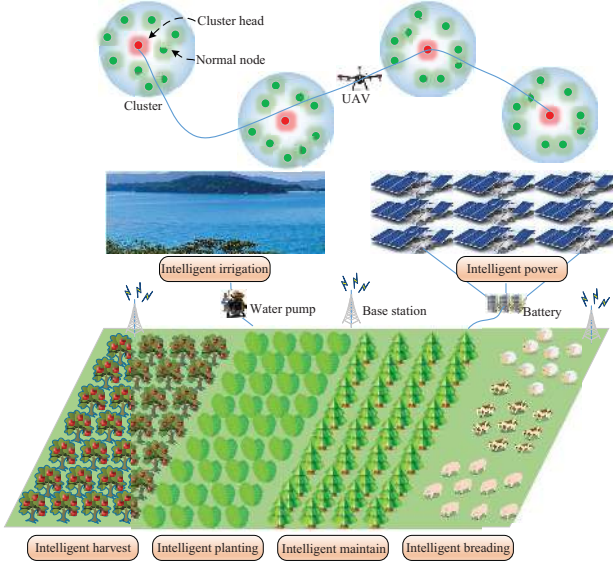
### A. Network Model



Fig. 1: Conceptual diagram of the studied UAV-IWSN scheme.

To enable intelligent, automatic, and informative industrial agriculture, we adopt an IIoT framework perspective to introduce a scheme for an UAV-IWSN based intelligent farm. In Fig. 1, the proposed UAV-IWSN is shown, consisting of an UAV and a series of wireless monitoring sensors divided by different clusters. The sensors are deployed for different agricultural monitoring purposes, e.g., soil moisture/acidity, temperature, air components, energy status, etc. Instead of using active data routing to collect and monitor the data at the base station, a pre-configured rechargeable UAV is scheduled to fly through each sensing field. Then, the UAV uploads the data to the data center, where the data are automatically computed/analyzed, and the agricultural management policy is adaptively adjusted, e.g., irrigation levels and adequate feed nutrient content.

As discussed, in the proposed network architecture, we divide the sensors into different clusters, each of which is managed by a CH. The $q$ normal nodes of the UAV-IWSN are denoted as $v_1$, $v_2$,..., $v_q$, while the $p$ CHs as $ch_1$, $ch_2$,...,$ch_p$

– which are related to clusters the $\mathcal{C}_1$, $\mathcal{C}_2$,..., $\mathcal{C}_p$ respectively. We utilize the notation

$$S(v_i) = \mathcal{C}_x, \qquad (1)$$

to indicate that the normal node $v_i$ belongs to the cluster $\mathcal{C}_x$. And

$$W(\mathcal{C}_x) = \{v_i | S(v_i) = \mathcal{C}_x\}, \qquad (2)$$

to denote all the nodes in the cluster $\mathcal{C}_x$ or the node set of cluster $\mathcal{C}_x$. To summarize, the architecture for UAV-IWSN presents the following properties:

i) The CH is equipped with a high-performance battery, data storing, communication components, and is more expensive than the normal node. As a result, in the proposed network architecture, the number of CH accounts for a very small percentage;

ii) Each normal node is assigned to a dedicated cluster for performing data collection, under the management of the CH. To improve the energy consumption efficiency, all the normal nodes forward their collected data to the CHs, instead of communicating directly with the UAV;

iii) The CHs store the gathered data and wait for the data uploading scheduling. The UAV performs non-real-time data collection from the CHs, based on a pre-configured data collection policy;

iv) The features (e.g., the location, the CH of each node) of each cluster are kept constant once the clusters are determined and deployed.

In a cluster $\mathcal{C}_x$, the communication range among the sensors is imposed by a threshold $r$ (determined when both of the energy consumption, packet loss rate, etc. are concurrently taken into account). And, the data can be forwarded to the $ch_x$ by utilizing multi-hop routing algorithm based on the spanning tree $G_x(V_x, E_x)$ derived from $\mathcal{C}_x$. In $G_x(V_x, E_x)$, $V_x = W(\mathcal{C}_x) \cup ch_x$, and $e_{ij} =< v_i, v_j >\in E_x$ is a corresponding edge if the distance between $v_i$ and $v_j$ does not exceed $r$.

### B. Energy Consumption Model

We refer to the energy consumption model in [27] and respectively use Eq. 3 and Eq. 4 to compute the energy consumption, when $n$ bits of data is sent and received.

$$C_s(n, r) = \begin{cases} n(\theta_{fs} r^2 + C_a) & r < \mathcal{R} \\ n(\theta_{mp} r^4 + C_a) & r \geq \mathcal{R}, \end{cases} \qquad (3)$$

$$C_r(n) = nC_a. \qquad (4)$$

In Eq. 3, both $\theta_{fs}$ and $\theta_{mp}$ represent the sending power coefficients; $\theta_{fs} r^2$ and $\theta_{mp} r^4$ respectively denote the consumed energy by the transmit amplifier for sending one bit of data within communication range $r$, when the transmit amplifier works in different modes. We notice that this mainly depends on both the communication range and the accepted bite error bit; $\mathcal{R}$ denotes the distance threshold in the free space model. $C_a$, both Eq. 3 and Eq. 4, denotes the consumed energy to activate the sending or receiving circuit. Both of

the aforementioned variables are determined based on the real features of the physical electronic components.

### C. Problem Definition

Given an UAV-IWSN, represented as graph $G(V, E)$, that is deployed in an open area of an intelligent farm, where $V$ is the set of nodes (including both normal nodes and CHs) and $E$ is the potential edge set. Based on the aforementioned network model in Sec. III-A, the proposed *optimal data collection* (ODC) problem aims at defining a suitable clustering method, together with a signal acquisition approach, to minimize the total energy consumption. Then, the ODC problem also requires to find a path planning algorithm to schedule the data gathering path for the UAV, such that the energy consumption for UAV is minimized.

The studied ODC problem is very common in real industrial agricultural monitoring systems, due to the wide deployment of the monitoring sensors, the un-rechargeable features of the sensors, etc.
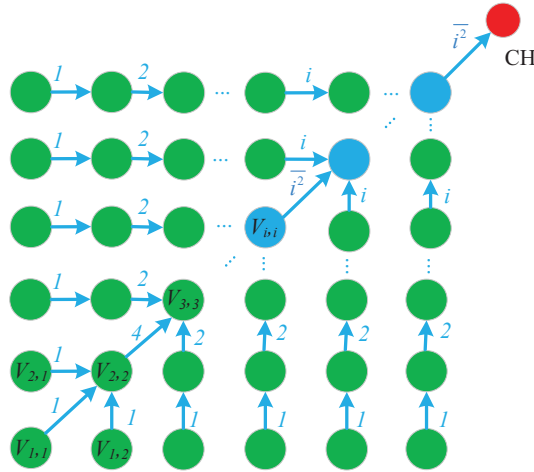
## IV. DATA COLLECTION SCHEME



Fig. 2: Proposed cluster-based data compression model.

### A. The Proposed Hybrid CS-based Cluster Model

In this paper, to improve the data collection efficiency and decrease the energy consumption, we employ the compressed sampling technology and propose to utilize the novel hybrid CS approach to address the signal acquisition problem. By hybrid CS technology, in a cluster, the normal sensing nodes can be divided into un-compressed and compressed nodes. In particular, as the intermediate nodes, the compressing nodes cannot only transmit the collected data following the routing algorithm, but also compress the data traversing them. To clearly explain our idea, we use the instance in Fig. 2 to clarify our cluster-based data compression model. Assuming the nodes in a given cluster-based spanning tree are uniformly distributed in a $N \times N$ squared region, where the CH is located in the lower-right corner, and the nodes in green are the compressing nodes. The data is forwarded along the shortest

path (in hops), i.e., following the sequence of blue arrows. Each node will produce one data unit, and the values in the blue arrows in Fig. 2 represent the total number of the data traversing the link. For instance, nodes $v_{1,1}$, $v_{1,2}$, and $v_{2,1}$ each transfer one unit of data to $v_{2,2}$, and $v_{2,2}$ transfers 4 units of data to $v_{3,3}$. Assuming the hybrid CS mechanism is carried out from $v_{i,i}$, $\overline{i^2}$ units of compressed data will be acquired at CH. The compressing result depends on the compression ratio or the sparsity of the data.

In this paper, we focus on the optimal energy consumption-based clustering problem based on the model in Sec. III-B. Therefore, we ignore the procedure of data compressing.
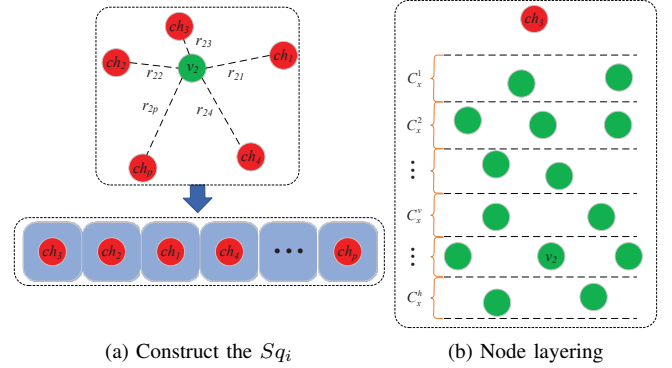


(a) Construct the $Sq_i$      (b) Node layering

Fig. 3: Hierarchical node clustering process.

### B. The Exact Approach

For a given cluster $\mathcal{C}_x$, $\gamma_x$, $D_x$ denote the data compression ratio of $\mathcal{C}_x$ and the scale of the gathered data respectively. Thus, the scale of the compressed data can be expressed by

$$\sigma_x = \frac{D_x}{\gamma_x}. \tag{5}$$

According to the depiction for the ODC problem, we can infer that different node clustering schemes will result in different spanning trees with different compression ratios. In the following, we will present the exact expression for addressing the ODC problem.

Let $e_{ij}$ be a potential link between the nodes $v_i$ and $v_j$ in the spanning tree $G_x(V_x, E_x)$, and $n_{ij}$ the data to be transmitted over $e_{ij}$. Following Eq. 3 and Eq. 4, the energy costs for transferring $n_{ij}$ units of data over link $e_{ij}$ can be computed as

$$C(e_{ij}) = C_s(n_{ij}, d_{ij}) + C_r(n_{ij}), \tag{6}$$

while the energy consumption for $\mathcal{C}_x$ is expressed as

$$C(\mathcal{C}_x) = \sum_{e_{ij} \in E_x} C(e_{ij}). \tag{7}$$

Thus, the entire consumed energy for the entire network is

$$C_{all} = \sum_{x=1}^{p} C(\mathcal{C}_x). \tag{8}$$

By following the hybrid CS theory, we acquire the following LP formulation to express the optimization model:

$$\min_{\mathcal{C}_x} \quad C_{all} = \sum_{x=1}^{p} C(\mathcal{C}_x),$$

$$\text{s.t} \quad \sum_{j:e_{ij}\in E_x} n_{ij} \geq \sum_{k:e_{ki}\in E_x} (|\delta\mathcal{C}_x| - \sigma_x) + \delta, \forall v_i \in \mathcal{C}_x,$$

$$\sum_{j:e_{ij}\in E_x} n_{ij} \geq \sum_{k:e_{ki}\in E_x} (\sigma_x - \delta)\xi_i + \delta, \forall v_i \in \mathcal{C}_x,$$

$$\sum_{j:e_{ij}\in E_x} \overline{n_{ij}} \geq \sum_{k:e_{ki}\in E_x} \overline{n_{ki}} + \frac{1}{|\mathcal{C}_x|}, \forall v_i \in \mathcal{C}_x,$$

$$m_{ij} \geq \overline{n_{ij}}, \forall e_{ij} \in E_x,$$

$$\sum_{e_{ij}\in E_x} m_{ij} = |\mathcal{C}_x| - 1,$$

$$n_{ij} \geq m_{ij}, \forall e_{ij} \in E_x,$$

$$m_{ij} \geq \sigma_x^{-1} n_{ij}, \forall e_{ij} \in E_x,$$

$$S(v_i) = \mathcal{C}_x, i = 1, 2, \cdots, q; x = 1, 2, \cdots, p,$$

$$W(\mathcal{C}_x) = \{v_i | S(v_i) = \mathcal{C}_x\}.$$

$$(9)$$

In Eq. 9, $m_{ij} \in \{0,1\}$ indicates whether $e_{ij}$ in $G_x(V_x, E_x)$ represents the last link connecting a normal node and $ch_x$; $\xi_i \in \{0,1\}$ is an indicator variable for an intermediate node $v_{ij}$; $|\mathcal{C}_i|$ denotes the number of the nodes in cluster $\mathcal{C}_i$; $\overline{n_{ij}}$ or $\overline{n_{ki}}$ denotes the flow over visual link (indicating the connectivity of two adjacent nodes). Eq. 9 is actually an extension of the minimum-cost flow problem. The optimization in Eq. 9 is aimed at seeking node clustering together by way of spanning tree construction – i.e., $\mathcal{C}_x, x = 1, 2, \cdots, p$, to minimize the entire energy consumption. In particular, the constraints present the rules for building $G_x$ for $\mathcal{C}_x, x = 1, 2, \cdots, p$. More in detail, the first two constraints guarantee the flow conservation at the intermediate node, where the data compressed sampling occurs. Constraints three to seven ensure the flow over each link in the network is non-negative. The last two constraints prevent the nodes ($v_i, i = 1, 2, \cdots, q$) from belonging to more than one of the clusters $\mathcal{C}_x, x = 1, 2, \cdots, p$.

Furthermore, from Eq. 9, we can see that the proposed exact approach is an integer LP-based solution, that cannot be run in polynomial time. In the following sub-section, we will present a greedy approach – an approximate optimal solution, to efficiently address the ODC problem.

### C. The Greedy Approach

From Eq. 5, the smaller $\gamma_x$ the larger the required $D_x$ – i.e., more data collection is required. As a result, more data requires to be forwarded to the CH, resulting in more energy consumption. Furthermore, following Eq. 3, larger data transfer distances result in the increase of energy consumption. Therefore, the energy consumption of cluster $\mathcal{C}_x$ depends on two factors: the compression factor $\gamma_x$ and the distance $r_{ix}$ between the normal node $v_i$ and $ch_x$.

To improve the energy consumption efficiency, $r_{ix}$ should be minimized while improving the compression ratio $\gamma_x$ of cluster $\mathcal{C}_x$, which involves a multi-objective trade-off. To

allocate $v_i$ into an approximate optimal cluster, we define the *balance factor* $B(i, x)$, which quantify the effect of allocating $v_i$ into a existing cluster $\mathcal{C}_x$:

$$B(i, x) = B_s(i, x)\rho + B_c(i, x)(1 - \rho), \quad (10)$$

where $B_s(i, x)$ is to compute the difference of data sparsity when $v_i$ is allocated to cluster $\mathcal{C}_x$, $B_c(i, x)$ is to compute the normalized distance between $v_i$ and $\mathcal{C}_x$, and $\rho \in [0, 1]$ aims at specifying the weight of $B_s(i, x)/B_c(i, x)$ in $B(i, x)$.

In particular, $B_s(i, x)$ and $B_c(i, x)$ can be respectively computed by Eq. 11 and Eq. 12.

$$B_s(i, x) = \frac{DD(i, x)}{\max\{|DD(:, :)|\}}; \quad (11)$$

$$B_c(i, x) = \frac{r_{ix}}{\max\{r_{\alpha\beta}, \alpha = 1, 2, \cdots, q, \beta = 1, 2, \cdots, p\}}. \quad (12)$$

In Eq. 11, $DD(i, x)$ represents the difference of data sparsity based on discrete cosine transform (DCT) algorithm and data compressed sampling, and $\{|DD(:, :)|\}$ denotes the set of the difference value when $v_i$ is allocated to different clusters. In Eq. 12, $\{r_{\alpha\beta}, \alpha = 1, 2, \cdots, q, \beta = 1, 2, \cdots, p\}$ re the set of distance value between each $v_\alpha$ and each $ch_\beta$.

In Eq. 10, $\rho$ specifies the slope of $v_i$, when $v_i$ is allocated to a cluster $\mathcal{C}_x$. The higher the value of $\rho$, the more inclined $v_i$ is to be allocated to the cluster with a higher compression ratio, else the cluster that is closer to $v_i$ is more inclined.

After $\rho$ is determined, the $\mathcal{C}_x$ that minimizes $B(i, x)$ is selected as the candidate cluster. Thus, this leads to a greedy approach which can cluster one node at a time, until the clustering problem of the ODC task addressed. However, as an approximate optimal scheme, the optimal approach based on computing $B(i, x)$ can sometimes acquire multiple locally optimal solutions.

In the following sub-section, we propose a hierarchical data collection scheme based on a proposed hybrid clustering scheme, integrating both the proposed exact and greedy approaches.

### D. The Hybrid Clustering-based Data Collection Scheme

In the proposed UAV-IWSN, all the computing operations (e.g., data sparsity computation based on DCT algorithm, constructing spanning tree, etc.) are performed onboard in the UAV, since it is equipped with high-performance computing and data storing capabilities. We propose Alg. 1, a hybrid optimal approach, to perform node clustering. In line 1 of Alg. 1, $\mathcal{V}_u$ represents the set of un-clustered nodes. Alg. 1 executes node clustering individually for $v_i \in \mathcal{V}_u$, as shown in lines 2 to 11; first, lines $3-7$ compute a set $ch_{\min}$ of optimal solution greedily. In particular, ithe LEACH algorithm [28] is invoked to execute the node clustering for data collection in the first round, such that the data compression ratio of each cluster can be computed. Then, in line 9, the proposed exact approach is utilized to select a solution that can minimize Eq. 9.

From Alg. 1, we can see that the DCT algorithm requires to be frequently invoked, which takes most of the running time. By Eq. 10, $v_i \in \mathcal{V}_u$ is preferentially allocated to $\mathcal{C}_x$, when $r_{ix}$ is very small, or smaller than a given threshold.

| parameter | description | value |
|---|---|---|
| $\|V\|$ | the number of nodes | 100 |
| $\theta_{fs}$ | sending power coefficient | 11 pJ/bit/m$^2$ |
| $\theta_{mp}$ | sending power coefficient | 0.00145 pJ/bit/m$^2$ |
| $C_a$ | consumed energy to activate the sending or receiving circuit | 65 nJ/bit |
| $r$ | communication range of each node | 200 m |
| $p$ | number of cluster | $[5, 6, \cdots, 24]$ |
| $h$ | number of layer in total | $[5, 6, \cdots, 24]$ |
| $\mathcal{R}$ | distance threshold in the free space model | 240 m |
| $\mu$ | number of layer leading to a node re-clustering phase | $[5, 6, 7, 8, 9]$ |
| $n_{ij}$ | scale of the forwarded data on each node | 40 |
| $\gamma_x$ | compression ratio of each cluster | 50 |

TABLE I: Parameter setting during model evaluation.

---

**Algorithm 1** Node Clustering($\mathcal{V}_u$)

1: Initialize $\mathcal{V}_u$.
2: **for** $v_i \in \mathcal{V}_u$ **do**
3:     **for** $\mathcal{C}_x, x = 1, 2, \cdots, p$ **do**
4:         Following Eq. 10, Eq. 11, and Eq. 12, computing $r_{ix}, DD(i,x), \max\{|DD(:,:)|\}$, and $\max\{r_{\alpha\beta}, \alpha = 1, 2, \cdots, q, \beta = 1, 2, \cdots, p\}$, respectively.
5:         Compute $B_s(i,x)$, $B_c(i,x)$, and $B(i,x)$.
6:     **end for**
7:     Compute a set $ch_{\min}$ of CHs minimizing $B(i,x)$.
8:     **if** $|ch_{\min}| \geq 2$ **then**
9:         Allocating $v_i$ to the $\mathcal{C}_m$ with $ch_m \in ch_{\min}$, leading Eq. 9 to be minimized.
10:     **end if**
11: **end for**
12: Return the computed node clustering policy.

---

With this property, in the following Alg. 2, we introduce a hierarchical data collection scheme to improve the data collection efficiency. In Alg. 2, after the initialization phase in lines $1-2$, line 4 constructs the sequence $Sq_i$ of $ch_x$ for $v_i$. Note that the sequence is in the ascending order of $r_{ix}, x = 1, 2, \cdots, p$. For instance, Fig. 3a shows the construction for the sequence $Sq_2$ of $ch_x$ $(i = 1, 2, \cdots, p)$ for $v_2$. Then, following line 5 of Alg. 2, allocate $v_i$ to the cluster whose CH is marked in $Sq_i[0]$, i.e., $\mathcal{C}_3$ in Fig. 3a. After that, we present a hierarchical node classification scheme shown in lines $8-9$. Line 8 groups the nodes in $\mathcal{C}_x$ into $h$ layers of equal range from 1 to $r_{\max}$ (Eq. 13). For instance, in Fig. 3b, the nodes in $\mathcal{C}_x$ are divided into $h$ parts: $\mathcal{C}_x^1, \mathcal{C}_x^2, \cdots, \mathcal{C}_x^h$. By specifying $\mu$ based on the practical scenario, allocating the nodes in $\mathcal{C}_x^\mu$, $\mathcal{C}_x^{\mu+1}, \cdots, \mathcal{C}_x^h$ $(1 < \mu < h)$ to the node set $\mathcal{V}_u$. This leads to a node re-clustering phase for the nodes in $\mathcal{V}_u$ by invoking Alg. 1, as shown in line 11 of Alg. 2. Thus, the running efficiency of the node clustering phase will be improved. Then, line 12 constructs $G_x(V_x, E_x), x = 1, 2, \cdots, p$ by evoking MECDA_GREEDY algorithm [29], and the data in each cluster follows the specified hybrid CS policy and is forwarded along the spanning tree to the CH. In our work, we assume the consumed power of the UAV is proportional to the path length, take the shortest possible route (that the UAV starts from the source, traverses each cluster head, and then goes back to the source) into account, and treat the path planning optimization problem as a travelling salesperson problem. Hence, in line 13, the ACO algorithm [30] can be selected as a candidate approach, to schedule the data collection path for the UAV when the energy consumption efficiency is taken into account.

---

**Algorithm 2** Hierarchical Data Collection($\mu, h$)

1: UAV collects the status of $v_i \in V$.
2: Initialize $\mathcal{V}_n$.
3: **for** $v_i \in \mathcal{V}_n$ **do**
4:     Construct the sequence $Sq_i$ for $v_i$.
5:     Allocate $v_i$ to the cluster whose CH is marked in $Sq_i[0]$.
6: **end for**
7: **for** $\mathcal{C}_x, x = 1, 2, \cdots, p$ **do**
8:     Based on $r_{\max}$, dividing the nodes in $\mathcal{C}_x$ into $h$ layers of equal range: $\mathcal{C}_x^1, \mathcal{C}_x^2$, and $\mathcal{C}_x^h$.
9:     Allocate the nodes in $\mathcal{C}_x^\mu, \mathcal{C}_x^{\mu+1}, \cdots, \mathcal{C}_x^h$ $(1 < \mu < h)$ to the node set $\mathcal{V}_u$.
10: **end for**
11: Evoke Node Clustering($\mathcal{V}_u$) algorithm to re-cluster the nodes in $\mathcal{V}_u$.
12: Build $G_x(V_x, E_x), x = 1, 2, \cdots, p$ by evoking MECDA_GREEDY algorithm.
13: Based on the clustering information, evoking ACO algorithm to schedule the data collection path.

---

$$r_{\max} = \max\{r_{ix}, i = 1, 2, \cdots, q\}. \tag{13}$$

## V. EVALUATIONS

In this section, we present simulation results to evaluate the performance of the proposed approach for tackling the ODC problem. We first test the performance of the proposed hierarchical clustering algorithm, and then we present some comparison results for demonstrating our proposal. Further, we survey some normal path planning schemes to demonstrate the data collection efficiency for the UAV, when the proposed data collection scheme is performed[1].

---

[1] The simulations were conducted on an Intel(R) Core i7-8565U 1.8 GHz machine with 8 Gb RAM.
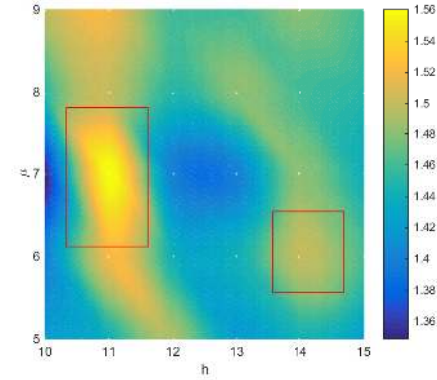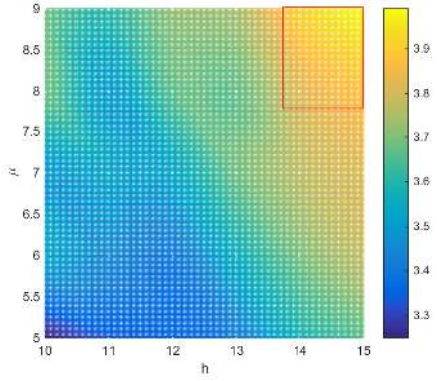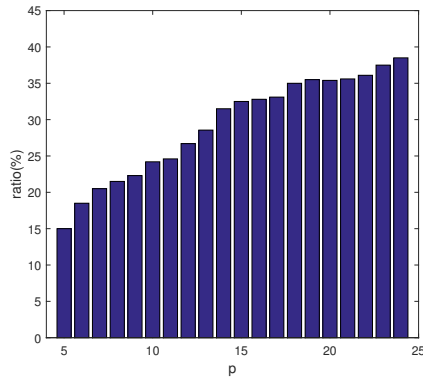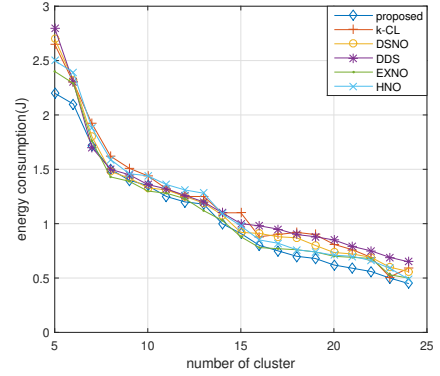
(a) Energy consumption with $h/\mu$



(b) Running time with $h/\mu$



(c) Ratio of multiple optimal solutions with $p$

Fig. 4: Test for the studied scheme.



(a) Energy consumption comparison



(b) Running time comparison

Fig. 5: Comparisons about the energy consumption and running time

number is fixed. As shown in Fig. 4b, different from the energy consumption test, the running time of the proposed scheme increases with $h/\mu$, especially $\mu$. This is due to the fact that larger $\mu$ increases the complexity of Alg. 1. In order to demonstrate the necessity of our proposal, we also test the probability of the greedy procedure in acquiring multiple locally optimal solutions – when the greedy approach is iterated at most 80 times. Then 200 are repetitions executed. when the nodes are divided into $5 - 24$ clusters. The ratio of acquiring multiple locally optimal solutions is shown in Fig. 4c; we can see that the ratio increases with the number of clusters – i.e., a larger number of clusters the greater the number of potential options.

Furthermore, to evaluate our approach, we also make comparisons with the other four schemes: DSNO, DDS, EXNO, HNO, and the scheme proposed in [19]. In particular, DSNO is the scheme that does not take the data compression into account; DDS is a scheme that only considers the distance between the normal node and CHs; EXNO is an approach that ignores the exact approach-based decision procedure; HNO is a method that is irrespective of the hierarchical node clustering procedure. In particular, we refer to the clustering scheme in [19] as the k-CL which ignores as well the data compression. We first test the energy consumption for the various algorithms. The result in Fig. 5a demonstrates
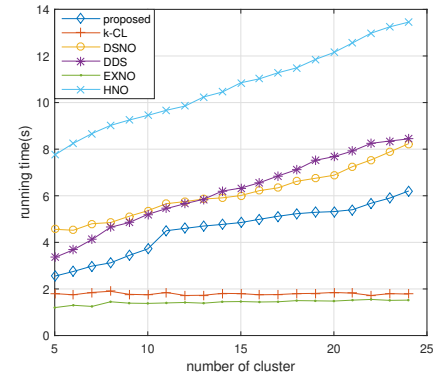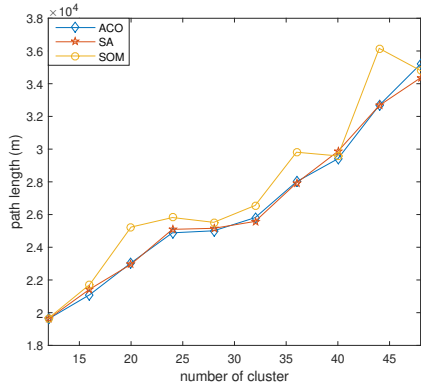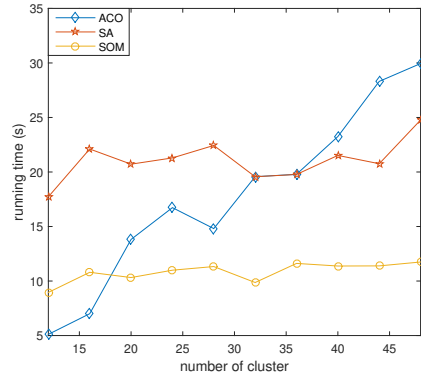
The simulations were implemented in Python 3.7 over a 2 dimensional array representing a $2000 \, \text{m}^2$ area, and where 100 nodes with $p$ CHs were randomly distributed. The communication among the nodes follows the Zigbee protocol [31], and the scale of each frame is limited to be 108 bytes. The forwarded data on each node is in float (4 bytes for each unit). All the clusters are with the same compression ratio. The simulation parameters are summarized in Table I.

Fig. 4a, displays the relationship between the energy consumption and $h/\mu$, when the 100 nodes are divided into 10 clusters by our approach. From Fig. 4a, we can see that $h/\mu$ is not correlated to energy consumption when the cluster

(a) Path length comparison



(b) Running time comparison

Fig. 6: Comparison among the ACO, SA, and SOM algorithms.



Fig. 7: Path planning test

the proposed scheme.

Finally, we utilize our proposal to plan the path for the data collecting UAV after the monitoring nodes are clustered. To understand why the ACO algorithm is selected to plan the path for the UAV in our proposed scheme, we compare ACO algorithm-based approach with the other two normal schemes: the simulated annealing (SA) algorithm [32] and the self-organizing-map (SOM) algorithm-based path planning schemes [33], especially in the aspect of path length and running efficiency. We select the coordinates in the first footnote as the CH coordinates that are distributed in a 8000 m*8000 m area [1]. In Fig. 6, we select $12 - 48$ CHs and respectively use the aforementioned schemes to plan the data collecting path and compute the path length together with the running time. In particular, all the results in Fig. 6 are the mid-values based on 1000 runs. As the test results for the path planning length shown in Fig. 6a, the ACO algorithm-based path planning scheme performs the best and is closest to the optimal solution. Meanwhile, in Fig. 6b, we also note that the running time of the ACO algorithm-based path planning scheme increases with the number of the CH and performs the best when the number of the CH is less than 18. From Fig. 6b, we can as well see that the performance of ACO is between SA and SOM. To summarized, we select ACO algorithm as the candidate approach to plan the data collecting path for the UAV in our scheme.

Besides, in Fig. 7, we also show a path planning case with 48 CHs based on the ACO algorithm-based approach. Assuming the data collection is started from CH 24, the path along the red arrow, i.e., $< 24, 10, 42, 5, 48, 39, 32, 21, 13, 25, 14, 23, 11, 12, 33, 46, 15, 40, 9, 1, 8, 38, 31, 44, 18, 7, 28, 36, 6, 37, 19, 27, 43, 17, 30, 20, 47, 3, 22, 16, 41, 34, 29, 2, 26, 4, 35, 45 >$ can be selected as an approximate-optimal path. The result in Fig. 7 demonstrates as well that our scheme can accurately guarantee an efficient data collection with no duplicate path, leading the ODC problem to be accurately addressed.

## VI. CONCLUSION

In this paper, we have employed the paradigm of intelligent agriculture to improve farming automation digitalization, and proposed a hierarchical data collection scheme to perform energy-optimal data collection in UAV-IWSN based agricultural monitoring systems. To improve the energy efficiency of data gathering from the UAV-IWSN nodes, the concept of data compressed sampling and node clustering have been
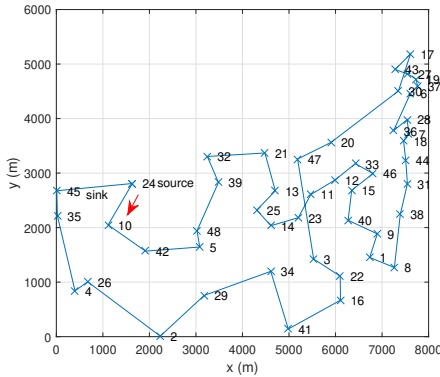
that the here proposed scheme benefits from the hierarchical hybrid clustering scheme and displays improved performance as compared to the other methods. In addition, we can see that the data compression skill in the proposed hierarchical hybrid clustering scheme has a maximum impact on the energy consumption result during our evaluation. Furthermore, the execution time results in Fig. 6b demonstrate that the proposed hierarchical node clustering framework can improve the running efficiency of the algorithm, and the exact approach-based decision procedure takes most of the running time in

[1]CH 1: (6734, 1453); CH 2: (2233, 10); CH 3: (5530, 1424); CH 4: (401, 841); CH 5: (3082, 1644); CH 6: (7608, 4458); CH 7: (7573, 3716); CH 8: (7265, 1268); CH 9: (6898, 1885); CH 10: (1112, 2049); CH 11: (5468, 2606); CH 12: (5989, 2873); CH 13: (4706, 2674); CH 14: (4612, 2035); CH 15: (6347, 2683); CH 16: (6107, 669); CH 17: (7611, 5184); CH 18: (7462, 3590); CH 19: (7732, 4723); CH 20: (5900, 3561); CH 21: (4483, 3369); CH 22: (6101, 1110); CH 23: (5199, 2182); CH 24: (1633, 2809); CH 25: (4307, 2322); CH 26: (675, 1006); CH 27: (7555, 4819); CH 28: (7541, 3981); CH 29: (3177, 756); CH 30: (7352, 4506); CH 31: (7545, 2801); CH 32: (3245, 3305); CH 33: (6426, 3173); CH 34: (4608, 1198); CH 35: (23, 2216); CH 36: (7248, 3779); CH 37: (7762, 4595); CH 38: (7392, 2244); CH 39: (3484, 2829); CH 40: (6271, 2135); CH 41: (4985, 140); CH 42: (1916, 1569); CH 43: (7280, 4899); CH 44: (7509, 3239); CH 45: (10, 2676); CH 46: (6807, 2993); CH 47: (5185, 3258); CH 48: (3023, 1942)

introduced. For that a hybrid *compressed technique* for clustering has been implemented. Based on the proposed model, we introduce a hierarchical data collection scheme integrating an exact approach and a greedy approach. By dividing the nodes into different layers, the exact approach and greedy approach can be intelligently matched. Particularly, for the exact approach, an improved version of the minimum cost flow problem – that can be expressed by LP formulation to model the energy-optimal problem – is designed. The greedy approach is based on a *balance factor* parameter, consisting of both, data sparsity and the distance from the cluster head to the normal nodes. The proposed scheme is tuned with an *ant-colony-optimization* path planning policy. Simulation results show that this method can efficiently gather the data than several normal schemes, especially in energy consumption, and plan the path for the UAV at a low energy cost.

For future work we plan to expand network architecture aspects. For instance, we intend to utilize software-defined networking (SDN) technology to improve the scalability of the monitoring network, such that the network status can be intensively monitored and surveyed, and the network operation can be uniformly deployed.

## References

[1] A. Antonacci, F. Arduini, D. Moscone, G. Palleschi, and V. Scognamiglio, "Nanostructured (Bio)sensors for Smart Agriculture," *TrAC Trends in Analytical Chemistry*, vol. 98, no. 1, pp. 95–103, Jan. 2018.

[2] S. Heble, A. Kumar, K. V. V. D. Prasad, S. Samirana, P. Rajalakshmi, and U. B. Desai, "A Low Power IoT Network for Smart Agriculture," in *Proc. IEEE World Forum on Internet of Things (WF-IoT)*, Coimbatore, India, 2018, pp. 609–614.

[3] O. Westermann, W. Frch, P. Thornton, J. Krner, L. Cramer, and B. Campbell, "Scaling up agricultural interventions: Case studies of climate-smart agriculture," *Agricultural Systems*, vol. 165, no. 9, pp. 283–293, Step. 2018.

[4] M. Martínez-García, Y. Zhang, K. Suzuki, and Y.-D. Zhang, "Deep recurrent entropy adaptive model for system reliability monitoring," *IEEE Transactions on Industrial Informatics*, 2020.

[5] R. Alcarria, B. Bordel, M. Á. Manso, T. Iturrioz, and M. Pérez, "Analyzing UAV-Based Remote Sensing and WSN Support for Data Fusion," in *Proc. International Conference on Information Technology & Systems (ICITS 2018)*, Cham, 2018, pp. 756–766.

[6] M. Aazam, K. A. Harras, and S. Zeadally, "Fog Computing for 5G Tactile Industrial Internet of Things: QoE-Aware Resource Allocation Model," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 3085–3092, May 2019.

[7] M. H. ur Rehman, I. Yaqoob, K. Salah, M. Imran, P. P. Jayaraman, and C. Perera, "The Role of Big Data Analytics in Industrial Internet of Things," *Future Generation Computer Systems*, vol. 99, no. 10, pp. 247–259, Oct. 2019.

[8] G. Han, X. Miao, H. Wang, L. Liu, J. Jiang, and Y. Peng, "A Dynamic Multipath Scheme for Protecting Source-Location Privacy Using Multiple Sinks in WSNs Intended for IIoT," *IEEE Transactions on Industrial Informatics*, DOI: 10.1109/TII.2019.2953937, to appear.

[9] L. Liu, G. Han, Y. He, and J. Jiang, "Fault-Tolerant Event Region Detection on Trajectory Pattern Extraction for Industrial Wireless Sensor Networks," *IEEE Transactions on Industrial Informatics*, DOI: 10.1109/TII.2019.2933238, to appear.

[10] M. Martínez-García, Y. Zhang, and T. Gordon, "Modeling lane keeping by a hybrid open-closed-loop pulse control scheme," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2256–2265, Oct. 2016.

[11] D. Popescu, F. Stoican, G. Stamatescu, O. Chenaru, and L. Ichim, "A Survey of Collaborative UAVWSN Systems for Efficient Monitoring," *Sensors*, vol. 19, no. 21, p. 4690, Oct. 2019.

[12] X. Wang, Q. Zhou, and C.-T. Cheng, "A UAV-Assisted Topology-Aware Data Aggregation Protocol in WSN," *Physical Communication*, vol. 34, no. 7, pp. 48–57, June 2019.

[13] Q. Wang, D. Lin, P. Yang, and Z. Zhang, "An Energy-Efficient Compressive Sensing-Based Clustering Routing Protocol for WSNs," *IEEE Sensors Journal*, vol. 19, no. 10, pp. 3950–3960, May 2019.

[14] V. K. Singh, M. Kumar, and S. Verma, "Node Scheduling and Compressed Sampling for Event Reporting in WSNs," *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 3, pp. 418–431, July 2019.

[15] P. Rajalakshmi and S. Devi Mahalakshmi, "IoT Based Crop-Field Monitoring and Irrigation Automation," in *Proc. IEEE International Conference on Intelligent Systems and Control (ISCO)*, Coimbatore, India, 2016, pp. 1–6.

[16] Chetan Dwarkani M, Ganesh Ram R, Jagannathan S, and R. Priyatharshini, "Smart Farming System Using Sensors for Agricultural Task Automation," in *Proc. IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR)*, Chennai, India, 2015, pp. 49–53.

[17] G. Sun, T. Hu, G. Yang, and J. Jia, "Real-Time and Clock-Shared Rainfall Monitoring With a Wireless Sensor Network," *Computers and Electronics in Agriculture*, vol. 119, no. 11, pp. 1–11, Nov. 2015.

[18] J. Bauer, B. Siegmann, T. Jarmer, and N. Aschenbruck, "On the Potential of Wireless Sensor Networks for the In-Situ Assessment of Crop Leaf Area Index," *Computers and Electronics in Agriculture*, vol. 128, no. 10, pp. 149–159, Oct. 2016.

[19] M. Abderrahim, H. Hakim, H. Boujemaa, and F. Touati, "A Clustering Routing based on Dijkstra Algorithm for WSNs," in *Proc. IEEE International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, Sousse, Tunisia, Tunisia, 2019, pp. 605–610.

[20] L. Hui, M. Zhijun, W. Hua, and X. Min, "Spatio-Temporal Variation Analysis of Soil Temperature Based on Wireless Sensor Network," *International Journal of Agricultural and Biological Engineering*, vol. 9, no. 6, pp. 131–138, June 2016.

[21] W.-H. Nam, T. Kim, E.-M. Hong, J.-Y. Choi, and J.-T. Kim, "A Wireless Sensor Network (WSN) Application for Irrigation Facilities Management Based on Information and Communication Technologies (ICTs)," *Computers and Electronics in Agriculture*, vol. 143, no. 12, pp. 185–192, Dece. 2017.

[22] J. G. Caicedo-Ortiz, E. D. la Hoz-Franco, R. M. Ortega, G. Pieres-Espitia, H. Combita-Nio, F. Estvez, and A. Cama-Pinto, "Monitoring System for Agronomic Variables Based in WSN Technology on Cassava Crops," *Computers and Electronics in Agriculture*, vol. 145, no. 2, pp. 275–281, Feb. 2018.

[23] K. Wang, W. Li, L. Deng, Q. Lyu, Y. Zheng, and S. Yi, "Rapid Detection of Chlorophyll Content and Distribution in Citrus Orchards Based on Low-Altitude Remote Sensing and Bio-Sensors," *International Journal of Agricultural and Biological Engineering*, vol. 11, no. 2, pp. 164–169, Feb. 2018.

[24] D. L. Ndzi, A. Harun, F. M. Ramli, M. L. Kamarudin, A. Zakaria, A. Y. M. Shakaff, M. N. Jaafar, S. Zhou, and R. S. Farook, "Wireless Sensor Network Coverage Measurement and Planning in Mixed Crop Farming," *Computers and Electronics in Agriculture*, vol. 105, no. 7, pp. 83–94, July 2014.

[25] J. Gago, C. Douthe, R. Coopman, P. Gallego, M. Ribas-Carbo, J. Flexas, J. Escalona, and H. Medrano, "UAVs Challenge to Assess Water Stress for Sustainable Agriculture," *Agricultural Water Management*, vol. 153, no. 5, pp. 9–19, May 2015.

[26] A. Malaver, N. Motta, P. Corke, and F. Gonzalez, "Development and Integration of a Solar Powered Unmanned Aerial Vehicle and a Wireless Sensor Network to Monitor Greenhouse Gases," *Sensors*, vol. 15, no. 2, pp. 4072–4096, Feb. 2015.

[27] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-Efficient Communication Protocol for Wireless Microsensor Networks," in *Proc. IEEE Annual Hawaii International Conference on System Sciences*, Maui, HI, USA, USA, 2000, pp. 1–10.

[28] Z. Cui, Y. Cao, X. Cai, J. Cai, and J. Chen, "Optimal LEACH protocol with modified bat algorithm for big data sensing systems in Internet of Things," *Journal of Parallel and Distributed Computing*, vol. 132, no. 10, pp. 217–229, Oct. 2019.

[29] L. Xiang, J. Luo, and A. Vasilakos, "Compressed Data Aggregation for Energy Efficient Wireless Sensor Networks," in *Proc. IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, Salt Lake City, UT, USA, 2011, pp. 46–54.

[30] Y. Liu, J. Ma, S. Zang, and Y. Min, "Dynamic Path Planning of Mobile Robot Based on Improved Ant Colony Optimization Algorithm," in *Proc. IEEE International Conference on Networks, Communication and Computing*, New York, NY, USA, 2019, pp. 248–252.

[31] A. Abane, M. Daoui, S. Bouzefrane, and P. Muhlethaler, "Ndn-over-zigbee: A zigbee support for named data networking," *Future Generation Computer Systems*, vol. 93, no. 4, pp. 792–798, April 2019.

[32] B. Basbous, "2D UAV Path Planning with Radar Threatening Areas using Simulated Annealing Algorithm for Event Detection," in *Proc. IEEE International Conference on Artificial Intelligence and Data Processing (IDAP)*, Malatya, Turkey, Turkey, 2018, pp. 1–7.

[33] D. Zhu, X. Cao, B. Sun, and C. Luo, "Biologically inspired self-organizing map applied to task assignment and path planning of an auv system," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 2, pp. 304–313, June 2018.