IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 7 December 2022, accepted 10 January 2023, date of publication 16 January 2023, date of current version 25 January 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3237024

RESEARCH ARTICLE

Adaptive Sampling Approach Exploiting Spatio-Temporal Correlation and Residual Energy in Periodic Wireless Sensor Networks

MARWA FATTOUM, ZAKIA JELLALI, AND LEILA NAJJAR ATALLAH
COSIM, Sup'Com, Carthage University, Tunis 2083, Tunisia

Corresponding author: Marwa Fattoum (marwa.fattoum@supcom.tn)

This work was supported by the European Union (EU) through the Education, la MObilité, la Recherche et l'Innovation (EMORI) Program and managed by the Agence Nationale de Promotion de la Recherche Scientifique [National Agency for the advancement of scientific research (ANPR)] under the MOBilité des doctorants (MOBIDOC) Scheme.

ABSTRACT Energy limitation is a major issue in wireless sensor networks where a high volume of redundant data is collected periodically and transmitted through the network. Therefore, efficient energy consumption is the key solution to maximize the network lifetime. This paper proposes an adaptive sampling approach based on spatio-temporal correlation of collected data and on nodes residual energy. This approach aims to optimize sampling rates of sensor nodes while ensuring a high quality of the collected data. In addition, a data reconstruction method based on linear regression is adopted in the sink to reconstruct the missing samples due to the sampling rate reduction and adaptation compared to the case of a constant maximal sampling rate. We compared our approach with recently proposed adaptive sampling benchmark methods in different scenarios of data temporal correlation. Simulation results demonstrate the effectiveness of our proposed method in optimizing energy consumption by reducing the sampling rate while maintaining data quality. Our contribution can be applied to several fields, particularly, the field of water resources management.

INDEX TERMS Wireless sensor network, adaptive sampling, spatio-temporal correlation, residual energy, data reconstruction.

I. INTRODUCTION

A. BACKGROUND

In recent years, the term Internet of Things (IoT) has been widely used to describe solutions developed with different devices with computational capacity and connected to the Internet [1]. With IoT, the entire physical infrastructure is closely coupled with information and communication technologies; where intelligent monitoring and management can be achieved via the usage of networked embedded devices. In such system, devices are interconnected to transmit useful measurement information and control instructions via distributed sensor networks.

Wireless Sensor Network (WSN) is an enabling technology for the IoT. WSNs are regarded as a revolutionary information gathering method to build the information and communication system which greatly improve the reliability and efficiency of infrastructure systems. Compared to the wired solution, WSNs feature easier deployment and better flexibility of devices [2].

WSN is an emerging technology which has a wide array of applications including infrastructure protection,industrial sensing and diagnostics, environmental monitoring and intelligent home and responsive environment [3]. This kind of network is composed of a large number of low-cost sensor nodes where each node is equipped with a sensor, able to detect physical or environmental data measurements. The main components of a WSN are the sensor nodes and the sink node where the monitored events data are collected. The sensor nodes are usually deployed randomly over the area to be sensed in severe conditions where no power source to aliment batteries exsists. Those small devices should perform their functions by measuring and transmitting a high volume of data periodically to the sink node which might be connected to the internet.

B. MOTIVATION

The design of sustainable WSN is a very challenging issue [4]. First, the resource limitation in terms of

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenzhou Tang.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

computational capacities and memory of sensor nodes prevent the use of complex algorithms. Second, the most challenging problem in WSN is the power consumption issue. In such way, energy-constrained sensors are expected to run autonomously for long periods. It may be cost-prohibitive to replace exhausted batteries or even impossible in hostile environments. Additionally, a huge amount of data is collected and transmitted through the network. Those data can be redundant and strongly correlated in space and time.

Therefore, minimizing the energy consumption for data agregation and transmission is one of the most important design considerations in WSNs. For this aim, this paper focuses on energy efficiency. In this manner, to design an energy-efficient scheme, it is necessary to guarantee a trade-off between energy consumption and data accuracy.

For this reason, a data reduction adaptive sampling approach can be used to adapt the sampling rate of each sensor node for the sake of energy conservation while keeping sufficiently good reconstruction performance at the sink node. The quantity of sampled data will be reduced when the correlation of data is important. In consequence, this method reduces redundant data transmission by collecting less data when correlated.

C. CONTRIBUTION

The main objective of this paper is to propose a new solution that guarantees the energy efficiency of WSN while preserving the accuracy of the collected data. We will propose a spatio-temporal adaptive sampling approach based on residual energy (STASRE) in long range wide area network (LoRaWAN). The gateway (sink) will be responsible for collecting data from long range (LoRa) end nodes in its range of communication. Then, it calculates the spatial correlation in terms of geographical proximity of nodes and the spatial and temporal correlation among sensor nodes data. The new sampling rate should increase when a low spatio-temporal correlation of data is detected to acquire enough information, and, should decrease when there is a high spatio-temporal correlation among data and/or when the node residual energy reaches a critical low level. We will compare our approach to spatial-temporal correlation based approach for sampling and transmission rate adaptation (STCSTA) [5], aggregation and adaptive sampling approach (AAS) [6] and fixed sampling rate method where data is be sampled and sent to the gateway directly at the same maximal rate without any processing operation. We will study the performance of our approach in three scenarios with varying the temporal correlation of data by evaluating the energy efficiency in terms of total energy consumption and number of dead nodes per period. We will present the effectiveness of our method in reducing data without losing significant data. The integrity of the data will be reconstructed in the gateway.

The rest of the paper is organized as follows: Section II provides a survey on the work related to energy conservation within WSN. Section III describes the system network model and the energy model that manages the energy consumption. Section IV details the proposed spatio-temporal adaptive approach based on residual energy (STASRE). Section V describes the used data reconstruction method. The performance evaluation and analysis are discussed in section VI. Finally, section VII summarizes the contribution and opens some perspectives.

II. RELATED WORKS

Several energy saving strategies are studied to manage the scarce energy resource of WSN [7]. Thoses strategies can be categorized as: duty cycling schemes, energy efficient routing and data driven approaches that optimize either the collected data volume or the data acquisition process itself.

Duty cycling is one of sleep/wake trading schemes where sensor nodes operate at low power and enter in sleep mode for most of the time whenever there is no task to be performed [8]. The radio transceiver then is switched off to save energy. The authors in [9] propose a duty cycle energy management scheme based on medium access control (MAC) protocol. One part of the duty cycle is allocated for transmitting the data of the sensor node itself and another part is allocated for receiving/transmitting data of neighbor nodes. The node turns off in sleep mode when no data is available in order to conserve energy. In [10], a dynamic duty cycle (DDC) scheme is proposed to minimize the transmission delay and, then, maximize the network lifetime.

In energy efficient routing technique [11], the distance between sensor nodes and their sink is a key factor to choose the optimal path to route the collected data. This energy efficient routing path should minimize the communication cost. Low-energy adaptive clustering hierarchy protocol (LEACH), proposed in [12], is one among the most popular clustering-based routing approaches in WSN which is based on two main steps. For the first step, which is the setup phase, the selection of cluster heads and the cluster formation are performed based on energy and probability function. In steady state phase, cluster heads collect data from nodes in their respective clusters using a time division multiple access (TDMA) scheduling and perform data aggregation before transmitting to the base station. This method ensures energy load distribution in all nodes of each cluster. A clustering and routing approach based on genetic algorithm for multi objective optimization (CRMOGA) is proposed in [13]. The goal is to find a near optimal network configuration by selecting optimal cluster heads, and, finding the optimal path to route the collected data. The multi objective fitness function aims to reduce the energy consumption of the network.

Data driven approaches are based on the exploitation of spatio-temporal correlation between sensor nodes' data. Two techniques of data reduction in WSN are compression and aggregation [14] where data will be reduced before being transmitted to the sink. For compression techniques, a recovery process will be performed in the base station [15], [16]. In [17], an energy-efficient data collection based on compressive sensing technique (CCS) in cluster-based network is

proposed. This method uses block diagonal matrices as measurements matrices. Then, the base station performs recovery process to reconstruct data. In [18], an energy efficient compressed data collection method based on compressive sensing theory and quantized compressed sensing is proposed. The authors in [19] propose a data aggregation method based on clustering network topology. The clustering approach is divided into two steps. The first step consists in dividing the sensed region into correlated contiguous regions on the basis of position and data correlation. The second step consists in performing clustering on each region on the basis of residual energy. Each cluster head collects, aggregates data and routes it toward the sink. This method allows an energy load distribution among sensor nodes in each cluster. In [20], authors introduce a method based on fuzzy logic model in two levels for energy efficient clustering (FLTLC) using similarity, distances and residual energy as input fuzzy logic parameters.

Considering that sampling is one of the most consuming activities in addition to transmission and processing, one of the efficient data acquisition techniques is the adaptive sampling method where sensor nodes adapt their sampling rate on the basis of data changes detection. This method reduces the number of collected data, and then, reduces the energy consumption when the data is highly correlated. In [6], authors propose a new method, AAS, to adapt sensor sampling rate in tow steps to reduce data collection and, thus, reduce the energy consumption. This method is applied at two levels: sensor node level for the aggregation phase and cluster head level for adaptive sampling phase. The sensor node data collection is performed periodically. In the first step, each sensor node calculates the similarity between the collected data in order to eliminate redundancy and reduce the amount of collected data. Then, the cluster head determines the new sampling rate on the basis of spatial and temporal correlation of collected data. In [5], authors propose a spatial-temporal correlation method in cluster-based network to adapt sampling and transmission rates. According to the correlation degree between sensor nodes' data, calculated using Pearson correlation coefficient, the sampling rate will be either increased or decreased. In [21], a data driven adaptive sampling algorithm (DDAS) is proposed. This method is based on a revised sigmoid function that allows to dynamically change the sampling frequency, depending on a number of the most recent data and the last two detected data.

Some researchers combine adaptive sampling approach and data prediction techniques. A prediction model is then built in the sink to reconstruct and predict the data using the historical collected measurements. In [22], a data prediction with cubic adaptive sampling method is proposed. This method combines an adaptive sampling technique, based on TCP CUBIC protocol, with a prediction model based on the simple exponential smoothing model. The authors in [23] propose a data reduction method based on adaptive sampling combined with a transmitted data reduction based on a prediction model. This method reduces the energy consumption by reducing the sensed data and the radio communication burden.

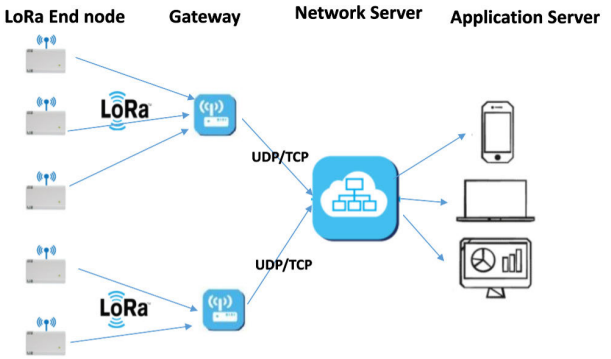

FIGURE 1. LoRa Network Architecture.

In this paper, we propose an adaptive sampling approach based first on spatio-temporal correlation of periodically collected data and based second on residual energy. The proposed scheme is developed in LoRa/LoRaWAN based architecture network. The algorithm is performed at the gateway level.

NOTATIONS

s_i	Sensor node s_i.
SR_i^p	Sampling rate of sensor s_i at period p.
SR_{max}	Maximal sampling rate.
M_{s_i}	Set of sensor nodes spatio-temporally correlated with s_i.
v_i^p	Vector of measurements of s_i at period p.
v_{ij}^p	j^{th} measurement of s_i at period $p, j = 1 \ldots SR_i$.
μ_{v_i},	Mean of measurements of node s_i.
σ_{v_i}	Standard deviation of measurements of node s_i.
R_s	Neighboring range.
d_{ij}	Euclidian distance between node s_i and node s_j.

III. NETWORK DESCRIPTION AND ENERGY MODEL

A. NETWORK DESCRIPTION

1) LoRaWAN-BASED IoT ARCHITECTURE

LoRa/LoRaWAN is one of the mostly used protocols in low power wide area network (LPWAN) [24], it provides a very low power consumption and a long range transmission. LoRaWAN uses a star topology illustrated in Fig. 1 where sensor nodes transmit the collected data to the gateway. Each received packet is relayed to the LoRaWAN network server. The latter checks the received packets to authenticate the nodes before transmitting them to the corresponding application server.

2) SYSTEM MODEL AND ASSUMPTIONS

We consider a LoRaWAN WSN composed of N sensor nodes randomly distributed in a monitored field and a gateway. We consider a periodic data collection in sensor network illustrated in Fig. 2 for sensor node s_i.

During the first period, N sensor nodes collect data using the maximal sampling rate SR_{max}. This data is stored in a

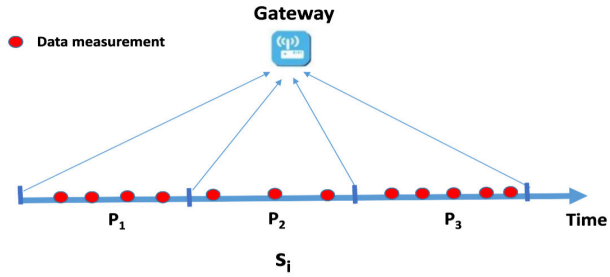

FIGURE 2. Periodic data collection of sensor node s_i.

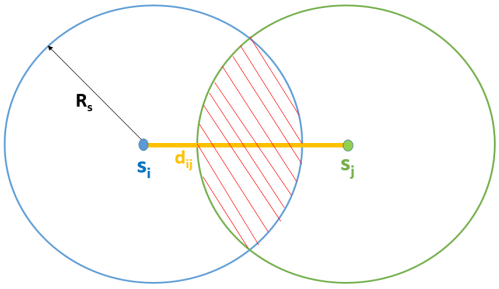

FIGURE 3. Overlap area between two sensor nodes.

matrix D. Each row corresponds to one sensor measurements during the period.

$$D = \begin{bmatrix} v_1 \\ v_2 \\ . \\ . \\ v_N \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1SR_{max}} \\ v_{21} & v_{22} & \dots & v_{2SR_{max}} \\ . & . & . & . \\ . & . & . & . \\ v_{N1} & v_{N2} & \dots & v_{NSR_{max}} \end{bmatrix}. \quad (1)$$

Then, the sampling rate will be updated through periods according to data dynamics. The stored matrix D during the period is given by the following equation:

$$D = \begin{bmatrix} v_{11} & v_{12} & \dots & Nan & .. & v_{1SR_{max}} \\ v_{21} & Nan & \dots & . & \dots & v_{2SR_{max}} \\ . & . & . & . & . & . \\ . & . & Nan & . & . & . \\ v_{N1} & v_{N2} & \dots & . & \dots & v_{NSR_{max}} \end{bmatrix}, \quad (2)$$

where '*Nan*' corresponds to missing samples whenever the sampling rate decreases with respect to its maximal value SR_{max}. Such non sampled data of the matrix D are reconstructed at the sink at each period using a reconstruction method. The possible herein envisaged reconstruction methods are the mean and median imputation in space or time and the regression imputation [29].

B. ENERGY MODEL DESCRIPTION

In order to compute the energy consumption of each sensor node, the energy consumption model based on LoRa/LoRaWAN architecture is presented in [25]. This model states that for a given node, the consumed energy E is given by:

$$E = E_{active} + E_{sleep}. \quad (3)$$

The total energy consumption E_{active} is the sum of dissipated energy by the sensor unit, the processing unit and the communication unit.

$$E_{active} = E_{wu} + E_m + E_{proc} + E_{wut} + E_T + E_R, \quad (4)$$

where E_{wu}, E_m, E_{proc}, E_{wut}, E_T and E_R are respectively, energies dissipated in the system wake-up, the data measurements, the microcontroller processing, the wake-up of the LoRa tranceiver, and the transmission and reception modes.

We assume that E_{sleep}, E_{wu}, E_{proc} and E_{wut} are negligible with respect to other energy components. In this way, Equation (3) reduces to

$$E = E_m + E_T + E_R, \quad (5)$$

where E_m is the energy dissipated by measuring SR values of the considered period.

$$E_m = SR(P_m T_m), \quad (6)$$

where P_m and T_m are the power consumption and the time duration of one measurement.

$$E_T = P_T T_T, \quad (7)$$

where P_T and T_T are the power consumption and the duration of transmission mode.

$$E_R = P_R T_R, \quad (8)$$

where P_R and T_R are the power consumption and the duration of reception mode.

In addition to the energy dissipated by the data communication, this energy model demonstrates that the sampling rate affects the energy consumption of the network in the phase of data measurements. In this paper, we will optimize the sampling rate of data measurements to enhance the lifetime of the network and reduce effectively the energy dissipation.

IV. PROPOSED SPATIO-TEMPORAL ADAPTIVE SAMPLING BASED ON RESIDUAL ENERGY (STASRE)

In this section, we will firstly detail how the spatial and temporal correlation between sensor nodes' measurements and the residual energy level are exploited in our proposed adaptive approach. Then, we will give a description of the sampling rate updating through data collection periods. For this aim, we will start by computing the spatial correlation between sensor nodes and measuring the data similarity.

A. CALCULATING SPATIAL CORRELATION BETWEEN SENSORS

In WSN generally, the closer neighbor nodes are, the more spatially correlated are their collected data. To this end, we introduce the overlap area between sensor nodes to calculate the degree of correlation between collected data of sensor nodes. The overlap area $A(s_i, s_j)$ between two neighbor nodes

s_i and s_j with spacing verifiying $d_{ij} < 2R_s$, as shown in Fig. 3, is defined by the following equation [26]:

$$A(s_i, s_j) = \frac{2}{\pi} \arccos\left(\frac{d_{ij}}{2R_s}\right) - \frac{d_{ij}}{\pi R_s^2}\sqrt{R_s^2 - \frac{d_{ij}^2}{4}}. \quad (9)$$

The parameter R_s is defined as the neighboring range, which is half the lower range with zero overlapping area. Two sensor nodes are spatially correlated when the overlap area between them is greater than a predefined threshold a.

The set of spatially correlated nodes with s_i is defined as:

$$spatial\_Corr(s_i) = \{s_j, d_{ij} < 2R_s \ and \ A(s_i, s_j) \geq a\}. \quad (10)$$

B. DATA SIMILARITY MEASURE

There are a variety of similarity functions that we can use to measure the similarity between collected data of different sensor nodes, such as Cosine similarity, Jaccard similarity, Pearson correlation coefficient [27], [28].

We define in Equation (11) the cosine similarity function $Similar(s_i, s_j)$ which measures the degree of similarity between two vectors of measurements of two sensor nodes s_i and s_j measured at the same period and maximal sampling rate SR_{max}.

$$Similar(s_i, s_j) = \frac{\sum_{k=1}^{SR_{max}} v_{ik} v_{jk}}{\sqrt{\sum_{k=1}^{SR_{max}} v_{ik}^2}\sqrt{\sum_{k=1}^{SR_{max}} v_{jk}^2}}. \quad (11)$$

The set of correlated nodes with sensor node s_i is defined as:

$$data\_Corr(s_i) = \{s_j, Similar(s_i, s_j) > s\}, \quad (12)$$

where s is the threshold for the correlation between sensor nodes.

After evaluating the spatial correlation and measuring the data similarity, a set of spatially-temporally correlated sensor nodes denoted as M_{s_i} is extracted for sensor node s_i by:

$$M_{s_i} = spatial\_Corr(s_i) \cap data\_Corr(s_i). \quad (13)$$

C. PROPOSED ADAPTIVE SAMPLING SCHEME

The different steps of our proposed adaptive sampling approach are presented by Algorithm 1 starting by the data acquisition until the sampling rate updating. Indeed, at each period the gateway adjusts the sampling rate of each sensor node taking into account the degree of temporal correlation between the nodes in the set M_{s_i} evaluated in Equation (13) and the level of residual energy E_r of the considered node. Hereafter, we give a detailed description of the three phases of our proposed STASRE scheme.

- **Phase 1: Initialization**
  At the first period, all the sensor nodes have the same initial energy equal to the maximal energy, $E_r = E_{max}$. The sampling rate takes the maximum value as follows:

$$SR_{s_i}^1 = SR_{max}, i = 1, .., N. \quad (14)$$

- **Phase 2: Sampling rate updating**
  We here discuss the update of the sampling rate at each period according to the residual energy E_{r,s_i}^p of node s_i at period p. In such way, two cases can be distinguished.
  -**Case 1:** $E_{min} < E_{r,s_i}^p < E_{max}$
  When the residual energy E_{r,s_i}^p of sensor node s_i at period p decreases and does not reach the minimum energy level E_{min}, its sampling rate is changed by exploiting both the spatial and temporal correlations. More precisely, its updated value at period $p+1$ is expressed as a weighted sum of a part related to spatial correlation and a part related to temporal correlation denoted respectively as SC^{p+1} and TC^{p+1}. In such way, the updated sampling rate for sensor node s_i at period $p+1$ combines $SC_{s_i}^{p+1}$ and $TC_{s_i}^{p+1}$ as:

$$SR_{s_i}^{p+1} = \alpha SC_{s_i}^{p+1} + (1-\alpha) TC_{s_i}^{p+1}, \quad (15)$$

where $\alpha \in [0, 1]$ is a ponderation coefficient. As already mentioned, the SC term depends on neighboring sensors spatial correlation. Therefore, if sensor s_i has one correlated node s_j, then $SC_{s_i,1}^{p+1}$ is expressed as:

$$SC_{s_i,1}^{p+1} = SR_{max}\left(1 - \tilde{\rho}_S^p(s_i, s_j)\right), \quad (16)$$

where $\tilde{\rho}_S^p(s_i, s_j)$ is the estimated spatial correlation between the two spatially-temporally correlated sensor nodes s_i and s_j given by the following equation:

$$\tilde{\rho}_S^p(s_i, s_j) = \frac{1}{SR_{max}} \sum_{k=1}^{SR_{max}} \left(\frac{|v_{ik}^p - \mu_{v_i}^p|}{\sigma_{v_i}^p}\right)\left(\frac{|v_{jk}^p - \mu_{v_j}^p|}{\sigma_{v_j}^p}\right), \quad (17)$$

If sensor node s_i has two correlated sensor nodes s_j and s_k, then $SC_{s_i,2}^{p+1}$ is expressed as:

$$\begin{aligned} SC_{s_i,2}^{p+1} &= SC_{s_i,1}^{p+1}(1 - \tilde{\rho}_S^p(s_i, s_j)) \\ &= SR_{max}(1-\tilde{\rho}_S^p(s_i, s_j))(1-\tilde{\rho}_S^p(s_i, s_k)) \end{aligned} \quad (18)$$

If sensor node s_i has more than two correlated sensor nodes, the expression of SC^{p+1} is given by:

$$SC_{s_i}^{p+1} = SR_{max} \prod_{s_j \in M_{s_i}} \left(1 - \tilde{\rho}_S^p(s_i, s_j)\right) \quad (19)$$

The expression of SC^{p+1} is inspired from the AAS method [6].
For TC temporal correlation term, we take into account the correlation between data collected by the same node in successive periods. Therefore, $TC_{s_i}^{p+1}$ for sensor s_i at period $p+1$ is expressed as:

$$TC_{s_i}^{p+1} = SR_{s_i}^p(1 - \tilde{\rho}_T^p(s_i)), \quad (20)$$

where $\tilde{\rho}_T^p(s_i)$ is the estimated temporal correlation between measurments of sensor node s_i in two successive periods $p-1$ and p given by the following equation:

$$\tilde{\rho}_T^p(s_i) = \frac{1}{SR_{max}} \sum_{k=1}^{SR_{max}} \left(\frac{|v_{ik}^p - \mu_{v_i}^p|}{\sigma_{v_i}^p} \right) \left(\frac{|v_{ik}^{p-1} - \mu_{v_i}^{p-1}|}{\sigma_{v_i}^{p-1}} \right). \tag{21}$$

In summary, the updated sampling rate SR_i^{p+1} of sensor node s_i is then expressed as:

$$SR_{s_i}^{p+1} = \alpha SR_{max} \prod_{s_j \in M_{s_i}} \left(1 - \tilde{\rho}_S^p(s_i, s_j)\right) + (1-\alpha) SR_{s_i}^p (1 - \tilde{\rho}_T^p(s_i)). \tag{22}$$

-**Case 2:** $0 < E_{r,s_i}^p \leq E_{min}$

When the residual energy of sensor s_i decreases enormously under a threshold energy level *min*, its new sampling rate will decrease according to the level of its residual energy. More precisely, its sampling rate update is discussed according to its mean spatial correlation ρ_{si}^{mean} as follows:

- If $\rho_{s_i}^{mean} \leq \epsilon$ then

$$SR_{s_i}^{p+1} = \frac{E_{r,s_i}^p}{E_{max}} SR_{max}, \tag{23}$$

- If $\rho_{s_i}^{mean} > \epsilon$ then

$$SR_{s_i}^{p+1} = \frac{E_{r,s_i}^p}{E_{max}} SR_{s_i}^p, \tag{24}$$

where

$$\rho_{s_i}^{mean} = \frac{1}{\overline{M}_{s_i}} \sum_{s_j \in M_{s_i}} \tilde{\rho}_S^p(s_i, s_j), \tag{25}$$

and ϵ is a predefined threshold, $\overline{*}$ indicates the cardinal of the set.

Throughout the different phases, the gateway calculates and sends the new sampling rate of each sensor node to be taken into account in the next period.

- **Phase 3: Data Acquisition**

Once the sampling rate is updated according to the previous discussions, each sensor node collects data using its own already updated sampling rate and transmits these readings to the gateway. The non-sampled data is replaced by $'Nan'$ when filling the sampled data in matrix **D**. However, in order to guarantee a good estimation of the spatial and temporal correlations given respectively by Equations (17) and (21), each *Nan* value is estimated by the mean imputation where the missing value is replaced by the mean value of the samples measurements.

Algorithm 1 Proposed Adaptive Sampling Scheme STASRE

Input: $SR_{max}, E_{max}, E_{min}$
$p \leftarrow 1$;
while *Alive nodes* **do**
  #Data acquisition
  $k \leftarrow 1$;
  while *Period not end* **do**
    for *(each sensor node s_i)* **do**
      if *Nothing is received* **then**
        $Data[i][k] \leftarrow Nan$;
      else
        $Data[i][k] \leftarrow v_{ik}$;
    $k \leftarrow k+1$;
  $Data \leftarrow reconstruction(missing\ \ Samples)$;
  #Spatio-temporal correlation evaluation
  for *each sensor node s_i* **do**
    $spatial\_Corr(s_i) \leftarrow find\_similar(Data)$;
    $data\_Corr(s_i) \leftarrow find\_temp\_corr(Data)$;
    $M(s_i) \leftarrow spatial\_Corr \cap data\_Corr$;
  #Sampling rate update
  for *each sensor node s_i* **do**
    $\tilde{\rho}_S \leftarrow Spatial\_Corr\_degree(M(s_i), Data)$;
    $\tilde{\rho}_T \leftarrow Temporal\_Corr\_degree(Data)$;
    if $E_{r,s_i}^p == E_{max}$ **then**
      $SR_{s_i}^p \leftarrow SR_{max}$;
    else
      if $E_{min} \leq E_{r,s_i}^p < E_{max}$ **then**
        $SR_{s_i}^p \leftarrow newSR(\tilde{\rho}_S, \tilde{\rho}_T, \alpha)$;
      else
        if $mean(\rho) \leq \epsilon$ **then**
          $SR_{s_i}^p \leftarrow newSR(SR_{max}, E_{r,s_i}^{p-1}, E_{max})$
        else
          $SR_{s_i}^p \leftarrow newSR(SR_{s_i}^{p-1}, E_{r,s_i}^{p-1}, E_{max})$
  $p \leftarrow p+1$

V. MISSING DATA RECONSTRUCTION

In this section, we describe the spatial correlation based data estimation method used in the gateway to reconstruct the missing data. Indeed, the missing data is the data that is not sampled as could be the case if the maximal sampling rate were used. This estimation method adopts the multiple linear regression model [29] to describe the data correlation among multiple neighbor nodes. Indeed, the data sensed by the sensor nodes whose locations are nearby are similar or have some relationships. For example, as shown in Fig.4, we can see that the data sensed by node 1 and node 30 which are neighbor nodes have similar variation curves. Therefore, when some sensor nodes data are missed, we can estimate them using its neighbor nodes data.

Equation (26) illustrates the linear regression model.

$$v_{it} = \beta_{i0} + \beta_{i1} v_{1t} + \beta_{i2} v_{2t} + \cdots + \beta_{im} v_{mt} + \mu_t, \tag{26}$$

where v_{it} is the data measurement of s_i at time t, and v_{kt} is the data measurement of neighbor node s_k $(k = 1, \ldots, m)$

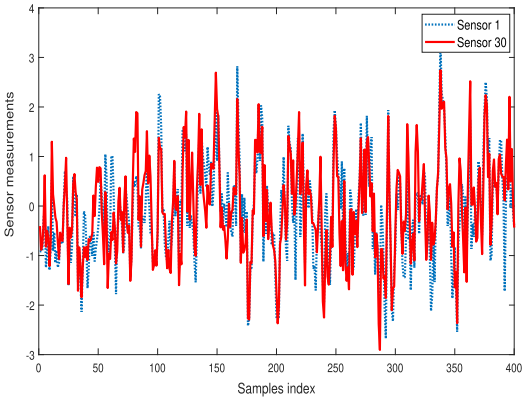

FIGURE 4. Measurements of two neighbor nodes.

at the same time, β_{ik} is the partial correlation coefficient corresponding to v_{kt}, β_0 represents a basic lag of the value at time t and μ_t is the random error.

In order to estimate the missing data, we collect the data in matrix $\mathbf{X}$ from h sets of m neighbor nodes measurements v_{ij}, $1 \leq i \leq m$ and $1 \leq j \leq h$, where $h - m \geq 2, j = 1, \ldots, h$ corresponds to time index.

$$\mathbf{X} = \begin{pmatrix} 1 & v_{11} & v_{21} & \cdots & v_{m1} \\ 1 & v_{12} & v_{22} & \cdots & v_{m2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & v_{1h} & v_{2h} & \cdots & v_{mh} \end{pmatrix}. \tag{27}$$

such that $\mathbf{X}\beta = \mathbf{V}$, where $\mathbf{V} = (v_{i1}, \ldots v_{ih})^T$ and $\beta = (\beta_{i0}, \ldots, \beta_{im})^T$.

After a learning phase about the sensed data of neighbor nodes, it is possible to estimate the coefficients as given:

$$(\tilde{\beta}_{i0}, \tilde{\beta}_{i1}, \ldots, \tilde{\beta}_{im}) = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{V}. \tag{28}$$

Therefore, the missing data can be estimated as follows:

$$\tilde{v}_{it} = \tilde{\beta_{i0}} + \tilde{\beta}_{i1}v_{1t} + \tilde{\beta}_{i2}v_{2t} + \cdots + \tilde{\beta}_{im}v_{mt}. \tag{29}$$

VI. PERFORMANCE EVALUATION AND ANALYSIS

A. SCENARIO DESCRIPTION

We consider a WSN composed of N sensor nodes randomly deployed in a squared field $(L \times L)m^2$ and one gateway placed in the center of the sensed field.

In our scenario, we assume that sensing nodes serve to collect environmental parameters such as temperature, humidity, light, pressure, which are generally highly correlated both in space and time domains. Also, we consider a periodic WSN where sensor nodes collect and store data during a predefined period p. After each period, they aggregate data and send a set of measurements to the gateway. The gateway receives data measurements and performs the adaptive sampling algorithm to determine the new sampling rate of each sensor node to apply at the next period.

We assume that all the sensor nodes have the same neighboring range, transmitting range and initial energy.

The parameters setting are fixed in Table. 1.

TABLE 1. Simulation parameters.

N	30
Sensed field	$(3000X3000)m^2$
Gateway	in the center of the sensed field
E_T [25]	$0.59mJ$
E_R [25]	$0.27mJ$
E_m [25]	$0.26mJ$
E_{init}	$0.5J$
SR_{max}	20 samples per period
R_s	$500m$

B. SPATIO-TEMPORAL CORRELATED SYNTHETIC DATA GENERATION

1) SPATIAL CORRELATION MODELS

To generate correlated readings of the sensors according to the spatial dimension, we adopt the most conventional correlation model, the exponential model [19] where the mutual spatial correlation between i^{th} and j^{th} nodes is given by:

$$\rho_S(s_i, s_j) = \exp\Big(-\Big(\frac{d_{ij}}{\theta_1}\Big)^{\theta_2}\Big),\ \theta_1 > 0,\ 0 < \theta_2 \leq 2. \tag{30}$$

From this correlation model, a network correlation matrix $\mathbf{B}$ is generated with $(i, j)^{th}$ element given by $\rho_S(s_i, s_j)$. Then, to generate correlated random variable, we use the Cholesky decomposition. The vector of correlated variables is given by $\mathbf{Cx}$ where $\mathbf{C}$ is the Cholesky decomposition of $\mathbf{B}$, with $\mathbf{B} = \mathbf{CC}^T$ and $\mathbf{x}$ is a vector of uncorrelated Gaussian random variables.

2) SPACE TIME CORRELATED SIGNAL MODEL

To enforce the temporal correlation, we adopt the autoregressive filter according to which the i^{th} node measurement evolves after a time interval ΔT as: [30]

$$v_i(t_0 + k\Delta T) = \rho_T(\Delta T)v_i(t_0 + (k-1)\Delta T) + \sqrt{1 - \rho_T(\Delta T)^2}\epsilon(t_0 + k\Delta T), \tag{31}$$

where $\epsilon(t_0 + k\Delta T)$ is an i.i.d. random Gaussian Noise, $k = 1, 2, .., SR_{max}$, ΔT denotes the duration of one period of time and $\rho_T(\Delta T)$ is a temporal correlation coefficient to be fixed according to requirements. If we are in the case of important correlation, ρ_T should take high values and for low correlation, ρ_T should have low values. In our study, we focus on evaluating the performance of our proposed adaptive sampling approach for different scenarios of temporal correlation. For this, threes scenarios, low, medium and high temporal correlation have been discussed by conserving the same spatial correlation degree. In such way, Fig.5 illustrates an estimate of the considered spatial correlation versus normalized distance evaluted according to Equation (17). Also, as shown in Fig. 6, an estimated temporal correlation is displayed for each scenario with a mean spatial correlation equal to 0.7. The displayed temporal correlation is evaluated according to Equation (21).

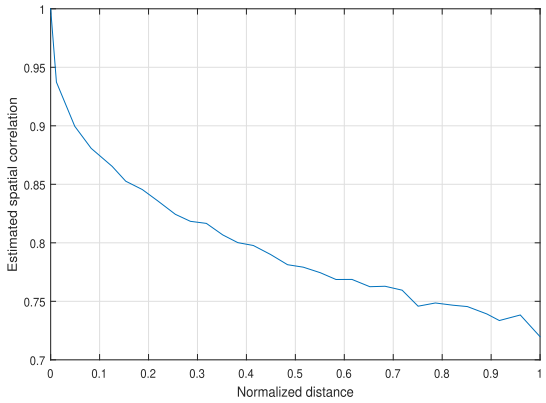

FIGURE 5. Estimated Spatial correlation.

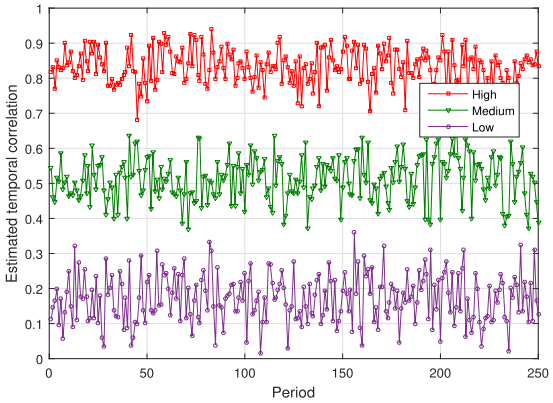

FIGURE 6. Estimated Temporal Correlation.

C. SIMULATION RESULTS: COMPARISON TO BENCHMARKS

We compare our proposed adaptive sampling method to two recently published methods AAS and STCSTA and to the Fixed SR method where all data will be sent to the gateway directly at a maximal rate and without any processing operation. We evaluate the performance of the proposed scheme and of the benchmark methods in different scenarios by varying the temporal correlation of generated data. In particular, we envisage 3 scenarios: high, medium and low temporal correlation as shown in Fig. 6.

1) SCENARIO 1: HIGH TEMPORAL CORRELATION

The first scenario represents the case of high temporal correlation of data generated of each sensor node through periods. The estimated temporal correlation corresponds to the curve High in Fig. 6.

We evaluate the energy efficiency by studying the lifetime of the network in terms of total energy consumption and number of dead nodes per period. We vary the minimun residual energy threshold E_{min} for our proposed algorithm. We also compare the efficiency of each algorithm in reducing the number of sampled data in percentage. Figures 7 and 8 display respectively the total residual energy of sensor nodes and the number of dead nodes through periods. The performances of our proposed method and STCSTA are close. After around 250 periods of communication, the network still has

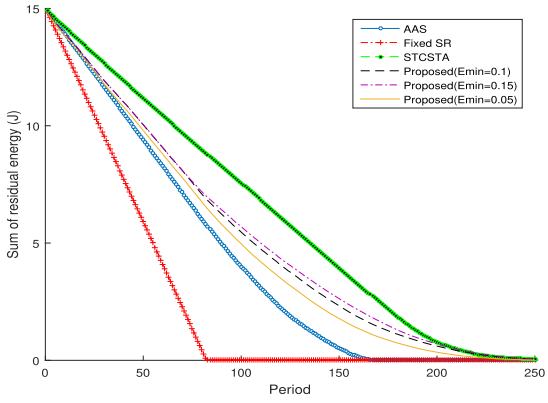

FIGURE 7. Scenario 1: Sum of residual energy vs. period.

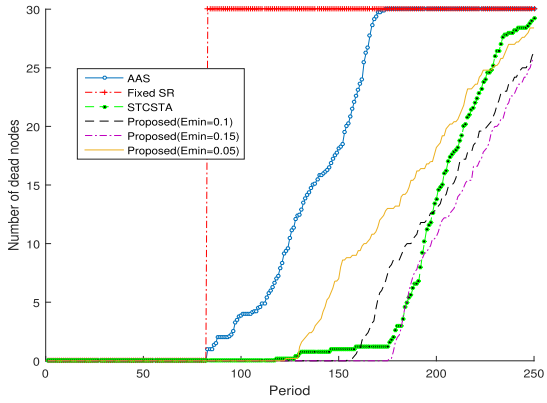

FIGURE 8. Scenario 1: Dead node vs. period.

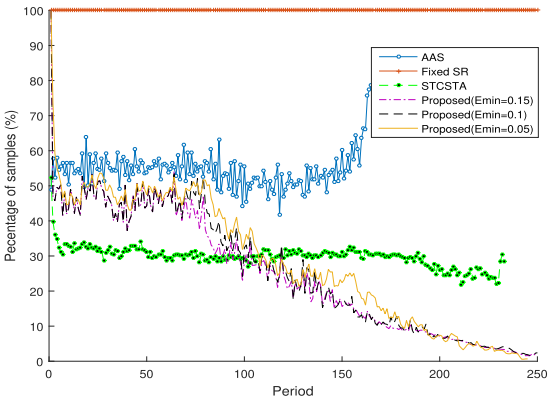

FIGURE 9. Scenario 1: Percentage of samples vs. period.

alive nodes while the AAS and the Fixed SR networks are totally disconnected respectively at around period 160 and period 80. Our proposed method outperforms all the other methods for all values of threshold energy level E_{min} in terms of remaining alive nodes and residual energy. Fig. 9 represents the percentage of reduction of samples number at a period p with respect to processing at a maximal sampling rate SR_{max}, and is calculated as:

$$\%Samples_p = \frac{\sum_{k=1}^{N_{alive}} SR_{s_k}^p \times 100}{N_{alive} \times SR_{max}}. \quad (32)$$

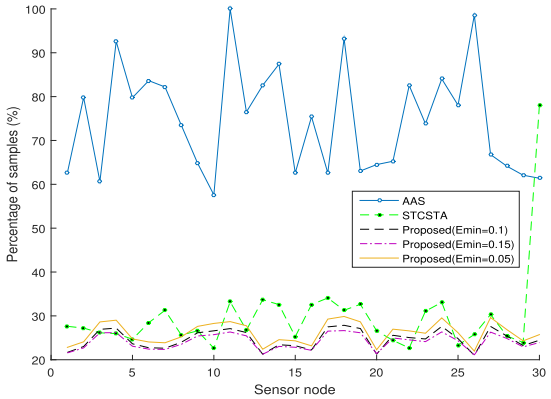

FIGURE 10. Scenario 1: Pecentage of samples per sensor node.

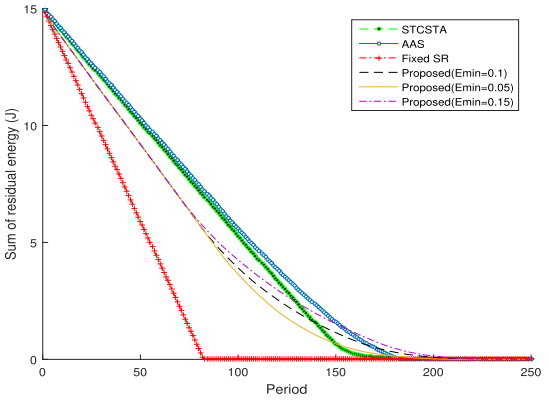

FIGURE 11. Scenario 2: Sum of residual energy.

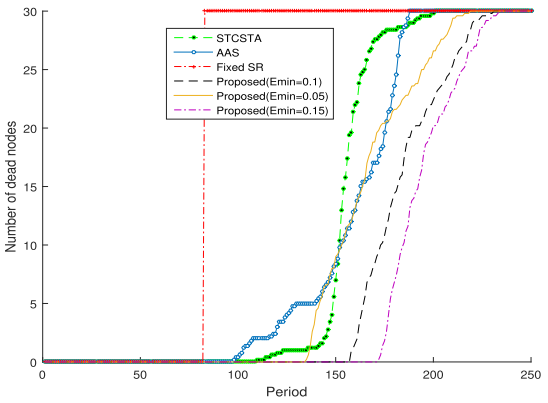

FIGURE 12. Scenario 2: Dead Nodes vs Period.

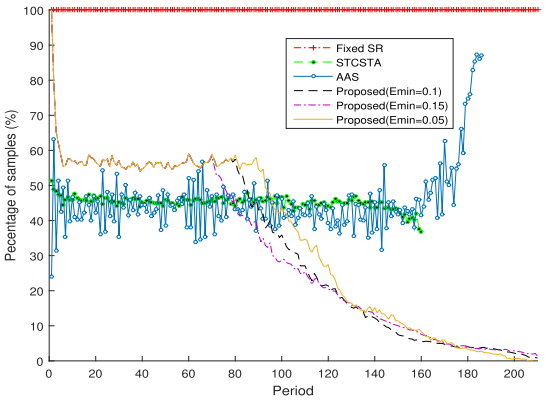

FIGURE 13. Scenario 2: Percentage of samples per period.

In addition to the data correlation, our proposed method uses the residual energy level as a factor to decrease the sampling rate when the nodes don't have enough residual energy to work at full capacity. This factor prolongs the netwok lifetime. Figures shows that AAS has the lowest percentage of samples volume reduction through periods. STCSTA has almost steady percentage of samples throughout periods whereas our scheme has a decreasing percentage of samples through periods as it accounts for energy depletion. Fig. 10 represents the percentage of reduction of samples number throughout P_{max} periods with respect to full maximal sampling rate SR_{max}, for a given i^{th} node, and is evaluated as:

$$\%Samples_i = \frac{\sum_{p=1}^{P_{max}} SR_{s_i}^p \times 100}{P_{max} \times SR_{max}}, \tag{33}$$

where P_{max} is the number of periods of the network lifetime.

Figures 10 shows that AAS has the lowest percentage of samples volume reduction for different sensors throughout the network lifetime. Our proposed method has the lowest percentage of samples between 20 to 30 percent by node, while STCSTA percentage of samples exceeds 30 percent.

2) SCENARIO 2: LOW TEMPORAL CORRELATION

The second scenario represents the case of low temporal correlation of data generated of each sensor node through periods. The estimated temporal correlation corresponds to the curve Low in Fig. 6. Our proposed method outperforms AAS, Fixed SR method and STCSTA in terms of energy preservation. This is due to the use of additional criteria to adapt the sampling rate such as the correlation of collected data by the same sensor in successive periods and the residual energy of sensor nodes. Fig. 11 illustrates the total residual energy of sensor nodes. The network is totally disconnected after around 170 periods for STCSTA, 180 periods for AAS and remains alive until more than 200 periods for our proposed method. In Fig. 13, the percentage of sampling rate for our proposed method is higher than the other methods because of the low correlation between data measurements. However, its number of dead nodes illustrated in Fig. 12 is lower. This is due to the use of the residual energy criterion to compute the sampling rate when the residual energy of a sensor node decreases under a threshold level. This allows to maximize the node lifetime and manage the remaining little amount of energy. It is better to collect a low number of measurements than not to collect any information.

3) SCENARIO 3: MEDIUM TEMPORAL CORRELATION

The third scenario represents the case of medium temporal correlation of data generated by each sensor node through periods. The estimated temporal correlation corresponds to the curve Medium in Fig. 6. Our proposed method outperforms STCSTA, AAS ans Fixed SR methods in terms of

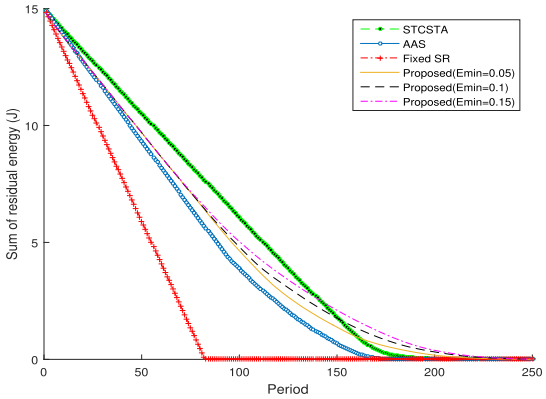

FIGURE 14. Scenario 3: Sum of residual energy.

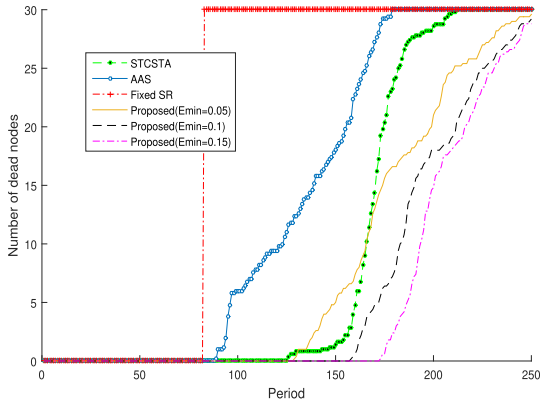

FIGURE 15. Scenario 3: Dead Nodes vs Period.

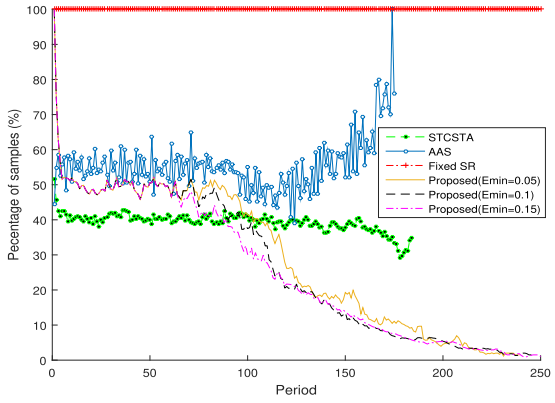

FIGURE 16. Scenario 3: Percentage of samples per period.

energy efficiency as illustrated in Figures 14 and 15. The networks of AAS and STCSTA methods are totally disconnected respectively at around periods 170 and 210, while our network of the proposed method is still connected after around 250 periods.

In TABLE 2, we have summarized the percentage of sampling rate $\%SR$ calculated using Equation (32) for the first 100 periods and the number of collected data SR_{tot} of our proposed method for the three scenarios where the temporal correlation of the generated data varies. Higher is the temporal correlation of data, lower is the percentage of sampling rate, which reduces the number of collected data, and then, reduces the energy consumption of the network.

TABLE 2. Summary of sampling rate of the proposed method.

Proposed method	Temporal correlation	%SR	$\mathbf{SR_{tot}}$
	High	45	39218
	Medium	50	39837
	Low	58	40935

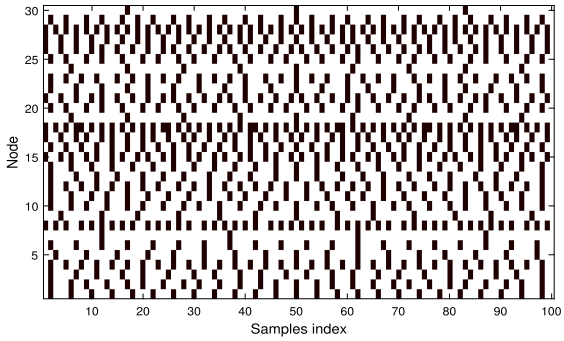

FIGURE 17. Node's sampling rate repartition.

D. DATA RECONSTRUCTION PERFORMANCE

Data quality is a crucial factor that trades off with the sampling rate reduction. In this way we evaluate for node s_i the normalized squared error denoted by reconstruction error (RE) given by the following equation to validate the accuracy of the estimated data at the sink level:

$$RE(s_i) = \frac{||v_i - \hat{v}_i||_2^2}{||v_i||_2^2}, \tag{34}$$

where v_i is the vector of effective measurements of sensor node s_i and $\hat{v}_i$ is the estimated vector of measurements, reconstructed at the sink level.

To study the impact of the sampling rate adaptation, we consider a periodic WSN using a maximum sampling rate $SR_{max} = 100$ and a low temporal correlation among sensing data.

In Fig.17, we present an example in 2D of the obtained sampling rate of each sensor node for a given period of our proposed adaptive sampling through the samples index. In such representation, the black color indicates that the sensor is in sleep mode, whereas in white zone the acquisition and then the sampling steps are done. In order to exploit the linear regression for missing data estimation, a learning phase is required. For this aim, we use a mean based imputation approach to recover the missing points. In such way, three illustrations based on linear regression are presented.

- **Illustration** 1**:** We here assume that we have only one sensor node with missing data and therefore we have a knowledge of the exact data transmitted by the other nodes. This case is the ideal case for missing data recovery, and is therefore studied here as a reference benchmark.
- **Illustration** 2**:** During each period, the missing data of each sensor node at time t are estimated according to the Equation (29). In such case, the mean based imputation is called to estimate the non sampled data of all the other sensor nodes still with missed points.

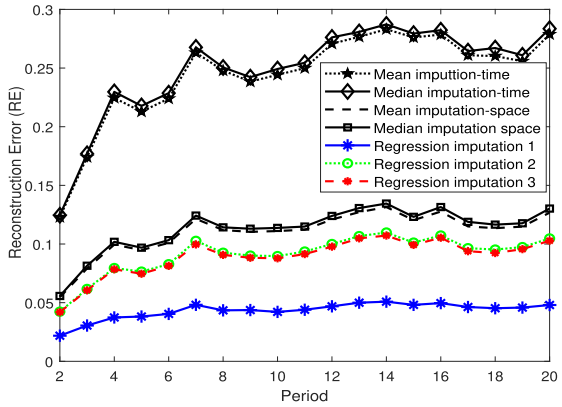

FIGURE 18. Reconstruction error of our proposed method.

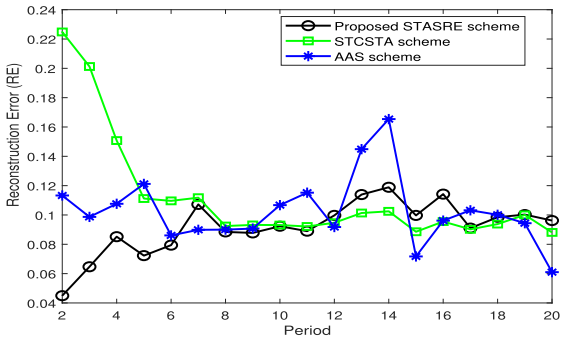

FIGURE 19. Reconstruction error of benchmark methods.

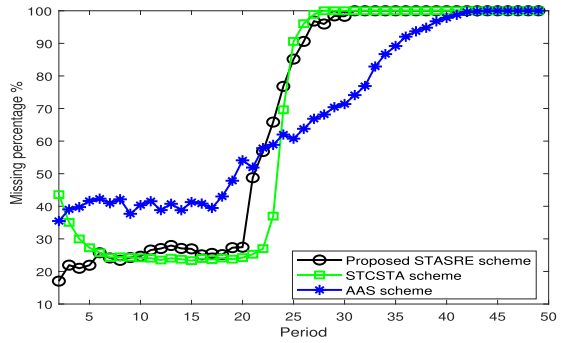

FIGURE 20. Missing rate percentage vs Period.

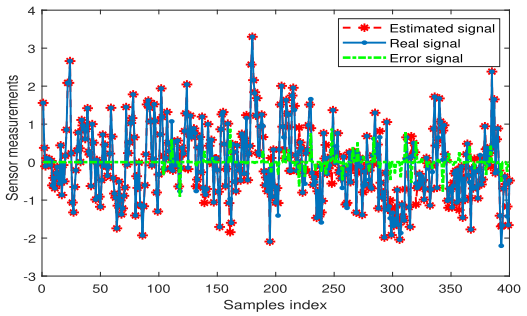

FIGURE 21. Reconstructed signal.

- **Illustration** 3**:** This illustration is similar to illustration 2 but we consider only the neighbor nodes of the considered sensor to estimate the non sampled points with Equation (29).

Fig.18 shows the reconstruction error, RE, of the estimated data of our proposed method using the regression imputation as described above, and the mean and median imputation that respectively replace the missing data by the mean and the median value of sampled measurements in space and time. The regression imputation presents a better signal reconstruction than the other methods. In the following simulations, we choose the case of regression imputation of illustration 2 referred to "*Regression imputation* 2" in Fig.18 to compare it with the reconstruction error of benchmark methods.

In Fig.19, we present the reconstruction error for the different adaptive sampling rate methods evaluated as the mean of the formula presented in Equation (34) over nodes. The three methods, AAS, STCSTA and our proposed STASRE, have an acceptable recovery error around 0.1 while our proposed STASRE allows a longer network lifetime as proved earlier. Fig.20 represents the percentage of missing samples through periods which is the complementary percentage with $\%Samples_p$ given by Equation (32). The curve increases when the nodes die. Fig.21 illustrates the reconstructed signal of our proposed method during 4 successive periods. The linear regression imputation used to reconstruct the signal is able to exploit spatial correlation and achieve a satisfactory reconstructed signal where the error signal is near to zero.

In conclusion, according to the last simulation, sensor nodes can reduce their sampling rate without losing significant data while the gateway reconstructs the missing samples. The integrity of the data will be recuperated in the gateway.

VII. CONCLUSION

In this paper, we propose an adaptive sampling rate algorithm for data acquisition reduction in periodic WSN. This algorithm is performed in the gateway using the spatio-temporal correlation of sampled data and residual energy of sensor nodes. After a number of periods, a reconstruction method based on linear regression is performed to reconstruct missing data. We compare our proposed method with recent benchmark methods in different scenarios of temporal correlation. First, we prove that our proposed method outperforms benchmark methods in maximizing the network lifetime by evaluating the sum of residual energy of sensor nodes and the number of remaining alive nodes. Then, to more emphasize the performance of our proposed method, we evaluate the quality of recovered data at the gateway using the regression imputation. We prove that the integrity of data is recuperated at the gateway.

In future work, we aim to introduce compressive sampling (CS) for spatial domain compression which allows to deactivate the sensing for some nodes until the active nodes attain a residual energy threshold. In this way, the sampling rate will be adaptive not only in time but also in space, and CS also allows the recovery of missing spatial data.

REFERENCES

[1] S. Villamil, C. Hernández, and G. Tarazona, "An overview of Internet of Things," *Telkomnika*, vol. 18, no. 5, pp. 2320–2327, 2020.

[2] K. W. Al-ani, A. S. Abdalkafor, and A. M. Nassar, "An overview of wireless sensor network and its applications," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 17, pp. 1480–1486, 2020.
[3] H. Jin, S. H. Nengroo, I. Kim, and D. Har, "Special issue on advanced wireless sensor networks for emerging applications," *Appl. Sci.*, vol. 12, no. 14, p. 7315, Jul. 2022.
[4] S. Villamil, C. Hernandez, and G. Tarazona, "Challenges and issues for wireless sensor networks: A survey," *J. Global Sci. Res.*, vol. 6, no. 1, pp. 1079–1097, 2021.
[5] G. B. Tayeh, A. Makhoul, C. Perera, and J. Demerjian, "A spatial–temporal correlation approach for data reduction in cluster-based sensor networks," *IEEE Access*, vol. 7, pp. 50669–50680, 2019.
[6] A. Karaki, A. Nasser, C. A. Jaoude, and H. Harb, "An adaptive sampling technique for massive data collection in distributed sensor networks," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 1255–1260.
[7] S. A. Inad, K. M. A. Alheeti, and S. S. Al-Rawi, "Energy conservation strategies in wireless sensor networks: A review," in *Proc. Int. Conf. Modern Trends Inf. Commun. Technol. Ind. (MTICTI)*, Dec. 2021, pp. 1–6.
[8] R. C. Carrano, D. Passos, L. C. S. Magalhaes, and C. V. N. Albuquerque, "Survey and taxonomy of duty cycling mechanisms in wireless sensor networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 181–194, 1st Quart., 2013.
[9] N. S. Usha, M. Hossen, and S. Saha, "Efficient duty cycle management for reduction of energy consumption in wireless sensor network," in *Proc. 2nd Int. Conf. Electr. Electron. Eng. (ICEEE)*, Dec. 2017, pp. 1–4.
[10] Y. Liu, A. Liu, N. Zhang, X. Liu, M. Ma, and Y. Hu, "DDC: Dynamic duty cycle for improving delay and energy efficiency in wireless sensor networks," *J. Netw. Comput. Appl.*, vol. 131, pp. 16–27, Jan. 2019.
[11] N. A. Pantazis, S. A. Nikolidakis, and D. D. Vergados, "Energy-efficient routing protocols in wireless sensor networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 551–591, 2nd Quart., 2013.
[12] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless micro sensor networks," in *Proc. 33rd Hawaii Int. Conf. Syst. Sci.*, Jan. 2000, p. 10.
[13] M. Fattoum, Z. Jellali, and L. N. Atallah, "A joint clustering and routing algorithm based on GA for multi objective optimization in WSN," in *Proc. IEEE 8th Int. Conf. Commun. Netw. (ComNet)*, Oct. 2020, pp. 1–5.
[14] S. Randhawa and S. Jain, "Data aggregation in wireless sensor networks: Previous research, current status and future directions," *Wireless Pers. Commun.*, vol. 79, pp. 1–71, Dec. 2017.
[15] Z. Jellali, L. N. Atallah, and S. Cherif, "Linear prediction for data compression and recovery enhancement in wireless sensors networks," in *Proc. Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Sep. 2016, pp. 779–783.
[16] H. Cui, S. Zhang, X. Gan, M. Shen, X. Wang, and X. Tian, "Information recovery via block compressed sensing in wireless sensor networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2016, pp. 1–6.
[17] M. T. Nguyen, K. A. Teague, and N. Rahnavard, "CCS: Energy-efficient data collection in clustered wireless sensor networks utilizing block-wise compressive sensing," *Comput. Netw.*, vol. 106, pp. 171–185, Sep. 2016.
[18] W. Wei, W. Dan, and J. Yu, "Energy efficient distributed compressed data gathering for sensor networks," *Ad Hoc Netw.*, vol. 58, pp. 112–117, Apr. 2017.
[19] M. Fattoum, Z. Jellali, and L. N. Atallah, "A two-level clustering based on position, data correlation and residual energy in WSN," in *Proc. 15th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Jun. 2019, pp. 961–966.
[20] M. Fattoum, Z. Jellali, and L. N. Atallah, "Fuzzy logic-based two-level clustering for data aggregation in WSN," in *Proc. 17th Int. Multi-Conf. Syst., Signals Devices (SSD)*, Jul. 2020, pp. 360–365.
[21] T. Shu, M. Xia, J. Chen, and C. de Silva, "An energy efficient adaptive sampling algorithm in a sensor network for automated water quality monitoring," *Sensors*, vol. 17, no. 11, p. 2551, Nov. 2017.
[22] G. B. Tayeha, A. Makhoula, D. Laiymania, and J. Demerjian, "DPCAS: Data prediction with cubic adaptive sampling for wireless sensor networks," in *Proc. Int. Conf. Green Pervasive Cloud Comput.*, vol. 10232, May 2017, pp. 353–368.
[23] G. B. Tayeh, A. Makhoul, D. Laiymani, and J. Demerjian, "A distributed real-time data prediction and adaptive sensing approach for wireless sensor networks," *Pervasive Mobile Comput.*, vol. 49, pp. 62–75, Sep. 2018.
[24] J. D. C. Silva, J. J. P. C. Rodrigues, A. M. Alberti, P. Solic, and A. L. L. Aquino, "LoRaWAN—A low power WAN protocol for Internet of Things: A review and opportunities," in *Proc. Int. Multidisciplinary Conf. Comput. Energy Sci.*, Jul. 2017, pp. 1–6.
[25] T. Bouguera, J.-F. Diouris, J.-J. Chaillout, R. Jaouadi, and G. Andrieux, "Energy consumption model for sensor nodes based on LoRa and LoRaWAN," *Sensors*, vol. 18, no. 7, p. 2104, Jun. 2018.
[26] R. K. Shakya, Y. N. Singh, and N. K. Verma, "Generic correlation model for wireless sensor network applications," *IET Wireless Sensor Syst.*, vol. 3, no. 4, pp. 266–276, Dec. 2013.
[27] C. Cassisi, P. Montalto, M. Aliotta, A. Cannata, and A. Pulvirenti, "Similarity measures and dimensionality reduction techniques for time series data mining," in *Advances in Data Mining Knowledge Discovery and Applications*, Sep. 2012, doi: 10.5772/49941.
[28] S. Dhimal and K. Sharma, "Energy conservation in wireless sensor networks by exploiting inter-node data similarity metrics," *Int. J. Energy, Inf. Commun.*, vol. 6, no. 2, pp. 23–32, 2015.
[29] L. Pan, H. Gao, H. Gao, and Y. Liu, "A spatial correlation based adaptive missing data estimation algorithm in wireless sensor networks," *Int. J. Wireless Inf. Netw.*, vol. 21, no. 4, pp. 280–289, Oct. 2014.
[30] D. Zordan, G. Quer, M. Zorzi, and M. Rossi, "Modeling and generation of space-time correlated signals for sensor network fields," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011, pp. 360–365.

MARWA FATTOUM received the engineering degree in computer science, in 2016, and the master's degree in information and communications technologies from the National School of Engineering of Carthage (ENICARTHAGE), in 2017. She is currently pursuing the Ph.D. degree with the Research Laboratory Communications, Signal and Image (COSIM), Higher School of Communication of Tunis (SUP'COM). She works on data aggregation and energy optimization problem in wireless sensors networks.

ZAKIA JELLALI received the bachelor's degree in physics from the University of Sfax, Tunisia, in 2010, the master's degree in telecommunication, in 2012, and the Ph.D. degree in information and communications technologies from the Higher School of Communication of Tunis (SUP'COM), University of Carthage, Tunisia, in 2017. She is an Assistant Professor with the Higher Institute for Technological Studies of Gabes, Tunisia. She carries her research activities with the Research Laboratory Communications, Signal and Image, SUP'COM. She works on data compression and aggregation and energy optimization problem in wireless sensors networks. Her research interests include sparse representation recovery and their applications to wireless communication.

LEILA NAJJAR ATALLAH received the degree in engineering from the Ecole Polytechnique de Tunisie, in 1997, and the M.Sc. and Ph.D. degrees in automatic and signal processing from Ecole Supérieure d'Electricité SUPELEC, Université Paris XI, France, in 1998 and 2002, respectively. Since July 2018, she has been a Full Professor with the Higher School of Communication of Tunis (SUP'COM), University of Carthage, Tunisia, where she leads her research in signal processing and wireless communications with the Communication, Signal and Image (COSIM) Laboratory.