Abstract—Understanding environmental conditions in different locations is crucial for addressing air-pollution issues. While wireless sensor networks offer the capability to monitor environmental quality locally, they face challenges related to power supply. This study introduces a low-power Wireless Sensor Network (WSN) employing distributed compressed sensing for a time-series environmental monitoring system. The proposed method achieves data compression at individual sensor nodes, mitigating power consumption during data transmission. Conversely, data restoration occurs on a server equipped with ample computing resources. This study investigates the power-saving impact of the proposed approach and identifies the optimal compression ratio. Experimental findings reveal a coefficient of determination of 0.9 or higher at a compression ratio of 90%. Our results indicate that the distributed compressed sensing-based WSN proposed in this study is effective for time-series environmental monitoring systems, offering valuable insights for future research endeavors.

Keywords—Wireless Sensor Network (WSN), compressed sensing, distributed compressed sensing, time-series environment monitoring, power saving

I. INTRODUCTION

In recent years, the evaluation of environmental qualities has been grounded in observations of the natural environment. Wireless Sensor Networks (WSN) frequently serve as tools for measuring these qualities, encompassing various data types such as time-series CO₂ concentration and atmospheric pollution. Despite their versatility, WSNs, predominantly reliant on battery power, confront challenges due to limited power resources. Consequently, numerous studies have focused on enhancing power efficiency in WSNs [1–9], employing strategies such as Receiver-Initiated Medium Access Control (RI-MAC) to adjust data transmission timing [5, 6] and compressed sensing to reduce data transmission costs [5–11]. Compressed sensing, commonly used in fields such as magnetic resonance imaging, compresses and restores data, offering potential applications in WSNs [10, 11].

While RI-MAC and compressed sensing contribute to power conservation, only minimal attention has been paid to the correlation between original measurement data and restored data in a time series. The accuracy of restoration process, which involves thinning out or linearly interpolating transmitted data cannot be guaranteed, posing a risk of errors in environmental values. Several studies investigate the consequences of decreasing data transmission frequency on power savings and the influence of reduced data transmission frequency on prediction accuracy [12, 13].

Advanced theories, such as distributed compressed sensing, have been explored [14–18]. In this theory, multiple sensor nodes measure target data, and the original data is recovered using fewer samples, leveraging the correlation among data measured by multiple nodes. Distributed compressed sensing technique is adapted to WSN, video coding, etc. [20, 21]. While distributed compressed sensing holds promise for maintaining the correlation between original and restored data in a time series, discussions on this aspect are lacking in WSNs using compressed sensing. Additionally, the accuracy and power consumption of restored data vary across WSN applications, underscoring the importance of integrating these factors a compression ratio aligned with application requirements.

This study introduces a power-saving method employing distributed compressed sensing in a time-series environmental monitoring system with WSNs. The proposed method enhances power consumption efficiency while considering the accuracy of restored data derived from raw observations. We assess the efficacy of distributed compressed sensing in WSNs, evaluating the accuracy of restored data concerning the number of sensor nodes and restoration parameters. Furthermore, we determine the optimal compression ratio, considering both power reduction and restored data accuracy. This endeavor successfully achieves the measurement and collection of
time-series environmental data while optimizing power consumption. Our research attempts to make local-scale environmental monitoring practical, which will contribute to solving environmental problems and predicting natural disasters. The remaining sections are organized as follows: Section II presents the related works. The proposed method is presented in Section III. Section IV describes the experimental model configuration. Section V presents experimental results. Section VI provides discussions and concluding remarks.

II. RELATED WORKS

A. Receiver-Initiated Medium Access Control (RI-MAC)

RI-MAC, a technique wherein a receiving node transmits a beacon and, upon receiving the beacon, the sending node with data transmits it [5, 6]. In the event of a packet collision preventing successful data reception the receiving node transmits and alternate beacon containing a back-off region. This region assigns a random wait time to each sending node, minimizing the likelihood of collision upon retransmission. This mechanism effectively mitigates packet collision recurrence. Subsequently, if a receiving node’s transmitted beacon goes unacknowledged, the node enters a sleep period. This approach optimizes throughput and minimizes power consumption by regulating transmission and reception, retransmission processing, and sleep timing. However, it may not be inherently compatible with time-series data.

B. Power Saving in Time-Series Environment Data Measurement

Bhandari et al. [12] and Engmann et al. [13] explored systems for measuring or predicting time-series environmental data with WSNs. These studies delve into the consequences of decreasing data transmission frequency on power savings and the influence of reduced data transmission frequency on prediction accuracy. The server complements the data, thinned due to reduced transmission, through a completion process (e.g., linear or spline interpolation). Nevertheless, as these methods solely thin out transmitted data or linearly interpolate thinned data, errors in environmental values are feasible compared to the actual data.

C. Compressed Sensing

Compressed Sensing (CS) is a technique that reduces data volume by applying a random matrix, known an observation matrix, to data acquired on the sensor side and subsequently restoring the data on the server side [7–11]. The random matrix must adhere to the Restricted Isometry Property (RIP), a condition typically satisfied by matrices with elements following a normal or Bernoulli distribution [14]. As only the matrix product calculation occurs on the sensor side, computational load and power consumption remain low. The restoration algorithm, which is computationally intensive, is executed on the server side, benefiting from ample computational resources. Consequently, the processing is tailored to the capabilities of each resource. The restoration algorithm addresses a problem with a non-uniquely determined solution. However, if the original data can be expressed as sparse data, the problem transforms into an L1-norm minimization problem [7]. A change-of-basis matrix, incorporating elements such as the Discrete Cosine Transformation (DCT) matrix and Discrete Wavelet Transformation (DWT) matrix, is utilized to represent this sparsity.

D. Compressed Sensing Based WSN

In WSN, transmitted data often contains various redundancies. To address this, a WSN system based on CS is introduced [9]. This CS based WSN method aims to diminish information volume, communication load, and power consumption by organizing the WSN into three layers: the sending layer (comprising sensor nodes), the processing layer (an intermediate layer for temporary data storage), and the application layer (responsible for processing and analyzing acquired data). Additionally, the compression algorithm is anticipated to offer encryption benefits, rendering it applicable to home-WSN devices. However, leveraging the correlation of information obtained by sensor nodes poses challenges.

To tackle the energy limitation issue in large underwater WSNs (UWSNs) for extended environmental monitoring, a solution is presented in [22, 23]. Wang et al. [22] proposes an energy-efficient data collection scheme utilizing CS in UWSNs, specifically designed for environmental monitoring over fading channels. The model introduced in Ref. [22] establishes a compressed sensing-based UWSNs data collection approach, capitalizing on the spatial sparsity of underwater environmental data to reduce the required number of sensor nodes.

III. PROPOSED METHOD

A. Proposed WSN

This study introduces a WSN system for environmental monitoring, employing distributed CS. The WSN comprises multiple sensor nodes designed to measure temperature, atmospheric pressure, and CO₂ concentration. The structure of the proposed WSN system is illustrated in Fig. 1. Each sensor node includes a microcomputer (Arduino UNO), a temperature-humidity-pressure sensor (AE-BME280), a CO₂ sensor (MH-Z19C), and a wireless communication module (XBees). In scenarios where several sensor nodes are utilized to measure indoor environmental qualities, a correlation between the values recorded by each sensor node is likely. Consequently, the adaptation of distributed CS to WSN focused on indoor environment monitoring is deemed highly effective. The proposed method follows the outlined flow.

- Each sensor node observes time-series data.
- Using a compression algorithm, each sensor node compresses the data and transmits it to the server.
- After transmitting data, sensor nodes enter a sleep mode.
- The server receives data from each sensor node.
- Using the restoration algorithm, the server reconstructs the compressed data.
The server employs moving averages to correct for errors.

This approach aims to enhance the efficiency of environmental monitoring by utilizing distributed CS in a WSN context.

Reducing power consumption during data transmission can be achieved by compressing the data subsequent to each sensor node’s data measurement. Data compression, involving the product of the observation matrix and data, demands minimal computational load and power consumption, rendering it feasible for implementation at each sensor node with low loads. Upon data transmission, the server receives data from each sensor node and proceeds to restore the compressed data utilizing a restoration algorithm, specifically employing a γ-weighted L1 norm minimization problem.

The restoration algorithm, being computationally intensive and power-consuming, presupposes a computationally resourceful server. Leveraging the commonality of data acquired by multiple sensor nodes, the restoration algorithm simultaneously restores multiple compressed data, enabling the process to be executed with fewer samples. Additionally, the restored data undergoes processing through a moving average to rectify potential errors. In the experimental setup, the average value is derived by summing the 10 data points preceding and following the relevant value, subsequently replacing the relevant value with the obtained average.

This study uniquely performs compression and restoration processes on the empirically observed data, enabling the determination of the accuracy of the restored data. The evaluation of restored data accuracy, using the coefficient of determination, serves as a focal point for assessing the practicality and efficacy of the proposed method.

### B. Compressed Process

If the number of sensor nodes is J, the data observed at the j-th sensor node is \( x_j \in \mathbb{R}^n \). The \( m \times n \) observation matrix used for compression is \( \Phi_j \). Additionally, the sensor data \( y_j \in \mathbb{R}^m \) after compression can be expressed as follows:

\[
y_j = \Phi_j x_j \in \mathbb{R}^m
\]

Typically, the observation matrix in distributed CS is a random matrix. Ideally, a random matrix following a normal or Bernoulli distribution is preferred. The choice of the random matrix is flexible, as long as the default transformation matrix \( \Psi \) and the observation matrix \( \Phi \) used in the restoration maintain an incoherent relationship. Commonly utilized random matrices include:

1. **Gaussian Random Matrix (GRM)**
   - A gaussian random matrix where each element of the \( m \times n \) matrix independently follows a normal distribution with a mean of 0 and variance of \( \frac{1}{m} \)

2. **Discrete Fourier Transformation Matrix (DFTM)**
   - A matrix with rows randomly selected from the discrete Fourier transformation matrix.

These random matrices are selected based on their ability to maintain an incoherent relationship with the default transformation matrix, ensuring effective performance in distributed CS applications. As an example of observation matrix, a sparse random matrix composed of 0 or 1 is used. An example of the relationship between the observation matrix and sensor data is shown in Fig. 2.

### C. Restoration Process

Distributed CS, premised on the assumption that data from multiple sensors exhibit correlations, is a methodology designed to leverage this correlation for the restoration of compressed data [16–18]. This approach is rooted in the concept of joint sparsity, wherein the sparsity is not attributed to individual signals but rather to the entire set of interrelated data. Three Joint Sparsity Models (JSMs) have been proposed in prior research [16]. Our experiment primarily focuses on JSM-1, deemed practical for environmental variables such as temperature.

In JSM-1, each dataset comprises a common component \( z_c \) and an innovation component \( z_i \) with the original data being the sum of these two components. In other words, the original data \( x_j \) at the \( j^{th} \) sensor node is expressed as follows:

\[
x_j = z_c + z_i.
\]

The restoration process involves solving a γ-weighted L1 norm minimization problem to compute \( Z = (z_0, z_1, z_2, \ldots, z_J) \). If \( Y \) represents the data post-compression and \( \phi \) is the observation matrix, then \( Z = (z_0, z_1, z_2, \ldots, z_J) \) can be determined by solving the equation under the constraint \( Y = \phi Z \). In this context, the respective weight, \( \gamma \), must be duly considered.
\[ Z = \text{argmin} \gamma_c \| z_c \|_1 + \sum_{j=1}^{N} \gamma_j \| z \|_1 \]  

Our study investigates the impact on restoration accuracy with variations in the number of sensor nodes and explores the accuracy variations corresponding to changes in the value of \( \gamma_c \). Additionally, we determine the optimal compression ratio by formulating an objective function that takes into account both the reduced power ratio and restoration accuracy.

### IV. EXPERIMENT

In our experiment, we employed sensor nodes equipped with the proposed method to observe data on indoor temperature, atmospheric pressure, and CO\(_2\) concentration. Specifically, indoor temperature and atmospheric pressure were measured by three sensor nodes, while CO\(_2\) concentration was measured by two sensor nodes. Additionally, a sensor node without the proposed method (denoted as “w/o compression algorithm”) was included in the setup, measuring indoor temperature, atmospheric pressure, and CO\(_2\) concentration. Each sensor node conducted measurements of environmental values at 20 s intervals. The sensor node implementing the compression algorithm retained data from 10 measurements and compressed it based on the specified compression ratio. For instance, with a compression ratio of 50%, five compressed data points were transmitted after 10 measurements. By contrast, the method under comparison (sensor node without CS) transmitted measured data every 20s. The compression of the time-series data for 10 measurements were performed to maintain the correlation between the time series, facilitating subsequent data recovery. The parameters of the sensor nodes are detailed in Table I.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Unit</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating Voltage</td>
<td>V</td>
<td>5.00</td>
</tr>
<tr>
<td>Consumed Current in Active Mode</td>
<td>mA</td>
<td>16.26</td>
</tr>
<tr>
<td>Consumed Current in Sleep Mode</td>
<td>mA</td>
<td>0.0657</td>
</tr>
<tr>
<td>A Packet Size</td>
<td>byte</td>
<td>127</td>
</tr>
<tr>
<td>Data Rate</td>
<td>kbps</td>
<td>9.6</td>
</tr>
</tbody>
</table>

In this study, the compression algorithm employs a random matrix as the observation matrix, where each element adheres to a standard normal distribution with a mean of 0 and a variance of 1. For instance, when the compression ratio is set at 30%, the data is compressed by an 7 \times 10 observation matrix. Additionally, the compression algorithm in this study utilizes a DCT matrix as the basis transformation matrix.

Let \( \phi_k[i] \) be the DCT basis. DCT basis for the \( i \)-th column and \( k \)-th line is defined as follows:

\[
\phi_k[i] = \begin{cases} 
1 \quad (k = 0) \\
\frac{2}{\sqrt{N}} \cos \left( \frac{(2i + 1)k\pi}{2N} \right) \quad (k = 1, 2, \ldots, N - 1),
\end{cases}
\]

where \( N \) is the number of lines. The sparse data can be reconstructed by solving a \( \gamma \)-weighted L1 norm optimization problem. The values of \( \gamma_1, \gamma_2, \) and \( \gamma_3 \) in the \( \gamma \)-weighted L1 norm minimization problem are fixed at 1.0, and the value of \( \gamma_c \) is varied from 0.1 to 2.0 to assess the accuracy of the recovery. Additionally, the accuracy is evaluated by altering the number of sensor nodes in the WSN that measure temperature and atmospheric pressure, ranging from one to three, and by adjusting the number of sensor nodes measuring CO\(_2\) concentration from one to two.

The accuracy is verified using the coefficient of determination. A higher the coefficient of determination indicates that the restored data closely aligns with the original data. The coefficient of determination (\( R^2 \)) is calculated using the following equation:

\[
R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2},
\]

where \( y_i (i=1, 2, \ldots, n) \) be the original data, \( \hat{y}_i \) be the restored data and \( \bar{y} \) be the average of original data. The coefficient of determination between measurements obtained at sensor nodes equipped with the compression algorithm and those from the sensor node without the compression algorithm (considered as correct values) is determined using Eq. (2). This coefficient signifies the degree to which the estimated results align with the actual data, with a value closer to 1 indicating a better fit of the restored results to the actual data.

Each sensor node captures environmental values, retaining a portion of the measured time-series data and compressing them using a defined observation matrix. The sensors transmit the compressed data, effectively reducing power costs during transmission by minimizing the amount of transmitted data. On the receiving end, the receiver node restores the received data by solving a weighted L1 norm minimization problem on the server side.

To assess accuracy, we established sensor nodes implementing the compression algorithm and sensor nodes without the compression algorithm at each observation point, as illustrated in Fig. 3. The values from sensor nodes without the compression algorithm were considered as the ground truth, and accuracy was evaluated using the coefficient of determination.
V. RESULTS

A. Result and Discussion About $\gamma_c$

In the parameters related to temperature and atmospheric pressure, $\gamma_1, \gamma_2$ and $\gamma_3$ are fixed at 1.0. For the parameters related to CO$_2$, $\gamma_1$ and $\gamma_2$ are set to 1.0, while the value of $\gamma_c$ is varied from 0.1 to 2.0. The accuracy of temperature, atmospheric pressure, and CO$_2$ data when $\gamma_c = 0.1$ and $\gamma_c = 1.0$ is depicted in Fig. 4. The figure illustrates the coefficient of determination on the vertical axis and the compression ratio (ranging from 80% to 99%) on the horizontal axis. The number of sensor nodes is 3 for temperature and atmospheric pressure data and 2 for CO$_2$ data, with the average coefficient of determination of the recovered data for each sensor node being presented.

For temperature data, $\gamma_c = 1.0$ demonstrates stability and accuracy, with a coefficient of determination close to 1 below a 90% compression ratio. Conversely, $\gamma_c = 0.1$ resulted in a coefficient of determination of approximately 0.8. In the case of atmospheric pressure data, the accuracy of $\gamma_c = 1.0$ is more consistently stable than that of $\gamma_c = 0.1$, albeit with a marginal difference. For CO$_2$ data, both $\gamma_c = 0.1$ and 1.0 exhibit instability at compression ratios above 90%, but $\gamma_c = 1.0$ displays stable accuracy at compression ratios below 90%. Notably, across all data types, setting $\gamma_c = 1.0$ consistently results in a coefficient of determination of 0.9 or higher, even at an 80% compression ratio.

B. Result When Changing the Sensor Node Number

In the proposed WSN, $\gamma_c$ is fixed at 1.0, and the number of sensor nodes in each WSN is varied. The average accuracy of each sensor node is depicted in Fig. 5, with the coefficient of determination on the vertical axis and the compression ratio (ranging from 80% to 99%) on the horizontal axis.

For temperature data, the accuracy stabilizes at high compression ratios of 97% or higher when there are many sensor nodes. However, no substantial difference is noted in the later compression ratios.

In the case of atmospheric pressure and CO$_2$ data, distributed CS by multiple sensor nodes proves effective at compression ratios of 90% or higher.

C. Result Regarding the Optimal Compression Ratio

Building on the presented outcomes, we contemplate deducing the optimal compression ratio by assessing the interplay between the power-saving effect and the accuracy of restored data. The objective function is formulated as follows, with the power-saving effect (power reduction ratio) denoted as $E (0 \leq E < 1)$ and the restoration accuracy as $R (R \leq 1)$.

$$f = \alpha E + (1 - \alpha)R \quad (0 \leq \alpha \leq 1)$$

The power-saving effect is directly proportional to the compression ratio. Power consumption $P$ is calculated using the following equation, where $I$ (mA) represents the current, $N$ (bit) is the data size and $D$ is the data rate. This power consumption is per node.

$$P = \frac{IN}{D}$$

The power saving effect refers to the percentage of power reduction during data transmission. Power consumption by the compression algorithm on the sensor side is not considered due to its negligible impact. The compression ratio that maximizes the function $f$ can be determined as the optimal compression ratio. This is achieved by increasing $\alpha$ when aiming to enhance the power-saving effect and decreasing $\alpha$ when prioritizing the accuracy of restored data.

Fig. 6 shows the value of $f$ when $\alpha = 0.5$. The figure presents the value of $f$ on the vertical axis and the compression ratio (ranging from 80% to 99%) on the horizontal axis.

Fig. 6 shows the compression ratio at which $f$ reaches the maximum values. This compression ratio can be deemed as the optimal compression ratio.
D. Power-Saving Efficiency

Table II provides an overview of the power consumption of the sensor nodes during a 12 h operation. The measured power consumption results demonstrate that implementing the proposed method leads to reduced power consumption. For instance, at a compression ratio of 70%, the power consumption of the sensor node implementing the compression algorithm is 78% of that of the sensor node without it. These findings indicate that despite the compression algorithm contributing to increased power consumption, the overall effect of compression results in significant power savings.

<table>
<thead>
<tr>
<th>Compression ratio</th>
<th>Energy Consumption for a Sending Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 % (without CS)</td>
<td>5.40 (Wh)</td>
</tr>
<tr>
<td>30%</td>
<td>4.26 (Wh)</td>
</tr>
<tr>
<td>50%</td>
<td>4.24 (Wh)</td>
</tr>
<tr>
<td>70%</td>
<td>4.20 (Wh)</td>
</tr>
</tbody>
</table>

E. Performance Comparison

Several studies regarding time-series environment data measurement with WSNs [12, 13] delve into the consequences of decreasing the data transmission frequency on power savings and the influence of reduced data transmission frequency on prediction accuracy. The server complements the data, thinned due to reduced transmission, through a completion process (e.g., linear or spline interpolations). An improvement in the reduction ratio of the transmission numbers leads to lower power consumption, and a coefficient of determination closer to 1 indicates a smaller error than the original environmental value. In other words, if the coefficient of determination remains close to 1 even with a higher reduction in the number of transmissions, the power is saved, and the values are kept close to the original data.

Fig. 7 shows the coefficient of determination as a function of the transmission-number reduction rate. The solid line shows the coefficient of determination for the method using the compression algorithm (proposed method), and the dashed line shows the coefficient of determination for the method using spline interpolation (conventional method). From this figure, the proposed method maintains a higher coefficient of determination than the method with spline interpolation, even when the transmission reduction rate is increased. This result shows the effectiveness of the proposed method.

VI. DISCUSSION AND CONCLUSION

We examine how adjusting parameters impacts data accuracy in environmental monitoring. Changing the weight parameter $\gamma_c$ in the $\gamma$-weighted L1 norm minimization problem enhances temperature, atmospheric pressure, and CO$_2$ data stability when $\gamma_c$ is around 1.0 or higher. However, excessively small values of $\gamma_c$ may compromise local handling, reducing accuracy.

Results regarding sensor nodes show significantly improved accuracy with two or more sensors, slightly better with a larger number. This underscores the effectiveness of distributed compressive sensing in time-series environmental monitoring.

We also consider optimal distributed compression sensing parameters to balance power-saving and data accuracy. Using $\alpha=0.5$, the temperature and atmospheric pressure data compression ratio ranges from 90% to 96%, and for CO$_2$ data, it is 90%. This suggests a well-balanced approach to considering power-saving effects and data accuracy at $\alpha=0.5$.

In conclusion, this study proposed power-saving methods using distributed CS in time-series environmental monitoring WSN, achieving high accuracy in restored data. The results indicate that the proposed method can obtain reliable restored data with significant power-saving effects at high compression ratios. Additionally, the study derived the optimal compression ratio based on the nature of WSN and observed data. Using a large $\alpha$ prioritizes power consumption, while a small $\alpha$ prioritizes data accuracy, enabling the derivation of the optimal compression ratio for specific data and applications. Additionally, the
proposed method maintains a better concordance rate between the original and restored data than the conventional method (with spline interpolation) with reduced transmission numbers. The obtained results underscore the effectiveness of the proposed distributed compressed sensing–used WSN for power savings in time-series environmental monitoring systems, offering valuable insights for future research endeavors.

ACKNOWLEDGMENT

This work was partly based on results obtained from a project, JPNP20004 and the joint research program of CERes, Chiba University (2023).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

N.K. suggested the basic concept of this study; S.M. suggested the algorithm demonstrated in this study; designed and performed the experiments; S.M. and N.K. analyzed the experiment data; all authors had approved the final version.

REFERENCES


Copyright © 2024 by the authors. This is an open access article distributed under the Creative Commons Attribution License (CC BY-NC-ND 4.0), which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.