Energy Performance of a Combined Horizontal and Vertical Compression Approach in Cluster-based WSNs

Jihane El Aasri, Samia Al Fallah, Mounir Arioua, and
Ahmed El Oualkadi

Laboratory of Information and Communication Technologies, National School of
Applied Science, Tangier
Abdelmalek Essaadi University, Morocco

Abstract

 Energy efficiency is an essential issue to be reckoned in wireless sensor networks development. Since the low-powered sensor nodes deplete their energy in transmitting the collected information, several strategies have been proposed to investigate the communication power consumption, in order to reduce the amount of transmitted data without affecting the information reliability. Lossy compression is a promising solution recently adapted to overcome the challenging energy consumption, by exploiting the data correlation and discarding the redundant information. In this paper, we propose a hybrid compression approach based on two dimensions specified as horizontal (HC) and vertical compression (VC), typically implemented in cluster-based routing architecture. The proposed scheme considers two key performance metrics, energy expenditure, and data accuracy to decide the adequate compression approach based on HC-VC or VC-HC configuration according to each WSN application requirement. Simulation results exhibit the performance of both proposed approaches in terms of extending the clustering network lifetime.

Keywords

WSN, Clustering, Vertical compression, Horizontal compression, Energy efficiency, Data reliability

1   Introduction

Wireless Sensor Networks (WSNs) are being increasingly used in a wide range of applications, such as seismic sensing, agriculture transforming, military tracking, healthcare and IoT [1] [2]. Sensor nodes in WSNs are energy-constrained and equipped with a sensing unit, a processing device, a transceiver and a small battery [3] [4]. In many applications, the replacement of sensor nodes batteries is eventually strenuous. Several

approaches have been proposed to enhance the network power consumption by designing energy-efficient methods in terms of information routing, MAC protocols and adaptive sensing [5][6][7]. Most of these active researches assume that the transmission cost dominates the energy expenditure in sensor networks [8]. To overcome this challenging issue, data reduction algorithms based on data compression and aggregation are considered to decline the number of bits forwarded through the network. Thus algorithms use the advantage of temporal and spatial correlation to eliminate statistical redundancies in space and time. Some of these techniques exploit the spatial correlation by grouping cluster nodes with similar sensing parameters in order to reduce the data traffic [9]. Research works on data reduction were focused on applying prediction models on time series temporally correlated to reduce the number of transmitted packets [10]. A compelling approach is proposed in [11] aims to analyse real-world datasets by investigating nodes with similar data evolution and bringing them to form groups in which only a representative node sends its data to the sink. In this performed work, the analysis of sensor datasets is performed vertically, this compression technique is considered in the rest of this paper as Vertical Compression (VC). In our previous work [12][13], a comparison of different compression algorithms is carried out and reveals that Piecewise Linear Approximation (PLA) techniques offer noteworthy enhancements in terms of energy consumption and data reliability [13]. These algorithms consider the collected information as time series and discard redundant signal using line segments, PLA is considered in the rest of this paper as Horizontal Compression (HC). In addition, network structure and routing schemes are the main components of data reduction algorithms that directly affect the energy efficiency and data accuracy in WSN. Cluster-based networks have proven to be an effective routing architecture that accomplishes a crucial aspect to accomplish a long-term monitoring network [14][15]. In such architecture, the network is divided to form a predetermined number of clusters in which a Cluster-Head (CH) is elected to collect information from the Cluster-Members (CMs) and communicate it to the sink [16].

Through this research paper, a performant data transmission structure is proposed exploiting the performance of Vertical and Horizontal compression schemes applied in a clustered network following two configurations (HC-VC and VC-HC). The proposed hybrid compression scheme is yielded to be energy efficient in the clustered network, maintaining a valuable data reliability estimated by root-mean-square error (RMSE) measure calculating the distortion of the original and reconstructed signal.

The rest of this paper is organized as follows: Section II describes the used system model detailing the type of signal, topology and data gathering, and the energy model. We briefly review in section III the horizontal and vertical approaches. The proposed compression scheme is discussed in section IV following the two suggested configurations. Section V exhibits the experiments we have conducted in real sensors data with discussion of results, in section VI, we summarize our main findings.

2 SYSTEM MODEL

2.1 Signal model

In this work, we consider as signal model real-world temperature measures, gathered from sensors in the frame of the project, that aims to study and share experimental results, obtained for the localization problem, and taken by Free University of Brussels researchers [17]. Time $t = 1, 2, 3$ is slotted with a fixed duration dt, to form a time period $P_{seg} = [S_1, S_1,..., S_T]$ of T time slots. In each period P_{seg}, each sensor node collects a vector of T data samples. We define a data collection round r as a set of m periods: $r = m\, P_{seg}$.

2.2 Topology and data gathering

We consider a cluster-based WSN of N nodes randomly deployed in an area D. Sensor nodes are grouped to form a number of clusters, in which one Cluster-Head is elected according to LEACH routing protocol [18]. In every period $Pseg$, each sensor node sense and transmit its sampling vector of T samples to its CH, which stores this incoming data from all cluster members. This process is repeated for each period $Pseg$ until achieving the end of the round when CH transmits the collected data from its CMs to the sink.

2.3 Energy model

For each compression technique, we identify two contributions of total energy expenditure: Energy for processing needed to accomplish the compression task, and energy cost to ensure the data transmission from nodes to CHs on the first hand and from CHs to the sink on the other hand.

Energy for processing: for each compression method, we counted the number of operations such as additions, multiplications, subtractions, comparisons, and divisions. Depending on the used micro-controller, the number of clock cycles corresponding to the obtained number of each type of operation represents the dissipated energy in executing the compression algorithm. In this work, Micro-controller MSP430 is selected for each sensor node, using 16 bits floating point package of calculation [19]. The adopted CPU consumes $E_0 = 0,726nJ$ per clock cycle. For each operation, a corresponding number of clock cycles is considered according to Table 1.

Table 1: CPU cycles needed for processing

Operation		Clock cycle
Addition	X+Y	184
Subtraction	X-Y	177
Multiplication	X×Y	395
Division	X /Y	405
Comparison	X<=>Y	37

***Energy for transmission*:** we consider a simple radio energy model, to evaluate hardware energy dissipation, where the transmitter depletes energy to operate the radio electronics (E_{ele-TX}) and the power amplifier (E_{amp}), and the receiver consume energy on running the radio electronics (E_{ele-RX}), as shown in Fig.1.
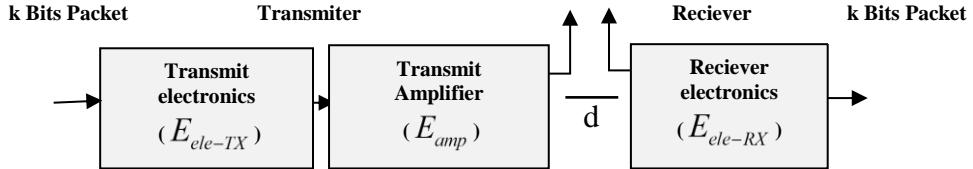

Figure 1: Energy dissipation model

In this work, both the free space (d^2 power loss) and the multipath (d^4 power loss) communication models were used, according to the distance between the transmitter and receiver [18]. If the distance is less than a threshold $d_{crossover}$, the free space model is used (f_s) otherwise, the multipath (f_{mp}) is used, in order to appropriate setting of the power amplifier. The crossover distance is related to antenna technology, for the two scenarios, it is defined as follows [20] [21]:

$$d_{crossover} = \frac{4\pi\sqrt{L}h_t h_r}{\lambda} \tag{1}$$

Where $h_r=h_t=1$m, *hr* is the heights of receiving antenna and *ht* is the heights of transmitting antenna relative to the ground, while *l* is the transmitted wavelength in proportion to the transmitting frequency and *L* is the system loss factor equal to one considering a lossless system. Thus, Energy E_{Tx} expended by radio to send data of *k-bit* packet over a distance *d* is displayed by equation (2) [22][18]:

$$E_{Tx}(k,d) = E_{Tx-ele}(k) + E_{Tx-amp}(k,d) = \begin{cases} k(E_{ele-Tx} + \varepsilon_{fs}d^2), d < d_{crossover} \\ k(E_{ele-Tx} + \varepsilon_{mp}d^4), d \geq d_{crossover} \end{cases} \tag{2}$$

Moreover, to receive this message, the radio expend:

$$E_{Rx}(k) = E_{ele-Rx} \cdot k \tag{3}$$

3   COMPRESSION TECHNIQUES BACKGROUD

3.1   Temporal Compression

Temporal compression algorithms, considered thereafter as horizontal compression, refer to data compression approaches typically based on exploiting the temporal correlation of collected information [14][23][24][13][25]. With horizontal compression, sensor nodes only transmit their compressed readings obtained from the sensing device so as to save transmission and reception energy by reducing the data size [25]. Some temporal compression methods named lossless provide an accurate reconstruction of the original

information [25] [26]. In other compression algorithms called Lossy, some degree of information loss is introduced in the compression/decompression process in order to achieve high compression ratio [14] [24]. In cluster-based WSNs, each sensor node compresses its data using horizontal compression in order to prolong the network lifespan [25]. Since the temporal correlation is considerably present in the sensing information, horizontal compression methods have inevitably enhanced the energy performance of WSNs [25]. Various compression methods have been proposed in the literature taken into account the collected information as time series to exploit the temporal data correlation [24] [13][26][27]. Some of these works are interested in transforming the original reading into a set of coefficients in order to facilitate the representation of the sensed signal [24]. For instance, some of these algorithms transform the time series from the time domain into the frequency domain, (Fast Fourier Transform [28], Discrete Cosine Transform [29] and Wavelet Transform [30]), however, each of these methods selects the coefficients differently [24]. In addition, other compression techniques are based on an adaptive modelling approach which converts the information to line segments [31], polynomial equation [32] or autoregressive schemes [33][24]. Considering the linear approaches, Piecewise Linear Approximation schemes aim to transform a time series of the sampled information into a set of segments depending on desired error tolerance [24]. For polynomial techniques, environmental measures are approximated via polynomial coefficients to minimize the number of original data samples [24]. Autoregressive approaches (M-AAR as an example) are based in a model of basic coefficients that exploit the history of data samples using the advantage of the signal temporal correlation [24]. Table 2 summarizes the performance of these compression schemes in terms of computational cost, data reliability and correlation impact on the algorithms complexity [14].

Table 2: Performance of Temporal Compression Schemes

Compression Schemes	Type	Computational cost	Data reliability	Complexity vs Correlation
FFT	Transform-based	Very high	Very high	Decreasing
DCT	Transform-based	High	High	Decreasing
LTC	PLA	Low	average	Decreasing
EPLAMLIS	PLA	Low	average	Decreasing
M-AAR	Autoregressive	High	High	Increasing
PR	Polynomial	Very high	Very high	Increasing

Authors in [34] and [24] confirm that PLA schemes ensure high energy saving with minimum energy for processing, in reverse to Polynomial Regression (PR), which induces a high computational expenditure nevertheless, it performs efficiently in terms of data reliability when the polynomial order is increased [24]. Concerning the Autoregressive methods, the model complexity is highly affected by the correlation signal length that may lead to high-energy consumption [24].In this work, we exploit the performance of PLA schemes considered as horizontal compression applied in cluster-members, in order to minimize the amount of the sensed data [25]. PLA schemes guarantee a significant performance in terms of power consumption and memory [24][25][34][24][13].

Lightweight Temporal Compression (LTC) is a PLA technique that has significantly proven its efficiency in terms of resource saving [24] [25][31]. It uses a linear model to approximate a given time series, depending on a predetermined error tolerance e. The algorithm works by approximating multiple successive readings through a single segment, in a way that the segment meets the error tolerance e for all points (algorithm 1). This method has linear complexity in the number of readings of time series. In addition, since the segment can be described by four coordinates, the compression is reached when the number of the covered readings is greater than 4.

Previous works revealed that LTC requires a slight energy consumption by reason of its lightweight algorithm, which requires a low number of operations while the processing (Algorithm 1) [14][24][13][34]. LTC loses its energy effectiveness when the temporal correlation decreases, due to the significant changes in data values [14][25][13][34].

Algorithm 1 Lightweight Temporal Compression

```
Inputs x, e // Sensed samples, Error bound
for i=1 to length(x) do
    j=i+1
    High-Seg=Seg[x(i),x(j)+e]
    Low-Seg= Seg[x(i),x(j)-e]
    while j < N do
        if High-Seg below x(j+1)- e or Low-Seg above x(j+1)+e then
            Save x(j)
            i=j
        else
        if High-Seg above x(j+1)+e then
            High-Seg=Seg[x(i),x(j+1)+e]
            end if
        if Low-Seg below x(j+1)- e then
            Low-Seg=Seg[x(i),x(j+1)-e]
            end if
        end if
        j=j+1
    end while
end for
```

3.2   *Data aggregation in clustered network*

Reducing data packets volume is a crucial process in WSN. Data aggregation algorithms, considered thereafter as Vertical compression, the attempt at analysing datasets coming from cluster members at the CHs level. Acting as an aggregator, every CH collects, processes and forwards aggregated packets to the base station. Many research papers were devoted to the processing of data collected by neighboring sensors to reduce the occurred redundancy, using the advantage of their significant spatial correlation. These works use mathematical theories [35], instead of basic operation (Average, SUM, MAX, MIN, Count) in order to maintain a trade-off between data accuracy and resource saving [36].

Researchers in [37] adopted an intriguing method for discarding redundancy between sets. This method is based on data aggregation for periodic sensor networks using similarity functions. The data reduction algorithm has two phases: a local aggregation is per-formed at the node level, followed by a second phase, in which datasets similarity exploited at the aggregator level using the Jaccard similarity function [24]. Their results revealed an important reduction of data, by handling the in-network redundancy and transmitting only relevant information to the sink. Authors in [9] proposed a cluster-based approximate aggregation (CAG) technique, intended at restricting the number of packets transmitted to the base station by creating clusters of nodes with similar sensing values, based on the spatial correlation between them, then only data recorded by a representative node is sent. In spite of the significant reduction in the transmission of packets during the data collection, constructing clusters by flooding is energy-consuming. In our work, we assume that clusters are created before, and we concentrate on the data aggregation process by using similarity functions.

Thus, Data processing at the CH level covers several algorithms varying in complexity and level of operations. starting from simple summarization methods to complex inference mechanisms, according to the requirements of the design for each specific WSN application.

3.2.1   Observation from datasets

A similarity measure based on distance functions aims to eliminate redundant transmissions by grouping similar data instances [38] [39]. It describes data samples as how close or how fare they lie to each other [40] [41]. However, collected measures highly depend on the monitored environment. Consequently, when the assessed condition speeds up or slows down, collected measures by each sensor are more correlated and redundant. As shown in Figure 2, temperature data generated by nearby sensors show similar evolution, whereas the data is different with some deviation along time. Therefore, spatial clustering based on data evolution is feasible and beneficial [11] [39].

Distance Measures largely applied to sensor network are Euclidean distance [36], Pearsons correlation, Mahalanobis distance (covariance matrix), Jaccard similarity function [9], and cosine similarity.

3.2.2   Vertical compression at the CH level

- Capture the data evolution

According to the results showed in Figure 2, we examine the evolution of data coming from cluster members periodically in rounds, adopting the same method performed by authors in [11], using piecewise linear regression (in the segment). In this method, at CHs level, for each cluster member, an approximate vector of regression coefficients should be associated $a = [a_{seg-1}, a_{seg-2}, ..., a_{seg-m}]$, it represents data evolution captured in a segment.
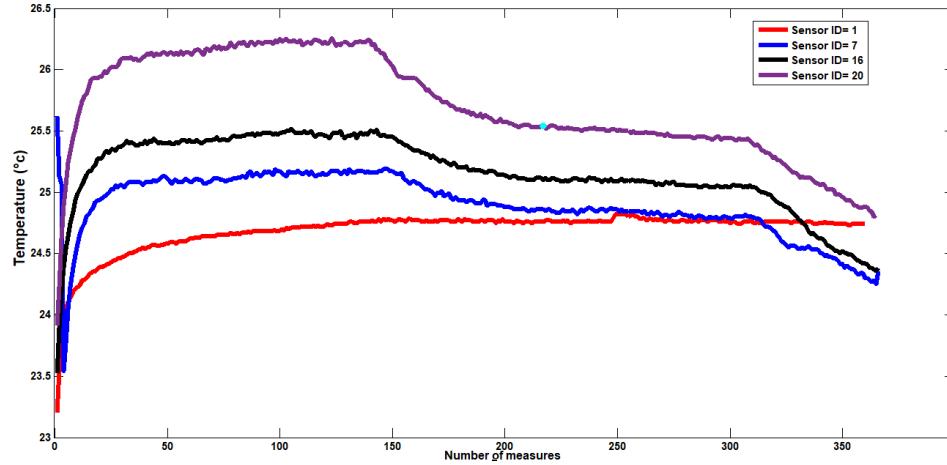

Figure 2: Distribution of sensors in the laboratory [17]

As presented in section II, Each round r is divided into periods P where P_{seg} ($seg \in [1, 2, ..., m]$). Each period $P_{seg} = [S_1, S_2, ..., S_T]$ is partitioned into a number T of time slots. At each slot S_i, each sensor S captures a new measure w_{it}, until it forms a vector of measures $W_i = [w_{i1}, w_{i2}, ..., w_{iT}]$ at the end of each period, then it transmits it to its CH. At the end of round r, and after receiving all data packets from the CMs, a vector of slops $a_{iseg} = [a_{i1}, a_{i2}, ..., a_{im}]$ is constructed at CH level, where a_{seg} is a slope calculated by a CH in a period P_{seg} for their entire attached CM, using equation (4). y is temperature value and i is the corresponding time sequence. The following equation expresses the regression coefficient a_{seg} in one segment (Si = 5 as an example).

$$a_{seg} = \frac{\sum_{i=1}^{si}(i-\bar{\imath})(y-\bar{y})}{\sum_{i=1}^{si}(i-\bar{\imath})^2} \tag{4}$$

Where $\bar{\imath} = \dfrac{1}{S_i}\sum_{i=1}^{S_i} i$, $\bar{y} = \dfrac{1}{S_i}\sum_{i=1}^{s_i} y_i$ and S_i is a length of a segment.
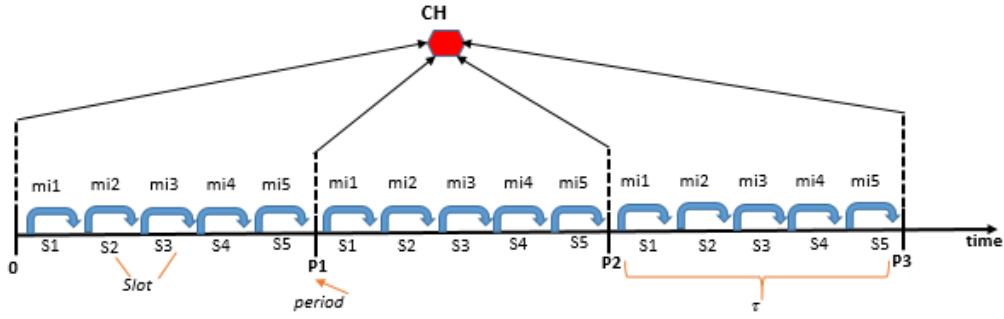
Figure 3: Data collection at the CH level for m=3 and T=5

- Datasets similarity measure

Similarity assessment between vectors of data evolution constructed for each sensor member is the goal of the second phase. Using the cosine similarity function, which is a measure of similarity between two non-zero vectors, it captures the orientation (the angle) between them based on the dot product, as illustrated in the following equation:

$$x \cdot y = \|x\| \|y\| \cos \vartheta \qquad (5)$$

Where q is the angle within vectors x and y. A cosine value of 0 means that the angle between the two vectors is equal to 90 degrees (orthogonal) and has no match. The nearer the vectors to each other, the cosine value closer to 1, therefore similar.

$$Cs(x, y) = \cos \vartheta = \frac{x \cdot y}{\|x\| \|y\|} \qquad (6)$$

At the CHs, sensor nodes groups are created according to a given threshold th_{cs}, determined by the user.

$$sim(x, y) = \begin{cases} 1 \longrightarrow Cs(x, y) \geq th_{cs} \\ 0 \longrightarrow Otherwise \end{cases} \qquad (7)$$

Only one data value per group is transmitted as shown in Figure 4, which decreases the number of transmitted packets from CHs to sink.
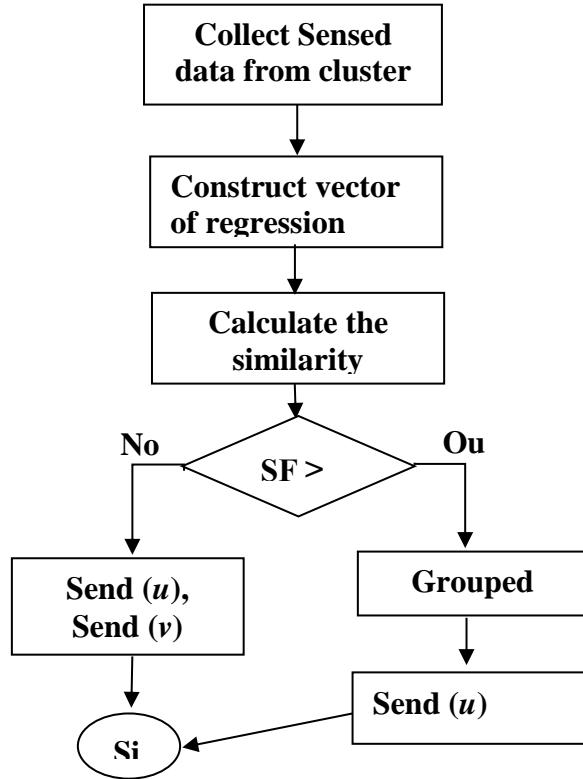
Figure 4: Data classification model in the CH

Algorithm 2 Data classification in the CH level

Inputs SensorData, Cluster Membre CM
for i=1 to CM do
   Slopes(i) Slope SensorData(i)
end for
for i=1 to CM do
  for j=i+1 to CM do
    sim=Similarity(Slopes(i),Slopes(j))
    if sim > th_{Cs} then
      Group(i,j)
    end if
  end for
end for

---

4   PROPOSED SCHEME

The main goal of the proposed scheme is to prolong energy lifetime as well as to improve the data reliability in cluster-based WSNs. A combined horizontal and vertical compression, the approach is designed to reduce the communication burden taking into account two configurations.

Considering the system model presented in Section II. Sensor nodes sense their environments periodically, in order to forward data collected packets to the appropriate CH after each time period. CH aggregates data packets at the edge of each round and sends it to the base station. The transmission process follows three scenarios:

- $\partial_1 : s \rightarrow c$: Sensing data sent from sources nodes to their assigned CHs, based on the nearest distance between them.

- $\partial_2 : c \rightarrow Bs$: At this level, data packets coming from source nodes are combined using an aggregation function to reduce redundant data, then the aggregated packet is sent to the base station via the direct channel.

- $\partial_3 : S \rightarrow Bs$: Two cases enable source nodes to communicate directly to the base station, the first one when no CH is elected, this occurred frequently in the latest rounds. The second instance if the distance between the source node and the base station is shorter than the distance between the source node and CH.

Our proposed configurations are based on two algorithms: vertical compression and horizontal compression, following two different schemes.

4.1   Configuration 1:

Considering a cluster-based architecture, a horizontal compression based on PLA compression technique is applied at sensor nodes level, in a manner that *k bits* message is transformed into *k HCR bits* message in the reduced packet version, where *HCR* is the Horizontal Compression Ratio, which is estimated based on a given Error bound as described in the third section. The horizontal compression ratio is given by the following equation:

$$HCR = \frac{Volume\ of\ compressed\ data}{volume\ of\ data} \tag{8}$$

At the end of each period *Pseg*, the data packet is sent to the elected CH to be stored. After receiving *m* data compressed packets from each cluster member, vertical compression is performed at CHs, by applying algorithm 2. We take an example of a CHA, which is associated to four Cluster members *S1*, *S2*, *S3* and *S4*. So as to illustrate the process of group forming, Table 3 Show an example of node *S1* where the column *Sensor$_{ID}$* is the neighbor *ID*, *Cs* and *Sim* are the values of Eq.6 and Eq.7 respectively.

Table 3: The cluster members table for S1 node

Sensor_ID	Cs	Sim
S_2	$c_s(\vec{R}_c^{s_1}, \vec{R}_c^{s_2})$	1
S_3	$c_s(\vec{R}_c^{s_1}, \vec{R}_c^{s_3})$	1
S_4	$c_s(\vec{R}_c^{s_1}, \vec{R}_c^{s_4})$	0

From the example above, node S_1 is similar with S_2 and S_3, considering a theorem proved by authors in [35], If R_c^i is similar with R_c^j and R_c^k, thus R_c^j should be similar with R_c^k, i.e.,

$$sim(\vec{R}_c^i, \vec{R}_c^j) = 1 \,\&\, sim(\vec{R}_c^i, \vec{R}_c^k) = 1$$
$$\Rightarrow sim(\vec{R}_c^j, \vec{R}_c^k) = 1, \forall_{j,k} \tag{9}$$

Thus, S_2 and S_3 are similar, a group of S_1, S_2, S_3 is constructed, and presented only by one node. The comparison process is performed for all cluster members, until grouping all similar nodes. As a result, each CH transmits to the base station a reduced version of the cluster data according the following equation:

$$ClusterData = m * k * HCR * N_{CM} * VCR \tag{10}$$

Where m is the number of periods in the round, N_{CM} is number of cluster-members and VCR is the vertical compression ratio, which represents the ratio of compressed packets (n_{vc}) respected to the total packets generated (N_{tot}):

$$VCR = \frac{n_{VC}}{N_{tot}} \tag{11}$$

As *HCR*, *VCR* restricts the amount of transmitted information at the cost of some Vertical Compression Distortion *VCD*. The efficiency of the proposed scheme in terms of energy saving owing to the reduced amount of transmitted data in the network. In fact, each sensor node sends only their compressed datasets to the CH, in which a vertical compression is introduced, obtaining a restricted number of transmitted packets. Moreover, the configuration presents some disadvantages in terms of data reliability, due to the loss of information engendered in every compression step especially at the sensor nodes level.

4.2   Configuration 2:

In this scenario, each sensor node transmits the collected information to its CH without applying any compression algorithm. At the CHs level, a vertical compression (*VCR,VCD*) is initially considered, in which groups of similar nodes are formed, following the specifications mentioned in section 2.4. Therefore, CH reduces the number of received vectors from *N* to *N*VCR*. Afterwards, the horizontal compression (HCR, HCD) is carried out on the data packet (*k bits* message) outcome from groups of nodes, at the same level of the CH, in order to reduce the maximum number of bits, before transmitting a reduced version to the base station (Algorithm 3).

Algorithm 3 Configuration 1

---
    for i=1 to N do
      SensorData(*i*) *k HCR*(*i*)
      CH   SensorData(*i*)
    end for
    ClusterData CH VCR

Algorithm 4 Configuration 2

---
    for i=1 to N do
       SensorData(*i*) k
       CH   SensorData(*i*)
    end for
    CH   CH VCR
    ClusterData CH HCR

This configuration ensures a high level of data reliability with some additional gains in terms of energy saving. Indeed, the loss of information is not presented at the sensor nodes level, which promotes the data accuracy at the CH. Moreover, horizontal compression is only applied for the obtained vectors after vertical compression, which makes the received data more reliable at the base station.

5   ENERGY CONSUMPTION MODEL

5.1   *Energy consumption for configuration 1:*

Based on the first order RF model described in section II, the total energy expended to transmit compressed data (*k.HCR*) using horizontal compression at the node level, over a distance *d* is described as follows:

$$E_{CM}(k \cdot HCR, d) = E_{Tx}(k.HCR) + E_{Rx}(k \cdot HCR) + E_{HC}(k) \tag{12}$$

Where $E_{Tx}(k.HCR)$ is the energy expended by radio electronics in free space and multipath communication, is given by the following equation:

$$E_{TX}(k.HCR, d) = \begin{cases} k.HCR\ (E_{elec-Tx} + \epsilon_{fs} * d^2) & if\ d < d_{crossover} \\ k.HCR\ (E_{elec-Tx} + \epsilon_{amp} * d^4) & if\ d \geq d_{crossover} \end{cases} \tag{13}$$

The energy consumed at the receiver (CH) to run radio electronics is expressed by the following equation:

$$E_{RX-elec} = E_{elec-Rx} * k * HCR \tag{14}$$

After receiving compressed packets from sensor nodes, the total energy consumed by CH is expressed as the sum of energies [21].

$$E_{CH} = E_{RX-CM} + E_{VC} + E_{Tx-BS} \tag{15}$$

E_{Tx-Bs} quantifies the energy needed to forward data packets from CH to BS, taking into account *T* samples collected at *m Pseg* per round, which are stored at CH before executing the vertical compression algorithm. We consider $G_{conf\,1}$ as the total number of bits received from cluster members, represented as follows:

$$G_{conf1} = \sum_{i}^{N_{cm}} HCR(i) * k * m \tag{16}$$

Then, E_{Tx-BS} is given by the following equation:

$$E_{Tx-Bs}(G_{conf1},d) = \begin{cases} G_{conf1}*VCR(E_{ele-Tx}+\varepsilon_{fs}.d^2) & \text{If d} < \text{d}_{crossover} \\ G_{conf1}*VCR(E_{ele-Tx}+\varepsilon_{mp}.d^4) & \text{If d} > \text{d}_{crossover} \end{cases} \qquad (17)$$

$E_{Rx\text{-}CM}$ is the required energy to receive compressed data packets from cluster member, expressed in Eq.14.

5.2   Energy consumption for configuration 2:

In the second scenario, sensed data sent from sensor nodes to the appropriate CH without compression, therefore, the energy consumed to transmit data over distance d is described as:

$$E_{CM}(k,d) = E_{Tx}(k) + E_{Rx}(k) \qquad (18)$$

At the CH level, horizontal and vertical compression occurred, the following equation express the total energy dissipated:

$$E_{CH} = E_{RX-CM} + E_{VC} + E_{Tx-BS} + E_{HC} \qquad (19)$$

Where $E_{Rx\text{-}CM}$ in this case, is expressed as follow:

$$E_{RX-CM} = E_{elec-Rx} * k \qquad (20)$$

In order to calculate the energy needed to transmits packets from CH to Bs $E_{Tx\text{-}Bs}$, we modeled the number of resulting bits, before applying horizontal compression $G_{conf\,2}$, as follows:

$$G_{conf2} = m*k*VCR*N_{CM} \qquad (21)$$

As a result, the following equation expresses $E_{Tx\text{-}Bs}$, the energy required for transmitting a compressed packet from CH to BS:

$$E_{Tx-Bs}(G_{conf2},d) = \begin{cases} G_{conf2}* \sum_{1}^{VCR*N_{CM}} HCR(i)*(E_{ele-Tx}+\varepsilon_{fs}.d^2) \\ G_{conf2}* \sum_{1}^{VCR*N_{CM}} HCR(i)(E_{ele-Tx}+\varepsilon_{mp}.d^4) \end{cases} \qquad (22)$$

5.3   Distortion

Since the designated compression methods are lossy at the receiver side, the recovered data is an approximation of the original information [42]. Hence, the distortion parameter, defined as an information loss, it represents a measure to estimate how much error between two data sets (a predicted value and an observed known value), calculated as follow:

$$RMSE = D = \frac{1}{N}\sum_{i=1}^{N}|\hat{x}(i) - x(i)| \qquad (23)$$

Where $x(i)$ is a data value of a given time series and $x\hat{}(i)$ represents its compressed version. The accuracy requirement depending on the WSN application. Hence, selected data compression methods aim to minimize energy consumption and meanwhile guarantee consistent reliability of reconstructed data.

6   SIMULATIONS AND RESULTS

Through Matlab simulations, we present the experimental results of the hybrid vertical and horizontal compression approaches proposed in a cluster-based wireless sensor network, in order to evaluate our proposed configurations in terms of the energy conservation and data accuracy. The network is formed by a set of N = 50 sensors arbitrarily scattered in a plan area of 100*100 m, at 50 meters from the base station, which is powerful in computing and storage capacity. A starting energy level of all nodes is 0,5 *J*. This value is widely used in other research papers since it provides small enough energy to quickly see the effect of the suggested algorithm [43]. A packet size of k = 4000 bits is used. Table 4 shows the simulation parameters adopted in this study. We selected the same clustering algorithm used in the LEACH protocol to generate clusters in the network.

Table 4: Simulation parameters

Parameter	Value
ϵ_{fs}	2,59 $*10^{-10}$ J/bit/m²
ϵ_{amp}	2,52 $*10^{-14}$ J/bit/m⁴
E_{elec}	50 nJ/bit
Initial node energy	50 mJ
Number of nodes	50
Frequency	2.4 GHz
Bit rate	250 kbps
Antenna gain	1dB
High of antennas	1m
$d_{crossover}$	100m
Packet size	4000 bits
Percentage of CH	0.5%
Th_{cs}	0.7

Based on the results revealed in our previous work [25], in which we assessed the performance of vertical (VC) and horizontal (HC) compression approaches distinctly in a cluster-based wireless sensor network, we ensure the effectiveness in promoting energy efficiency using the two approaches separately. Therefore, the aim of this work is to evaluate the performance of combining the two approaches. The comparison we carried-out is between the LEACH-HC-VC according to the first configuration, LEACH-VC-HC following the second configuration and LEACH algorithm without compression scheme. This comparison is referred to as certain performance metrics namely: First Node Dies

(FND), Half of Nodes Alive (HNA) and Last Node Dies (LND). Table 5 summarizes the results of these metrics (FND, HND and LND) for LEACH and our proposed approaches.

Table 5: DIE-OFF STATISTICS OF COMPRESSION ALGORITHMS IN CLUSTER BASED WSN

Compression Approaches	Metrics			
	FND	HND	LND	Distortion
LEACH	35	69	153	--------
LEACH-HC 1st config	41	147	622	0.0103
LEACH-VC	179	359	568	0.38
LEACH-HC-VC	139	353	966	0.84
LEACH-HC 2nd config	39	117	415	0.0129
LEACH-VC-HC	181	363	662	0.45

6.1 LEACH-HC-VC results (configuration 1)

Figure 5 shows the total system energy of clustered sensor networks per transmission rounds using horizontal, vertical and combined (horizontal and vertical) compressions regarding the first configuration (LEACH-HC-VC). It is clearly shown that using selected compression algorithms perform energetically better than LEACH without compression. Therefore, compression is mandatory to prolong the network lifetime. In fact, applying only a horizontal compression at the cluster-members level (LEACH-HC) reduces the amount of transmit-ted data for a high number of nodes, which promotes the network lifetime achieving 622 rounds.
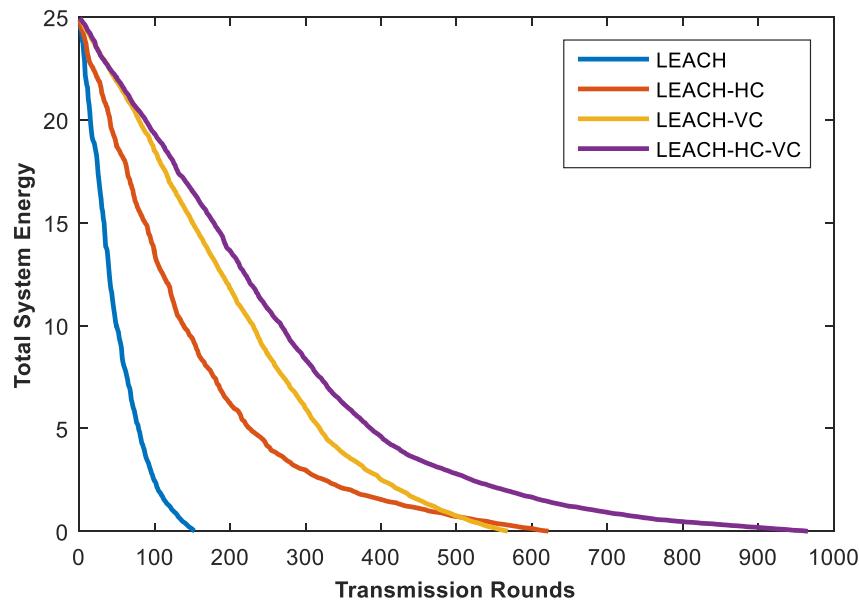

Figure 5: Total energy of cluster network for LEACH-HC-VC compression

For the vertical compression (LEACH-VC), the CH discards the redundant information collected from the cluster-members with high spatial correlation and transmits only reduced amount of data to the base station, which restricts the energy consumption at the CH level reaching 568 rounds. Regarding the combined horizontal and vertical compression following the first configuration (LEACH-HC-VC), it extends the network lifetime achieving 966 rounds by reducing the transmission energy dissipated in the entire network. Consequently, an energy is gained up to 84%, 41% and 35% from LEACH, LEACH-VC, and LEACH-HC respectively.

The maximum amount of time between (FND) and (LND) is defined as the network lifetime [44]. As the loss of any sensor node affects the whole network, the longer stability period is required. From Table 5, we assert that the FND parameter for LEACH-VC is enhanced and reaches 179 rounds, more important than other compression methods. Since there is no energy loss for processing horizontal compression. Therefore, vertical compression keeps nodes alive for a long time, while improving the stability period, which is required for some applications. The proposed scheme ensures important energy saving owing to the fact that the data transmission is integrally reduced in the clustered network. Effectively, CMs compress horizontally their collected data before transmitting it to their CH, in which another compression method (VC) takes place removing redundant information.

As a result, the communication burden is relieved at the CH, promoting subsequently the network lifetime.

6.2 LEACH-VC-HC results (configuration 2)

Figure 6 shows the total system energy per transmission rounds for the selected compression approaches following the second configuration.
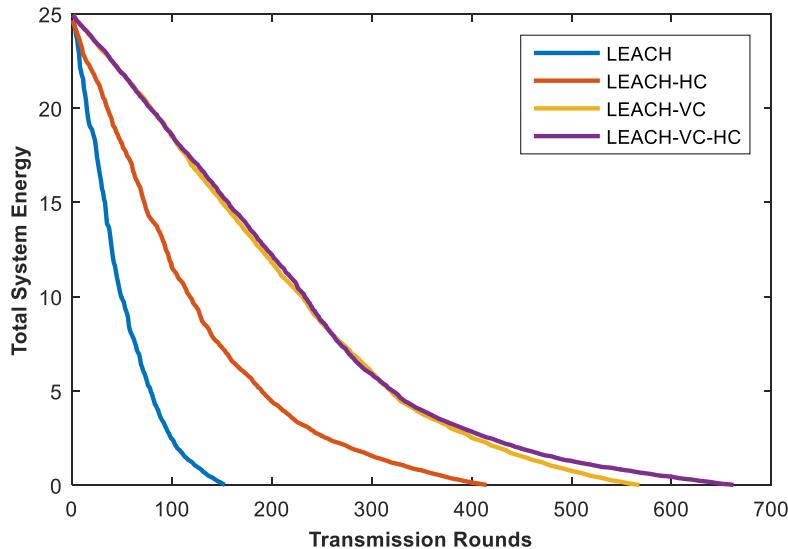

Figure 6: Total energy of cluster network for LEACH-VC-HC compression

The obtained results reveal that adopting selected compression algorithms extends the cluster-based network lifetime in various ways. In fact, applying only a horizontal compression at the CH level promotes the energy efficiency of the network achieving 415 rounds, which corresponds to the lifetime increase of 260 rounds when it is compared to LEACH without compression. With vertical compression, the total network energy is entirely consumed in 568 rounds, attaining a lifetime increase of 27% when compared to LEACH with HC. The vertical compression discards the data redundancy by constructing groups of nodes with similar data evolution. The CH sends only one representative information of each group of nodes to the base station, which makes the VCR more efficient than the HCR. Hence, the proposed algorithm based on combining horizontal and vertical compression (LEACH-VC-HC) at the CH level displays remarkable energy saving with a network lifetime reaching 662 rounds. Hence, the performances of LEACH-VC and LEACH-VC-HC are almost similar to a miniature increase of 14% in the network lifetime for the proposed solution, except when the number of alive nodes becomes small. In this phase, clustering will no longer be applied and each node compresses horizontally its collected data and transmits it directly to the base station.

Regarding the stability period for the network, for both LEACH-VC and the proposed LEACH-VC-HC, the FND increases of about 180 rounds. Consequently, the second proposed configuration extends the network lifetime, as well as the stability period, is broaden.

6.3 Distortion

The reliability of each compression method likely measured by the distortion parameter. Figure 7 shows the distortion means of horizontal, vertical and combined compression approaches following the proposed configurations. The results show the effect of adopting
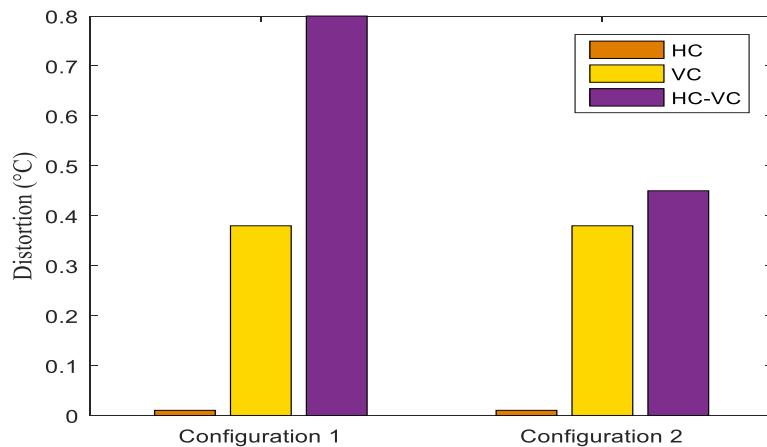

Figure 7: Distortion of selected compression algorithms in different configurations

Lossy compression algorithms on data reliability of the cluster-based network. As mentioned in Figure 7, horizontal compression presents the lowest level of distortion when compared to the vertical compression for both configurations. Furthermore, the combined algorithm is more accurate in the second configuration since HC is deployed at the CH level, which make the loss of information minimum compared to the first configuration when HC is applied in all the cluster members.

In the first configuration, each sensor node uses HC to compress its collected data before sending it to the CH. Since HC only reduces the number of samples in the time series exploiting the temporal correlation to discard data redundancies, the information accuracy is slightly affected by the data reduction. For the Vertical compression, The CH receives the raw information from each cluster-member and eliminates the entire vectors with similar data evolution taking advantage of spatial correlation. Hence, the discarded data is substantial compared to HC, which explains the difference between VC and HC in terms of distortion. Concerning the combined approach HC-VC, the distortion of collected data is significant owing to the fact that the information loss is considered at the entire network in various levels (HC in cluster-members and VC in CH).

In the second configuration, the combined algorithm VC-HC is based on applying HC to the obtained vectors from VC at the CH level, before sending it to the base station. As a result, the information loss is not considered on a large scale but only in each CH, which justifies the efficiency in data reliability of the proposed algorithm in this configuration.

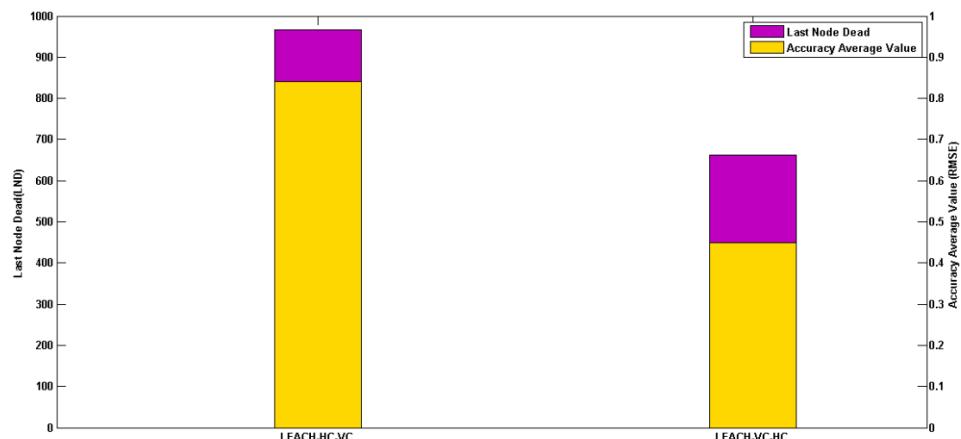- LEACH HC-VC vs LEACH VC-HC

Figure 8: Energy Vs Accuracy for the two proposed configuration

Figure 8 summarizes the results of the proposed data reduction solution in terms of energy consumption and data accuracy. The proposed compression approach promotes energy saving in both configurations. Indeed, the performance of the combined algorithm in terms of energy is more significant in the first configuration as observed in the curve, at the cost of some distortion, due to the elimination of redundant information in different levels

(cluster-members and CH). When vertical and horizontal compression algorithms are implemented in the CH (2nd configuration), the data reduction is centralized in the CH, which makes the collected information less affected by distortion. As a result, the first configuration can be used in a range of applications that demand high resource availability, and the second configuration can be suitable for applications that necessitate high information accuracy.

7   CONCLUSION

In this paper, we have evaluated the effectiveness of vertical and horizontal compression schemes in cluster-based WSNs following two proposed scenarios. Obtained results have revealed that the proposed compression algorithm HC-VC is more energy efficient in the first configuration due to the fact that HC is applied in the cluster-members, which reduces the data transmission in the entire network. However, with the second configuration, VC -HC provides consistent performance in terms of data accuracy owing to the restricted information loss, which took place only in the CH level. Hence, the proposed approach can be employed depending on the application requirements (resources availability or data reliability) solving, as a result, the trade-off between energy saving and data accuracy.

REFERENCES

[1]   M. M. N. Aldeer, "A summary survey on recent applications of wireless sensor networks," *Proceeding - 2013 IEEE Student Conf. Res. Dev. SCOReD 2013*, no. December, pp. 485–490, 2015.

[2]   M. Bagaa, Y. Challal, A. Ksentini, A. Derhab, and N. Badache, "Data Aggregation Scheduling Algorithms in Wireless Sensor Networks: Solutions and Challenges," *IEEE Commun. Surv. Tutorials*, vol. 16, no. 3, pp. 1339–1368, 2014.

[3]   A. Dâmaso, D. Freitas, N. Rosa, B. Silva, and P. Maciel, "Evaluating the power consumption of Wireless Sensor Network applications using models," *Sensors (Switzerland)*, vol. 13, no. 3, pp. 3473–3500, 2013.

[4]   R. M. Curry and J. C. Smith, "A survey of optimization algorithms for wireless sensor network lifetime maximization," *Comput. Ind. Eng.*, vol. 101, pp. 145–166, 2016.

[5]   H. A. Marhoon, M. Mahmuddin, and S. A. Nor, "Chain-based routing protocols in wireless sensor networks: A survey," *ARPN J. Eng. Appl. Sci.*, vol. 10, no. 3, pp. 1389–1398, 2015.

[6]   S. Borbèly, L. Cser, Y. M. Ostanevich, and S. Vass, "Influence of hydrocarbon chain length on micellar size," *J. Phys. Chem.*, vol. 93, no. 24, pp. 7967–7969, 1989.

[7]   P. Padhy, R. K. Dash, K. Martinez, and N. R. Jennings, "A utility-based adaptive sensing and multihop communication protocol for wireless sensor networks," *ACM Trans. Sens. Networks*, vol. 6, no. 3, pp. 1–39, 2010.

[8]   C. Y. Cho, C. L. Lin, Y. H. Hsiao, J. S. Wang, and K. C. Yang, "Data aggregation with spatially correlated grouping technique on cluster-based WSNs," *Proc. - 4th Int. Conf. Sens. Technol. Appl. SENSORCOMM 2010*, pp. 584–589, 2010.

[9]   Y. Yin, F. Liu, X. Zhou, and Q. Li, "An Efficient Data Compression Model Based on Spatial Clustering and Principal Component Analysis in Wireless Sensor Networks," *Sensors*, vol. 15, no. 8, pp. 19443–19465, 2015.

[10]   H. Harb, A. Makhoul, S. Tawbi, and R. Couturier, "Comparison of Different Data Aggregation Techniques in Distributed Sensor Networks," *IEEE Access*, vol. 5, no. 0, pp. 4250–4263, 2017.

[11]   J. Cui and F. Valois, "Simba: Similar-evolution based aggregation in Wireless Sensor Networks," *IFIP Wirel. Days*, vol. 2016–April, pp. 1–6, 2016.

[12]   S. Al Fallah and A. El Oualkadi, "On the Performance of Piecewise Linear Approximation Techniques in WSNs," *Int. Conf. Adv. Commun. Technol. Netw.*, 2018.

[13]   S. Al Fallah, M. Arioua, A. El Oualkadi, and J. El Asri, "PLA Compression Schemes Assessment in Multi-hop Wireless Sensor Networks," in *Procedia Computer Science, 9th International Conference on Ambient Systems, Networks and Technologies'(ANT 2018)*, 2018, vol. 130, pp. 279–286.

[14]   D. Zordan, B. Martinez, I. Vilajosana, and M. Rossi, "On the Performance of Lossy Compression Schemes for Energy Constrained Sensor Networking," *ACM Trans. Sens. Networks*, vol. 11, no. 1, pp. 1–34, 2014.

[15]   M. A. Razzaque, C. Bleakley, and S. Dobson, "Compression in wireless sensor networks: A Survey and Comparative Evaluation," *ACM Trans. Sens. Networks*, vol. 10, no. 1, pp. 1–44, 2013.

[16]   S. Nithyakalyani, "Data Aggregation in Wireless Sensor Network Using Node Clustering Algorithms - a Comparative Study," *Proc. 2013 IEEE Conf. Inf. Commun. Technol. (ICT 2013)*, no. Ict, pp. 508–513, 2013.

[17]   "Temprerature Data." [Online]. Available: http://di.ulb.ac.be/labo/datasets.html.

[18]   W. B. Heinzelman, A. P. Chandrakasan, S. Member, and H. Balakrishnan, "An Application-Specific Protocol Architecture for Wireless Microsensor Networks," vol. 1, no. 4, pp. 660–670, 2002.

[19]   L. Bierl, "MSP430 Family Mixed-Signal Microcontroller Application Reports," *Architecture*, no. January, p. 1088, 2000.

[20]   A. El *et al.*, "ScienceDirect PLA Assessment Sensor Networks in PLA Compression Compression Schemes Schemes Assessment in Multi-hop Multi-hop Wireless Wireless Sensor Sensor b , Networks Networks PLA Compression Schemes Assessment in Multi-hop Wireless," *9th Int. Conf. Ambient Syst. Networks Technol.*, vol. 130, pp. 279–286, 2018.

[21]   I. Ez-zazi, M. Arioua, A. El Oualkadi, and P. Lorenz, "On the performance of adaptive coding schemes for energy efficient and reliable clustered wireless sensor networks," *Ad Hoc Networks*, vol. 64, pp. 99–111, 2017.

[22]   D. Mantri, N. R. Prasad, and R. Prasad, "Grouping of clusters for efficient data aggregation (GCEDA) in wireless sensor network," *Proc. 2013 3rd IEEE Int. Adv. Comput. Conf. IACC 2013*, pp. 132–137, 2013.

[23]   M. A. Razzaque, C. Bleakley, and S. Dobson, "Compression in wireless sensor networks: A survey and comparative evaluation," *ACM Trans. Sens. Networks*, vol. 10, no. 1, pp. 5–44, 2013.

[24]   S. Al Fallah, M. Arioua, A. El Oualkadi, and J. El Asri, "On the Performance of Piecewise Linear Approximation Techniques in WSNs," in *IEEE International Conference on*

Advanced Communication Technologies and Networking' (CommNet'18).

[25]   J. El, S. Al, M. Arioua, A. El, and A. Zekriti, "Vertical and Horizontal Compression Scheme Assessment in Cluster-Based WSNs," *2018 IEEE 5th Int. Congr. Inf. Sci. Technol.*, pp. 650–655, 2018.

[26]   Y. Liang and Y. Li, "An efficient and robust data compression algorithm in wireless sensor networks," *IEEE Commun. Lett.*, vol. 18, no. 3, pp. 439–442, 2014.

[27]   C. Liu, K. Wu, and J. Pei, "An Energy-Efficient Data Collection Framework for Wireless Sensor Networks by Exploiting Spatiotemporal Correlation," *IEEE Trans. Parallel Distrib. Syst.*, vol. 18, no. 7, pp. 1010–1023, 2007.

[28]   M. Computing, "Research Article Comparison of Dwt and Dft for Energy Efficiency in Underwater," vol. 4, no. 7, pp. 128–137, 2015.

[29]   G. Quer, R. Masiero, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "On the interplay between routing and signal representation for Compressive Sensing in wireless sensor networks," *Inf. Theory Appl. Work. ITA 2009*, pp. 206–215, 2009.

[30]   G. Shen and A. Ortega, "Transform-based distributed data gathering," *Signal Process. IEEE Trans.*, vol. 58, no. 7, pp. 3802–3815, 2010.

[31]   R. Baxter, N. Hastings, A. Law, and E. J. . Glass, "Lightweight Temporal Compression of Microclimate Datasets," *Anim. Genet.*, vol. 39, no. 5, pp. 561–563, 2008.

[32]   S. Ozdemir, "Polynomial Regression Based Secure Data Aggregation for Wireless Sensor Networks," *2011 IEEE Glob. Telecommun. Conf. - GLOBECOM 2011*, no. December, pp. 1–5, 2011.

[33]   J. Lu, F. Valois, M. Dohler, and M. Y. Wu, "Optimized data aggregation in WSNs using adaptive ARMA," *Proc. - 4th Int. Conf. Sens. Technol. Appl. SENSORCOMM 2010*, pp. 115–120, 2010.

[34]   D. Zordan, B. Martinez, I. Vilajosana, and M. Rossi, "To Compress or Not To Compress: Processing vs Transmission Tradeoffs for Energy Constrained Sensor Networking," 2012.

[35]   J. Cui, "DATA AGGREGATION IN WIRELESS SENSOR NETWORKS," INSA lyon.

[36]   H. Harb, A. Makhoul, S. Tawbi, and R. Couturier, "Comparison of Different Data Aggregation Techniques in Distributed Sensor Networks," *IEEE Access*, vol. 5, pp. 4250–4263, 2017.

[37]   J. M. Bahi, A. Makhoul, and M. Medlej, "Data aggregation for periodic sensor networks using sets similarity functions," *IWCMC 2011 - 7th Int. Wirel. Commun. Mob. Comput. Conf.*, pp. 559–564, 2011.

[38]   K. T. M. Tran, S. H. Oh, and J. Y. Byun, "Well-suited similarity functions for data aggregation in cluster-based underwater wireless sensor networks," *Int. J. Distrib. Sens. Networks*, vol. 2013, 2013.

[39]   H. Harb, A. Makhoul, D. Laiymani, and A. Jaber, "A distance-based data aggregation technique for periodic sensor networks," *ACM Trans. Sens. Networks*, vol. 13, no. 4, 2017.

[40]   M. Ashouri, H. Yousefi, J. Basiri, A. M. A. Hemmatyar, and A. Movaghar, "PDC: Prediction-based data-aware clustering in wireless sensor networks," *J. Parallel Distrib. Comput.*, vol. 81–82, pp. 24–35, 2015.

[41]   N. C. Devi, S. Prabeela, V. Palanisamy, and K. Baskaran, "Distance Functions For Clustering in Wireless Sensor Networks," *Int. J. Wirel. Commun. Netw.*, vol. 3, no. 1, pp. 73–78, 2011.

[42]   M. Dev, T. Chai, and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature," pp. 1–11, 2014.

[43]   M. Arioua, Y. El Assari, I. Ez-Zazi, and A. El Oualkadi, "Multi-hop Cluster Based Routing Approach for Wireless Sensor Networks," Procedia Comput. Sci., vol. 83, no. Ant, pp. 584–591, 2016.

[44]   M. Elshrkawey, S. M. Elsherif, and M. Elsayed Wahed, "An Enhancement Approach for Reducing the Energy Consumption in Wireless Sensor Networks," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 30, no. 2, pp. 259–267, 2018.

AUTHORS

Jihane EL AASRI received the engineer degree in Network & Telecommunications engineering from the National School of Applied Sciences, Abdelmalek Essaadi University in 2014. She is currently a Ph.D. student at the Department of Information and Communication Systems at the National School of Applied Sciences of Tangier. Her research interests include in-network data processing in wsn, design of energy-efficient data aggregation and compression algorithms in WSNs.

Samia AL FALLAH received the engineer degree in Electronics and Telecommunications from the National High School of Electricity and Mecanics in 2014, Hassan II University, Casablanca. She is currently Ph.D student at the Department of Information and Communication Technologies at National School of Applied Sciences of Tangier, Abdelmalek Essaadi University, Morocco. Her research interests include wireless sensor networks, data compression and agregation methods in wireless communications, routing algorithms and signal processing.

Mounir Arioua received his Ph.D. degree in Telecommunications and Computer Science from Cadi Ayyad University of Marrakech in 2012. In 2012, he joined the National School of Applied Sciences of Marrakech as research assistant. Currently, he is an associate professor at the National School of Applied Sciences of Tetouan, University of Abdelmalek Essaadi. His research interests include wireless sensor networks, Internet of things, wireless networking and communication, wireless communications and mobile computing, real-time processing and embedded systems.

Ahmed El Oualkadi is a full professor at the department of information and communication systems at the National school of applied sciences of Tangier, Abdelmalek Essaadi University in Morocco. He is the coordinator of the Graduate engineering program in Telecommunication Systems and Networks. His research interests include analog IC, microwave & RFIC design for wireless communication, power electronics, embedded system and wireless communications applications.