Optimization of submodularity and BBO-based routing protocol for wireless sensor deployment

Wang, Yaoli; Duan, Yujun; Di, Wenxia; Chang, Qing; Wang, Lipo

Optimization of Submodularity and BBO-Based Routing Protocol for Wireless Sensor Deployment

Yaoli Wang [1,*] 🔴, **Yujun Duan** [1], **Wenxia Di** [2], **Qing Chang** [1] **and Lipo Wang** [3] 🔴

[1]   College of Information and Computer, Taiyuan University of Technology, Jinzhong 030600, China; duanyujun0248@link.tyut.edu.cn (Y.D.); changqing@tyut.edu.cn (Q.C.)
[2]   Foreign Languages Department, Taiyuan Normal University, Jinzhong 030600, China; wendy_cn@263.net
[3]   School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore; elpwang@ntu.edu.sg
[*]   Correspondence: wangyaoli@tyut.edu.cn

Abstract: Wireless sensors are limited by node costs, communication efficiency, and energy consumption when wireless sensors are deployed on a large scale. The use of submodular optimization can reduce the deployment cost. This paper proposes a sensor deployment method based on the Improved Heuristic Ant Colony Algorithm-Chaos Optimization of Padded Sensor Placements at Informative and cost-Effective Locations (IHACA-COpSPIEL) algorithm and a routing protocol based on an improved Biogeography-Based Optimization (BBO) algorithm. First, a mathematical model with submodularity is established. Second, the IHACA is combined with pSPIEL-based on chaos optimization to determine the shortest path. Finally, the selected sensors are used in the biogeography of the improved BBO routing protocols to transmit data. The experimental results show that the IHACA-COpSPIEL algorithm can go beyond the local optimal solutions, and the communication cost of IHACA-COpSPIEL is 38.42%, 24.19% and 8.31%, respectively, lower than that of the greedy algorithm, the pSPIEL algorithm and the IHACA algorithm. It uses fewer sensors and has a longer life cycle. Compared with the LEACH protocol, the routing protocol based on the improved BBO extends the life cycle by 30.74% and has lower energy consumption.

Keywords: wireless sensor deployment; submodularity; ant colony algorithm; routing protocol; biogeography-based optimization

1. Introduction

Wireless sensors are widely deployed on a large scale in commercial fields [1,2], but are limited by node costs, communication efficiency between nodes, and energy consumption [3–5], e.g., in forest and grassland fire risk monitoring and early warning. The problem of wireless sensor deployment is considered as deploying a certain number of nodes to meet monitoring needs, that is, finding the number and location of deployed nodes. The goal of solving this problem is to find as few sensors as possible to meet the monitoring requirements and reduce the communication cost. It is transformed into an optimal sensor node solution set, which is an NP-hard problem. The sensor deployment problem has diminishing returns, e.g., submodularity [6–8]. Initially, when a small number of sensors are deployed, each new sensor will significantly improve its deployment utility. As more sensors are placed, the improvement in utility from adding new sensors diminishes. Krause [9] showed that for problems with submodularity, at least the $(1 - 1/e)$ approximation of the optimal solution can be obtained using the greedy algorithm.

Many methods have been proposed for sensor deployment. In [10], Huang et al. assumed that the node's perception ability is a circular area. That is, targets within the circular area are fully perceived,

and targets outside the circular area will not be perceived. In [11], Guestrin et al. proposed the use of mutual information-based optimization criteria so that the set of deployed nodes contains information about unselected points, and the monitoring is quite accurate. In [12], Cheng et al. proposed a Markov random field model to describe the data correlation between sensor nodes. In [9], Krause deployed sensors with a greedy algorithm to maximize the amount of information, but neglecting the influence of the communication distance between nodes. In [13], Krause et al. improved the greedy algorithm and proposed the Padded Sensor Placements at Informative and cost-Effective Locations (pSPIEL) algorithm to solve the problem of sensor deployment optimization under the constraint of communication distance; however, a large number of sensors needed to be deployed. In [14], Mariohat et al. established a Gaussian model, improved the greedy algorithm under the constraint of fixed costs, and proposed the SUPSUB method to minimize the submodular set function, while neglecting the influence of the communication distance between nodes on the deployment. The sensor placement problem considering communication distance is a constrained optimization problem. The bi-projection neural network proposed by Xia et al. [15] can effectively solve large-scale constrained optimization problems, and has good stability and faster convergence [16]. Liu et al. [17] proposed a ML-OAXSMT-PSO construction algorithm, which can significantly reduce the total cost.

During communication transmission, wireless sensors have limited energy, but effective clustering nodes can better save energy and extend the life cycle of the entire network. There are various energy-saving methods. Guo et al. [18] proposed the FTAOA algorithm to minimize task completion time to save node energy. Cheng et al. [12] proposed the NSA algorithm to reasonably deploy nodes and significantly improve network lifetime. Liu et al. [19] proposed the KPNS algorithm to appropriately select more active nodes for monitoring, so that the energy is fully utilized. Effective node clustering can greatly save energy and extend the life cycle of the entire network. The LEACH protocol balances the energy of each sensor in the entire network by randomly selecting cluster heads [20]. However, it has the problem of uneven number and distribution of cluster heads [21], for not having considered the transmission distance. This causes either the nodes far away from the base station to be selected as the cluster head or the nodes far away from the cluster head to die prematurely [22]. In [23], Simon proposed the biogeography-based optimization algorithm with advantages of simple operation, few parameters, and high search accuracy [24]. In [25], Pal and others used the Biogeography-Based Optimization (BBO) algorithms to select cluster heads and cluster nodes, and obtained good energy efficiency. However, the authors only took the distance between cluster heads and the distance between nodes in the cluster into consideration, while neglecting the energy consumed by data transmission between nodes.

Deploying wireless sensors is limited by cost and power consumption [26]. Therefore, the following two issues need to be considered during deployment: one is to achieve efficient data collection; the other is to use as few sensors as possible and minimize the communication distance between sensors to reduce the total energy consumption. Because of the existing problem of deploying fewer sensors in terms of the distance between nodes, and the deficiency of some popular algorithms in the field of sensor deployment, we shall proceed as follows in this paper. The mutual information is used to describe the correlation between the observed and the unobserved points. The communication distance is described by the connection of the graph and the Improved Heuristic Ant Colony Algorithm-Chaos Optimization of Padded Sensor Placements at Informative and cost-Effective Locations (IHACA-COpSPIEL) algorithm is used to choose the optimized submodular model. By considering the distance between clusters, the distance between nodes in the cluster, and the energy consumption of data transmission by the nodes, we obtain an optimized routing protocol in which the BBO algorithm is used to transmit data with an improved cost performance.

The structure of this article is as follows. In Section 2, wireless sensor deployment optimization is introduced. The routing protocol of the wireless sensor network based on the BBO algorithm is presented in Section 3. In Section 4, we introduce the experimental verification, discuss the deployment

effects using different algorithms, and analyze the performance of the protocol for the BBO algorithm. Finally, this article is summarized in Section 5.

2. Wireless Sensor Deployment Optimization

2.1. Wireless Sensor Deployment

2.1.1. Problem Description

Given a certain area of interest, $V(|V| = N)$ is the set of monitored locations in the area. The goal of node deployment is to select a subset $A \subset V(|A| = K < N)$, and the base station can efficiently estimate the value of any element in the set $V \backslash A(|V \backslash A| = N - K)$ based on the observed values of the subset.

Mutual information is a way to describe the correlation between two sets of events. This paper uses mutual information to describe the correlation between set A with deployed sensors and set $V \backslash A$ without deployed sensors. Suppose set $V = [V_1, V_2, \ldots, V_N]$ represents N positions, and $X_V = [X_1, X_2, \ldots, X_N]$ describes the random variables of the observation results of these positions. For any subset $A \subset V$, X_A is used to represent the set of random variables associated with the location subset A. The objective function is to maximize the amount of information containing the unselected location set through the selected deployment set A.

$$\max_{A \subseteq V} F(A) = \max_{A \subseteq V} (H(X_{V \backslash A}) - H(X_{V \backslash A}|X_A)). \tag{1}$$

In Equation (1), $H(X_{V \backslash A})$ represents the entropy of the random variable $X_{V/A}$, and $H(X_{V \backslash A}|X_A)$ represents the random variable $X_{V/A}$ relative to the conditional entropy of X_A.

2.1.2. Objective Function of Sensor Deployment Considering Communication Cost

When deploying sensors, we must consider not only the number of sensors, but also the energy consumption during wireless sensor network transmission, because energy consumption is related to the distance between sensors. However, we cannot determine which sensors will communicate to accurately reduce the communication distance between them. Therefore, the communication distance between all the sensors is selected to be reduced so that the communication distance between subsequent sensors is reduced. Assume that there are N optional points. Node i and node j are selected to deploy sensors. The communication cost between the two nodes is defined as follows.

$$d_{i,j} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}. \tag{2}$$

In Equation (2), (x_i, y_i) is the node i coordinate, and (x_j, y_j) is the node j coordinate.

$$C(A) = \sum_{k=1}^{n=|A|} \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}. \tag{3}$$

In Equation (3), (x_i, y_i) and (x_j, y_j) are the coordinates of the nodes in the set A.

In this paper, the problem of improving the efficiency of the sensor submodularity and reducing the communication cost (C) is transformed into a combination optimization problem. The objective function is as follows.

$$\max_{A \subseteq V} F(A), s.t. C(A) \leq B. \tag{4}$$

For the communication cost budget $B > 0$, Equation (4) aims to find the solution set with the maximum mutual information within a low communication cost.

2.2. IHACA-COpSPIEL Deployment Method

Krause et al. proposed the pSPIEL algorithm, which is an improved greedy algorithm, but with both a larger number of sensors and a longer communication distance. The ant colony algorithm is easy to combine with other methods and performs well in path optimization. Therefore, this paper combines the improved ant colony algorithm with the chaotic operator improved pSPIEL algorithm, and proposes the IHACA-COpSPIEL.

2.2.1. Chaos Optimized pSPIEL Algorithm

The standard pSPIEL algorithm applies non-decreasability, submodularity, and locality to solving the problem of sensor node deployment. Compared with the greedy algorithm, pSPIEL can optimize the sensor layout and reduce the communication cost, but with a large number of sensors and slightly higher communication costs. The ergodicity of chaotic motion can effectively traverse each state within a specified range. Therefore, this paper introduces a chaotic operator to traverse all cluster numbers to determine the optimal cluster number. We propose a chaotic optimized pSPIEL algorithm (COpSPIEL). The basic idea of the chaotic locality parameter r adjustment strategy is to use a chaos generator to generate a set of chaotic variables and then use the carrier transform method to map to the locality parameters, and map it to the value range of the locality parameters. Logistic mapping is a typical chaotic system.

$$z_{i+1} = \mu z_i(1 - z_i), i = 0, 1, 2, \ldots, z_i \in (0, 1). \tag{5}$$

In Equation (5), μ is the control parameter. When $\mu = 4$, the system is completely chaotic.
The search r_i is mapped to the domain (0,1) of the logistic equation by Equation (6).

$$z_i = (r_i - r_{\min})/(r_{\max} - r_{\min}), r_i \in (r_{\min}, r_{\max}). \tag{6}$$

It iteratively generates a chaotic sequence by Logistic equation:

$$z^m (m = 1, 2, 3, \ldots). \tag{7}$$

The generated chaotic sequence is inversely mapped by Equation (8):

$$r_i^m = r_{\min} + (r_{\max} - r_{\min})z^m. \tag{8}$$

This returns to the original solution space and produces a solvable chaotic sequence containing chaotic variables.

$$r_i^m = (r_1^m, r_2^m, \ldots, r_i^m) \tag{9}$$

The locality parameter r optimizes the search space in this sequence.

2.2.2. Improved Heuristic Ant Colony Algorithm

The ant colony algorithm was inspired by the research on real ant colony behavior, and has been applied to the optimization of communication networks and others. The essence of the ant colony algorithm is to use pheromone as a medium for ants in an ant colony to communicate. In the sensor layout, the ants must be moved toward the sensor node with a large submodular gain. The traditional ACA algorithm has the problem of blind search and it is easy to fall into local optimal solutions. To improve the heuristic function and pheromone, a new mechanism, improved heuristic ant colony algorithm, IHACA is proposed.

The heuristic function of the traditional ant colony algorithm takes no consideration of the distance relationship between the next node j and the adjacent cluster head, and the search is blind. Therefore,

this paper adds the Euclidean distance between the next node j and the cluster head of the adjacent cluster. The improved heuristic function is as follows.

$$\eta_{ij} = \frac{1}{wdist(i,j) + (1-w)dist(j,g_{i1})}, w \in (0,1). \tag{10}$$

where g_{i1} is the first node of cluster i and w is the weight.

In order to avoid premature, stagnation or local optimization problems due to excessive pheromone concentration [27], this paper introduces a local and global pheromone update mechanism. Local update of pheromone helps ants to select unselected points, and a full update of pheromone helps to enhance the global search ability of the algorithm.

Each ant moves from node i to node j, and needs to update the pheromone on the path (i,j) that it just walked.

$$\tau_{ij}(n+1) = \xi\tau_{ij}(n) + \varepsilon\tau_0. \tag{11}$$

In Equation (11), n is the number of iterations, ξ is a local pheromone evaporation coefficient, τ_0 is a pheromone under initial conditions, and ε is a constant.

When all the ants complete this iteration, we select the shortest path and the longest path in this iteration to globally update the pheromone on the path.

$$\tau_{ij}(n+1) = (1-\rho)\tau_{ij}(n) + \rho\sum_{k=1}^{m}\Delta\tau_{ij}^k. \tag{12}$$

$$\Delta\tau_{ij}^k = \begin{cases} \frac{Q}{L_{best}}, L \in L_{best} \\ -\frac{Q}{L_{worst}}, L \in L_{worst} \\ 0, other \end{cases} \tag{13}$$

In Equations (12) and (13), m is the number of ants, ρ is the evaporation coefficient of the global information system, τ_{ij}^k is the pheromone left by ant k on (i,j), Q is the pheromone quality coefficient, L_{best} is the shortest path, and L_{worst} is the longest path.

2.2.3. IHACA-COpSPIEL Algorithm

(1) Clustering

Using the local parameter αr chaotic sequence pair of Equation (9) randomly divides the position set V into small clusters of diameter αr, where $\alpha \in (0,1]$. The nodes near their cluster boundaries are stripped, so the clusters are well separated. The locality of F makes the clusters almost independent and provides a wealth of information [13].

(2) Establishing module approximation

In the ith cluster (C_i), a greedy algorithm is used to obtain the ranks of $g_{i,1}, g_{i,2}, \ldots, g_{i,n_i}$ on the ith cluster's nodes (n_i), and the nodes are connected in this order to form a chain of clusters. A module approximation graph G' is created from G through these chains. A modular directional arithmetic algorithm is used on G' to solve the corresponding objective function, the selected path in G' is extended according to the corresponding shortest path in G, and the solution set A' is output.

(3) Select the next position

The initial node of A' is used as the initial value of the Improved Heuristic Ant Colony Algorithm (IHACA). The IHACA algorithm selects the next position from the first node according to Equation (14), and adds the selected position to the taboo table or $tabu_k$ of ant k. η_{ij}^β is calculated by Equation (10), and τ_{ij} is calculated by Equation (15).

$$P_{ij}^k = \frac{\tau_{ij}^\alpha \eta_{ij}^\beta}{\sum\limits_{j \in A} \tau_{ij}^\alpha \eta_{ij}^\beta} \tag{14}$$

$$\tau_{ij} = F(C_i \cup k_j) - F(C_i). \tag{15}$$

In Equation (14), α is the weight of the path, β is the weight of the heuristic information, and τ_{ij} represents the pheromone intensity of the path from the cluster C_i to the k_j sensor.

(4) Pheromone update

After the next position is determined, the pheromone traversed by the ant (i, j) is updated according to Equation (11). When all ants reach the endpoint, the global pheromone is updated according to Equation (12), and the tabu list is cleared.

The pseudo-code of the IHACA-COpSPIEL algorithm is as follows. Line 2 calls clustering with complexity $O(m_{c\,\max})$. Lines 3–6 form a chain with complexity $O(NlogN)$. Line 7 calls the block-oriented algorithm with complexity $O(nlogN)$ where n is the number of connectable edges of graph nodes ($n \leq N$). Lines 8–10 select nodes for A'' of the greedy algorithm with complexity $O(log(N - n_i))$. Lines 13–17 reach the given maximum mutual information with complexity $O(kN^2)$ where k is iteration times. Line 22 calls updating global pheromone with complexity $O(N^2)$. The computational complexity of Algorithm 1 is $O(kN^2)$ approximately.

Algorithm 1 Improved Heuristic Ant Colony Algorithm-Chaos Optimization of Padded Sensor Placements at Informative and cost-Effective Locations (IHACA-COpSPIEL).

Input: Position set V and covariance matrix
Output: Solution set A
1: Initialize parameters: α, β, w, ξ, τ_0, ε, $n_{\max}$
2: Divide V into m_{cmax} clusters $\{C_i | i \in [1, m_{c\,\max}]\}$
3: **for** each cluster C_i **do**
4:    Sort position points in C_i by greedy algorithm and then get the ranks of $g_{i,1}, g_{i,2}, \ldots, g_{i,n_i}$
5:    Connect $g_{i,1}, g_{i,2}, \ldots, g_{i,n_i}$ to form a chain which is then included into G_i'
6: **end**
7: Uses G' as input of block-oriented algorithm to solve $F(A)$ and then get the solution A'',
    where $G' = \{G_i' | i \in [1, m_{c\,\max}]\}$
8: **while** a given maximum mutural information in A'' is not reached **do**
9:     Select nodes for A'' with greedy algorithm
10: **end**
11: $A' = A''$
12: **for** n=1:n_{max} **do**
13:    **while** a given maximum mutural information in A_n' is not reached **do**
14:       Select IHACA initial points in A_n' from head nodes in A'
15:       Select next point with Equation (14)
16:       Update local pheromone $\tau_{ij}(n)$ with Equation (11)
17:    **end**
18:    **if** $C(A_n') \leq B$ **then**
19:       $A = A_n'$ and output A
20:       **return**
21:    **end**
22:    Update global pheromone $\tau_{ij}(n)$ with Equation (12)
23: **end**
24: $A = A_n'$ and output A
25: **return**

---

3. Routing Protocols for Wireless Sensor Networks

3.1. Communication Model

The energy consumption of data sent by sensor nodes is shown in Equation (16).

$$E_{TX}(k,d) = E_{elec}(k) + E_{amp}(k,d) = \begin{cases} kE_{elec} + kE_{fs}d^2, d < d_0 \\ kE_{elec} + kE_{mp}d^4, d \geq d_0 \end{cases}, \tag{16}$$

where k is the number of bits of transmitted data, d is the transmission distance, $E_{elec}(k)$ is the energy consumption of the transmitting circuit to send k bit data, and $E_{amp}(k,d)$ is the transmission power amplifier transmitting k bit data when the transmission distance is d. E_{elec} is the unit energy consumption of the transmitting or receiving circuit, and d_0 is the threshold. E_{fs} is the energy consumption parameter of the transmission power amplifier under the free space channel model and E_{mp} is the energy consumption parameter of the transmission power amplifier under the multipath fading channel model.

The calculation of the energy consumption of the receiving circuit to receive k bit data is shown in Equation (17).

$$E_{RX}(k) = E_{TX-elec}(k) = kE_{elec}. \tag{17}$$

3.2. Optimal Clustering

The number of cluster heads has a great impact on network performance. According to [14], the optimal number of cluster heads is shown in Equation (18).

$$k_{opt} = \frac{\sqrt{N_A \cdot E_{fs}}}{\sqrt{2\pi \cdot E_{mp}}} \frac{M}{d_{toBS}^2}. \tag{18}$$

In Equation (19), N_A is the number of nodes in set A, M is the area side length, and d_{toBS} is the distance from the node to the base station.

The probability of a node being elected as a cluster head is shown in Equation (19).

$$p = \frac{k_{opt}}{N_{rounds}}. \tag{19}$$

3.3. Fitness Function

The fitness values are based on parameters used to achieve the best solution. It considers intra-cluster compactness, inter-cluster separation and total energy consumption.

Tightness refers to the internal distance, that is, the distance between the nodes in the cluster and the cluster head (CH).

$$C = \sum_{i=1}^{CHs} \sum_{\forall n \in C_i} d(n, CH_i). \tag{20}$$

Separability refers to the distance between clusters, that is, the minimum distance between cluster heads.

$$S = \min_{\forall C_i, C_j, C_i \neq C_j} (d(CH_i, CH_j)). \tag{21}$$

The total energy consumption refers to both the cluster head communication energy consumption E_{CH} and ordinary node communication energy consumption E_{NN}, of which the energy consumption of the cluster head includes the energy consumption E_{RN} required to receive data sent by the nodes in the cluster, the energy consumption E_{HD} required to collect data for fusion, and the energy consumption E_{TB} required to send data to the base station. The energy consumption of ordinary nodes includes the

energy consumption E_{TH} required to send data to the cluster head. Assume that the total number of nodes is N_A, the number of cluster heads is m, and the ordinary nodes in each cluster are $n_1, n_2, ..., n_m$.

$$E_{CH} = E_{RN} + E_{HD} + E_{TB} = kE_{elec}n_i + kE_{DA}(n_i + 1) + (kE_{elec} + kE_{mp}l_{HB}^4). \qquad (22)$$

In Equation (16), E_{DA} is the energy consumed by unit bit data fusion, and l_{HB} is the distance between the cluster head and the base station.

$$E_{NN} = E_{TH} = kE_{elec} + kE_{fs}l_{NH}^2. \qquad (23)$$

In Equation (17), l_{NH} is the distance between the nodes in the cluster and the cluster head.

$$E_{total} = \sum_{i=1}^{m} E_{CH_i} + \sum_{j=1}^{N} E_{NN_j}. \qquad (24)$$

In Equation (18), the closer the distance between the cluster nodes and the cluster head in a cluster, the better. The greater the separation between cluster heads, the better the total energy consumption. The fitness function is as follows.

$$F = w_1 C + \frac{w_2}{S} + w_3 E_{total}, w_1, w_2, w_3 \in (0,1). \qquad (25)$$

In Equation (19), $w_1 + w_2 + w_3 = 1$.

3.4. Routing Protocol Based on BBO Algorithm

BBO algorithm is an information intelligence heuristic algorithm first proposed by Dan Simon in 2008. The habitats of biological populations have their corresponding Habitat Suitability Index (HSI), which is used to describe the quality of the habitat environment, and the factors that affect the fitness index are called Suitable Index Variables (SIVs). The BBO algorithm has the advantages of simple operation, fast convergence, and fewer parameters [28]. The standard BBO algorithm uses a simple linear migration model, but in the real biogeographic environment, species migration often occurs randomly and does not follow this rule. Complex and natural migration models are much better than simple migration models [23,29]. In this paper, a cosine migration model is used. When the number of species in the habitat is either large or small, the change in the immigration rate λ and the emigration rate μ is relatively stable. When the number of species in the habitat is neither large nor small, the immigration rate λ and the emigration rate μ changes quickly. The expression of the cosine migration model is shown in Equations (26) and (27).

$$\lambda_k = \frac{I}{2}(\cos(\frac{k\pi}{n}) + 1) \qquad (26)$$

$$\mu_k = \frac{E}{2}(-\cos(\frac{k\pi}{n}) + 1). \qquad (27)$$

In Equations (26) and (27), I is the maximum value of the immigration rate, E is the maximum value of the emigration rate, k is population number and n is the maximum population number.

The mutation operator provides a certain global search capability for the algorithm through the mutation of the habitat's own information.

$$m_i = m_{\max}(1 - \frac{p_s}{p_{\max}}). \qquad (28)$$

In Equation (28), m_{max} is the maximum mutation rate, p_s is the probability that habitat i has s species, and $p_{max} = max(p_s)$.

The steps of optimizing wireless sensor network routing protocol based on the BBO algorithm

are as follows. Lines 3–7 select CH with complexity $O(n^2)$. Lines 9–31 reach the minimum fitness value with complexity $O(qn^2)$ where q is the number of iterations. Lines 10–20 calculate the migration rate with complexity $O(n^2)$. Lines 21–28 calculate the mutation rate with complexity $O(n)$ and lines 32–35 calculate with complexity $O(m_2)$ where m_2 is the number of nodes ($m_2 \geq n$). The computational complexity of Algorithms 2 is $O(qn^2)$ per round approximately.

Algorithm 2 Biogeography-Based Optimization (BBO)-based routing protocol process.

Input: node coordinates, energy model
Output: residual energy per round, number of dead nodes, number of surviving nodes
1: Initialize parameters: number of habitats n, maximum emigration rate E, maximum immigration rate I,
    probability of species number for each habitat p_s, maximum number of species p_{max},
    maximum number of rounds $round_{max}$
2: **for** l = 1: $round_{max}$ **do**
3:    **for** j = 1: n **do**
4:       Select CH according to Equation (19)
5:       Initialize population randomly
6:       Calculate the fitness value of habitat j according to Equation (25)
7:    **end**
8:    Keep habitat with the smallest fitness values as elite habitat
9:    **while** habitat does not reach minimum fitness value **do**
10:       **for** k = 1: n **do**
11:          Calculate the migration rate λ_k according to Equation (26)
12:          **If** λ_k is greater than a uniformly distributed pseudo random number in [0,1] **then**
13:             **for** t = 1: n **do**
14:                Calculate the migration rate μ_t according to Equation (27)
15:                **If** μ_t is greater than a uniformly distributed pseudo random number in [0,1] **then**
16:                   The roulette selection method is used to select the population to move out of the habitat t
                      and move into the habitat k
17:                **end**
18:             **end**
19:          **end**
20:       **end**
21:       **for** i = 1: n **do**
22:          **If** Habitat i is not an elite habitat **then**
23:             Calculate the mutation rate m_i according to Equation (28)
24:             **If** m_i is greater than a uniformly distributed pseudo random number in [0,1] **then**
25:                Select population mutations in habitat i randomly
26:             **end**
27:          **end**
28:       **end**
29:       Calculate fitness value
30:       Replace the worst habitats with elite habitats
31:    **end**
32:    Calculate the shortest distance from ordinary nodes to CH
33:    Calculate the energy consumed by ordinary nodes to CH to transmit and receive data packets
34:    Calculate the energy consumed by CH to sink nodes to transmit and receive data packets
35:    Calculate the remaining energy, dead nodes, and surviving nodes of the sensor network
36:    **If** All network nodes are dead **then**
37:       **return**
38:    **end**
39: **end**
40: **return**

---

4. Experimental Verification

4.1. Parameter Settings

One of the application backgrounds of this research is for forest and grassland fire risk monitoring and early warning. The precision and accuracy of forest and grassland fire risk monitoring depend

on the multi-source and space-time data which are collected under the high-resolution sensor layout. How to deploy as few sensors as possible to monitor a large area of forests and grasslands and how to measure such multi-source parameters related to fire risk warning as atmospheric temperature, light, soil temperature and humidity, wind speed and rainfall are the application problems that need to be solved. To meet the application requirements, this paper takes the distance between and among the wireless sensor nodes into consideration. The service time and life of the whole network are extended to the maximum extent under such constraints as communication energy consumption and node distances. In order to verify the comprehensive performance of the algorithm proposed in this paper, we conducted simulation experiments and comparisons among the four algorithms, i.e., the greedy algorithm, the pSPIEL algorithm, the ant colony algorithm, and the improved heuristic ant colony algorithm. The BBO-based routing protocol is used for data transmission. In this paper, the forest environment monitoring area is separated into $|V| = N = 86$ locations, and a subset is selected to deploy sensors. The experimental parameter settings are shown in Table 1. We explain the experimental parameter settings as follows.

In Algorithm 1, α is a parameter in Equation (14). If α is large, it would make the ant search according to the pheromone and fall into the local minimum easily, whereas if α is small and its value is 0.1, it would increase the randomness of the search. For the same reason, β in Equation (14), ξ in Equation (11), and ρ in Equation (12) are all 0.1. w is the weight parameter of the heuristic function in Equation (10). To achieve equilibrium of the distance effects from node j to node i and from node j to head of adjacent cluster on the heuristic function, w takes the value of 0.5. Q is a parameter in Equation (13) and its value is 1 in order to strengthen the positive feedback mechanism of the algorithm. τ_0 is a parameter in Equation (11) and it takes a small value of 0.0003 to increase the probability for ant to choose an optimal path. ε is a parameter in Equation (11), which takes constant value 1. n_{max} is greater than N.

In Algorithm 2, p_s is a parameter in Equation (28). The smaller the p_s value is, the more it is prone to mutation. Hence the p_s value is 0.1. p_{max} is a parameter in Equation (28). It is the maximum value of p_s and it takes a value of 1. I is a parameter in Equation (26) and E is a parameter in Equation (27). In order to make both the immigration rate and the emigration rate take value at [0,1], both I and E take value 1. E_0, E_{elec}, E_{mp} and E_{fs} take the commonly used default values. The value $round_{max}$ depends on the lifetime of all nodes of the network.

Table 1. Parameter settings.

Parameter	Description	Value
N	Total of location number for possibly-deployed sensors	86
α	Path weight	0.1
β	Heuristic information weight	0.1
ξ	Local pheromone evaporation coefficient	0.1
ρ	Global information system evaporation coefficient	0.1
w	Weight	0.5
Q	Pheromone quality coefficient	1
τ_0	Pheromone under initial conditions	0.0003
ε	Constant	1
p_s	Probability that habitat i has species s	0.1
p_{max}	Maximum of p_s	1
I	Maximum immigration rate	1
E	Maximum emigration rate	1
E_0	Sensor node's initial energy	0.5 J
E_{elec}	Energy consumption per bit for transmitting data	50 nJ/bit
E_{mp}	Energy consumption of multipath model transmitter	0.013 pJ/bit/m^4
E_{fs}	Energy consumption of free space model transmitter	10 pJ/bit/m^2
$round_{max}$	Maximum round number	2500
n_{max}	Maximum iteration number	100

4.2. Results and Analyses

(1)   Comparison of communication cost and the number of sensors

Table 2 is a set of experimental data of the simulation. Tables 2 and 3 compare the greedy algorithm, the pSPIEL algorithm, and the IHACA algorithm, respectively. The algorithm proposed in this paper not only meets the same deployment requirements but also achieves the best results, that is, the lowest communication cost and the number of sensors. With the same amount of mutual information and compared with the greedy algorithm, the pSPIEL algorithm, and the IHACA algorithm, the communication cost of the IHACA-COpSPIEL algorithm is also the lowest. When mutual information is 0.16, sensor deployment has a high-cost performance, and the communication cost of the IHACA-COpSPIEL algorithm is 38.42% lower than that of the greedy algorithm, 24.19% lower than that of the pSPIEL algorithm, and 8.31% lower than that of the IHACA algorithm. The sensor deployments of the four algorithms are shown in Figure 1, where the blue dots indicate the possibly-deployed points and the red squares indicate the selected points for deployment.

The r value of the pSPIEL algorithm is randomly selected and the number of clusters is also random. Thus the number of clusters affects the selection of nodes and it is difficult to obtain the optimal number of clusters. Therefore the communication cost is high. The IHACA-COpSPIEL algorithm adds a chaotic operator, which can traverse the local parameter r value to obtain the number of clusters under different r values. The deployed nodes are selected within the optimal number of clusters. The heuristic function concerns the distance between the next node and the adjacent cluster head to minimize the communication distance between sensors. The first node of the solution set of the COpSPIEL algorithm is used as the first node of the IHACA. The node with maximum mutual information is selected as the deployment point. This can reduce the number of sensors and reduce the total communication cost, so it has better performance than any other algorithm.

The objective function of sensor deployment is submodular. When mutual information was 0.14 and the number of sensors was small, or when mutual information was 0.20 and the number of sensors was large, a sensor is added each time and the communication cost was the lowest compared with the greedy algorithm, the pSPIEL algorithm, and the IHACA algorithm. The more the amount of mutual information increases as the number of sensors increases, the better the sensor deployment effect. When the number of sensors is small, the increment of submodular benefit is large by adding a new sensor each time. As the number of sensors increases, the increment of submodular benefit starts to decrease whenever a new sensor is added.

Table 2. Comparison of communication cost for the four algorithms.

Mutual Information	Communication Cost			
	Greedy	**pSPIEL**	**IHACA**	**IHACA-COpSPIEL**
0.14	49.22	36.01	39.7	35.42
0.15	59.47	50.4	45.16	40.06
0.16	71.83	58.34	48.24	44.23
0.17	74.31	62.44	60.36	53.26
0.18	78.48	70.43	67.75	64.11
0.19	80.82	78.28	78.41	75.22
0.20	98.57	95.21	97.76	89.05

Table 3. Comparison of sensor number for the four algorithms.

Mutual Information	Sensor Number			
	Greedy	pSPLIE	IHACA	IHACA-COpSPIEL
0.14	5	6	7	6
0.15	6	8	8	7
0.16	7	10	10	8
0.17	10	13	12	10
0.18	12	14	13	12
0.19	17	25	25	20
0.20	25	30	30	28
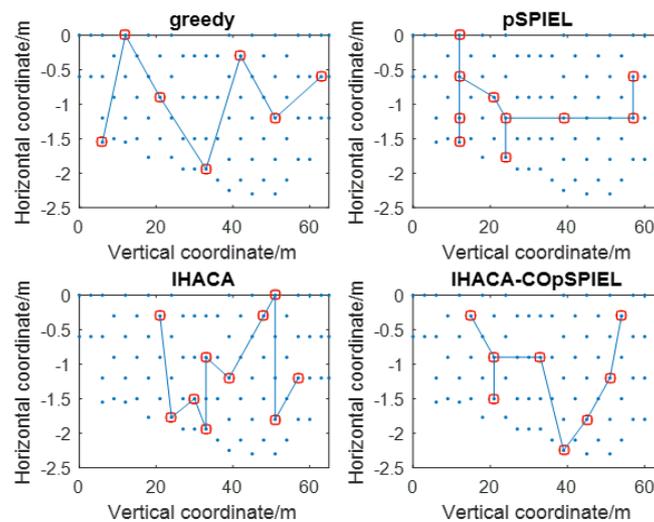
Figure 1. Sensor deployments of the four algorithms.

Figure 2 shows that under the mutual information of 0.14–0.20, the ratio of the IHACA-COpSPIEL algorithm is higher than that of the greedy algorithm, the pSPIEL algorithm, and the IHACA algorithm. Therefore, IHACA-CpSPIEL achieves the best cost-benefit ratio.
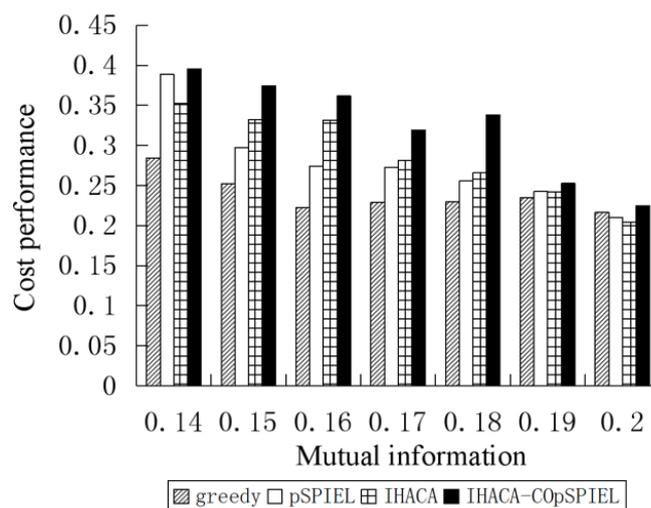

Figure 2. Cost performance and mutual information of the four algorithms.

(2) Comparison of the lifecycle and average energy

The life cycle refers to the time from which the wireless sensor starts to work until the death of the first node. Figure 3 shows the comparison results of the life cycle of the sensor nodes selected by the routing protocol based on the BBO algorithm for the greedy algorithm, the pSPIEL algorithm, the IHACA algorithm, and the IHACA-COpSPIEL algorithm proposed in this paper. It shows from Figure 3 that the first dead node under the greedy algorithm appears in round 1368, the first dead node in the pSPIEL algorithm appears in round 1430, the first dead node in the IHACA algorithm appears in round 1272, and the first dead node of the IHACA-COpSPIEL algorithm proposed in this paper appears in round 1681, which indicates that the wireless sensors deployed in the IHACA-COpSPIEL algorithm have a longer life cycle. The reason for this is that the communication distance of the wireless sensors deployed by the IHACA-COpSPIEL algorithm is the shortest, which reduces the energy consumption of transmission. Figure 4 shows a comparison of the remaining energy percentages of data transmission after the sensors are deployed in each of the four algorithms. The IHACA-COpSPIEL algorithm has a higher percentage of remaining energy per round than the greedy, pSPIEL, and IHACA algorithms. Therefore, the overall energy consumption of the proposed algorithm is lower than any of the other three algorithms.
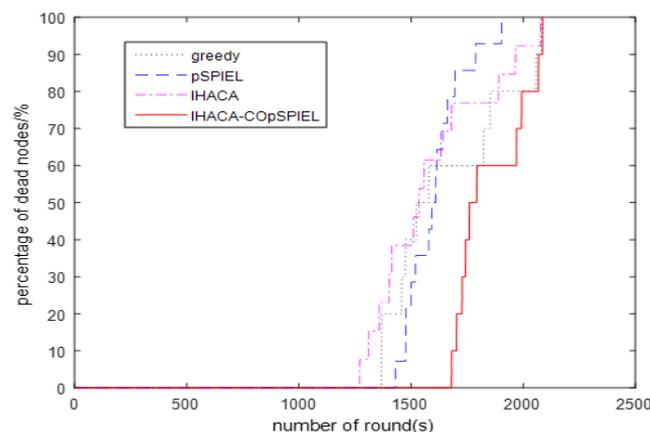
Figure 3. Comparison of the percentage of dead nodes in the four algorithms.
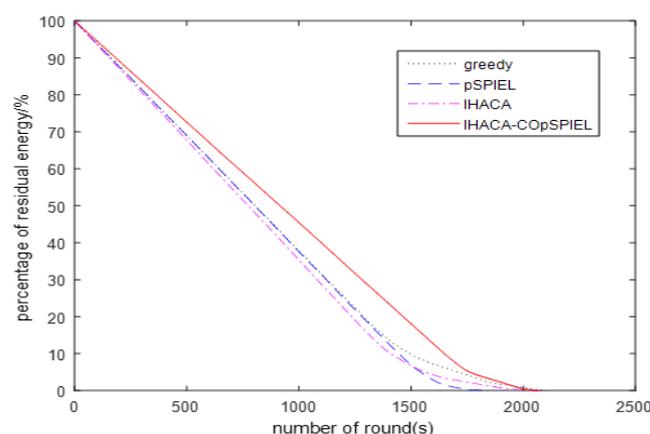
Figure 4. Percentage of remaining energy in the four algorithms.

As the number of simulation rounds increases, Figures 5 and 6 show both the deaths of nodes by using the LEACH routing protocol and the BBO routing protocol, respectively, and the network energy consumption results by using the LEACH routing protocol and the BBO routing protocol, respectively, to deploy sensors in the IHACA-COpSPIEL algorithm. As is seen from Figure 5, under the LEACH

protocol, the first node died in round 1435 and the last node died in round 1584, and under the BBO routing protocol, the first node died in round 1659 and all nodes died in round 2071. The network survival time is longer than the former. As is seen from Figure 6 that the remaining energy under the routing protocol based on the BBO algorithm has always been more than that under the LEACH protocol. This is because the routing protocol based on the BBO algorithm fully takes the cluster into account. The distance between the nodes and the cluster heads, the distance between the cluster heads and the cluster heads, and the total energy consumption effectively balances the network load. Hence the life of the entire wireless sensor network is significantly extended.
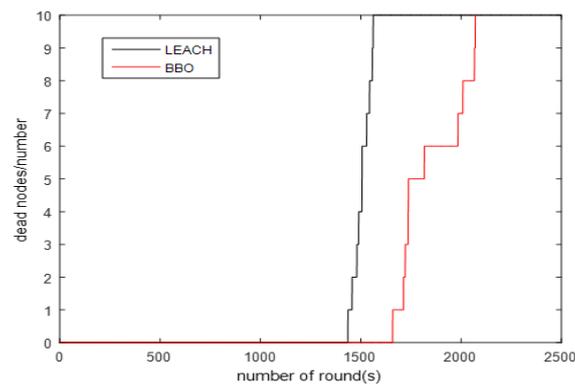

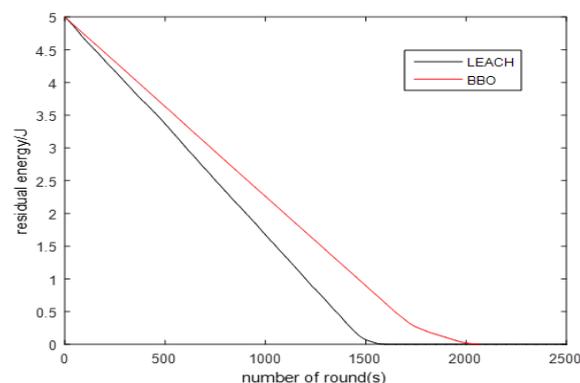**Figure 5.** Comparison of dead nodes under two protocols.

Figure 6. Comparison of residual energy under two protocols.

5. Conclusions

In order to reduce costs and save energy, this paper proposes a large-scale sensor deployment method called the IHACA-COpSPIEL algorithm and a routing protocol based on the BBO algorithm. Mutual information is introduced to describe the correlation between observed points and unobserved points, a mathematical model with submodularity is established, and the edges of graph theory are used to represent communication costs. The pSPIEL algorithm with enhanced optimization ability by a chaos operator and the ant colony algorithm with improved heuristic function and pheromone update mechanism are used to find the optimal path. What has been studied can further solve the sensor deployment problem under the constraint of communication cost. Finally, the BBO algorithm-based routing protocol transmits data to the deployed sensors. The computational complexity of the IHACA-COpSPIEL is $O(kN^2)$, and the computational complexity of the routing protocol based on the BBO algorithm is $O(qn^2)$. The experiments show that the deployment algorithm proposed in this paper has better sensor deployment capabilities. This deployment algorithm reduces the communication cost by 38.42% compared with the greedy algorithm. It also reduces the number of sensors and has a longer life cycle. Compared with the LEACH protocol, the BBO algorithm-based routing protocol has

lower energy consumption and longer network life.

In the future, we intend to use a discrete event simulator (DES) such as NS-3 to further combine practical application scenarios to improve the effectiveness of the algorithm. Our vision for future work is as follows.

We will complete the IHACA-COpSPIEL protocol design in the NS-3. We will refer to the RFC document of Multi-Protocol Label Switching protocol, and elaborate on the design and implementation of each basic component of IHACA-COpSPIEL, including the forwarding equivalence class (FEC), next hop label forwarding entry (NHLFE), FEC to NHLFE mapping (FTN), etc. By statically configuring the label forwarding table, the communication between private networks through the backbone network by IHACA-COpSPIEL forwarding will be realized.

Author Contributions: Conceptualization, Y.W., Y.D.; methodology, Y.W., L.W.; software, Y.W., Y.D.; visualization, Y.D., W.D.; investigation, Y.D., W.D.; validation, Y.D., Q.C.; data curation, Q.C., L.W.; original draft preparation, W.D., Q.C.; supervision, Y.W. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

BBO	Biogeography-Based Optimization
IHACA	Improved Heuristic Ant Colony Algorithm
pSPIEL	padded Sensor Placements at Informative and cost-Effective Locations
IHACA-COpSPIEL	Improved Heuristic Ant Colony Algorithm-Chaos Optimization of padded Sensor Placements at Informative and cost-Effective Locations
SIVs	Suitable Index Variables

References

1. Hussein, A.A.; Khalid, R.A. An Improved Energy Efficient Chain-Based Routing Protocol Using Aco. *Iraqi J. Inf. Commun. Technol.* **2019**, *2*, 11–29. [CrossRef]

2. Wang, Y.; Wang, L.; Chang, Q.; Yang, C. Effects of direct input–output connections on multilayer perceptron neural networks for time series prediction. *Soft Comput.* **2019**, 1–10. [CrossRef]

3. Aitsaadi, N.; Achir, N.; Boussetta, K.; Pujolle, G. Multi-objective WSN deployment: Quality of monitoring, connectivity and lifetime. In Proceedings of the IEEE International Conference on Communications, Cape Town, South Africa, 23–27 May 2010; pp.1–6.

4. Debbagh, M. Development of a Low-Cost Wireless Sensor Network for Passive in situ Measurement of Soil Greenhouse Gas Emissions. Ph.D. Thesis, McGill University, Montrea, QC, Canada, 2019.

5. Kaddi, M.; Benahmed, K.; Boussetta, K.; Omari, M. A Novel Communication Mode for Energy-Efficient Based Chain in Wireless Sensor Networks. In *International Conference in Artificial Intelligence in Renewable Energetic Systems*; Springer: Cham, Switzerland, 2019; pp. 397–407.

6. Caselton, W.F.; Zidek, J.V. Optimal monitoring network designs. *Stat. Probab. Lett.* **1984**, *2*, 223–227. [CrossRef]

7. Niazadeh, R.; Roughgarden, T.; Wang, J. Optimal algorithms for continuous non-monotone submodular and dr-submodular maximization. In *Advances in Neural Information Processing Systems*; Springer: Cham, Switzerland, 2018; pp. 9594–9604.

8. Ilves, K.; Harnefors, L.; Norrga, S.; Nee, H.P. Predictive sorting algorithm for modular multilevel converters minimizing the spread in the submodule capacitor voltages. *IEEE Trans. Power Electron.* **2014**, *30*, 440–449. [CrossRef]

9. Krause, A. Optimizing Sensing: Theory and Applications. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA, USA, 2008.

10. Huang, C.F.; Tseng, Y.C. The coverage problem in a wireless sensor network. *Mob. Netw. Appl.* **2005**, *10*, 519–528. [CrossRef]

11. Guestrin, C.; Krause, A.; Singh, A.P. Near-optimal sensor placements in gaussian processes. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 7–11 August 2005; pp. 265–272.

12. Cheng, H.; Su, Z.; Xiong, N. Energy-efficient node scheduling algorithms for wireless sensor networks using Markov Random Field model. *Inf. Sci.* **2016**, *329*, 461–477. [CrossRef]

13. Krause, A.; Guestrin, C. Near-optimal Nonmyopic Value of Information in Graphical Models. In Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence, Edinburgh, UK, 26–29 July 2005; pp. 324–331.

14. Coutino, M.; Chepuri, S.P.; Leus, G. Submodular sparse sensing for gaussian detection with correlated observations. *IEEE Trans. Signal Process.* **2018**, *66*, 4025–4039. [CrossRef]

15. Xia, Y.; Wang, J. A Bi-Projection Neural Network for Solving Constrained Quadratic Optimization Problems. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *329*, 214–224. [CrossRef]

16. Zhang, S.; Xia, Y.; Wang, J. A complex-valued projection neural network for constrained optimization of real functions in complex variables. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 3227–3238. [CrossRef]

17. Liu, G.; Huang, X.; Guo, W.; Niu, Y.; Chen, G. Multilayer obstacle-avoiding x-architecture steiner minimal tree construction based on particle swarm optimization. *IEEE Trans. Cybern.* **2015**, *45*, 989–1002. [PubMed]

18. Guo, W.; Li, J.; Chen, G.; Niu, Y.; Chen, C. A PSO-optimized real-time fault-tolerant task allocation algorithm in wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **2015**, *26*, 3236–3249. [CrossRef]

19. Liu, X.; Zhao, S.; Liu, A.; Xiong, N.; Vasilakos, A.V. Knowledge-aware Proactive Nodes Selection approach for energy management in Internet of Things. *Future Gener. Comput. Syst.* **2019**, *92*, 1142–1156. [CrossRef]

20. Heinzelman, W.R.; Kulik, J.; Balakrishnan, H. Adaptive protocols for information dissemination in wireless sensor networks. In Proceedings of the 5th annual ACM/IEEE International Conference on Mobile Computing and Networking, Seattle, WA, USA, 15–20 August 1999; pp.174–185.

21. HK, D.W.; Kurniawan, A.; Arifianto, M.S. Improving Topology of LEACH Cluster Using Reinforcement Learning Method. In Proceedings of the IEEE International Conference on Sensors and Nanotechnology, Penang Island, Malaysia, 24–25 July 2019; pp.1–4.

22. Daanoune, I.; Baghdad, A.; Balllouk A. BRE-LEACH: A New Approach to Extend the Lifetime Of Wireless Sensor Network. In Proceedings of the Third International Conference on Intelligent Computing in Data Sciences (ICDS), Marrakech, Morocco, 28–30 October 2019; pp. 1–6.

23. Simon, D. Biogeography-Based Optimization. *IEEE Trans. Evol. Comput.* **2008**, *12*, 702–713. [CrossRef]

24. Nomosudro, P.; Mehra, J.; Naik, C.; Pushparaj, S.D. ECABBO: Energy-efficient clustering algorithm based on Biogeography optimization for wireless sensor networks. In Proceedings of the IEEE Region 10 Conference (TENCON), Kochi, India, 17–20 October 2019; pp. 828–834.

25. Pal, R.; Pandey, H.M.A.; Saraswat, M. BEECP: Biogeography optimization-based energy efficient clustering protocol for HWSNs. In Proceedings of the Ninth International Conference on Contemporary Computing (IC3), Noida, India, 11–13 August 2016; pp.1–6.

26. Wang, Z.; Xie, H.; He, D.; Chan, S. Wireless Sensor Network Deployment Optimization Based on Two Flower Pollination Algorithms. *IEEE Access* **2019**, *7*, 180590–180608. [CrossRef]

27. Wang, Z. Application of an Improved Ant Colony Algorithm in Coastal Tourism Route Optimization. *J. Coast. Res.* **2019**, *98*, 84–87.

28. Kaveh, M.; Mesgari, M.S. Improved biogeography-based optimization using migration process adjustment: An approach for location-allocation of ambulances. *Comput. Ind. Eng.* **2019**, *135*, 800–813. [CrossRef]

29. Saligheh, S.; Arasteh, B. Quality of Service Aware Service Composition Method Using Biogeography-Based Optimization (BBO) Algorithm. *NASHRIYYAH -I MUHANDISI -I BARQ VA MUHANDISI -I KAMPYUTAR -I IRAN, B- MUHANDISI -I KAMPYUTAR* **2019**, *17*, 190–198.