Energy Aware Smart Sensing and Implementation in Green Air Pollution Monitoring System

Sushmita Ghosh[a], Payali Das[b], Shakthipriya Murugesh[c], Swades De[b], Shouri Chatterjee[b], Marius Portmann[d]
[a]UQ-IITD Academy, IIT Delhi, New Delhi, India
[b]Department of EE, IIT Delhi, New Delhi, India
[c]Department of ECE, Government College of Technology, Coimbatore, India
[d]School of ITEE, University of Queensland, Australia

***Abstract*—The field-deployed Internet-of-Things (IoT) sensor nodes are powered by rechargeable batteries. The nodes are equipped with energy harvesters to harvest energy from the environment to replenish the batteries and continue the sensing operations. However, due to the high energy consumption of the power-hungry sensors, the nodes still suffer from energy depletion issues. Towards a green IoT system, an energy aware adaptive sensing algorithm is proposed in this paper. For a multi-sensing node, a learning-aided smart sensing strategy is developed to find a set of optimal sensors to be activated in the next measurement cycle depending on the cross-correlation factors and the sensing energy consumption. The parameters of inactive sensors are predicted from cross-correlated parameters of active sensors using Gaussian process regressor model. Further, the algorithm is implemented in a solar powered air pollution monitoring system to analyze the performance of this method. The proposed method saves 68% energy of the node compared to the nearest competitive method, while the sensing error is within the limit.**

***Index Terms*—Energy awareness, energy harvesting, energy sustainability, green IoT, smart sensing, temporal and cross-correlation.**

I. Introduction

Internet-of-Things (IoT) is gathering a huge momentum in every sector like industries, health-care, agriculture, border surveillance, education, transport, etc. The IoT networks are consisting of many sensor nodes, edge nodes, and base stations that are wirelessly connected to each other. The sensor nodes are deployed in the field to sense the environmental parameters and transmit the sensed data at the edge node for further processing. Limited battery capacity of the wireless sensor nodes limits their continuous sensing operations. Therefore, the nodes are equipped with energy harvesters such as solar, Rf, wind, piezoelectric, etc. to harvest energy from ambient. In the context of massive IoT deployment scenario, the fifth-generation technology has become a very challenging in wireless research. Due to its huge applications, the demand for green IoT communication systems are growing day by day. It has been observed that the number of IoT products has surpassed the number of humans. Here many aspects comes into picture in terms of energy consumption, recharging, security etc. which are gaining significant interest in research. Massive IoT devices will consume massive energy in both sensing and data transmission which has to accommodate with efficient energy replenishment policy [1].

A. Related Works

The environmental parameters being monitored by a sensor node possess temporal and cross-correlation, which makes the system predictive. Significant studies have been observed in the literature on node-level and network-level adaptive sensing. In the context of node-level adaptive sensing, the work in [2] developed an Adaptive sampling algorithm to estimate the optimal sampling frequencies online using the Nyquist criteria, where the sampling rate is calculated from maximum frequency of the sensing signal. Similarly, three data collection algorithms were proposed in [3] to adapt the sampling interval with the variation of the environment using the one-way Anova model, Euclidean Distance Function, and Jaccard Similarity Function. An age of sample based data collection framework was proposed in [4], which exploits the similarity between two temporal samples and the energy efficiency of the node to estimate optimum sensing instances for a time window. Another node-level adaptive sensing algorithm was proposed in [1], which selects a set of sensors to collect samples in the next measurement cycle by exploiting the cross-correlation among the different sensing parameters and the energy consumptions of the sensors of a multi-sensing node (a node having multiple sensors to monitor multiple parameters in the environment). The parameters belong to the inactive sensors are predicted from the parameters of the active sensor set using Gaussian Process Regressor (GPR) model. The sampling interval of the active sensors in [1] was decided based on the Nyquist criteria, whereas, the work in [5] exploits the temporal correlation of each sensing parameter to find its sampling interval.

In case of a densely deployed Wireless Sensor Networks, both the spatial and temporal correlations of the sensing signals can be used to select a subset of sensor nodes for collecting samples of a specific sensing parameter. The framework proposed in [6] outlined an optimization function for selecting an optimal set of sensing nodes to collect samples for the subsequent measurement cycle, taking into account sensing quality, dynamic energy resources available at the node, and process dynamics.

B. Literature Gap and Motivation

The algorithm proposed in [2] and [3] were implemented in a snow monitoring application and a humidity sensing node, respectively, where the node consists of a single sensor.

Moreover, the sampling interval of the sensor is not adaptive to the energy availability of the node. However, the learning-based adaptive sensing algorithms studied in [1] and [5] considered multi-sensing nodes and exploit the correlation among the data to further optimize the energy consumption. Although the learning-based methods studied in [4], [1], [5], and [6] are more efficient in the context of sensing energy consumption, the nodes are considered as battery operated. Since the energy available in the battery reduces with time, the optimization function reduces the sensing accuracy by turning off the power-hungry sensors to increase the battery's lifetime. However, the sensor selection scenario is different in the case of energy harvesting nodes. Moreover, the learning-based algorithms are not implemented in any real-life system to evaluate the efficiency of these methods. To this end, an energy aware smart sensing framework is developed for an energy harvesting multi-sensing node. The measured parameter values at the multi-sensing node are sent to the edge node to exploit the cross-correlation and temporal correlation of the sensing parameters to find the optimal active set of sensors and the sampling interval of the active sensors, respectively. The missing time series samples of the parameters belong to the active set are predicted from the collected samples using GPR model. The parameters of the sleep set are also predicted from the parameters of the active set using the GPR model. The performance of the algorithm is implemented in green air pollution monitoring multi-sensing node, where the node is powered by a solar energy harvester.

C. Contributions

The main features and contributions of this work are:

1) An energy aware, learning-aided adaptive sensing strategy to optimize the energy sustainability for energy harvesting wireless sensor nodes is presented in this work. The cross-correlation and temporal correlation among the parameters monitored in the same environment are exploited to remove the sparsity in the data.
2) The proposed framework presents an UCB-based optimization function to find the optimal sampling interval of the sensors and an optimal set of sensors in a multi-sensing node to activate in the next measurement period by exploiting the cross-correlation factor, energy consumed by the active sensors, and energy availability at the node. The parameters belong to the inactive sensor set are predicted from the cross-correlated parameters of the active sensors using GPR models.
3) The proposed sensor selection framework is adaptive to the energy harvested at the sensor node.
4) The proposed algorithm is implemented in an air pollution monitoring system powered by a solar energy harvester, where the sensor node consists of six sensors monitoring eight parameters (temperature, humidity, NO_2, Ozone, CO, SO_2, $PM_{2.5}$, and PM_{10}). At the end of each measurement period, the field deployed sensor node transmits the collected data in a batch to the edge node using WiFi. The edge node apply the adaptive sensing algorithm to find the optimal system parameters and transmits the information to the sensor node. The experimental results validate the efficiency of the algorithm.

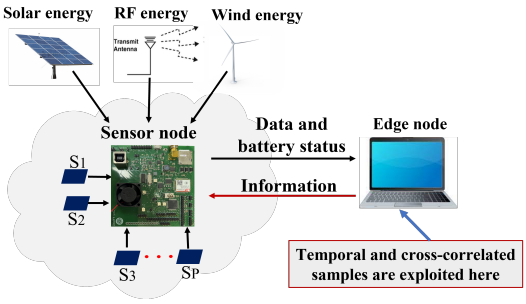


Fig. 1: Data acquisition model.

Organization: The energy aware adaptive sensing framework and its implementation are discussed in Section II and Section III, respectively. Section IV presents the experimental results and Section V concludes the work.

II. Proposed System Model and Smart Sensing Framework

The proposed system model and the energy adaptive and learning-aided sensing framework are discussed in this section. Section II-A describes the system model. Section II-B and Section II-C describe the GPR-based signal prediction model and the UCB-based optimization function, respectively. Section II-D presents the proposed energy aware adaptive sensing framework.

A. System Model

The proposed data acquisition model, shown in Fig. 1, consists of a green multi-sensing node powered by energy harvester and an edge node to collect data from the sensor node wirelessly. The sensor node consists of P sensors to measure the variation of P different parameters in the system. The adaptive sensing algorithm is implemented at the edge node to optimize the processing complexity of the sensor node. The algorithm is implemented over a batch of data collected at a measurement cycle to find the optimal sensor set to be activated at the next measurement cycle.

Let, the sensing parameter set be represented as $\mathcal{P} = \{P_p; 1 \leq p \leq P\}$ and $z_p^x(i)$ is the i^{th} sample measured by the p^{th} sensor in the x^{th} measurement cycle. The sampling interval of each sensor can be decided from their Nyquist sampling rate [3], which can be denoted as tn_p. Let, $N^x = \{N_p^x; 1 \leq p \leq P\}$ be the vector contains the total number of samples measured in the x^{th} cycle τ^x at the Nyquist sampling rate of the respective parameters. Thus, $\mathbf{Z}_p^x \in \mathbb{R}^{N_p^x \times 1}$ comprises the time series samples of the p^{th} parameter in the x^{th} measurement cycle. However, a strong temporal correlation is observed when the data are collected at the Nyquist sampling rate, such that the signals can be reconstructed with $M_p^x \ll N_p^x$ samples. Since the multiple sensors in the node are monitoring the same process dynamics,

a strong cross-correlation is also observed among the different sensing signals. Therefore, by activating a fewer sensors, the other sensing parameters can be predicted.

The total number of sensor sets created using P sensors is $N' = (2^P - 2)$, except the set containing all the parameters and the null set. Let, the i^{th} active set in the x^{th} measurement cycle be represented as $\mathcal{A}_i^x$. Thus, $\mathcal{B}_i^x = \mathcal{P} - \mathcal{A}_i^x$ is the i^{th} sleep set in the x^{th} measurement cycle. $\mathcal{Q}^x = \{(\mathcal{A}_i^x, \mathcal{B}_i^x); 1 \leq i \leq N'\}$ contains the active-sleep pair of sets.

Let, $c(a, b)$ be the cross-correlation coefficient between parameter a and parameter b and c_{th} be the cross-correlation threshold. If $|c(a, b)| \geq c_{th}$, the parameters a and b are considered to have a good correlation. The i^{th} sensor set is correlated if each parameter of $\mathcal{B}_i^x$ is correlated with at least one parameter of $\mathcal{A}_i^x$. The average cross-correlation of $\mathcal{A}_i^x$ and $\mathcal{B}_i^x$, denoted as cross-correlation factor C_i^x is:

$$C_i^x = \frac{1}{A_i^x B_i^x} \sum_{k=1}^{B_i^x} \Big(\sum_{m=1}^{A_i^x} |c^x(m, k)| \Big); \forall k \in \mathcal{B}_i^x \text{ and } m \in A_i^x. \tag{1}$$

$C_i^x = 0$ for $C_i^x < c_{th}$.

If $ct_p^x(s)$ is the temporal correlation between $Z_p^x(i)$ and $Z_p^x(i - s)$, where the temporal samples are collected at the Nyquist sampling rate tn_p. If $ct_p^x(s) \geq ct_{p,th}$, s consecutive samples can be predicted from the previously collected sample with prediction error below et_{th}. If $\zeta^x = s$ is the lag value in the x^{th} cycle, the optimum sampling interval in the proposed framework is set as $t_p^x = \zeta^x tn$. Thus, $M_p^x = \frac{\mathcal{T}^x}{t_p^x}$ is the optimum number of samples required to collect for the p^{th} parameter in the x^{th} cycle.

B. GPR-based Sensing Signal Recovery

As discussed in Section II-A, the temporal signal of each parameter can be reconstructed using $M_p^x \ll N_p^x$ samples. The intermediate samples need to be predicted from the immediate past samples collected by the sensor to maintain the accuracy of the signal. Therefore, GPR is used to predict the temporal samples, as it is more efficient to predict non-stationary noisy signals compared to the other regressors available in literature. P GPR models are used to predict the temporal samples of P parameters, which are named as GPR_1 models. The time series samples of the inactive sensors $\in \mathcal{B}_i^x$ are also predicted using GPR from the cross-correlated time series samples of the active sensors $\in \mathcal{A}_i^x$, which are named as GPR_2 models.

Let, the GPR_1 model consists of P submodels for P different parameters. If $Z_p^x(j)$ is the j^{th} temporal sample, the $(j+1)^{th}$ sample is $Z_p^x(j+1) = f_p(Z_p^x(j))$. GPR estimates the underlying function f_p that fits best with the data.

Similarly GPR_2 model is consisting of N' submodels for N' active sets. Let, the optimal active set selected in the x^{th} cycle is $\mathcal{A}_i^x$ and the i^{th} submodel consists of B_i^x number of regressors for predicting the cross-correlated parameters of the sleep set $\mathcal{B}_i^x$. Let, the measurement vectors of $\mathcal{A}_i^x$ and $\mathcal{B}_i^x$ in the x^{th} cycle for j^{th} sampling instant are respectively $\mathbf{Z}_{\mathcal{A}_i}^x(j) = \{Z_{\mathcal{A}_i}^x(j,1), Z_{\mathcal{A}_i}^x(j,2), \cdots, Z_{\mathcal{A}_i}^x(j, A_i^x)\} \in \mathbb{R}^{1 \times A_i^x}$ and $\mathbf{Z}_{\mathcal{B}_i}^x(j) = \{Z_{\mathcal{B}_i}^x(j,1), Z_{\mathcal{B}_i}^x(j,2), \cdots, Z_{\mathcal{B}_i}^x(j, B_i^x)\} \in \mathbb{R}^{1 \times B_i^x}$. If $\mathbf{Z}_{\mathcal{A}_i}^x(*)$ is the test vector, the underlying function $f_{i,*}^{x,k}$ to estimate the k^{th} parameter of $\mathcal{B}_i^x$ is:

$$Z_{\mathcal{B}_i}^x(*, k) = f_k(\mathbf{Z}_{\mathcal{A}_i}^x(*)) = f_{i,*}^{x,k} \forall k \in \mathcal{B}_i{}^x(*), \tag{2}$$

where $f_i^k \sim N(0, K_{n \times n})$. $K_{n \times n}$ is the covariance matrix. The elements of this matrix in this framework are derived from a squared exponential covariance function [7], [1].

C. UCB-based Active Sensor Set Selection

The Upper Confidence Bound (UCB) algorithm is used to solve the multi-arm bandit problems to find the optimal arm. This algorithm is modified in the proposed work to develop an optimization function to find the optimal active sensor set by exploiting the cross-correlation factor, energy consumed by the active sensors, and energy availability at the node.

As discussed in Section II-A, C_i^x defines the cross-correlation strength between $\mathcal{A}_i^x$ and $\mathcal{B}_i^x$. Let, $\underline{\mathbf{Es}} \in \mathbb{R}^{P \times P}$ be a diagonal matrix, with diagonal elements containing the sensing energy consumption to collect one sample. Thus, $Es_i^x = \mathbf{A}_i^x \underline{\mathbf{Es}} \mathbf{M}^x$ is the total energy consumption of the sensors $\in i^{th}$ active set. where $\mathbf{A}_i^x$ is the binary sensing vector such that $\mathbf{A}_i^x(k) = 1 \forall k \in \mathcal{A}_i^x$, otherwise, $\mathbf{A}_i^x(k) = 0$. To incorporate energy consciousness, the harvested energy available in the battery is considered as an additional performance criterion when determining the optimal sensor set. If E_{batt} is the capacity of the battery and E_0^x is the available energy in x^{th} measurement cycle. $\lambda^x \triangleq \frac{E_0^x}{E_{batt}}$ is the normalized energy of node. Therefore, the objective is to minimize Es_i^x and maximize C_i^x and λ^x.

Therefore, the reward obtained for selecting the i^{th} active set is defined as $R_i^x = \frac{\lambda^x C_i^x}{\nu^x Es_i^x}$, where $\nu^x = \max_{i \in \mathcal{Q}^x} \frac{C_i^x \tau_{\mathbf{A}_i}^x}{Es_i^x}$. Thus, the reward is bounded to $[0, 1]$. R_i^x is computed $\forall i \in \mathcal{Q}^x$ for every measurement cycle to choose an optimal set [5].

According to [8], the rewards for i^{th} sensor set up to x^{th} cycle $R_i^1, R_i^2, \cdots, R_i^x$ is a sequence of independent and identically distributed Gaussian random variables with true mean μ and variance 1. Hence, the empirical mean $\widehat{\mu}_i^x$ of the distribution $P_{R_i}^x$ is estimated as:

$$\widehat{\mu}_i^x = \frac{1}{x} \sum_{t=1}^{x} R_i^t = \frac{1}{x} \sum_{t=1}^{x} \frac{\lambda^x C_i^x}{\nu^x Es_i^x}. \tag{3}$$

The UCB-based optimization function to select an optimal sensor set for the $(x+1)^{th}$ measurement cycle is:

$$\begin{aligned} \mathcal{A}_i^{x+1} = & \underset{i \in \mathcal{Q}^x}{\text{maximize}} \frac{1}{x} \sum_{t=1}^{x} \frac{\lambda^x C_i^x}{\nu^x Es_i^x} + \sqrt{\frac{2 \ln \frac{1}{\delta}}{T_i^x}} \\ & \text{s. t. } C_i^x > c_{th} \text{ and } Es_i^x < E_0^x. \end{aligned} \tag{4}$$

In (4), δ is defined as a confidence bound. T_i^x denotes the number of times i^{th} sensor set is selected to collect data. The optimization function is adaptive to the harvested energy availability using the parameter λ. When the battery is charged by the energy harvester, E_0 increases and C_i^x gets relatively more weight, whereas, the Es_i^x gets more weight for decreasing E_0.

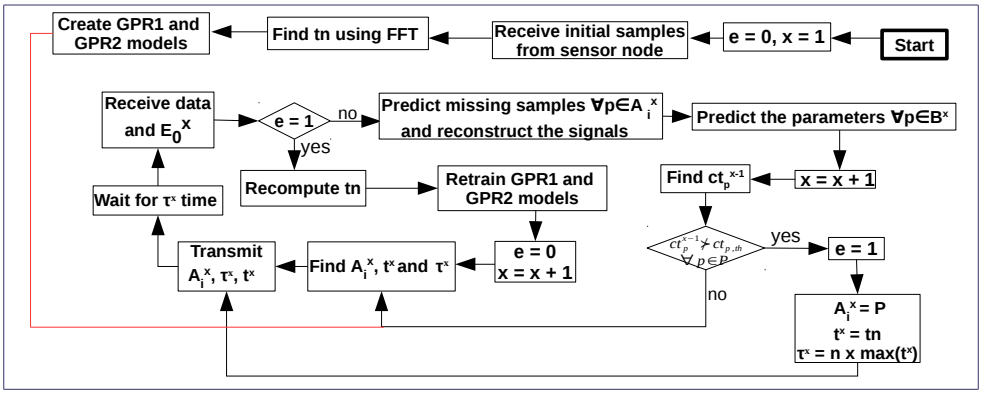


Fig. 2: Flow chart of the smart sensing algorithm programmed in the edge node.

D. Proposed Learning-Aided Adaptive Sensing Framework

The system model, depicted in Fig. 1, consists of two main modules. The first module is the sensor node powered by an energy harvester to harvest energy from the environment to replenish the battery. The node collects data from the sensors over a batch and transmits them to the data collector node/edge node, which is the second module in the system. To reduce the processing complexity of the end node, the smart sensing algorithm is programmed in the edge node; the flow diagram of the smart sensing framework is presented in Fig. 2.

At the beginning, the sensor node collects n samples for each parameter and transmits them to the edge node to train the GPR models, as discussed in Section II-B. The node also transmits the energy available in the battery which is recharged by the energy harvester. The edge node finds the Nyquist sampling interval tn for each parameter which provides a faithful reconstruction of the signal [3]. Then it finds the optimum sampling interval of each parameter by exploiting the temporal correlation and the temporal prediction error; as discussed in Section II-A. Then the UCB-based optimization function, developed in Section II-C, solves the trade-off among the cross-correlation factor, energy consumed by the active sensors, and energy availability at the node, and selects an optimal active sensor set using (4). To reduce the complexity of the algorithm, the length of measurement τ is considered as fixed. The information contains the active set $\mathcal{A}_i^x$, sampling interval vector t^x, and the length of the x^{th} measurement cycle τ^x are transmitted to the sensor node. After receiving the information, the field deployed sensor node activates the sensors belong to the active set, while the other sensors remain off. At the end of the cycle, the node again transmits the collected samples and the value of E_0^x to the edge node. The edge node first predicts the missing temporal samples of $\mathcal{A}_i^x$ using GPR_1 model and then, it predicts the time series samples of $\mathcal{B}_i^x$ from the cross-correlated parameters of $\mathcal{A}_i^x$ using GPR_2 model. After reconstructing the sensing signals, the algorithm draws the samples from reconstructed signals for all the parameters using Nyquist sampling rate. Then it estimates the optimal sampling interval and the optimal active set for the next cycle, which again transmitted back to the sensor node. The algorithm retrains the prediction models using original data if the temporal correlation of the reconstructed signals for each parameter falls below a threshold for any parameter.

III. Implementation of the Proposed Algorithm

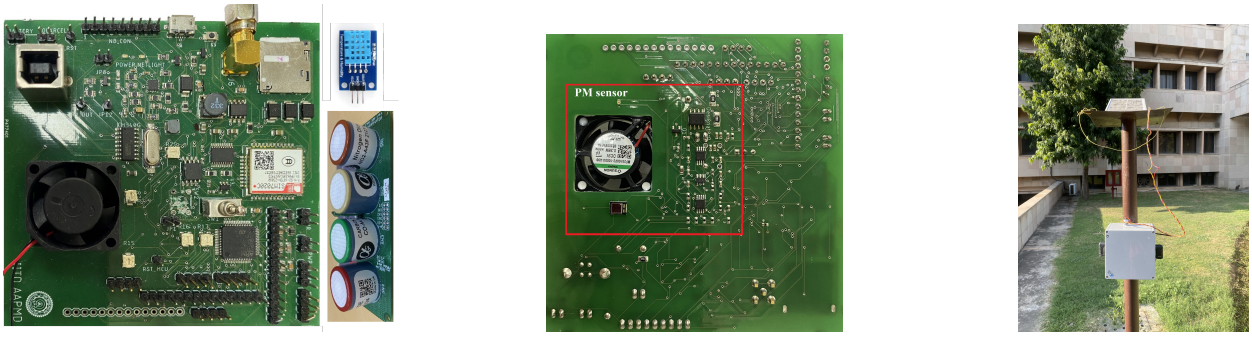
This section describes the experimental setup developed for implementation of the algorithm, discussed in Section II.

A. Experimental Setup

The developed air pollution monitoring sensor node is shown in Fig. 3(a) and (b). The designed system is capable of measuring $\text{PM}_{2.5}$, PM_{10}, CO, NO_2, SO_2, and Ozone, along with temperature and humidity. The on-board low power PM sensor is a LED light scattering-based optical sensor which consumes 64 mJ energy to collect one sample [9]. Alphasense electrochemical gas sensors are used to monitor gas pollutants based on the significant results reported in [10]. The energy consumption of the sensor set $\{$DHT11, NO_2, Ozone, CO, SO_2, PM$\}$ to collect one sample are respectively $\{0.012, 0.1, 0.02, 0.05, 0.18, 0.646\}$ J, where DHT11 measure the temperature and humidity, and PM sensor measure the $\text{PM}_{2.5}$ and PM_{10}.

The system is green and self sustainable using a 2W solar panel. A Li-ion battery of 3.7 V, 10000 mAh is used for night hours, rainy, and foggy days. The on-board solar energy harvester is used for recharging the battery. The node uses a very low power ARM cortex STM32L476 microcontroller which wakes up periodically only at the time of sampling and data transmission and remains in deep sleep mode otherwise. The module hosts SIMCOM 7020C narrow band-IoT (NB-IoT) module and ESP8266 WiFi module to add communication complaisance and utilizes the local high computing system of IIT Delhi called 'Baadal' to accumulate all the sensed parameters and run the algorithm.

Fig. 3(c) shows the prototype sensor module deployed at IIT Delhi campus. The water resistant IP67 rated system is designed to be mounted easily to poles or lamp posts. The fan must draw outside air to take a new air pollutant measurement. As shown in the figure, there are two vents installed in the prototype case for drawing fresh air for every new measurement and clearing the inside air after each cycle.



(a) (b) (c)

Fig. 3: The designed air pollution monitoring board, (a) the front side of the board with off-the-shelf gas sensors and DHT sensor, (b) the back side of the board with on-board PM sensor circuitry, and (c) Deployed IP67 rated prototype with solar panel in IIT Delhi campus.

B. Energy Harvesting Strategy of the System

To enable the solar energy harvesting capability in the system, a 9 cm×12 cm solar panel is connected to the node, mounted on the top of the pole as shown in Fig. 3(c). The power harvested from the solar panel has been measured hourly. During the day, the solar cell provides the required energy and the battery stores access energy. The hourly solar energy has been evaluated from the voltage-current profile of the panel to determine the captured energy from the solar cell. The harvested energy from the solar energy harvester has been computed using a 10% efficiency factor.

IV. Results and Discussions

The energy aware and learning-aided adaptive sensing framework, discussed in Section II-D, has been implemented in an air pollution monitoring system, as presented in Section III. The solar energy harvesting profile is shown in Fig. 5. The amount of harvested power increases gradually from morning 7AM, peaks at 12 noon, and then decreases.

With six sensors in the node, the parameter set is $\mathcal{P} = \{$temperature, humidity, PM_{10}, $PM_{2.5}$, NO_2, Ozone, CO, $SO_2\}$. Applying Nyquist adaptive sampling algorithm in the initially collected data, the Nyquist sampling interval vector is $tn = \{35, 9, 15, 15, 15, 16, 14, 15\}$ Sec. The prediction error is computed in terms of normalized Mean Squared Error (nMSE). If $\mathbf{Z}_p$ and $\widehat{\mathbf{Z}}_p$ contains the actual and reconstructed temporal samples of the p^{th} parameter, respectively, the nMSE is expressed as, $nMSE_p = \frac{||\mathbf{Z}_p - \widehat{\mathbf{Z}}_p||}{||\mathbf{Z}_p||}$. Fig. 4 shows the variation of the temporal correlation and the temporal prediction error of $PM_{2.5}$ (using GPR_2 model) with ζ. It can be observed that the prediction error increases and the temporal correlation decrease with the increasing ζ, which provides a correlation threshold ct_{th} for the corresponding temporal prediction error threshold et_{th}. Considering $et_{th} = 0.00001$, the temporal correlation threshold for all the parameters are $ct_{th} = \{0.98, 0.98, 0.97, 0.97, 0.96, 0.98, 0.96, 0.97\}$ to find the optimum values of ζ at every measurement cycle [5]. Thus, the optimum interval vector for the x^{th} cycle is $t^x = \zeta^x tn$. According to [9], the measurement cycle is set as $\tau = 1800$ Sec. The cross-correlation threshold in (4) is set as $c_{th} = 0.7$ to get the overall reconstruction error e_{th} below 10^{-3}.

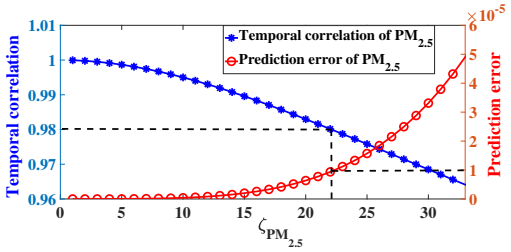


Fig. 4: Prediction error and temporal correlation versus ζ.

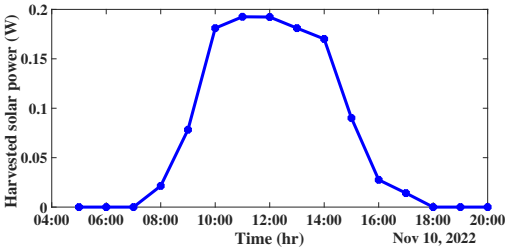
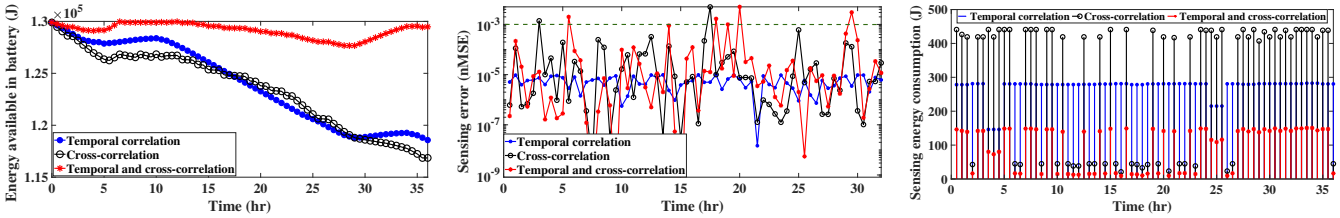


Fig. 5: Solar power harvested in a day.

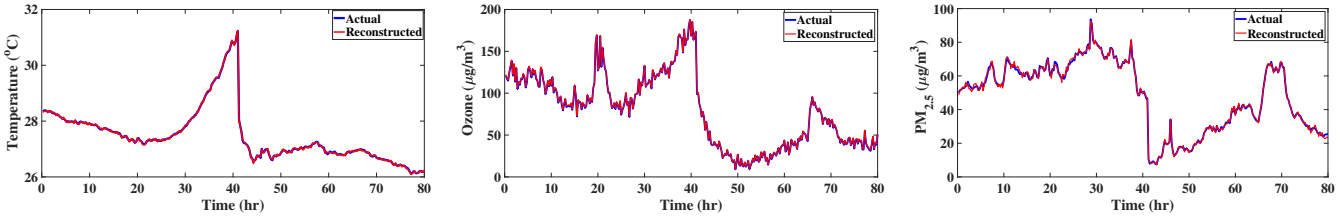
The algorithm has been implemented for three cases: (1) only temporal correlation is exploited for each parameter to find the optimum sampling interval and the node activates all the sensors to collect data, (2) only cross-correlation among the parameters is exploited to find the active sensor set and the sensors collect data at Nyquist sampling interval [1], and (3) both cross-correlation and temporal correlation are exploited to find the optimum system parameters. The node is deployed with a fully charged battery of 3.7 V and 10000 mAh, which stores 130000 J energy. The APMD is designed such that it can measure the voltage and the current drawn from the battery. At the end of every cycle, the node transmits the voltage-current profile of the battery to the edge node, which is a local system to compute the remaining energy of the battery.

Fig. 6 shows the performance of the algorithm for three cases. A performance comparison is listed in Table I. Unlike the simulation studies presented in [5], the implementation of these methods incorporates all the vulnerabilities of the system and the real-time delays, such as processing time, transmission time, etc. Fig. 6 and Fig. 7 validate the efficiency of the



(a) (b) (c)

Fig. 6: Performance comparison of the algorithm implemented in the air pollution monitoring system.



(a) Temperature (b) Ozone (c) $PM_{2.5}$

Fig. 7: Actual and reconstructed signals at the edge node.

TABLE I: Performance comparison among different cases

	Energy consumption of node	Sensing error
case 1	271 J	5.5×10^{-6}
case 2 [1]	295 J	1.4×10^{-4}
case 3	95 J	2.4×10^{-4}

algorithm in real systems. From Fig. 6(a), case 1 performs better than case 2, whereas, case 3 outperforms the other methods. The node is fully sustainable for case 3, as the energy utilized by the node is replenished by the solar in the daytime. Whereas, the energy available in the battery is decreasing for case 1 and case 2. Since the Nyquist sampling interval is very low, the sensors need to collect many samples in case 2. Therefore, the sensing energy consumption is very high, which can be observed in Fig. 6(c). Fig. 6(b) shows that the average sensing errors or reconstruction errors of all the parameters for three cases are below the threshold. The prediction error in case 1 is below et_{th}, which was set to find the optimum values of ζ. The actual and reconstructed signals of temperature, Ozone, and $PM_{2.5}$ are shown in Fig. 7. The proposed method saves 68% energy of the battery compared to the nearest competitive method case 2, which was proposed in [1], while the sensing error is within the limit.

V. Concluding Remarks

An energy aware and learning-aided smart sensing algorithm has been presented in this article. The cross and temporal correlation among the parameters have been exploited to remove the redundancy in the data and optimize the energy sustainability of the node. A green air pollution monitoring system has been developed to implement the proposed learning-based algorithm, where the sensor node is powered by solar energy harvester. The algorithm has been programmed in a local system that considered as the edge node in this experiment. The experimental results validate that the proposed temporal and cross-correlation based algorithm can make the system fully sustainable, whereas, the other cases cannot.

Acknowledgement

This work was supported in part by the Science and Engineering Board, Department of Science and Technology (DST), Government of India, under Grant CRG/2019/002293; in part by the Indian National Academy of Engineering (INAE) through the Abdul Kalam Technology Innovation National Fellowship; and in part by the Prime Minister's Doctoral Research Fellowships.

References

[1] S. Ghosh, S. De, S. Chatterjee, and M. Portmann, "Learning-based adaptive sensor selection framework for multi-sensing wsn," *IEEE Sensors J.*, vol. 21, no. 12, pp. 13 551–13 563, 2021.

[2] C. Alippi, G. Anastasi, M. Di Francesco, and M. Roveri, "An adaptive sampling algorithm for effective energy management in wireless sensor networks with energy-hungry sensors," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 2, pp. 335–344, 2009.

[3] H. Harb and A. Makhoul, "Energy-efficient sensor data collection approach for industrial process monitoring," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 661–672, 2017.

[4] V. Gupta and S. De, "Energy-efficient temporal sensing: An age-of-sample-based approach," *IEEE Internet Things J.*, vol. 9, no. 3, pp. 1806–1817, 2022.

[5] S. Ghosh, S. De, S. Chatterjee, and M. Portmann, "Edge intelligence framework for data-driven dynamic priority sensing and transmission," *IEEE Transactions on Green Communications and Networking*, pp. 1–1, 2021.

[6] V. Gupta and S. De, "SBL-based adaptive sensing framework for WSN-assisted IoT applications," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4598–4612, 2018.

[7] C. Rasmussen and C. Williams, "Gaussian processes for machine learning.,(mit press: Cambridge, ma)," 2006.

[8] T. Lattimore and C. Szepesvári, *Bandit algorithms*. Cambridge University Press, 2020.

[9] P. Das, S. Ghosh, S. Chatterjee, and S. De, "A low cost outdoor air pollution monitoring device with power controlled built-in pm sensor," *IEEE Sensors Journal*, vol. 22, no. 13, pp. 13 682–13 695, 2022.

[10] L. Sun, K. C. Wong, P. Wei, S. Ye, H. Huang, F. Yang, D. Westerdahl, P. K. Louie, C. W. Luk, and Z. Ning, "Development and application of a next generation air sensor network for the hong kong marathon 2015 air quality monitoring," *Sensors*, vol. 16, no. 2, p. 211, 2016.