International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 03, March-2022

Improvement of Indoor Air Quality Data of Low-Cost IoT Sensors using Kalman Filter

AL-Khawlani Akram Saleh
School of Informatics,
Zhejiang Sci-Tech University, China,

Prof. Zheng Junbao
School of Informatics,
Zhejiang Sci-Tech University,China

Waleed M. Ismael
College of IoT Engineering,
Hohai University,Changzhou Campus, China,

Zaid Yemeni
College of IoT Engineering,
Hohai University,Changzhou Campus, China

***Abstract*—The rapid technological development of the mod- ern world brings a significant change to our life. Everythingsurrounds us has become smarter and makes our daily activ-ities more comfortable. Unfortunately, such rapidly developed technologies have caused negative consequences to humans. One of them is indoor air pollution, which leads to various issues,such as cardiovascular and respiratory diseases. Consequently, air quality status knowledge is crucial since it affects well-being lives. Therefore, it is imperative to monitor indoor air quality to measure the degree of danger to human health. IoT is widely used as an application for the deployment of air quality monitoring systems. When the data of indoor air quality are collected by IoT low-cost and inaccurate sensors, the collected data will be inaccurate and noisy and lead to decision errors. Therefore, the air quality data are required to be improved and noise-free. This paper employs Kalman filter to improve the indoor air quality data collected by IoT low-cost sensors. The proposed system is evaluated by the IoT paradigm to monitor indoor pollutants' concentration, such as temperature, humidity, Carbon Dioxide (CO2), Carbon Monoxide (CO), and Sulfur Dioxide (SO2). The results indicate that the proposed system, based on the IoT paradigm, can effectively and timely monitor indoor air quality.**

Index Terms*—*Internet of Things (IoT); Indoor Air Quality; Data Improvement; Kalman Filter.

I. INTRODUCTION

Due to the expansion of civilization and rapidly increase in the production of contaminated emissions from industrial process of raw materials and automobiles exhausts, atmo- spheric conditions keep deteriorating each year. In spite of the fact that air is an essential element for life, a lot of people are unconcerned about to the seriousness of air pollution or only recently realized the problem [1]. Air pollution is considered the most serious and most precarious cause of climate emergency and mortal diseases than the other types of pollutants such as noise, soil, water, and heat. As stated by the World Health Organization (WHO), the very high portion of the world population (90%) inhales contaminated air, and the death of seven million persons every year are attributedto the air pollution [2]. Moreover, As stated by the American Association for the Advancement of Science, Air pollution is now the world's crucial-leading fatal health risk factor [3] [4], leading to lung cancer, asthma attacks, chronic bronchitis, or heart disease. air pollution severely affects health, causing stroke, lung cancer, and heart disease. Consistent with the United States Environmental Protection Agency (EPA), the contamination of indoor air is about one hundred times higher than the outdoor air contamination [5]. Since the majority of people spend most of their time inside, the indoor air pollution has a superior negative effects on human health than outdoor air pollution. Therefore, robust indoor air quality monitoring plays a very vital role in preventing exposure through proactive and preventive measures. However, it requires high-precision monitoring system, which are often higher in terms of costs, thus, there is a trade-off between the accuracy of air quality data and the cost of monitoring system. Therefore, low-cost and efficient indoor air quality monitoring is indispensable to correctly and accurately manage air quality.

The indoor air quality relies on the various types of gases presented or generated and particulate matter of pollutants in semi-enclosed locations. There are diverse sources by which pollutants are generated, such as chemical, biological, toxic, and airborne particulates [6] [7] [8]. These pollutant sources can cause health issues, discomfort, or lead to death in some serious situations. Here, some examples are given. The level of CO2 can be used to specify the exchange rate of air. A high level of CO2 indicates a higher level of contaminants owing to a lack of fresh air circulation. Monitoring SO2 deficiency is also important indoor. Increasing in the its level can cause severe effects, such as respiratory issues, including bronchitis, nose, throat, and lung irritation. It has a bond with cardiovascular disease. Moreover, SO2 detection is indispensable in providing a safe and healthy working environment [9] [10]. Moreover, CO has severe effects on human life. Long exposure to CO causes vomiting, heart irregularity, brain damage, breathing difficulties, abortion, muscle, and even death. Moreover, monitoring indoor temperature and humidity is not less important than monitoring other indoor gasses. Indoor air temperature monitoring is a vital procedure providing helpful information in numerous fields, including human being thermal comfort, food preservation, and energy reservation [11]. The indoor relative humidity monitoring is also critical to the health of dwellers [12] [13]. A lowhumidity level is considered one of the leading causes of health problems, such as dry, bloody noses, cracked skin, chapped lips, and dry sinuses [14]. Besides, it can aggravate pre-existing health conditions, such as bronchitis, asthma. In contrast, high humidity helps create a suitable environment for many microorganisms to thrive, such as bacteria, dust mites, and mold, which would increase the effects of respiratoryissues and Legionnaires' disease [15] [16].

Air quality (AQ) is the most important in indoor environ ments. In fact, the prediction of AQ relies on the analysis of the collected data from monitoring sensors. The accuracy of collected data affects the accuracy of the decision-making process to a certain extent. The indoor dwellers should take appropriate protective measures when the monitored AQ is going above a certain predefined threshold. However, when the data of AQ are collected by the low-cost and inaccurate sensors, the produced data will be inaccurate and lead to decision errors. To obtain highly accurate AQ data, various AQ monitoring schemes utilize high-accuracy sensors. Conversely, high-accuracy sensors are often more expensive. Under relatively low-cost sensors, the collected data have various problems, such as noisy, missing , and inaccuratedata. Different methods have been there for cleaning data, estimating a system current state and predicting the futurestate. Among them is Kalman filter (KF). Over the past few decades, KF has been broadly exploited in different appli- cations, such as navigation and guidance application, radar tracking, floodgate control and multi-robot motion estimation,air pollution emission prediction etc. The KF algorithm works effectively on a series of measurements that are statistically noisy and/or accompanied with other inaccuracies; it produces the states of unknown variables that are likely to be more accurate than those that are single-measurement-based [17]. The algorithm of KF is based on two phases: prediction and update phases. In the prediction phase, it makes estimates of the current state variables, accompanied by uncertainties. When the result of the next measurement is obtained, the weighted average is used to update these estimates, giving more weight to the estimates with more certainty [18]. Inorder to improve indoor air quality data collected by low-cost sensors flexibly arranged throughout the interested area, this paper employs the KF algorithm in developing a method for improving indoor air quality data. The developed method can be applied on the collected AQ data before performing data analysis and visualization.

The contributions of this paper are represented in two folds, as follows:

- Proposing a solution for improving the air quality data collected by IoT low-cost sensors, removing the noisy data, and correcting incorrect data.
- Implementing, deploying, and testing the proposed solution on an IoT real-life scenario. The obtained results will be analyzed to evaluate its performance.

The remainder of this paper is organized as follows: Section II presents state-of-the-art approaches related to the proposed solution. Section III introduces the methods of the proposed solution. Section IV elaborates the evaluation of the proposed solution and discusses the obtained results. Finally, Section VI concludes this work.

II. Related work

In recent years, many researches have been conducted on improving the quality of data collected by low-cost sensors. Air quality data has also been given such attention for improv-ing its quality to accurately assess the pollutant concentrationsand air quality in a certain area.

Authors in [19] employed data fusion techniques and machine learning approaches to improve the data quality of IoT sensor in environmental monitoring. In their study, the factors affecting the data quality of low-cost IoT sensors are identified using three features selection approaches, including Forward Feature Selection (FFS), Forward Feature Selection (FFS) and Exhaustive Feature Selection (EFS. Then, the sensor data are fused with the identified factors and integrated into one equation used to calibrate the given sensors using Artificial Neural Networks (ANN) and Linear Regression (LR). Since the proposed solution is based on machine learning techniques, it suffers from computational overhead and requires more resources and much for performing training. In addition, the final fused data lacks interpretability of each pollutant. In [20], the authors proposed air quality monitoring and prediction system based on edge computing. The authors implemented the Kalnam filter for improving to reduce the dependency of IoT application on cloud computing. The designed system are used to monitor outdoor pollutants such as SO2, NO2 and PM2.5. The study is evaluated on outdoor air quality data. This limit the proposed solution to improve the outdoor air quality, which might be applicable to indoor air quality data.

To improve the data quality of IoT low-cost sensors, the authors in [21] introduced a data fusion method, known as multi-sensor space-time data fusion framework. The proposed method was build depending on the Optimum Linear Data Fusion theory. For spatial-temporal estimation, their method was also integrated with a multi-time step Kriging method. The proposed method showed an ability in improving the estimation of PM2.5 concentration in space and time. The proposed method was only evaluated on PM2.5 concentrations, so that it lack a comprehensive evaluation on the concentrations of other gasses. Other authors applied statistical methods to correct air quality data collected by low-cost sensors. The authors in [22] conducted an investigation on random and linear forest models for the purpose of correcting PM2.5 measurements collected by network of low-cost sensors of the Denver Department of Public Health and Environment (DDPHE) in comparison with measurements from U.S. Environmental Protection Agency Federal Equivalence Method (FEM) monitors. The prediction models that are based on statistical methods are only applicable on single feature and are not suitable for complex prediction process.

III. Methods

Data Quality Improvement Module (DQIM)

DQIM is proposed to improve the quality of data collected by low-cost sensors. It is based on the KF, which is mathematically described in Equation (1). Equation (1) describes the implementation of the proposed system.

$$x_t = H_t x_{t-1} + B_t u_t + w_t \tag{1}$$

where x_t represents the signal value, which is a combination of its prior value x_{t-1}, a control signal u_t , and a process noise w_t. u_t is the control signal, which is a particular external factor affecting the system. H_t represents a state transition matrix to model x_{t-1}, and B_t represents control matrix to model u_t. w_t

denotes the process noise at time t, representing the effects of influential external factors on the system. In the proposed model, the statistical characteristics of w_t are assumed to be normally distributed with a normal mean ($\mu \approx 0$), and a covariance matrix ($\rho(w_t)$) follows the normal distribution of Q_t , as given by Equation (2).

$$\rho(w_t) \sim N(0, Q_t) \quad (2)$$

So that the sensor measured value z_t to the real state at time t is mathematically represented by Equation (3).

$$z_t = H_t x_t + v_t \quad (3)$$

In Equation (3), the value of z_t whose accuracy is unknown is a linear combination of the signal value x_t and the measurement noise v_t. In the proposed model, the statistical characteristics of v_t are assumed to be normally distributed with a normal mean ($\mu \approx 0$), and a covariance matrix ($\rho(v_t)$) follows the normal distribution of R_t, as given by Equation (4).

$$\rho(v_t) \sim N(0, R_t) \quad (4)$$

where w_t and v_t are independent.

To obtain the system's optimal state estimate at time t, a linear combination of the system's optimal estimated state at time t-1 and the observation of the system state at time t. Since the proposed model is based on the KF, the state of the KF is denoted by $\hat{x}_{t|t}$, which is the estimated state of the system at time t, and $P_{t|t}$, a covariance matrix of the estimated state error at time t. The value of $P_{t|t}$ is used to measure the accuracy of the estimation.

The KF is represented in two phases: prediction and correction phases.

1) Time Update (Prediction): The time update phase isthe first step in the KF algorithm. The state estimate andestimate error covariance matrix at time t is obtained based on the optimal estimation and estimated error covariance matrix of the system state at time t-1. The following equations mathematically describe this step:

$$\hat{x}_{t|t-1} = H_t \hat{x}_{t-1|t-1} + B_t u_t \quad (5)$$

$$P_{t|t-1} = H_t P_{t|t-1} H_t^T + Q_t \quad (6)$$

where H_t the state transition matrix, B_t is the control input matrix, and Q_t is the covariance matrix of the process noise, respectively.

2) Measurement Update (Correction): To get the optimal state of the system at time t, $\hat{x}_{t|t-1}$ is required to be corrected by combining it with the current sensor observation. To achieve this objective, the correction phase is implemented. The following equations mathematically describe this step:

$$F_t = H_t P_{t|t-1} H_t^T + R_t \quad (7)$$

$$G_t = P_{t|t-1} H_t^T F_t^{-1} \quad (8)$$

$$\hat{z}_t = z_t - H_t \hat{x}_{t|t-1} \quad (9)$$

where H_t is the identity matrix, R_t is the covariance matrix of the observed noise, and G_t is Kalman gain. z_t is the sensor observation at time t, and $\hat{z}_t$ is the observation margin, calculated as the difference between z_t and $\hat{x}_{t|t-1}$ obtained by the a priori estimation. Based on the above equations, the values of $\hat{x}_{t|t-1}$ and $P_{t|t-1}$ are updated to obtain the optimal estimates of $\hat{x}_{t|t}$ and $P_{t|t}$ of the estimation error of the system state at time t.

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + G_t \hat{z}_t \quad (10)$$

$$P_{t|t} = (1 - G_t H_t) P_{t|t-1} \quad (11)$$

The following algorithm outlines the proposed module for improving the quality of sensor observation.

Algorithm 1: DQIM process.

Input: Sensors' readings
Output: Estimated data
1 Initialize the system.
2 Perform system training
3 **while** *true* **do**
4 Perform system evaluation
5 **if** *not reach the best values of the parameters* **then**
6 Continue
7 **else**
8 Break
9 Apply prediction.

IV. Experiments and Results

This section articulates the evaluation of the proposed system in indoor air quality (IAQ) monitoring application which architecture have been presented in Section III. Since the proposed system is a real-time system, the system's overall performance is evaluated based on collection latency.

A. Setup

After implementing the proposed system, the next step is to evaluate the proposed system's performance by deploying the system in real-life scenarios. The selected scenario for the deployment of the proposed system was the Dormitory of Friendship Building, Second Life Zone, Zhejiang Sci-Tech University, Xiasha District, Hangzhou City, Zhejiang Province. The experiments have been held primarily on the 14th floor, room No. 1405.

Based on the IoT paradigm, The proposed system is deployed in real-life scenarios to evaluate its performance. The following sections are dedicated to describing the architecture of the proposed system.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 03, March-2022

1) The Proposed System Architecture: The overall architecture of the proposed indoor air quality monitoring system is shown in Fig. 1. The system can simultaneously collect AQ data from several sensors and transmit them an IoT server to be made available to remote users both in graphical and tabular forms. The system components are described as follows:

- Hardware Modules
- Cloud Platform
- Mobile App

a) Hardware Modules: The hardware module consists of the sensors and the controller board. The sensors are used to sense indoor air quality data. The controller mainly is to collect data from given sensors and propagate the collected data via the Internet (TCP/IP) to the cloud platform for online visualization and storage. Finally, the end-user can access the stored data using a mobile App developed for real-time monitoring. In the proposed system, three different MQ sensors (MQ- 4, MQ-136, MQ-7) are used to measure CO2, SO2, CO, respectively. For the total Air Quality Index, we use the MQ-135 sensor, which is sensitive to a range of gases, along with DHT sensors for temperature and relative humidity. We used Arduino Ethernet Shield and Arduino Leonardo board. The Arduino Ethernet Shield is used to enable connecting an the Arduino Leonardo board to the Internet [22]. The Arduino Leonardo Board is characterized by 20 digital I/O pins, 12 pins of which as analog inputs. Its clock speed is up to 16 MHz. It has a power jack, USB connection, and an ICSP header.

b) IoT Cloud Platform: We utilized the Thinger.io platform for our proposed system, an open-source platform dedicated to providing an attainable scalable infrastructure for linking things. The cloud platform facilitates the definition of numerous resources for the sensors used by the proposed system that can sense for making our device accessible over the Internet. Also, it facilitates the interaction with our devices defined using API functionalities. Thus, the input sensors' readings are timely uploaded to the cloud platform. Fig. 2 show online visualization of the collected data from the CO2, SO2, CO, and Temperature and Relative Humidity Sensors used for the evaluation of the proposed system.

c) Mobile Application: The proposed system's objective is to enable the end-user to remotely monitor data produced by indoor sensors using handheld devices to take protective measures for time-sensitive emergencies happening indoors. The mobile app is implemented using Cordova framework.

B. Experimental Dataset

Air quality data employed in this paper were collected by the sensors used in the design of the proposed system from May 1, 2021, to June 30, 2021. The whole data containsdifferent air quality datasets: CO, CO2, SO2, Temperature, and relative humidity datasets. The gasses concentrations are mea-sured in reference to [34]. CO2, CO, and SO2 were measured by PPM while the temperature was measured by Celcius, and Relative Humidity was measured in percentage. The monthly average concentrations for CO2, CO, SO2, all together with Temperature and Relative Humidity, were calculated using the hourly and daily average concentrations. Datasets are supplied in CSV files that are transformed into big datasets, which are compiled in tables, and the related outlier-free data were extracted and mathematically manipulated. The sampling rate of the collected data sets is fixed to 30 seconds. The number of samples used in this analysis is 4550 measurements for each gas.

C. Result Analysis

In this section, the result analysis is performed based on the system performance the collected datasets for gasses: CO, CO2, SO2, along with temperature and relative humidity.

1) System Performance: In this section, different experiments with N gasses sensors were conducted varied from 1 to 5 sensors connected to Arduino board to calculate the difference of data production time and the data visualization time to the end-user.

a) Collection Latency: The performance of the proposed system is estimated by measuring the time of data collection from N sensors and the data visualization on the developed mobile App to the end-user. Furthermore, the connection time between the Arduino board and the cloud platform, the connection time between the developed mobile App and the cloud platform, and receiving the data from all IoT devices to be visualized on the mobile App is known as collection latency. The experiments were conducted using Samsung Galaxy 8 and iPhone 11pro max. The connection time taken to establish a connection between the Arduino board and the cloud platform is approximately 15 seconds. The collection latency is calculated by Equation (1).

$$CL = DR - DP \qquad (12)$$

where CL is the collection latency. DR is the time of receiving data by the developed mobile App, and DP is the time of data production by N sensors.

Two experiments were conducted to measure the collection latency. In the first experiment, the collection latency is measured while the data propagated to the cloud platform sequentially for N sensors. However, in the second experiment, the collection latency is measured while data are propagated simultaneously. The average values of a sequential connection are depicted in Fig. 3. As it can be seen, both Samsung Galaxy 8 and iPhone 11pro max show the same trend. Still,the connection established by iPhone 11 Pro max takes ashorter time than Samsung Galaxy 8 owing to its relativelylow capacity. As observed in Fig. 4, the collection latency in simultaneous connection is more considerable and faster than the sequential connection.

Finally, the collection latency measurement in both sequential and simultaneous connection for both Samsung Galaxy 8 and iPhone 11pro max is subject to the device's performance. Therefore, it is recommended the devices with higher capacity are preferable. We assume that there is no data loss in the experiments, and the network bandwidth is fixed.

2) DQIM Performance: One of the critical points for evaluating the trustability of DQIM is the value of Q , which means that if its value is small, it indicates the process noiseof the system is less and trustable. However, if the value of

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 03, March-2022

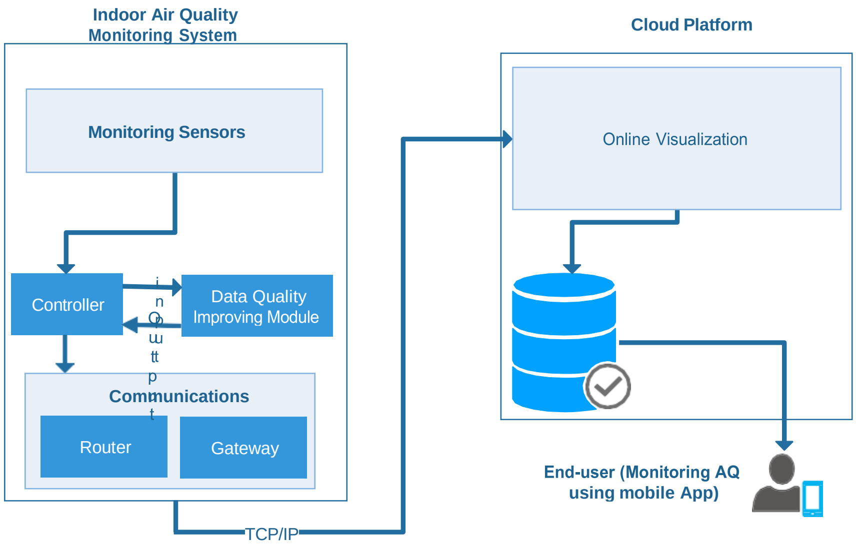

Fig. 1: Proposed system architecture.

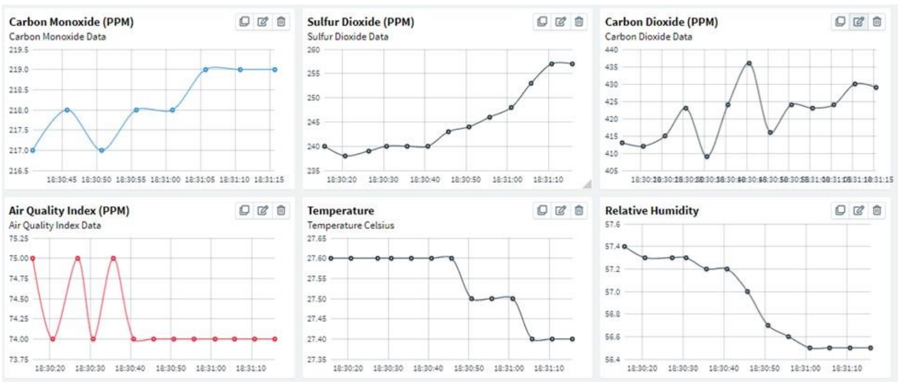

Fig. 2: Dashboard of the cloud platform.

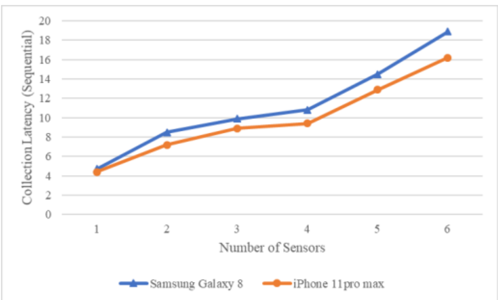

Fig. 3: Collection latency for sequential connection.

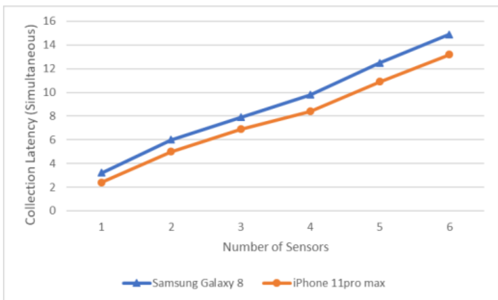

Fig. 4: Collection latency for simultaneous connection.

Q is high, it indicates that the process noise of the system is high, and the prediction system becomes untrustable. When the value of Q is zero, the estimated value x_t is more trustable than the sensors' observations. However, when the value of is one, the sensors' observations are more trustable than the estimated value x_t. In this paper, we applied the tuning

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 03, March-2022

TABLE I: Statistical summary statistics of CO2, CO, SO2, temperature, and Relative Humidity concentrations.

	CO2	CO	SO2	Temperature	Relative Humidity
Samples Count	1183	1183	1183	1183	1183
Standard Dev.	123.2365	95.731961	36.56394	1.664489792	5.181274747
Average	297.3982	167.8639	208.7922	27.71851395	72.06706385
Coeff. of variation	41.43822	57.029512	17.51212	6.004974852	7.189518305
Minimum	173	116	133.75	22.20000076	46.25
Maximum	922.75	955.25	321.5	30.35000038	80.10000229
Stnd. skewness	3.853599	6.7418114	-0.0368	-0.844601963	-2.138055498
Stnd. kurtosis	15.88695	49.932513	-0.40763	-0.240041953	6.987424644

TABLE II: Pearson's correlation coefficients matrix for CO2, CO, CO, Temperature, and Relative Humidity.

	CO2	CO	SO2	Temperature	Relative Humidity
CO2	–	-0.266392	0.210426	0.198774903	0.047736248
	–	p=6.59214799479572E-06	p=7.60741418093521E-13	p=1.3535224271778E-11	p=5.25381476661487E-12
CO	-0.26639	–	0.261419	-0.159277328	0.019543338
	p=6.59214799479572E-06	–	p=3.20287266558599E-19	p=6.68876076769264E-08	p=0.107663193847883
SO2	0.210426	0.2614193	–	-0.051714505	0.688097868
	p=7.60741418093521E-13	p=3.20287266558599E-19	–	p=0.0813280256625801	p=0.510327222058107
Temperature	0.198775	-0.159277	-0.05171	–	0.688097868
	p=1.3535224271778E-11	p=6.68876076769264E-08	p=0.0813280256625801	–	p=2.75316830192132E-160
Relative Humidity	0.047736	0.0195433	0.688098	0.688097868	–
	p=5.25381476661487E-12	p=0.107663193847883	p=0.510327222058107	p=2.75316830192132E-160	–

model proposed by the authors in [23] for tuning the Q , in which the predictive model reaches the best point of process convergence. During the tuning process, RMSD (root mean square deviation) between the predicted value $\hat{x}$ produced by the KF and its corresponding actual value x collected by sensors, as given by Equation (13).

$$RMSD(x,\hat{x}) = \sqrt{\frac{\sum_{i=0}^{N}(x-\hat{x})}{N}} \tag{13}$$

During the tuning process, when the state of the KF is non-converged, the RMSD will be relatively large between the predicted value $\hat{x}$ and the actual value x ; the value is rather large. Therefore, the RMSDs for CO2, CO, SO2, all together with temperature and relative humidity values, respectively, are normalized by the min-max normalization method ranging between (0,1). In this paper, we used the value of RMSD to select the best optimal value of Q (global minimum). The smaller value of RMSD indicates that the value of $\hat{x}$ will be less unconverted, indicting faster convergence of the model. Accordingly, the value of Q will be more suitable for estimating more accurate values of CO2, CO, SO2, temperature, and relative humidity. The best optimal values of Q obtained from the experiment are 0.563, 0.079,0.453, 0.172, 0.144 for CO2, CO, SO2, temperature, and relative humidity, respectively.

Figures 5-9 give an illustrative demonstration to the obtain results of DQIM after its application on a portion of the collected data of SO2, CO2, CO, Temp, and Relative Humidity.

3) Descriptive statistics and Correlation Analysis: Descriptive statistics of CO2, CO, SO2, Temperature, and Relative Humidity in the studied area have been carried out. Table 1 lists a statistical summary of these gasses, including the measurements of variability, central tendency, and form. From Table 1, we can observe that the maximum monthly averages for CO2, CO, and SO2 are 297.398197 ppm, 167.8638962 ppm, and 208.7922164 ppm, respectively. These values exceeded the monthly standard average threshold of 100 ppm for

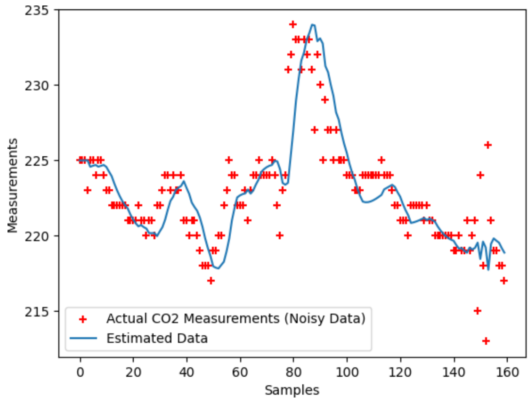

Fig. 5: The output of DQIM for CO2.

human health. Furthermore, the highest standard deviation was found for CO2, and the highest variation coefficient was found for CO. These values indicate heterogeneous concentrations of CO2 and CO. However, SO2 showed a slight deviation and variation. Moreover, CO2 and CO had shown higher skewness of their probability distributions. Moreover, the averages of Temperature and Relative Humidity are 27.71851395°C and 72.06706385%, respectively. These indicate that Temperature is moderate and Relative Humidity id very poor.

Correlations analysis enables the investigation of chemical or environmental relations among the gasses concerned inthis research. It reflects the relationship possibility among the gasses resources. Table 2 shows the analysis results using Pearson's correlation coefficients with p values$\leq$ 0.05. CO2 shows a negative correlation with CO (-0.266392) and a positive and insignificant correlation with SO2 (0.210426). This suggests that there is no good associations or common source among the gasses. Furthermore, CO2 and CO show insignificant correlation with Temperature and Relative Humidity, while

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 03, March-2022

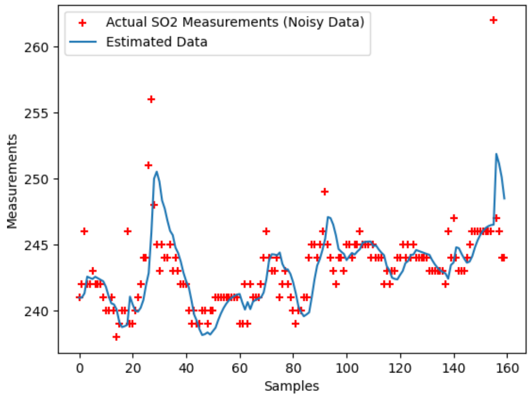

Fig. 6: The output of DQIM for SO2.

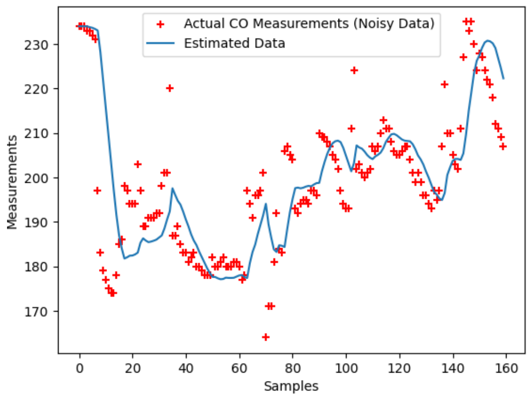

Fig. 7: The output of DQIM for CO.

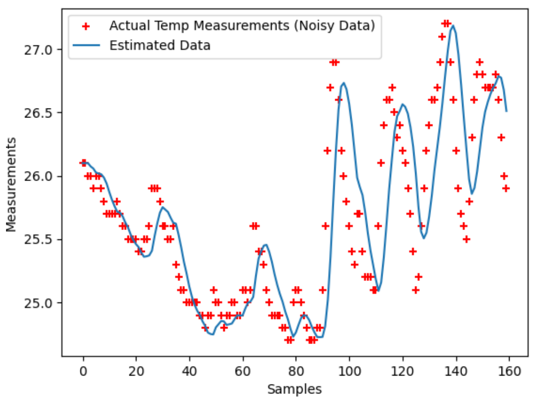

Fig. 8: The output of DQIM for Temperature.

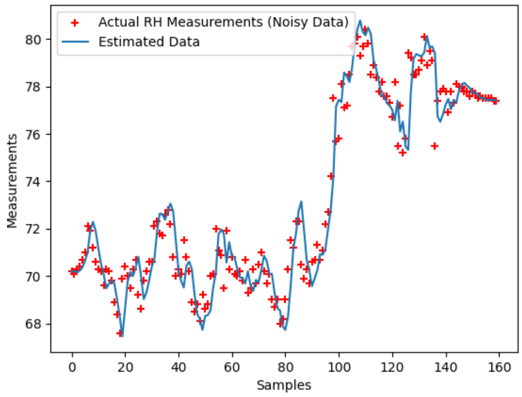

Fig. 9: The output of DQIM for RH.

SO2 shows positive and significant correlation with Relative Humidity. This indicates that increasing in Relative Humidity helps increasing in SO2 concentration.

V. Conclusion

This paper focuses on developing an IoT-based system to efficiently collect data from sensing IoT devices using a set of sensors connected to the Arduino board to propagate the collected data to the cloud for storage and analysis. A mobile app also is developed to help the end-users to visualize the collected data in real-time, anywhere, and anytime. For this purpose, it selects the most cost-efficient devices to collect data. Data analysis services are developed for the collected data to measure the levels of indoor gasses and inform the end-users about their effect on human health. The proposed system encompasses three layers: hardware, cloud, and mobile layers. In the hardware layer, we developed our perception layer consisting of a set of sensors for data collection connected to a controller (Arduino board). The main task of the controller is to propagate the collected data from the sensors to thecloud. The second layer is the cloud, which is responsiblefor providing data analysis and data storage. The third layer is the mobile App, which is responsible for collecting data from the cloud and visualizing it to the end-users in real-time.

The proposed system was implemented using a prototype and evaluated based on the data collection and visualization latency using different environments and different connec-tion modes (Sequential and simultaneous connections). The proposed solution is based on the Kalman filter, which is characterized by fast convergence and higher estimation ac-curacy. This strengthens the performance and the accuracy of our proposed solution. The obtained results showed that the proposed system highlights good performance in improving the indoor air quality data through cleaning the noisy dataand correcting incorrect data.

References

[1] G. Parmar, S. Lakhani, and M. K. Chattopadhyay, “An iot based low cost air pollution monitoring system,” in *2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE)*, pp. 524–528, IEEE, 2017.

[2] A. P. WHO, "Child health: Prescribing clean air," *World Healt h Organization, Geneva*, 2018.
[3] N. H. Motlagh, M. A. Zaidan, E. Lagerspetz, S. Varjonen, J. Toivonen, J. Mineraud, A. Rebeiro-Hargrave, M. Siekkinen, T. Hussein, P. Nurmi, *et al.*, "Indoor air quality monitoring using infrastructure-based motion detectors," in *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, vol. 1, pp. 902–907, IEEE, 2019.
[4] S. Mad Saad, A. M. Andrew, A. Y. Md Shakaff, M. A. Mat Dzahir, M. Hussein, M. Mohamad, and Z. A. Ahmad, "Pollutant recognition based on supervised machine learning for indoor air quality monitoring systems," *Applied Sciences*, vol. 7, no. 8, p. 823, 2017.
[5] U. S. E. P. Agency, "Managing air quality - air pollutant types," *https://www.epa.gov/air-quality-management-process/managing- air-quality-air-pollutant-types.*, 2018.
[6] X. Yang, L. Yang, and J. Zhang, "A wifi-enabled indoor air quality monitoring and control system: The design and control experiments," in *2017 13th IEEE International Conference on Control & Automation (ICCA)*, pp. 927–932, IEEE, 2017.
[7] V. V. Tran, D. Park, and Y.-C. Lee, "Indoor air pollution, related human diseases, and recent trends in the control and improvement of indoor air quality," *International journal of environmental research and public health*, vol. 17, no. 8, p. 2927, 2020.
[8] "A guide to air quality and your health - us environmental protection agency office of air quality planning and standards outreach and information division research triangle park, nc.." https://www3.epa.gov/airnow/aqi_brochure_02_14.pdf. Accessed: 2021-01-07.
[9] A. Andrade and F. H. Dominski, "Indoor air quality of environments used for physical exercise and sports practice: Systematic review," *Journal of environmental management*, vol. 206, pp. 577–586, 2018.
[10] U. Jaimini, T. Banerjee, W. Romine, K. Thirunarayan, A. Sheth, and M. Kalra, "Investigation of an indoor air quality sensor for asthma management in children," *IEEE sensors letters*, vol. 1, no. 2, pp. 1–4, 2017.
[11] F. Lachhab, M. Bakhouya, R. Ouladsine, and M. Essaaidi, "Monitoring and controlling buildings indoor air quality using wsn-based technologies," in *2017 4th International Conference on Control, Decision and Information Technologies (CoDIT)*, pp. 0696–0701, IEEE, 2017.
[12] À. Boso, B. Álvarez, C. Oltra, J. Garrido, C. Muñoz, and Á. Hofflinger, "Out of sight, out of mind: participatory sensing for monitoring indoor air quality," *Environmental monitoring and assessment*, vol. 192, no. 2, pp. 1–15, 2020.
[13] A. Cincinelli and T. Martellini, "Indoor air quality and health," 2017.
[14] S. Zanni, F. Lalli, E. Foschi, A. Bonoli, and L. Mantecchini, "Indoor air quality real-time monitoring in airport terminal areas: An opportunity for sustainable management of micro-climatic parameters," *Sensors*, vol. 18, no. 11, p. 3798, 2018.
[15] M. Gola, G. Settimo, and S. Capolongo, "Indoor air in healing environments: Monitoring chemical pollution in inpatient rooms," 2019.
[16] R. Mumtaz, S. M. H. Zaidi, M. Z. Shakir, U. Shafi, M. M. Malik, A. Haque, S. Mumtaz, and S. A. R. Zaidi, "Internet of things (iot) based indoor air quality sensing and predictive analytic—a covid-19 perspective," *Electronics*, vol. 10, no. 2, p. 184, 2021.
[17] Y. V. Oktaviana, E. Apriliani, and D. K. Arif, "Fractional kalman filter to estimate the concentration of air pollution," in *Journal of Physics: Conference Series*, vol. 1008, p. 012008, IOP Publishing, 2018.
[18] D. K. Arif, H. N. Fadhilah, and P. Aditya, "Estimation of air pollutant transportation equation in surabaya using kalman filter method," *IJCSAM (International Journal of Computing Science and Applied Mathematics)*, vol. 6, no. 1, pp. 18–22, 2020.
[19] N. U. Okafor, Y. Alghorani, and D. T. Delaney, "Improving data quality of low-cost iot sensors in environmental monitoring networks using data fusion and machine learning approach," *ICT Express*, vol. 6, no. 3, pp. 220–228, 2020.
[20] X. Lai, T. Yang, Z. Wang, and P. Chen, "Iot implementation of kalman filter to improve accuracy of air quality monitoring and prediction," *Applied Sciences*, vol. 9, no. 9, p. 1831, 2019.
[21] Y.-C. Lin, W.-J. Chi, and Y.-Q. Lin, "The improvement of spatial-temporal resolution of pm2. 5 estimation based on micro-air quality sensors by using data fusion technique," *Environment international*,vol. 134, p. 105305, 2020.
[22] E. M. Considine, C. E. Reid, M. R. Ogletree, and T. Dye, "Improving accuracy of air pollution exposure measurements: Statistical correction of a municipal low-cost airborne particulate matter sensor network," *Environmental Pollution*, vol. 268, p. 115833, 2021.
[23] Z. Chen, N. Ahmed, S. Julier, and C. Heckman, "Kalman filter tuning with bayesian optimization," *arXiv preprint arXiv:1912.08601*, 2019.