

Improving Accuracy of the Shewhart-based Data-Reduction in IoT Nodes using Piggybacking

Anish Shastri*, Vivek Jain*, Sachin Chaudhari*, Shailesh Singh Chouhan[†], Stefan Werner**

*International Institute of Information Technology (IIIT) - Hyderabad, India

[†]Luleå University of Technology, Luleå, Sweden

** Norwegian University of Science and Technology (NTNU), Trondheim, Norway

Email: anish.shastri@research.iiit.ac.in, jain.vivek@students.iiit.ac.in,

sachin.chaudhari@iiit.ac.in, shailesh.chouhan@ltu.se, stefan.werner@ntnu.no

Abstract—This paper proposes the use of Shewhart test to reduce the number of data-transmissions in IoT networks. It is shown to outperform the widely-used least mean square (LMS) based data reduction method in terms of the number of data-transmissions, implementation complexity and mean square error (MSE) in prediction of time-series data at the sink node based on the partial transmissions of the measured time-series data from the sensor node. The paper also proposes the use of piggybacking and interpolation to further reduce the MSE of the estimated time-series data at the sink node without increasing the number of packet transmissions. The time-series data used for the comparison of data reduction algorithms is a set of measured temperature values in indoor and outdoor scenarios for four days using custom-designed wireless sensor nodes. To express the effectiveness of the piggybacked transmissions on battery lifetime, the total current consumption of the sensor node is measured for different number of piggybacks and corresponding battery lifetime is estimated. It is shown that the proposed piggyback approach significantly reduces the MSE at the cost of slight decrease in battery-lifetime.

Index Terms—battery-lifetime, data-transmission reduction, IEEE 802.15.4, MSE, piggybacking, Shewhart test.

I. INTRODUCTION

Most of the internet of things (IoT) networks deploy a large number of sensor nodes, which are often placed in hazardous or inaccessible locations of the observation environment. Therefore, providing wired power to these nodes or changing their batteries frequently may be inefficient or prohibitive. As such, increasing the lifetime of the battery and consequently that of the sensor node has been an important design constraint for the IoT networks. Out of the three power consuming activities in a sensor node, the power consumption cost for communication is much higher compared to that for sensing and computation [1]. Therefore, one of the effective ways of increasing the battery lifetime of a sensor node is to reduce the number of data-transmissions, also termed as *data reduction* for convenience.

Several data reduction algorithms have been proposed in wireless sensor network (WSN) literature

[1]–[5]. Different techniques like *naive*, *least mean squares (LMS)*, *auto-regressive integrated moving average (ARIMA)* based time-series prediction algorithms were compared in [2], [3]. Among these algorithms, the LMS-based data reduction is widely used due to lower computational overhead and greater flexibility [1]–[5]. In [4], a convex combination of two decoupled variable-length LMS filters was used for better prediction of the data. In our earlier work [5], LMS algorithm was tested using a low-cost customized wireless sensor node with low power consumption.

The contributions of this paper are:

- The use of the Shewhart test [6] is proposed for data-reduction in IoT networks and its performance is compared with the widely-used LMS algorithm in terms of data reduction, MSE and complexity based on the temperature measurements in [5].
- The use of piggybacking and interpolation is proposed along with the Shewhart algorithm. Experimental results are presented for different number of piggybacks and different interpolation schemes.
- Current consumption measurement is carried out for different number of piggybacks on the custom hardware implementation used in [5].
- Battery-lifetime prediction is performed for the Shewhart algorithm with different number of piggyback values.

Motivation for the Shewhart test is that it is simpler and has much lower complexity than the LMS algorithm. The Shewhart test has been applied previously for data reduction in smart metering application [7]. Motivation for piggybacking is that it will improve the MSE performance without increasing the number of packet transmissions. Given the significant overhead of the header in each packet, increasing the payload with a few additional data points does not significantly increase the energy consumption required for the packet transmission as will be shown in this paper.

The paper is organized as follows. Section II briefly presents the LMS and Shewhart based data reduction algorithms along with the proposed piggybacking approach. Section III briefs about the implementation of the

algorithm on the customized wireless sensor node and experimentation setup to collect the temperature data. Section IV presents the results and Section V concludes the paper.

II. DATA REDUCTION ALGORITHMS

This section begins with description of the standard LMS-based data reduction algorithm followed by the Shewhart based algorithm. In the last part, piggyback-based Shewhart test is explained.

A. LMS based data reduction

Consider a data stream $\{x[n]\}$, where the data at the n th instant is given as $x[n]$. The LMS based estimate $\hat{x}[n]$ at the n th instant is given as [8]

$$\hat{x}[n] = \mathbf{w}_{n-1}^T \mathbf{y}_{n-1} \quad (1)$$

where $\mathbf{w}_{n-1} = [w_{n-1}[1] \dots w_{n-1}[N]]^T$ is the weight coefficient vector of the N tap linear prediction filter associated with the update instant $n-1$ and $\mathbf{y}_n = [y[n] \dots y[n-N+1]]^T$ is a vector that consists of values taken from the set of N previous values $\{x[k]\}_{k=n-N}^{n-1}$, or their corresponding estimates $\{\hat{x}[k]\}_{k=n-N}^{n-1}$. The *a priori* prediction error $\epsilon[n]$ is given as

$$\epsilon[n] = x[n] - \hat{x}[n] \quad (2)$$

If $|\epsilon[n]|$ falls below an error threshold ϵ_t , then $y[n] = \hat{x}[n]$ and no transmission happens from the node to the sink. If $|\epsilon[n]|$ is larger than the error threshold, then $y[n] = x[n]$ and the current value $x[n]$ is transmitted from the node to the sink.

The filter weight coefficients are adapted as follows

$$\mathbf{w}_n = \mathbf{w}_{n-1} + \mu \epsilon[n] \mathbf{y}_n, \quad (3)$$

where μ represents the step size of the filter and must satisfy the constraint $0 \leq \mu \leq \frac{1}{E_x}$. Here, E_x is the mean power of the signal and given by,

$$E_x = \frac{1}{K} \sum_{n=1}^K |x[n]|^2, \quad (4)$$

where K is the number of samples taken for calculating E_x at the start of the experiment. The parameter μ controls the speed of convergence of the algorithm [8].

B. Shewhart based data reduction

In this work, the Shewart test that has been used for change detection in [6], is deployed for the purpose of data reduction. In the Shewhart test, the data $x[n]$ is transmitted by the sensor node only if the absolute of the error, i.e., $|\epsilon[n]|$ exceeds the error threshold ϵ_t , which is a user-defined application-specific parameter. At the sink node, the prediction is done according to the received data as follows

$$\hat{x}[n] = \begin{cases} x[n], & \text{if data received.} \\ \hat{x}[n-1], & \text{otherwise.} \end{cases} \quad (5)$$

This way, the magnitude of the prediction error, i.e., $|x[n] - \hat{x}[n]|$ at the sink node is ensured to be upper bounded by the error threshold.

Note that the Shewhart test not only has significantly lower implementation complexity than the LMS scheme but also ensures the same performance in terms of bounded error. The complexity of Shewhart amounts to a single comparison as only the last value is sent as opposed to the linear time complexity of weight update in LMS. This allows the use of the algorithm for a wide-range of applications, where the sensor node is constrained in terms of power and computation. Also, unlike LMS, the Shewhart algorithm does not have a tunable parameter like step size which will affect its convergence.

C. Proposed piggybacking-based Shewhart test

Let us denote n_t as the sample-index when t th transmission happens for the Shewhart test from the node to the sink. Now consider a window of sensor values $(x[n_{t-1}], x[n_{t-1}+1], \dots, x[n_t-1], x[n_t])$ of length $W_t = n_t - n_{t-1} + 1$. This window consists of the two end-points $x[n_{t-1}]$ and $x[n_t]$, which are transmitted from the node to the sink. On the other hand, the intermediate sensor values in this window, which lie in between n_{t-1} and n_t sample-indices are not transmitted. In the proposed method, p (approximately) equispaced values $x[n_{t-1} + \lfloor \frac{m(W_t-1)}{p+1} \rfloor]$ for $m = 1, \dots, p$ are piggybacked with the transmission $x[n_t]$. Here, $\lfloor \cdot \rfloor$ denotes the round-off operation. Note that the window length W_t , corresponding to the transmission instances n_t and n_{t-1} , changes with the value of t as it depends on the time-series data and the threshold. If $2 < W_t < p$, all intermediate values are sent as piggybacks.

The extra sensor values will result in a reduction of the MSE for the time-series data at the sink as p of the estimates are getting replaced by the actual values. To further reduce the MSE, interpolation is carried out on a window-by-window basis at the sink using the end-points and the piggybacked values. If we let $\mathbf{x}_t = [x[n_{t-1}], x[n_{t-1} + \lfloor \frac{(W_t-1)}{p+1} \rfloor], \dots, x[n_{t-1} + \lfloor \frac{p(W_t-1)}{p+1} \rfloor], x[n_t]]^T$ denote the vector consisting of the end-points of the window and the piggybacked values, then the predicted values at the sink node for the sample-indices $n_{t-1} < n < n_t$ are given as

$$\hat{x}[n] = \begin{cases} x[n], & \text{if data received.} \\ f(\mathbf{x}_t), & \text{otherwise.} \end{cases} \quad (6)$$

where $f(\mathbf{x}_t)$ is an interpolation function. In this paper, only simple and exact interpolation schemes such as *linear*, *cubic* and *previous* (also called as *naive*) are considered.

Note that piggybacking only increases the payload of each packet and not the number of packets transmit-

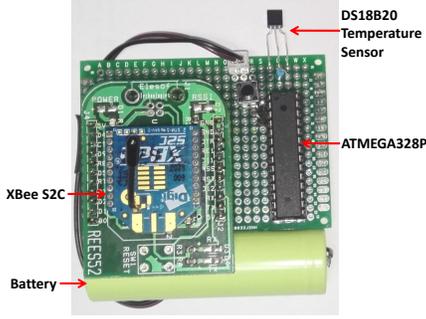


Fig. 1: Hardware design of the sensor node [5].

ted. This is because more data points are now getting conveyed in the same number of packets. The increased payload leads to increase in processing and transmission power of a packet. However, the presence of significant packet-overhead means that the increase in the packet length is not linearly increasing with the increase in the payload. For example, in our XBee S2C based implementation [5], the header consists of 9 bytes while each temperature value is represented by 2 bytes. So using $p = 1$ and $p = 4$ piggybacks would result in an increase of 100% and 400%, respectively, in the length of payload while the increase in the packet length would be only 18.18% and 72.72%, respectively.

III. MEASUREMENT SETUP

In this section, the design of the sensor node is first discussed followed by the experimental setup used for collecting the measurements. Note that the measurement setup, including the sensor node and the measurements, is the same as in our previous work [5] and is briefly presented here for the sake of completeness.

A. Sensor node

The custom designed sensor node is shown in Fig. 1. The main components of the sensor node include ATmega328P as the microcontroller [9], DS18B20 as the temperature sensor [10] and XBee S2C Zigbee module [11] as the radio module. The XBee radio modules are configured to operate using IEEE 802.15.4 protocol. The radio module at the sensor node is configured as an end-device while the one at the sink node is configured as the coordinator. A 2800 mAh Lithium ion battery is used to power the sensor node. The sink node is supplied with uninterrupted power, as commonly assumed in the WSN literature.

B. Setup

The measured temperature data-points were collected from 13th through 16th June 2018 for two scenarios: indoor and outdoor of our lab in IIIT-Hyderabad. The deployment of the sensor nodes and the coverage area are shown in Fig. 2. The first sensor node is placed in the

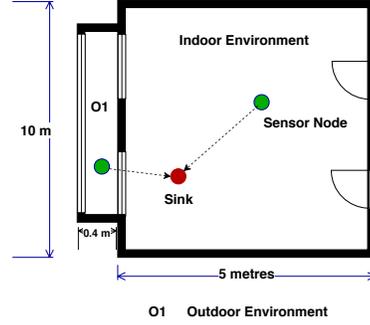


Fig. 2: Floor plan of sensor node deployment at IIIT-Hyderabad [5].

indoor environment, which is affected by the occupancy of the room. The other node is placed at the outer ledge of the window just outside the lab. The sink is placed inside the lab and the data is collected from the nodes every 30 s in a typical star topology. The API-2 XBee data packet header has the following contents: start delimiter (1 byte), length till checksum (2 bytes), frame type (1 byte), frame ID (1 byte), destination address (2 bytes), options (1 byte), checksum (1 byte) and the data payload is 2 bytes. The sink adds timestamps along the received instances of the data and then pushes the data to the cloud.

IV. RESULTS

This section presents results in four parts. The first part compares the LMS and the Shewhart algorithms in terms of MSE and reduction in the number of data-transmissions. In the second part, improvement in the MSE obtained by the proposed piggybacking approach is presented. Finally the current consumption analysis and prediction of the lifetime extension of the battery are presented. The data reduction algorithms in the first two parts have been run on the measurements in [5] using MATLAB, while the later parts have been carried out on the hardware in [5].

A. Comparison of Shewhart and LMS without piggybacking

Fig. 3 shows the temperature values collected in indoor and outdoor scenarios as well as the transmission instances with LMS and Shewhart. The error threshold ϵ_t was set to 0.5°C . The parameters for the LMS algorithm are as follows: $K = 50$, $\mu = \frac{1}{50 E_x}$. Table I summarizes the reduction in data-transmissions for these datasets using the two algorithms. It can be seen that both the algorithms significantly reduce the data-transmissions from the node to the sink in indoor as well as outdoor scenarios. However, the Shewhart-based method outperforms the LMS algorithm. Also, the MSE for the Shewhart test is lower than that in the LMS case for both the datasets. This result is important as the

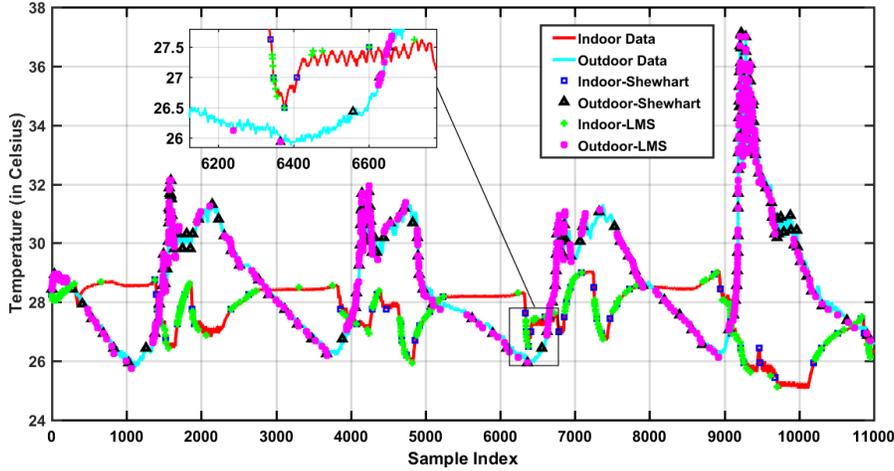


Fig. 3: Measured temperature data in the indoor and outdoor scenarios along with transmission instances while using LMS and Shewhart-based data reduction schemes.

Shewhart test has less complexity compared with widely used LMS algorithm as discussed earlier. As such, only Shewhart test is considered in rest of the section.

B. Shewhart with piggybacking

Fig. 4 presents interpolated temperature data for indoor scenario over one of the windows (between two actual transmissions) using various interpolation schemes with four number of piggybacks for Shewhart method. Fig. 5 shows the variation of the MSE as a function of the number of piggybacks. With no interpolation, the reduction in MSE is negligible and as the number of piggybacks increases, the MSE is decreasing. It can be seen from Figs. 5 and 6 that the linear and cubic interpolation schemes perform best for the indoor temperature data among the considered interpolation schemes. Fig. 6 presents the relative decrease in the MSE of the interpolated data after piggybacking with respect to no piggybacking. It can be seen that significant reduction in MSE of around 40% can be achieved by using a single piggyback and a simple linear interpolation scheme. The relative decrease in the MSE increases with the number of piggybacks, in general, and gives around 90% decrease for $p = 10$. Similar results have been obtained for the outdoor scenario, but not shown here for brevity. It

TABLE I: Comparison of the Shewhart and LMS methods for data-transmission reduction for the two data-sets.

	No. of Transmissions	% savings	MSE
Indoor Scenario			
With LMS	408	96.29	0.0906
With Shewhart	58	99.47	0.0489
Outdoor Scenario			
With LMS	682	93.80	0.0645
With Shewhart	201	98.17	0.0442

has been observed during the measurements that current consumption increases with increasing piggybacks and hence, reduces the battery-lifetime. Thus, this approach introduces a trade-off between the MSE and the battery-lifetime.

C. Current consumption measurement

The current consumption of the sensor node is measured for different number of piggybacks. The current measurements were performed using the Agilent 34410A Digital Multimeter [12]. The time duration was approximately analysed using the Keysight BenchVue software [13], where the data from the digital multimeter is logged along with the timestamps.

The different current consumption states observed in one duty cycle of operation of the sensor node are: *Processing State-1* (I_{P1}), *Radio ON* (I_R), *Transmit* (I_{Tx}), *Processing State-2* (I_{P2}), *Sleep* (I_S). The operation of the sensor node in different current states is summarized in the Table II.

Fig. 7 shows the measured current profiles of the sensor for one actual transmission instance, i.e., transmit state for different number of piggybacks. The red curve is for $p = 1$ while the blue curve is for $p = 4$. For this figure, the blue and red curves are not synced to start at the same instance for easy visualization. It can

TABLE II: Summary of the operation of the sensor node in different operating states

State	Operation
I_{P1}	Sense temperature and computation
I_R	Turn ON XBee if data needs to be sent to the sink
I_{Tx}	Data transmission instant
I_{P2}	Increment for further iterations
I_S	Sensor node sleeps for fixed time duration

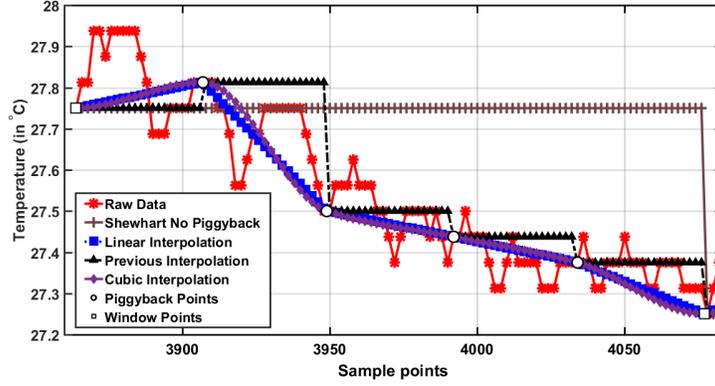


Fig. 4: Illustration of interpolated temperature data for indoor scenario over one window (between two actual transmissions) using various interpolation schemes with $p = 4$ number of piggybacks for the Shewhart method. Here, $W_{15} = 214$ $n_{14} = 3864$ and $n_{15} = 4077$.

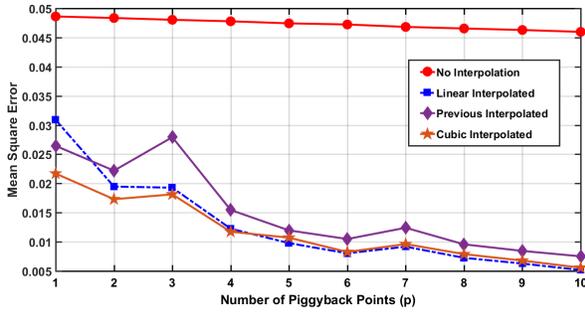


Fig. 5: MSE for the indoor temperature data at the sink as a function of number of piggybacks using various interpolation schemes for the proposed piggyback-based Shewhart test.

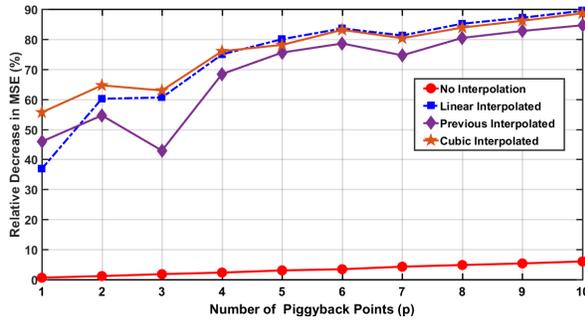


Fig. 6: Relative decrease in the MSE of the indoor temperature data at the sink using the proposed piggyback-based Shewhart test with respect to no-piggybacking based Shewhart test.

be seen that the increase in the number of piggybacks only slightly increases the duration of Radio ON and transmit states while affecting the current levels during those states. This result can also be corroborated from Table III, which shows the average current values and duration for different states for Shewhart test for different

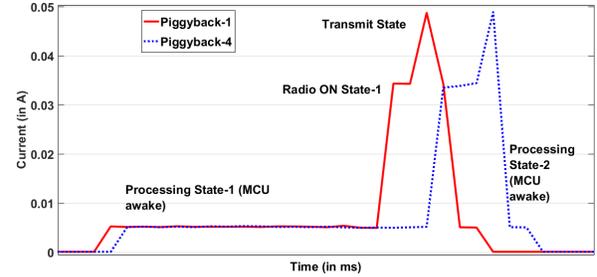


Fig. 7: Measured current profile for 1 transmission instant with a payload of 4 bytes ($p = 1$) and payload of 10 bytes ($p = 4$). The blue and red curves are not completely overlapped for easy visualization.

TABLE III: Average current consumption for different states for payload (2 bytes for Shewhart case, 4 bytes for 1 Piggyback and 10 bytes for 4 Piggybacks).

State	Avg. Current (in mA)	Duration ($p = 0$) (in ms)	Duration ($p = 1$) (in ms)	Duration ($p = 4$) (in ms)
I_{P1}	5	118	120	124
I_R	34.8	30	31	35
I_{Tx}	47	5	5	5
I_{P2}	5	10	10	10
I_S	0.0061	29837	29834	29826

values of piggybacking ($p = 0, 1, 4$). Note that $p = 0$ corresponds to no-piggybacking.

D. Prediction of battery lifetime

The duty-cycle controlled operation of the sensor node ensures the easy calculation of the average current consumption and thus, an estimation of the extension in the battery lifetime. Average current consumption of the sensor node (I_{Avg}) in 1 hour, is given as

$$I_{Avg} = \frac{(I_{P1} t_{P1} + I_R t_R + I_{Tx} t_{Tx} + I_{P2} t_{P2} + I_S t_S) L}{3600}, \quad (7)$$

TABLE IV: Comparison of average current consumption and expected lifetime of a node in the two scenarios of Shewhart algorithm with and without piggybacking.

Method	Avg. Current (mA)	Avg. Lifetime (days)
Indoor Scenario Dataset		
Without Shewhart	0.0735	1513
Shewhart with $p = 0$	0.02599	4488
Shewhart with $p = 1$	0.0263	4430
Shewhart with $p = 4$	0.027026	4316
Outdoor Scenario Dataset		
Without Shewhart	0.0735	1513
Shewhart with $p = 0$	0.02657	4390
Shewhart with $p = 1$	0.0269	4332
Shewhart with $p = 4$	0.02767	4215

where L = total number of transmissions in 1 hour. The I_{Avg} calculated in the above equation is used to calculate the expected lifetime of the sensor node. As given in [14], the lifetime of the battery of the sensor node in days γ is given as $\gamma = \frac{C}{24 I_{Avg}}$, where C is the capacity of the battery in mAh.

Table IV shows the average current consumption and the expected lifetime of the sensor node for Shewhart test for different number of piggybacks $p = 0, 1, 4$. These calculations have been done as presented in [14] and [15]. It can be seen from Table IV that Shewhart based data reduction schemes (with $p = 0, 1, 4$) significantly increase the lifetime as compared to the scheme "Without Shewhart", which is basically transmitting all the values from the node to the sink.

To see the effect of piggybacking on the battery-lifetime while using Shewhart, it can be seen from the Table IV that the average lifetime is decreasing with the increase in the number of piggybacks. In indoor scenario, Shewhart with $p = 1, 4$ decreases the lifetime by 1.29% and 3.8%, respectively in comparison with Shewhart with no-piggybacking ($p = 0$). In outdoor scenario, Shewhart with $p = 1, 4$ decreases the lifetime by 1.32% and 3.99%, respectively when compared to the lifetime for Shewhart with $p = 0$. However, it can also be clearly seen that the decrease in the battery-lifetime due to piggybacking is negligible compared to the Shewhart algorithm without piggybacking. This is a reasonable cost to pay for the significant reduction in the MSE of the predicted time-series temperature data at the sink node.

V. CONCLUSION

This paper demonstrates that a simple Shewhart test is much better than the widely-used LMS in terms of MSE, complexity and reduction in data-transmissions. The piggyback approach proposed along with interpolation in this paper is shown to improve the MSE of the time-series data (temperature in this paper) for the Shewhart test based data reduction. There is an inherent

trade-off between the reduction in MSE and reduction in the battery-lifetime while choosing the number of piggybacks, which can be decided based on the application. A significant decrease in MSE of around 38% and 75%, respectively for one piggyback and four piggyback values with simple linear interpolation scheme has been observed. On the other hand, the decrease in the lifetime of the sensor nodes in these cases are much smaller as 1.29% and 3.8%, respectively. Similar results have been obtained for the outdoor scenario as well.

VI. ACKNOWLEDGEMENT

This work was supported in part by the Research Council of Norway.

REFERENCES

- [1] S. Santini and K. Romer, "An adaptive strategy for quality-based data reduction in wireless sensor networks," in *Proc. 3rd Int. Conf. Networked Sensing Systems*, 2006, pp. 29–36.
- [2] F. Aderohunmu, G. Paci, D. Brunelli, J. Deng, and L. Benini, "Prolonging the lifetime of wireless sensor networks using light-weight forecasting algorithms," in *IEEE 8th Int. Conf. Intelligent Sensors, Sensor Networks and Inform. Process.*, 2013, pp. 461–466.
- [3] F. Aderohunmu, G. Paci, D. Brunelli, J. Deng, L. Benini, and M. Purvis, "An application-specific forecasting algorithm for extending WSN lifetime," in *IEEE Int. Conf. Distributed Computing in Sensor Systems*, 2013, pp. 374–381.
- [4] Y. Fathy, P. Barnaghi, and R. Tafazolli, "An adaptive method for data reduction in the Internet of Things," in *Proc. IEEE 4th World Forum on Internet of Things*, 2018.
- [5] A. Shastri, V. Jain, R. Singh, S. Chaudhari, and S. Chouhan, "On the implementation of LMS-based algorithm for increasing the lifetime of IoT networks," in *12th IEEE Int. Conf. Advanced Telecommun. Syst. and Networks*, 2018.
- [6] W. Shewhart, *Economic control of quality of manufactured product*. New York: American Society for Quality Control, 1980.
- [7] J. Lundén and S. Werner, "Real-time smart metering with reduced communication and bounded error," in *IEEE Int. Conf. Smart Grid Communications*, 2014, pp. 326–331.
- [8] S. Haykin, *Adaptive Filter Theory*. Pearson Education India, 2008.
- [9] "ATmega328P datasheet," Nov 2017. [Online]. Available: http://ww1.microchip.com/downloads/en/DeviceDoc/Atmel-42735-8-bit-AVR-Microcontroller-ATmega328-328P_Datasheet.pdf
- [10] "DS18B20, programmable resolution 1-wire digital thermometer," 2015. [Online]. Available: <https://datasheets.maximintegrated.com/en/ds/DS18B20.pdf>
- [11] "XBee/XBee-PRO S2C 802.15.4 radio frequency (RF) module," Feb 2018. [Online]. Available: <https://www.digi.com/resources/documentation/digidocs/pdfs/90001500.pdf>
- [12] "Keysight Technologies Digital Multimeters," May 2018. [Online]. Available: <https://literature.cdn.keysight.com/litweb/pdf/5991-1983EN.pdf>
- [13] "Keysight technologies Benchvue software 2017 (bv0000a)," March 2017. [Online]. Available: <http://literature.cdn.keysight.com/litweb/pdf/5991-3850EN.pdf>
- [14] G. Bag, Z. Pang, M. Johansson, X. Min, and S. Zhu, "Engineering friendly tool to estimate battery life of a wireless sensor node," *J. of Industrial Inf. Integration*, vol. 4, pp. 8–14, 2016.
- [15] F. Wu, C. Tan, M. Sarvi, C. Rudiger, and M. Yuce, "Design and implementation of a low-power wireless sensor network platform based on XBee," in *IEEE 85th Veh. Technol. Conf.*, 2017, pp. 1–5.