Placement optimization in sensors network using a linear discernment analyses-based clustering for smart greenhouse.

BENTOUMI Adel
Department of Electrical Engineering
Laboratory of Electronics and New Technologies –LENT-
University of Larbi Ben M'hidi, Oum El Bouaghi
Oum El Bouaghi, Algeria
bentoumi.adel@univ-oeb.dz/Bentoumi.adel@outlook.com

BELHANI Ahmed
Department of electronics
Université Frères Mentouri - Constantine 1
Constantine, Algeria
ahmed.belhani@umc.edu.dz

***Abstract*—The location of the sensors is generally determined based on the experience of greenhouse producers and designers. To monitor accurately the interior environment of a greenhouse, the installation locations of sensors must be carefully chosen. In this study, we propose a linear discriminant analysis (LDA) which chooses an optimal configuration of sensors for a particular application from a whole set of Networks. The proposed method finds the direction w via the LDA such that when data are projected onto this direction, The scheme aims to split the large network into irregular subareas, regarded as blocks. In each of such blocks, the nodes measurements are as uniformly correlated as possible in placement from each cluster.**

Keywords: IoT, Sensors, Clustering, OSP, Smart greenhouse, LDA.

I. INTRODUCTION

Smart agriculture or smart farming is regarded as an important axis and priority that can be developed and implemented to ensure the food safety.

Based on information and communication technology (ICT) capabilities and IoT, environmental parameters can be measured and communicated in order to supervise and control internal parameters microclimate (example: greenhouse) like temperature, humidity, CO_2 under some constraints like solar radiation, UV, CO_2 concentration and wind speed.

In large greenhouses it is difficult to control the internal environment in a consistent and appropriate manner using information technology and communication (ICT) , , so a communication between sensors must be guaranteed by IoT in order to ensure a best control and supervision quality, furthermore, the sensor placement must be optimized in smart agriculture system.

The arrangement of the sensors is a crucial factorthat must be given proper attention, this arrangement creates a layout notion that means a physical topology. Due to the dynamic nature of greenhouse environment where parameters vary spatially and temporally [1], the plant will grow over time and eventually affect the performance of the sensors. A large greenhouse itself contains many microclimatic zones. It can have heterogeneous areas where parameters differ from surrounding areas and the global environment within a single greenhouse. These microclimates exist both horizontally and vertically in a greenhouse. Thus, monitoring the parameters requires inconsistent deployment of sensors. Irrigation and fertilization methods can also affect the sensors location decision. The arrangement of the sensors in a greenhouse has been broadly categorized as horizontal and vertical arrangement [1].

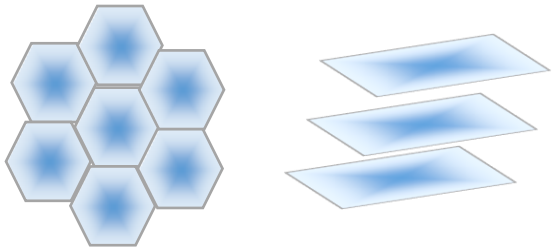

Fig. 1.Tessellations.

Several techniques and methods are used to place the sensors, Javier López et José [2]. the sensors formed a grid of 12 measuring stations was distributed along the greenhouse to record data. Each station is composed of 9 sensors, placed at 0.23 m, 0.93 m and 1.56 m from the ground, (air temperature, black globe temperature, air velocity and relative humidity -117 sensors at all). In addition, another measurement station was placed outside the greenhouse to measure the outdoor climatic conditions.

Konstantinos and P.Ferentinos [3] have adopted the sensors nodes strategy, with sensors are placed symmetrically at the level of the canopy (plant) (1.5 m high) and second four nodes «in caisson» are placed at the four corners of the 20 m by 8 m greenhouse, 3 lanes from the sides, at a height of about 1.8 m. The main goal of the experiments was to investigate the accuracy and reliability of measurements as influenced by sunlight inside the greenhouse.

Marco Mancuso and Franco Bustaffa [1] have proposed asensor network as a grid of wireless nodes organized in a 20 by 50 meters greenhouse of

tomatoes. 6 nodes were used and arranged in two rows 12,5 m apart. The mutual distance between the row nodes is 6.5 m.. The nodes are placed at the bottom of the field with an average height of 0,25 m. A similar configuration developed by Ferentinos and Tsiligiridis [2] they made and implemented to cover larger greenhouses a network of 30 by 30 units using 900 sensors placed on all junctions' blocks.

For Min-Sheng Liao and Shih-Fang Chen [3], the wireless imaging platform based on the IoT is built using 12 wireless camera modules, each of which includes a 5-megapixel sensor camera . The IoT-based environmental monitoring system consists of 12 nodes of wireless sensor and gateway.

Because all crops grow vertically, either upwards or downwards, the sensors are not deployed at ground level, resulting in a vertical arrangement system. The vertical arrangement is of particular importance for greenhouse climate control. The growth and foliage of the crops significantly affect the communication range of the sensor, so the vertical layout seems to be one of the most important solutions. This survey classifies all variants of vertical arrangements in a typical greenhouse, while other authors suggested placement the sensors. K. Mesmoudi and A. Soudani [4] used the model to monitor parameters at the canopy level and above the canopy level of the crop. In another variant of vertical layout, the sensors were positioned on the ground and only the coordinators were placed at a highier level or in the center in the case of a single coordinator Marco Mancuso and Franco Bustaffa [1]. Hybrid of both layouts can also prove to be quiet promising in both the monitoring accuracy and the number of sensors, cropping and the measured parameter is an important factor. Although the topology of the grid seems simple to deploy, it can be about deployment at the periphery and wasting nodes sometimes in a greenhouse environment.

Overlap is a serious problem in the layout of the square grid, few environmental parameters do not change in a few meters in a greenhouse.

Thus, for the measurement of parameters such as temperature, humidity and brightness, the grid layout does not work optimally. For parameters with a range of variability of few meters such as soil temperature, soil moisture, soil pH the grid layout is suitable but requires more sensors [5].

In this work, the main goal is to design an efficient clustering and energy optimization network and placement of sensors to estimate the optimal number of clusters. We illustrate an optimal algorithm for clustering the sensor nodes such that each cluster (which has a master) is balanced and the total distance between sensor nodes and master nodes is minimized. Then, we conceive an efficient clustering scheme, based on Linear Discriminant Analysis (LDA) technique thus exploiting the correlation between clusters.

The remainder of this paper is organized as follows, Section II describes the deferent methods of optimal sensors placement. Section III we explain LDA optimization approach for clustering layout senors and the problem statement where the adopted scenario and the considered assumptions will be presented. Section IV we present sensor layout based on LDA approach for a smart greenhouse and we present the simulation results and discussions, section V offers our conclusion and future work.

II. Sensor Placement Optimiszation Methods

The issue of optimal sensor placement (OSP) in modal testing has attracted considerable attention over the decades. A large body of research has been published contributing to two basic components in determining OSP: the endpoint, which involves a specified performance modal test requirements, and the solution strategy, which explains how to find the best OSP for requirement given by Sang-yeon Lee and In-bok Lee [6].And the optimization which was performed using two methods: placement of the sensor based on error and placement based on entropy. Locations of the sensors for which the monitored data was close to the reference, i.e., the mean data of all measurement locations, were selected. Sensor locations influenced by external weather conditions in poor environmental control were selected. Using these methods, to determine optimal sensor locations to represent the entire facility environment and to detect areas with significant changes in air temperature.

Rainald Lohner and Fernando Camelli [7] The method considers the general problem of sensor placement. Assuming a given number of sensors, each version scenario leads to a sensor input. The data recorded from all possible release scenarios at all possible sensor locations identifies the optimal sensor locations. Clearly, if a single sensor is to be placed, it should be at the location that has recorded the greatest number of rejections. This argument can be used recursively by removing from any further consideration all versions already recorded by previously placed sensors.

new approach developed by M. Arnesano and G.M. Revel [8], based on index measurement performance, to help the HVAC engineers optimize air temperature monitoring, in terms of number of sensors and placement in large spaces. The methodological workflow consists essentially of two main steps:

-Generation of a data set characterizing the horizontal distribution of air temperature in space, which is a key climatic condition that must be measurable physically, virtually or both.

-Optimization of the number and placement of sensors along the perimeter of space.

Khairul and Mohd N [9] proposed another method that is more practical to implement for wired and wireless light sensors in small and large buildings, as well as for new buildings and renovation projects.

- Formulation of a new model for calculating the dimming levels of the LED luminaire, which is also incorporated into the PSO algorithm.

Three contributions related to the lighting control strategy are presented as follows:

-Development of a new method for placing light sensors in terms of numbers and positions using the particle swarm optimization algorithm (PSO) to minimize costs (ie, sensors and electrical energy);

-The proposed method offers superior performance in terms of computational effort and optimal solution (i.e., the number and position of light).

Worden and A.P. Burrows [10], choose a defect detection and classification approach using neural networks and combinatorial optimization methods. Given the existence of an effective defect detection procedure, place the sensors for optimum detection efficiency, a neural network is used to locate and classify defects and a variety of methods are applied to determine an optimal (or almost optimal) sensor distribution.

Ting Hua Yi1 and Guang Dong [11] they propose the optimal placement of the triaxial sensor which plays a crucial role in three-dimensional modal identification. To efficiently find the optimal configuration of the triaxial sensor with the proposed three-dimensional optimal criterion, the hierarchical wolf algorithm (HWA) is developed by mimicking the swarm intelligence embedded in the wolf pack.

Dongmin Guo and David Zhang [12] propose a linear discriminant analysis (LDA) based sensor selection technique (LDASS) which chooses an optimal configuration of sensors for a particular application from a whole set of available configurations. The proposed method finds the direction **w** via the LDA such that when data are projected onto this direction, the samples from two classes are as separate as possible. It is found that after projection, the difference of means of the two distinct sample classes can be expressed as the linear combination of the responses of all the sensors in the system, and **w** can be regarded as the weight vectors for these sensors which indicate the contribution weight of each sensor.

For this work we will use a linear discriminant analysis (LDA) to determine the placement of sensors and their cluster in the network.

.

TABLE I. Method of optimal sensor placement

Reference	Domain	Strategies sensors placement		Hauteur	Parameters of measurement	Approach
		Fixe	*Optimal*			
[1]	greenhouse	X		0.25m	-Temperature -Relative humidity	-WSN Réseaux de capteurs sans fil
[13]	Greenhouse		X		-WSN	-Genetic algorithms
[6]	Greenhouse		X	0.9 m	-Temperature	-Error-based method.
						-Entropy-based method
[7]	Building		X		- Contamination	Computational Fluid Dynamics (CFDs) model simulation
[8]	Salle de sport		X		-Temperature	-Zonal model -The mean deviation. -The standard deviation.
						-The outliers period.
						-The Z index.
[9]	Building		X		-La lumière	- Particle swarm optimization (PSO) algorithm.
[10]	Materials		X		-Default detection	- Réseaux de neurones -Méthodes d'optimisation combinatoire
[14]	Greenhouse	X			-Solar radiation - Temperature	Computational Fluid Dynamics (CFDs) model simulation
[15]	Greenhouse	X			-Solar radiation - Temperature	The thermal energy stored in plants and soil is transferred to the greenhouse environment through respiration, convection and radiation, the curtain and roof reduce losses through the upper walls.
[16]	Greenhouse	X		0.23m 0.93m	-Temperature -Ultraviolet	ISO7726 obligation for favorable conditions of humans works.
				1.56m	-humidity	
					-Speed of Wind	
[11]	Génie civil		X		-Position	-L'algorithme de loup hiérarchique (HWA)

III. LDA Optimization Approach For Cluastring Lyaout Senor

The linear discriminant analysis (LDA) is a fundamental data analysis method originally proposed by R. Fisher [17] in 1936 ,for discriminating between different types of Cluster that can be used for supervised or unsupervised learning [18]. It consists in finding the projection hyperplane that minimizes the interclass variance and maximizes the distance between the projected means of the classes. The LDA can determine the optimum direction of projecting ω, such that when projecting onto ω, the samples from two different classes are as separate as possible and the samples from the same classes are as close as possible. The intuition behind LDA. Data samples in two dimensions are projected in a Lower dimension space Fig 2 (line ω). The line has to be chosen so that the projection maximizes the "separability" of the projected samples.

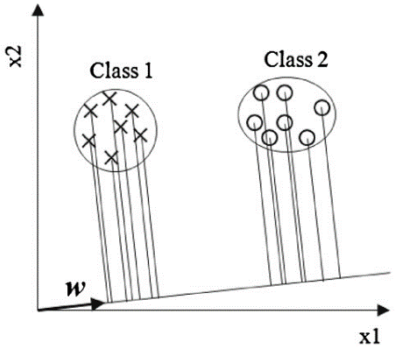

Fig. 2. Data from two classes that are projected onto ω [1].

For Network we consider a large-scale wireless sensor network WSN composed by a large number of sensors, noted *N*, randomly deployed over the monitored area to collect environmental information such as temperature, humidity and solar radiation, as well as soil moisture, which are generally highly correlated both in space and time domains.

For large scale networks, the gathered data volume and needed memory for storage are huge, to reduce the amount of data transmission and then energy consumption, we seek to use a clustering that partitions the considered network into sub-networks, regarded as subareas, by exploiting the inherent data spatial correlation. We here consider WSN with locally uniformly correlated data. After clustering, the network will be partitioned into sub-areas such that the sensor readings in the same spatially contiguous sub-network are uniformly spatially correlated even if the spatial data correlation throughout the network is non-uniform.

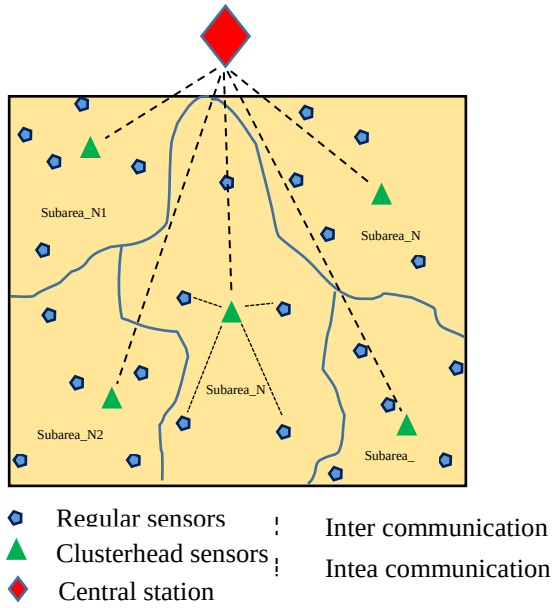

Fig. 3.Network model description.

In the example shown in Fig 3, the network comprises 5 irregular sub-areas with random distribution of nodes. Each region equipped with N_i sensor nodes, where $i = 1, 2., 5$, is presented by a blue line boundary thus designating region with a different uniform data correlation.

A. Assumptions

In our framework, we make the following assumptions:

- The *N* sensor nodes are supposed randomly distributed in a squared monitored area.
- Provided that a sensor may operate as a cluster head transmitting at an appropriate signal range (CH sensor) which allows the communication with the remote base station, or it may operate as a ‘‘regular sensor’’ (RS sensor)
- All sensor nodes know their own geographic location.
- The length unit is defined as the distance between the positions of two neighboring sensor nodes in the horizontal or vertical dimension
- The network topology and the data correlation stay unchanged over the required processing period. In our scenario, sensor nodes are partitioned into clusters, and each cluster has an aggregator node (local sink), represented.

B. Balanced k-Clustering

There are several sophisticated clustering methodologies in the literature of WSNs towards energy saving[19] . However, our work tackles the energy saving issue through the optimization of energy aware clustering in sensor networks nodes. A simple approach of clustering sensors in regular operating modes with their closest CH sensor is adopted for clusters design in the network. The balanced clustering formulation can overcome this drawback by having an upper/lower bound on the size of each cluster, we transform the balanced k-clustering problem to a min-cost flow instance. This instance, can be solved optimally using existing techniques, we use in our work pre-sort with K-means methods [20] for clustering. Suppose we are given N points and we want to group them into k clusters. For the sake of simplicity, we assume that N is a multiple of k. We want to obtain a strictly balanced solution, i.e., each cluster has to contain exactly N/K sensors. After that we apply LDA for optimization methodology for self-organizing, and for validation K optimal Classes we apply five cluster validity Metrics indices for clustering performance evaluation,

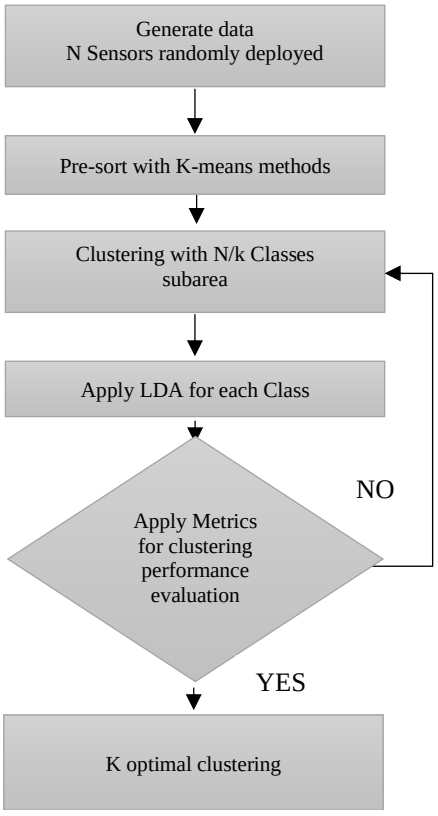

Fig. 4. The working flowchart of Approach K optimal clustering.

Fig 5. demonstrates an example of clustering effect on communication energy consumption. In this figure, black filled points indicate master nodes while the empty circles represent the sensor nodes. Assuming the capacity constraint of 3 for each master node, two sample clustering options, namely AB and C-D have been shown. The total square of distances between master nodes and sensors in clustering A-B is 65 unit, while it is 117 units for clustering C-D. Considering the relation of energy dissipation and this simplified metric (total square of distances), LDA and optimization problem.

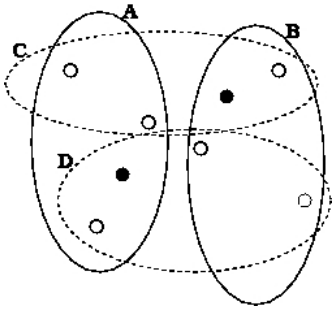

Fig. 5.Partitioning the nodes into clusters A and B leads to a solution dissipating less communication energy compared to clusters C and D [18].

C. Data expression

Assume that we have N sensors samples. They are labeled as classes C_1, C_2,..,C_k, respectively. Each subarea consists of Ni sensors, i.e., S1,.., S_{Ni}, where Ni= N/K our system. each sensor is represented by a discrete by three coordinates 3D (X,Y,Z)(where dimensional d = 3 in our case the ith sensor from class C_1, So the training set including Ni sensors subarea C_1 is expressed as

$$\begin{bmatrix} S_1^1 \\ \vdots \\ S_{Ni}^1 \end{bmatrix} = \begin{bmatrix} x_1^1 & y_1^1 & z_1^1 \\ \vdots & \vdots & \vdots \\ x_{Ni}^1 & y_{Ni}^1 & z_{Ni}^1 \end{bmatrix} \quad (1)$$

Where $S_i^k = [x_i^k \quad y_i^k \quad z_i^k]$ it is sensor i in Cluster k With coordinate (x,y,z).
Same as Eq. (1), the subarea set consisting of Ni sensors from class C_2 is expressed as

$$\begin{bmatrix} S_1^2 \\ \vdots \\ S_{Ni}^2 \end{bmatrix} = \begin{bmatrix} x_1^2 & y_1^2 & z_1^2 \\ \vdots & \vdots & \vdots \\ x_{Ni}^2 & y_{Ni}^2 & z_{Ni}^2 \end{bmatrix} \quad (2)$$

Hence, the sensors set from the whole area can be written as

$$X^T = \begin{bmatrix} S_1^1 \\ \vdots \\ S_{Ni}^1 \\ S_1^2 \\ \vdots \\ S_{Ni}^2 \\ \vdots \\ \vdots \\ S_1^k \\ \vdots \\ S_{Ni}^k \end{bmatrix} = \begin{bmatrix} x_1^1 & y_1^1 & z_1^1 \\ \vdots & \vdots & \vdots \\ x_{Ni}^1 & y_{Ni}^1 & z_{Ni}^1 \\ x_1^2 & y_1^2 & z_1^2 \\ \vdots & \vdots & \vdots \\ x_{Ni}^2 & y_{Ni}^2 & z_{Ni}^2 \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ x_1^k & y_1^k & z_1^k \\ \vdots & \vdots & \vdots \\ x_{Ni}^k & y_{Ni}^k & z_{Ni}^k \end{bmatrix} \quad (3)$$

where $N = Ni.K$ is the total number of sensors.

D. Find out the optimum direction by LDA

Given the sensors introduced in Section 1,

$$Y = \omega^T.X \quad (4)$$

Where $\omega = \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{23} \\ \omega_{31} & \omega_{23} & \omega_{33} \end{bmatrix}$

is the projection of X onto ω,
ω is the $3 * 3$ dimensional matrix
Y is the $3 * Ni$ dimensional matrix given by

$$Y = [Y_1^1, \dots, Y_{Ni}^1; Y_1^2, \dots, Y_{Ni}^2; Y_1^3, \dots, Y_{Ni}^3] \quad (5)$$

Each entry of row in Y represents projection of a sensor, or $Y = [\mathrm{LDA1}, \mathrm{LDA2}, \mathrm{LDA3}]$.

Where $LDA_i = [\omega_{11}X(1) + \omega_{21}X(2) + \omega_{31}X(3)]$.

And- $X(1)$ row number 1 of matrix X.
-$X(2)$ row number 2 of matrix X.
-$X(3)$row number 3 of matrix X.

The means of the sensors with the two different subareas after projection are denoted by μ_1 and μ_2.

$$\mu_1 = \frac{1}{Ni}\sum_{i=1}^{Ni} S_i^1 \quad (6)$$

$$\mu_2 = \frac{1}{Ni}\sum_{i=1}^{Ni} S_i^2 \quad (7)$$

The means of entire area are denoted by μ. Thus, we have

$$\mu = \mu_1 + \mu_2 + \cdots + \mu_k \quad (8)$$

After projection, we would like the samples with different placement to be well separated, which implies that the means of the classes with different places should be as far apart as possible and the sensors with the same placements are scattered in as small a region as possible. As we know, the LDA determines the ω by maximizing,
According to Fisher's [17] intuition it is desired to find a hyperplane in order to maximize the distance between the means of the two classes and at the same time to minimize the variance in each class. Mathematically this can be described by maximization of Fisher's criterion:

$$J(\omega) = max \left\| \frac{\omega^T Sb\ \omega}{\omega^T Sw\ \omega} \right\| \quad (9)$$

where Sb_i is the between classes scatter of subarea i .

$$Sb_i = (\mu_i - \mu))(\mu_i - \mu)^T \quad (10)$$

The between scatter matrix Sb is defined as

$$Sb = \sum_{i=1}^{k} Sb_i \quad (11)$$

Furthermore, Sw_i is the Within class scatter matrix of subarea i.

$$Sw_i = \left(S_1^i - \mu_i\right))(S_1^i - \mu_i)^T \quad (12)$$

And Within scatter matrix is expressed as

$$Sw = \sum_{i=1}^{k} Sw_i \quad (13)$$

This optimization problem can have infinitely many solutions That is for a solution ω all the vectors $c\ \omega$ give exactly the same value. If, without loss of generality, we replace the denominator with an equality constraint in order to choose only one solution. Then the problem becomes:

$$\max_{\omega} \|\omega^T Sb\ \omega\| \quad (14)$$

$$\text{If } \omega^T Sw\ \omega = 1 \quad (15)$$

The Lagrangian associated with this problem is:

$$\mathcal{L}_{LDA}(x, \lambda) = \omega^T Sb\ \omega - \lambda(\omega^T Sw\ \omega - 1) \quad (16)$$

where λ is the Lagrange multiplier that is associated with the constraint equation (15). Since S_b is positive semidefinite the problem is convex and the global minimum will be at the point for which

$$\frac{\delta \mathcal{L}_{LDA}(x,\lambda)}{\delta x} = 0 \Leftrightarrow Sb\ \omega - \lambda\ Sw = 0 \quad (17)$$

The optimal ω can be obtained as the eigenvector that corresponds to the greatest eigenvalue of the following generalized eigensystem

$$Sb.\omega = \lambda.Sw.\omega \quad (18)$$

E. Raisonment algorithm

Algorithm 1 summarizes the complete LDA selection procedure.

Algorithm 1

Input: N sensors $[\boldsymbol{S_1} \quad \dots \quad \boldsymbol{S_N}]$.
1) Partition with k subarea by $K=N.N_i$ and with k means (we compute k with divide N/Ni and Ni is variable with number of clusters it is satisfied $K=N.N_i$).
2) Compute Sb et Sw matrix.
3) Compute the eigenvector $\boldsymbol{\omega}$ by the LDA.
4) Sort eigenvector with decreasing order of eigenvalues in matrix $\boldsymbol{\omega}$.
5) Calcule projection basis $= \boldsymbol{\omega^T}.\boldsymbol{X}$.
6) Standardization of Y with Z index $\boldsymbol{Z_{i=1,k}} = \frac{Y_i - \mu}{\sqrt{Var(Y_i)}}$
7) Calcule the decision boundary with divide the means of each adjacent two subareas in one dimensional linear combination.

8) Apply Metrics for clustering performance evaluation Check if the classification accuracy if not go to step 1
9) Output: K optimal clustering.

IV. Sensor Layout Based On LDA Approach For A Smart Greenhouse

A. Greenhouse

Wireless Sensors can form a dense network and provide the possibility for continuous monitoring of relevant parameters in a dense grid for a reasonable price, providing a decision support system (DSS) for delivering insight into possible treatments, field-wide or for specific parts of a field.Challenges for the implementation of a wireless network are on one hand due to the fact that large-scale greenhouses have large distributed areas, high variance of temperature and humidity. Thus, it is difficult to lay wires and power supplements. On the other hand, the workload of wired equipment installation and maintenance is heavy. So, in large-scale greenhouses, transmitting signals with wires is not suitable. Finally, the network made up of wireless sensor nodes has some good characters such as mobility, reliable stability and good maintainability. However, in order to attain the desired results, the variables with an indirect influence on the climate also need to be modeled fig 6. in this work we model with Computational Fluid Dynamics (CFD) model developed by ANSYS fluent the solar radiation and the soil heat, for a tomato greenhouse of 10 by 50 meters.

B. Cluster generation

We consider a WSN field composed of N sensor nodes randomly distributed in a large square monitored area with irregular topology. For uniform correlation regions generation, we use the K means to divide the whole network into K contiguous subareas. In each block, there are Ni, sensor nodes whose data are assumed to be uniformly correlated.
The clustering approach proposed is based on LDA . As mentioned before, the network state is assumed to be stationary over a long period of time. After collecting the data from sensors, the sink calculates the data matrix as given in (3) and computes its Eigenvalue Decomposition (EVD).
Then, the matrix ω is constructed from the 3 main eigenvectors of the data matrix.
The WSN is clustered into K subareas, since the number of clusters K is not available in practice, it must be estimated. The optimal number K of clusters is of great importance as it fixes the amount of inter and

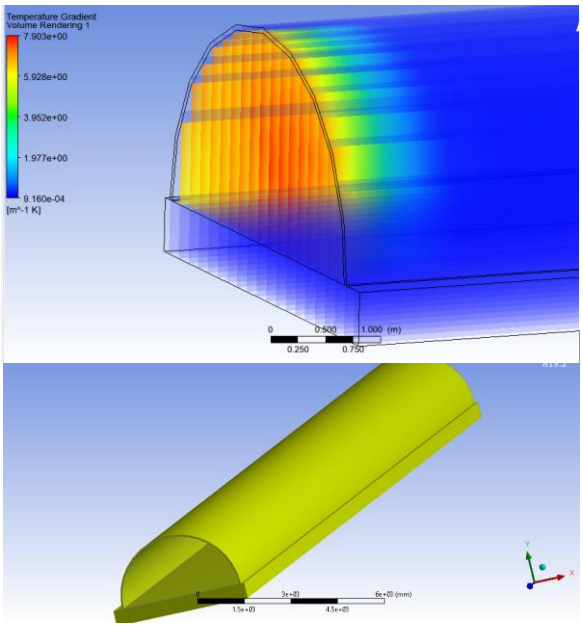

Fig. 6. Greenhouse CFD model

intra-cluster communications. To estimate K, we propose the exploiting the spatial data correlation local uniformity by using LDA technique.
Where $\omega \in R^{3*3}$ is the projection matrix with orthonormal columns formed by the sorted eigenvectors of the equation (18) (in decreasing order of eigenvalues), Y can be considered in one dimensional degree by using dimensionality redaction of LDA when neglecting the d-2;
Therefore, we propose estimating the number of subarea K (corresponding to the clusters number). For this purpose, we recover the optimal K that minimizes LDA reconstruction error or mean square error (MSE), expressed, for the assumed K as

$$MSE(K) = E\left\{\left\|S_i^k - Y_i^K\right\|^2\right\}_{i\,=\,1,2,..,K} \qquad (19)$$

Where Y_i^K is the reconstructed signal using LDA for an assumed K subarea, given by equation (4). this way, we evaluate the error MSE for each possible value of K ranging from 1 to N by using $k = N.N_i$. The optimal number of clusters, Kopt lowest value of MSE.

C. Metrics for clustering performance evaluation

The performance of clustering approaches is evaluated in terms of the validity indices, variance analysis and the global silhouette value. More precisely, five cluster validity indices [21], as expressed in the following, have been evaluated.

1) Calinski-Harabasz

$$CH_{index} = \frac{\frac{1}{K-1}\sum_{l=1}^{K}\|Z_l - Z\|^2}{\frac{1}{N-1}\sum_{l=1}^{K}\sum_{i=1}^{Ni}\|x_l - Z_l\|^2} \qquad (20)$$

- N_i number of points in the ith cluster.
- z: the centroid of the entire data set.
- z_i: represents the ith cluster centroid.

2) Sum of Squared Error (SSE)

the variance analysis is evaluated in terms of the Sum of Squared Error (SSE) given by

$$SEE = \frac{1}{N}\sum_{l=1}^{K}\sum_{i=1}^{Ni}\left\|Y_{ij} - \bar{C}_i\right\|^2 \qquad (21)$$

- Yij is the jth element in C_i .
- $\bar{C}_i$ is the mean measure of Ci nodes.

3) Davies-Bouldin (DB)

$$DB = \frac{1}{K}\sum_{i=1}^{k}\max_{j,j\neq i}\left(\frac{S_i+S_j}{d'_{ij}}\right) \qquad (22)$$

With

- $S_i = \frac{1}{N_i}\sum_{x\in C_i}\|x - z_i\|$
- $d'_{ij} = \|z_i - z_j\|$: distance between clusters C_i and C_j respective centroids.
- z_i: represents the ith cluster centroid.

4) Dunn's (V_D)

$$v_D = \min_{1\leq i\leq K}\left(\min_{1\leq j\leq K, j\neq i}\left(\frac{\delta(C_i,C_j)}{\max_{1\leq l\leq K}\Delta(C_i)}\right)\right) \qquad (23)$$

- $\delta(C_i, C_j) = \min_{x\in C_i, y\in C_j}(d(x,y))$ set distance between C_i and C_j .
- $\Delta(C_i) = \max_{x,y\in C_i,}(d(x,y))$ the diameter of C_i.

5) Global silhouette

for the global silhouette evaluation, we compute the silhouette score over clusters, which corresponds to the mean of silhouette coefficient for each sample calculated using the mean intra-cluster distance and the mean nearest-cluster distance.

$$S_{sil} = \frac{1}{K}\sum_{l=1}^{K}\frac{1}{Ni}\sum_{i=1}^{Ni} S_{sil}(i) \qquad (24)$$

- $S_{sil}(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))}$
- $a(i) = \frac{1}{N_i-1}\sum_{x\in C_i, y\in C_j} d(x,y)$ be the mean distance between x and all other data points in the same cluster.
- $b(i) = \min_{k'\neq k}\frac{1}{N_i}\sum_{x\in C_i, y\in C_j} d(x,y)$ the mean dissimilarity of point x to some cluster C_k as the mean of the distance from x to all points in C_i (where $C_k \neq C_i$).

The objective here is to minimize the values of DB, SSE and maximize the indices values of CH, v_D and the global silhouette value to achieve proper clustering best value which is 1 [22], [23].

D. Evaluation of optimal number of clusters in greenhouse

In order to effectively detect the number of clusters K, we use the dimensionality reduction through LDA technique application. In this approach, we consider a large WSN with N = 30 sensor nodes that are partitioned into $k = N.N_i$ uniform correlation and spatially contiguous subarea. Fig. 7 displays the error MSE(K) as a function of the supposed sparsity level (K). We can see that MSE(K) minimum value with K =2. This means first that LDA captures the useful information through the one independent dominant component, thus showing the effectiveness of LDA application for clusters number detection.

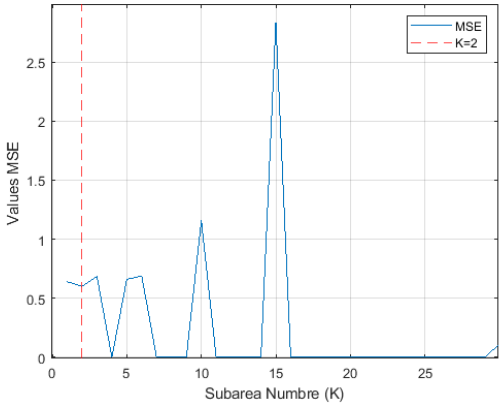

Fig. 7. MSE evaluation as a function of the supposed sparsity level (K).

Also, in order to estimate Kopt, we use method is called the Elbow [21] and the validity indices are evaluated for different values of K as shown in Fig 8,9,10,11&12 In such study, the number of clusters that maximizes CH index, global silhouette value is taken as Kopt. We notice the effectiveness of such proposed metrics as they lead to Kopt = 2.

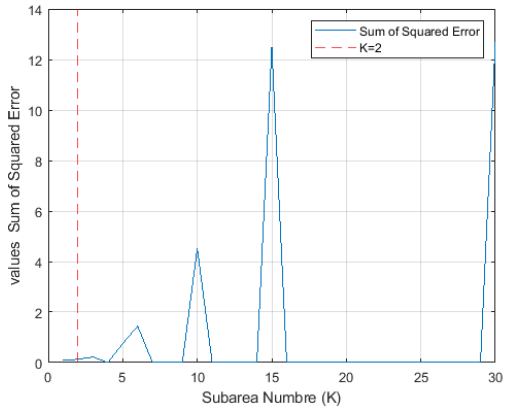

Fig. 8. Validity indices, SSE versus the clusters number (K).

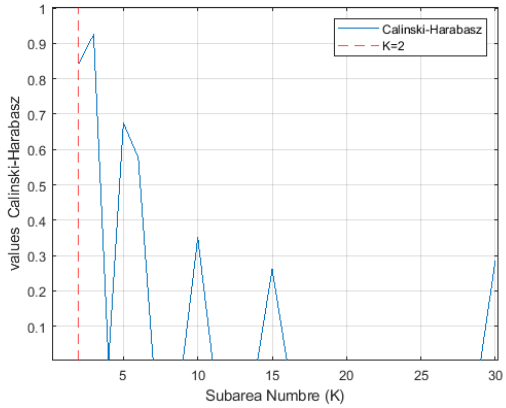

Fig. 9. Validity indices, Calinski-Harabasz versus the clusters number (K).

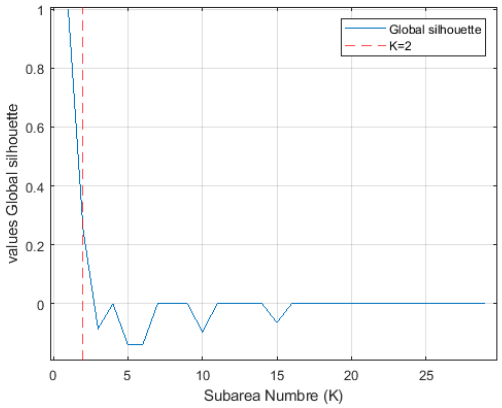

Fig. 11. Validity indices, Global silhouette versus the clusters number (K).

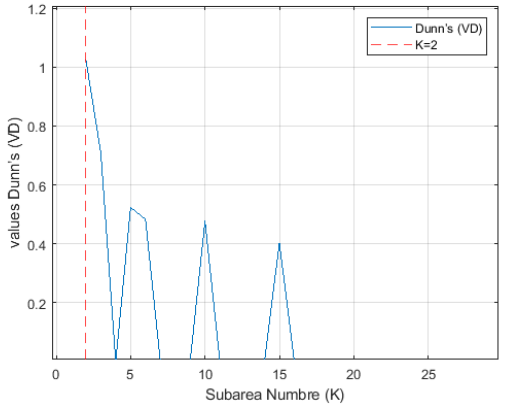

Fig. 10. Validity indices, Dunn's versus the clusters number (K).

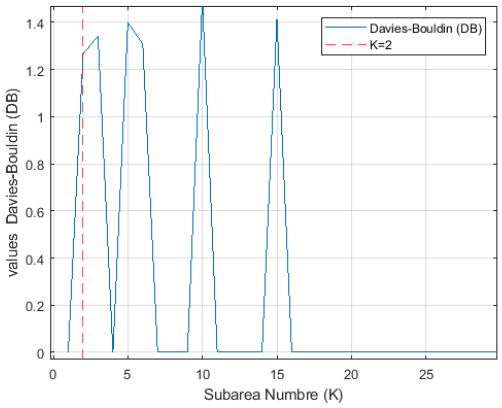

Fig. 12. Validity indices, Davies-Bouldin (DB) versus the clusters number (K).

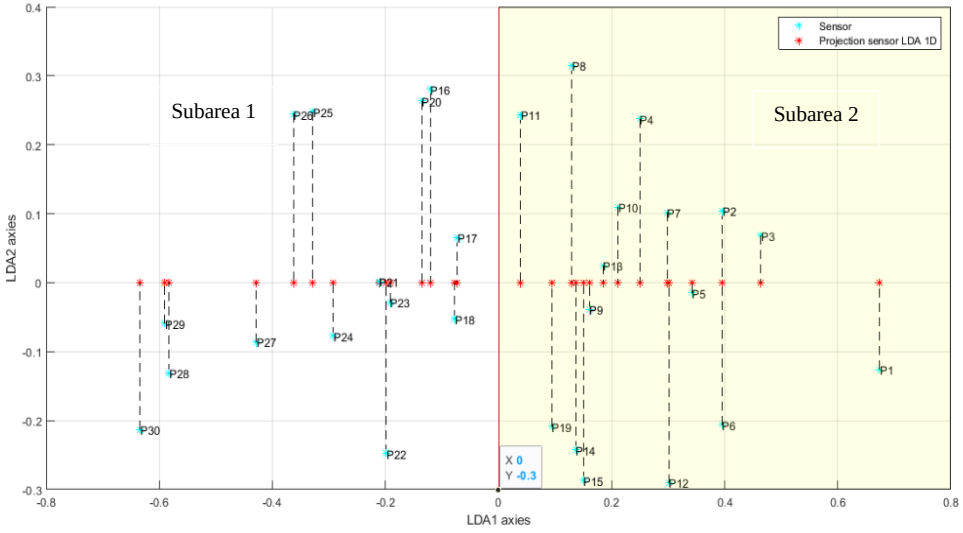

Fig. 13. Optimal partition K=2, Projection Data of sensors in LDA bias with line decision boundary

For finding the projection hyperplane that minimizes the interclass variance and maximizes the distance between the projected means of the classes. The LDA can determine the optimum direction of projecting ω, such that when projecting onto ω, Table II show the variation within and between scatter for each K are considered when setting K to Kopt =2 the samples from two different subarea are as separate as possible and the sensors from the same classes are as close as possible Var=0.08 and sbet =273.2.
Var: interclass variance /Sbet: between scatter

TABLE II. Within and between scatter for each K

	K=1	K=2	K=3	K=5	K=6	K=10	K=15	K=30
Var	0.03	**0.08**	0.13	0.31	0.38	0.92	2.85	1.07
Sbet	1.59e-29	**273.2**	226.9	168.1	140.6	92.24	63.20	32.25

Data samples in two dimensions are projected in a Lower dimension space present in Fig 13 the line has to be chosen so that the projection maximizes the "separability" of the projected samples.
The obtained results K=2 show that the proposed LDA-based clustering achieves the lowest SSE rates compared to others as shown In Fig. 7 Also, our approach realizes the best performance by reaching the higher global silhouette value as well as v_D and CH rates as seen in Fig 9,10 & 11.

V. CONCLUSION

This paper proposes a method for obtaining the optimal sensor configuration in a system. Placement and best clustering When seeking the best direction, for LDA, Experiments show that this approach could significantly increase the classification accuracy and energy management.
This clustering approach aims to divide the large network into irregular subareas, regarded as blocks, in each of which the measurements nodes are as uniformly correlated as possible. This heterogeneity is exploited for estimating the number of possible clusters number in the network. The best we obtained Kopt=2.
Some improvements such as using the Quadratic Discriminant Analysis (QDA) could be investigated in the future to achieve better results for boundary lines.

REFERENCES

[1] M. Mancuso and F. Bustaffa, 'A wireless sensors network for monitoring environmental variables in a tomato greenhouse', in *2006 IEEE International Workshop on Factory Communication Systems*, Jun. 2006, pp. 107–110. doi: 10.1109/WFCS.2006.1704135.

[2] K. P. Ferentinos, N. Katsoulas, A. Tzounis, T. Bartzanas, and C. Kittas, 'Wireless sensor networks for greenhouse climate and plant condition assessment', *Biosyst. Eng.*, vol. 153, pp. 70–81, Jan. 2017, doi: 10.1016/j.biosystemseng.2016.11.005.

[3] M.-S. Liao *et al.*, 'On precisely relating the growth of Phalaenopsis leaves to greenhouse environmental factors by using an IoT-based monitoring system', *Comput. Electron. Agric.*, vol. 136, pp. 125–139, Apr. 2017, doi: 10.1016/j.compag.2017.03.003.

[4] M. Kamel *et al.*, 'Modèle de bilan énergétique d'une serre agricole sans couvert végétal', vol. 11, pp. 1–51, Jan. 2008.

[5] S. Lee, I. Lee, U. Yeo, R. Kim, and J. Kim, 'Optimal sensor placement for monitoring and controlling greenhouse internal environments', *Biosyst. Eng.*, vol. 188, pp. 190–206, Dec. 2019, doi: 10.1016/j.biosystemseng.2019.10.005.

[6] R. Löhner and F. Camelli, 'Optimal placement of sensors for contaminant detection based on detailed 3D CFD simulations', *Eng. Comput.*, vol. 22, no. 3, pp. 260–273, Jan. 2005, doi: 10.1108/02644400510588076.

[7] M. Arnesano, G. M. Revel, and F. Seri, 'A tool for the optimal sensor placement to optimize temperature monitoring in large sports spaces', *Autom. Constr.*, vol. 68, pp. 223–234, Aug. 2016, doi: 10.1016/j.autcon.2016.05.012.

[8] Khairul Rijal Wagiman, Mohd Noor Abdullah, Mohammad Yusri Hassan, Nur Hanis Mohammad Radzi, 'A new optimal light sensor placement method of an indoor lighting control system for improving energy performance and visual comfort', *Building Engineering*, 2020.

[9] K. Worden and A. P. Burrows, 'Optimal sensor placement for fault detection', *Eng. Struct.*, vol. 23, no. 8, pp. 885–901, Aug. 2001, doi: 10.1016/S0141-0296(00)00118-8.

[10] T.-H. Yi, G.-D. Zhou, H.-N. Li, and C.-W. Wang, 'Optimal placement of triaxial sensors for modal identification using hierarchic wolf algorithm: Optimal placement of triaxial sensors using HWA', *Struct. Control Health Monit.*, vol. 24, no. 8, p. e1958, Aug. 2017, doi: 10.1002/stc.1958.

[11] D. Guo, D. Zhang, and L. Zhang, 'An LDA based sensor selection approach used in breath analysis system', *Sens. Actuators B-Chem. - Sens. ACTUATOR B-CHEM*, vol. 157, pp. 265–274, Sep. 2011, doi: 10.1016/j.snb.2011.03.061.

[12] K. P. Ferentinos and T. A. Tsiligiridis, 'Adaptive design optimization of wireless sensor networks using genetic algorithms', *Comput. Netw.*, vol. 51, no. 4, pp. 1031–1051, Mar. 2007, doi: 10.1016/j.comnet.2006.06.013.

[13] A. Saberian and S. M. Sajadiye, 'The effect of dynamic solar heat load on the greenhouse microclimate using CFD simulation', *Renew. Energy*, vol. 138, pp. 722–737, Aug. 2019, doi: 10.1016/j.renene.2019.01.108.

[14] A. Abene, M. Mefoued, L. Maacha, and V. Dubois, 'Modélisation thermique d'une serre agricole chauffée par l'énergie géothermale', in *Revue des énergies renouvelables*, 2007, pp. 207–215. Accessed: Jul. 27, 2021. [Online]. Available: http://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=20792914

[15] J. López-Martínez, J. L. Blanco-Claraco, J. Pérez-Alonso, and Á. J. Callejón-Ferre, 'Distributed network for measuring climatic parameters in heterogeneous environments: Application in a greenhouse', *Comput. Electron. Agric.*, vol. 145, pp. 105–121, Feb. 2018, doi: 10.1016/j.compag.2017.12.028.

[16] 'Ronald Aylmer Fisher', *Wikipédia*. Jun. 29, 2021. Accessed: Jul. 30, 2021. [Online]. Available: https://fr.wikipedia.org/w/index.php?title=Ronald_Aylmer_Fisher&oldid=184222594

[17] P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, 'Linear Discriminant Analysis', in *Robust Data Mining*, P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, Eds. New York, NY: Springer, 2013, pp. 27–33. doi: 10.1007/978-1-4419-9878-1_4.

[18] S. Ghiasi, A. Srivastava, X. Yang, and M. Sarrafzadeh, 'Optimal Energy Aware Clustering in Sensor Networks', *Sensors*, vol. 2, no. 7, Art. no. 7, Jul. 2002, doi: 10.3390/s20700258.

[19] 'K-moyennes', *Wikipédia*. Sep. 09, 2020. Accessed: Jul. 30, 2021. [Online]. Available: https://fr.wikipedia.org/w/index.php?title=K-moyennes&oldid=174559088

[20] 'Performance evaluation of some clustering algorithms and validity indices | IEEE Journals & Magazine | IEEE Xplore'. https://ieeexplore.ieee.org/abstract/document/1114856 (accessed Aug. 02, 2021).

[21] Z. Jellali, L. N. Atallah, and S. Cherif, 'Principal Component Analysis based Clustering Approach for WSN with Locally Uniformly Correlated Data', in *2019 15th International Wireless Communications Mobile Computing Conference (IWCMC)*, Jun. 2019, pp. 174–179. doi: 10.1109/IWCMC.2019.8766477.

[22] U. Maulik and S. Bandyopadhyay, 'Performance evaluation of some clustering algorithms and validity indices', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1650–1654, Dec. 2002, doi: 10.1109/TPAMI.2002.1114856.